BAYESIAN SMOOTHING SPLINE WITH DEPENDENCY MODELS

A Dissertation presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

BENQIAN ZHANG

Dr. Zhuoqiong He and Dr. Dongchu Sun, Dissertation Supervisors

May 2021

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

BAYESIAN SMOOTHING SPLINE WITH DEPENDENCY MODELS

presented by Benqian Zhang,
a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Zhuoqiong He

Dr. Dongchu Sun

Dr. (Tony) Jianguo Sun

Dr. Sounak Chakraborty

Dr. X.H. Wang

ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere gratitude to my advisors, Dr. Dongchu Sun and Dr. Zhuoqiong He. Without their guidance, inspiration and patience, this project would have never been made possible.

I would also like to thank my committee members Dr. Tony Sun, Dr. Sounak Chakraborty, and Dr. X.H. Wang for their constructive suggestions and comments on my research.

In addition, I want to thank other faculty members in the Statistics Department who taught me, helped me, and worked with me, and I want to thank Judy, Kathleen and Abbie for their help.

Most importantly, I am indebted to my family, especially my parents, for their love, understanding, and moral support through my study. It is to them that I dedicate this work.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

BAYESIAN SMOOTHING SPLINE WITH DEPENDENCY MODELS

Benqian Zhang

Dr. Dongchu Sun and Dr. Zhuoqiong He, Dissertation Supervisors

ABSTRACT

The smoothing spline model is widely used for fitting a smooth curve because of its flexibility and smoothing properties. Our study is motivated by estimating the long-term trend of the U.S. unemployment level. In this dissertation, a class of Bayesian smoothing spline with dependency models is developed.

The unemployment level and other labor-force time series, which are often used to analyze market and economic conditions, are strongly influenced by seasonality, as well as irregular or short-term fluctuations. We apply the basic Bayesian smoothing spline model to obtain the smooth estimation of the trend from a time series, which captures the fundamental tendency of general economic expansions and contractions. We further generalize the basic Bayesian smoothing spline model with dependence structure. This generalization significantly improves the boundary performance and elevates the overall accuracy and precision by borrowing information from different cycles. Finally, we construct the multivariate Bayesian smoothing spline with dependency model, which enables us to estimate the trends of the unemployment and employment level simultaneously. The accuracy and precision are improved by the joint model.

Chapter 1

Introduction

The smoothing spline model is widely used for fitting a smooth curve because of its flexibility and smoothing properties. Given the time series data $\boldsymbol{y}$ with time $\boldsymbol{t} = (t_1, \ldots, t_n)'$, the smoothing spline model can be used to enhance the visual appearance of the scatterplot of $\boldsymbol{y} = (y_1, \ldots, y_n)'$ vs $\boldsymbol{t}$, and to estimate the trend in the plot objectively.

The univariate smoothing spline problems have been thoroughly studied from the frequentist perspective. Consider the univariate nonparametric regression model,

$$y_i = f(t_i) + \epsilon_i, \ i = 1, \ldots, n, \tag{1.1}$$

where $\epsilon_i \overset{\text{iid}}{\sim} N(0, \delta_0)$, $f(t)$ is an unknown smooth function of time t, and $a \leq t_1 < \cdots < t_n \leq b$. The unknown smooth function $f(t)$ could be estimated by minimizing the penalized least squares,

$$\sum_{i=1}^{n} [y_i - f(t_i)]^2 + \eta \int_a^b [f^{(k)}(t)]^2 dt, \tag{1.2}$$

where $f^{(k)}(t) = \frac{d^k}{dt^k} f(t)$, and $\eta > 0$ is a smoothing parameter that balances the trade-off between goodness-of-fit and smoothness of the estimate. The cross-validation method is often used to choose the smoothing parameter η. See, for example, Hastie and Tibshirani (1990) and Gu (2013) for a comprehensive review. From the Bayesian perspective, certain forms of smoothing spline correspond to Bayesian estimates under a class of improper Gaussian prior distributions on function spaces, which is fully explained in Wahba (1990). Sun, Tsutakawa, and Speckman (1999) first introduced the Partially Informative Normal (PIN) prior, which lays the foundation for this work. See Speckman and Sun (2003) and Cheng and Speckman (2012) for full Bayesian analysis.

In Chapter 2, we introduce a motivating example to estimate the long-term trend of the U.S. unemployment level. We apply a Bayesian smoothing spline model to obtain the smooth estimation of the trend from a time series to capture the fundamental tendency of general economic expansions and contractions.

The Current Population Survey (CPS), conducted jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS), provides numerous labor force statistics for the U.S. population. For example, the unemployment level counts all persons who had no employment with certain criteria, which is formally defined in Current Population Survey Technical Paper 77.

Those labor-force time series, which are often used to analyze market and economic conditions, however, are strongly influenced by seasonality, as well as irregular or short-term fluctuations. Therefore we want to estimate the trend, which is the long term movement in a time series without periodic related and irregular effects, as a reflection of the underlying tendency.

Traditionally, people decompose a time series into a trend, a seasonal component, and an irregular component. In order to obtain the seasonal effect and the long-term

trend of the unemployment level and other labor force time series, the original X-11 software was developed by the US Census Bureau in the 1960s, and later improved by Statistics Canada (Dagum 1980). Subsequent software packages by the US Census Bureau were called X-12-ARIMA (Findley et al. 1998) and X-13ARIMA-SEATS (or X-13, for short) (Monsell 2007).

The most recent version X-13 offers two seasonal adjustment methods: X-11, which is used as a name for filter-based seasonal adjustment methods, and SEATS (signal extraction in ARIMA time series). Details can be found from Ladiray and Quenneville (2012), and Dagum and Bianconcini (2016). We compare our result with the current X-13 method.

In Chapter 3, we further generalize the Bayesian smoothing spline model with dependence structure. We design a correlation matrix, see Appendix 3.10, as the dependence structure for the seasonal pattern. This generalization significantly improves the boundary performance and elevates the overall accuracy and precision by borrowing information from different cycles. We develop an efficient MC sampling for the posterior distribution with the bivariate ratio-of-uniform method.

In Chapter 4, we construct the multivariate Bayesian smoothing spline with dependency model, which enables us to estimate the trends of the unemployment and employment levels simultaneously. We are interested in the trend estimates of multiple time series with the same period, which may or may not share a similar seasonal pattern. We further improve the accuracy and precision with the joint model by borrowing information across multiple series.

Chapter 2

Estimating the Long-term Trend of Unemployment Level Using Bayesian Smoothing Spline Model

2.1 Introduction

We often see economic time series like the unemployment rate being strongly influenced by seasonality—periodic fluctuations associated with recurring calendar-related events such as weather, holidays, and the opening and closing of schools. Traditionally, people decompose a time series into a trend, a seasonal and an irregular component. In order to obtain the seasonal effect and the long-term trend of the unemployment level and other labor force time series, the original X-11 software was developed by the US Census Bureau in the 1960s, and later improved by Statistics Canada (Dagum 1980). Subsequent software packages by the US Census Bureau were called X-12-ARIMA (Findley et al. 1998) and X-13ARIMA-SEATS (or X-13, for short) (Monsell 2007).

The most recent version X-13 offers two seasonal adjustment methods: X-11, which is used as a name for filter-based seasonal adjustment methods, and SEATS (signal extraction in ARIMA time series). Details can be found from Ladiray and Quenneville (2012), and Dagum and Bianconcini (2016). X-13 procedures are based on "filters" that successively average a shifting timespan of data, thereby providing estimates of seasonal factors that change in a smooth fashion from year to year. To seasonally adjust recent data, shorter filters with less desirable properties must be used. Every time an observation is added, previous estimates will be revised. A final adjustment may wait up to 5 years when standard options are used.

In this chapter, we propose a Bayesian smoothing spline model to estimate the long-term trend of the U.S. unemployment level time series. In Section 2.2, we introduce the model construction and computation. In Section 2.3, we show the Bayesian smoothing spline trend estimation for our motivating application, compared with the X-13 result. In Section 2.4.1, we discuss the unstable boundaries and improvement procedures. In Section 2.4.2, we study the effect of newly added data to previous trend estimates. We compare the stability of our Bayesian smoothing spline trend estimates with the X-13 result, when new data are available. In Section 2.5, we estimate the changing points, where the trend switches from increasing to decreasing, or vice versa, as well as the uncertainty of the estimated changing points. In Section 2.6, we perform a simulation study.

2.2 Bayesian Smoothing Spline Model

2.2.1 Smoothing Spline Model

We consider the univariate nonparametric regression model in Speckman and Sun (2003)

$$y_i = z_i + \epsilon_i,\ z_i = f(t_i),\ i = 1, \ldots, n, \tag{2.1}$$

where $\epsilon_i \overset{\text{iid}}{\sim} N(0, \delta_0)$, $f(t)$ is an unknown smooth function of time t, and $a \leq t_1 < \cdots < t_n \leq b$. The unknown smooth function $f(t)$ is estimated by minimizing the penalized least squares

$$\sum_{i=1}^{n} [y_i - f(t_i)]^2 + \eta \int_a^b [f^{(k)}(t)]^2 dt, \tag{2.2}$$

where $f^{(k)}(t) = \frac{d^k}{dt^k} f(t)$, and $\eta > 0$ is a smoothing parameter that balances the trade-off between goodness-of-fit and smoothness of the estimate.

When $k = 2$, equation (2.1) is the natural cubic smoothing spline. We consider equally spaced knots at $t = 1, 2, ..., n$, and let $\boldsymbol{Q} = \boldsymbol{F}_0' \boldsymbol{F}_1^{-1} \boldsymbol{F}_0$, where the $(n-2) \times n$ matrix $\boldsymbol{F}_0$ and $(n-2) \times (n-2)$ matrix $\boldsymbol{F}_1$ are

$$\boldsymbol{F}_0 = \begin{pmatrix} 1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1 \end{pmatrix}, \quad \boldsymbol{F}_1 = \frac{1}{6} \begin{pmatrix} 4 & 1 & \ldots & 0 & 0 \\ 1 & 4 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 4 \end{pmatrix}. \tag{2.3}$$

Let $\boldsymbol{y} = (y_1, ..., y_n)'$, $\boldsymbol{z} = (z_1, ..., z_n)'$, and $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)'$. We rewrite the model (2.1)

as

$$\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \delta_0 \boldsymbol{I}_n). \tag{2.4}$$

The solution to the minimization (2.2) is

$$\boldsymbol{\mu}_z = (\boldsymbol{I}_n + \eta \boldsymbol{Q})^{-1} \boldsymbol{y}. \tag{2.5}$$

One major difficulty of the smoothing spline method is how to estimate the smoothing parameter η. Wahba (1990) explains the generalized cross validation (GCV) method and its limits. The GCV method is not reliable with small sample sizes. Even for large sample sizes, unwanted extreme estimates ($\hat{\eta} = 0$ *or* $\hat{\eta} = \infty$) may still occur.

Following Speckman and Sun (2003) and Cheng and Speckman (2012), we consider the full Bayesian analysis of the smoothing spline model.

2.2.2 Bayesian Analysis and Computation

We apply the Partially informative Normal prior in Sun, Tsutakawa, and Speckman (1999) for $\boldsymbol{z} = (z_1, ..., z_n)'$,

$$p(\boldsymbol{z}|\delta_1) \propto \big(\frac{1}{\delta_1}\big)^{\frac{n-2}{2}} exp\big(-\frac{1}{2\delta_1}\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{z}\big). \tag{2.6}$$

Here $\delta_1 = \delta_0/\eta$ for δ_0 and η given in (2.1) and (2.2). For the following computation, we re-parameterize the variance component as $\delta_1 = \delta_0/\eta$, where $\eta = \delta_0/\delta_1$ represents the noise-signal ratio.

We choose the Jeffreys prior for δ_0,

$$p(\delta_0) \propto \frac{1}{\delta_0}, \tag{2.7}$$

and the scaled Pareto prior (Cheng and Speckman 2012) for η,

$$p(\eta) = \frac{c}{(c+\eta)^2}, \eta > 0. \tag{2.8}$$

Its median is the scale parameter c. As c increases, the prior tends flatter and less concentrated to 0. For the nonparametric model (2.4) with prior densities (2.6), (2.7) and (2.8), we have the joint posterior for $(\boldsymbol{z}, \delta_0, \eta)$,

$$\begin{aligned} p(\boldsymbol{z}, \delta_0, \eta|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{z}, \delta_0, \eta)p(\boldsymbol{z}|\delta_0, \eta)p(\delta_0)p(\eta) \\ &\propto \frac{\eta^{\frac{n-2}{2}}}{\delta_0^n}\frac{c}{(c+\eta)^2}exp\big[-\frac{1}{2\delta_0}(\boldsymbol{z}-\boldsymbol{\mu}_z)'(\boldsymbol{I}_n+\eta\boldsymbol{Q})(\boldsymbol{z}-\boldsymbol{\mu}_z)\big] \\ &\quad exp\{-\frac{1}{2\delta_0}[\boldsymbol{y}'\boldsymbol{y}-\boldsymbol{\mu}_z'(\boldsymbol{I}_n+\eta\boldsymbol{Q})\boldsymbol{\mu}_z]\}, \end{aligned} \tag{2.9}$$

where $\boldsymbol{\mu}_z = (\boldsymbol{I}_n + \eta\boldsymbol{Q})^{-1}\boldsymbol{y}$, which is exactly the solution (2.4). Therefore we have the fact as follows.

Fact 2.2.1. *From the joint posterior density (2.9), we have the following results:*

(a) The conditional distribution of $\boldsymbol{z}$ *given* $(\delta_0, \eta; \boldsymbol{y})$ *is*

$$(\boldsymbol{z}|\delta_0, \eta; \boldsymbol{y}) \sim N_n((\boldsymbol{I}_n + \eta\boldsymbol{Q})^{-1}\boldsymbol{y}, \delta_0(\boldsymbol{I}_n + \eta\boldsymbol{Q})^{-1}). \tag{2.10}$$

(b) The conditional distribution of δ_0 *given* $(\eta; \boldsymbol{y})$ *is*

$$(\delta_0|\eta; \boldsymbol{y}) \sim IG(\frac{n}{2} - 1, \frac{1}{2}\eta\boldsymbol{y}'(\boldsymbol{I}_n + \eta\boldsymbol{Q})^{-1}\boldsymbol{Q}\boldsymbol{y}). \tag{2.11}$$

(c) The marginal posterior density of η given data $\boldsymbol{y}$ is

$$p(\eta|\boldsymbol{y}) \propto \frac{\eta^{\frac{n-2}{2}}}{|\boldsymbol{I}_n + \eta\boldsymbol{Q}|^{\frac{1}{2}}[\eta\boldsymbol{y}'(\boldsymbol{I}_n + \eta\boldsymbol{Q})^{-1}\boldsymbol{Q}\boldsymbol{y}]^{\frac{n}{2}-1}} \frac{c}{(c+\eta)^2}. \tag{2.12}$$

The proof is omitted as this is a special case in Tong, He, and Sun (2018), which has no replicates at each knot.

Based on the Fact 2.2.1, the Monte Carlo algorithm is as follows:

Algorithm 1 Posterior sampling of density (2.9)

1: Sample η given data $\boldsymbol{y}$ from density (2.12).

2: Sample δ_0 given $(\eta; \boldsymbol{y})$ from distribution (2.11).

3: Sample $\boldsymbol{z}$ given $(\delta_0, \eta; \boldsymbol{y})$ from distribution (2.10).

For the choice of hyper-parameter c, we follow Cheng and Speckman (2012) that c can be chosen based on a desirable prior degrees of freedom, which is defined as the trace of the smoother matrix $d(\eta) = tr((\boldsymbol{I}_n + \eta\boldsymbol{Q})^{-1})$. Here we take $c = 10$, and MC result is pretty robust with different choice of c.

2.3 The Long-term Trend of Unemployment Level

We use the public data of the U.S. unadjusted unemployment level from January of 2004 to December of 2018. The length of the time series is 180 months. Following the algorithm above, we draw 10,000 samples from the joint posterior distribution.

We show the posterior mean of $\boldsymbol{z}$ as the Bayesian smoothing spline trend estimates in Figure 2.3. Compared with seasonal adjusted and trend estimates from X-13, our trend estimation thoroughly removes the short-term fluctuations, and is considered a better trend estimation in terms of smoothness.

To access the uncertainty of our trend estimates, we show the 95% credible interval of the trend estimates in Figure 2.1, and the posterior coefficient of variation in Figure 2.2. We note that the uncertainty increases sharply as time approaches the two boundaries. We will discuss this phenomenon in the following section.

2.4 Stability

2.4.1 Boundary Performance and Improvement

We have noticed that the estimates around the boundary are not stable. A simple solution is to discard the two undesirable boundaries; i.e. if we are interested in the trend within time (t_a, t_b), we first model the trend within a larger range $(t_a - c, t_b + c)$, and then keep the trend estimation in (t_a, t_b) as the final result.

The lower bound may not be an issue, as the historical data before the range of interest are often available. For the upper bound, the estimate at the latest time point is needed most often. Our method is to simulate some data for the future time points.

2.4.1.1 Future Data Simulation

Note that we do not expect to achieve precise prediction for the future. Instead, we just want the simulated data to have the similar seasonal pattern with real data. The most similar ones available are the historical data. Thus, for the simulated data, we would like to repeat the month-to-month changes from previous cycles, which contain the pattern of seasonal effect. Considering the continuity, we may have a class of simulators of $\hat{\boldsymbol{y}}^{(k)} = (\hat{y}^{(k)}_{n+1}, \ldots, \hat{y}^{(k)}_{n+J})'$ using the month-to-month changes at

the same position in each cycle as

$$\hat{y}_{n+j}^{(k)} = \hat{y}_{n+j-1} + (y_{n+j-ks} - y_{n+j-ks-1}), \qquad k = 1, 2, ...; j = 1, 2, ..., J. \tag{2.13}$$

Note that $y_1, ..., y_n$ are the real data used, and we would like to simulate additional J months of data as $\hat{y}_{n+1}, ..., \hat{y}_{n+J}$. Given the period $s = 12$ months in a year, suppose the real data contain at least k cycles, as $n > ks$.

The k-th simple simulator will use the month-to-month change $(y_{n+j-ks} - y_{n+j-ks-1})$ from the real data k year(s) before. If the real series is long enough, eg. $k > 3$, we may use the weighted average of those k simple simulators to achieve a better guess of the real month-to-month difference.

We simulate the values of one month at a time, given the starting value $\hat{y}_n = y_n$, and generate the next one based on the real data and values simulated previously. Depending on the length of data used, we may include one or two or more cycles in the simulator. Here we generate $J = 12$ months' data points out of the upper bound, and propose four different simulators for comparison.

Simulator 1 borrows the information from one resent year of real data.Simulator 2 borrows the information from two resent years of real data.Simulators 3 & 4 borrow the information from three resent years of real data.

Simulator 1: $\hat{y}_{n+j} = \hat{y}_{n+j-1} + (y_{n+j-s} - y_{n+j-s-1}),\ j = 1, 2, ..., 12.$

Simulator 2: $\hat{y}_{n+j} = \hat{y}_{n+j-1} + 0.75(y_{n+j-s} - y_{n+j-s-1}) + 0.25(y_{n+j-2s} - y_{n+j-2s-1}),\ j = 1, ..., 12.$

Simulator 3: $\hat{y}_{n+j} = \hat{y}_{n+j-1} + 0.6(y_{n+j-s} - y_{n+j-s-1}) + 0.3(y_{n+j-2s} - y_{n+j-2s-1}) + 0.1(y_{n+j-3s} - y_{n+j-3s-1}), \qquad j = 1, ..., 12.$

Simulator 4: $\hat{y}_{n+j} = \hat{y}_{n+j-1} + \frac{1}{3}[(y_{n+j-s} - y_{n+j-s-1}) + (y_{n+j-2s} - y_{n+j-2s-1}) + (y_{n+j-3s} -$

$y_{n+j-3s-1})]$, $j = 1, ..., 12$.

The real data from 01/2006 to 01/2009 are used in Figures 2.4 and 2.5. In Figure 2.4, based on the real data from 01/2006 to 01/2009, we are interested in the trend estimates between 01/2007 and 01/2009. The real data from 01/2006 to 01/2009, and simulated data from 02/2009 to 01/2010 with different simulators are used. We also plot the real data from 02/2009 to 01/2010, which are not used, just for reference.

The simulators copy both the long-term tendency and the short-term seasonal pattern from previous cycles. When the long-term tendency itself changes rapidly over time, the four simulators perform differently due to different numbers of cycles involved for simulation, but we only want the simulated data to have a similar pattern compared with real data. All four simulators serve this purpose, and the estimates are almost identical.

In Figure 2.5, we show the posterior coefficient of variation of the four trend estimates in Figure 2.4. As expected, we achieve the low and stable variation after removing the lower and upper boundaries.

2.4.1.2 Predictive Posterior

We apply posterior predictive distribution to improve boundary performance. Given the known real data $\boldsymbol{y} = (y_1, ..., y_n)'$, we define $\boldsymbol{y}_* = (\hat{y}_{n+1}, ..., \hat{y}_{n+J})'$ as the additional J months data simulated. Let $\tilde{n} = n + J$, $\tilde{\boldsymbol{y}} = (\boldsymbol{y}', \ \boldsymbol{y}_*')' = (y_1, ..., y_n, \hat{y}_{n+1}, ..., \hat{y}_{\tilde{n}})'$, $\tilde{\boldsymbol{z}} = (z_1, ..., z_{\tilde{n}})'$, $\tilde{\boldsymbol{Q}} = \tilde{\boldsymbol{F}}_0' \tilde{\boldsymbol{F}}_1^{-1} \tilde{\boldsymbol{F}}_0$, which is the matrix $\boldsymbol{Q}$ in (2.3) with dimension $\tilde{n}$ instead of n.

The full conditional distributions for Gibbs sampling are given by

$$\begin{aligned}
(\tilde{\boldsymbol{z}}|\delta_0,\eta,\nu,\boldsymbol{y}_*;\boldsymbol{y}) &\sim N_{\tilde{n}}((\boldsymbol{I}_{\tilde{n}}+\eta\tilde{\boldsymbol{Q}})^{-1}\tilde{\boldsymbol{y}},\delta_0(\boldsymbol{I}_{\tilde{n}}+\eta\tilde{\boldsymbol{Q}})^{-1}),\\
(\delta_0|\tilde{\boldsymbol{z}},\eta,\nu,\boldsymbol{y}_*;\boldsymbol{y}) &\sim IG(n-1,\frac{1}{2}[(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{z}})'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{z}})+\eta\tilde{\boldsymbol{z}}'\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{z}}]),\\
(\eta|\tilde{\boldsymbol{z}},\delta_0,\nu,\boldsymbol{y}_*;\boldsymbol{y}) &\sim Gamma(\frac{n}{2},\nu+\frac{1}{2\delta_0}\tilde{\boldsymbol{z}}'\tilde{\boldsymbol{Q}}\tilde{\boldsymbol{z}}),\\
(\nu|\tilde{\boldsymbol{z}},\delta_0,\eta,\boldsymbol{y}_*;\boldsymbol{y}) &\sim Gamma(2,c+\eta),\\
(\hat{y}_i|\tilde{\boldsymbol{z}},\delta_0,\eta,\nu;\boldsymbol{y}) &\sim N(z_i,\delta_0), \qquad i=n+1,...,n+J.
\end{aligned}$$

We implement $J=12$ new time points for the Posterior Predictive Distribution. We also consider the influence of different initial values. In Figure 2.6, the initial values are borrowed from Simulators 1, 2, and 3 from the previous section.

We also explore the difference between different length of simulation. In Figure 2.7, we take $J=4,6,8,12$, and the initial values are shared from the initial values (2) in the previous example. The estimates don't change much.

2.4.2 Revisions as New Data are Available

The unemployment level and other labor force data are collected monthly by the Census Bureau, and we may obtain new trend estimates for the previous time points each time when new data are available. Our question is whether the current trend estimates will coincide with the previous one.

Note that the boundary improvement discussed in the previous section is an attempt to reduce the model uncertainty. Because we have to borrow information from all the series to estimate the trend at each time point, adding new data will inevitably affect the previous estimates. The current X-13 method also has this issue, and it is a common practice to revise the estimates after new data are collected.

In this section, we study the stability in terms of adding new data of Bayesian smoothing spline trend estimation and compare it with the X-13 method.

In Figures 2.8 and 2.9, we denote the non-revision trend estimates as lag 0. For lag 0, we first perform analysis with 01/2004 - 12/2008 data, and record the result for the time point of 12/2008 only. Then we perform analysis with 02/2004 - 01/2009 data, and record the result for the time point of 01/2009 only, and so on.

The 3-month revised trend is denoted as lag 3. For lag 3, we first perform analysis with 04/2004 - 03/2009 data, and record the result for the time point of 12/2008 only. Then we perform analysis with 05/2004 - 04/2009 data and record the result for the time point of 01/2009 only, and so on.

Similarly, we construct the 6-month, 12-month, 24-month and 36-month revised trends, which are shown as lag 6, lag 12, lag 24 and lag 36 respectively. We keep the estimation window, or the range of data used, as 60 months for each estimate for consistency.

Ideally, we wish to see that all trend estimates overlap each other. However, by comparing the revised and non-revised trend estimates, it is clear that the revision is necessary for both our Bayesian smoothing spline method and the X-13 method. But our method is more stable and requires less correction as new data are added. To better illustrate our strength, we calculate the relative correction between each revised trend estimation. For example, the relative correction between the 3-month revised trend (lag 3) and the 6-month revised trend (lag 6) is defined as:

$$Relative\ correction_{3,6} = \left| \frac{Trend_{lag3} - Trend_{lag6}}{Trend_{lag3}} \right|. \tag{2.14}$$

We summarize the relative correction for all the time points from 12/2008 to 12/2013 with Histograms 2.10 - 2.13.

For the comparison between the 3-month revised trend (lag 3) and the 6-month revised trend (lag 6) in Histogram 2.10, the relative correction with Bayesian smoothing spline ($mean = 0.4325\%$) is smaller than the one with X-13 ($mean = 0.5138\%$).

For the comparison between the 6-month revised trend (lag 6) and the 12-month revised trend (lag 12) in Histogram 2.11, the relative correction with Bayesian smoothing spline ($mean = 0.3296\%$) is smaller than the one with X-13 ($mean = 0.6886\%$).

For the comparison between the 12-month revised trend (lag 12) and the 24-month revised trend (lag 24) in Histogram 2.12, the relative correction with Bayesian smoothing spline ($mean = 0.2996\%$) is smaller than the one with X-13 ($mean = 0.6589\%$).

For the comparison between the 24-month revised trend (lag 24) and the 36-month revised trend (lag 36) in Histogram 2.13, the relative correction with Bayesian smoothing spline ($mean = 0.1970\%$, $median = 0.1558\%$, $quantile_{0.95} = 0.4949\%$) is much smaller than the one with X-13 ($mean = 0.5816\%$, $median = 0.4263\%$, $quantile_{0.95} = 1.486\%$). We also notice that relative correction of X-13 has a much larger discrepancy in Figures 2.11, 2.12 and 2.13.

From Histograms 2.10 - 2.13, with equally added data, the Bayesian smoothing spline method requires much less amount of revision than X-13. As more data are available, the amount of revision with Bayesian smoothing spline model decreases significantly. From the 24-month revised trend (lag 24) to the 36-month revised trend (lag 36), more than 95% of time points require correction that is smaller than 0.5%. The majority have almost no revision, which can be hardly identified from those Histograms.

2.5 Changing Point Detection

In this section, we would like to explore at which time point the unemployment tendency is changed, i.e. the unemployment level switches from decreasing to increasing or vice versa, which may be considered as an indicator of the economic crisis' coming or leaving.

Recall the smoothing spline model (Fessler 1991),

$$f(t) = \begin{cases} z_0 + b_0(t - t_0), & t \in [t_0, t_1), \\ z_i + b_i(t - t_i) + \dfrac{c_i}{2}(t - t_i)^2 + \dfrac{d_i}{6}(t - t_i)^3, & t \in [t_i, t_{i+1}), \\ z_n + b_n(t - t_n), & t \in [t_n, t_{n+1}]. \end{cases} \tag{2.15}$$

After we obtain the estimate of $\boldsymbol{z} = (z_1, ..., z_n)' = (f(t_1), ..., f(t_n))'$ from MCMC, with the following equations, we calculate the parameters for the natural cubic spline smoothing curve (2.15),

$$\begin{aligned} \mathbf{c} &= (c_2, ..., c_{n-1})' = \boldsymbol{F}_1^{-1}\boldsymbol{F}_0\boldsymbol{z}, \quad c_1 = 0, \\ d_{i-1} &= c_i - c_{i-1}, \\ b_{i-1} &= (z_i - z_{i-1}) - \frac{c_i + 2c_{i-1}}{6}, \quad b_n = b_{n-1} + c_{n-1} + \frac{d_{n-1}}{2}. \end{aligned}$$

Here we are interested in the b_i's, which are the values of the first derivative of the smoothing spline at each time point t_i, because they quantify the increasing/decreasing rates at those points. The changing points where the unemployment rate turns to be increasing from decreasing or vice versa, which are the indications of the top or bottom of economic cycles, are achieved by continuity approximately at the time when the curve of b_i crosses the horizontal line. In addition, our method is able to access the uncertainty of b_i efficiently with the posterior samples.

In Figure 2.14, we compute the point estimates of b_i's with respect to the trend estimates in Figure 2.1, along with the 95% credible interval. Figure 2.14 shows two changing points. In December of 2006, the unemployment level began to increase, indicating the coming of the recession. In February of 2010, the unemployment level started to decrease, indicating the ending of the recession period.

Note that, from the previous section, the boundary issue cannot be avoided for the estimation of the unemployment level, and this is also true for the estimation of b's. We expect larger variation here as they are the estimation of first derivatives.

2.6 Simulation

In this section, we explore the accuracy of our Bayesian smoothing spline method and compare it with the X-13 Program.

Here we use a simple additive model to simulate the data of 48 time points with 1 observation at each time.

$$y_i = f(t_i) + S(t_i), \qquad i = 1, 2, ..., 48,$$

where $f(t_i)$'s are the fixed true trend which are defined as

$$f(t) = \begin{cases} 0.2(t-15)^2 + 1000, & t \in [1, 15), \\ -2/3(t^3 - 58.5t^2 + 1080t) + 5275, & t \in [15, 24), \\ -0.2(t-24)^2 + 1243, & t \in [24, 48]. \end{cases}$$

The above function is manually created to simulate the tendency during economic crisis.

Following Pfeffermann (1991), the seasonal effect $S(t_{ij})$'s are simulated by

$$S_t = -\sum_{j=t-s+1}^{t-1} S_j + \omega_t,\ \omega_t \sim N(0, \delta),$$

where the number of seasons $s = 12$, and variance $\delta = 100$.

Figure 2.15 shows the true trend and the estimates, along with the simulated data. Our method overall performs better.

2.7 Comments

Our Bayesian smoothing spline model provides a method to remove the seasonal fluctuations thoroughly, achieves the smooth estimates from a time series, and captures the fundamental tendency of general economic expansions and contractions.

We propose a new procedure, which borrows the pattern of historical data to reduce the model uncertainty of estimates on the boundary, and to achieve the overall stable estimates. Compared with X-13, the method currently applied by BLS, our method obtains the desired continuity properties and produces efficient uncertainty assessment.

When new data are available, the previous trend estimates may need to be revised. Compared with the X-13 method, the Bayesian smoothing spline trend estimation requires less correction with equally added data. As more data are added, the amount of revision of the Bayesian smoothing spline estimates reduces sharply, whereas the X-13 result is revised slowly. Therefore, in terms of adding new data, our method is more stable and reliable than X-13.

Also, the continuity properties help us to further study the changing rate of a trend as well as the uncertainty, and to detect the changing point, where the trend

switches from increasing to decreasing, or vice versa.

In addition, our method does not require the data to be at least 3 year-long (36 observations) for monthly series. 13 months (observations) is the minimum for our boundary improvement procedure.

2.8 Figures

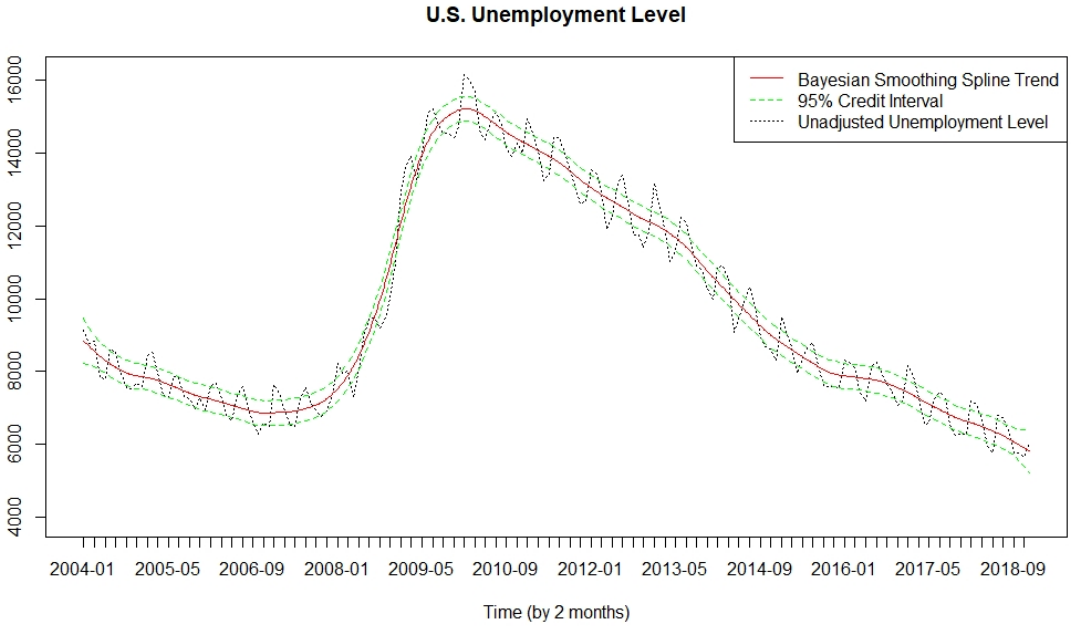

Figure 2.1: 95% credible interval of Bayesian Smoothing Spline Estimates

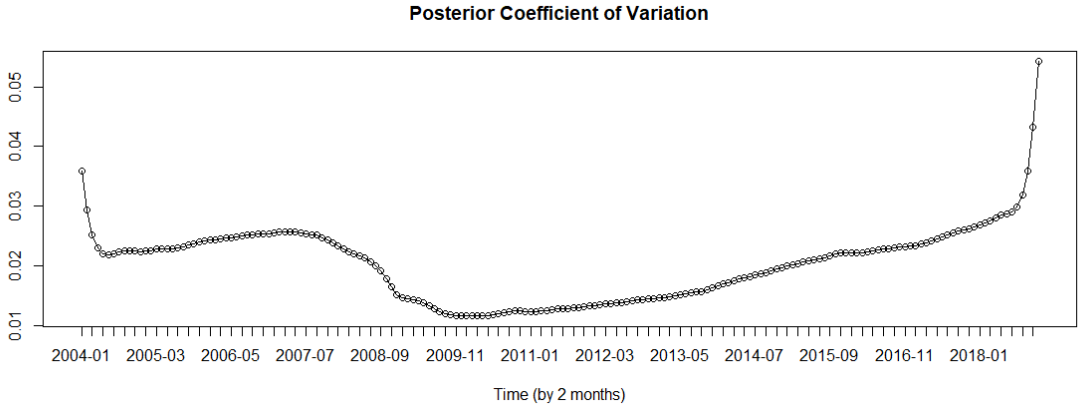

Figure 2.2: Posterior Coefficient of Variation of Bayesian Smoothing Spline Estimates

