

Bayesian causal inference with a single treated unit

by

Shuangshuang Liang

B.S., Dalian Medical University, China, 2015

M.S., Emporia State University, 2018

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Abstract

Causal inference with observational data has been of great interest in various areas of study. One of the important research questions in observational studies is how to estimate the treatment effect when the treatment is assigned to a single unit. To address this problem, the Synthetic Control Method (SCM) was proposed to construct a weighted combination of control units that approximates the treated unit. However, SCM has several limitations, including difficulties in conducting formal statistical inference. The Generalized Synthetic Control Method (GSCM) was proposed to address the issues of SCM. GSCM is based on a factor modeling framework that requires rank selection and creates some difficulties with making statistical inferences. We address these issues by developing a Bayesian GSCM. The proposed approach enables Bayesian inference through Markov chain Monte Carlo (MCMC). We further extend the model to allow for flexible, time-varying treatment effects using B-spline representations. To improve computational efficiency, we incorporate parameter expansion (PX) within the MCMC framework to enhance mixing and convergence. In addition, rank selection is performed using posterior model probabilities. To demonstrate the validity of the proposed method, we conduct simulation studies and compare the proposed method with existing SCM and GSCM. Finally, the proposed method is applied to two different real datasets to study treatment effects in distinct empirical settings.

This dissertation consists of five chapters. In Chapter 1, we discuss the background and motivation for our study. In Chapter 2, we develop a fully Bayesian approach for estimating the treatment effect when the treatment is assigned to a single unit. Chapter 3 extends the Bayesian framework to model time-varying treatment effects and introduces a parameter expansion technique to improve the efficiency of the Gibbs sampler. Chapter 4 introduces a Bayesian procedure for rank selection in the generalized synthetic control framework. Chapter 5 discusses some remarks and future directions.

Bayesian causal inference with a single treated unit

by

Shuangshuang Liang

B.S., Dalian Medical University, China, 2015

M.S., Emporia State University, 2018

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2026

Approved by:

Co-Major Professor
Gyuhyeong Goh

Approved by:

Co-Major Professor
Perla E. Reyes

Copyright

© Shuangshuang Liang.

Abstract

Causal inference with observational data has been of great interest in various areas of study. One of the important research questions in observational studies is how to estimate the treatment effect when the treatment is assigned to a single unit. To address this problem, the Synthetic Control Method (SCM) was proposed to construct a weighted combination of control units that approximates the treated unit. However, SCM has several limitations, including difficulties in conducting formal statistical inference. The Generalized Synthetic Control Method (GSCM) was proposed to address the issues of SCM. GSCM is based on a factor modeling framework that requires rank selection and creates some difficulties with making statistical inferences. We address these issues by developing a Bayesian GSCM. The proposed approach enables Bayesian inference through Markov chain Monte Carlo (MCMC). We further extend the model to allow for flexible, time-varying treatment effects using B-spline representations. To improve computational efficiency, we incorporate parameter expansion (PX) within the MCMC framework to enhance mixing and convergence. In addition, rank selection is performed using posterior model probabilities. To demonstrate the validity of the proposed method, we conduct simulation studies and compare the proposed method with existing SCM and GSCM. Finally, the proposed method is applied to two different real datasets to study treatment effects in distinct empirical settings.

This dissertation consists of five chapters. In Chapter 1, we discuss the background and motivation for our study. In Chapter 2, we develop a fully Bayesian approach for estimating the treatment effect when the treatment is assigned to a single unit. Chapter 3 extends the Bayesian framework to model time-varying treatment effects and introduces a parameter expansion technique to improve the efficiency of the Gibbs sampler. Chapter 4 introduces a Bayesian procedure for rank selection in the generalized synthetic control framework. Chapter 5 discusses some remarks and future directions.

Table of Contents

List of Figures	viii
List of Tables	x
Acknowledgements	xi
1 Introduction	1
1.1 The Synthetic Control Method and its challenges	2
1.2 The Generalized Synthetic Control Method and its challenges	4
1.3 Motivation and dissertation outline	6
2 Bayesian Generalized Synthetic Control Method with Single Treatment Unit	8
2.1 Model setup and prior specification	9
2.2 Posterior Inference	11
2.3 Simulation study	14
2.4 Empirical example	16
2.5 Discussion	21
3 Bayesian Generalized Synthetic Control Method with Treatment Modeling	23
3.1 Model setup and prior specification	24
3.2 Posterior inference	26
3.3 Simulation study	30
3.4 Empirical example	34
3.5 Discussion	40

4	Rank selection for the Bayesian Generalized Synthetic Control Method	42
4.1	Background and motivation	43
4.2	Model setup for rank selection	45
4.3	Bayesian rank selection method	46
4.4	Simulation study	51
4.5	Discussion	53
5	Conclusion	54
	Bibliography	56
A	Calculation of the Normal Linear Factor Model	62
B	The full conditional distributions for Chapter 2	64
C	The full conditional distributions for Chapter 3	67
C.1	Step 1. Parameter estimation	67
C.2	Step 3. Derivation of the posterior distribution for the treatment effect . . .	72
D	Derivations for parametric treatment effect specifications	74
D.1	Polynomial treatment effect specification	74
D.2	Gibbs sampling for the polynomial specification	75

List of Figures

2.1	Trends in per-capita cigarette sales: California vs. synthetic California by the SCM	18
2.2	Per-capita cigarette sales gap between California and synthetic California by the SCM.	19
2.3	Trends in per-capita cigarette sales: California vs. synthetic California by the GSCM.	19
2.4	Per-capita cigarette sales gap between California and synthetic California by the GSCM.	20
2.5	Trends in per-capita cigarette sales: California vs. synthetic California by the B-GSCM.	20
2.6	Per-capita cigarette sales gap between California and synthetic California by the B-GSCM.	21
3.1	True treatment effect, Bayesian B-spline estimate, and 95% credible interval from the simulated example.	32
3.2	Path Plot of Per Capita CO ₂ Emissions from Transport during 1960–2005: Sweden versus Synthetic Sweden by the SCM	36
3.3	Gap in per capita CO ₂ emissions from transport between Sweden and synthetic Sweden by the SCM	37
3.4	Path Plot of Per Capita CO ₂ Emissions from Transport during 1960–2005: Sweden versus Synthetic Sweden by the GSCM	37
3.5	Gap in per capita CO ₂ emissions from transport between Sweden and Synthetic Sweden by the GSCM	38

3.6	Gap in per capita CO ₂ emissions from transport between Sweden and Synthetic Sweden by the B-spline GSCM	38
3.7	Path Plot of Per Capita CO ₂ Emissions from Transport during 1960–2005: Sweden versus Synthetic Sweden by the B-spline GSCM	39
4.1	Bar plots of the selected factor rank for the Bayesian GSCM (left) and GSCM (right) for the 100 simulated datasets described in Section 2.3.	51
4.2	Bar plots of the rank for the Bayesian GSCM and GSCM for the 100 simulated datasets described in Section 3.3.	52

List of Tables

2.1	Summary of simulation study over 100 runs.	15
3.1	Summary of simulation study over 100 runs with a lower-noise setting. . . .	31
3.2	Summary of simulation study over 100 runs with a higher-noise setting. . . .	33
3.3	Comparison of Estimated Treatment Effects: SCM, GSCM, and BGSCM. Cell values represent the estimated treatment effect (Sweden - Synthetic Sweden)	39
3.4	95% CI of Treatment Effects (Sweden - Synthetic Sweden) and Interval Length	40

Acknowledgments

Looking back on my doctoral journey, I realize that it was filled with many challenges that I could not have overcome on my own. I am deeply grateful to the many people who supported and guided me along the way. I would like to express my sincere appreciation to all of them.