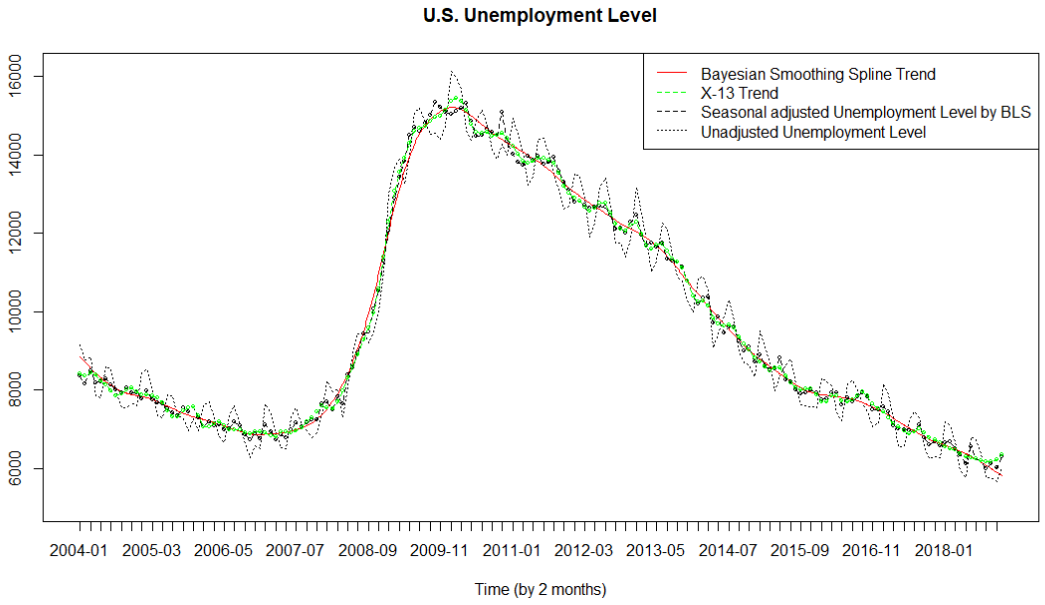

Figure 2.3: Unemployment Level: BSS Trend estimates compared with X-13 trend

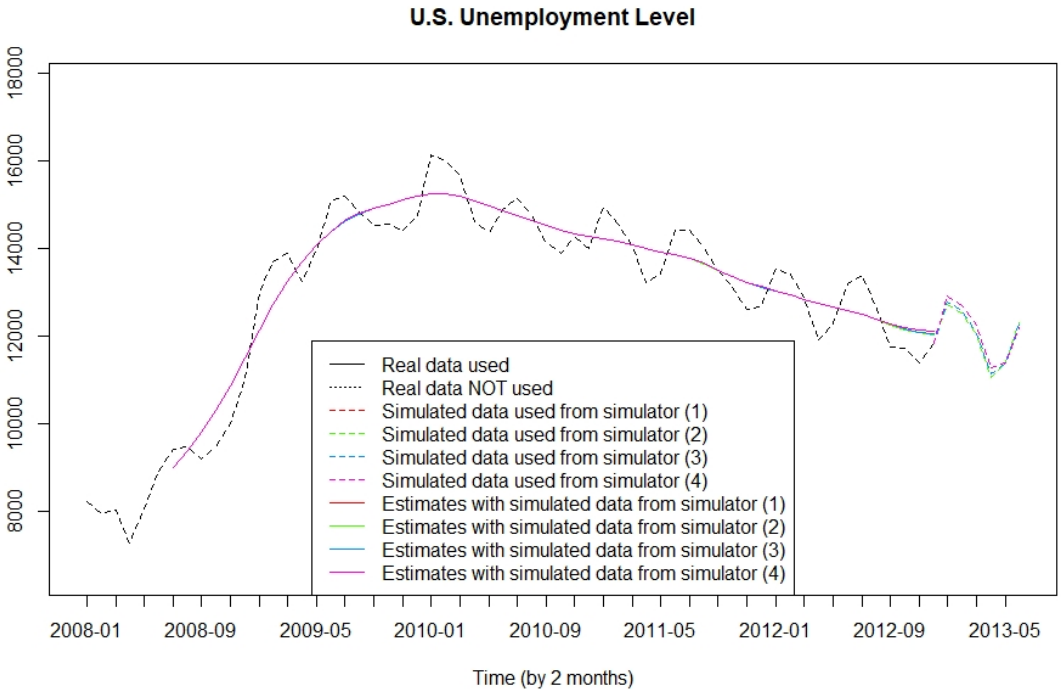

Figure 2.4: Boundary Improved BSS Trend Estimates with Simulators

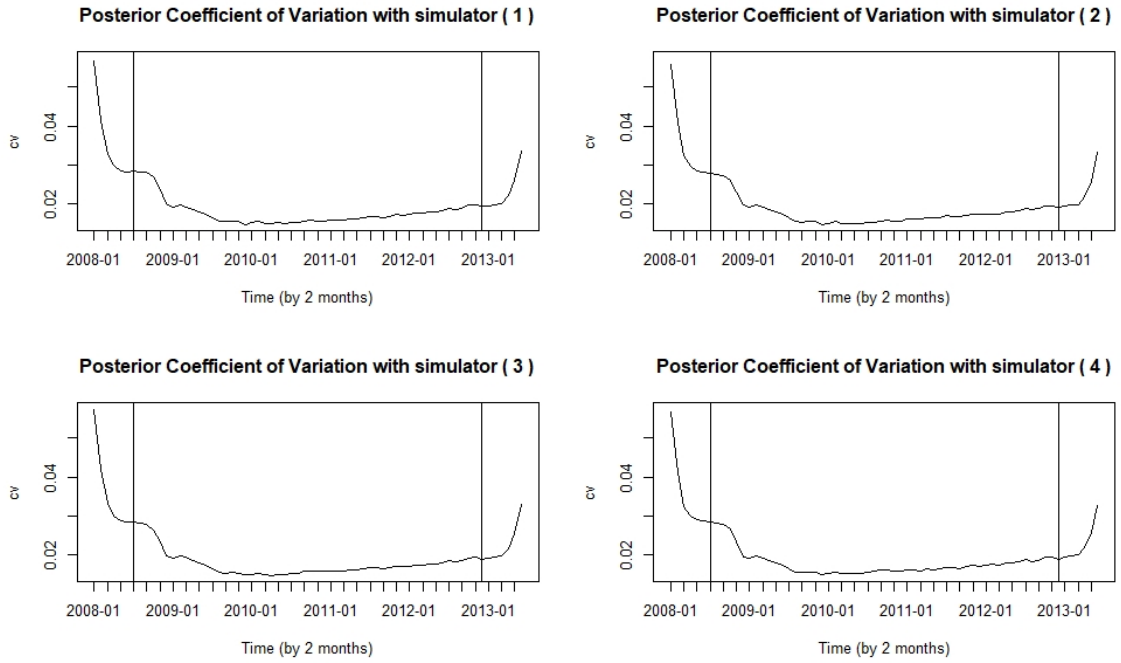

Figure 2.5: Posterior Coefficient of Variation of Bayesian Smoothing Spline Trend Estimates with Simulators

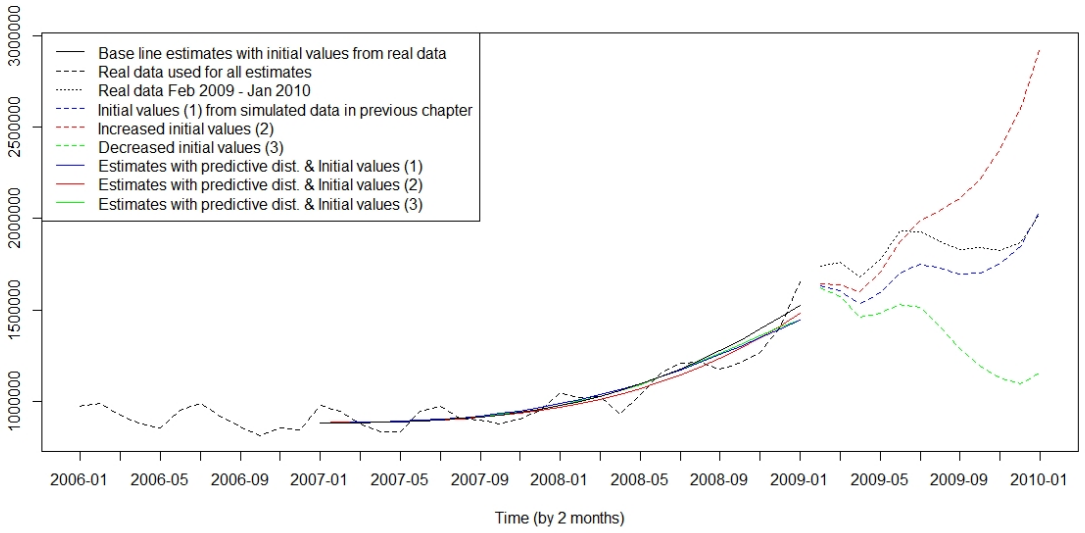

Figure 2.6: Comparison of Estimates with Predictive Distribution and Different Initial Values

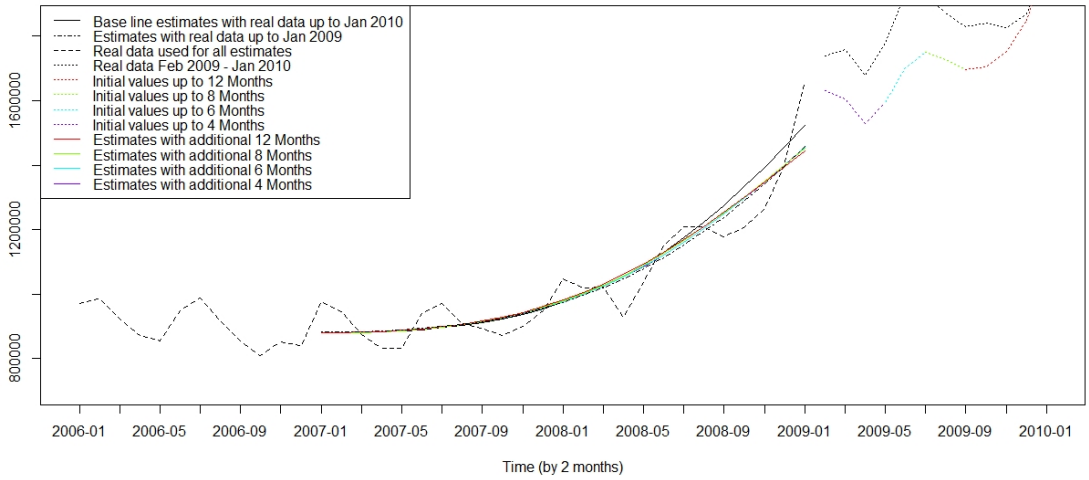

Figure 2.7: Comparison of Estimates with different length of simulation

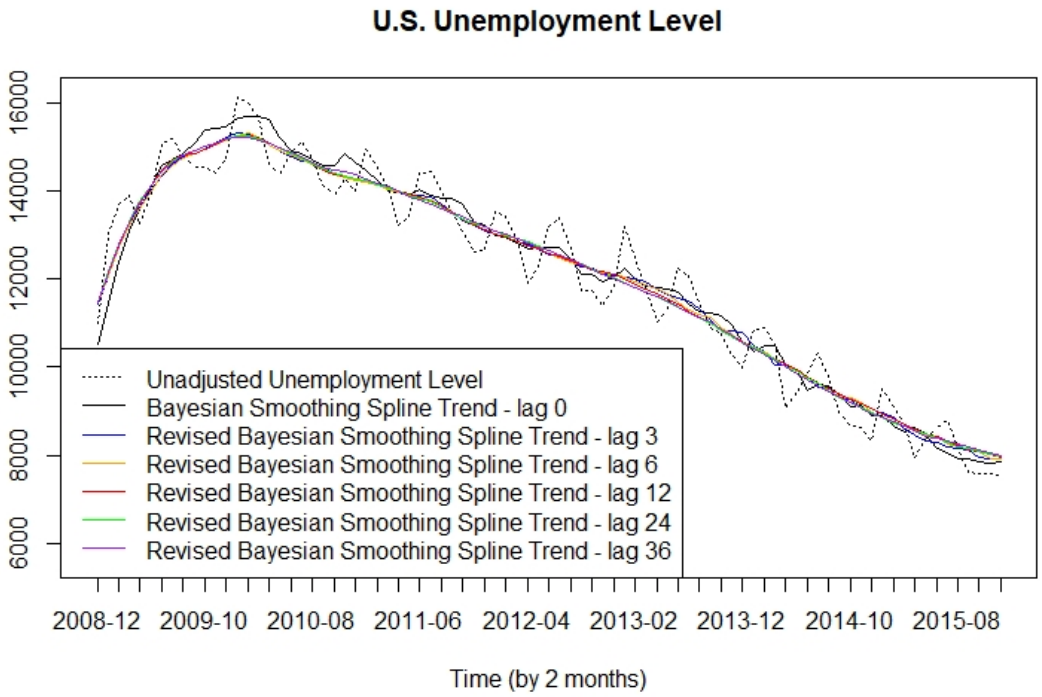

Figure 2.8: Comparison of Revised Bayesian Smoothing Spline Trend Estimates

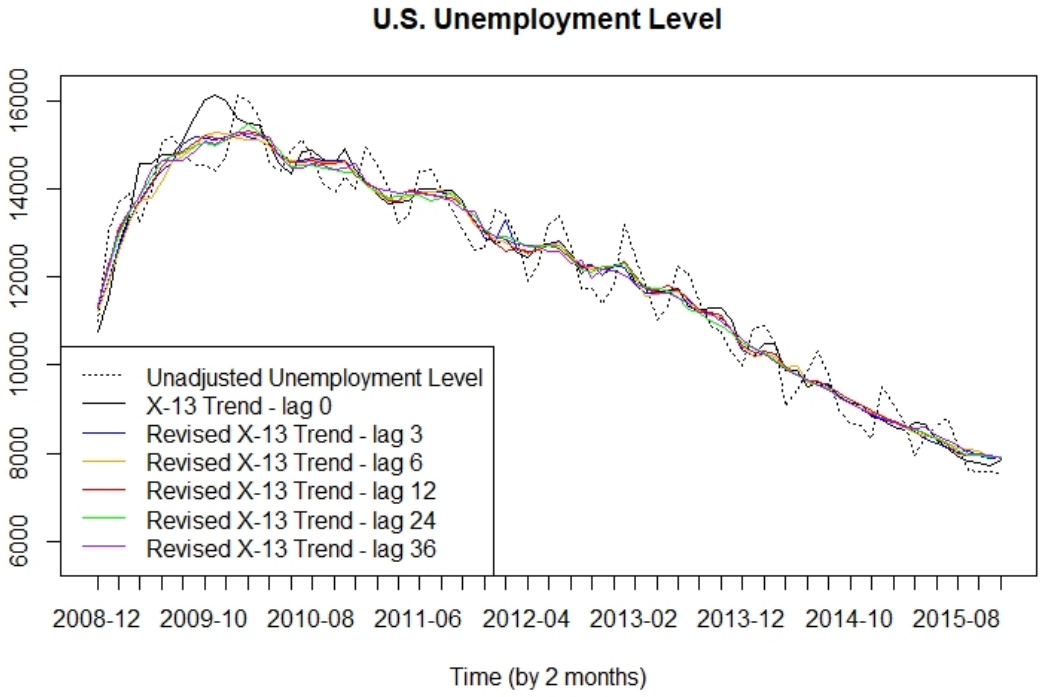

Figure 2.9: Comparison of Revised X-13 Trend Estimates

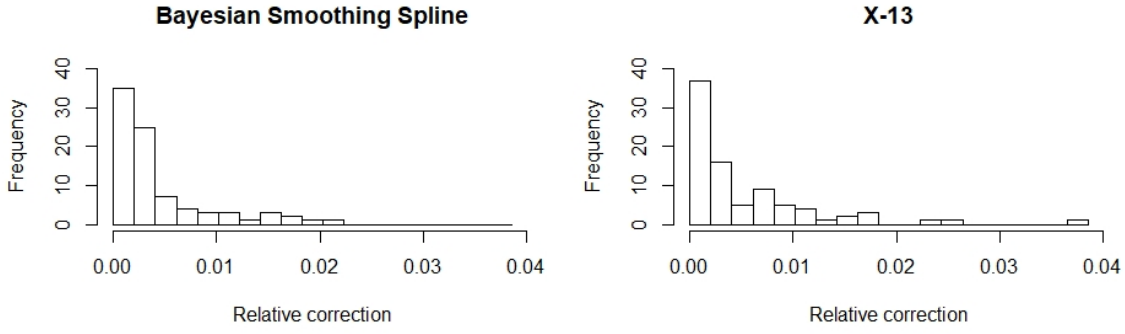

Figure 2.10: Relative correction between 3-month revised trend (lag 3) and 6-month revised trend (lag 6)

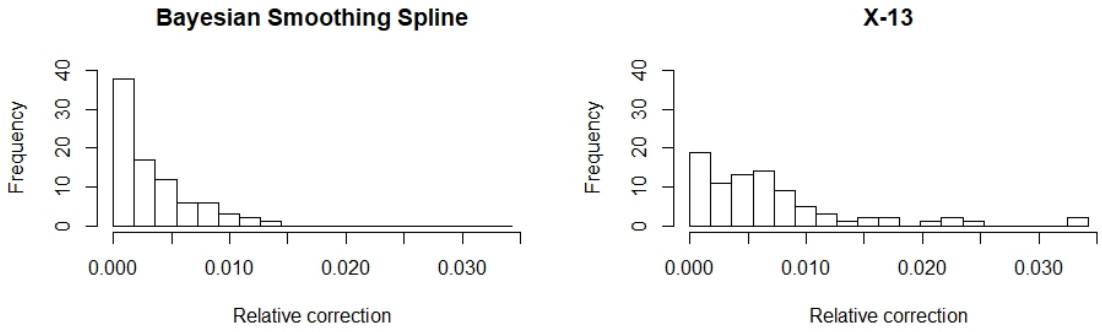

Figure 2.11: Relative correction between 6-month revised trend (lag 6) and 12-month revised trend (lag 12)

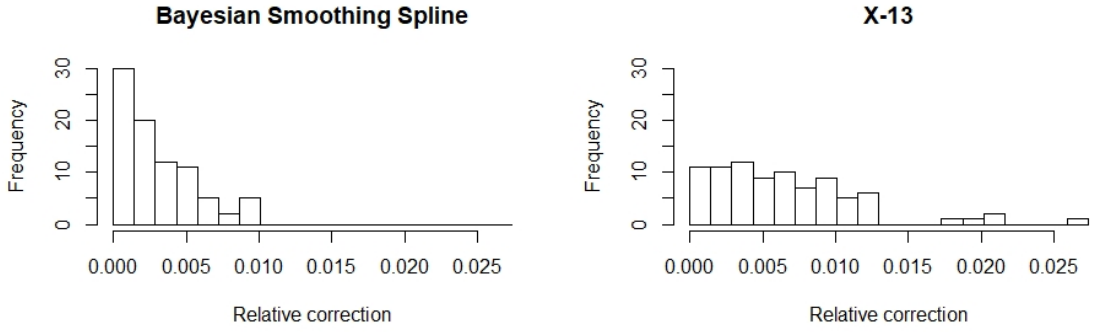

Figure 2.12: Relative correction between 12-month revised trend (lag 12) and 24-month revised trend (lag 24)

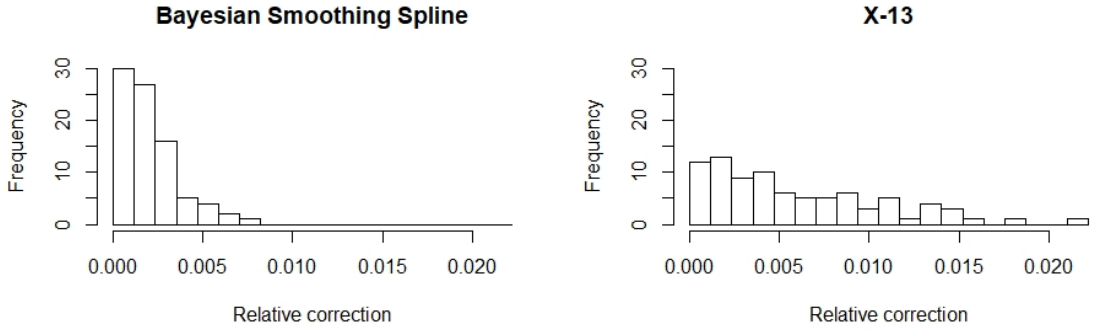

Figure 2.13: Relative correction between 24-month revised trend (lag 24) and 36-month revised trend (lag 36)

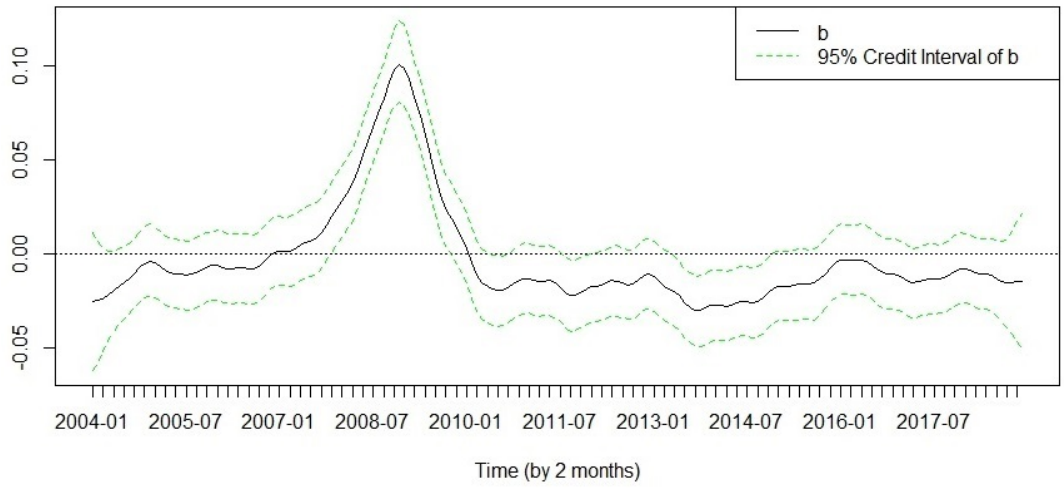

Figure 2.14: Detection of Changing Points

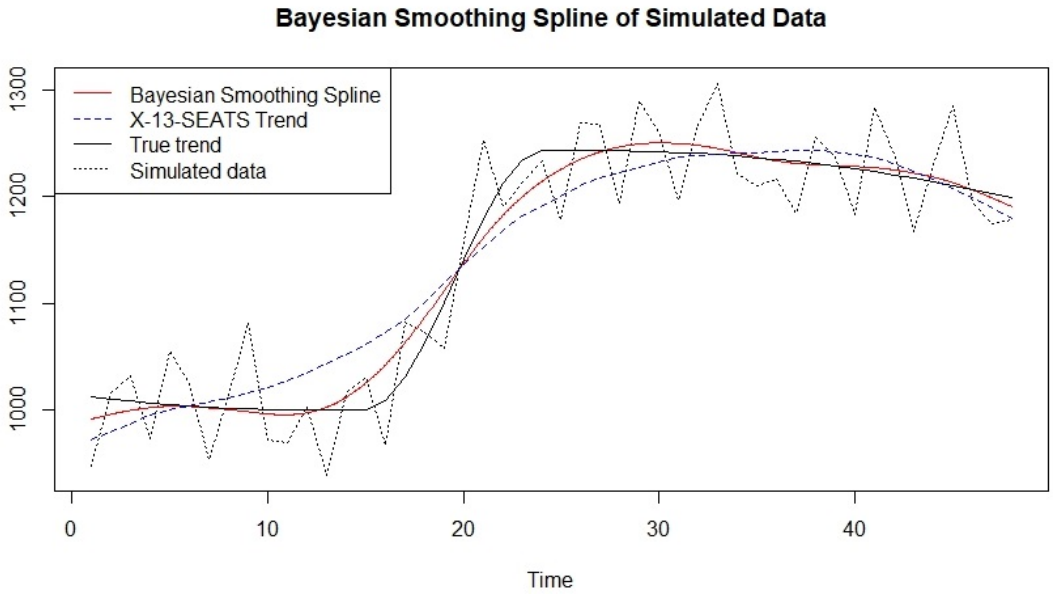

Figure 2.15: Simulation: Bayesian Smoothing Spline Compared with X-13

Chapter 3

Bayesian Smoothing Spline with Dependency Model

3.1 Introduction

The Bayesian smoothing spline model in Chapter 2 provides the long-term trend of the unemployment level with good properties. In the previous Chapter, we apply the basic Bayesian smoothing spline model to estimate the U.S. unemployment level from January of 2004 to December of 2018. To review the example 2.3, we analyze the residuals, which are defined in (3.1).

$$Residual = Unadjusted\ unemployment\ level\ -\ Trend\ estimate. \tag{3.1}$$

Figure 3.1 shows the residuals of 180 months from 01/2004 to 12/2018. Then we represent the 15 years' residuals in the domain of 12 months in a year in Figure 3.2. It is clear that a seasonal pattern of period $T = 12$ exists in the residuals. Thus it is reasonable to assume that the residuals of all Januaries are correlated, and the

residuals of all Februaries are correlated, etc.

In this chapter, we are interested in generalizing the basic model we use in Chapter 2, and propose a dependence structure to improve accuracy, precision and stability. In Section 3.2, we introduce the generalized model with the dependence structure adapted to the seasonal pattern, which we discovered in Figures 3.1-3.2. Section 3.3 explores the sampling procedure for the posterior distribution of our Bayesian smoothing spline with dependency (BSSD) model. In Section 3.4, we revisit the example 2.3, and the comparison shows that for real data, our new model has much better boundary performance than the basic model. Also the BSSD trend estimation is more stable than the result from the X-13 method. Section 3.5 and 3.6 discuss the model selection with DIC. In Section 3.7, we perform three simulation studies with different combinations of tendency and periodic errors. In this chapter, we refer to the univariate model, when the Bayesian smoothing spline with dependency (BSSD) model is mentioned without specification, and we will discuss the multivariate BSSD model in Chapter 4.

3.2 Bayesian smoothing spline with dependency

The basic Bayesian smoothing spline model consider the nonparametric regression model (Speckman and Sun 2003),

$$y_i = z_i + \epsilon_i, \;\; z_i = f(t_i), \tag{3.2}$$

where $\epsilon_i \overset{\text{iid}}{\sim} N(0, \delta_0)$, $i = 1, ..., n$. Let $\boldsymbol{y} = (y_1, ..., y_n)'$, $\boldsymbol{z} = (z_1, ..., z_n)'$, and $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)'$. We rewrite (3.2) as

$$\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \delta_0 \boldsymbol{I}_n). \tag{3.3}$$

Following the dependence assumption, we generalize the fully independent structure $\boldsymbol{I}_n$ to a specified dependent structure $\mathbf{R}$. Model (3.3) is generalized to

$$\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \delta_0 \mathbf{R}), \tag{3.4}$$

where $\mathbf{R} = (r_{ij})_{n \times n}$ is the correlation matrix. Assuming the period is T, the upper-triangular elements ($1 \leq i \leq j \leq n$) are

$$r_{ij} = \begin{cases} 1, & i = j, \\ \rho, & j = i + T \times k, k = 1, 2, 3, ..., \\ 0, & otherwise. \end{cases} \tag{3.5}$$

For example, we take three full years' data, the length of which is $n = 36$ (months), and the period is $T = 12$ (months). It is convenient to construct the correlation matrix as

$$\mathbf{R} = \begin{bmatrix} \boldsymbol{I}_T & \rho \boldsymbol{I}_T & \rho \boldsymbol{I}_T \\ \rho \boldsymbol{I}_T & \boldsymbol{I}_T & \rho \boldsymbol{I}_T \\ \rho \boldsymbol{I}_T & \rho \boldsymbol{I}_T & \boldsymbol{I}_T \end{bmatrix}. \tag{3.6}$$

In the case that the length of data n is not an integer multiple of T, for example $n = 32$, it is also easy to obtain the correlation matrix $\mathbf{R}$ by taking the sub-matrix of the matrix (3.6), which keeps the first 32 columns and rows.

Formally, suppose univariate observations $\boldsymbol{y} = (y_1, ..., y_n)$ are taken at points $\boldsymbol{t} = (t_1, \cdots, t_n)'$, where $-\infty < a \leq t_1 < \cdots < t_n \leq b < \infty$. An unknown function $f(t)$ is estimated to minimize the generalized loss function with a penalty on smoothness,

$$(\boldsymbol{y} - \boldsymbol{z})'\mathbf{R}^{-1}(\boldsymbol{y} - \boldsymbol{z}) + \eta \int_a^b [f^{(2)}(t)]^2 dt, \tag{3.7}$$

where $\boldsymbol{z} = (f(t_1), \ldots, f(t_n))'$, $f^{(2)}(t)$ is the second derivative, for Cubic natural smoothing spline. For equal spaced knots at $t = 1, 2, ..., n$, as a special case in Fessler (1991), the smoothness penalty term can be rewrite as $\eta \boldsymbol{z}'\boldsymbol{Q}\boldsymbol{z}$, where

$$\boldsymbol{Q} = \boldsymbol{F}_0' \boldsymbol{F}_1^{-1} \boldsymbol{F}_0, \tag{3.8}$$

and the $(n-2) \times n$ matrix $\boldsymbol{F}_0$ and $(n-2) \times (n-2)$ matrix $\boldsymbol{F}_1$ are

$$\boldsymbol{F}_0 = \begin{pmatrix} 1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1 \end{pmatrix}, \quad \boldsymbol{F}_1 = \frac{1}{6} \begin{pmatrix} 4 & 1 & \ldots & 0 & 0 \\ 1 & 4 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 4 \end{pmatrix}. \tag{3.9}$$

The minimizer (3.7) can be written as

$$(\boldsymbol{y} - \boldsymbol{z})'\mathbf{R}^{-1}(\boldsymbol{y} - \boldsymbol{z}) + \eta \boldsymbol{z}'\boldsymbol{Q}\boldsymbol{z}. \tag{3.10}$$

The solution to (3.10) can be found by taking the first derivative with respect to $\boldsymbol{z}$ and set it to zero.

$$\boldsymbol{\mu}_z = (\mathbf{R}^{-1} + \eta \boldsymbol{Q})^{-1} \mathbf{R}^{-1} \boldsymbol{y}. \tag{3.11}$$

3.2.1 Choice of Priors

Following Sun, Tsutakawa, and Speckman (1999), we use partially informative Normal prior for $\boldsymbol{z} = (z_1, ..., z_n)^T, z_i = f(t_i)$,

$$[\boldsymbol{z}|\delta_1] \propto \big(\frac{1}{\delta_1}\big)^{\frac{n-2}{2}} exp\big(-\frac{1}{2\delta_1}\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{z}\big). \tag{3.12}$$

The variance component is $\delta_1 = \delta_0/\eta$, where $\eta = \delta_0/\delta_1$ represents the noise-signal ratio.

We choose the Jeffreys prior for δ_0,

$$[\delta_0] \propto \delta_0^{-1}. \tag{3.13}$$

For the choice of prior of ρ, given the assumption of positive correlation, we use

$$[\rho] \propto (1-\rho)^{-\frac{1}{2}},\ 0 < \rho < 1, \tag{3.14}$$

with our belief that the correlation parameter ρ is closer to 1 than 0. We may also consider a class of priors

$$[\rho] \propto (1-\rho)^{-\frac{1}{2}}(1+\frac{\rho}{k-1})^{-\frac{1}{2}},\ 0 < \rho < 1, k = 2, 3, \ldots \tag{3.15}$$

We choose scaled Pareto prior (Cheng and Speckman 2012) for η

$$[\eta] = \frac{c}{(c+\eta)^2}, \eta > 0. \tag{3.16}$$

Its median is the scale parameter c. As c increases, the prior tends flatter and less concentrated to 0.

Following White (2006), we choose the hyper-parameter c based on the effective degrees of freedom. In Hastie and Tibshirani (1990), the degrees of freedom of a smoother is defined as the trace of the smoother matrix, i.e.

$$e.d.f = tr(\boldsymbol{S}_\eta), \tag{3.17}$$

where $\boldsymbol{S}_\eta = (\boldsymbol{I}_n - \eta \mathbf{R}\boldsymbol{Q})^{-1}$ for our BSSD model. Different from the discussion in White 2006 and Yue, Speckman, and Sun 2012, there is another unknown parameter ρ existing in the smoother matrix. We find that the choice of c and the performance of prior (3.16) are not very sensitive to different values of ρ. We select the value c that sets trace $tr(\boldsymbol{S}_c)$ equal to a desired prior degrees of freedom and correlation ρ.

3.2.2 Posterior

For the nonparametric model (3.4) with prior densities (3.12), (3.13), (3.14) and (3.16), we have the joint posterior density of $(\boldsymbol{z}, \delta_0, \rho, \eta)$

$$\begin{aligned}
[\boldsymbol{z}, \delta_0, \rho, \eta | \boldsymbol{y}] &\propto [\boldsymbol{y}|\boldsymbol{z}, \delta_0, \eta][\boldsymbol{z}|\delta_0, \eta][\eta][\delta_0][\rho] \\
&\propto [\rho]\frac{\eta^{\frac{n-2}{2}}}{\delta_0^n |\mathbf{R}|^{\frac{1}{2}}}\frac{c}{(c+\eta)^2} exp\big\{-\frac{1}{2\delta_0}(\boldsymbol{z}-\boldsymbol{\mu}_z)'(\mathbf{R}^{-1}+\eta\boldsymbol{Q})(\boldsymbol{z}-\boldsymbol{\mu}_z)\big\} \\
&\quad exp\big\{-\frac{1}{2\delta_0}[\boldsymbol{y}'\mathbf{R}^{-1}\boldsymbol{y} - \boldsymbol{\mu}_z'(\mathbf{R}^{-1}+\eta\boldsymbol{Q})\boldsymbol{\mu}_z]\big\},
\end{aligned} \tag{3.18}$$

where $\boldsymbol{\mu}_z = (\mathbf{R}^{-1} + \eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}$, given in (3.11).

Theorem 1. *Suppose the length of data $n > T$, where integer T is the period. Consider the nonparametric model (3.4) with prior distribution $[\boldsymbol{z}, \delta_0|\eta]$ given by (3.12), (3.13). Assume the correlation matrix $\boldsymbol{R}$ is positive definite(p.d.). Then we have the following results:*

(a) The joint posterior $(\boldsymbol{z}, \delta_0, \rho, \eta|\boldsymbol{y})$ *is proportion to*

$$\begin{aligned}[\boldsymbol{z}, \delta_0, \rho, \eta|\boldsymbol{y}] \quad &\propto \quad [\rho][\eta]\frac{\eta^{\frac{n}{2}-1}}{\delta_0^n\,|\boldsymbol{R}|^{\frac{1}{2}}}exp\big\{-\frac{1}{2\delta_0}(\boldsymbol{z}-\boldsymbol{\mu}_z)'(\boldsymbol{R}^{-1}+\eta\boldsymbol{Q})(\boldsymbol{z}-\boldsymbol{\mu}_z)\big\} \\ &\qquad exp\big\{-\frac{1}{2\delta_0}\left[\eta\boldsymbol{y}'\boldsymbol{Q}(\boldsymbol{R}^{-1}+\eta\boldsymbol{Q})^{-1}\boldsymbol{R}^{-1}\boldsymbol{y}\right]\big\}, \qquad (3.19)\end{aligned}$$

where $\boldsymbol{\mu}_z = (\boldsymbol{R}^{-1}+\eta\boldsymbol{Q})^{-1}\boldsymbol{R}^{-1}\boldsymbol{y}$*, given in (3.11).*

(b) The conditional posterior of $(\boldsymbol{z}|\delta_0, \rho, \eta; \boldsymbol{y})$ *is*

$$(\boldsymbol{z}|\delta_0, \rho, \eta; \boldsymbol{y}) \sim N_n\left(\boldsymbol{\mu}_z, \delta_0(\boldsymbol{R}^{-1}+\eta\boldsymbol{Q})^{-1}\right). \qquad (3.20)$$

(c) The conditional posterior of $(\delta_0|\rho, \eta; \boldsymbol{y})$ *is*

$$(\delta_0|\rho, \eta; \boldsymbol{y}) \sim IG\big(\frac{n}{2}-1, \frac{\eta}{2}\boldsymbol{y}'\boldsymbol{Q}(\boldsymbol{R}^{-1}+\eta\boldsymbol{Q})^{-1}\boldsymbol{R}^{-1}\boldsymbol{y}\big). \qquad (3.21)$$

(d) The joint posterior $p(\rho, \eta|\boldsymbol{y})$ *is proportional to*

$$[\rho, \eta|\boldsymbol{y}] \propto [\rho][\eta]\,|\boldsymbol{I}_n+\eta\boldsymbol{R}\boldsymbol{Q}|^{-\frac{1}{2}}\,[\boldsymbol{y}'\boldsymbol{Q}(\boldsymbol{R}^{-1}+\eta\boldsymbol{Q})^{-1}\boldsymbol{R}^{-1}\boldsymbol{y}]^{1-\frac{n}{2}}. \qquad (3.22)$$

(e) The marginal likelihood of (ρ, η)

$$\begin{aligned}L(\rho, \eta) \quad &= \quad \int [\boldsymbol{y}|\boldsymbol{z}, \delta_0, \rho][\boldsymbol{z}, \delta_0|\eta]d\boldsymbol{z}d\delta_0 \\ &\propto \quad |\boldsymbol{I}_n+\eta\boldsymbol{R}\boldsymbol{Q}|^{-\frac{1}{2}}\,[\boldsymbol{y}'\boldsymbol{Q}(\boldsymbol{R}^{-1}+\eta\boldsymbol{Q})^{-1}\boldsymbol{R}^{-1}\boldsymbol{y}]^{1-\frac{n}{2}}. \qquad (3.23)\end{aligned}$$

is upper-bounded, i.e. there exists some constant $C < +\infty$ *such that* $L(\rho, \eta) \leq C$*. The joint posterior is proper, given proper priors* $[\rho]$ *and* $[\eta]$*.*

Proof of Theorem 1. From (3.18), the joint posterior density of $(\boldsymbol{z}, \delta_0, \rho, \eta|\boldsymbol{y})$ is

$$\begin{aligned}[\boldsymbol{z}, \delta_0, \rho, \eta|\boldsymbol{y}] \quad &\propto \quad [\rho][\eta]\frac{\eta^{\frac{n}{2}-1}}{\delta_0^n \left|\mathbf{R}\right|^{\frac{1}{2}}} exp\big\{-\frac{1}{2\delta_0}(\boldsymbol{z}-\boldsymbol{\mu}_z)'(\mathbf{R}^{-1}+\eta\boldsymbol{Q})(\boldsymbol{z}-\boldsymbol{\mu}_z)\big\} \\ &\qquad exp\big\{-\frac{1}{2\delta_0}\left[\eta\boldsymbol{y}'\boldsymbol{Q}(\boldsymbol{I}_n+\eta\mathbf{R}\boldsymbol{Q})^{-1}\boldsymbol{y}\right]\big\},\end{aligned}$$

where $\boldsymbol{\mu}_z = (\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}$, which gives (3.19).

From (3.19), the conditional posterior of $(\boldsymbol{z}|\delta_0, \rho, \eta; \boldsymbol{y})$ is

$$[\boldsymbol{z}|\delta_0, \rho, \eta; \boldsymbol{y}] \propto exp\{-\frac{1}{2\delta_0}[(\boldsymbol{z}-\boldsymbol{\mu}_z)'(\mathbf{R}^{-1}+\eta\boldsymbol{Q})(\boldsymbol{z}-\boldsymbol{\mu}_z)]\}, \tag{3.24}$$

which proves part (b).

We integral (3.19) over $\boldsymbol{z}$, and obtain

$$\begin{aligned}[\delta_0, \rho, \eta|\boldsymbol{y}] \quad &\propto \quad [\rho][\eta]\frac{\eta^{\frac{n}{2}-1}}{\delta_0^{\frac{n}{2}}\left|\mathbf{R}\right|^{\frac{1}{2}}\left|\mathbf{R}^{-1}+\eta\boldsymbol{Q}\right|^{\frac{1}{2}}} exp\Big\{-\frac{1}{2\delta_0}\left[\eta\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}\right]\Big\} \\ &\propto \quad \frac{\eta^{\frac{n}{2}-1}\left[\mathbf{R}\right]\left[\eta\right]}{\left|\boldsymbol{I}_n+\eta\mathbf{R}\boldsymbol{Q}\right|^{\frac{1}{2}}}\delta_0^{-\frac{n}{2}} exp\Big\{-\frac{1}{\delta_0}\left[\frac{\eta}{2}\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}\right]\Big\},\end{aligned} \tag{3.25}$$

which gives the conditional distribution of $(\delta_0|\boldsymbol{y}, \rho, \eta)$,

$$[\delta_0|\rho, \eta; \boldsymbol{y}] \propto \delta_0^{-\frac{n}{2}} exp\Big\{-\frac{1}{\delta_0}\left[\frac{\eta}{2}\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}\right]\Big\}, \tag{3.26}$$

which implies part (c).

We integral (3.25) over δ_0, and obtain (3.22) as follows

$$[\rho, \eta|\boldsymbol{y}] \propto [\rho][\eta]\left|\boldsymbol{I}_n+\eta\mathbf{R}\boldsymbol{Q}\right|^{-\frac{1}{2}}\left[\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}\right]^{1-\frac{n}{2}}.$$

The likelihood of (ρ, η) as $L(\rho, \eta) = \int [\boldsymbol{y}|\boldsymbol{z}, \delta_0, \rho][\boldsymbol{z}, \delta_0|\eta] d\boldsymbol{z} d\delta_0$ is proportional to (3.23)

as

$$L(\rho,\eta) \propto |\boldsymbol{I}_n + \eta \mathbf{R}\boldsymbol{Q}|^{-\frac{1}{2}} \left[\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1} + \eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}\right]^{1-\frac{n}{2}}.$$

As the precision matrix $\boldsymbol{Q}$ in prior (3.12) has the rank $n-2$. There exists an orthogonal matrix $\boldsymbol{\Gamma}$ such that $\boldsymbol{Q} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'$, $\boldsymbol{\Lambda} = Diag(0, 0, \lambda_3, \ldots, \lambda_n)$, $0 < \lambda_3 \leq \cdots \leq \lambda_n$ are the non-zero eigenvalues of $\boldsymbol{Q}$. $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ are known given data.

Given $\mathbf{R}$ is p.d. and symmetric. There exists an orthogonal matrix $\boldsymbol{P}$ such that $\mathbf{R} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$, $\boldsymbol{D} = Diag(d_1, \ldots, d_n)$, $0 < d_1 \leq \cdots \leq d_n$ are the eigenvalues of $\mathbf{R}$. Let $\boldsymbol{P} = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n)$, $\mathbf{R}\boldsymbol{p}_i = d_i\boldsymbol{p}_i$. The $\boldsymbol{p}_i, i = 1, \ldots, n$ does not depend on parameter ρ. (see Appendix: 3.10)

Let $\boldsymbol{\Lambda}_+ = diag(\lambda_3, \ldots, \lambda_n)$, $\boldsymbol{\Lambda}_* = diag(\boldsymbol{I}_2, \boldsymbol{\Lambda}_+)$. Then $\boldsymbol{\Lambda}_*^{-1} = diag(\boldsymbol{I}_2, \boldsymbol{\Lambda}_+^{-1})$, $\boldsymbol{\Lambda}\boldsymbol{\Lambda}_*^{-1} = diag(\mathbf{0}_2, \boldsymbol{I}_{n-2})$, and

$$\begin{aligned}
\boldsymbol{Q}(\mathbf{R}^{-1} + \eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1} &= \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'(\boldsymbol{I}_n + \eta\mathbf{R}\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}')^{-1}\boldsymbol{\Gamma}\boldsymbol{\Gamma}' \\
&= \boldsymbol{\Gamma}\boldsymbol{\Lambda}(\boldsymbol{I}_n + \eta\boldsymbol{\Gamma}'\mathbf{R}\boldsymbol{\Gamma}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Gamma}' \\
&= \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Lambda}_*^{-1}\boldsymbol{\Lambda}_*(\boldsymbol{I}_n + \eta\boldsymbol{\Gamma}'\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'\boldsymbol{\Gamma}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Gamma}' \\
&= \boldsymbol{\Gamma}\begin{pmatrix} \mathbf{0}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-2} \end{pmatrix}\left[\boldsymbol{\Lambda}_*^{-1} + \eta\boldsymbol{\Gamma}'\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'\boldsymbol{\Gamma}\begin{pmatrix} \mathbf{0}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{n-2} \end{pmatrix}\right]^{-1}\boldsymbol{\Gamma}'. \qquad (3.27)
\end{aligned}$$

Let

$$\boldsymbol{\Gamma}'\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{B}_{2\times 2} & \boldsymbol{E}_{2\times(n-2)} \\ \boldsymbol{E}'_{(n-2)\times 2} & \boldsymbol{H}_{(n-2)\times(n-2)} \end{pmatrix} > 0, \qquad (3.28)$$

then $\boldsymbol{H} > 0$ by Lemma 3.10.5. Then there exists orthogonal matrix $\boldsymbol{U}_{(n-2)\times(n-2)}$, $\boldsymbol{D}_* = diag(d_3^*, \ldots, d_n^*)$, such that $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{D}_*\boldsymbol{U}'$.

By Lemma 3.10.10,

$$d_{i-2} \leq d_i^* \leq d_i. \tag{3.29}$$

The (3.27) can be written as

$$\begin{aligned}
\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1} &= \boldsymbol{\Gamma}\begin{pmatrix}\mathbf{0}_2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-2}\end{pmatrix}\left[\begin{pmatrix}\boldsymbol{I}_2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Lambda}_+^{-1}\end{pmatrix}+\eta\begin{pmatrix}\mathbf{0} & \boldsymbol{F}\\ \mathbf{0} & \boldsymbol{H}\end{pmatrix}\right]^{-1}\boldsymbol{\Gamma}'\\
&= \boldsymbol{\Gamma}\left[\begin{pmatrix}\mathbf{0}_2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-2}\end{pmatrix}\begin{pmatrix}\boldsymbol{I}_2 & \eta\boldsymbol{F}\\ \mathbf{0} & \boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H}\end{pmatrix}^{-1}\right]\boldsymbol{\Gamma}'\\
&= \boldsymbol{\Gamma}\left[\begin{pmatrix}\mathbf{0}_2 & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}_{n-2}\end{pmatrix}\begin{pmatrix}\boldsymbol{I}_2 & -\eta\boldsymbol{F}(\boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H})^{-1}\\ \mathbf{0} & (\boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H})^{-1}\end{pmatrix}\right]\boldsymbol{\Gamma}'\\
&= \boldsymbol{\Gamma}\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & (\boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H})^{-1}\end{pmatrix}\boldsymbol{\Gamma}'\\
&= \boldsymbol{\Gamma}\begin{pmatrix}\mathbf{0}_{2\times(n-2)}\\ \boldsymbol{I}_{n-2}\end{pmatrix}(\boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H})^{-1}\begin{pmatrix}\mathbf{0}_{(n-2)\times 2} & \boldsymbol{I}_{n-2}\end{pmatrix}\boldsymbol{\Gamma}'.
\end{aligned} \tag{3.30}$$

As $\boldsymbol{\Lambda}_+^{-1} = diag(\lambda_3^{-1},\ldots,\lambda_n^{-1})$ and $\boldsymbol{H}$ are symmetric and positive definite, $\boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H}$ is also symmetric and positive definite.

$$\lambda_n^{-1}\boldsymbol{I}+\eta\boldsymbol{H} \leq \boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H} \leq \lambda_3^{-1}\boldsymbol{I}+\eta\boldsymbol{H}. \tag{3.31}$$

By Lemma 3.10.7,

$$(\lambda_3^{-1}\boldsymbol{I}+\eta\boldsymbol{H})^{-1} \leq (\boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H})^{-1} \leq (\lambda_n^{-1}\boldsymbol{I}+\eta\boldsymbol{H})^{-1}. \tag{3.32}$$

Let $\boldsymbol{v} = \big(\mathbf{0}_{(n-2)\times 2}\ \boldsymbol{I}_{n-2}\big)\boldsymbol{\Gamma}'\boldsymbol{y}$, which is known given data,

$$\begin{aligned}
\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y} &= \boldsymbol{v}'(\boldsymbol{\Lambda}_+^{-1}+\eta\boldsymbol{H})^{-1}\boldsymbol{v} \\
&\geq \boldsymbol{v}'(\lambda_3^{-1}\boldsymbol{I}+\eta\boldsymbol{H})^{-1}\boldsymbol{v} \\
&= \lambda_3(\boldsymbol{U}'\boldsymbol{v})'(\boldsymbol{I}+\eta\lambda_3\boldsymbol{D}_*)^{-1}(\boldsymbol{U}'\boldsymbol{v}), \qquad (3.33)
\end{aligned}$$

and

$$\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y} \leq \lambda_n(\boldsymbol{U}'\boldsymbol{v})'(\boldsymbol{I}+\eta\lambda_n\boldsymbol{D}_*)^{-1}(\boldsymbol{U}'\boldsymbol{v}). \qquad (3.34)$$

Let $\boldsymbol{v}_* = \boldsymbol{U}'\boldsymbol{v} = (v_{*3},\ldots,v_{*n})$, then $\boldsymbol{v}_*'\boldsymbol{v}_* = \boldsymbol{v}'\boldsymbol{v}$,

$$\begin{aligned}
\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y} &\geq \lambda_3\boldsymbol{v}_*'(\boldsymbol{I}+\eta\lambda_3\boldsymbol{D}_*)^{-1}\boldsymbol{v}_* \\
&= \lambda_3\sum_{i=3}^{n}\frac{v_{*i}^2}{1+\eta\lambda_3 d_i^*} \\
&\geq \lambda_3\sum_{i=3}^{n}\frac{v_{*i}^2}{1+\eta\lambda_3 d_i}, \qquad (3.35)
\end{aligned}$$

$$\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y} \leq \lambda_n\sum_{i=3}^{n}\frac{v_{*i}^2}{1+\eta\lambda_n d_i^*} \leq \lambda_n\sum_{i=3}^{n}\frac{v_{*i}^2}{1+\eta\lambda_n d_{i-2}}. \qquad (3.36)$$

By Lemma 3.10.4,

$$d_1^{-1}d_2^{-1}\prod_{j=3}^{n}(d_j^{-1}+\eta\lambda_j) \leq |\mathbf{R}^{-1}+\eta\boldsymbol{Q}| \leq \big[\prod_{j=1}^{n-2}(d_j^{-1}+\eta\lambda_{n-j+1})\big]d_{n-1}^{-1}d_n^{-1}. \qquad (3.37)$$

Since $|\mathbf{R}| = \prod_{j=1}^{n} d_j$,

$$\prod_{j=3}^{n}(1+\eta d_j\lambda_j) \leq |\mathbf{R}||\mathbf{R}^{-1}+\eta\boldsymbol{Q}| \leq \prod_{j=1}^{n-2}(1+\eta d_j\lambda_{n-j+1}). \qquad (3.38)$$

Therefore

$$|\boldsymbol{I}_n + \eta \mathbf{R}\boldsymbol{Q}|^{-\frac{1}{2}} \leq \big[\prod_{j=3}^{n}(1+\eta\lambda_j d_j)\big]^{-\frac{1}{2}} \leq \big[\prod_{j=3}^{n}(1+\eta\lambda_3 d_j)\big]^{-\frac{1}{2}}, \tag{3.39}$$

and

$$|\boldsymbol{I}_n + \eta \mathbf{R}\boldsymbol{Q}|^{-\frac{1}{2}} \geq \big[\prod_{j=1}^{n-2}(1+\eta d_j\lambda_{n-j+1})\big]^{-\frac{1}{2}}. \tag{3.40}$$

We now prove that the likelihood of (η, ρ) is upper-bounded in two separated cases.

Case 1. First we consider the simple case that $n = mT$, where integer $m \geq 2$ and T is the period. The eigenvalues of $\mathbf{R}$ defined in (3.5) have the close form as (see Appendix 3.10)

$$d_i = \begin{cases} 1-\rho, & i = 1, \ldots, (m-1)T, \\ 1-\rho+m\rho, & i = (m-1)T+1, \ldots, n. \end{cases} \tag{3.41}$$

For eigenvalues of $\boldsymbol{H}$, the $n-2$ order principal submatrix of $\mathbf{R}$, by (3.29)

$$d_3^* = \cdots = d_{(m-1)T}^* = 1-\rho = d_3 = \cdots = d_{(m-1)T}. \tag{3.42}$$

By Lemma 3.10.11, the corresponding eigenvectors $\boldsymbol{\xi}_i, i = 3, \ldots, (m-1)T$ only depend on the first $(m-1)T-2 = n-T-2$ columns $\boldsymbol{\Gamma}'\boldsymbol{P}$. $\boldsymbol{P}$ does not depend on ρ.(see Appendix 3.10) Then the first $n-T-2$ columns of $\boldsymbol{U}$ also do not depend on parameter ρ. For $\boldsymbol{v}_* = \boldsymbol{U}'[\mathbf{0}_{(n-2)\times 2}\ \boldsymbol{I}_{n-2}]\boldsymbol{\Gamma}'\boldsymbol{y} = (v_{*3}, \ldots, v_{*n})$, the $(v_{*3}, \ldots, v_{*(m-1)T})$ is known given data.

Let $C_1 = \sum_{i=3}^{(m-1)T} v_{*i}^2 > 0$, and $C_2 = \boldsymbol{v}'\boldsymbol{v} = \sum_{i=3}^{mT} v_{*i}^2 > 0$. The upper bound is

$$
\begin{aligned}
& \left| \boldsymbol{I}_n + \eta \mathbf{R}\boldsymbol{Q} \right|^{-\frac{1}{2}} [\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1} + \eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}]^{1-\frac{n}{2}} \\
\leq & \left[\prod_{j=3}^{n} (1 + \eta\lambda_3 d_j) \right]^{-\frac{1}{2}} \left[\lambda_3 \sum_{i=3}^{n} \frac{v_{*i}^2}{1 + \eta\lambda_3 d_i} \right]^{1-\frac{n}{2}} \\
= & (\lambda_3)^{-\frac{n-2}{2}} \left\{ [1 + \eta\lambda_3(1-\rho)]^{n-T-2} [1 + \eta\lambda_3(1-\rho+m\rho)]^{T} \right\}^{-\frac{1}{2}} \\
& \left\{ \left[\sum_{i=3}^{(m-1)T} \frac{v_{*i}^2}{1 + \eta\lambda_3(1-\rho)} + \sum_{i=(m-1)T+1}^{mT} \frac{v_{*i}^2}{1 + \eta\lambda_3(1-\rho+m\rho)} \right]^{(n-T-2)+T} \right\}^{-\frac{1}{2}} \\
= & (\lambda_3)^{-\frac{n-2}{2}} \left\{ \left[\sum_{i=3}^{(m-1)T} v_{*i}^2 + \sum_{i=(m-1)T+1}^{mT} v_{*i}^2 \frac{1 + \eta\lambda_3(1-\rho)}{1 + \eta\lambda_3(1-\rho+m\rho)} \right]^{n-T-2} \right\}^{-\frac{1}{2}} \\
& \left\{ \left[\sum_{i=3}^{(m-1)T} v_{*i}^2 \frac{1 + \eta\lambda_3(1-\rho+m\rho)}{1 + \eta\lambda_3(1-\rho)} + \sum_{i=(m-1)T+1}^{mT} v_{*i}^2 \right]^{T} \right\}^{-\frac{1}{2}}. \qquad (3.43)
\end{aligned}
$$

Furthermore, the upper bound is less than

$$
\begin{aligned}
(3.43) \leq & (\lambda_3)^{-\frac{n-2}{2}} \left\{ \left[\sum_{i=3}^{(m-1)T} v_{*i}^2 \right]^{n-T-2} \right\}^{-\frac{1}{2}} \left\{ \left[\sum_{i=3}^{(m-1)T} v_{*i}^2 + \sum_{i=(m-1)T+1}^{mT} v_{*i}^2 \right]^{T} \right\}^{-\frac{1}{2}} \\
= & (\lambda_3)^{-\frac{n-2}{2}} \left(C_1^{n-T-2} C_2^{T} \right)^{-\frac{1}{2}}, \qquad (3.44)
\end{aligned}
$$

which is a constant given data $\boldsymbol{y}$.

Note that by (3.43),

$$
\begin{aligned}
& |\boldsymbol{I}_n + \eta \mathbf{R}\boldsymbol{Q}|^{-\frac{1}{2}} [\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1} + \eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}]^{1-\frac{n}{2}} \\
\leq & (\lambda_3)^{-\frac{n-2}{2}} \left\{ \left[\sum_{i=3}^{(m-1)T} v_{*i}^2 \right]^{n-T-2} \right\}^{-\frac{1}{2}} \left\{ \left[\sum_{i=3}^{(m-1)T} v_{*i}^2 \frac{\kappa_2}{\kappa_1} + \sum_{i=(m-1)T+1}^{mT} v_{*i}^2 \right]^{T} \right\}^{-\frac{1}{2}} \\
= & (\lambda_3)^{-\frac{n-2}{2}} C_1^{-\frac{n-T-2}{2}} \left[C_1 \frac{\kappa_2}{\kappa_1} + (C_2 - C_1) \right]^{-\frac{T}{2}}, \qquad (3.45)
\end{aligned}
$$

where $\kappa_1 = 1 + \eta\lambda_3(1-\rho)$, $\kappa_2 = 1 + \eta\lambda_3(1-\rho+m\rho)$.

Let

$$g(\eta,\rho) = \frac{1+\eta\lambda_3(1-\rho+m\rho)}{1+\eta\lambda_3(1-\rho)}, \tag{3.46}$$

$$h(\eta,\rho) = (\lambda_3)^{-\frac{n-2}{2}} C_1^{-\frac{n-T-2}{2}} \left[C_1 g(\eta,\rho) + (C_2 - C_1)\right]^{-\frac{T}{2}}. \tag{3.47}$$

Then

$$\frac{\partial g}{\partial \rho} = \frac{\eta\lambda_3(m-1)[1+\eta\lambda_3(1-\rho)] + \eta\lambda_3}{[1+\eta\lambda_3(1-\rho)]^2} > 0, \tag{3.48}$$

$$\frac{\partial g}{\partial \eta} = \frac{\lambda_3 m\rho}{[1+\eta\lambda_3(1-\rho)]^2} > 0. \tag{3.49}$$

When $\eta \to +\infty, \rho \to 1$, we have $g(\eta,\rho) \to +\infty$ and $h(\eta,\rho) \to 0$.

Thus, the likelihood of (η,ρ) is upper-bounded as in (3.44), but does not have positive lower bound.

Case 2. We consider the case that $n \neq mT$. There exists an integer $m \geq 2$, $1 \leq l \leq T-1$ such that $n = mT - l$. The eigenvalues of $\mathbf{R}$ defined in (3.5) have the close form as (see Appendix 3.10)

$$d_i = \begin{cases} 1-\rho, & i = 1,\ldots,(m-1)T-l, \\ 1-\rho+(m-1)\rho, & i = (m-1)T-l+1,\ldots,(m-1)T, \\ 1-\rho+m\rho, & i = (m-1)T+1,\ldots,n. \end{cases} \tag{3.50}$$

For eigenvalues of $\boldsymbol{H}$, the $n-2$ order principal submatrix of $\mathbf{R}$, by (3.29)

$$d_3^* = \cdots = d_{(m-1)T-l}^* = 1-\rho = d_3 = \cdots = d_{(m-1)T-l}. \tag{3.51}$$

By Lemma 3.10.11, the corresponding eigenvectors $\boldsymbol{\xi}_i, i = 3, \ldots, (m-1)T - l$ only depend on the first $(m-1)T-l-2 = n-T-2$ columns of $\boldsymbol{\Gamma}'\boldsymbol{P}$. $\boldsymbol{P}$ does not depend on ρ. (See Appendix 3.10.) Then the first $n-T-2$ columns of $\boldsymbol{U}$ also do not depend on parameter ρ. For $\boldsymbol{v}_* = \boldsymbol{U}'[\mathbf{0}_{(n-2)\times 2}\ \boldsymbol{I}_{n-2}]\boldsymbol{\Gamma}'\boldsymbol{y} = (v_{*3}, \ldots, v_{*n})$, the $(v_{*3}, \ldots, v_{*(m-1)T-l})$ is known given data.

Thus the upper bound is

$$
\begin{aligned}
|\boldsymbol{I}_n + \eta\mathbf{R}\boldsymbol{Q}|^{-\frac{1}{2}}\,[\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}]^{1-\frac{n}{2}} &\le \Big[\prod_{j=3}^{n}(1+\eta\lambda_3 d_j)\Big]^{-\frac{1}{2}}\Big[\lambda_3\sum_{i=3}^{n}\frac{v_{*i}^2}{1+\eta\lambda_3 d_i}\Big]^{1-\frac{n}{2}} \\
&= (\lambda_3)^{-\frac{n-2}{2}}\left\{[\kappa_1]^{n-T-2}[\kappa_3]^{l}[\kappa_2]^{T-l}\right\}^{-\frac{1}{2}} \\
&\quad \Big[\sum_{i=3}^{n-T}\frac{v_{*i}^2}{\kappa_1}+\sum_{i=n-T+1}^{(m-1)T}\frac{v_{*i}^2}{\kappa_3}+\sum_{i=(m-1)T+1}^{n}\frac{v_{*i}^2}{\kappa_2}\Big]^{-\frac{n-2}{2}} \\
&= (\lambda_3)^{-\frac{n-2}{2}}\Big[\sum_{i=3}^{n-T}v_{*i}^2+\sum_{i=n-T+1}^{(m-1)T}v_{*i}^2\frac{\kappa_1}{\kappa_3}+\sum_{i=(m-1)T+1}^{n}v_{*i}^2\frac{\kappa_1}{\kappa_2}\Big]^{-\frac{n-T-2}{2}} \\
&\quad \Big[\sum_{i=3}^{n-T}v_{*i}^2\frac{\kappa_3}{\kappa_1}+\sum_{i=n-T+1}^{(m-1)T}v_{*i}^2+\sum_{i=(m-1)T+1}^{n}v_{*i}^2\frac{\kappa_3}{\kappa_2}\Big]^{-\frac{l}{2}} \\
&\quad \Big[\sum_{i=3}^{n-T}v_{*i}^2\frac{\kappa_2}{\kappa_1}+\sum_{i=n-T+1}^{(m-1)T}v_{*i}^2\frac{\kappa_2}{\kappa_3}+\sum_{i=(m-1)T+1}^{n}v_{*i}^2\Big]^{-\frac{T-l}{2}}, \qquad (3.52)
\end{aligned}
$$

where $\kappa_1 = 1+\eta\lambda_3(1-\rho)$, $\kappa_2 = 1+\eta\lambda_3(1-\rho+m\rho)$, $\kappa_3 = 1+\eta\lambda_3(1-\rho+(m-1)\rho)$. Furthermore, the upper bound is less than

$$
\begin{aligned}
(3.52) &\le (\lambda_3)^{-\frac{n-2}{2}}\left[\sum_{i=3}^{n-T}v_{*i}^2\right]^{-\frac{n-T-2}{2}}\left[\frac{m-1}{m}\sum_{i=3}^{n}v_{*i}^2\right]^{-\frac{l}{2}}\left[\sum_{i=3}^{n}v_{*i}^2\right]^{-\frac{T-l}{2}} \\
&= (\lambda_3)^{-\frac{n-2}{2}}\Big(\frac{m-1}{m}\Big)^{-\frac{l}{2}}\left(C_1^{n-T-2}C_2^{T}\right)^{-\frac{1}{2}}, \qquad (3.53)
\end{aligned}
$$

which is a constant given data $\boldsymbol{y}$, where $C_1 = \sum_{i=3}^{(m-1)T-l} v_{*i}^2 = \sum_{i=3}^{n-T} v_{*i}^2 > 0$, and $C_2 = \boldsymbol{v}'\boldsymbol{v} = \sum_{i=3}^{n} v_{*i}^2 > 0$. The inequality in (3.53) is given by $\frac{1+\eta\lambda_3(1-\rho+(m-1)\rho)}{1+\eta\lambda_3(1-\rho+m\rho)} \ge$

$\frac{m-1}{m}$ as follows. Let

$$g_*(\eta, \rho) = \frac{1 + \eta\lambda_3(1 - \rho + (m-1)\rho)}{1 + \eta\lambda_3(1 - \rho + m\rho)}. \tag{3.54}$$

Then

$$\frac{\partial h}{\partial \rho} = \frac{-\eta\lambda_3 - \eta^2\lambda_3^2}{[1 + \eta\lambda_3(1 - \rho + m\rho)]^2} < 0, \tag{3.55}$$

$$\frac{\partial h}{\partial \eta} = \frac{-\lambda_3\rho}{[1 + \eta\lambda_3(1 - \rho + m\rho)]^2} < 0. \tag{3.56}$$

Thus $g_*(\eta, \rho) \geq \lim_{\eta \to +\infty, \rho \to 1} g_*(\eta, \rho) = \frac{m-1}{m}$.

Similar to Case 1, the upper bound (3.52) goes to 0 as (η, ρ) approaches $(+\infty, 1)$. Thus the likelihood of (η, ρ) is upper-bounded as in (3.53), but does not have positive lower bound.

Therefore for both cases, the likelihood of (ρ, η) is upper-bounded by some constant. If the priors for ρ, η are proper, then the joint posterior will be proper as well.

□

3.3 Calculation and Sampling method

Based on the posterior distributions derived in the previous section, we draw posterior samples and obtain the trend estimates $\boldsymbol{z}$ with Algorithm 2 as follows.

3.3.1 Simplify the matrix calculation

We want to simplify the calculation because a large amount of numerical matrix inversion is time consuming and will lower the accuracy.

Algorithm 2 Posterior sampling of density (3.18)

1: Given data $\boldsymbol{y}$, sample (ρ, η) from density (3.22).

2: Given $(\rho, \eta; \boldsymbol{y})$, sample δ_0 from distribution (3.21).

3: Given $(\rho, \eta, \delta_0; \boldsymbol{y})$, sample $\boldsymbol{z}$ from distribution (3.20).

Note that both determinant and inverse for n dimensional matrix $\mathbf{R}^{-1} + \eta \boldsymbol{Q}$ is needed in (3.20), (3.21), and (3.22). Given $\boldsymbol{Q} = \boldsymbol{F}_0' \boldsymbol{F}_1^{-1} \boldsymbol{F}_0$,
by the matrix determinant lemma,

$$\left|\mathbf{R}^{-1} + \eta \boldsymbol{Q}\right| = \eta^{n-2} \frac{\left|\eta^{-1} \boldsymbol{F}_1 + \boldsymbol{F}_0 \mathbf{R} \boldsymbol{F}_0'\right|}{\left|\boldsymbol{F}_1\right| \left|\mathbf{R}\right|}, \tag{3.57}$$

and by the Sherman–Morrison–Woodbury formula,

$$(\mathbf{R}^{-1} + \eta \boldsymbol{Q})^{-1} = \mathbf{R} - \mathbf{R} \boldsymbol{F}_0' (\eta^{-1} \boldsymbol{F}_1 + \boldsymbol{F}_0 \mathbf{R} \boldsymbol{F}_0')^{-1} \boldsymbol{F}_0 \mathbf{R}. \tag{3.58}$$

Thus

$$\left|\mathbf{R}\right|^{-\frac{1}{2}} \left|\mathbf{R}^{-1} + \eta \boldsymbol{Q}\right|^{-\frac{1}{2}} = \eta^{1-\frac{n}{2}} \left[\frac{\left|\boldsymbol{F}_1\right|}{\left|\eta^{-1} \boldsymbol{F}_1 + \boldsymbol{F}_0 \mathbf{R} \boldsymbol{F}_0'\right|}\right]^{\frac{1}{2}}, \tag{3.59}$$

$$(\mathbf{R}^{-1} + \eta \boldsymbol{Q})^{-1} \mathbf{R}^{-1} = I_n - \mathbf{R} \boldsymbol{F}_0' (\eta^{-1} \boldsymbol{F}_1 + \boldsymbol{F}_0 \mathbf{R} \boldsymbol{F}_0')^{-1} \boldsymbol{F}_0. \tag{3.60}$$

The $n \times n$ Correlation matrix $\mathbf{R}$ is defined in (3.5), and nonsingular matrices $\boldsymbol{F}_0, \boldsymbol{F}_1$ are defined in (2.3). By using (3.59) and (3.60), we reduce the numbers of matrix inverse calculation and compute the inverse for a $n-2$ dimensional matrix instead of a n dimensional matrix.

3.3.2 Sample (ρ, η)

We choose prior for ρ as

$$[\rho] \propto (1-\rho)^{-\frac{1}{2}}, \tag{3.61}$$

and use the multivariate ratio-of-uniform method (Wakefield, Gelfand, and Smith 1991) to sample from the marginal joint posterior density,

$$[\rho, \eta|\boldsymbol{y}] \propto f(\rho,\eta) = \frac{(1-\rho)^{-\frac{1}{2}}}{|\mathbf{R}|^{\frac{1}{2}}\left|\mathbf{R}^{-1}+\eta\boldsymbol{Q}\right|^{\frac{1}{2}}(c+\eta)^2}[\boldsymbol{y}'\boldsymbol{Q}(\mathbf{R}^{-1}+\eta\boldsymbol{Q})^{-1}\mathbf{R}^{-1}\boldsymbol{y}]^{1-\frac{n}{2}}. \tag{3.62}$$

Based on equations (3.59) and (3.60),

$$\begin{aligned} log f(\rho,\eta) &= -2log(c+\eta) - \frac{1}{2}log(1-\rho) + \frac{1}{2}[log\,|\boldsymbol{F}_1| - log\left|\eta^{-1}\boldsymbol{F}_1 + \boldsymbol{F}_0\mathbf{R}\boldsymbol{F}_0'\right|] \\ &\quad +(1-\frac{n}{2})log(\boldsymbol{y}'\boldsymbol{Q}[I_n - \mathbf{R}\mathbf{F}_\mathbf{0}'(\eta^{-\mathbf{1}}\mathbf{F_1} + \mathbf{F_0}\mathbf{R}\mathbf{F_0}')^{-\mathbf{1}}\mathbf{F_0}]\mathbf{y}). \end{aligned} \tag{3.63}$$

Note that if $(u, \boldsymbol{v})$ are uniform on

$$\mathcal{A} = \left\{(u,\boldsymbol{v}) : \boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2, 0 < u < [f(\boldsymbol{\mu} + \boldsymbol{v}/u^r)]^{\frac{1}{2r+1}}\right\}, \tag{3.64}$$

for some $\boldsymbol{\mu} = (\mu_1, \mu_2) \in \mathbb{R}^2$, then $\boldsymbol{\theta} = \boldsymbol{\mu} + \boldsymbol{v}/u^r$ has the density $p(\rho, \eta|\boldsymbol{y}) \propto f(\rho, \eta)$.

We first construct the minimal bounding rectangle for uniform sampling

$$
\begin{aligned}
a^+ &= \sup\left([f(\rho,\eta)]^{\frac{1}{2r+1}}\right), && (3.65)\\
b_1^+ &= \sup\left((\rho-\mu_1)[f(\rho,\eta)]^{\frac{r}{2r+1}}\right), && (3.66)\\
b_1^- &= \inf\left((\rho-\mu_1)[f(\rho,\eta)]^{\frac{r}{2r+1}}\right) = -\sup\left((\mu_1-\rho)[f(\rho,\eta)]^{\frac{r}{2r+1}}\right), && (3.67)\\
b_2^+ &= \sup\left((\eta-\mu_2)[f(\rho,\eta)]^{\frac{r}{2r+1}}\right), && (3.68)\\
b_2^- &= \inf\left((\eta-\mu_2)[f(\rho,\eta)]^{\frac{r}{2r+1}}\right) = -\sup\left((\mu_2-\eta)[f(\rho,\eta)]^{\frac{r}{2r+1}}\right). && (3.69)
\end{aligned}
$$

We take $r = 1$, which yields the standard ratio-of uniform method.

The sampling algorithm is as follows.

Algorithm 3 Bivariate ration-of-uniform sampling of ρ, η

1: Independently sampling u from $Unif(0, a^+)$, v_1 from $Unif(b_1^-, b_1^+)$, v_2 from $Unif(b_2^-, b_2^+)$.

2: If $0 < \mu_1+v_1/u^r < 1$, $\mu_2+v_2/u^r > 0$, and $0 < u < (f(\mu_1+v_1/u^r, \mu_2+v_2/u^r))^{\frac{1}{2r+1}}$, report $(\rho, \eta) = (\mu_1 + v_1/u^r, \mu_2 + v_2/u^r)$; otherwise return to Step 1.

3.3.2.1 Computing the minimal bounding rectangle without relocation

We take $(\mu_1, \mu_2) = (0, 0)$. As $f(\rho, \eta)$ is bounded over the domain of (ρ, η), $b_1^- = b_2^- = 0$. We compute a^+, b_1^+, b_2^+ as follows.

(1) Calculation the mode of $logf(\rho, \eta)$ as (ρ^*, η^*).

(2) $a^+ = [f(\rho^*, \eta^*)]^{\frac{1}{2r+1}}$.

(3) Numerically search $b_1^{*+} = \sup_{\rho \geq \rho^*} \left[log\rho + \frac{r}{2r+1} logf(\rho, \eta)\right]$ from initial value (ρ^*, η^*).

(4) Numerically search $b_2^{*+} = \sup_{\eta \geq \eta^*} \left[log\eta + \frac{r}{2r+1} logf(\rho, \eta)\right]$ from initial value (ρ^*, η^*).

(5) $b_1^+ = exp(b_1^{*+})$, $b_2^+ = exp(b_2^{*+})$.

We take the unemployment level data from 01/2004 to 12/2008 for example. The acceptance rate is 4.971%, and the accepted points are plotted in Figure 3.5.

3.3.2.2 Computing the minimal bounding rectangle with relocation

We take $(\mu_1, \mu_2) = (\rho^*, \eta^*)$, where (ρ^*, η^*) is the mode of $f(\rho, \eta)$. We compute $a^+, b_1^+, b_1^-, b_2^+, b_2$ as follows.

(1) Calculation the mode of $logf(\rho, \eta)$ as (ρ^*, η^*).

(2) $a^+ = [f(\rho^*, \eta^*)]^{\frac{1}{2r+1}}$.

(3) Numerically search $b_1^{*+} = \sup_{\rho>\rho^*} \left[log(\rho - \rho^*) + \frac{r}{2r+1} logf(\rho, \eta)\right]$ from initial value $((\rho^* + 1)/2, \eta^*)$.

(4) Numerically search $b_1^{*-} = \sup_{\rho<\rho^*} \left[log(\rho^* - \rho) + \frac{r}{2r+1} logf(\rho, \eta)\right]$ from initial value $(\rho^*/2, \eta^*)$.

(5) Numerically search $b_2^{*+} = \sup_{\eta>\eta^*} \left[log(\eta - \eta^*) + \frac{r}{2r+1} logf(\rho, \eta)\right]$ from initial value $(\rho^*, \eta^* * 2)$.

(6) Numerically search $b_2^{*-} = \sup_{\eta<\eta^*} \left[log(\eta^* - \eta) + \frac{r}{2r+1} logf(\rho, \eta)\right]$ from initial value $(\rho^*, \eta^*/2)$.

(7) $b_1^+ = exp(b_1^{*+})$, $b_1^- = -exp(b_1^{*-})$, $b_2^+ = exp(b_2^{*+})$, $b_2^- = -exp(b_2^{*-})$.

We take the unemployment level data from 01/2004 to 12/2008 for example, just as in the previous section. The acceptance rate rises to 36.34%, and the accepted points are plotted in Figure 3.6.

We have shown that the sampling with relocation is much more efficient than that without relocation. Then we further explore the quality of the sampling.

In Figure 3.7, we show the contour plot of the joint marginal posterior density $f(\rho, \eta)$, compared with the 2-d histogram of the posterior samples of (ρ, η) with relocation and those without relocation. We use the same dataset as in the previous section, the unemployment level data from 01/2004 to 12/2008. 10,000 samples are taken for each sampling procedure.

Both sampling methods work well and match the contour plot of density. The samples with relocation contain fewer extreme values. Considering both the performance and efficiency, we recommend to choose the ratio-of-uniform sampling with relocation.

3.4 Trend Estimation of Unemployment Level

We use the public data of the unadjusted unemployment level from 01/2004 to 12/2018. The length of the time series is 180 months. Following the algorithm above, we draw 10,000 samples from the joint posterior distribution.

We show the posterior mean of $\boldsymbol{z}$ as the trend estimates with the 95% credible intervals in Figure 3.8. Compared with the X-13 trend, our Bayesian smoothing spline estimation throughly removes the short-term fluctuations.

3.4.1 Comparison with Basic Model

We compare the posterior coefficient of variation (CV) between the generalized Bayesian smoothing spline model and the basic model in Figure 3.9. Note that the posterior CV is directly related to the divergence of data $\boldsymbol{y}$ from the trend estimates $\boldsymbol{z}$. Thus the posterior CV are similar at the majority of knots. However, our generalized Bayesian smoothing spline model performs much more precisely at the boundaries than the

basic model.

3.4.2 Stability

The unemployment level and other labor force data are collected monthly by the Census Bureau, and we may obtain new trend estimates for the previous time points each time when new data are available. Our question is whether the current trend estimates will coincide with the previous one.

As previously discussed, because we have to borrow information from all the series to estimate the trend at each time point, adding new data will inevitably affect the previous estimates. The current X-13 method also has this issue, and it is a common practice to revise the estimates after new data are collected.

In this section, we study the stability in terms of adding new data of Bayesian smoothing spline trend estimation, and compare it with X-13 method and the basic model.

In Figures 3.10, 3.11 and 3.12, we denote the non-revision trend estimates as lag 0. For lag 0, we first perform analysis with 01/2004 - 12/2008 data, and record the result for the time point of 12/2008 only. Then we perform analysis with 02/2004 - 01/2009 data, and record the result for the time point of 01/2009 only, and so on.

The 3-month revised trend is denoted as lag 3. For lag 3, we first perform analysis with 04/2004 - 03/2009 data, and record the result for the time point of 12/2008 only. Then we perform analysis with 05/2004 - 04/2009 data, and record the result for the time point of 01/2009 only, and so on.

Similarly, we construct the 6-month, 12-month, 24-month and 36-month revised trends, which are shown as lag 6, lag 12, lag 24 and lag 36 respectively.

3.5 Alternative of matrix Q

Following Speckman and Sun (2003), another choice of matrix $\boldsymbol{Q}$ is constructed by the approximation of the smoothing penalty term in (3.7) as below.

For cubic spline $p = 2$, and equal spaced knots at $t = t_1, t_2, ..., t_n$, $h = t_k - t_{k-1}$, $f^{(2)}(t_k) \approx h^{-2}\nabla^2 f(t_k)$, where

$$\nabla^2 f(t_k) = \sum_{i=0}^{2} (-1)^i C_i^2 f(t_{k-i}) = z_{k-2} - 2z_{k-1} + z_k$$

is the second order backward difference operator. Then the penalty structure is approximately

$$\int [f^{(2)}(t)]^2 dt \approx \sum_{k=3}^{n} h[h^{-2}\nabla^2 f(t_k)]^2 = h^{-3} \sum_{k=3}^{n} [\nabla^2 f(t_k)]^2. \tag{3.70}$$

Let $z_k = f(t_k)$, (3.70) can be written as $\boldsymbol{z}'\boldsymbol{Q}\boldsymbol{z} = (\boldsymbol{F}_0\boldsymbol{z})'(\boldsymbol{F}_0\boldsymbol{z})$, where F_0 is defined in (3.9), and

$$\boldsymbol{F}_0\boldsymbol{z} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}_{(n-2)\times n} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} \vdots \\ \nabla^2 z_k \\ \vdots \end{pmatrix}_{n-2}.$$

Thus $\boldsymbol{Q} = \boldsymbol{F}_0'\boldsymbol{F}_0$, which is an $n \times n$ second order operator matrix of rank $n - 2$.