First, I would like to sincerely thank my co-major advisor, Dr. Gyuhyeong Goh, for his guidance, patience, and support throughout my doctoral studies. His knowledge and expertise in statistics helped me greatly in completing my research. I am very grateful for his encouragement, helpful suggestions, and the many discussions we had during this process. I learned a lot from his dedication to academic work and his passion for research. His mentorship was very important in helping me finish this dissertation.

I would also like to express my sincere gratitude to Dr. Perla Reyes for serving as my co-major advisor, for her valuable advice, and for her help with important paperwork during this process. In addition, I would like to thank Dr. Mike Higgins, Dr. Jisang Yu, and Dr. Shengwu Shang for serving on my dissertation committee and for providing helpful comments and constructive feedback on my research.

I would like to thank the Department of Statistics at Kansas State University for providing me with a graduate assistantship, which made it possible for me to pursue and complete my graduate studies. I am grateful to the professors and staff members in the department, both past and present, for their support. In particular, I would like to thank Dr. Abigail Jager, Dr. Christopher Vahl, Dr. Mike Higgins, and Dr. Perla Reyes, whom I worked with while serving as a graduate teaching assistant. I also appreciate Jo Blackburn and Alisha Dean for their help with administrative matters. In addition, I am thankful to the professors from whom I learned through courses and lectures, including Dr. Christopher Vahl, Dr. James Neill, Dr. Juan Du, Dr. Weixing Song, Dr. Haiyan Wang, Dr. Gyuhyeong Goh, Dr. Jisang Yu, and Dr. Mike Higgins.

I would like to express my deepest gratitude to my family for their unconditional love, patience, and continuous support throughout my academic journey. Their encouragement has been a constant source of strength for me.

Last but not least, I am especially grateful to my husband, Changkun, for his understanding, support, and encouragement during this challenging process. His patience and belief in me gave me the motivation to persevere and complete this work.

Chapter 1

Introduction

Causal inference is a fundamental approach for understanding relationships between causes and their effects. A key challenge is that we only observe one potential outcome for each unit, while the counterfactual outcome is unobserved, and estimating it is the fundamental problem. A key assumption in many causal methods is the parallel trends assumption, which means that, in the absence of treatment, the treated and control groups would have followed similar outcome trends before the intervention. Difference-in-differences is often used for causal inference and requires the parallel trends assumption. In practice, it can be difficult to establish the parallel trends assumption, especially in the real world. The synthetic control method ([Abadie and Gardeazabal, 2003](#)) ([Abadie et al., 2010](#)) and the generalized synthetic control method ([Xu, 2017](#)) do not require this often-violated assumption.

The Synthetic Control Method (SCM) constructs a weighted combination of groups as controls, to compare the treatment group. The weights are determined using a statistical optimization algorithm, which aims to find weights that make the synthetic control group as similar as possible to the treatment unit prior to the intervention. The Synthetic Control Method is a powerful and flexible tool for causal inference. Earlier SCM faced limitations in formal statistical inference, sometimes depending on placebo or permutation tests to assess uncertainty. Recent methodological developments have introduced more formal inferential frameworks, including approaches based on conformal inference and prediction intervals.

Despite these advances, SCM still faces several challenges, such as sensitivity to donor pool selection, the choice of predictors and optimization procedures, and its traditional focus on settings with a single treated unit.

The generalized synthetic control method (GSCM) extends the synthetic control framework and addresses several of its limitations. By incorporating an interactive fixed effects structure, GSCM can accommodate multiple treated units and allow treatment effects to vary over time. These extensions broaden the applicability of synthetic control methods in empirical research. However, despite these improvements, several methodological challenges remain. In particular, determining the appropriate factor rank is still an important issue. In practice, the rank is typically selected using cross-validation, but this approach may produce unstable results across samples, which can affect the reliability and robustness of the estimated treatment effects.

1.1 The Synthetic Control Method and its challenges

The Synthetic Control Method (SCM), developed by [Abadie and Gardeazabal \(2003\)](#), provides an effective approach for estimating effects when there is only one treated unit. SCM constructs a synthetic control unit as a weighted combination of untreated units. SCM does not rely on strong assumptions about the Data Generate Process (DGP) of the outcomes, making it very flexible for causal inference in observational studies.

In the field of causal inference, the Synthetic Control Method has attracted the attention and admiration of an increasing number of researchers from a variety of disciplines. This introductory section draws on the research of many experts ([Abadie et al., 2015](#); [Ben-Michael et al., 2021](#); [Kreif et al., 2016](#); [McClelland and Gault, 2017](#)) to explain why the SCM is important and describe its scope and requirements.

The Synthetic Control Method has become one of the most widely used approaches in causal inference because it solves the issue of determining the effect of an approach on a single unit, such as a region, nation, or entity. Traditional statistical methods frequently fail when there are no control groups with comparable characteristics prior to treatment

([Abadie and Gardeazabal, 2003](#)). The SCM provides a principled framework for estimating treatment effects under such challenging conditions.

The Synthetic Control Method builds a synthetic unit to represent the counterfactual outcome without treatment. The control unit is thoroughly constructed by combining untreated units in a weighted sum until their pre-treatment characteristics match those of the treated unit. By comparing the treated unit with the synthetic control, researchers can estimate the causal effect.

[Abadie et al. \(2010\)](#) illustrate that the SCM is appealing because it can be adapted to a wide range of research questions and is data-driven, allowing researchers to more easily estimate causal effects from observational data. Since it does not rely on specific parameter values, it can be used to solve problems from a variety of disciplines. When randomized experiments cannot be conducted for practical or ethical reasons, the SCM becomes a crucial tool in a number of fields of study, including policy evaluation, economics, epidemiology, and the social sciences ([Amjad et al., 2018](#); [Bonander et al., 2021](#); [Freire, 2018](#); [Goh and Yu, 2022](#)).

The SCM offers a promising path to causal inference, but it is not lacking of challenges. [Abadie and Gardeazabal \(2003\)](#) cite the absence of a sufficient collection of control units that replicate the pre-treatment trajectory of the treated unit as a significant challenge. This can be challenging, particularly when dealing with unconventional or rare approaches. In addition, although SCM provides point estimates of treatment effects, conducting formal statistical inference can be challenging. Recent methodological developments, such as prediction interval approaches implemented in R’s `scpi` package ([Cattaneo et al., 2025](#)), have begun to address this limitation by providing more formal measures of uncertainty.

However, SCM has several limitations, including challenges in formal statistical inference and in capturing time-varying treatment effects. Despite recent advances, reliable inference remains difficult in practice. Researchers have begun to place greater emphasis on probabilistic methods that provide not only point estimates but also quantifications of the uncertainty in causal impact estimations. The Bayesian Synthetic Control Method (BSCM) provides a promising framework for overcoming these challenges. Researchers explore the motivation,

development, and application of the Bayesian SCM (Kim et al., 2020; Kurz et al., 2019; Pang et al., 2022; Pinkney, 2021) aiming to advance the field of causal inference with observational data.

Recent developments in SCM have increasingly incorporated Bayesian frameworks to improve flexibility and uncertainty quantification. Chandra et al. (2023) propose a Bayesian nonparametric framework that generates synthetic control groups by matching treated units with similar subpopulations in external datasets. Pang et al. (2022) develop a Bayesian alternative to the synthetic control method based on a dynamic latent factor model, allowing for heterogeneous and time-varying treatment effects. Ma et al. (2026) introduce a Bayesian synthetic control approach with horseshoe priors to handle sparsity and multiple treated units. Klinenberg (2023) proposes a synthetic control method with time-varying coefficients using a state-space model with Bayesian shrinkage. Xu and Zhou (2025) relaxes the simplex constraint and incorporates spike-and-slab priors for variable selection.