We compare the performance of $\boldsymbol{Q}_1 = \boldsymbol{F}_0'\boldsymbol{F}_1^{-1}\boldsymbol{F}_0$ and $\boldsymbol{Q}_2 = \boldsymbol{F}_0'\boldsymbol{F}_0$.

Figures 3.13 and 3.14 compare the posterior samples of η and ρ. There is no obvious difference for η and ρ between different choice of $\boldsymbol{Q}$.

Figure 3.15 shows the trend estimates with different $\boldsymbol{Q}$, and they are almost identical. Figure 3.16 further shows the relative difference between the trend estimates

with different $\boldsymbol{Q}$, which is mostly under 0.2%. The relative difference is defined as $(Trend\ 1 - Trend\ 2)\ /\ Trend\ 1$.

Figure 3.17 values the uncertainty of the trend estimates with different $\boldsymbol{Q}$. The model with $\boldsymbol{Q}_1 = \boldsymbol{F}_0'\boldsymbol{F}_1^{-1}\boldsymbol{F}_0$ performs a little more precisely in terms of posterior coefficient of variation. Thus we prefer to use $\boldsymbol{Q} = \boldsymbol{F}_0'\boldsymbol{F}_1^{-1}\boldsymbol{F}_0$ for the approximation of the penalty term.

3.6 Model comparison with DIC

We compare models using the deviance information criterion (DIC). Following Gelman et al. (2013),

$$DIC = -2logp(y|\hat{\boldsymbol{\theta}}_{Bayes}) + 2p_{DIC},$$

where, with the posterior mean $\hat{\boldsymbol{\theta}}_{Bayes}$, the effective number of parameters p_{DIC} is defined as:

$$p_{DIC} = 2\left(logp(y|\hat{\boldsymbol{\theta}}_{Bayes}) - E_{post}(logp(y|\boldsymbol{\theta}))\right). \tag{3.71}$$

Given the posterior samples $\boldsymbol{\theta}^s$, $s = 1, ..., S$, (3.71) is computed as:

$$p_{DIC} = 2\left(logp(y|\hat{\boldsymbol{\theta}}_{Bayes}) - \frac{1}{S}\sum_{s=1}^{S}(logp(y|\boldsymbol{\theta}^s))\right). \tag{3.72}$$

We use data from 01/2004 to 12/2018. From model (3.4), given the log-likelihood

$$logp(\boldsymbol{y}|\boldsymbol{z}, \delta_0, \rho) = -\frac{n}{2}log(2\pi) - \frac{1}{2}log(|\delta_0\mathbf{R}|) - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{z})'(\delta_0\mathbf{R})^{-1}(\boldsymbol{y}-\boldsymbol{z}). \tag{3.73}$$

Table 3.1: DIC with different h, $\boldsymbol{Q}$

	$\boldsymbol{Q} = \boldsymbol{Q}_1$	$\boldsymbol{Q} = \boldsymbol{Q}_2$
h=1	2441.239	2437.465
h=1/n	2441.77	2438.826

Table 3.2: DIC with different length, $\boldsymbol{Q}$

Length of data	36	48	60	120	180
$\boldsymbol{Q} = \boldsymbol{Q}_1$	549.5731	706.8194	831.9119	1643.99	2441.77
$\boldsymbol{Q} = \boldsymbol{Q}_2$	550.0193	707.0836	831.6877	1640.148	2438.826

we compute the $DIC_{new} = -668.1519$. Compared with the basic model (3.3), $DIC_{basic} = -329.2693$, our generalized Bayesian Smoothing Spline model with dependency is preferred.

3.7 Simulation Study

In this section, we demonstrate the accuracy and precision of our BSSD trend estimation through three simulation studies and compare it with the X-13 Program.

3.7.1 Simulation 1

To generate data, we first specify a "true" trend m_t, and then apply this trend to the basic structure model (Pfeffermann 1991),

$$y_t = m_t + s_t + e_t, e_t \sim N(0, \sigma_1^2),\ t = 1, \ldots, n, \tag{3.74}$$

where the seasonal component s_t follows

$$s_t = -\sum_{j=t-12+1}^{t-1} s_j + \epsilon_t,\ \epsilon_t \sim N(0, \sigma_2^2). \tag{3.75}$$

We use two trends with different shapes. For each trend, 1000 series are generated. After we obtain the estimates of trend $\hat{m}_t$ at knot t for every series, we evaluate the precision by calculating the mean squared deviation (MSD) for each model M as,

$$MSE_t^M = \frac{1}{1000} \sum_{iter=1}^{1000} (\hat{m}_t^M - m_t)^2. \tag{3.76}$$

Trend 1. The true trend is set as:

$$m_t = 9000exp\left\{-\frac{(t-20)^2}{2\times 10^2}\right\} + 6000. \tag{3.77}$$

We take $t = 1, ..., 60$, $\sigma_1^2 = 4096$, $\sigma_2^2 = 40000$, and we generate a monthly series of 5 years through (3.77),(3.74),(3.75). We denote the time as 01/2020,...,12/2024. Note that the time label will not affect our Bayesian smoothing spline model, but it is required by the X-13 program. The X-13 trends are obtained by the R Package Seasonalview (Sax 2017). For our BSSD method, the prior (3.14) for ρ is used, and we select the hyper-parameter for the prior of η as $c = 50$.

In Figure 3.18, we show the simulated data, the true trend, and our Bayesian smoothing spline trend estimate, compared with the X-13 trend estimate for one of the series. To see the overall performance over 1000 simulations, we first show the relative bias in Figure 3.19. At each knot, the relative bias is defined as $\frac{1}{1000}\sum_{i=1}^{1000}\frac{E_i-T}{T}$. We also compare the relative standard deviation of the trend estimates between the two methods in Figure 3.20. The variation of our BSSD method is uniformly smaller than that of X-13.

We further show the relative root mean squared deviation of the trend estimates between the two method in Figure 3.21. Compared with the X-13 method, our BSSD is much more accurate at the boundaries. Also at the majority knots, our generalized

Bayesian smoothing spline with dependency model uniformly outperforms the X-13 method.

3.7.2 Simulation 2

In this section, we use trigonometric functions to construct seasonal patterns, instead of (3.75).

$$s_t = \sum_{k=1}^{K} A_k sin(2\pi \frac{k}{T} t - \mu_k),\ \mu_k \overset{\text{iid}}{\sim} unif(0, 2\pi). \tag{3.78}$$

Given the period $T = 12$, in order to create a harmonic series, we take $K = T/2 = 6$. The amplitudes are selected as $A_1 = 1000$, $A_k \sim N(0, (\frac{k}{K} A_1)^2), k = 2, \ldots, K$. The data are still generated from the additive model (3.74) for 1000 times. For each time, the μ_k and A_k are randomly selected, and we take $t = 1, ..., 60$, $\sigma_1^2 = 10000$ to simulate a series of length 60.

Trend 2. The true trend is set as,

$$m_t = 5000 sin(\frac{\pi}{60} t) + 10000. \tag{3.79}$$

Figure 3.22 shows one simulated series with the true trend and trend estimates. Note that for different series, the simulated seasonal pattern will be different.

After we obtain the trend estimates for 1000 simulated series. We compare the relative bias of the average of 1000 trend estimates from the true trend for the two methods: BSSD and X-13. Generally our model outperforms the X-13 method.

We also compare the relative standard deviation of the trend estimates between two methods in Figure 3.24. The variation of our BSSD method is uniformly smaller than that of X-13.

In Figure 3.26, we compare the relative root mean squared deviation of the trend estimates between two methods. Note that the majority of the RMSD comes from the variation. Our BSSD model outperforms X-13 uniformly.

3.7.3 Simulation 3

In this section, we use trigonometric functions to construct seasonal patterns,

$$s_t = \sum_{k=1}^{K} A_k sin(2\pi \frac{k}{T} t - \mu_k). \tag{3.80}$$

Given the period $T = 12$, in order to create a harmonic series, we take $K = T/2 = 6$. The amplitudes are selected as $A_k \sim N(0, 160000), k = 1, \ldots, K$. The phase shift in each function is selected as $\mu_k \overset{\text{iid}}{\sim} unif(0, 2\pi), k = 1, \ldots, K$.

After the seasonal pattern s_t been chosen, we generate data of length 60 from the additive model

$$y_t = m_t + s_t + e_t, \tag{3.81}$$

$$m_t = 5000 sin(\frac{\pi}{60} t) + 10000. \tag{3.82}$$

for 500 times, where $e_t \sim N(0, 40000)$.

Figure 3.25 shows the seasonal pattern, one simulated series with the true trend and trend estimates. Note that the true trend and seasonal pattern remain the same in all 500 simulation series.

After we obtain the trend estimates for 500 simulated series, we compare the relative bias of the average of 500 trend estimates from the true trend for the two methods: BSSD and X-13. Generally our model outperforms the X-13 method.

We also compare the relative standard deviation of the trend estimates between

the two methods in Figure 3.28. The variation of our BSSD method is uniformly smaller than that of X-13.

In Figure 3.29, we compare the relative root mean squared deviation of the trend estimates between two methods. Note that the majority of the RMSD comes from the variation. and our BSSD model outperforms X-13 uniformly.

Sometimes, we are interested in obtaining the one number summary to quantify and compare the performance of the BSSD and X-13 method. One way is to use the Frechet distance, which calculates the farthest that the curves separate. However, in the simulation studies, even though we observe that the trend estimates at the two end points have large uncertainty, there is no guarantee and it is obviously not the case that the largest bias occurs at the same single point for 1000 times. Thus we obtain the maximum deviation of the average bias or, to remove the effect of scale, use the maximum deviation of the average relative bias (MDRB). For the Simulation 3, the MDRB for the BSSD trend is 0.00315, whereas the MDRB for the X-13 trend is as large as 0.01497, almost three times greater than the deviation of the BSSD method. Another way is to compute the sum of squares of the relative bias (SSRB) for all knots. The SSRB for the BSSD trend is 2.57×10^{-5}, whereas the SSRB for the X-13 trend is as large as 1.49×10^{-3}, 58 times greater than the deviation of the BSSD method.

3.8 Comments

In this chapter, we generalize the Bayesian smoothing spline model with dependence structure. The correlation matrix as the dependence structure, which is specially designed for the seasonal variation, is easily interpreted and computationally efficient. This generalization significantly improves the boundary performance and elevates the

overall accuracy and precision by borrowing information from different cycles.

We develop an efficient MC sampling for the posterior distribution. The bivariate ratio-of-uniform method with relocation is preferred, with the sampling performance compared in Figure 3.5 and 3.6. The MC sampling method we propose can be naturally programed by parallel computing. In each step of the Algorithm 2, the parameters can be sampled simultaneously. It saves tremendous time of computation to process long series. Besides efficiency, the MC method draws samples from the exact posterior distribution, and avoids the convergence issue in MCMC sampling.

We conduct simulation studies with different types of trends and seasonal variation components, and our Bayesian smoothing spline with dependency model outperforms the X-13 method in terms of both accuracy and precision.

3.9 Figures

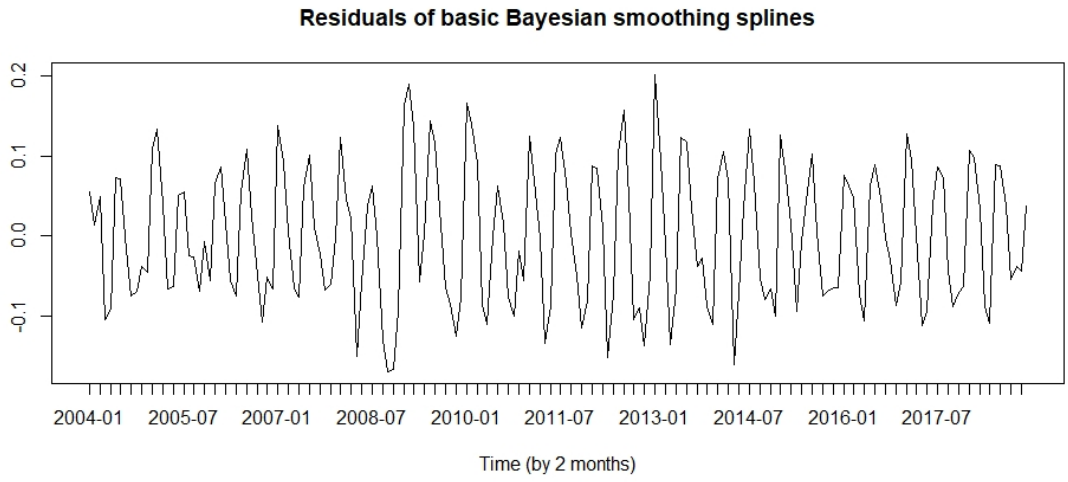

Figure 3.1: Relative residuals from 01/2004 to 12/2018

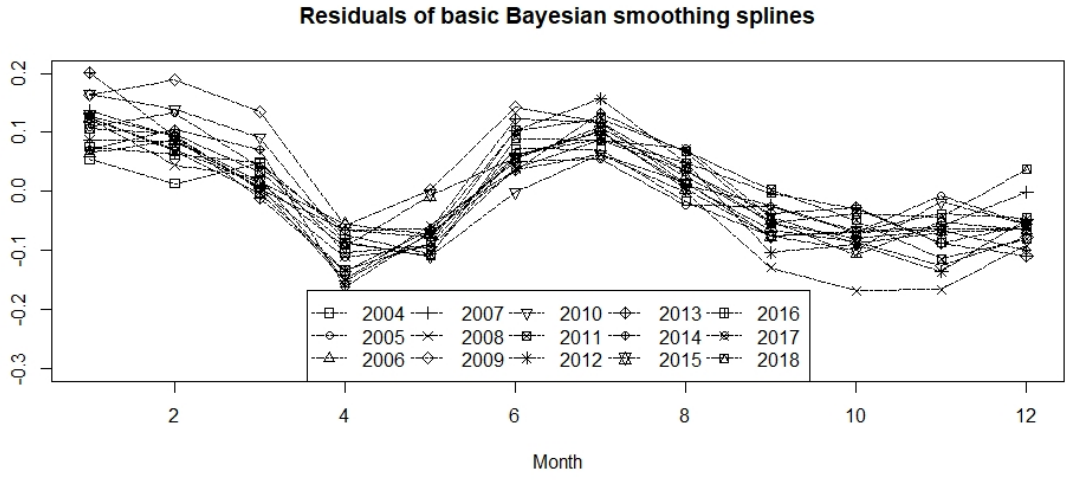

Figure 3.2: Residuals by month

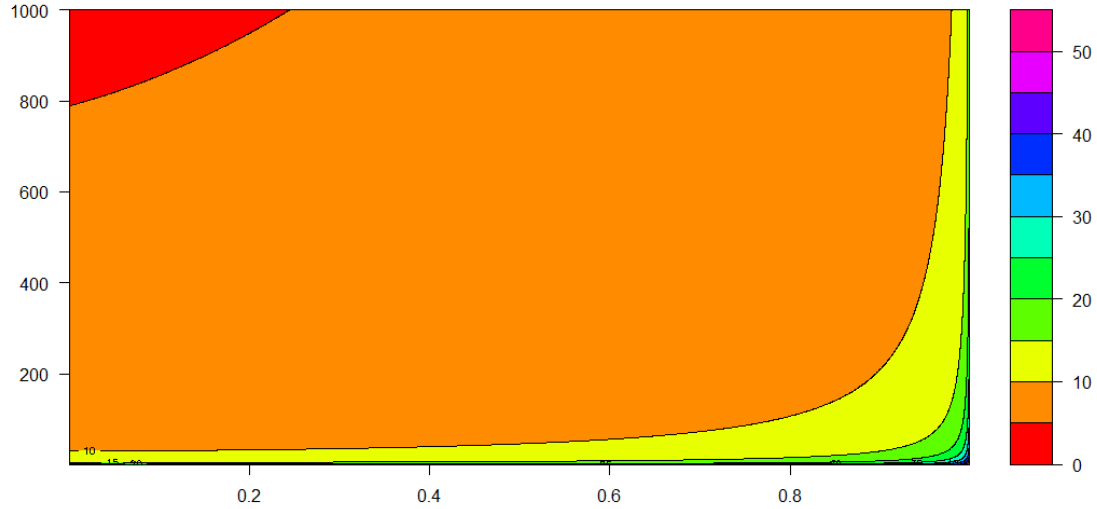

Figure 3.3: The Contour Plot of Effective Degrees of Freedom

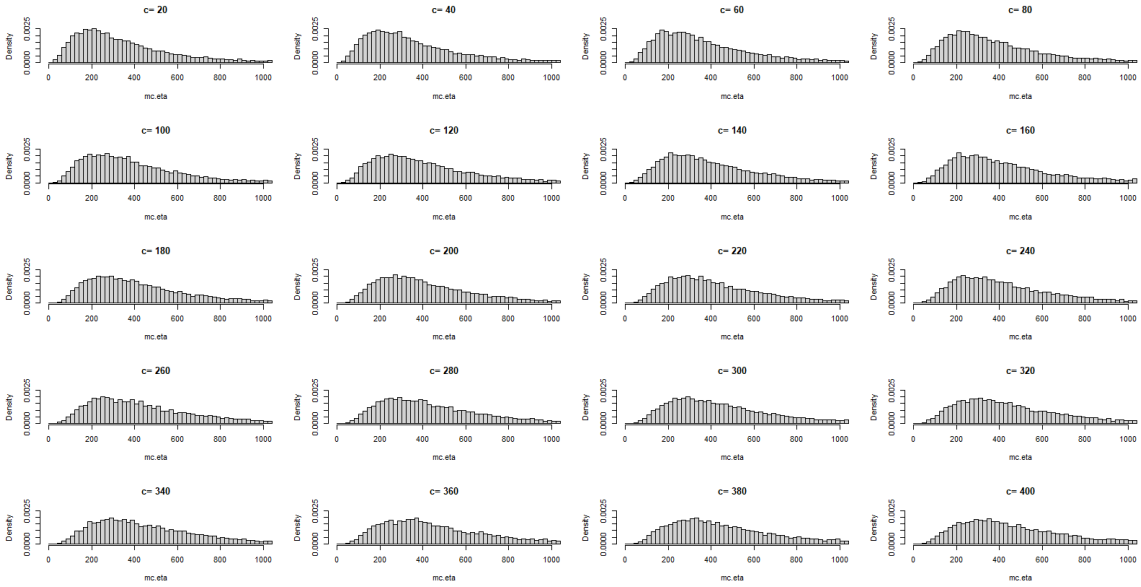

Figure 3.4: The Histogram of Posterior Samples of η

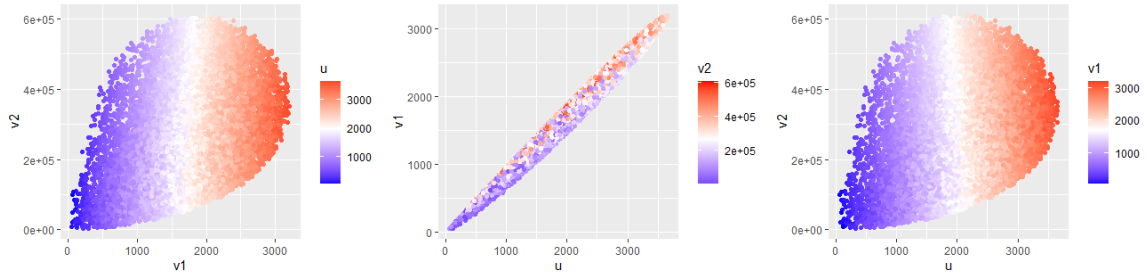

Figure 3.5: Scatter plot of (v_1, v_2), (u, v_1) and (u, v_2) without relocation

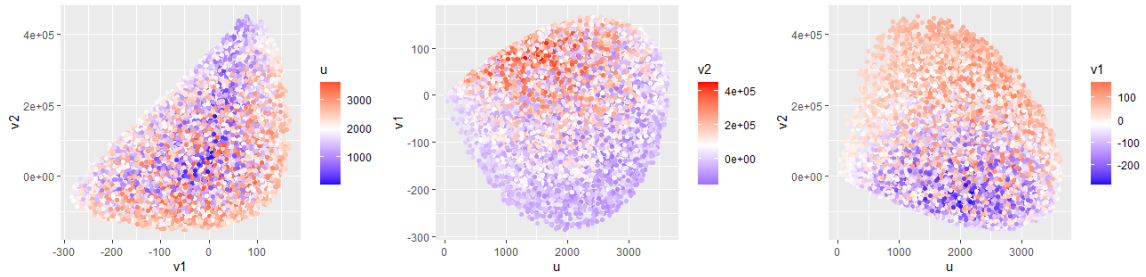

Figure 3.6: Scatter plot of (v_1, v_2), (u, v_1) and (u, v_2) with relocation

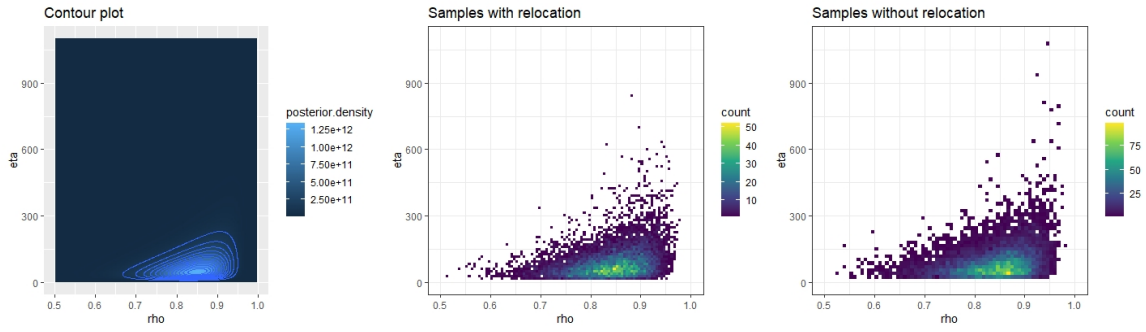

Figure 3.7: Contour plot of posterior density $f(\rho, \eta)$, 2-d histograms of the posterior samples of (ρ, η) with and without relocation

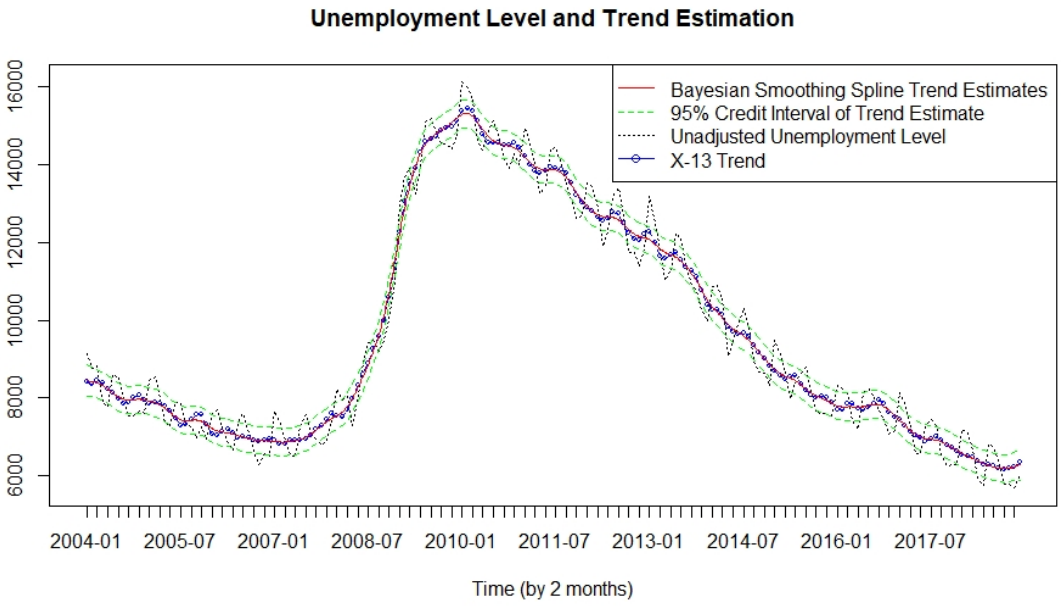

Figure 3.8: Bayesian Smoothing Spline Trend with 95% credible interval Compared with X-13 Trend

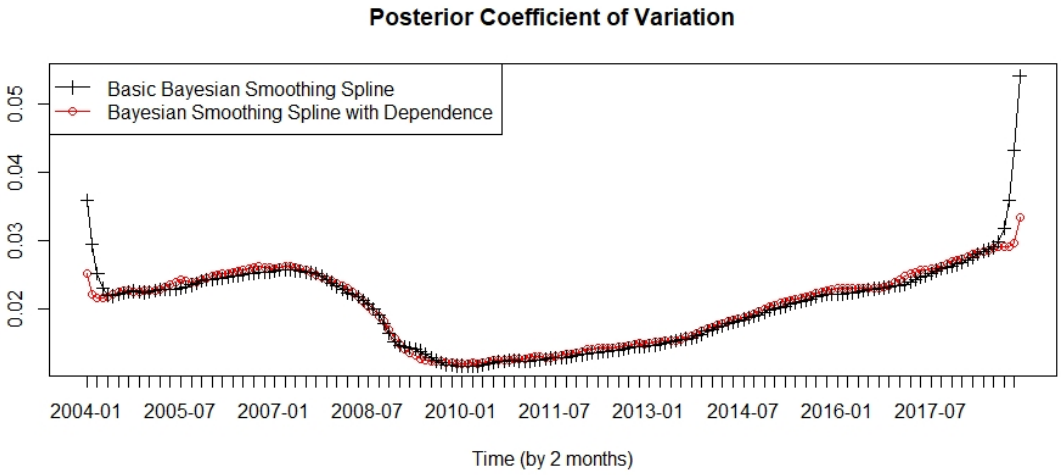

Figure 3.9: Posterior Coefficient of Variation of Trend Estimate

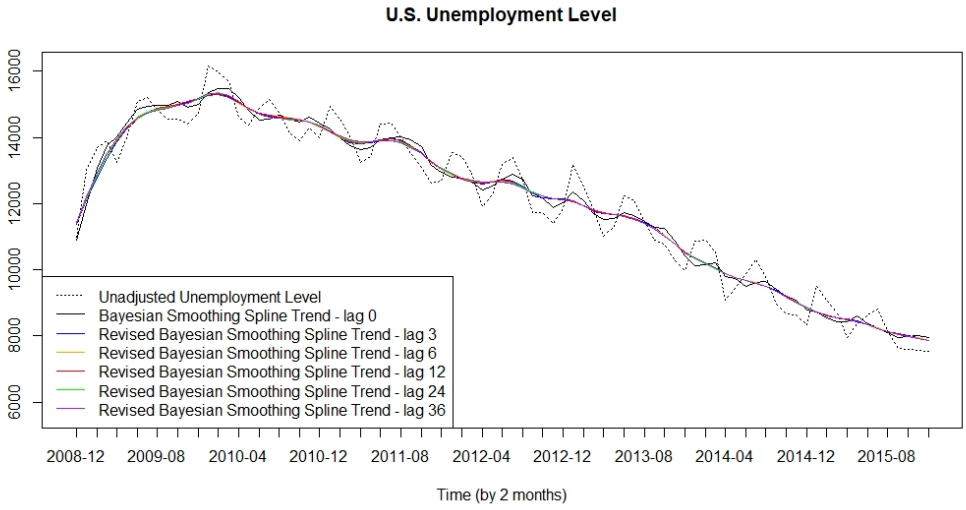

Figure 3.10: Comparison of Revised Trend Estimates of Bayesian Smoothing Spline with Dependency

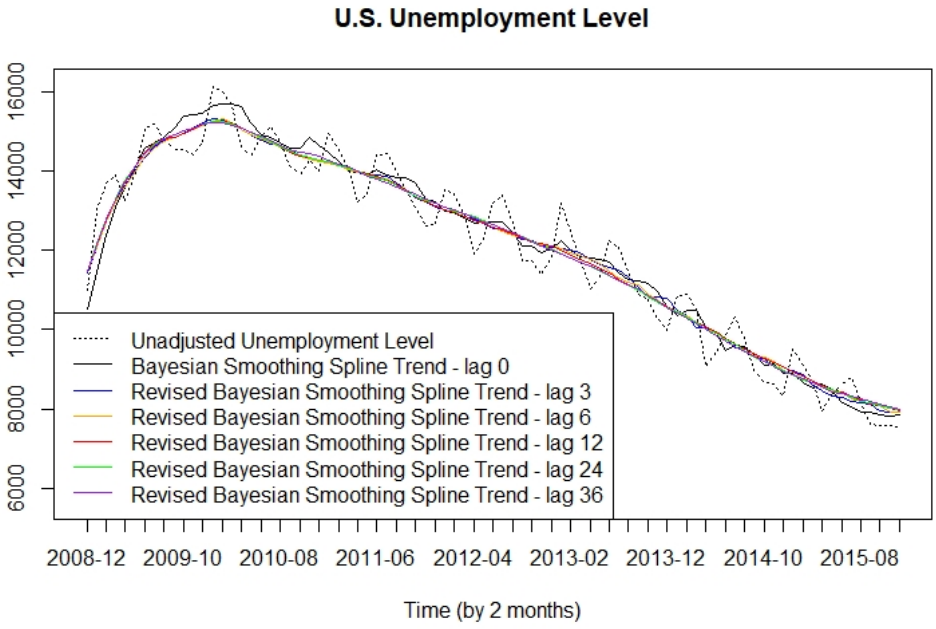

Figure 3.11: Comparison of Revised Trend Estimates of Basic Bayesian Smoothing Spline

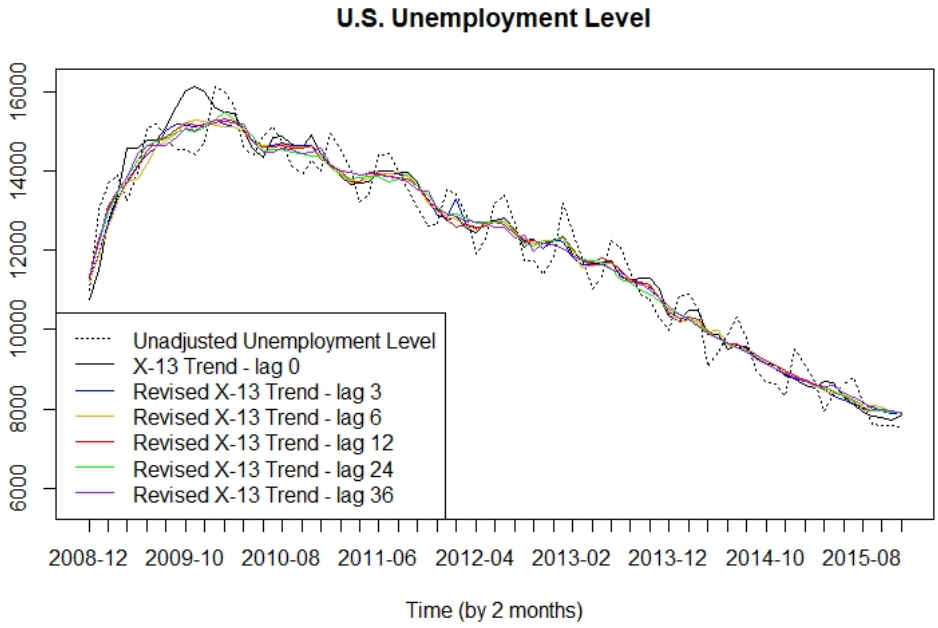

Figure 3.12: Comparison of Revised X-13 Trend Estimates

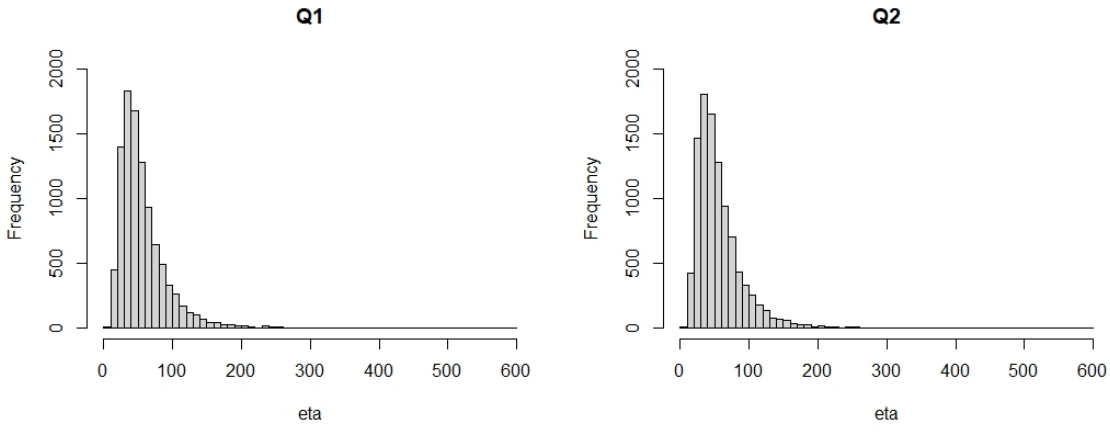

Figure 3.13: Histogram of posterior samples η with different $\boldsymbol{Q}$

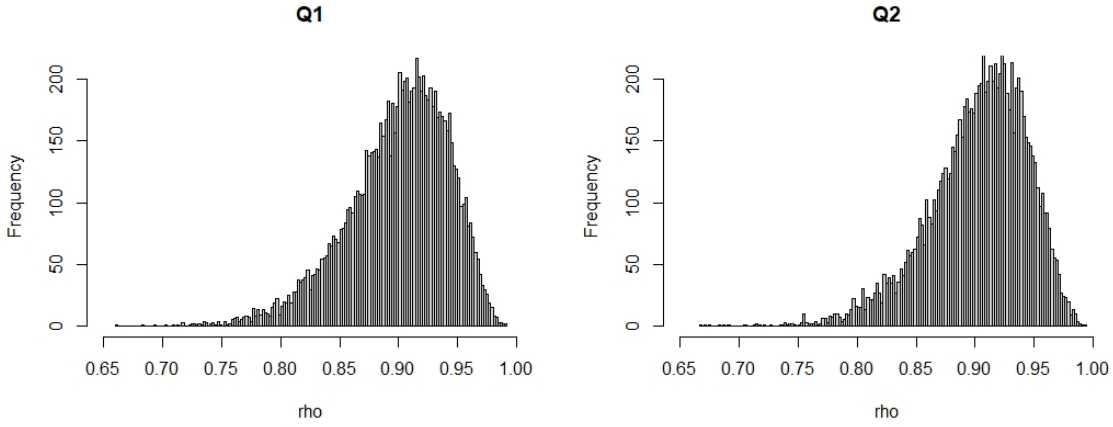

Figure 3.14: Histogram of posterior samples ρ with different $\boldsymbol{Q}$

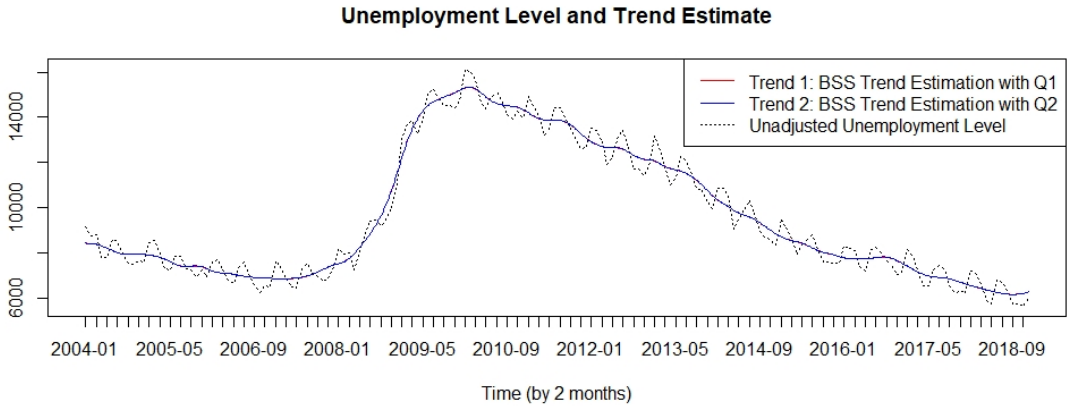

Figure 3.15: Comparison of Trend Estimates with different $\boldsymbol{Q}$

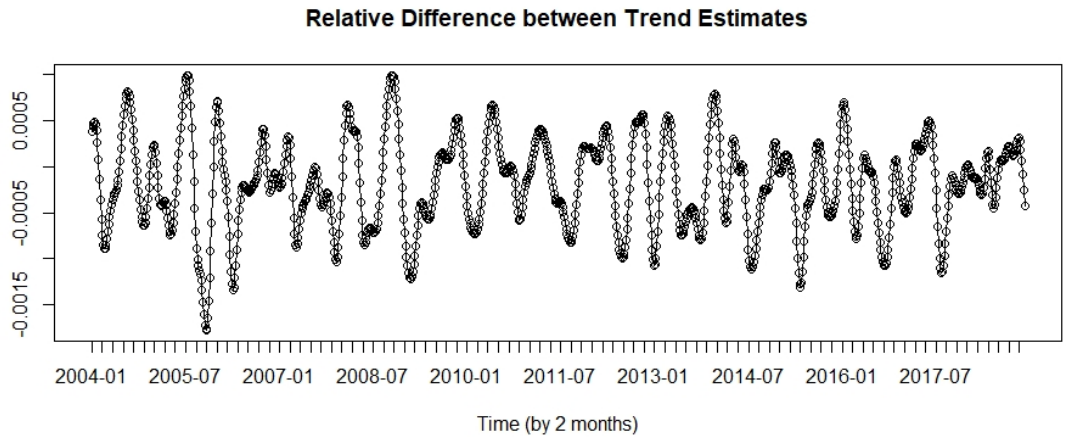

Figure 3.16: Relative Difference between the Trend Estimates with different $\boldsymbol{Q}$

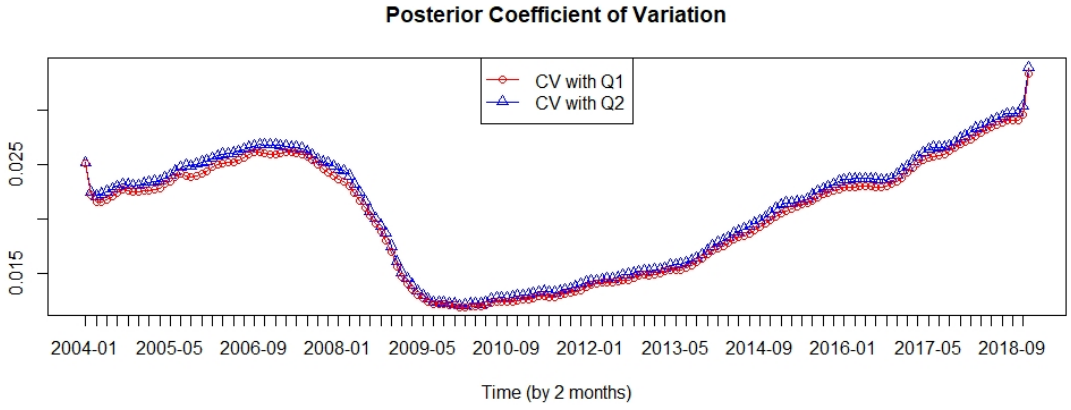

Figure 3.17: Comparison of posterior coefficient of variation with different $\boldsymbol{Q}$

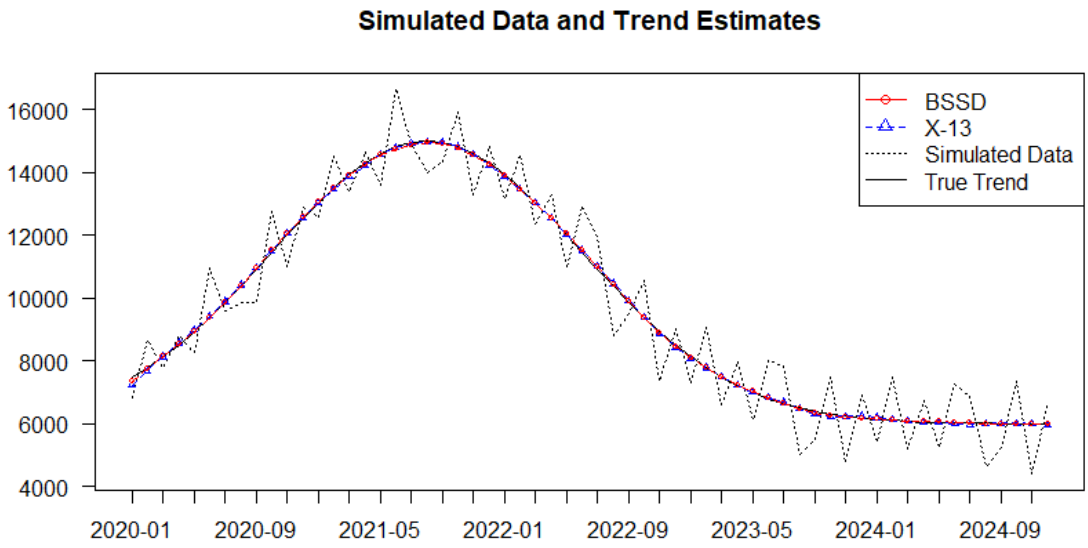

Figure 3.18: One Series of Simulated Data with Trend Estimates Compared with True Trend

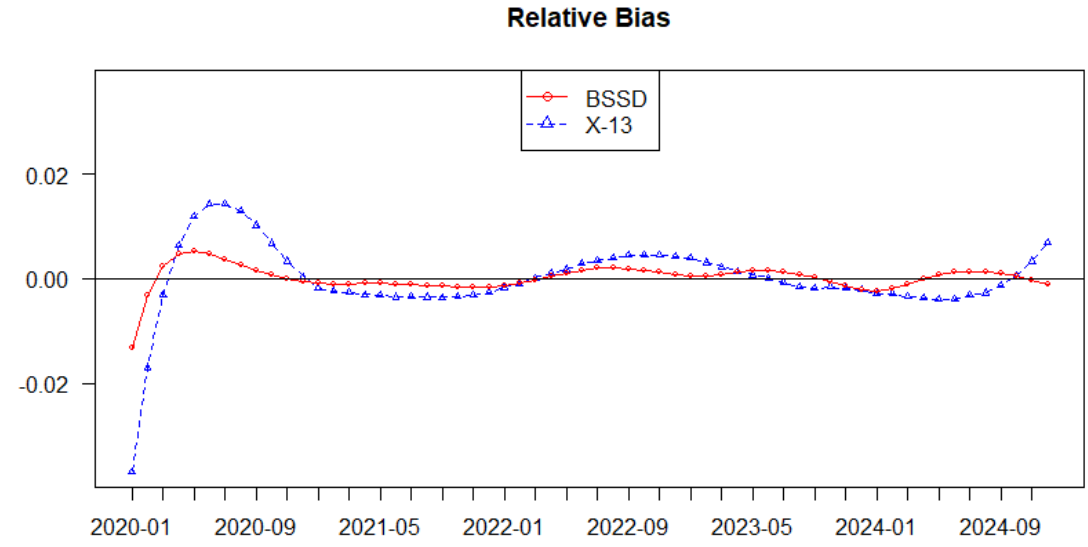

Figure 3.19: Comparison of the Relative Bias of the Trend Estimates

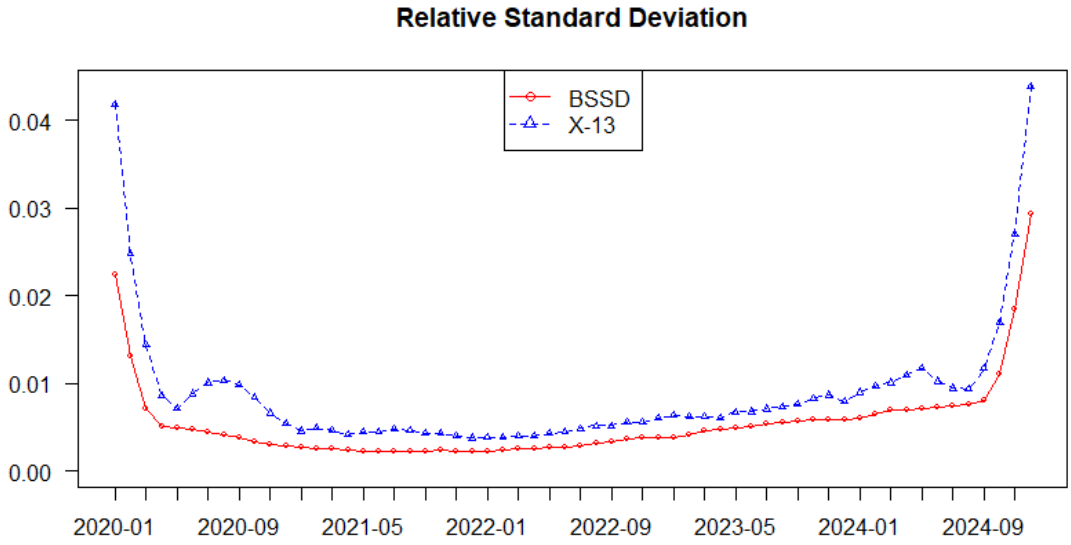

Figure 3.20: Comparison of the Relative Standard Deviation of the Trend Estimates

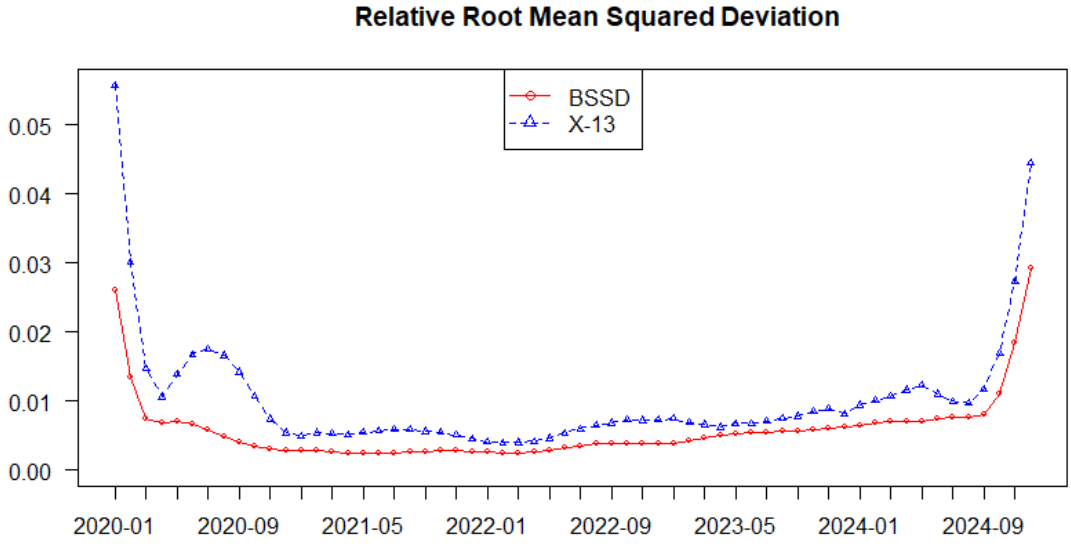

Figure 3.21: Comparison of the Relative RMSD of the Trend Estimates

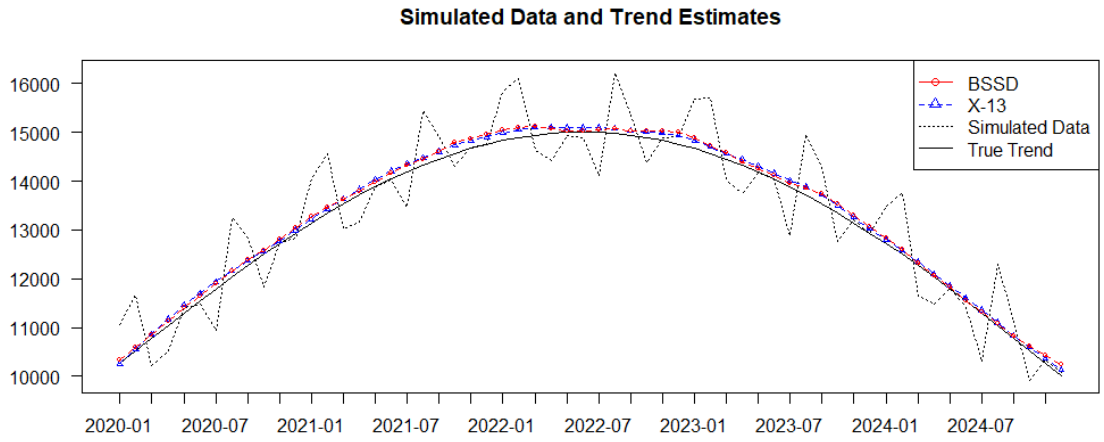

Figure 3.22: One Series of Simulated Data with Trend Estimates Compared with True Trend

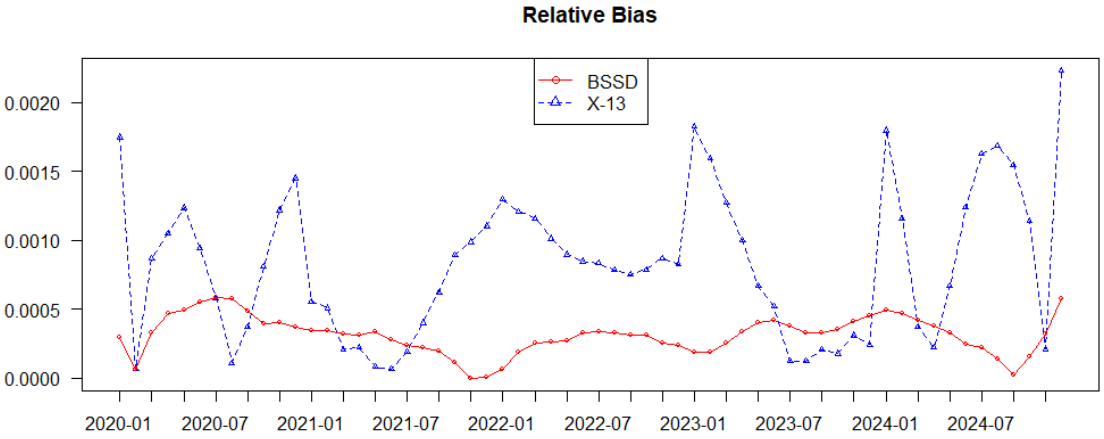

Figure 3.23: Comparison of the Relative Bias of the Trend Estimates

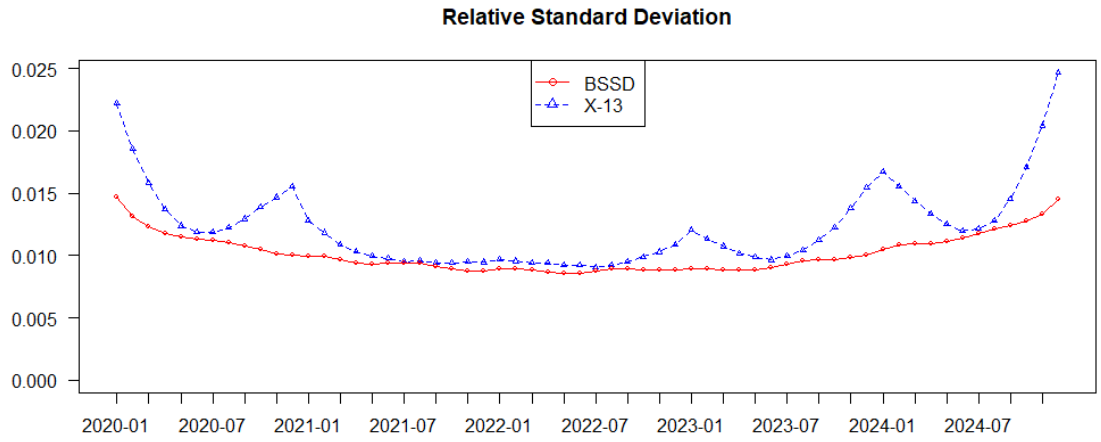

Figure 3.24: Comparison of the Relative Standard Deviation of the Trend Estimates

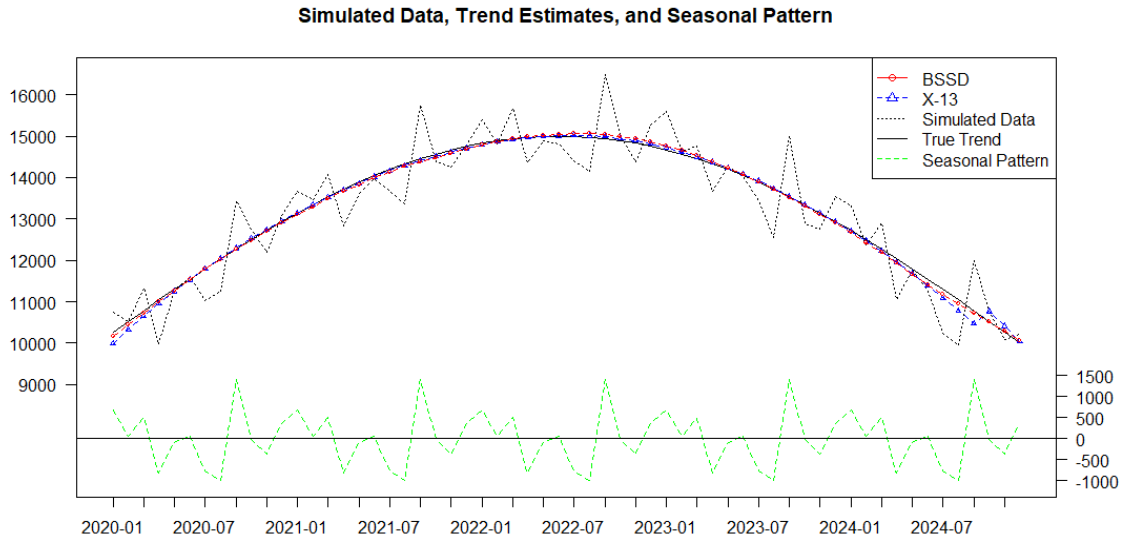

Figure 3.25: One Series of Simulated Data with Trend Estimates Compared with True Trend

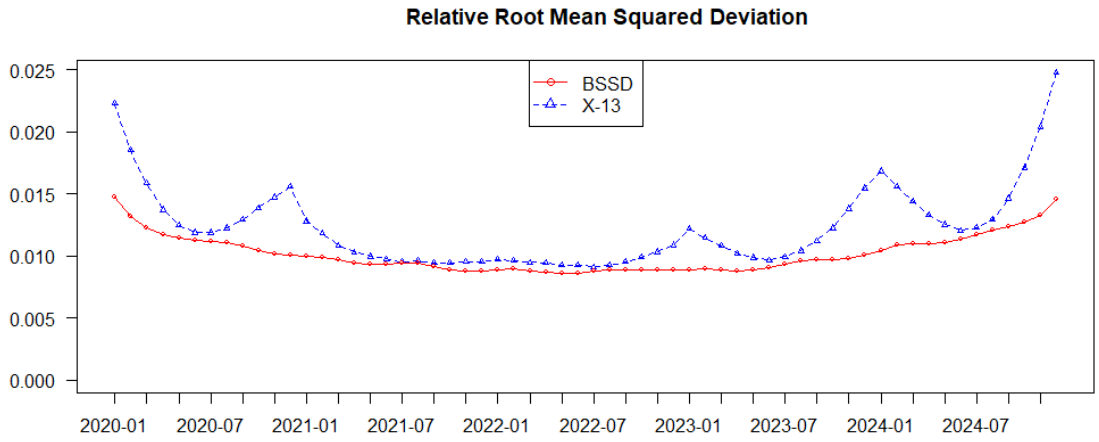

Figure 3.26: Comparison of the Relative RMSD of the Trend Estimates

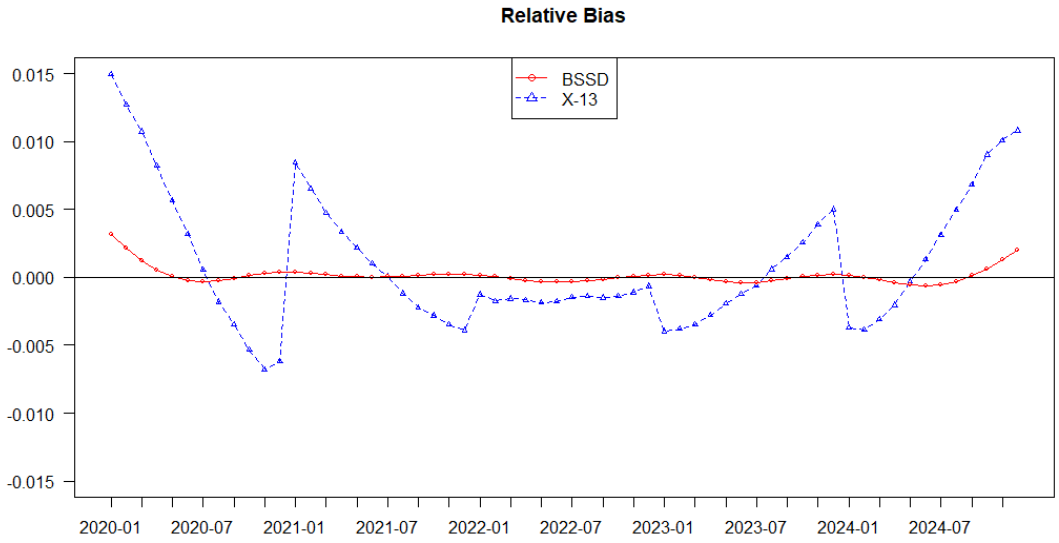

Figure 3.27: Comparison of the Relative Bias of the Trend Estimates

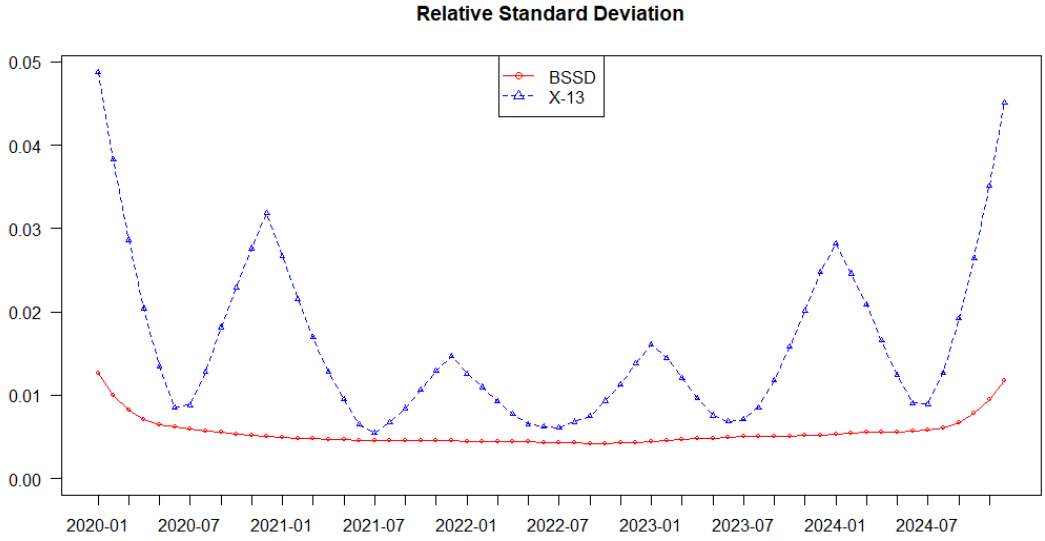

Figure 3.28: Comparison of the Relative Standard Deviation of the Trend Estimates

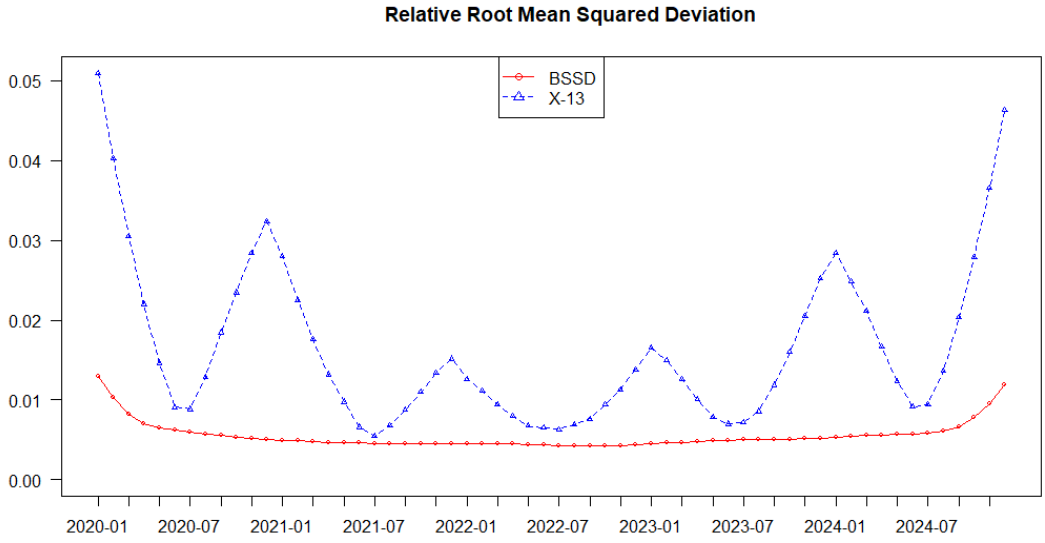

Figure 3.29: Comparison of the Relative RMSD of the Trend Estimates

3.10 Appendix

3.A Lemmas

To begin with, we introduce the following Lemmas.

Lemma 3.10.1 (Corollary 18.1.2 (Harville 2008)). *For any $n \times m$ matrix $\boldsymbol{A}$ and any $m \times n$ matrix $\boldsymbol{B}$, $|\boldsymbol{I}_n + \boldsymbol{AB}| = |\boldsymbol{I}_m + \boldsymbol{BA}|$.*

Lemma 3.10.2 (Theorem 18.1.6 (Harville 2008)). *For any $n \times n$ symmetric positive definite matrix $\boldsymbol{A}$ and any $n \times n$ symmetric nonnegative definite matrix $\boldsymbol{B}$, $|\boldsymbol{A}+\boldsymbol{B}| \geq |\boldsymbol{A}|$.*

Lemma 3.10.3 (Theorem 1.7 (Schott 1996)). *Suppose $\boldsymbol{A}_{m\times m}$ and $\boldsymbol{B}_{n\times n}$ are nonsingular matrices. For any matrices $\boldsymbol{C}_{m\times n}$ and $\boldsymbol{D}_{n\times m}$, if $\boldsymbol{A}+\boldsymbol{CBD}$ is nonsingular, then $(\boldsymbol{A}+\boldsymbol{CBD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{C}(\boldsymbol{B}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{C})^{-1}\boldsymbol{DA}^{-1}$.*

Lemma 3.10.4 (Marshall, Olkin, and Arnold (2011)). *Assume that two $n \times n$ symmetric matrices $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are both non-negative definite. Let $\lambda_1(\boldsymbol{S}_i) \leq \lambda_2(\boldsymbol{S}_i) \leq \cdots \leq \lambda_n(\boldsymbol{S}_i)$ be the eigenvalues of $\boldsymbol{S}_i$ for $i = 1, 2$. Then*

$$\prod_{j=1}^{n} \left[\lambda_j(\boldsymbol{S}_1) + \lambda_j(\boldsymbol{S}_2)\right] \leq |\boldsymbol{S}_1 + \boldsymbol{S}_2| \leq \prod_{j=1}^{n} \left[\lambda_j(\boldsymbol{S}_1) + \lambda_{n-j+1}(\boldsymbol{S}_2)\right].$$

Lemma 3.10.5 (Theorem 9.1.6 (Albert 1972)). *Let $\boldsymbol{A}$ be symmetric and the partition*

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix}.$$

Then $\boldsymbol{A} > 0$ if and only if (1) $\boldsymbol{A}_{22} > 0$, and (2) $\boldsymbol{A}_{11} > \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}$.

Lemma 3.10.6 (Theorem A.9 (Schur complement) (De Klerk 2002))**.** *Let the partition*

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & \boldsymbol{C} \end{pmatrix}.$$

If $\boldsymbol{A}$ *is positive definite and* $\boldsymbol{C}$ *is symmetric, then the following are equivalent.*

(1) $\boldsymbol{M}$ *is positive (semi)definite,*

(2) $\boldsymbol{C} - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}$ *is positive (semi)definite.*

If $\boldsymbol{C}$ *is positive definite and* $\boldsymbol{A}$ *is symmetric, then the following are equivalent.*

(1) $\boldsymbol{M}$ *is positive (semi)definite,*

(2) $\boldsymbol{A} - \boldsymbol{B}'\boldsymbol{C}^{-1}\boldsymbol{B}$ *is positive (semi)definite.*

Lemma 3.10.7. *Let* $\boldsymbol{A}_{n\times n}, \boldsymbol{B}_{n\times n}$ *be positive definite and symmetric. If* $\boldsymbol{A} \leq \boldsymbol{B}$*, then* $\boldsymbol{A}^{-1} \geq \boldsymbol{B}^{-1}$*.*

Proof. Let the partition

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{I} \\ \boldsymbol{I}' & \boldsymbol{A}^{-1} \end{pmatrix}.$$

As $\boldsymbol{B} - \boldsymbol{A} \geq 0$ and $\boldsymbol{A} > 0$, by Lemma 3.10.6, $\boldsymbol{M} \geq 0$. Also $\boldsymbol{A}^{-1} > 0$, and by Lemma 3.10.6, $\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1} \geq 0$ □

Lemma 3.10.8 (Theorem 21.11.1 (Harville 2008))**.** *Let* λ *represent an eigenvalue of matrix* $\boldsymbol{A}_{m\times m}$ *with the corresponding eigenvector* $\boldsymbol{x}$*. Let* τ *represent an eigenvalue of matrix* $\boldsymbol{B}_{p\times p}$ *with the corresponding eigenvector* $\boldsymbol{y}$*. Then* $\lambda\tau$ *is an eigenvalue of* $\boldsymbol{A} \otimes \boldsymbol{B}$ *with the corresponding eigenvector* $\boldsymbol{x} \otimes \boldsymbol{y}$*.*

Lemma 3.10.9 (Corollary 21.11.5 (Harville 2008)). *Suppose that $\boldsymbol{A}_{m\times m}$ has m (not necessarily distinct) eigenvalues, say $d_1,\ldots,d_m$, and $\boldsymbol{B}_{p\times p}$ has p (not necessarily distinct) eigenvalues, say $f_1,\ldots,f_p$. Then $\boldsymbol{A}\otimes\boldsymbol{B}$ has mp (not necessarily distinct) eigenvalues: d_if_j $(i=1,\ldots,m; j=1,\ldots,p)$.*

Lemma 3.10.10 (Eigenvalue Interlacing Theorem, See Hwang (2004), and Theorem 4.3.28 (Horn and Johnson 2012)). *Suppose $\boldsymbol{A}_{n\times n}$ is Hermitian. Let $\boldsymbol{B}_{m\times m}$, $m<n$, be a principal submatrix of $\boldsymbol{A}$. Suppose $\boldsymbol{A}$ has eigenvalues $\lambda_1\le\cdots\le\lambda_n$ and $\boldsymbol{B}$ has eigenvalues $\beta_1\le\cdots\le\beta_m$. Then*

$$\lambda_k\le\beta_k\le\lambda_{k+n-m}, k=1,\ldots,m. \tag{3.83}$$

WOLG, let $\boldsymbol{A}$ be partitioned as

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{B}_{m\times m} & \boldsymbol{C}_{m\times(n-m)}\\ \boldsymbol{C}^*_{(n-m)\times m} & \boldsymbol{D}_{(n-m)\times(n-m)}\end{pmatrix}.$$

In (3.83), the equality in the lower bound for some i if and only if there is a nonzero vector $\boldsymbol{\xi}$ such that $\boldsymbol{B}\boldsymbol{\xi}=\beta_i\boldsymbol{\xi}$ and $\boldsymbol{C}^*\boldsymbol{\xi}=\boldsymbol{0}$; equality in the upper bound occurs for some i if and only if there is a nonzero vector $\boldsymbol{\xi}$ such that $\boldsymbol{B}\boldsymbol{\xi}=\beta_{i+n-m}\boldsymbol{\xi}$ and $\boldsymbol{C}^*\boldsymbol{\xi}=\boldsymbol{0}$.

If $i\in\{1,\ldots,m\}, 1\le r\le i$, and

$$\lambda_{i-r+1}=\cdots=\lambda_i=\beta_i, \tag{3.84}$$

then $\beta_{i-r+1}=\cdots=\beta_i$ and there are orthonormal vectors $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_r$ such that $\boldsymbol{B}\boldsymbol{\xi}_j=\beta_i\boldsymbol{\xi}_j$ and $\boldsymbol{C}^*\boldsymbol{\xi}_j=\boldsymbol{0}$ for each $j=1,\ldots,r$.

If $i \in \{1, \ldots, m\}, 1 \leq r \leq m - i + 1$, *and*

$$\beta_i = \lambda_{i+n-m} = \cdots = \lambda_{i+n-m+r-1}, \tag{3.85}$$

then $\beta_i = \cdots = \beta_{i+n-m+r-1}$ *and there are orthonormal vectors* $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_r$ *such that* $\boldsymbol{B}\boldsymbol{\xi}_j = \beta_i \boldsymbol{\xi}_j$ *and* $\boldsymbol{C}^* \boldsymbol{\xi}_j = \boldsymbol{0}$ *for each* $j = 1, \ldots, r$.

Lemma 3.10.11. *Let* $\boldsymbol{A}_{n \times n}$ *be Hermitian, partitioned as*

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{B}_{m \times m} & \boldsymbol{C}_{m \times (n-m)} \\ \boldsymbol{C}^*_{(n-m) \times m} & \boldsymbol{D}_{(n-m) \times (n-m)} \end{pmatrix}.$$

Suppose $\boldsymbol{A}$ *has eigenvalues* $\lambda_1 \leq \cdots \leq \lambda_n$, *and* $\boldsymbol{B}$ *has eigenvalues* $\beta_1 \leq \cdots \leq \beta_m$.

If for some $K \in \{1, \ldots, m\}$, $\{i_1, \ldots, i_K\} \subseteq \{1, \ldots, n\}$, $\{j_1, \ldots, j_K\} \subseteq \{1, \ldots, m\}$, *and*

$$\lambda_{i_1} = \cdots = \lambda_{i_K} = \beta_{j_1} = \cdots = \beta_{j_K}. \tag{3.86}$$

Then there exist corresponding orthogonal eigenvectors $\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_K}$ *of* $\boldsymbol{B}$ *of* $\boldsymbol{A}$, *and corresponding orthogonal eigenvectors* $\boldsymbol{\xi}_{j_1}, \ldots, \boldsymbol{\xi}_{j_K}$ *of* $\boldsymbol{B}$, *satisfying that*

$$\boldsymbol{x}_{i_k} = \begin{pmatrix} \boldsymbol{\xi}_{j_k} \\ \boldsymbol{0} \end{pmatrix} \quad k = 1, \ldots, K. \tag{3.87}$$

Proof. By Lemma 3.10.10, there are orthonormal vectors $\boldsymbol{\xi}_{j_1}, \ldots, \boldsymbol{\xi}_{j_K}$ such that $\boldsymbol{B}\boldsymbol{\xi}_{j_k} = \beta \boldsymbol{\xi}_{j_k}$ and $\boldsymbol{C}^* \boldsymbol{\xi}_{j_k} = \boldsymbol{0}$ for each $k = 1, \ldots, K$, and $\beta = \beta_{j_1} = \cdots = \beta_{j_K} = \lambda_{i_1} = \cdots = \lambda_{i_K}$.

Then

$$\begin{pmatrix} \boldsymbol{B} & \boldsymbol{C} \\ \boldsymbol{C}^* & \boldsymbol{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{\xi}_{j_k} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}\boldsymbol{\xi}_{j_k} \\ \boldsymbol{C}^*\boldsymbol{\xi}_{j_k} \end{pmatrix} = \begin{pmatrix} \beta\boldsymbol{\xi}_{j_k} \\ \mathbf{0} \end{pmatrix} = \beta \begin{pmatrix} \boldsymbol{\xi}_{j_k} \\ \mathbf{0} \end{pmatrix} \quad k = 1, \ldots, K \qquad (3.88)$$

Thus the eigenvector $\boldsymbol{x}_{i_k}$ of $\boldsymbol{A}$ corresponding to λ_{i_k} and the eigenvector $\boldsymbol{\xi}_{j_k}$ of $\boldsymbol{B}$ corresponding to β_{j_k} satisfy that:

$$\boldsymbol{x}_{i_k} = \begin{pmatrix} \boldsymbol{\xi}_{j_k} \\ \mathbf{0} \end{pmatrix}, k = 1, \ldots, K.$$

□

3.B Correlation Matrix

We introduce an efficient way to generate the correlation matrix $\mathbf{R}$ defined in (3.5), and compute the eigenvalues of $\mathbf{R}$.

Case 1. $n = mT$

We first consider the case that the length of series $n = mT$, where T is the period, and the integer $m > 1$.

We define $T \times T$ identity matrix $\boldsymbol{I}_T$, and $m \times m$ matrix $\boldsymbol{A}_m$ as

$$\boldsymbol{A}_m = (1-\rho)\boldsymbol{I}_m + \rho \mathbf{1}'_m \mathbf{1}_m = \begin{pmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & & \rho \\ \rho & \ldots & \rho & 1 \end{pmatrix},$$

where $\mathbf{1}'_m = (1, \ldots, 1)'$ is a vector of length m. Then the correlation matrix $\mathbf{R}$ is

$$\mathbf{R} = \boldsymbol{A}_m \otimes \boldsymbol{I}_T = \begin{pmatrix} \boldsymbol{I}_T & \rho\boldsymbol{I}_T & \ldots & \rho\boldsymbol{I}_T \\ \rho\boldsymbol{I}_T & \boldsymbol{I}_T & \ddots & \vdots \\ \vdots & \ddots & & \rho\boldsymbol{I}_T \\ \rho\boldsymbol{I}_T & \ldots & \rho\boldsymbol{I}_T & \boldsymbol{I}_T \end{pmatrix}. \tag{3.89}$$

The eigenvalues of $\boldsymbol{A}_m$ can be obtained by

$$|\lambda\boldsymbol{I} - \boldsymbol{A}_m| = (\lambda - 1 + \rho)^{m-1}(\lambda - 1 + \rho - m\rho) = 0. \tag{3.90}$$

$$\lambda_i(\boldsymbol{A}_m) = \begin{cases} 1 - \rho, & i = 1, \ldots, m-1, \\ 1 - \rho + m\rho, & i = m. \end{cases} \tag{3.91}$$

The corresponding eigenvectors $\boldsymbol{v}_i$ are:

$$\begin{aligned} \boldsymbol{v}_1 &= (1, -1, 0, \ldots, 0)', \\ \boldsymbol{v}_2 &= (1, 1, -2, 0, \ldots, 0)', \\ \boldsymbol{v}_3 &= (1, 1, 1, -3, 0, \ldots, 0)', \\ &\ldots \\ \boldsymbol{v}_{m-1} &= (1, 1, \ldots, 1, -(m-1))', \\ \boldsymbol{v}_m &= (1, 1, \ldots, 1)'. \end{aligned}$$

i.e.

$$\boldsymbol{v}_i = \begin{cases} (\mathbf{1}_i', -i, 0, \ldots, 0)', & i = 1, \ldots, m-1, \\ \mathbf{1}_m, & i = m. \end{cases} \tag{3.92}$$

Note that all $\boldsymbol{v}_i$'s are orthogonal with each other and linear independent. Also, the $\boldsymbol{v}_i$'s do not depend on the parameter ρ.

Let $\boldsymbol{P}_m = (\frac{\boldsymbol{v}_1}{|\boldsymbol{v}_1|}, \ldots, \frac{\boldsymbol{v}_m}{|\boldsymbol{v}_m|})$, and $\boldsymbol{D}_m = diag(\lambda_1(\boldsymbol{A}_m), \ldots, \lambda_m(\boldsymbol{A}_m))$. $\boldsymbol{P}_m$ is an orthogonal matrix and $\boldsymbol{A}_m$ is diagonalized as $\boldsymbol{A}_m = \boldsymbol{P}_m \boldsymbol{D}_m \boldsymbol{P}_m'$.