Despite these developments, existing approaches primarily focus on either flexible weighting schemes or dynamic outcome modeling, with limited attention to integrating flexible treatment effect specifications within a Bayesian factor model framework. This dissertation addresses this gap by combining Bayesian generalized synthetic control methods with flexible modeling of dynamic treatment effects.

1.2 The Generalized Synthetic Control Method and its challenges

While the synthetic control method (SCM) is a useful tool for estimating causal effects in settings with a single treated unit, it has several limitations that restrict its use in more complex situations. In particular, SCM relies on fixed weights to construct the counterfactual and does not explicitly account for unobserved heterogeneity. It may perform poorly when the relationship between treated and control units evolves over time.

To address these limitations, Xu (2017) proposed the Generalized Synthetic Control

Method (GSCM), which extends the SCM framework by incorporating an interactive fixed effects factor model structure. Under this framework, the outcome is modeled as

$$y_{it} = \tau_{it}D_{it} + \mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i\mathbf{f}_t + \epsilon_{it},$$

where the latent factor structure captures unobserved heterogeneity across units and over time. The factor structure flexibly accounts for unobserved heterogeneity, helping the model more closely reflect the true data-generating process.

The GSCM builds on earlier work on interactive fixed effects models (Bai, 2009) and relates to approaches that combine synthetic control ideas with panel data methods (Gobillon and Magnac, 2016). It also relates to matrix completion approaches for causal inference, which leverage low-rank structures in panel data (Athey et al., 2021).

By introducing a model-based structure, GSCM improves estimation accuracy and allows for time-varying treatment effects. It works well for high-dimensional panel data. The method extracts the essential structure more efficiently. This helps improve the accuracy of treatment effect estimation. Given these advantages, GSCM has been widely applied in empirical research across a variety of fields. For example, Sheridan et al. (2022) uses GSCM to estimate the impact of extreme weather events on population health, demonstrating its usefulness in environmental and epidemiological settings. Wang et al. (2024) study the impact of informal payments on doctor engagement in an online health community, finding evidence of a crowding-out effect of monetary incentives. Letellier et al. (2025) extend the generalized synthetic control method to a two-stage framework to estimate heterogeneous health effects of extreme weather events, demonstrating its flexibility in environmental applications. Some other applications of the GSCM (Goin et al., 2025; Mader and Rüttenauer, 2022; Sheridan et al., 2022; Wen et al., 2023).

The GSCM also introduces new challenges. A key issue is the selection of the number of latent factors, which can substantially influence estimation results. In practice, the factor rank is typically chosen using cross-validation, this approach may produce unstable results across samples and does not fully account for model uncertainty. Furthermore, statistical inference remains challenging, as uncertainty arising from factor estimation and model selection is difficult to quantify within the standard framework.

1.3 Motivation and dissertation outline

The generalized synthetic control approach offers a framework that is flexible for predicting counterfactual outcomes through latent factor models. Its conventional applications mainly depend on point estimation and large-sample approximations for inference. In empirical applications, however, the number of treated units is generally small, and uncertainty in the calculation of latent components and factor loadings can be significant. A Bayesian approach to GSCM addresses these limitations by treating all unknown quantities including latent factors, factor loadings, and counterfactual outcomes as random variables and integrating over their posterior distributions. This method creates a consistent framework for uncertainty quantification and enables for valid finite sample inference of treatment effects.

[Xu \(2017\)](#) proposes selecting the number of latent factors in the generalized synthetic control method using cross-validation. From a Bayesian approach, selecting the number of factors can be viewed as a model comparison problem, where each potential rank corresponds to a separate latent factor model. Rather than relying on cross-validation, the Bayesian framework gives a formal mechanism for comparing models through their marginal likelihoods. In this framework, uncertainty about the factor dimension is accounted for directly within the inferential procedure, rather than being resolved through external validation methods. We develop a Bayesian rank selection procedure in this dissertation, which is based on the maximization of the posterior probability of the rank r using a Laplace approximation of the marginal likelihood.

In numerous statistical problems, it is challenging to sample directly from the joint distribution of the parameters. This problem is neatly solved by the Gibbs sampler ([Casella and George, 1992](#); [Smith and Roberts, 1993](#)). This dissertation develops Bayesian extensions of the generalized synthetic control method for causal inference in panel data. It addresses two central challenges: recovering unobserved potential outcomes for treated units and modeling treatment effects in the post-treatment period.

Chapter [2](#) focuses on completing the missing potential outcomes using a Bayesian factor model. We develop a Gibbs sampling based Bayesian GSCM framework to impute counter-

factual outcomes for treated units during post-treatment period. This chapter emphasizes model estimation, uncertainty quantification, and the recovery of unobserved components of the outcome matrix.

Chapter 3 parameterizes the Average Treatment Effect on the Treated (ATT) using a flexible B-spline basis function. This technique allows the model to capture nonlinear, time-varying treatment effects, which is important for studying long-term interventions. A significant challenge in Bayesian latent factor models is the slow convergence of Markov Chain Monte Carlo (MCMC) samplers, often caused by high correlation between factors and their corresponding loadings. To address this issue, Chapter 3 incorporates the Parameter Expansion (PX) algorithm into the Gibbs sampling framework.

Chapter 4 is about the marginal likelihood-based method for selecting the number of latent factors. Rather than relying on cross-validation, we use a Laplace approximation to construct a Bayesian criterion that determines the factor dimension from the pre-treatment data. Two rank selection simulations based on settings of Chapters 2 and 3 are conducted.

We close with a final Chapter 5 that summarizes our conclusions.

Chapter 2

Bayesian Generalized Synthetic Control Method with Single Treatment Unit

Causal inference methods are essential for identifying and understanding causal relationships between variables in various fields, such as epidemiology, economics, social sciences, and public policy. There are some common and popular methods in causal inference including difference-in-difference from [Card and Krueger \(1993\)](#) or propensity score matching by [Rosenbaum and Rubin \(1983\)](#). Both approaches suffer from selection bias and restrictive assumptions.

The Synthetic Control Method, introduced by [Abadie and Gardeazabal \(2003\)](#) and [Abadie et al. \(2010\)](#), offers an alternative by constructing a weighted combination of control units that closely matches the treated unit’s pre-treatment characteristics and outcome path. By emphasizing pre-treatment fit, SCM provides a transparent and data-driven approach to counterfactual construction and reduces reliance on extrapolation.

The Generalized Synthetic Control Method is widely used by researchers to assess the causal effects of treatments, interventions, or policy changes in observational studies. The GSCM extends the functionality of the SCM to handle more intricate scenarios. Examples

of the usage of GSCM in the literature include [Huang et al. \(2022\)](#) who studied air pollution control measures of G20 Hangzhou summit and [Silva et al. \(2021\)](#) study of the policies to combat poverty in Brazil. While the GSCM has been a valuable tool, it represents only one perspective of the multifaceted inference of causality. To fill this gap and provide an alternative perspective, we proposed a new method in a Bayesian framework.

The structure of this chapter is as follows. In Section [2.1](#), we specify the model settings and the priors for unknown parameters. In Section [2.2](#), we introduce a fully Bayesian approach to the GSCM including details of posterior computation. Simulation studies are presented in Section [2.3](#). The proposed method is applied to real data in Section [2.4](#). Finally, concluding remarks are given in Section [2.5](#).