Then, by Lemma 3.10.9, Suppose $\mathbf{R} = \boldsymbol{A}_m \otimes \boldsymbol{I}_T$ has eigenvalues $d_1 \leq \cdots \leq d_{mT}$,

$$d_i = \begin{cases} 1-\rho, & i = 1, \ldots, (m-1)T, \\ 1-\rho+m\rho, & i = (m-1)T+1, \ldots, mT. \end{cases} \tag{3.93}$$

Let $\boldsymbol{D} = diag(d_1, \ldots, d_{mT})$, and $\boldsymbol{P} = \boldsymbol{P}_m \otimes \boldsymbol{I}_T$. By Lemma 3.10.8, the spectrum decomposition of $\mathbf{R}$ is $\mathbf{R} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$, where the orthogonal matrix $\boldsymbol{P}$ does not depend on parameter ρ.

Case 2. $n \neq mT$

We consider the case that $n \neq mT$. There exist integers $m > 1$ and $1 \leq l \leq T-1$, such that $n = mT - l$. We first construct the $mT \times mT$ correlation matrix $\mathbf{R}_0 = \boldsymbol{A}_m \otimes \boldsymbol{I}_T$. Then $\mathbf{R}$ is the leading principal submatrix of order n of $\mathbf{R}_0$. Suppose $\mathbf{R}_0$ has eigenvalues $r_1^0 \leq \cdots \leq r_{mT}^0$ and $\mathbf{R}$ has eigenvalues $d_1 \leq \cdots \leq d_n$. Then

$$r_i^0 = \begin{cases} 1-\rho, & i = 1, \ldots, (m-1)T, \\ 1-\rho+m\rho, & i = (m-1)T+1, \ldots, mT. \end{cases} \tag{3.94}$$

By Lemma 3.10.10,

$$r_k^0 \le d_k \le r_{k+l}^0, k = 1, \ldots, n. \tag{3.95}$$

Meanwhile, if we construct a $(m-1)T \times (m-1)T$ correlation matrix $\mathbf{R}_1 = \boldsymbol{A}_{m-1} \otimes \boldsymbol{I}_T$. Then $\mathbf{R}_1$ is the leading principal submatrix of $\mathbf{R}$, by removing the last $T-l$ rows and columns. Suppose $\mathbf{R}_1$ has eigenvalues $r_1^1 \le \cdots \le r_{(m-1)T}^1$. Then

$$r_i^1 = \begin{cases} 1-\rho, & i = 1, \ldots, (m-2)T, \ if \ m \ge 3, \\ 1-\rho+(m-1)\rho, & i = (m-2)T+1, \ldots, (m-1)T. \end{cases} \tag{3.96}$$

By Lemma 3.10.10,

$$d_k \le r_k^1 \le d_{k+T-l}, k = 1, \ldots, (m-1)T. \tag{3.97}$$

Therefore, by (3.94), (3.95), (3.96), (3.97), the eigenvalues of $\mathbf{R}$

$$d_i = \begin{cases} 1-\rho, & i = 1, \ldots, (m-1)T-l, \\ 1-\rho+(m-1)\rho, & i = (m-1)T-l+1, \ldots, (m-1)T, \\ 1-\rho+m\rho, & i = (m-1)T+1, \ldots, n. \end{cases} \tag{3.98}$$

Let $\boldsymbol{p}_i$ be the eigenvector corresponding to d_i. By Lemma 3.10.10, the $\boldsymbol{p}_i$ havs the close form as follows.

Let the partitions of $\mathbf{R}_0$ and $\mathbf{R}$ be

$$\mathbf{R}_0 = \begin{pmatrix} \mathbf{R} & \boldsymbol{B}_{(mT-l)\times l} \\ \boldsymbol{B}'_{l\times(mT-l)} & \boldsymbol{I}_l \end{pmatrix}, \ \mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \boldsymbol{E}_{(m-1)T\times(T-l)} \\ \boldsymbol{E}'_{(T-l)\times(m-1)T} & \boldsymbol{I}_{T-l} \end{pmatrix}. \tag{3.99}$$

Note that

$$d_1 = \cdots = d_{(m-1)T-l} = r_1^0 = \cdots = r_{(m-1)T-l}^0, \tag{3.100}$$

$$d_{(m-1)T-l+1} = \cdots = d_{(m-1)T} = r_{(m-1)T-l+1}^1 = \cdots = r_{(m-1)T}^1, \tag{3.101}$$

$$d_{(m-1)T+1} = \cdots = d_n = r_{(m-1)T+1}^0 = \cdots = r_n^0. \tag{3.102}$$

By Lemma 3.10.11, let $\boldsymbol{I}_T = [\boldsymbol{e}_1 \ \ldots \ \boldsymbol{e}_T]$, $n \times mT$ matrix $\boldsymbol{\Gamma}_1 = [\boldsymbol{I}_n \ \mathbf{0}]$, $n \times (m-1)T$ matrix $\boldsymbol{\Gamma}_2 = [\boldsymbol{I}_{(m-1)T} \ \mathbf{0}]'$, and $\boldsymbol{v}_i$ be defined as (3.92), then

$$\begin{aligned}
\boldsymbol{p}_1 &= \boldsymbol{\Gamma}_1(\boldsymbol{v}_1 \otimes \boldsymbol{e}_1), \\
&\vdots \\
\boldsymbol{p}_T &= \boldsymbol{\Gamma}_1(\boldsymbol{v}_1 \otimes \boldsymbol{e}_T), \\
\boldsymbol{p}_{T+1} &= \boldsymbol{\Gamma}_1(\boldsymbol{v}_2 \otimes \boldsymbol{e}_1), \\
&\vdots \\
\boldsymbol{p}_{(m-1)T-l} &= \boldsymbol{\Gamma}_1(\boldsymbol{v}_{m-1} \otimes \boldsymbol{e}_{T-l}), \\
\boldsymbol{p}_{(m-1)T-l+1} &= \boldsymbol{\Gamma}_2(\mathbf{1}_{m-1} \otimes \boldsymbol{e}_{T-l+1}), \\
&\vdots \\
\boldsymbol{p}_{(m-1)T} &= \boldsymbol{\Gamma}_2(\mathbf{1}_{m-1} \otimes \boldsymbol{e}_T), \\
\boldsymbol{p}_{(m-1)T+1} &= \boldsymbol{\Gamma}_1(\boldsymbol{v}_m \otimes \boldsymbol{e}_1), \\
&\vdots \\
\boldsymbol{p}_n &= \boldsymbol{\Gamma}_1(\boldsymbol{v}_m \otimes \boldsymbol{e}_{T-l}).
\end{aligned}$$

Note that all $\boldsymbol{p}_i$'s are orthogonal with each other and linear independent. Also, the $\boldsymbol{p}_i$'s do not depend on the parameter ρ. Thus let the spectrum decomposition of $\mathbf{R}$ be $\mathbf{R} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$, and the orthogonal matrix $\boldsymbol{P}$ does not depend on parameter ρ.

3.C Calculation for Programs

As the precision matrix $\boldsymbol{Q}$ in prior (3.12) has the rank $n-2$. There exists a matrix $\boldsymbol{G}_{(n-2)\times n}$ with rank $n-2$, such that $\boldsymbol{Q} = \boldsymbol{G}'\boldsymbol{G}$. Note that $\boldsymbol{Q} = \boldsymbol{F}_0'\boldsymbol{F}_1^{-1}\boldsymbol{F}_0$. Let the LU decomposition $\boldsymbol{F}_1 = \boldsymbol{L}\boldsymbol{L}'$, where $\boldsymbol{L}$ is a $(n-2)\times(n-2)$ lower triangular matrix. Then $\boldsymbol{G} = \boldsymbol{L}^{-1}\boldsymbol{F}_0$.

Given $\mathbf{R}$ is p.d. and symmetric, and $\boldsymbol{G}\boldsymbol{W}\mathbf{R}\boldsymbol{W}\boldsymbol{G}' = (\boldsymbol{G}\boldsymbol{W})\mathbf{R}(\boldsymbol{G}\boldsymbol{W})'$ is also p.d. and symmetric. There exists an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}' = \boldsymbol{G}\boldsymbol{W}\mathbf{R}\boldsymbol{W}\boldsymbol{G}'$, $\boldsymbol{D} = Diag(d_1, \ldots, d_{n-2})$, $0 < d_1 \leq \cdots \leq d_{n-2}$ are the eigenvalues of $\boldsymbol{G}\boldsymbol{W}\mathbf{R}\boldsymbol{W}\boldsymbol{G}'$.

By Lemma 3.10.1 ,

$$\begin{aligned} |\boldsymbol{I}_n + \eta\boldsymbol{W}\mathbf{R}\boldsymbol{W}\boldsymbol{Q}| &= |\boldsymbol{I}_n + \eta\mathbf{R}\boldsymbol{W}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{W}| \\ &= |\boldsymbol{I}_n + \eta\boldsymbol{G}\boldsymbol{W}\mathbf{R}\boldsymbol{W}\boldsymbol{G}'| \\ &= |\boldsymbol{I}_n + \eta\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'| \\ &= |\boldsymbol{I}_n + \eta\boldsymbol{D}| \\ &= \prod_{j=1}^{n-2}(1+\eta d_j). \end{aligned} \tag{3.103}$$

By Lemma 3.10.3,

$$\begin{aligned} [(\boldsymbol{W}\mathbf{R}\boldsymbol{W})^{-1} + \eta\boldsymbol{Q}]^{-1} &= [(\boldsymbol{W}\mathbf{R}\boldsymbol{W})^{-1} + \boldsymbol{G}'(\eta\boldsymbol{I})\boldsymbol{G}]^{-1} \\ &= (\boldsymbol{W}\mathbf{R}\boldsymbol{W}) - (\boldsymbol{W}\mathbf{R}\boldsymbol{W})\boldsymbol{G}'(\eta^{-1}\boldsymbol{I} + \boldsymbol{G}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})\boldsymbol{G}')^{-1}\boldsymbol{G}(\boldsymbol{W}\mathbf{R}\boldsymbol{W}). \end{aligned} \tag{3.104}$$

Thus

$$\begin{aligned}
& \boldsymbol{Q}((\boldsymbol{W}\mathbf{R}\boldsymbol{W})^{-1}+\eta\boldsymbol{Q})^{-1}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})^{-1} \\
= \; & \boldsymbol{G}'\boldsymbol{G}[(\boldsymbol{W}\mathbf{R}\boldsymbol{W})-(\boldsymbol{W}\mathbf{R}\boldsymbol{W})\boldsymbol{G}'(\eta^{-1}\boldsymbol{I}+\boldsymbol{G}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})\boldsymbol{G}')^{-1}\boldsymbol{G}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})](\boldsymbol{W}\mathbf{R}\boldsymbol{W})^{-1} \\
= \; & \boldsymbol{G}'\boldsymbol{G}-\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})\boldsymbol{G}'(\eta^{-1}\boldsymbol{I}+\boldsymbol{G}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})\boldsymbol{G}')^{-1}\boldsymbol{G} \\
= \; & \boldsymbol{G}'[\boldsymbol{I}-\boldsymbol{G}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})\boldsymbol{G}'(\eta^{-1}\boldsymbol{I}+\boldsymbol{G}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})\boldsymbol{G}')^{-1}]\boldsymbol{G} \\
= \; & \boldsymbol{G}'[\boldsymbol{I}-\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'(\eta^{-1}\boldsymbol{I}+\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}')^{-1}]\boldsymbol{G} \\
= \; & \boldsymbol{G}'\boldsymbol{P}[\boldsymbol{I}-\boldsymbol{D}(\eta^{-1}\boldsymbol{I}+\boldsymbol{D})^{-1}]\boldsymbol{P}'\boldsymbol{G}. \qquad (3.105)
\end{aligned}$$

Let $\boldsymbol{v}=(v_1,\ldots,v_{n-2})=\boldsymbol{P}'\boldsymbol{G}\boldsymbol{y}$, $\boldsymbol{B}=\boldsymbol{I}-\boldsymbol{D}(\eta^{-1}\boldsymbol{I}+\boldsymbol{D})^{-1}=Diag(b_1,\ldots,b_{n-2})$, $b_i=1-\frac{d_i}{\eta^{-1}+d_i}=\frac{1}{1+\eta d_i}$, then

$$\begin{aligned}
\boldsymbol{y}'\boldsymbol{Q}((\boldsymbol{W}\mathbf{R}\boldsymbol{W})^{-1}+\eta\boldsymbol{Q})^{-1}(\boldsymbol{W}\mathbf{R}\boldsymbol{W})^{-1}\boldsymbol{y} &= \boldsymbol{v}'\boldsymbol{B}\boldsymbol{v} \\
&= \sum_{i=1}^{n-2} b_i v_i^2 = \sum_{i=1}^{n-2}\frac{v_i^2}{1+\eta d_i}. \qquad (3.106)
\end{aligned}$$

Therefore $p(\mathbf{R},\eta|\boldsymbol{y})$ has the closed form as

$$\begin{aligned}
p(\mathbf{R},\eta|\boldsymbol{y}) &= \boldsymbol{C}[\mathbf{R}][\eta]\left[\prod_{j=1}^{n-2}(1+\eta d_j)\right]^{-\frac{1}{2}}\left[\sum_{i=1}^{n-2}\frac{v_i^2}{1+\eta d_i}\right]^{1-\frac{n}{2}} \\
&= \boldsymbol{C}[\mathbf{R}][\eta]\left\{\prod_{j=1}^{n-2}\left[\sum_{i=1}^{n-2} v_i^2\frac{1+\eta d_j}{1+\eta d_i}\right]\right\}^{-\frac{1}{2}}. \qquad (3.107)
\end{aligned}$$

Chapter 4

Multivariate Bayesian Smoothing Spline with Dependency Model

4.1 Introduction

In the real world, we would like to estimate the trend of both the unemployment and the employment level simultaneously, because although the unemployment and the employment level tend to have a negative correlation, there is no determinative relation between these two values. Another example that we are interested in is the developmental tendency of the Covid-19 epidemic, given the daily case data from multiple states in the US.

Many authors have studied the multivariate Bayesian smoothing spline models with independent error assumption. However, for the data with seasonal fluctuations, the results are not stable and have unpleasant boundary performance, which happens in the univariate case as well. In this chapter, we construct the multivariate Bayesian smoothing spline with dependency (MBSSD) model to borrow information across all series and to improve the performance of accuracy and stability. In section 4.2, we

provide the full Bayesian analysis of the MBSSD model, explore the sampling method, and conduct the model selection with the deviance information criterion (DIC). In section 4.4, we present two simulation studies to evaluate the performance of our MBSSD model. In section 4.5, we apply the MBSSD model with real datasets and compare with the results from the univariate BSSD model in Chapter 3.

4.2 Multivariate Bayesian smoothing spline model

We consider a nonparametric regression model with p spline curves,

$$y_{ij} = z_{ij} + \epsilon_{ij},\ z_{ij} = f_j(t_i), \tag{4.1}$$

at knots t_i, where $i = 1, \ldots, n$, $j = 1, \ldots, p$, and $-\infty < a \leq t_1 < \cdots < t_n \leq b < \infty$. We denote $\boldsymbol{Y} = (y_{ij})_{n\times p}$, $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{nj})'$, $\boldsymbol{y} = vec(\boldsymbol{Y}) = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_p')'$, $\boldsymbol{Z} = (z_{ij})_{n\times p}$, $\boldsymbol{z}_j = (z_{1j}, \ldots, z_{nj})'$, $\boldsymbol{z} = vec(\boldsymbol{Z})$, $\boldsymbol{f}(t) = (f_1(t), \ldots, f_p(t))'$.
The randoms errors $\boldsymbol{\epsilon} = (\epsilon_{11}, \ldots, \epsilon_{n1}, \ldots, \epsilon_{1p}, \ldots, \epsilon_{np})' \sim N_{np}(\boldsymbol{0}, \boldsymbol{\Sigma}_{00} \otimes \boldsymbol{\Sigma}_{01})$, where $\boldsymbol{\Sigma}_{00}$ is a $p \times p$ covariance matrix, and $\boldsymbol{\Sigma}_{01}$ is a $n \times n$ covariance matrix.

To estimate the unknown function vector $\boldsymbol{f}(t)$, we minimize a squared error loss function with a penalty on smoothness,

$$(\boldsymbol{y} - \boldsymbol{z})'(\boldsymbol{\Sigma}_{00} \otimes \boldsymbol{\Sigma}_{01})^{-1}(\boldsymbol{y} - \boldsymbol{z}) + \int_a^b \left(\boldsymbol{f}^{(k)}(s)\right)' \boldsymbol{\Sigma}_1^{-1} \left(\boldsymbol{f}^{(k)}(s)\right) ds, \tag{4.2}$$

where $\boldsymbol{f}^{(k)}(t)$ is the kth derivative. We take $k = 2$ for Cubic natural smoothing spline. The smoothness penalty term can be rewrite as

$$\text{tr}\left[\boldsymbol{\Sigma}_1^{-1} \int_a^b \left(\boldsymbol{f}^{(2)}(s)\right)\left(\boldsymbol{f}^{(2)}(s)\right)' ds\right] = \text{tr}\left(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{Z}\right) = \boldsymbol{z}'(\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q})\boldsymbol{z}. \tag{4.3}$$

The problem of multivariate smoothing splines with unrestricted covariance structure has not been fully studied in the past. Wang, Guo, and Brown (2000) restricted $\mathbf{\Sigma}_1$ to be diagonal, and Sun, Ni, and Speckman (2014) proposed a full Bayesian analysis with general $\mathbf{\Sigma}_{00}, \mathbf{\Sigma}_1$. However in these papers, the random errors ϵ_{ij} were assumed be independent for different i's.

In this chapter, we are interested in the series with seasonal effects. We assume that random errors ϵ_{ij} are correlated at different knots but are independent over different series. Considering the identifiable issue that $k\mathbf{\Sigma}_{00} \otimes \mathbf{\Sigma}_{01} = \mathbf{\Sigma}_{00} \otimes k\mathbf{\Sigma}_{01}$ for any constant $k > 0$, we let

$$\mathbf{\Sigma}_{00} = diag(\delta_1, \ldots, \delta_p), \qquad \mathbf{\Sigma}_{01} = \boldsymbol{D}^* \mathbf{R}_0 \boldsymbol{D}^*, \tag{4.4}$$

where the correlation $\mathbf{R}_0$ describes the seasonal dependency (see Appendix 4.8), and the weight matrix $\boldsymbol{D}^* = \text{diag}(d_1^*, \ldots, d_n^*)$ is considered to be known. We usually take $\boldsymbol{D}^* = \boldsymbol{I}_n$.

For equal spaced knots at $t = 1, 2, ..., n$, as a special case in Fessler (1991), the smoothness structure matrix $\boldsymbol{Q}$ can be defined as

$$\boldsymbol{Q} = \boldsymbol{F}_0' \boldsymbol{F}_1^{-1} \boldsymbol{F}_0, \tag{4.5}$$

where the $(n-2) \times n$ matrix $\boldsymbol{F}_0$ and $(n-2) \times (n-2)$ matrix $\boldsymbol{F}_1$ are

$$\boldsymbol{F}_0 = \begin{pmatrix} 1 & -2 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & -2 & 1 \end{pmatrix}, \quad \boldsymbol{F}_1 = \frac{1}{6} \begin{pmatrix} 4 & 1 & \ldots & 0 & 0 \\ 1 & 4 & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 4 \end{pmatrix}. \tag{4.6}$$

Then (4.2) can be written as

$$(\boldsymbol{y}-\boldsymbol{z})'(\boldsymbol{\Sigma}_{00}^{-1}\otimes\boldsymbol{\Sigma}_{01}^{-1})(\boldsymbol{y}-\boldsymbol{z})+\boldsymbol{z}'(\boldsymbol{\Sigma}_{1}^{-1}\otimes\boldsymbol{Q})\boldsymbol{z}. \tag{4.7}$$

The solution of $\boldsymbol{z}$ to minimize (4.7) is

$$\boldsymbol{\mu}_z=\left[\boldsymbol{I}_{np}+(\boldsymbol{\Sigma}_{00}\boldsymbol{\Sigma}_{1}^{-1})\otimes(\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\right]^{-1}\boldsymbol{y}. \tag{4.8}$$

4.3 Bayesian analysis

The density of $\boldsymbol{y}$ given $\boldsymbol{z}$, $\boldsymbol{\Sigma}_{00}$ and $\boldsymbol{\Sigma}_{01}$ based on model (4.1) is

$$\begin{aligned} f(\boldsymbol{y}\mid\boldsymbol{z},\boldsymbol{\Sigma}_{00},\boldsymbol{\Sigma}_{01}) &= (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}_{01}|^{-\frac{p}{2}}\mathrm{etr}\big[-\frac{1}{2}\boldsymbol{\Sigma}_{00}^{-1}(\boldsymbol{Y}-\boldsymbol{Z})'\boldsymbol{\Sigma}_{01}^{-1}(\boldsymbol{Y}-\boldsymbol{Z})\big] \\ &= (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}_{01}|^{-\frac{p}{2}} \\ &\quad \exp\big[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{z})'\left(\boldsymbol{\Sigma}_{00}^{-1}\otimes\boldsymbol{\Sigma}_{01}^{-1}\right)(\boldsymbol{y}-\boldsymbol{z})\big]. \end{aligned} \tag{4.9}$$

4.3.1 Fixed $(\boldsymbol{\Sigma}_{00},\boldsymbol{\Sigma}_{01},\boldsymbol{\Sigma}_{1})$

Following Sun, Ni, and Speckman (2014), we consider the partially improper prior for $\boldsymbol{z}$

$$f(\boldsymbol{z}\mid\boldsymbol{\Sigma}_1)\propto|(\boldsymbol{\Sigma}_{1}^{-1}\otimes\boldsymbol{Q})|_{+}^{1/2}\exp\big[-\frac{1}{2}\boldsymbol{z}'(\boldsymbol{\Sigma}_{1}^{-1}\otimes\boldsymbol{Q})\boldsymbol{z}\big], \tag{4.10}$$

where $|\boldsymbol{A}|_+$ is the product of positive eigenvalues of a nonnegative definite matrix $\boldsymbol{A}$. For the $n\times n$ matrix $\boldsymbol{Q}$ defined in (4.5), $\boldsymbol{Q}$ has 2 eigenvalues as 0, and $n-2$ positive

eigenvalues. Then the prior (4.10) can be written as

$$f(\boldsymbol{z} \mid \boldsymbol{\Sigma}_1) \propto |\boldsymbol{\Sigma}_1|^{-\frac{n-2}{2}} \exp\big[-\frac{1}{2}\boldsymbol{z}'(\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q})\boldsymbol{z}\big]. \tag{4.11}$$

Theorem 2. *Consider model (4.9) with prior* (4.10)*. For fixed* $(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1)$*, the conditional posterior distribution of* $\boldsymbol{z}$ *given* $\boldsymbol{y}$ *is*

$$(\boldsymbol{z} \mid \boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1; \boldsymbol{y}) \;\; \sim \;\; N_{pn}(\boldsymbol{\mu}_z, \boldsymbol{\Omega}^{-1}), \tag{4.12}$$

where $\boldsymbol{\mu}_z = \big[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}\boldsymbol{\Sigma}_1^{-1}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\big]^{-1} \boldsymbol{y}$ *and* $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q}$.

Proof. Given $(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1)$,

$$\begin{aligned}
&[\boldsymbol{z} \mid \boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1; \boldsymbol{y}] \propto [\boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}][\boldsymbol{z} \mid \boldsymbol{\Sigma}_1] \\
&\propto \;\; (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{01}|^{-\frac{p}{2}} \exp\big[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{z})' \left(\boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1}\right)(\boldsymbol{y}-\boldsymbol{z})\big] \\
&\qquad |\boldsymbol{\Sigma}_1|^{-\frac{n-2}{2}} \exp\big[-\frac{1}{2}\boldsymbol{z}'(\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q})\boldsymbol{z}\big] \\
&= \;\; (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{01}|^{-\frac{p}{2}} |\boldsymbol{\Sigma}_1|^{-\frac{n-2}{2}} \\
&\qquad \exp\big[-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_z)' \left(\boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q}\right)(\boldsymbol{z}-\boldsymbol{\mu}_z)\big] \\
&\qquad \exp\big\{-\frac{1}{2}\boldsymbol{y}' \left(\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q}\right)\big[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}\boldsymbol{\Sigma}_1^{-1}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\big]^{-1} \boldsymbol{y}\big\},
\end{aligned} \tag{4.13}$$

where $\boldsymbol{\mu}_z = \big[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}\boldsymbol{\Sigma}_1^{-1}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\big]^{-1} \boldsymbol{y}$. □

4.3.2 Joint likelihood of $(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1)$

We consider the joint likelihood of $(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1 \mid \boldsymbol{y})$ from (4.9) and (4.11)

$$L(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1 \mid \boldsymbol{y}) = \int_{\mathbb{R}^{np}} f(\boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}) f(\boldsymbol{z} \mid \boldsymbol{\Sigma}_1) d\boldsymbol{z}, \tag{4.14}$$

$$
\begin{aligned}
& L(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1 \mid \boldsymbol{y}) \\
\propto\ & |\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{01}|^{-\frac{p}{2}} |\boldsymbol{\Sigma}_1|^{-\frac{n-2}{2}} \exp\Big\{-\frac{1}{2}\boldsymbol{y}' \left(\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q}\right) \left[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}\boldsymbol{\Sigma}_1^{-1}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\right]^{-1} \boldsymbol{y}\Big\} \\
& \int_{\mathbb{R}^{np}} \exp\big[-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_z)' \left(\boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q}\right) (\boldsymbol{z} - \boldsymbol{\mu}_z)\big] d\boldsymbol{z} \\
\propto\ & |\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{01}|^{-\frac{p}{2}} |\boldsymbol{\Sigma}_1|^{-\frac{n-2}{2}} \exp\Big\{-\frac{1}{2}\boldsymbol{y}' \left(\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q}\right) \left[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}\boldsymbol{\Sigma}_1^{-1}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\right]^{-1} \boldsymbol{y}\Big\} \\
& |\boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q}|^{-\frac{1}{2}}. \qquad (4.15)
\end{aligned}
$$

4.3.3 Re-parameterization of $(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_1)$

Note that in (4.13) and (4.15), the matrix

$$
\begin{aligned}
& \left(\boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q}\right) \left[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}\boldsymbol{\Sigma}_1^{-1}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\right]^{-1} \\
=\ & \boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1} - (\boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1})(\boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q})^{-1}(\boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1}) \qquad (4.16)
\end{aligned}
$$

is symmetric, because the matrices $\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \boldsymbol{\Sigma}_1, \boldsymbol{Q}$ are all symmetric.

However, matrices $\boldsymbol{\Sigma}_{00}\boldsymbol{\Sigma}_1^{-1}, \boldsymbol{\Sigma}_{01}\boldsymbol{Q}$ are asymmetric. Thus we apply the following re-parameterization for efficient computation and model interpretation.

We decompose the matrix $\boldsymbol{\Sigma}_1$ with the corresponding correlation matrix $\mathbf{R}_1$, which will be discussed later, as

$$
\boldsymbol{\Sigma}_1 = \boldsymbol{\Delta}\mathbf{R}_1\boldsymbol{\Delta}, \qquad (4.17)
$$

where $\boldsymbol{\Delta} = diag(\sigma_1, \ldots, \sigma_p)$ describes the prior variation for each carve. We consider the smoothing parameter, as the noise-to-signal ratio between model variation over time $\boldsymbol{\Sigma}_{00} = diag(\delta_1, \ldots, \delta_p)$ and $\boldsymbol{\Delta}$, to be

$$
\mathbf{H} = diag(\eta_1, \ldots, \eta_p) = diag(\delta_1/\sigma_1^2, \ldots, \delta_p/\sigma_p^2). \qquad (4.18)
$$

For notation, We define $\boldsymbol{D}^{\frac{1}{2}} = \mathrm{diag}(d_1^{\frac{1}{2}}, \ldots, d_p^{\frac{1}{2}})$, for any diagonal matrix $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_p)$.

Let $\boldsymbol{\Xi} = \mathbf{H}^{\frac{1}{2}}\mathbf{R}_1^{-1}\mathbf{H}^{\frac{1}{2}}$. Clearly, $\boldsymbol{\Xi}$ is symmetric, and $|\boldsymbol{\Xi}| = |\mathbf{R}_1|^{-1}\prod_{j=1}^{p}\eta_j$. Then

$$\boldsymbol{\Sigma}_{00} = \boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{\Sigma}_{00}^{\frac{1}{2}},\ \boldsymbol{\Sigma}_1^{-1} = \boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}. \tag{4.19}$$

For the smoothing structure matrix $\boldsymbol{Q}$ defined in (4.5), which has the rank $n-2$, there exists a matrix $\boldsymbol{G}_{(n-2)\times n}$ with rank $n-2$, such that $\boldsymbol{Q} = \boldsymbol{G}'\boldsymbol{G}$. Note that $\boldsymbol{Q} = \boldsymbol{F}_0'\boldsymbol{F}_1^{-1}\boldsymbol{F}_0$. Let the LU decomposition $\boldsymbol{F}_1 = \boldsymbol{L}\boldsymbol{L}'$, where $\boldsymbol{L}$ is a $(n-2)\times(n-2)$ lower triangular matrix. Then $\boldsymbol{G} = \boldsymbol{L}^{-1}\boldsymbol{F}_0$.

Thus the matrix (4.16) can be rewritten as

$$\begin{aligned}
& \left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G}'\boldsymbol{G}\right)\left[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\boldsymbol{G})\right]^{-1} \\
= & \left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{\Xi} \otimes \boldsymbol{G}'\right)\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G}\right) \\
& \left[\boldsymbol{I}_{np} - \left(\boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{\Xi} \otimes \boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)(\boldsymbol{I}_{(n-2)p} + \boldsymbol{\Xi} \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}')^{-1}\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G}\right)\right] \\
= & (\boldsymbol{I}_p \otimes \boldsymbol{G})'\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2}\right)\boldsymbol{W}\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2}\right)(\boldsymbol{I}_p \otimes \boldsymbol{G}),
\end{aligned} \tag{4.20}$$

where, by Lemma 4.8.1, the symmetric matrix $\boldsymbol{W}$ is

$$\begin{aligned}
\boldsymbol{W} &= \boldsymbol{\Xi} \otimes \boldsymbol{I}_{n-2} - \left(\boldsymbol{\Xi}^2 \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)\left(\boldsymbol{I}_{(n-2)p} + \boldsymbol{\Xi} \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)^{-1} \\
&= (\boldsymbol{\Xi} \otimes \boldsymbol{I}_{n-2})\left(\boldsymbol{I}_{(n-2)p} + \boldsymbol{\Xi} \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)^{-1} \\
&= \left(\boldsymbol{\Xi}^{-1} \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)^{-1} \\
&= \left(\mathbf{H}^{-\frac{1}{2}}\mathbf{R}_1\mathbf{H}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)^{-1}.
\end{aligned} \tag{4.21}$$

Let the spectral decompositions that $\mathbf{H}^{-\frac{1}{2}}\mathbf{R}_1\mathbf{H}^{-\frac{1}{2}} = \boldsymbol{P}_1\boldsymbol{D}_1\boldsymbol{P}_1'$, $\boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}' = \boldsymbol{P}_0\boldsymbol{D}_0\boldsymbol{P}_0'$.

Note that $\boldsymbol{P}_1\boldsymbol{P}_1' = \boldsymbol{I}_p$, $\boldsymbol{P}_0\boldsymbol{P}_0' = \boldsymbol{I}_{n-2}$. From (4.21),

$$\begin{aligned} \boldsymbol{W} &= (\boldsymbol{P}_1\boldsymbol{D}_1\boldsymbol{P}_1' \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{P}_0\boldsymbol{D}_0\boldsymbol{P}_0')^{-1} \\ &= (\boldsymbol{P}_1 \otimes \boldsymbol{P}_0)\,(\boldsymbol{D}_1 \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{D}_0)^{-1}\,(\boldsymbol{P}_1 \otimes \boldsymbol{P}_0)'. \end{aligned} \tag{4.22}$$

Therefore, from (4.15), the joint likelihood of $(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \mathbf{R}_1, \mathbf{H} \mid \boldsymbol{y})$ is

$$\begin{aligned} & L(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \mathbf{R}_1, \mathbf{H} \mid \boldsymbol{y}) \\ \propto\ & C_0 \exp\big[-\tfrac{1}{2}\boldsymbol{y}'(\boldsymbol{I}_p \otimes \boldsymbol{G})' \left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2}\right) \boldsymbol{W} \left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2}\right) (\boldsymbol{I}_p \otimes \boldsymbol{G})\boldsymbol{y}\big], \end{aligned} \tag{4.23}$$

where

$$\begin{aligned} C_0 &= |\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_{01}|^{-\frac{p}{2}} |\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \boldsymbol{\Xi} \boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}|^{\frac{n-2}{2}} |\boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \boldsymbol{\Xi} \boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G}'\boldsymbol{G}|^{-\frac{1}{2}} \\ &= |\boldsymbol{\Sigma}_{00}|^{-\frac{n-2}{2}} |\boldsymbol{\Xi}^{-1} \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'|^{-\frac{1}{2}} \\ &= |\boldsymbol{\Sigma}_{00}|^{-\frac{n-2}{2}} |\boldsymbol{D}_1 \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{D}_0|^{-\frac{1}{2}}. \end{aligned} \tag{4.24}$$

4.3.4 Structure of Correlation Matrix $\mathbf{R}_1$

To simplify the model, we choose the exchangeable structure for correlation matrix $\mathbf{R}_1$,

$$\mathbf{R}_1 = \begin{pmatrix} 1 & \rho_1 & \dots & \rho_1 \\ \rho_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & & \rho_1 \\ \rho_1 & \dots & \rho_1 & 1 \end{pmatrix}_{p \times p}. \tag{4.25}$$

Note that, in order to guarantee that the correlation matrix $\mathbf{R}_1$ is positive definite, ρ_1 should satisfy that

$$-\frac{1}{p-1} < \rho_1 < 1. \tag{4.26}$$

4.3.5 Posterior density

We choose the independent Jeffreys prior for $\mathbf{\Sigma}_{00} = \text{diag}(\delta_1, \ldots, \delta_p)$

$$[\delta_j] \propto \delta_j^{-1}, \quad j = 1, \ldots, p. \tag{4.27}$$

We consider the independent scaled Pareto prior (Cheng and Speckman 2012) for $\mathbf{H} = \text{diag}(\eta_1, \ldots, \eta_p)$,

$$[\eta_j] = \frac{c}{(c+\eta_j)^2}, \quad \eta_j > 0, \quad j = 1, \ldots, p. \tag{4.28}$$

We use constant prior for correlation parameters ρ_0, ρ_1

$$[\rho_0] \propto 1, \quad 0 < \rho_1 < 1. \tag{4.29}$$

$$[\rho_1] \propto 1, \quad -\frac{1}{p-1} < \rho_1 < 1. \tag{4.30}$$

With the likelihood (4.23), and priors (4.27), (4.28), (4.29), (4.30), the joint posterior density of parameters $(\mathbf{\Sigma}_{00}, \mathbf{\Sigma}_{01}, \mathbf{R}_1, \mathbf{H})$ is

$$\begin{aligned}
&[\mathbf{\Sigma}_{00}, \mathbf{\Sigma}_{01}, \mathbf{R}_1, \mathbf{H} \mid \boldsymbol{y}] \\
&\propto \prod_{j=1}^{p} (c+\eta_j)^{-2} |\mathbf{\Sigma}_{00}|^{-\frac{n}{2}} |\boldsymbol{D}_1 \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{D}_0|^{-\frac{1}{2}} \\
&\quad \exp\Big\{-\frac{1}{2}\boldsymbol{y}'(\mathbf{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G}')\boldsymbol{W}(\mathbf{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G})\boldsymbol{y}\Big\}.
\end{aligned} \tag{4.31}$$

4.3.6 Sampling method for $(\mathbf{\Sigma}_{00}, \mathbf{\Sigma}_{01}, \mathbf{R}_1, \mathbf{H})$

As defined in (4.4) and (4.18), $(\mathbf{\Sigma}_{00}, \mathbf{\Sigma}_{01}, \mathbf{R}_1, \mathbf{H})$ are uniquely determined by $k = 2p + 2$ parameters $(\delta_1, \ldots, \delta_p, \eta_1, \ldots, \eta_p, \rho_0, \rho_1)$. We use the multivariate ratio-of-uniform method (Wakefield, Gelfand, and Smith 1991) to sample from the marginal joint posterior density (4.31), which is denoted as unnormalized density $f(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) = (\delta_1, \ldots, \delta_p, \eta_1, \ldots, \eta_p, \rho_0, \rho_1)$.

Note that if $(u, \boldsymbol{v})$ are uniform on

$$\mathcal{A} = \left\{(u, \boldsymbol{v}) : \boldsymbol{v} = (v_1, \ldots, v_k) \in \mathbb{R}^k, 0 < u < [f(\boldsymbol{\mu} + \boldsymbol{v}/u^r)]^{\frac{1}{rk+1}}\right\}, \tag{4.32}$$

for some $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k) \in \mathbb{R}^k$, then $\boldsymbol{\theta} = \boldsymbol{\mu} + \boldsymbol{v}/u^r$ has the density $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{\theta})$. Usually we take the posterior mode of $f(\boldsymbol{\theta})$ as $\boldsymbol{\mu}$ to maximize the acceptance rate.

We first construct the minimal bounding rectangle for uniform sampling

$$\begin{aligned} a^+ &= \sup\left([f(\boldsymbol{\theta})]^{\frac{1}{rk+1}}\right), && (4.33)\\ b_l^+ &= \sup\left((\theta_l - \mu_l)[f(\mu_1)]^{\frac{r}{rk+1}}\right),\ l = 1, \ldots, k, && (4.34)\\ b_l^- &= \inf\left((\theta_l - \mu_l)[f(\mu_1)]^{\frac{r}{rk+1}}\right),\ l = 1, \ldots, k. && (4.35) \end{aligned}$$

We take $r = 1$, which yields the standard ratio-of uniform method.

The sampling algorithm is as follows.

Algorithm 4 Multivariate ration-of-uniform sampling of $\boldsymbol{\theta} = (\delta_1, \ldots, \delta_p, \eta_1, \ldots, \eta_p, \rho_0, \rho_1)$

1: Independently sampling u from $Unif(0, a^+)$, v_l from $Unif(b_l^-, b_l^+)$, $l = 1, \ldots, k$;

2: If $0 < u < (f(\boldsymbol{\mu} + \boldsymbol{v}/u^r))^{\frac{1}{rk+1}}$, report $\boldsymbol{\theta} = \boldsymbol{\mu} + \boldsymbol{v}/u^r$; otherwise return to Step 1.