2.1 Model setup and prior specification

Suppose y_{it} is the outcome for unit i at time t , for units $i = 1, \dots, N$ and time periods $t = 1, \dots, T$. Let $\mathbf{T}_{0,i}$ represent the number of pretreatment periods for unit i . To simplify the notation, we assume that all treated units begin receiving the treatment at a common time point, denoted by $\mathbf{T}_0$, so that $\mathbf{T}_{0,i} = \mathbf{T}_0$ for all treated units. Let $\mathcal{T}$ and $\mathcal{C}$ denote the sets of units in treatment and control groups, respectively. Let N_{tr} and N_{co} denote the numbers of treated and control units, respectively. While this framework accommodates multiple treatment units, the focus of this dissertation is on a more specific case: the single treatment unit scenario, in which only one unit is exposed to the treatment. Assume that y_{it} follows a linear factor model:

$$y_{it} = \tau_{it}D_{it} + \mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i\mathbf{f}_t + \epsilon_{it},$$

where τ_{it} is the treatment effect on unit i at time t , $D_{it} = 1$ when $i \in \mathcal{T}$ and $t > \mathbf{T}_0$, and $D_{it} = 0$ otherwise. $\mathbf{x}_{it}$ is a $(k \times 1)$ vector of observed covariates, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of unknown parameters, $\mathbf{f}_t$ is an $(r \times 1)$ vector of unobserved common factors, $\boldsymbol{\lambda}_i$ is an $(r \times 1)$ vector of unknown factor loadings, and we assume the error terms $\epsilon_{it} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$.

For unit i , the data generating process when no treatment is assigned can be written as:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\lambda}_i + \boldsymbol{\epsilon}_i, \quad (2.1)$$

where $\mathbf{Y}_i = (y_{i1}, \dots, y_{iT})'$ is a $(T \times 1)$ vector, $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})'$ is a $(T \times k)$ matrix, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector, $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)'$ is a $(T \times r)$ matrix, $\boldsymbol{\lambda}_i$ is an $(r \times 1)$ vector, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{iT})'$ is a $(T \times 1)$ vector with $\boldsymbol{\epsilon}_i \sim \mathcal{N}_T(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_T$.

For time t , we can rewrite the model as:

$$\tilde{\mathbf{Y}}_t = \tilde{\mathbf{X}}_t\boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{f}_t + \tilde{\boldsymbol{\epsilon}}_t, \quad (2.2)$$

where $\tilde{\mathbf{Y}}_t = (y_{1t}, \dots, y_{Nt})'$ is a $(N \times 1)$ vector, $\tilde{\mathbf{X}}_t = (\mathbf{x}_{1t}, \dots, \mathbf{x}_{Nt})'$ is a $(N \times k)$ matrix, $\boldsymbol{\beta}$ is a $(k \times 1)$ vector, $\mathbf{f}_t$ is an $(r \times 1)$ vector, $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_N)'$ is a $(N \times r)$ matrix, and $\tilde{\boldsymbol{\epsilon}}_t = (\epsilon_{1t}, \dots, \epsilon_{Nt})'$ is a $(N \times 1)$ vector with $\tilde{\boldsymbol{\epsilon}}_t \sim \mathcal{N}_N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$.

Likelihood identifiability is an important property of statistical models because it ensures that the parameters of the model can be estimated accurately and uniquely from the observed data. We assume $\mathbf{F}$ has the following form to guarantee likelihood identification ([Geweke, 1996](#)):

$$\mathbf{F} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{F}^* \end{bmatrix},$$

where $\mathbf{F}^*$ is a $(T - r) \times r$ full-rank matrix.

The choice of priors is an important aspect of Bayesian factor analysis model and should be carefully considered, inverse gamma priors are often used for modeling the error variances, and normal or truncated normal priors are commonly used for modeling the factor loadings. We also use normal priors for observed covariates and the corresponding parameters.

To complete the model specification, we assign prior distributions to the model parameters $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$, $\mathbf{F}$, and $\boldsymbol{\Sigma}$.

We first specify a multivariate Gaussian prior for the regression coefficients:

$$\boldsymbol{\beta} \sim \mathcal{N}_k(\mathbf{0}, v_0 \mathbf{I}_k),$$

where v_0 is a pre-specified hyperparameter.

For the factor loadings, we assume independent Gaussian priors:

$$\boldsymbol{\lambda}_i \stackrel{\text{iid}}{\sim} \mathcal{N}_r(\mathbf{0}, \xi_0 \mathbf{I}_r), \quad i = 1, \dots, N,$$

where ξ_0 is a hyperparameter.

Similarly, we assign a multivariate Gaussian prior to the latent factors:

$$\mathbf{f}_t \stackrel{\text{iid}}{\sim} \mathcal{N}_r(\mathbf{0}, C_0 \mathbf{I}_r), \quad t = 1, \dots, T,$$

where C_0 is a hyperparameter.

For the idiosyncratic variance, we assume an inverse-gamma prior:

$$\sigma^2 \sim \text{IG}\left(\frac{a_0}{2}, \frac{b_0}{2}\right),$$

where a_0 and b_0 are hyperparameters.

2.2 Posterior Inference

Let $Y_{it}(1)$ and $Y_{it}(0)$ denote the potential outcomes under treatment and control, respectively.

The model is specified as

$$Y_{it}(0) = \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\lambda}_i'\mathbf{f}_t + \epsilon_{it},$$

and

$$Y_{it}(1) = \tau_{it} + \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\lambda}_i'\mathbf{f}_t + \epsilon_{it}.$$

The individual treatment effect for treated unit i at time t is

$$\tau_{it} = Y_{it}(1) - Y_{it}(0), \quad i \in \mathcal{T}, t > T_0.$$

For treated units in the post-treatment period, $Y_{it}(1)$ is observed, while $Y_{it}(0)$ is not directly observed. Under the factor model, the counterfactual outcome is determined by the model parameters, allowing inference to proceed through the posterior distribution of the parameters.

Specifically, Bayesian inference is based on the posterior

$$\pi(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \mathbf{F}, \boldsymbol{\Sigma} \mid \mathbf{Y}),$$

where $\mathbf{Y}$ denotes the observed outcome matrix.

To sample from this posterior distribution, we employ a Gibbs sampler that iteratively draws from the full conditional distributions:

1. $\boldsymbol{\beta} \sim \pi(\boldsymbol{\beta} \mid \mathbf{Y}, \mathbf{F}, \boldsymbol{\Sigma}),$
2. $\boldsymbol{\Lambda} \sim \pi(\boldsymbol{\Lambda} \mid \mathbf{Y}, \mathbf{F}, \boldsymbol{\Sigma}, \boldsymbol{\beta}),$
3. $\mathbf{F} \sim \pi(\mathbf{F} \mid \mathbf{Y}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}, \boldsymbol{\beta}),$
4. $\boldsymbol{\Sigma} \sim \pi(\boldsymbol{\Sigma} \mid \mathbf{Y}, \boldsymbol{\Lambda}, \mathbf{F}, \boldsymbol{\beta}).$

Given posterior draws of the model parameters, the counterfactual outcomes for treated units in the post-treatment period are obtained as

$$Y_{it}(0) \sim \mathcal{N}(\mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i\mathbf{f}_t, \sigma^2), \quad i \in \mathcal{T}, t > T_0.$$

The full conditional posterior distributions are derived using standard conjugate Bayesian updating. First, it can be shown that the conditional distribution of y_i given $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ is multivariate normal,

$$y_i \mid \boldsymbol{\beta}, \boldsymbol{\Omega} \sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega}), \tag{2.3}$$

where $\mathbf{\Omega} = \xi_0 \mathbf{F} \mathbf{F}' + \mathbf{\Sigma}$. The proof is provided in Appendix A.

Based on this result, the full conditional posterior distributions can be obtained in closed form. The detailed process of the full conditional posterior distribution is provided in Appendix B.