4.3.7 Efficient conditional distribution of $\boldsymbol{z}$

In the previous chapter, we obtain the posterior sample of , and we may directly sample the trend estimates from Theorem 2. However, in order to reduce the computation load, we introduce the following efficient sampling method for $\boldsymbol{z}$, which is inspired by Sun, Ni, and Speckman (2014).

With the re-parameterization (4.19) and decomposition $\mathbf{H}^{-\frac{1}{2}}\mathbf{R}_1\mathbf{H}^{-\frac{1}{2}} = \boldsymbol{P}_1\boldsymbol{D}_1\boldsymbol{P}_1'$, where $\boldsymbol{D}_1 = diag(d_{11}, \ldots, d_{1p})$, estimator (4.8)

$$\begin{aligned}
\hat{\boldsymbol{z}} &= \left[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\mathbf{H}^{\frac{1}{2}}\mathbf{R}_1^{-1}\mathbf{H}^{\frac{1}{2}}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\right]^{-1}\boldsymbol{y} \\
&= \left[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{P}_1\boldsymbol{D}_1^{-1}\boldsymbol{P}_1'\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{Q})\right]^{-1}\boldsymbol{y} \\
&= (\boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{P}_1 \otimes \boldsymbol{I}_n)(\boldsymbol{I}_{np} + \boldsymbol{D}_1^{-1} \otimes \boldsymbol{\Sigma}_{01}\boldsymbol{Q})^{-1}(\boldsymbol{P}_1'\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_n)\boldsymbol{y}. \qquad (4.36)
\end{aligned}$$

Let $\boldsymbol{\Psi} = \boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{P}_1$, $(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{I}_n)\hat{\boldsymbol{z}} = vec(\hat{\boldsymbol{Z}}\boldsymbol{\Psi}^{-\top})$, $(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} = vec(\boldsymbol{Y}\boldsymbol{\Psi}^{-\top})$,

$$vec(\hat{\boldsymbol{Z}}\boldsymbol{\Psi}^{-\top}) = (\boldsymbol{I}_{np} + \boldsymbol{D}_1^{-1} \otimes \boldsymbol{\Sigma}_{01}\boldsymbol{Q})^{-1}vec(\boldsymbol{Y}\boldsymbol{\Psi}^{-\top}). \qquad (4.37)$$

Let $\hat{\boldsymbol{V}} = \hat{\boldsymbol{Z}}\boldsymbol{\Psi}^{-\top} = [\hat{\boldsymbol{v}}_1 \cdots \hat{\boldsymbol{v}}_p]$, $\boldsymbol{U} = \boldsymbol{Y}\boldsymbol{\Psi}^{-\top} = [\boldsymbol{u}_1 \cdots \boldsymbol{u}_p]$,

$$\hat{\boldsymbol{v}}_j = (\boldsymbol{I}_n + d_{1j}^{-1}\boldsymbol{\Sigma}_{01}\boldsymbol{Q})^{-1}\boldsymbol{u}_j. \qquad (4.38)$$

The precision matrix in (4.12)

$$\begin{aligned}
\boldsymbol{\Omega} &= \boldsymbol{\Sigma}_{00}^{-1} \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{\Sigma}_1^{-1} \otimes \boldsymbol{Q} \\
&= (\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{P}_1 \otimes \boldsymbol{I}_n)(\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{D}_1^{-1} \otimes \boldsymbol{Q})(\boldsymbol{P}_1'\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_n) \\
&= (\boldsymbol{\Psi}^{-\top} \otimes \boldsymbol{I}_n)(\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}_{01}^{-1} + \boldsymbol{D}_1^{-1} \otimes \boldsymbol{Q})(\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{I}_n). \qquad (4.39)
\end{aligned}$$

Note that the matrices $\boldsymbol{I}_{np}+\boldsymbol{D}_1^{-1}\otimes\boldsymbol{\Sigma}_{01}\boldsymbol{Q}$ and $\boldsymbol{I}_p\otimes\boldsymbol{\Sigma}_{01}^{-1}+\boldsymbol{D}_1^{-1}\otimes\boldsymbol{Q}$ are block diagonal. Thus, the conditional distribution (4.12) is equivalent to

$$\begin{aligned}((\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{I}_n)\boldsymbol{z} \quad &| \quad \boldsymbol{\Sigma}_{00},\boldsymbol{\Sigma}_{01},\boldsymbol{\Sigma}_1;\boldsymbol{y})\\ &\sim \quad N_{np}((\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{I}_n)\hat{\boldsymbol{z}},(\boldsymbol{\Psi}^{-1}\otimes\boldsymbol{I}_n)\boldsymbol{\Omega}^{-1}(\boldsymbol{\Psi}^{-\top}\otimes\boldsymbol{I}_n)),\end{aligned} \tag{4.40}$$

or

$$(\boldsymbol{v}_j \mid \boldsymbol{\Sigma}_{00},\boldsymbol{\Sigma}_{01},\boldsymbol{\Sigma}_1;\boldsymbol{y}) \overset{\text{indep.}}{\sim} N_n(\hat{\boldsymbol{v}}_j,\boldsymbol{\Omega}_j^{-1}), \quad j=1,\ldots,p, \tag{4.41}$$

where $\hat{\boldsymbol{v}}_j=(\boldsymbol{I}_n+d_{1j}^{-1}\boldsymbol{\Sigma}_{01}\boldsymbol{Q})^{-1}\boldsymbol{u}_j$, and $\boldsymbol{\Omega}_j=(\boldsymbol{\Sigma}_{01}^{-1}+d_{1j}^{-1}\boldsymbol{Q})$.

Algorithm 5 Sampling $\boldsymbol{z}$ given $\boldsymbol{y}$, $\boldsymbol{\theta}=(\delta_1,\ldots,\delta_p,\eta_1,\ldots,\eta_p,\rho_0,\rho_1)$

1: Given $\boldsymbol{y},\boldsymbol{\theta}$, obtain $\boldsymbol{u}_j,\hat{\boldsymbol{v}}_j,\boldsymbol{\Omega}_j$, $j=1,\ldots,p$;

2: Sample $\boldsymbol{V}=[\boldsymbol{v}_1\cdots\boldsymbol{v}_p]$ from independent multivariate normal distribution (4.41);

3: $\boldsymbol{Z}=\boldsymbol{V}\boldsymbol{\Psi}'$, $\boldsymbol{z}=vec(\boldsymbol{Z})$.

4.3.8 Model Selection

Our full model is flexible and easy to interpret. However we are also interested in the possibility to simplify the model and to reduce the number of parameters.

We consider two reduced models, which are simplified from the full model:

Reduced Model 1. Let

$$\mathbf{R}_1=\boldsymbol{I}_p, \tag{4.42}$$

which is equivalent to apply the univariate smoothing spline model for each series.

Reduced Model 2. Let

$$\mathbf{H} = \eta \boldsymbol{I}_p, \tag{4.43}$$

which assumes the p splines share the same smoothness.

We compare the full model with the two reduced models using the deviance information criterion (DIC). Following Gelman et al. (2013),

$$DIC = -2logp(y|\hat{\boldsymbol{\theta}}_{Bayes}) + 2p_{DIC},$$

where, with the posterior mean $\hat{\boldsymbol{\theta}}_{Bayes}$, the effective number of parameters p_{DIC} is defined as

$$p_{DIC} = 2\left(logp(y|\hat{\boldsymbol{\theta}}_{Bayes}) - E_{post}(logp(y|\boldsymbol{\theta}))\right). \tag{4.44}$$

Given the posterior samples $\boldsymbol{\theta}^s$, $s = 1, ..., S$, (4.44) is computed as

$$p_{DIC} = 2\left(logp(y|\hat{\boldsymbol{\theta}}_{Bayes}) - \frac{1}{S}\sum_{s=1}^{S}(logp(y|\boldsymbol{\theta}^s))\right). \tag{4.45}$$

We use the Covid-19 daily data in chapter 4.5.2 to compute the DIC for each model. $DIC_{Full} = -423.6028$, $DIC_{ReduceModel_1} = -372.5568$, $DIC_{ReduceModel_2} = -229.715$, which suggests that the full model is preferred.

4.4 Simulation Study

In this section, we evaluate the performance of our multivariate Bayesian smoothing spline with dependency model through two simulation studies. In the first simulation

study, we are interested in the accuracy of estimating the trend. In the second simulation study, we are interested in the accuracy of recovering the model parameters $\mathbf{\Sigma}_{00}, \mathbf{\Sigma}_{01}$ in (4.1).

4.4.1 Simlulation 1

We generate data from the basic structure model (Pfeffermann 1991),

$$y_t = m_t + s_t + e_t, e_t \sim N(0, \sigma_1^2),\ t = 1, \ldots, n, \tag{4.46}$$

where the seasonal component s_t follows

$$s_t = -\sum_{j=t-12+1}^{t-1} s_j + \epsilon_t,\ \epsilon_t \sim N(0, \sigma_2^2). \tag{4.47}$$

We simulate two series of data for multivariate analysis, and apply the univariate Bayesian smoothing spline with dependency model to each series for comparison. For each series, we take 20 knots with the period 5.

For Series 1, we specify the "true" trend m_t in (4.46) as
Trend 1:

$$m_t = 10 \times exp\{-\frac{(t-16)^2}{2 \times 10^2}\}. \tag{4.48}$$

The variation parameters are $\sigma_1^2 = 0.01$, $\sigma_2^2 = 0.04$, which are select to obtain the reasonable signal-noise ratio. Figure 4.1 shows one of the 1000 simulated data with the true trend and estimation.

For Series 2, we specify the "true" trend m_t in (4.46) as

Trend 2:

$$m_t = 10 \times exp\{\frac{(t-16)^2}{2 \times 10^2}\}. \tag{4.49}$$

The variation parameters are $\sigma_1^2 = 1$, $\sigma_2^2 = 1$. Figure 4.2 shows one of the 1000 simulated data with the true trend and estimation.

In Figure 4.3, we compare the relative bias (rBias) of trend estimation between the multivariate method with univariate method. At each knot i with the estimate $\hat{m}_i$ and true trend m_i, the relative bias at the knot i is defined as

$$rBias = \frac{1}{1000} \sum_{iter=1}^{1000} \frac{\hat{m}_i - m_i}{m_i}. \tag{4.50}$$

The multivariate method improves the trend estimation accuracy, the average bias is smaller compared with the univariate method. We also show the standard deviation of estimates in Figure 4.4. In Figure 4.5, we compare the root-mean-square deviation (RMSD) of trend estimation between the multivariate method with univariate method. At each knot i with the estimate $\hat{m}_i$ and true trend m_i, the relative bias at the knot i is defined as

$$RMSE = \sqrt{\frac{1}{1000} \sum_{iter=1}^{1000} (\hat{m}_i - m_i)^2}. \tag{4.51}$$

Figure 4.5 and Figure 4.4 are almost identical, because for the MSD, the contribution from bias is one order smaller than the contribution from standard deviation. For almost all knots, the multivariate BSSD model outperform the univariate BSSD model.

4.4.2 Simlulation 2

In the second simulation study, we are interested in the accuracy of recovering the model parameters $\mathbf{\Sigma}_{00}, \mathbf{\Sigma}_{01}$ in (4.1). We define the same two true trends as
Trend 1:

$$m_t = 10 \times exp\{-\frac{(t-16)^2}{2 \times 10^2}\}.$$

Trend 2:

$$m_t = 10 \times exp\{\frac{(t-16)^2}{2 \times 10^2}\}.$$

The seasonal patterns, which are the random component, are simulated from normal distribution

$$\boldsymbol{\epsilon} = (\epsilon_{11}, \ldots, \epsilon_{n1}, \ldots, \epsilon_{1p}, \ldots, \epsilon_{np})' \sim N_{np}(\mathbf{0}, \mathbf{\Sigma}_{00} \otimes \mathbf{\Sigma}_{01}). \tag{4.52}$$

We take $n = 20, p = 2$. The $\mathbf{\Sigma}_{00} = diag(\delta_1, \delta_2)$, and $\mathbf{\Sigma}_{01}$ is only depend on correlation parameter ρ_0. We simulate 1000 datasets for each case, and show the averaged bias and root-mean-square deviation (RMSD) in Table 4.1. For different combinations of variation parameter $\mathbf{\Sigma}_{00}$ and correlation parameter ρ_0, the estimated parameters from the multivariate BSSD model have smaller bias and RMSD compared with the univariate model.

Table 4.1: Comparison of model parameter estimation

True Model	Multivariate		Univariate			
			Series 1		Series 2	
	Bias	RMSD	Bias	RMSD	Bias	RMSD
Case 1						
$\rho_0 = 0.9$	-0.07561	0.12048	-0.09514	0.16930	-0.09449	0.15271
$\delta_1 = 1$	-0.17270	0.37077	-0.13155	0.38160		
$\delta_2 = 4$	0.02647	2.18356			0.36921	2.42706
Case 2						
$\rho_0 = 0.8$	-0.07381	0.13943	-0.08381	0.16533	-0.08992	0.17743
$\delta_1 = 1$	0.04246	0.44531	0.11306	0.53739		
$\delta_2 = 4$	0.37871	2.12850			0.77972	2.60481
Case 3						
$\rho_0 = 0.9$	-0.05915	0.09872	-0.09107	0.14970	-0.07157	0.13461
$\delta_1 = 2$	-0.16159	0.89638	-0.17332	0.95827		
$\delta_2 = 6$	0.27714	3.34669			0.96570	4.25412

4.5 Applications

4.5.1 U.S. Unemployment and Employment Levels

We are interested in modeling the trends for both the unemployment and employment levels simultaneously. For the data between 01/2007 to 12/2010, which is shown in Figure 4.7, the Pearson correlation coefficient (-0.95) between the two series of raw data suggests that we expect a strong negative correlation between the trends of the Unemployment and Employment Levels.

We compare our multivariate estimation with the univariate estimation for each series. In Figure 4.9, the mean estimates from the two methods are very close. However, Figure 4.10 shows that the variation differs. For the posterior standard deviation (Post. SD) of the employment level, the multivariate BSSD model uniformly outperforms the univariate model. For the unemployment level, the Post. SD of the multivariate BSSD model is smaller than the Post. SD of the univariate BSSD model at most knots. We also notice that the Post. SD of the unemployment trend tends

to be stable, whereas the Post. SD of the employment trend seems to fluctuate with a periodic pattern. We further show the comparisons of the signal variation and the noise variation for the unemployment case and the employment case in Figure 4.11 and 4.12 respectively. The posterior standard deviation, which indicates the signal variation for the trend, is plausibly stable for each case given the scale of the noise variation. Thus the patterns in Figure 4.10 are reasonable.

4.5.2 Covid-19 daily confirmed new cases in FL and NY

We are interested in the recent development of U.S. Covid-19 cases. We often use daily (confirmed new) cases to describe the tendency of the disease's progression . For the daily cases reported in the U.S., nationwide or statewide, we often observe a seven-day pattern, which is due to the testing and reporting procedure. In order to capture the development tendency of the epidemic, we want to estimate the trend with our multivariate smoothing spline model instead of the moving average, which presents the delayed information. In Figure 4.13, we obtain the data of Florida and New York from 10/14/2020 to 11/17/2020, and compare the trend estimation from both the multivariate model and the univariate model. Figure 4.14 shows the comparison between our trend estimation and the moving average, along with the Florida data.

In Figure 4.15, we compare the posterior standard deviation between the multivariate model and the univariate model. By modeling jointly, we obtain a smaller estimation variation.

4.6 Comments

In this chapter, we constructed the multivariate Bayesian smoothing spline with dependency model, which enables us to estimate the multiple trends with seasonal errors of the same period simultaneously. We used the multivariate ratio-of-uniform method to sample from the joint posterior distribution of the smoothing, correlation and variation parameters, and proposed an efficient sampling algorithm for trend estimates.

We performed the simulation studies with different types of seasonal variation components. The accuracy and precision are improved by the joint model. The performance of recovering the simulation parameter, in terms of unbiasedness and deviation, is also elevated by modeling simultaneously. For the real data application, the multivariate Bayesian smoothing spline with dependency model achieves the trend estimations with smaller variation, compared with the univariate BSSD model.

4.7 Figures

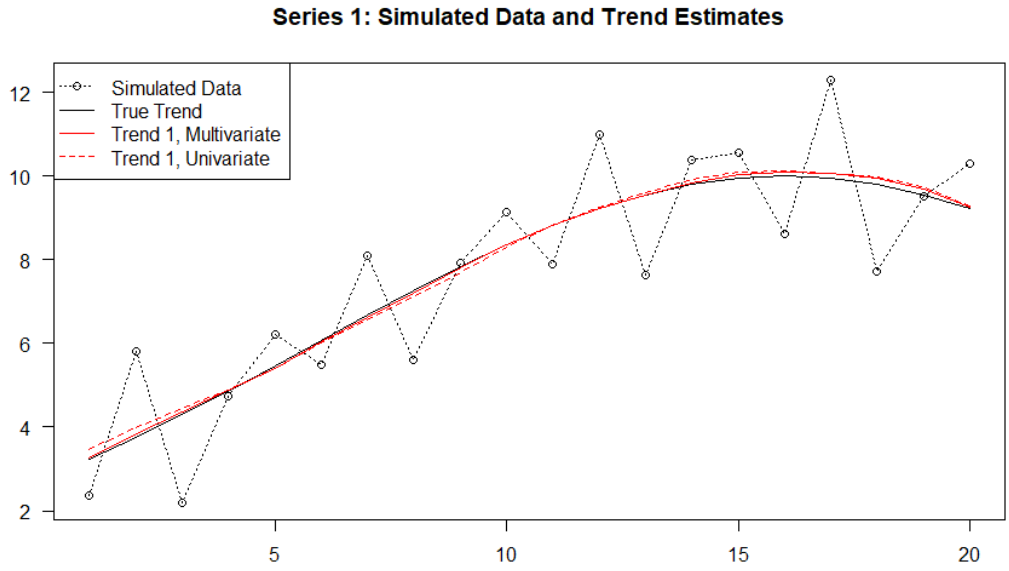

Figure 4.1: Simulation 1: Data of series 1 with true trend and estimation

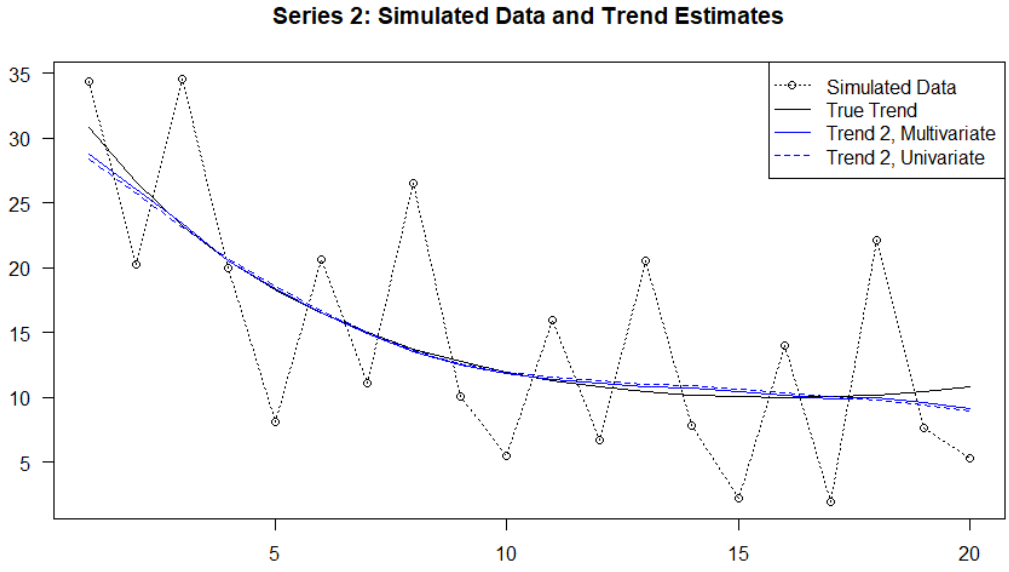

Figure 4.2: Simulation 1: Data of series 2 with true trend and estimation

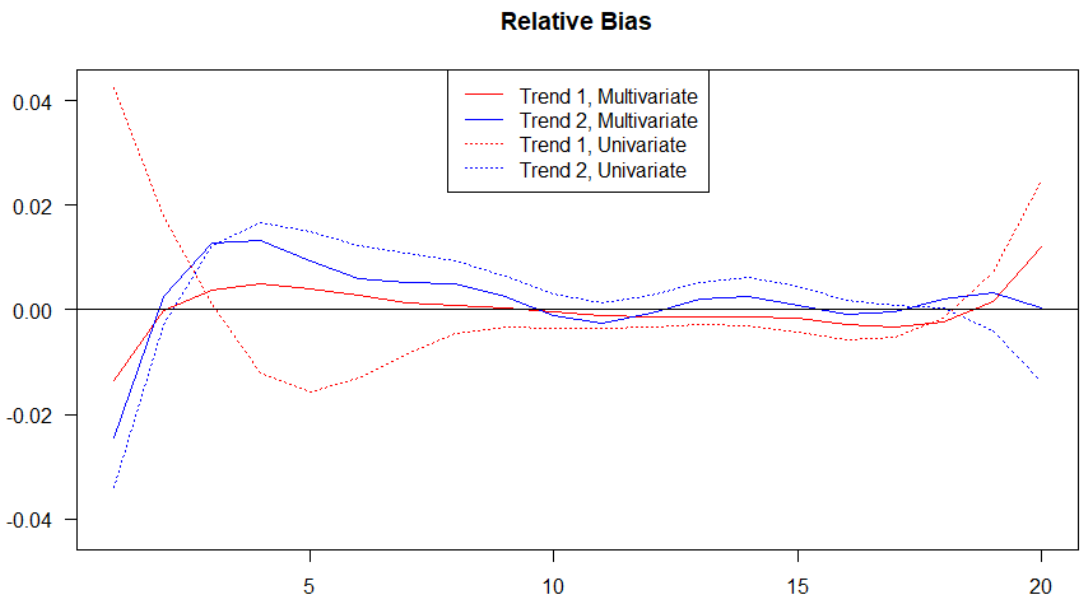

Figure 4.3: Simulation 1: Comparison of the relative bias

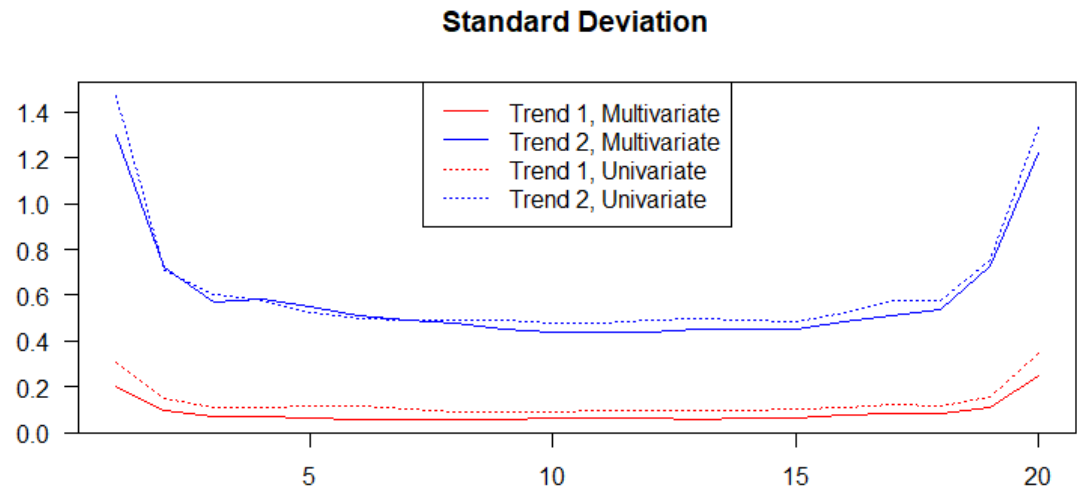

Figure 4.4: Simulation 1: Comparison of the standard deviation

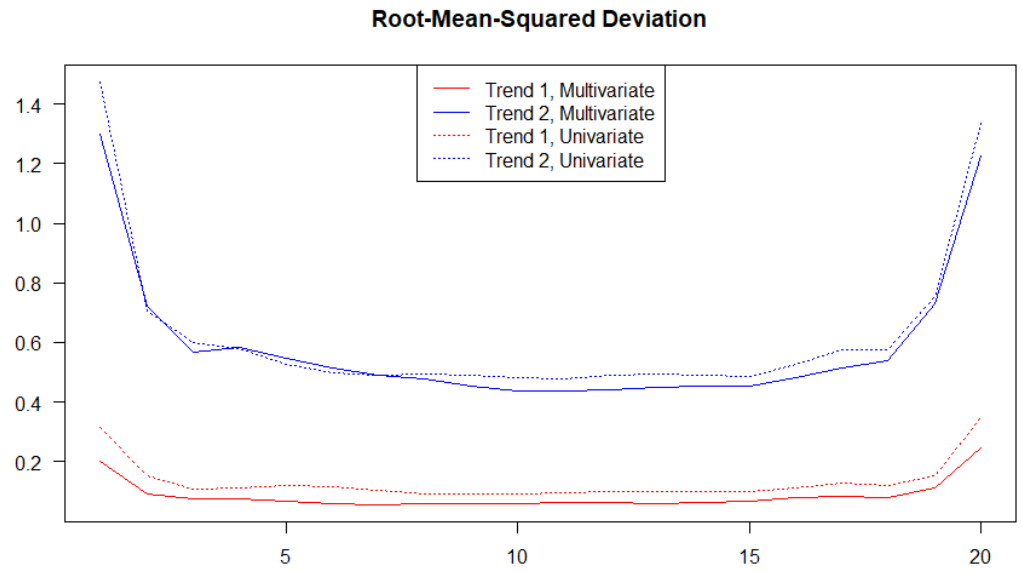

Figure 4.5: Simulation 1: Comparison of the root-mean-square deviation

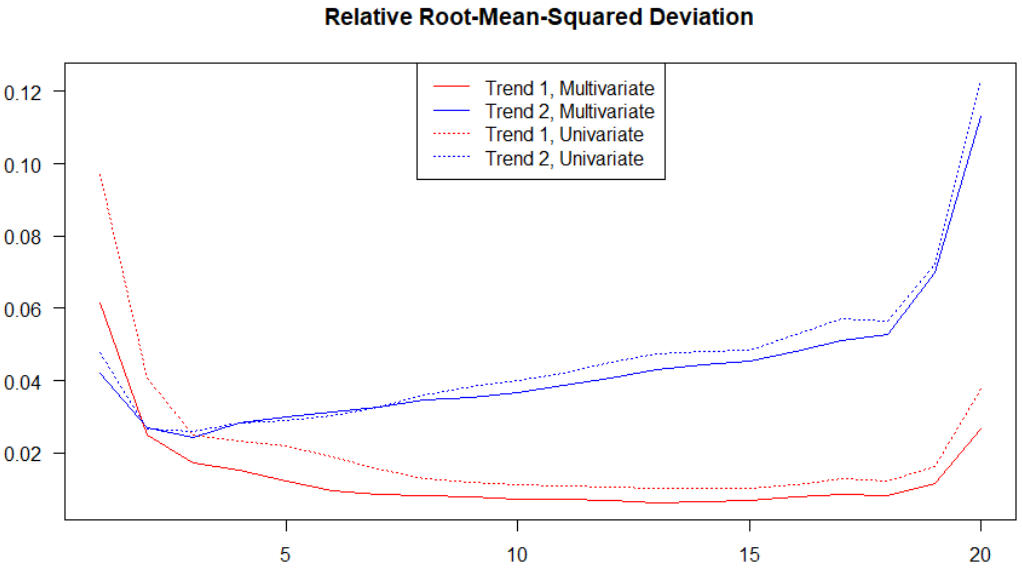

Figure 4.6: Simulation 1: Comparison of the relative root-mean-square deviation

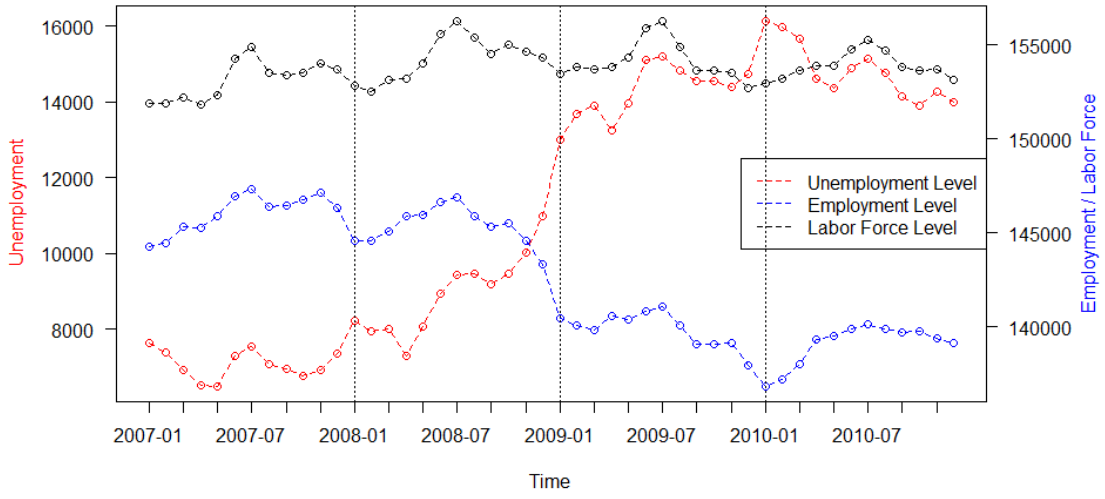

Figure 4.7: U.S. Unemployment and Employment Level from 01/2007 to 12/2010

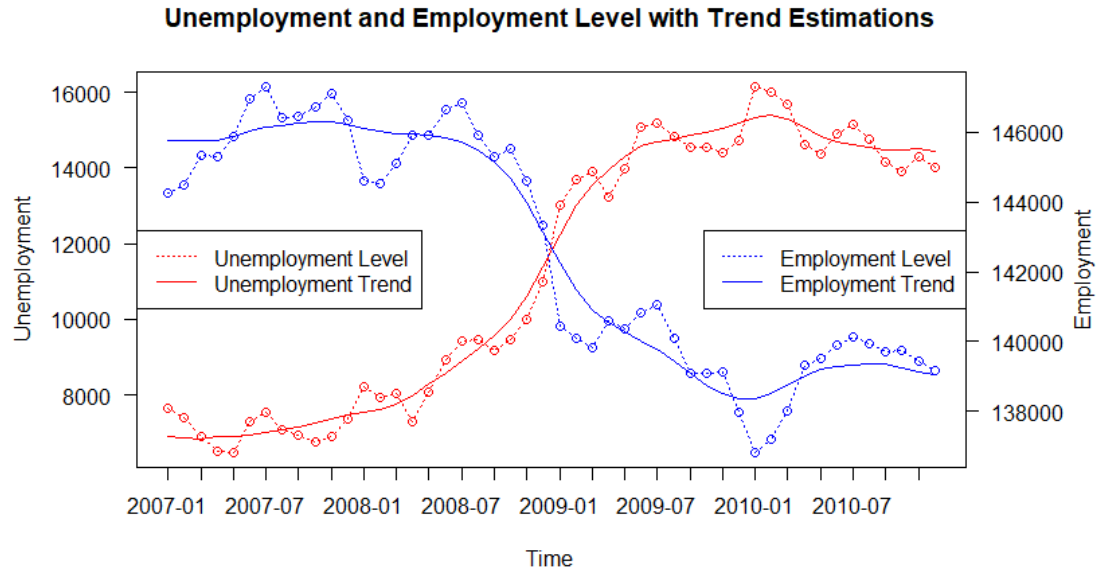

Figure 4.8: U.S. Unemployment and Employment Level with Trend Estimation from 01/2007 to 12/2010