The prior $\boldsymbol{\beta}$ and the likelihood function are derived from equation (2.1). By Bayes' theorem, the conditional posterior distribution of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta} \mid \mathbf{Y}, \mathbf{\Omega} \sim \mathcal{N}(\mathbf{B}_\beta \mathbf{b}_\beta, \mathbf{B}_\beta),$$

where

$$\mathbf{B}_\beta = \left(\sum_{i=1}^N \mathbf{X}_i' \mathbf{\Omega}^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right)^{-1}, \quad \mathbf{b}_\beta = \sum_{i=1}^N \mathbf{X}_i' \mathbf{\Omega}^{-1} \mathbf{Y}_i.$$

The conditional posterior distribution of the factor loadings is

$$\boldsymbol{\lambda}_i \mid \mathbf{F}, \boldsymbol{\beta}, \mathbf{\Sigma}, \mathbf{y} \sim \mathcal{N}(\mathbf{B}_\lambda \mathbf{b}_{\lambda,i}, \mathbf{B}_\lambda),$$

where

$$\mathbf{B}_\lambda = \left(\mathbf{F}' \mathbf{\Sigma}^{-1} \mathbf{F} + \frac{1}{\xi_0} \mathbf{I}_r \right)^{-1}, \quad \mathbf{b}_{\lambda,i} = \mathbf{F}' \mathbf{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}).$$

The prior $\mathbf{f}_t$ and the likelihood function are derived from equation (2.2). By Bayes' theorem, the conditional posterior distribution of $\mathbf{f}_t$ is given by

$$\mathbf{f}_t \mid \mathbf{\Lambda}, \boldsymbol{\beta}, \mathbf{\Sigma}, \mathbf{y} \sim \mathcal{N}(\mathbf{B}_f \mathbf{b}_{f,t}, \mathbf{B}_f),$$

where

$$\mathbf{B}_f = \left(\sigma^{-2} \mathbf{\Lambda}' \mathbf{\Lambda} + C_0^{-1} \mathbf{I}_r \right)^{-1}, \quad \mathbf{b}_{f,t} = \sigma^{-2} \mathbf{\Lambda}' \left(\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} \right).$$

The conditional posterior distribution of the variance parameter is

$$\sigma^2 \mid \mathbf{\Lambda}, \boldsymbol{\beta}, \mathbf{F}, \mathbf{y} \sim \text{IG} \left(\frac{a_0 + TN}{2}, \frac{S + b_0}{2} \right),$$

where

$$S = \sum_{t=1}^T \left\| \tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\Lambda} \mathbf{f}_t \right\|^2.$$

These conditional distributions admit closed-form expressions, enabling efficient posterior inference via a Gibbs sampling algorithm.

2.3 Simulation study

In this section, we conduct a Monte Carlo simulation study to examine the performance of the proposed generalized synthetic control method. Given $N=40$, $T=100$, $T_0=60$, and $p=4$, we simulate the data $\{(Y_{it}, \mathbf{Z}_{it}) : i = 1, \dots, N, t = 1, \dots, T\}$ from the following linear factor model:

$$Y_{it} = \alpha_i + \gamma_t + \mathbf{c}' \mathbf{Z}_{it} + (0.5 + \sqrt{t/2}) D_{it} + \epsilon_{it},$$

where $D_{it}=1(i=1, t > T_0)$, $\alpha_i \stackrel{\text{iid}}{\sim} N(0, \sigma_\alpha^2)$ with $\sigma_\alpha^2 = 0.5$, $\gamma_t \stackrel{\text{iid}}{\sim} N(\gamma_{t-1}, \sigma_\gamma^2)$ with $\sigma_\gamma^2 = 0.1$ and $\gamma_0 = 0$, $\mathbf{c} = (1, 1, 1, 1)^T$, $\mathbf{Z}$ is a three-dimensional $(N \times p \times T)$ matrix with $\mathbf{Z}_{it} \sim N_p(\mathbf{1}_p, 2\mathbf{I}_p)$, and $\epsilon_{it} \stackrel{\text{iid}}{\sim} N(0, 1)$.

We simulate the data $\{(\mathbf{Y}_i, \mathbf{X}_i) : i = 1, \dots, N\}$ from the data $\{(Y_{it}, \mathbf{Z}_{it}) : i = 1, \dots, N, t = 1, \dots, T\}$. $\mathbf{X}$ is a three-dimensional $(T \times p \times N)$ matrix, $\mathbf{X}_i$ is a $(T \times p)$ matrix with $\mathbf{X}_i = \mathbf{Z}'_i$. $\mathbf{Y}_i = (Y'_{i1}, \dots, Y'_{iT})$ is a $(T \times N)$ matrix. We use hyperparameters $v_0 = 100$, $\xi_0 = 1$, $C_0 = 1$, $a_\sigma = 1$, and $b_\sigma = 1$.

In the data generating process, the treatment effect of unit $i=1$ at time $t > T_0$ is given by:

$$\tau_{1t} = 0.5 + \sqrt{t/2}.$$

For posterior computation, we run the Gibbs sampler for 10,000 iterations, with the first 2,000 iterations discarded as burn-in. The remaining samples are used for posterior inference.

To evaluate the performance of estimation among SCM, GSCM, and Bayesian GSCM,

Table 2.1: *Summary of simulation study over 100 runs.*

Method	MSE	S.E.	Avg. 95% CI Width
SCM	10.05	0.21	10.36
GSCM with CV rank	1.10	0.02	4.25
GSCM with true rank 2	1.03	0.02	4.01
Bayesian GSCM	0.94	0.05	3.37

we consider the mean squared error (MSE) and the standard error (S.E.):

$$\begin{aligned} \text{MSE} &= \frac{1}{T - T_0} \sum_{t=T_0+1}^T (\hat{\tau}_{1t} - \tau_{1t})^2, \\ \text{S.E.} &= \frac{\sqrt{\frac{1}{N-1} \sum_{s=1}^N \left(\text{MSE}^{(s)} - \overline{\text{MSE}} \right)^2}}{\sqrt{N}}, \end{aligned}$$

where $\hat{\tau}_{1t}$ denotes the estimated treatment effect and τ_{1t} is the true treatment effect. The mean squared error (MSE) quantifies the average squared deviation between the estimated and true treatment effects over the post-treatment period. In addition, the standard error (S.E.) is calculated as the standard deviation of the MSE across simulation runs divided by the square root of the number of simulations. The average 95% confidence interval width is defined as the mean length of the 95% intervals for the estimated treatment effects, computed over the post-treatment period and averaged across simulation replications.

Table 2.1 summarizes the results of simulations that are conducted to compare the performance of different methods in estimating the average treatment effects over 100 runs. The proposed method, Bayesian GSCM, has the smallest MSE demonstrating superior performance in terms of MSE compared to both SCM and GSCM, highlighting its potential for enhanced prediction accuracy. However, it is crucial to take into account the trade-off associated with the standard error, since it offers valuable information regarding the accuracy of the estimations. In addition, the Bayesian GSCM produces the narrowest intervals, indicating more precise estimation compared to SCM and GSCM.

Using the true factor rank leads to improved performance compared to the rank selection with cross-validation, indicating the importance of accurate rank specification. This moti-

vates the development of a Bayesian rank selection approach, which is presented in Chapter 4.

2.4 Empirical example

For empirical application, we used the California tobacco control program dataset originally analyzed [Abadie et al. \(2010\)](#) in their study of the effects of California’s tobacco control policy. The dataset contains annual state-level observations on per capita cigarette sales for the treated unit, California, and a set of potential control states in the United States. The sample period is from 1970 to 2000. The policy intervention of interest is the implementation of California Proposition 99, a voter-approved tobacco control program passed in 1988. This policy substantially increased cigarette taxes and provided funding for anti-smoking education and media campaigns aimed at reducing tobacco consumption. In this study, California is considered the treated unit, while the remaining U.S. states serve as potential donor units used to construct the synthetic control. The outcome variable is annual per capita cigarette sales, measured in packs per person. The pre-treatment period spans 1970–1988 and is used to construct the synthetic control unit, while the post-treatment period covers 1989–2000 and is used to estimate the effect of the tobacco control program. In this real data application, the Gibbs sampler is run for 15,000 iterations, with the first 30% of iterations discarded as burn-in.

Figure 2.1 presents the path plot obtained from the classical synthetic control method. The solid black line represents the observed per-capita cigarette sales in California, while the dashed blue line represents the synthetic California constructed from a weighted combination of control states. As shown in the figure, the synthetic control closely tracks the observed California trajectory during the pre-treatment period, indicating that the model provides a reasonable counterfactual approximation for California prior to the implementation of Proposition 99. A visible divergence between California and the synthetic California after the passage of Proposition 99. Specifically, the observed cigarette sales in California decline more rapidly than those predicted by the synthetic control. This divergence suggests

that Proposition 99 had a substantial negative effect on cigarette consumption in California relative to the counterfactual trajectory implied by the donor pool.

The estimated treatment effect is illustrated more directly in the gap plot Fig. 2.2. The gap is defined as the difference between the observed outcome and the synthetic control. During the pre-treatment period, the gap fluctuates around zero, indicating a good pre-treatment fit. Following the policy implementation, the gap becomes negative and grows in magnitude throughout the 1990s, implying that cigarette sales in California were significantly lower than what would have occurred in the absence of the policy. The shaded region represents the post-treatment prediction interval obtained using the `scpi` package. These intervals are constructed using the inferential framework for synthetic control methods developed by Cattaneo et al. (2021), which quantifies the uncertainty in the counterfactual prediction.

Figure 2.3 displays the results obtained using the generalized synthetic control method. Similar to the SCM results, the pre-treatment trajectories of California and its estimated counterfactual align closely, demonstrating that the factor model underlying GSCM captures the underlying structure of the panel data effectively. The observed path diverges noticeably from the estimated counterfactual after the treatment year. The gap between the two series grows steadily over time, suggesting a sustained reduction in cigarette sales following Proposition 99.

The gap plot Fig. 2.4 provides a direct visualization of the treatment effects. Prior to the treatment year, the estimated effects remain close to zero, confirming the adequacy of the pre-treatment fit. After the treatment year, the treatment effect becomes increasingly negative. The shaded region represents the 95% confidence interval obtained from the bootstrap procedure implemented in the generalized synthetic control framework. This interval quantifies the statistical uncertainty associated with the estimated treatment effects.

Figures 2.5 and 2.6 present the path and gap plots obtained from the proposed Bayesian generalized synthetic control method. The path plot displays the observed California series together with the posterior mean of the counterfactual trajectory and its associated 95% credible band. The close alignment of the observed and counterfactual paths in the pre-

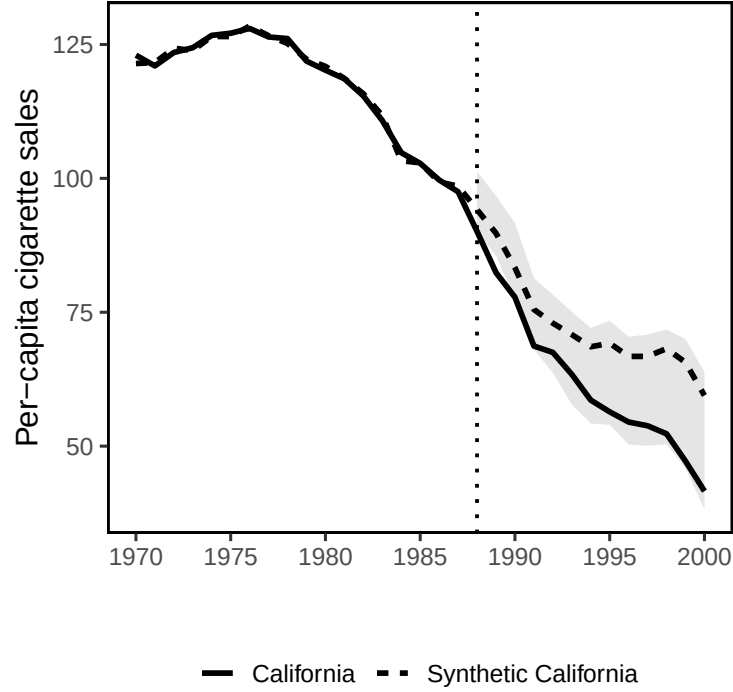


Figure 2.1: Trends in per-capita cigarette sales: California vs. synthetic California by the SCM. The shaded area presents 95% prediction interval for the estimated counterfactual outcome in the post-treatment period.

treatment period indicates that the Bayesian model provides a close fit before the policy intervention.

After the implementation of Proposition 99, the observed trajectory falls well below the estimated counterfactual path, suggesting a substantial reduction in cigarette consumption attributable to the policy. The shaded region around the counterfactual trajectory represents the 95% posterior credible interval, which reflects the uncertainty in the counterfactual predictions derived from the posterior distribution of the model parameters.

The gap plot Fig. 2.6 presents the posterior mean treatment effects along with their corresponding 95% credible intervals. Before the treatment year, the estimated effects fluctuate around zero, indicating that the model fits the pre-treatment data well. After 1989, the treatment effects become strongly negative, and the credible intervals suggest that the reduction in cigarette consumption is statistically significant during most of the post-treatment period.



Figure 2.2: *Per-capita cigarette sales gap between California and synthetic California by the SCM. The shaded area presents 95% prediction interval for the estimated treatment effect in the post-treatment period.*

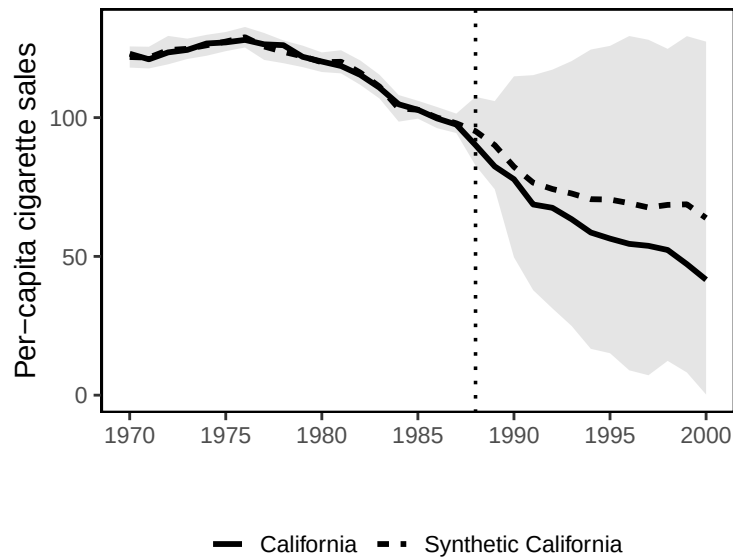


Figure 2.3: *Trends in per-capita cigarette sales: California vs. synthetic California by the GSCM. The shaded region marks the 95% C.I. obtained by a bootstrap approach.*