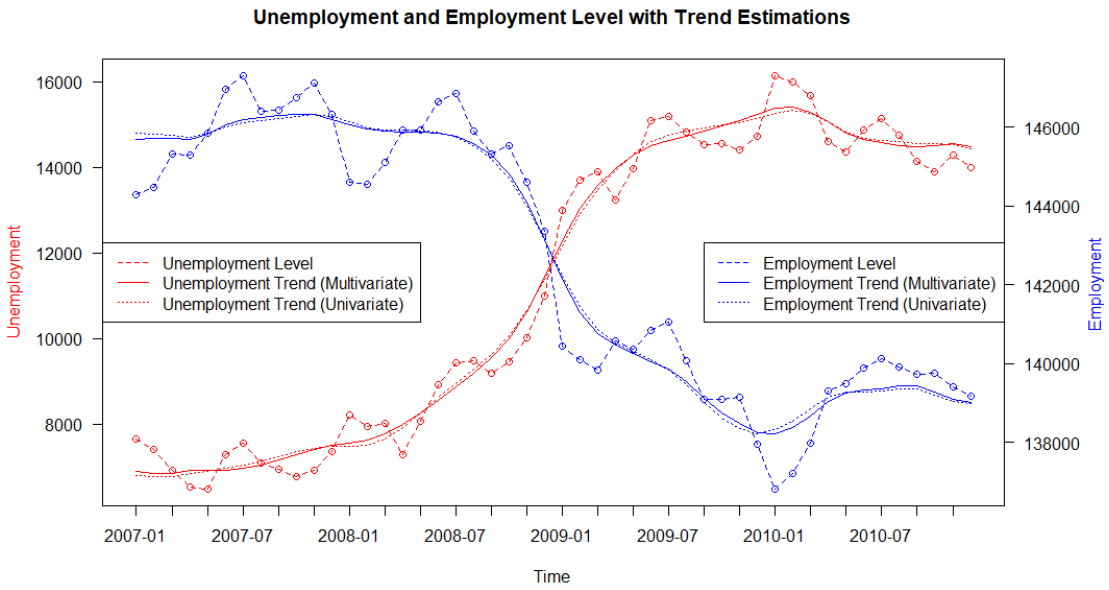

Figure 4.9: Comparison of Trend Estimation

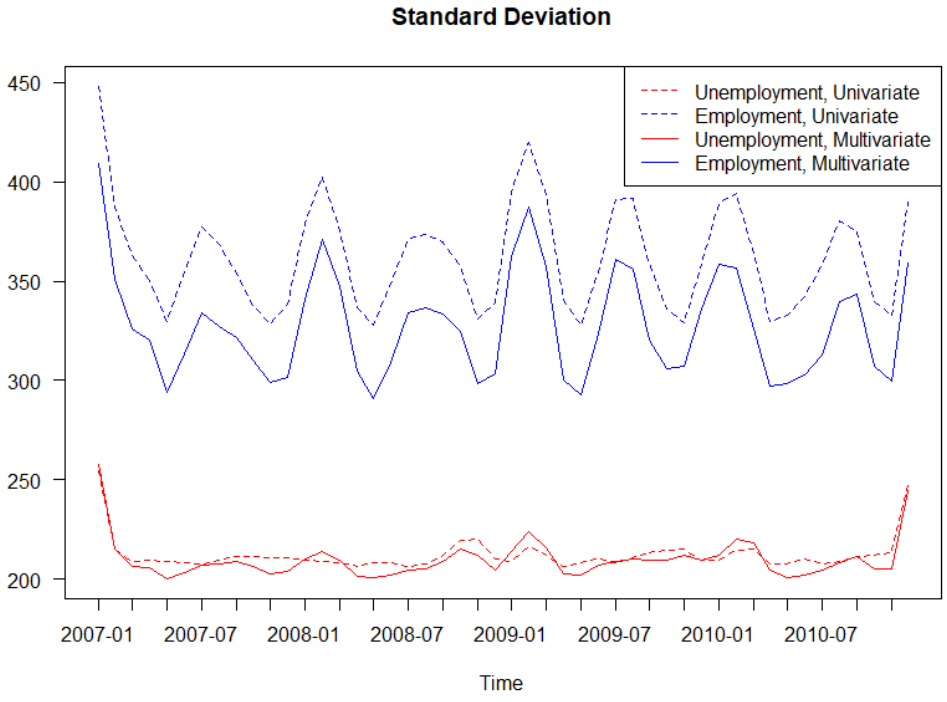

Figure 4.10: Comparison of Posterior Standard Deviation

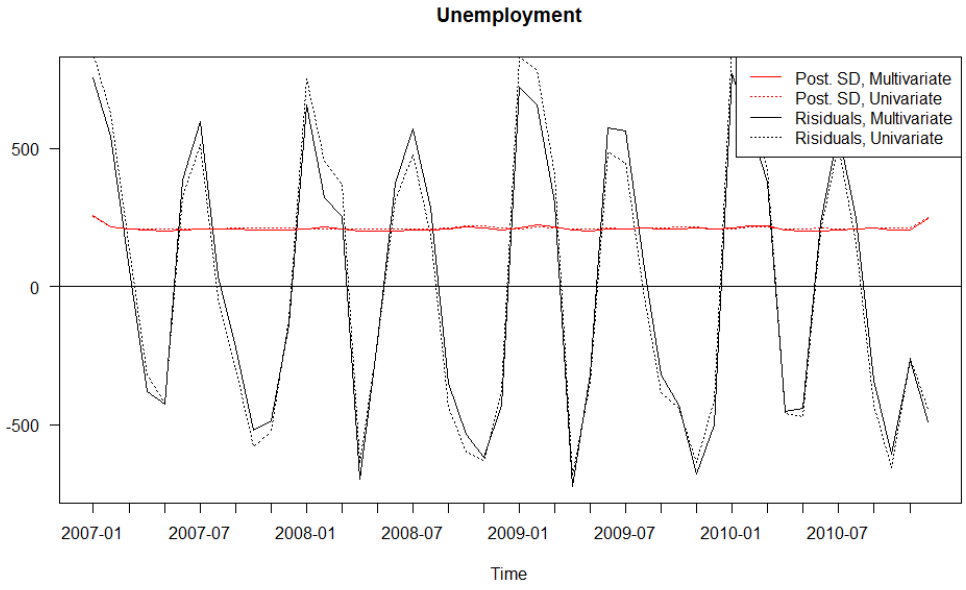

Figure 4.11: Comparison of signal variation and noise variation for unemployment

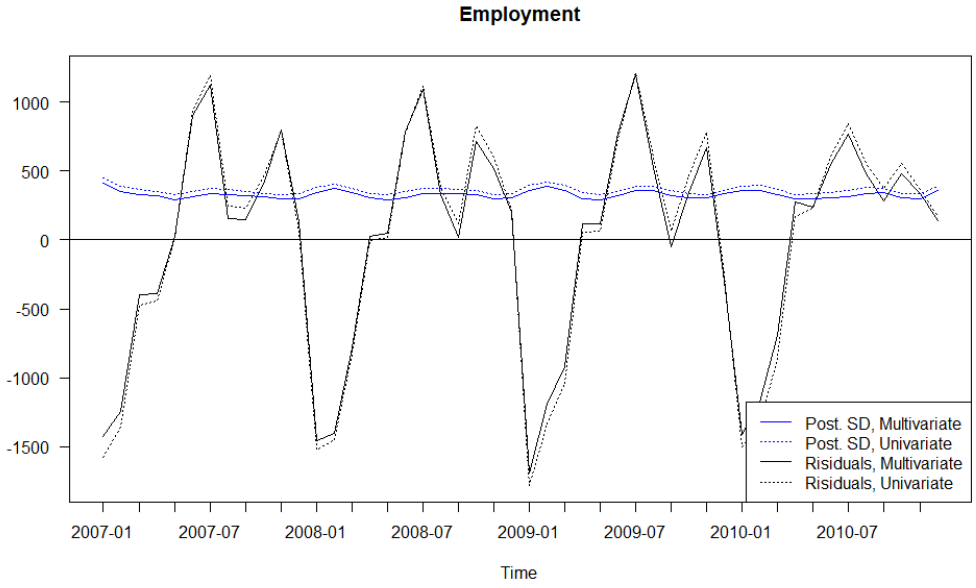

Figure 4.12: Comparison of signal variation and noise variation for employment

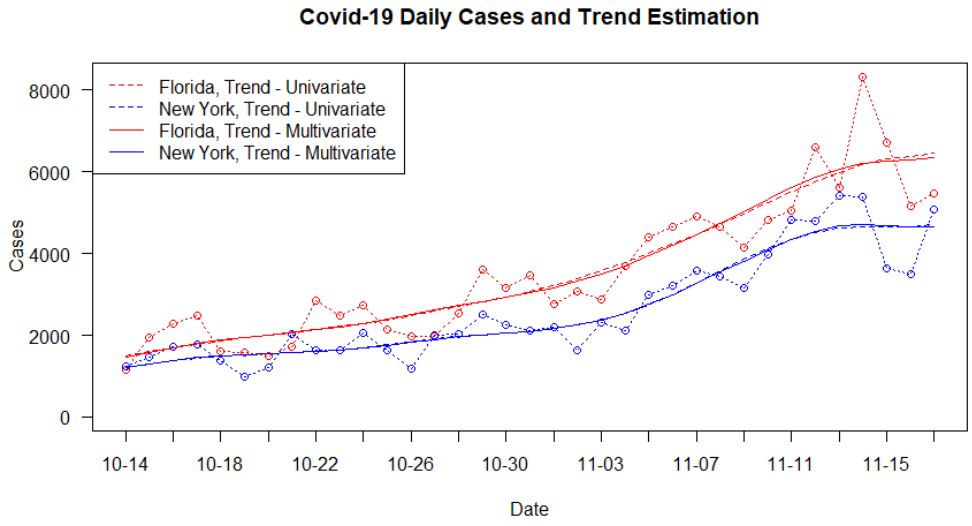

Figure 4.13: Covid-19 daily increased cases in FL and NY from Oct 14 to Nov 17

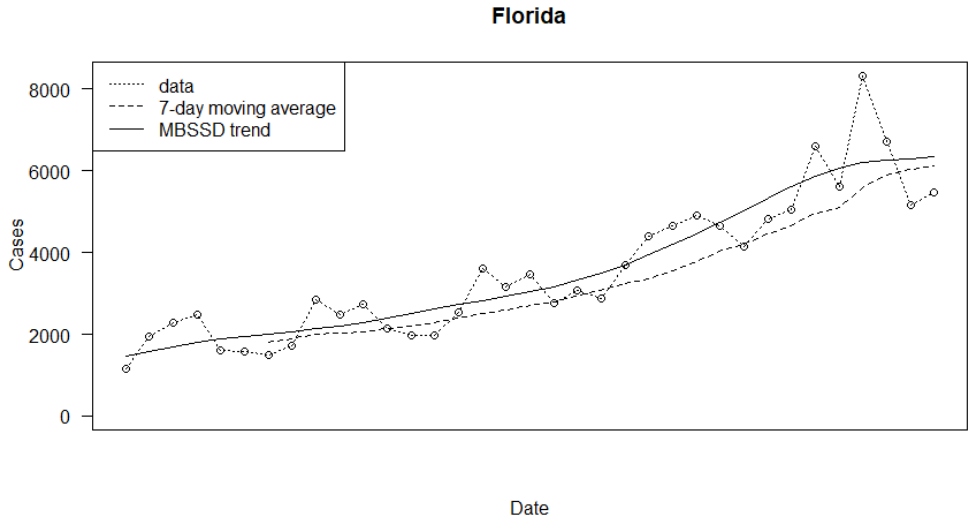

Figure 4.14: Comparison of MBSSD trend with Moving average

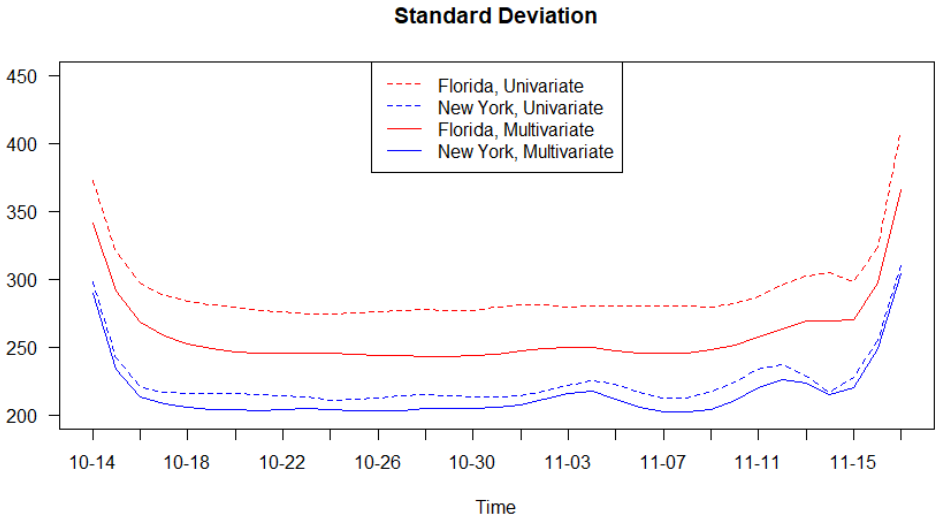

Figure 4.15: Comparison of Posterior Standard Deviation

4.8 Appendix

4.A Lemmas

To begin with, we introduce the following Lemmas.

Lemma 4.8.1 (Theorem 1.7 (Schott 1996)). *Suppose* $\boldsymbol{A}_{m\times m}$ *and* $\boldsymbol{B}_{n\times n}$ *are nonsingular matrices. For any matrices* $\boldsymbol{C}_{m\times n}$ *and* $\boldsymbol{D}_{n\times m}$*, if* $\boldsymbol{A}+\boldsymbol{CBD}$ *is nonsingular, then*

$$(\boldsymbol{A}+\boldsymbol{CBD})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{C}(\boldsymbol{B}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{C})^{-1}\boldsymbol{DA}^{-1}.$$

Lemma 4.8.2 (Theorem 16.2.1 (Harville 2008)). *For any* $m\times n$ *matrix* $\boldsymbol{A}$*,* $n\times p$ *matrix* $\boldsymbol{B}$*,* $p\times q$ *matrix* $\boldsymbol{C}$*,*

$$vec(\boldsymbol{ABC})=(\boldsymbol{C}'\otimes\boldsymbol{A})vec(\boldsymbol{B}).$$

Lemma 4.8.3 (Theorem 16.2.2 (Harville 2008)). *For any $m \times n$ matrix $\boldsymbol{A}$, $m \times p$ matrix $\boldsymbol{B}$, $p \times q$ matrix $\boldsymbol{C}$, and $n \times q$ matrix $\boldsymbol{D}$*

$$tr(\boldsymbol{A}'\boldsymbol{B}\boldsymbol{C}\boldsymbol{D}') = (vec\boldsymbol{A})'(\boldsymbol{D} \otimes \boldsymbol{B})vec\boldsymbol{C}.$$

4.B Correlation Matrix

We introduce two kinds of correlation matrices in the multivariate Bayesian smoothing spline model. The correlation matrix over time $\mathbf{R}_0$ is defined similar as the $\mathbf{R}$ in Chapter 3, see Appendix 3.B. Another prior correlation $\mathbf{R}_1$ over splines is defined in Section 4.3.4.

4.C Alternative Re-parameterization of $\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_1$

We consider another re-parameterization of $\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_1$ to explore the possibility of reducing the model. We decompose the matrix $\boldsymbol{\Sigma}_1$ with the corresponding correlation matrix $\mathbf{R}_1$, as in (4.17) and the smoothing parameter (4.18). $\boldsymbol{\Sigma}_{00}$ is re-parameterized as

$$\boldsymbol{\Sigma}_{00} = diag(\eta_1\sigma_1^2, \ldots, \eta_p\sigma_p^2) = \mathbf{H}\ diag(\sigma_1^2, \ldots, \sigma_p^2). \tag{4.53}$$

The joint likelihood (4.23) is rewritten as

$$\begin{aligned} &L(\boldsymbol{\Sigma}_1, \mathbf{H}, \boldsymbol{\Sigma}_{01} \mid \boldsymbol{y}) \\ &\quad \propto C_0 \exp\{-\tfrac{1}{2}\boldsymbol{y}'(\boldsymbol{I}_p \otimes \boldsymbol{G})'[\boldsymbol{\Sigma}_1 \otimes \boldsymbol{I}_{n-2} + \boldsymbol{\Sigma}_{00} \otimes \boldsymbol{G}\Sigma_{01}\boldsymbol{G}']^{-1}(\boldsymbol{I}_p \otimes \boldsymbol{G})\boldsymbol{y}\}, \end{aligned} \tag{4.54}$$

where

$$C_0 = |\boldsymbol{\Sigma}_1 \otimes \boldsymbol{I}_{n-2} + \boldsymbol{\Sigma}_{00} \otimes \boldsymbol{G}\Sigma_{01}\boldsymbol{G}'|^{-\frac{1}{2}}. \tag{4.55}$$

We define $\boldsymbol{S} = diag(s_1, \ldots, s_p)$ as the ratio of signal variation, where $\sigma_j = s_j\sigma, j = 1, \ldots, p$, and $s_1 = 1$. Often the $\boldsymbol{S}$ is considered known. We consider the reduced model,

$$\boldsymbol{\Sigma}_1 = \sigma^2\boldsymbol{S}\mathbf{R}_1\boldsymbol{S},\ \boldsymbol{\Sigma}_{00} = \sigma^2 diag(\eta_1 s_1^2, \ldots, \eta_p s_p^2). \tag{4.56}$$

We choose the independent Jeffreys prior for $(\sigma_1^2, \ldots, \sigma_p^2)$

$$\left[\sigma_j^2\right] \propto (\sigma_j^2)^{-1}, \quad j = 1, \ldots, p. \tag{4.57}$$

For reduced model we choose the Jeffreys prior for σ^2

$$\left[\sigma^2\right] \propto (\sigma^2)^{-1}. \tag{4.58}$$

For notation, We define $\boldsymbol{D}^{\frac{1}{2}} = \mathrm{diag}(d_1^{\frac{1}{2}}, \ldots, d_p^{\frac{1}{2}})$, for any diagonal matrix $\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_p)$.

Let $\boldsymbol{\Xi} = \mathbf{H}^{\frac{1}{2}}\mathbf{R}_1^{-1}\mathbf{H}^{\frac{1}{2}}$. Clearly, $\boldsymbol{\Xi}$ is symmetric, and $|\boldsymbol{\Xi}| = |\mathbf{R}_1|^{-1}\prod_{j=1}^p \eta_j$. Then

$$\boldsymbol{\Sigma}_{00} = \boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{\Sigma}_{00}^{\frac{1}{2}},\ \boldsymbol{\Sigma}_1^{-1} = \boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \tag{4.59}$$

For the smoothing structure matrix $\boldsymbol{Q}$ defined in (4.5), which has the rank $n-2$, there exists a matrix $\boldsymbol{G}_{(n-2)\times n}$ with rank $n-2$, such that $\boldsymbol{Q} = \boldsymbol{G}'\boldsymbol{G}$. Note that $\boldsymbol{Q} = \boldsymbol{F}_0'\boldsymbol{F}_1^{-1}\boldsymbol{F}_0$. Let the LU decomposition $\boldsymbol{F}_1 = \boldsymbol{L}\boldsymbol{L}'$, where $\boldsymbol{L}$ is a $(n-2)\times(n-2)$ lower triangular matrix. Then $\boldsymbol{G} = \boldsymbol{L}^{-1}\boldsymbol{F}_0$.

Thus the matrix (4.16) can be rewritten as

$$\begin{aligned}
& \left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G}'\boldsymbol{G}\right)\left[\boldsymbol{I}_{np} + (\boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}) \otimes (\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\boldsymbol{G})\right]^{-1} \\
= \; & \left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{\Xi} \otimes \boldsymbol{G}'\right)\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G}\right) \\
& \left[\boldsymbol{I}_{np} - \left(\boldsymbol{\Sigma}_{00}^{\frac{1}{2}}\boldsymbol{\Xi} \otimes \boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)(\boldsymbol{I}_{(n-2)p} + \boldsymbol{\Xi} \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}')^{-1}\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{G}\right)\right] \\
= \; & (\boldsymbol{I}_p \otimes \boldsymbol{G})'\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2}\right)\boldsymbol{W}\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2}\right)(\boldsymbol{I}_p \otimes \boldsymbol{G}), \qquad (4.60)
\end{aligned}$$

where, by Lemma 4.8.1, the symmetric matrix $\boldsymbol{W}$ is

$$\begin{aligned}
\boldsymbol{W} &= \boldsymbol{\Xi} \otimes \boldsymbol{I}_{n-2} - \left(\boldsymbol{\Xi}^2 \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)\left(\boldsymbol{I}_{(n-2)p} + \boldsymbol{\Xi} \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)^{-1} \\
&= (\boldsymbol{\Xi} \otimes \boldsymbol{I}_{n-2})\left(\boldsymbol{I}_{(n-2)p} + \boldsymbol{\Xi} \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)^{-1} \\
&= \left(\boldsymbol{\Xi}^{-1} \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)^{-1} \\
&= \left(\mathbf{H}^{-\frac{1}{2}}\mathbf{R}_1\mathbf{H}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'\right)^{-1}. \qquad (4.61)
\end{aligned}$$

Let spectral decomposition that $\mathbf{H}^{-\frac{1}{2}}\mathbf{R}_1\mathbf{H}^{-\frac{1}{2}} = \boldsymbol{P}_1\boldsymbol{D}_1\boldsymbol{P}_1'$, $\boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}' = \boldsymbol{P}_0\boldsymbol{D}_0\boldsymbol{P}_0'$. Note that $\boldsymbol{P}_1\boldsymbol{P}_1' = \boldsymbol{I}_p$, $\boldsymbol{P}_0\boldsymbol{P}_0' = \boldsymbol{I}_{n-2}$. From (4.61),

$$\begin{aligned}
\boldsymbol{W} &= (\boldsymbol{P}_1\boldsymbol{D}_1\boldsymbol{P}_1' \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{P}_0\boldsymbol{D}_0\boldsymbol{P}_0')^{-1} \\
&= (\boldsymbol{P}_1 \otimes \boldsymbol{P}_0)(\boldsymbol{D}_1 \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{D}_0)^{-1}(\boldsymbol{P}_1 \otimes \boldsymbol{P}_0)'. \qquad (4.62)
\end{aligned}$$

Therefore, from (4.54), the joint likelihood of $(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \mathbf{R}_1, \mathbf{H} \mid \boldsymbol{y})$ is

$$\begin{aligned}
& L(\boldsymbol{\Sigma}_{00}, \boldsymbol{\Sigma}_{01}, \mathbf{R}_1, \mathbf{H} \mid \boldsymbol{y}) \\
\propto \; & C_0 \exp\left\{-\tfrac{1}{2}\boldsymbol{y}'(\boldsymbol{I}_p \otimes \boldsymbol{G})'\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2}\right)\boldsymbol{W}\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}} \otimes \boldsymbol{I}_{n-2}\right)(\boldsymbol{I}_p \otimes \boldsymbol{G})\boldsymbol{y}\right\}, \qquad (4.63)
\end{aligned}$$

where

$$\begin{aligned} C_0 &= |\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}}|\boldsymbol{\Sigma}_{01}|^{-\frac{p}{2}}|\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}|^{\frac{n-2}{2}}|\boldsymbol{\Sigma}_{00}^{-1}\otimes\boldsymbol{\Sigma}_{01}^{-1}+\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\boldsymbol{\Xi}\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\otimes\boldsymbol{G}'\boldsymbol{G}|^{-\frac{1}{2}} \\ &= |\boldsymbol{\Sigma}_{00}|^{-\frac{n-2}{2}}|\boldsymbol{\Xi}^{-1}\otimes\boldsymbol{I}_{n-2}+\boldsymbol{I}_p\otimes\boldsymbol{G}\boldsymbol{\Sigma}_{01}\boldsymbol{G}'|^{-\frac{1}{2}} \\ &= |\boldsymbol{\Sigma}_{00}|^{-\frac{n-2}{2}}|\boldsymbol{D}_1\otimes\boldsymbol{I}_{n-2}+\boldsymbol{I}_p\otimes\boldsymbol{D}_0|^{-\frac{1}{2}}. \end{aligned} \tag{4.64}$$

Joint posterior density

$$\begin{aligned} &[\boldsymbol{\Sigma}_{00},\boldsymbol{\Sigma}_{01},\mathbf{R}_1,\mathbf{H}\mid\boldsymbol{y}] \\ &\quad\propto \prod_{j=1}^{p}(c+\eta_j)^{-2}|\boldsymbol{\Sigma}_{00}|^{-\frac{n}{2}}|\boldsymbol{D}_1\otimes\boldsymbol{I}_{n-2}+\boldsymbol{I}_p\otimes\boldsymbol{D}_0|^{-\frac{1}{2}} \\ &\qquad \exp\Big\{-\frac{1}{2}\boldsymbol{y}'(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\otimes\boldsymbol{G}')\boldsymbol{W}(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\otimes\boldsymbol{G})\boldsymbol{y}\Big\}, \end{aligned} \tag{4.65}$$

Let $\tilde{\boldsymbol{y}} = (\tilde{y}_1,\ldots,\tilde{y}_{(n-2)p})' = (\boldsymbol{I}_p\otimes\boldsymbol{G})\boldsymbol{y} = vec(\boldsymbol{G}\boldsymbol{Y})$, $\widetilde{\boldsymbol{Y}} = \mathrm{diag}(\tilde{y}_1,\ldots,\tilde{y}_{(n-2)p})$; $\boldsymbol{\delta} = (\delta_1,\ldots,\delta_p)'$. We employ the trick that

$$\left(\boldsymbol{\Sigma}_{00}^{-\frac{1}{2}}\otimes\boldsymbol{I}_{n-2}\right)(\boldsymbol{I}_p\otimes\boldsymbol{G})\boldsymbol{y} = \widetilde{\boldsymbol{Y}}(\boldsymbol{\delta}^{-\frac{1}{2}}\otimes\mathbf{1}_{n-2}), \tag{4.66}$$

and rewrite the likelihood (4.63) as

$$L(\boldsymbol{\delta},\boldsymbol{\Xi},\boldsymbol{\Sigma}_{01}\mid\boldsymbol{y}) \propto C_0\exp\Big\{-\frac{1}{2}(\boldsymbol{\delta}^{-\frac{1}{2}}\otimes\mathbf{1}_{n-2})'\left(\widetilde{\boldsymbol{Y}}\boldsymbol{W}\widetilde{\boldsymbol{Y}}\right)(\boldsymbol{\delta}^{-\frac{1}{2}}\otimes\mathbf{1}_{n-2})\Big\} \tag{4.67}$$

Let the partition of the $(n-2)p\times(n-2)p$ matrix $\widetilde{\boldsymbol{Y}}\boldsymbol{W}\widetilde{\boldsymbol{Y}}$ be

$$\widetilde{\boldsymbol{Y}}\boldsymbol{W}\widetilde{\boldsymbol{Y}} = \begin{pmatrix} \boldsymbol{W}_{11}^* & \cdots & \boldsymbol{W}_{1p}^* \\ \vdots & & \vdots \\ \boldsymbol{W}_{p1}^* & \cdots & \boldsymbol{W}_{pp}^* \end{pmatrix}, \tag{4.68}$$

with each submatrix $\boldsymbol{W}_{ij}^*$ to be a $(n-2)\times(n-2)$ matrix, and $\boldsymbol{W}_{ij}^* = \boldsymbol{W}_{ji}^{*'}$. Let $\tilde{w}_{ij} = \mathbf{1}_{n-2}'\boldsymbol{W}_{ij}^*\mathbf{1}_{n-2}$ be the summation of all elements of $\boldsymbol{W}_{ij}^*$, $1 \le i,j \le p$. Define the matrix $\widetilde{\boldsymbol{W}} = (\tilde{w}_{ij})_{p\times p}$, and the matrix $\widetilde{\boldsymbol{W}}$ is symmetric.

The likelihood (4.67) can be written as

$$L(\boldsymbol{\delta}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01} \mid \boldsymbol{y}) \propto C_1 \left[\prod_{j=1}^{p} \delta_j^{-\frac{n-2}{2}}\right] \exp\{-\frac{1}{2}(\boldsymbol{\delta}^{-\frac{1}{2}})'\widetilde{\boldsymbol{W}}\boldsymbol{\delta}^{-\frac{1}{2}}\}, \tag{4.69}$$

where $C_1 = |\boldsymbol{D}_1 \otimes \boldsymbol{I}_{n-2} + \boldsymbol{I}_p \otimes \boldsymbol{D}_0|^{-\frac{1}{2}}$.

4.C.1 Conditional distribution of $\boldsymbol{\delta}$ and joint likelihood of $\boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01}$

We choose the independent Jeffreys prior for $\boldsymbol{\delta}$

$$[\boldsymbol{\delta}] \propto \prod_{j=1}^{p} \delta_j^{-1} \tag{4.70}$$

From likelihood (4.67) and prior (4.58), the conditional distribution of $\boldsymbol{\delta}$ given $\boldsymbol{\Xi}$ and $\boldsymbol{\Sigma}_{01}$ is,

$$\begin{aligned} [\boldsymbol{\delta} \mid \boldsymbol{y}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01}] &\propto [\boldsymbol{\delta}]\, L(\boldsymbol{\delta}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01} \mid \boldsymbol{y}) \\ &\propto \left[\prod_{j=1}^{p} \delta_j^{-\frac{n}{2}}\right] \exp\{-\frac{1}{2}(\boldsymbol{\delta}^{-\frac{1}{2}})'\widetilde{\boldsymbol{W}}\boldsymbol{\delta}^{-\frac{1}{2}}\} \end{aligned} \tag{4.71}$$

The joint likelihood of $\boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01}$ is

$$L(\boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01} \mid \boldsymbol{y}) = \int_{\mathbb{R}_+^p} [\boldsymbol{\delta}] L(\boldsymbol{\delta}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01} \mid \boldsymbol{y}) d\boldsymbol{\delta}. \tag{4.72}$$

$$L(\boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01} \mid \boldsymbol{y}) \propto C_1 \int_{\mathbb{R}_+^p} \left[\prod_{j=1}^{p} \delta_j^{-\frac{n}{2}}\right] \exp\{-\frac{1}{2}(\boldsymbol{\delta}^{-\frac{1}{2}})'\widetilde{\boldsymbol{W}}\boldsymbol{\delta}^{-\frac{1}{2}}\} d\boldsymbol{\delta}, \tag{4.73}$$

where $C_2 = \int_{\mathbb{R}_+^p} \left[\prod_{j=1}^p \delta_j^{-\frac{n}{2}}\right] \exp\{-\frac{1}{2}(\boldsymbol{\delta}^{-\frac{1}{2}})'\widetilde{\boldsymbol{W}}\boldsymbol{\delta}^{-\frac{1}{2}}\} d\boldsymbol{\delta}$ is dependent on $\boldsymbol{\Xi}, \boldsymbol{\Sigma}_{01}, \boldsymbol{y}$.

4.C.2 Structure of $\mathbf{R}_0, \mathbf{R}_1$

Assuming the period of seasonal effect is T, we define the correlation matrix $\mathbf{R}_0 = (r_{ij})_{n\times n}$ with the upper-triangular elements $(1 \le i \le j \le n)$ that

$$r_{ij} = \begin{cases} 1, & \text{if } i = j, \\ \rho_0, & \text{if } j = i + T \times k,\ k = 1, 2, 3, \ldots,\ \rho_0 \in (0, 1), \\ 0, & otherwise. \end{cases} \tag{4.74}$$

If $n = mT$ for some $m \in \mathbb{N}$,

$$\mathbf{R}_0 = [(1 - \rho_0)\boldsymbol{I}_m + \rho_0 \mathbf{1}_m' \mathbf{1}_m] \otimes \boldsymbol{I}_T. \tag{4.75}$$

If $n < mT$, $\mathbf{R}_0$ is obtain by the first n columns and rows from matrix (4.75) of order mT.

For the case that $p = 2$, we define that

$$\mathbf{R}_1 = \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix},\ \rho_1 \in (-1, 1). \tag{4.76}$$

4.C.3 Prior and posterior of $\mathbf{H}, \rho_0, \rho_1$

We consider the case that $p = 2$. For $\mathbf{H} = (^{\eta_1}{}_{\eta_2})$, we consider the independent scaled Pareto prior (Cheng and Speckman 2012) for $\eta_j, j = 1, 2$

$$[\eta_j] = \frac{c}{(c + \eta_j)^2}, \eta_j > 0, \tag{4.77}$$

For $\rho_0 \in (0,1)$, we consider prior 1

$$[\rho_0] \propto (1-\rho_0)^{-\frac{1}{2}},\ 0 < \rho_0 < 1, \tag{4.78}$$

or prior 2

$$[\rho_0] \propto 1,\ 0 < \rho < 1. \tag{4.79}$$

For $\rho_1 \in (-1,1)$, we consider constant prior

$$[\rho_1] \propto 1,\ -1 < \rho < 1. \tag{4.80}$$

4.C.4 Semi-correlated mulitivariate gamma distribution

We define that the $p \times 1$ vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)'$ follows semi-correlated mulitivariate gamma distribution, denoted as

$$\boldsymbol{\delta} \sim SGam(\boldsymbol{\Omega}, k), \tag{4.81}$$

where $\boldsymbol{\Omega}_{p\times p}$ is symmetric and positive definite, integar $k > 2$, $\delta_j > 0, j = 1, \ldots, p$; if the density is

$$[\boldsymbol{\delta} \mid \boldsymbol{\Omega}, k] \propto f_{\boldsymbol{\delta}}(\boldsymbol{\delta} \mid \boldsymbol{\Omega}, k) = \left[\prod_{j=1}^{p} \delta_j^{-\frac{k}{2}}\right] \exp\{-\frac{1}{2}(\boldsymbol{\delta}^{-\frac{1}{2}})'\boldsymbol{\Omega}\boldsymbol{\delta}^{-\frac{1}{2}}\}. \tag{4.82}$$

We are interested in the quantity $m = \int_{\mathbb{R}_+^p} f_{\boldsymbol{\delta}}(\boldsymbol{\delta} \mid \boldsymbol{\Omega}, k) d\boldsymbol{\delta}$, and the sampling procedure for $\boldsymbol{\delta}$. Notice that m is the multivariate integration over hypercubes. There exist efficient methods to obtain numerical results. See Clenshaw and Curtis (1960) and Narasimhan et al. (2020).

For small p, we may to integral by substitution. Let $\boldsymbol{\tau} = \boldsymbol{\delta}^{-\frac{1}{2}} = (\delta_1^{-\frac{1}{2}}, \ldots, \delta_p^{-\frac{1}{2}})'$, $J(\boldsymbol{\delta} \to \boldsymbol{\tau}) = (-2)^p \prod \tau_j^{-3}$. Then

$$f_{\boldsymbol{\tau}}(\boldsymbol{\tau} \mid \boldsymbol{\Omega}) = 2^p \left(\prod \tau_j^{k-3} \right) \exp\{-\frac{1}{2} \boldsymbol{\tau}' \boldsymbol{\Omega} \boldsymbol{\tau}\}. \tag{4.83}$$

Let the the Cholesky decomposition $\boldsymbol{\Omega} = \boldsymbol{L}\boldsymbol{L}'$, $\boldsymbol{L}_{p \times p} = (l_{ij})$, $l_{11} > 0$. Let $\boldsymbol{\varsigma} = \boldsymbol{L}' \boldsymbol{\tau}$. Then $\boldsymbol{\tau} = \boldsymbol{L}^{-\top} \boldsymbol{\varsigma}$, $J(\boldsymbol{\tau} \to \boldsymbol{\varsigma}) = |\boldsymbol{L}|^{-1}$.

$$f_{\boldsymbol{\varsigma}}(\boldsymbol{\varsigma} \mid \boldsymbol{L}) = 2^p |\boldsymbol{L}|^{-1} \left(\prod \tau_j(\boldsymbol{\varsigma}) \right)^{k-3} \exp\{-\frac{1}{2} \boldsymbol{\varsigma}' \boldsymbol{\varsigma}\}. \tag{4.84}$$

We consider the case that $p = 2$. Denote $\boldsymbol{\varsigma} = (\varsigma_1, \varsigma_2)'$ and $\boldsymbol{L} = \left(\begin{smallmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{smallmatrix} \right)$. The density (4.84) is

$$f_{\boldsymbol{\varsigma}}(\boldsymbol{\varsigma} \mid \boldsymbol{L}) = \frac{2^2}{(l_{11} l_{22})^{k-2}} \left(\varsigma_1 \varsigma_2 - \frac{l_{21}}{l_{22}} \varsigma_2^2 \right)^{k-3} \exp\{-\frac{1}{2}(\varsigma_1^2 + \varsigma_2^2)\}, \tag{4.85}$$

We take the polar coordinates mapping $\boldsymbol{\varsigma} = (\varsigma_1, \varsigma_2)' = (r \cos\phi, r \sin\phi)'$, $J(\boldsymbol{\varsigma} \to (r, \phi)) = r$. Note that $\tau_j > 0$, then $\varsigma_1 > \frac{l_{21}}{l_{22}} \varsigma_2, \varsigma_2 > 0$, i.e. $r > 0, 0 < \phi < \phi_0 < \pi$, where $\cot \phi_0 = \frac{l_{21}}{l_{22}}$. The density (4.85) is transformed to

$$\begin{aligned} & f_{r,\phi}((r, \phi) \mid \boldsymbol{L}) \\ = \; & \frac{2^2}{(l_{11} l_{22})^{k-2}} \left(\cos\phi \sin\phi - \frac{l_{21}}{l_{22}} \sin^2\phi \right)^{k-3} r^{2(k-3)} \exp\{-\frac{1}{2} r^2\} r. \end{aligned} \tag{4.86}$$

Then the normalizing constant

$$\begin{aligned} m &= \frac{2}{(l_{11}l_{22})^{k-2}} \int_0^{\phi_0} \left(\frac{\cos(2\phi - \phi_0) - \cos\phi_0}{2\sin\phi_0} \right)^{k-3} d\phi \int_0^{+\infty} (r^2)^{k-3} \exp\{-\frac{1}{2}r^2\} dr^2 \\ &= \frac{2^{k-2}\Gamma(k-2)}{(l_{11}l_{22})^{k-2}} \int_{-\phi_0}^{\phi_0} \left(\frac{\cos\psi - \cos\phi_0}{2\sin\phi_0} \right)^{k-3} d\psi \\ &= \frac{2^2(l_{21}^2 + l_{22}^2)^{\frac{k-3}{2}}\Gamma(k-2)}{l_{11}^{k-2}l_{22}^{2k-5}} \int_0^{\phi_0} (\cos\psi - \cos\phi_0)^{k-3} d\psi. \end{aligned} \tag{4.87}$$

Chapter 5

Future Study

The studies in this dissertation are on-going. The univariate Bayesian smoothing spline with dependency model discussed in Chapter 3 has great advantages over the X-13 method and the basic Bayesian smoothing spline model for the trend estimation. The accuracy, precision, and boundary performance are all elevated by borrowing information from different cycles. However, the multivariate BSSD model needs improvement.

First, if the number of knots or series is large, the number of parameters sampled from Algorithm 4 increases dramatically, which requires a significant burden of computation. Even though, theoretically, we may borrow more information from a larger number of knots or series, the multivariate ratio-of-uniform sampling makes the model much less efficient compared with the univariate BSSD model. Currently, we choose the MC sampling method instead of MCMC, due to the convergence problem and nonstandard densities.

Second, we would like to find some priors for the correlation parameter to simplify the posterior distribution and improve the computational efficiency. Also, with

another choice of re-parameterization and appropriate priors, efficient MC or MCMC sampling methods may be possible.

Third, the current dependence structure is designed for the time series with periodic variation. This can be generalized to other time dependent structures, which may have broader applications. For example, with more ancillary information of time dependency, more complex structures may be designed instead of using just one correlation parameter. Non-periodic variations may also be modeled.

Bibliography

Albert, Arthur (1972). *Regression and the Moore-Penrose Pseudoinverse*. ISSN. Elsevier Science. ISBN: 9780080956039. URL: https://books.google.com/books?id=-kOivHeTIWQC.

Cheng, Chin-I and Paul L Speckman (2012). "Bayesian smoothing spline analysis of variance". In: *Computational Statistics & Data Analysis* 56.12, pp. 3945–3958.

Clenshaw, Charles W and Alan R Curtis (1960). "A method for numerical integration on an automatic computer". In: *Numerische Mathematik* 2.1, pp. 197–205.

Dagum, Estela Bee (1980). *The X-11-ARIMA seasonal Adjustment Method, statistics Canada*.

Dagum, Estela Bee and Silvia Bianconcini (2016). "Seasonal Adjustment Based on ARIMA Model Decomposition: TRAMO-SEATS". In: *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. Springer, pp. 115–145.

De Klerk, Etienne (2002). *Aspects of semidefinite programming: interior point algorithms and selected applications*. Vol. 65. Springer.

Fessler, Jeffrey A (1991). "Nonparametric fixed-interval smoothing of nonlinear vector-valued measurements". In:

Findley, David F et al. (1998). "New capabilities and methods of the X-12-ARIMA seasonal-adjustment program". In: *Journal of Business & Economic Statistics* 16.2, pp. 127–152.

Gelman, Andrew et al. (2013). *Bayesian data analysis*. CRC press.

Gu, Chong (2013). *Smoothing spline ANOVA models*. Vol. 297. Springer Science & Business Media.

Harville, David A. (2008). *Matrix Algebra from a Statistician's Perspective*. Springer.

Hastie, Trevor J and Robert J Tibshirani (1990). *Generalized additive models*. Vol. 43. CRC press.

Horn, R.A. and C.R. Johnson (2012). *Matrix Analysis*. Cambridge University Press. ISBN: 9781139788885. URL: https://books.google.com/books?id=O7sgAwAAQBAJ.

Hwang, Suk-Geun (2004). "Cauchy's Interlace Theorem for Eigenvalues of Hermitian Matrices". In: *The American Mathematical Monthly* 111.2, pp. 157–159. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/4145217.

Labor Statistics, U.S. Bureau of and U.S. Census Bureau (2019). "Current Population Survey Technical Paper 77". In:

Ladiray, Dominique and Benoit Quenneville (2012). *Seasonal adjustment with the X-11 method*. Vol. 158. Springer Science & Business Media.

Marshall, Albert W, Ingram Olkin, and Barry Arnold (2011). "Inequalities: Theory of Majorization and Its Applications". In:

Monsell, B (2007). "The X-13A-S seasonal adjustment program". In: *Proceedings of the 2007 Federal Committee On Statistical Methodology Research Conference. URL http://www. fcsm. gov/07papers/Monsell. II-B. pdf.*

Narasimhan, Balasubramanian et al. (2020). "Package 'cubature'". In:

Pfeffermann, Danny (1991). "Estimation and seasonal adjustment of population means using data from repeated surveys". In: *Journal of Business & Economic Statistics* 9.2, pp. 163–175.

Sax, Christoph (2017). *seasonalview: Graphical User Interface for Seasonal Adjustment*. R package version 0.3. URL: `https://CRAN.R-project.org/package=seasonalview`.

Schott, James R. (1996). *Matrix analysis for statistics*. John Wiley & Sons.

Speckman, Paul L and Dongchu Sun (2003). "Fully Bayesian spline smoothing and intrinsic autoregressive priors". In: *Biometrika* 90.2, pp. 289–302.

Sun, Dongchu, Shawn Ni, and Paul L Speckman (2014). "Bayesian Analysis of Multivariate Smoothing Splines". Manuscript.

Sun, Dongchu, Robert K Tsutakawa, and Paul L Speckman (1999). "Posterior distribution of hierarchical models using CAR (1) distributions". In: *Biometrika* 86.2, pp. 341–350.

Tong, Xiaojun, Zhuoqiong Chong He, and Dongchu Sun (2018). "Estimating Chinese Treasury yield curves with Bayesian smoothing splines". In: *Econometrics and statistics* 8, pp. 94–124.

Wahba, Grace (1990). *Spline models for observational data*. Vol. 59. Siam.

Wakefield, JC, AE Gelfand, and AFM Smith (1991). "Efficient generation of random variates via the ratio-of-uniforms method". In: *Statistics and Computing* 1.2, pp. 129–133.

Wang, Yuedong, Wensheng Guo, and Morton B Brown (2000). "Spline smoothing for bivariate data with applications to association between hormones". In: *Statistica Sinica*, pp. 377–397.

White, Gentry (2006). "Bayesian semiparametric spatial and joint spatio-temporal modeling". PhD thesis. University of Missouri–Columbia.

Yue, Yu Ryan, Paul L Speckman, and Dongchu Sun (2012). "Priors for Bayesian adaptive spline smoothing". In: *Annals of the Institute of Statistical Mathematics* 64.3, pp. 577–613.

VITA

Benqian Zhang was born on January 20, 1989 in Jinan City, Shandong Province of China. He graduated with a Bachelor of Science in Information & Computing Science from the Department of Mathematics, Beijing Institute of Technology in 2013 with a second degree of economics. Then he received a Master in Statistics from University of Missouri-Columbia in 2015. He began his doctoral studies in August 2015.