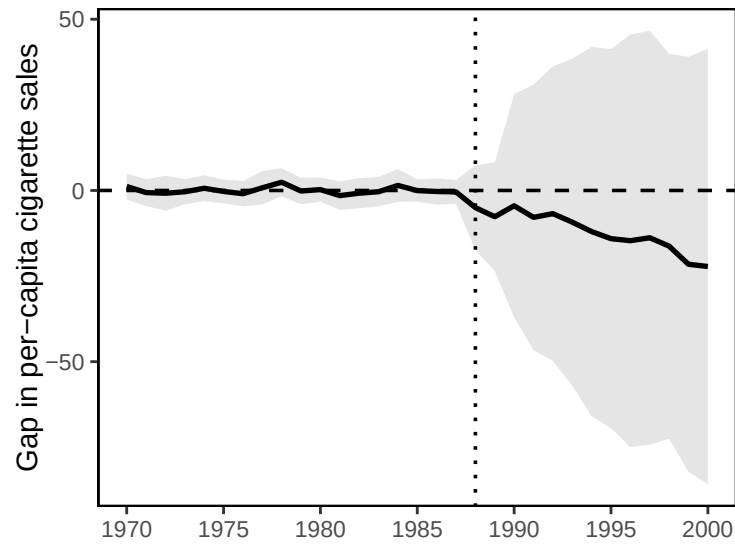


Figure 2.4: *Per-capita cigarette sales gap between California and synthetic California by the GSCM. The shaded region marks the 95% C.I. obtained by a bootstrap approach.*

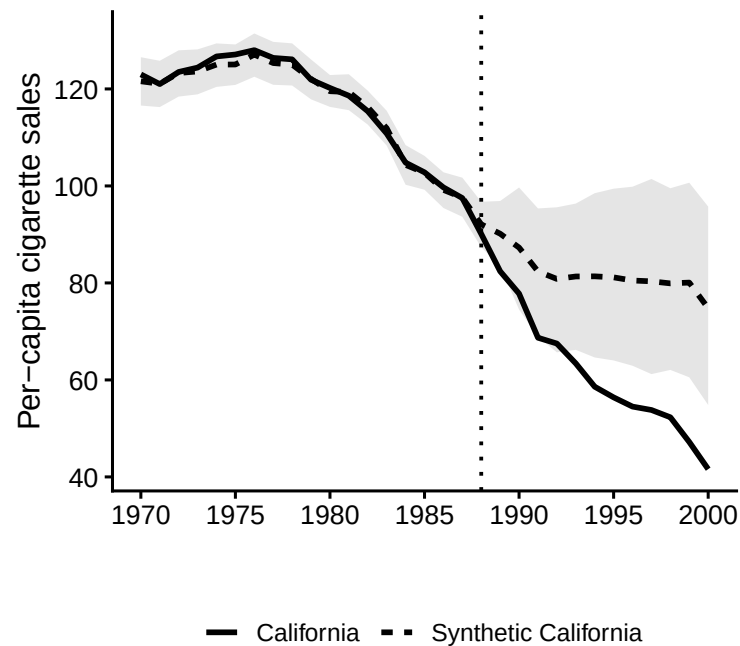


Figure 2.5: *Trends in per-capita cigarette sales: California vs. synthetic California by the B-GSCM. The shaded region marks the 95% credible interval for synthetic CA.*

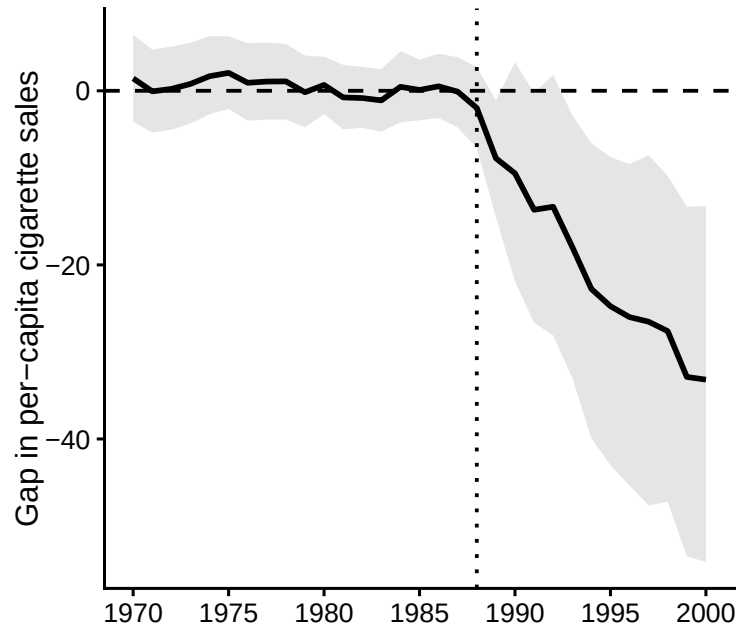


Figure 2.6: *Per-capita cigarette sales gap between California and synthetic California by the B-GSCM. The shaded region marks the 95% credible interval for treatment effect.*

2.5 Discussion

In observational studies, the synthetic control method and the generalized synthetic control method are regarded as the most influential technologies. These two methods attempt to deal with a problem that the parallel trend assumption frequently fails to address. The GSCM and SCM are based on a modeling framework, but SCM ignore the model when comes to estimate the weights. The SCM can also provide some statistical inferences by applying placebo tests.

This chapter presents the Bayesian generalized synthetic control method from a matrix completion perspective. By treating the untreated post-treatment outcomes of the treated unit as missing entries in a partially observed outcome matrix, the proposed framework connects causal inference with matrix completion and offers a probabilistic approach to recovering the missing counterfactual outcomes. This perspective is especially useful because it emphasizes the role of latent structure in reconstructing the unobserved portion of the panel

and provides a natural motivation for the Bayesian modeling framework developed in the next chapter. Although the proposed MCMC algorithm performs reasonably well, there is still room for improvement. In particular, computational efficiency and convergence may be further enhanced through more advanced sampling strategies, such as parameter expansion (PX), which will be introduced in Chapter 3.

Chapter 3

Bayesian Generalized Synthetic Control Method with Treatment Modeling

In this chapter, we introduce an alternative approach based on treatment modeling. It directly models the treatment process rather than focusing only on outcome reconstruction. This method allows for a more structured incorporation of covariate information and temporal dynamics. Additionally, a key innovation in our framework is the incorporation of Parameter Expansion (PX) into Bayesian factor models, following [Ghosh and Dunson \(2009\)](#). PX introduces auxiliary parameters that temporarily expand the parameter space during posterior sampling. This expansion reduces the posterior dependence between factor loadings and latent factors, leading to improved convergence and efficiency in Markov chain Monte Carlo (MCMC) sampling, while preserving proper prior distributions and stable posterior inference. By leveraging PX, our Bayesian generalized synthetic control model achieves faster convergence, more reliable estimates of latent factors, and more accurate treatment effect estimation compared with the standard GSCM.

The structure of this chapter is as follows: Section [3.1](#) describes the model setup and prior specifications, including the introduction of the parameter expansion approach. Sec-

tion 3.2 presents a fully Bayesian approach to GSCM with treatment modeling, including technical details of posterior computation. The calculation of the full conditional posterior distributions is shown in Appendix C. Section 3.3 conducts a simulation study, followed by an empirical example involving carbon taxes and CO₂ emissions data in Section 3.4. We conclude this chapter in Section 3.5.

3.1 Model setup and prior specification

Suppose y_{it} is the outcome for unit i at time t , for units $i = 1, \dots, N$ and time periods $t = 1, \dots, T$. Let $\mathbf{T}_{0,i}$ represent the number of pretreatment periods for unit i . The sets $\mathcal{T}$ and $\mathcal{C}$ denote the treatment and control groups, respectively. Let N_{tr} and N_{co} denote the number of treated and control units. The total number units $N = N_{tr} + N_{co}$. This framework accommodates multiple treatment units, while the focus of this dissertation is on a more specific case: the single treatment unit scenario, in which only one unit is exposed to the treatment. Assume that y_{it} is given by a linear factor model:

$$y_{it} = \tau_{it}D_{it} + \mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i\mathbf{f}_t + \epsilon_{it}, \quad (3.1)$$

where τ_{it} is the treatment effect on unit i at time t , $D_{it} = 1$ when $i \in \mathcal{T}$ and $t > \mathbf{T}_{0,i}$ and $D_{it} = 0$ otherwise. $\mathbf{x}_{it}$ is a $(k \times 1)$ vector of observed covariates, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ is a $(k \times 1)$ vector of unknown parameters. $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{ir})'$ is an $(r \times 1)$ vector of unknown factor loadings, $\mathbf{f}_t = (f_{1t}, \dots, f_{rt})'$ is an $(r \times 1)$ vector of unobserved common factors, ϵ_{it} is the random error term for unit i at time t and we assume $\epsilon_{it} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_t^2)$.

Gibbs sampling is applied for posterior inference in next section. However, the Gibbs sampler can exhibit poor convergence properties, resulting in slow mixing. To reduce the burden of computation, we define the following PX-factor model:

$$y_{it} = \tau_{it}D_{it} + \mathbf{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\lambda}'_i\mathbf{f}_t^* + \epsilon_{it}, \quad \boldsymbol{\lambda}_i \stackrel{\text{iid}}{\sim} \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Psi}). \quad (3.2)$$

where $\mathbf{f}_t^* = (f_{t1}^*, \dots, f_{tr}^*)'$ is a vector of working factors, and $\mathbf{\Psi} = \text{diag}(\psi_1, \dots, \psi_r)$.

To relate the working model parameters in (3.2) to the inferential model parameters in (3.1), we use the following transformation:

$$f_{tl} = S(f_{tl}^*) \psi_l^{-1/2} f_{tl}^*, \quad t = 1, \dots, T, \quad l = 1, \dots, r,$$

where $S(x) = -1$ for $x < 0$ and $S(x) = 1$ for $x \geq 0$.

The selection of priors is a crucial aspect of the Bayesian factor analysis model that should be carefully considered. To complete our Bayesian modeling, we need to attribute appropriate priors to unknown parameters. For algebraic and computational convenience, we assign the multivariate Gaussian prior for $\boldsymbol{\beta}$ as follows:

$$\boldsymbol{\beta} \sim \mathcal{N}_k(\mathbf{0}, v_0 \mathbf{I}_k),$$

where v_0 is a hyperparameter.

We induce a prior distribution on $\mathbf{F}$ through a prior distribution for $\mathbf{F}^*$, we have

$$f_{tl}^* \stackrel{iid}{\sim} \mathcal{N}(0, 1), \quad t = 1, \dots, T, \quad l = 1, \dots, \min(t, r),$$

$$f_{tl}^* \sim \delta_0, \quad t = 1, \dots, r-1, \quad l = t+1, \dots, r,$$

where δ_0 is the point mass distribution at 0.

Given the structure of the priors, we can now express the resulting matrix $\mathbf{F}^*$ as a $T \times r$ lower triangular matrix. This leads us to the following form for $\mathbf{F}^*$

$$\mathbf{F}^* = \begin{pmatrix} f_{11}^* & 0 & 0 & \cdots & 0 \\ f_{21}^* & f_{22}^* & 0 & \cdots & 0 \\ f_{31}^* & f_{32}^* & f_{33}^* & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ f_{T1}^* & f_{T2}^* & f_{T3}^* & \cdots & f_{Tr}^* \end{pmatrix}$$

Inverse gamma priors are frequently used to characterise error variances. We assign the following inverse gamma priors to σ_t^2, ψ_l :

$$\begin{aligned}\sigma_t^2 &\overset{iid}{\sim} IG(a_\sigma/2, b_\sigma/2), \\ \psi_l &\overset{iid}{\sim} IG(a_\psi/2, b_\psi/2),\end{aligned}$$

where $a_\sigma, b_\sigma, a_\psi$, and b_ψ are hyperparameters.

Our final goal is to obtain the main quantity of interest τ_{it} , for $i = 1$ and $t = T_{0,1} + 1, \dots, T$. For simplicity, we denote this as $\boldsymbol{\tau}_1$, which we assume follows a normal distribution with mean $\mu(t)$ and variance θ^2 , $\theta^2 \sim IG\left(\frac{a_\theta}{2}, \frac{b_\theta}{2}\right)$, expressed as

$$\boldsymbol{\tau}_1 \sim \mathcal{N}(\mu(t), \theta^2).$$

Now, we model $\mu(t)$ using a B-spline basis expansion:

$$\mu(t) = \sum_{j=1}^K \alpha_j B_j(t),$$

where $B_j(t)$ are the B-spline basis functions and α_j are the corresponding basis coefficients. We can express this in matrix form as:

$$\mu(t) = \mathbf{B}\boldsymbol{\alpha},$$

where $\mathbf{B}$ is a matrix with elements $[\mathbf{B}]_{ij} = B_j(t_i)$ and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)'$ is the $K \times 1$ vector of basis coefficients.

3.2 Posterior inference

In order to derive the posterior distribution and implement the Gibbs sampler efficiently, it is convenient to rewrite the PX-factor model in matrix form. This representation allows the

likelihood and the full conditional distributions to be expressed more compactly and makes the posterior computation easier to present. Specifically, we consider two equivalent forms of the model: a unit-specific representation obtained by stacking observations over time for each unit, and a time-specific representation obtained by stacking observations across units for each time period. These two forms are useful for deriving the conditional posterior distributions of different blocks of parameters in the Gibbs sampler.

For unit i , we can rewrite the (3.2) model as:

$$\mathbf{y}_i = \boldsymbol{\tau}_i \odot \mathbf{D}_i + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{F}^* \boldsymbol{\lambda}_i + \boldsymbol{\epsilon}_i, \quad (3.3)$$

where $\boldsymbol{\epsilon}_i \sim \mathcal{N}_T(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_T^2)$. Additionally, $\mathbf{y}_i \sim \mathcal{N}_T(\boldsymbol{\tau}_i \odot \mathbf{D}_i + \mathbf{X}_i \boldsymbol{\beta}, \boldsymbol{\Omega}^*)$, with $\boldsymbol{\Omega}^* = \mathbf{F}^* \mathbf{F}^{*'} + \boldsymbol{\Sigma}$.

For time t , we can rewrite the model in (3.2) as:

$$\tilde{\mathbf{y}}_t = \boldsymbol{\tau}_t \odot \mathbf{D}_t + \tilde{\mathbf{X}}_t \boldsymbol{\beta} + \boldsymbol{\Lambda} \mathbf{f}_t^* + \tilde{\boldsymbol{\epsilon}}_t, \quad (3.4)$$

where $\tilde{\boldsymbol{\epsilon}}_t \sim \mathcal{N}(\mathbf{0}, \sigma_t^2 \mathbf{I}_N)$.

We propose that the main quantity of interest τ_{it} can be generated from the posterior distribution by the following steps:

Step 1. Obtain the estimated values $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$, $\mathbf{F}^*$, and $\boldsymbol{\Sigma}$. We propose using a parameter expansion approach (Liu and Wu, 1999) to improve the behavior of the Gibbs samplers and using only the control group data.

Step 2. Derive the inferential factors $\mathbf{F}$ from the working factors $\mathbf{F}^*$ through the transformation.

Step 3. Draw the treatment effect τ_{it} from the posterior distribution, where $\tau_{it} \sim \mathcal{N}(\mu(t), \theta^2)$, for $i = 1$ and $t = T_{0,1} + 1, \dots, T$.

To accomplish Step 1, we used a Gibbs sampler under the PX-factor model that generates samples from the joint posterior distribution by iterating the following steps until convergence:

1. Generate β from $\pi(\beta \mid \mathbf{Y}, \mathbf{F}^*, \Sigma)$,
2. Generate Λ from $\pi(\Lambda \mid \mathbf{Y}, \Psi, \mathbf{F}^*, \Sigma, \beta)$,
3. Generate $\mathbf{F}^*$ from $\pi(\mathbf{F}^* \mid \mathbf{Y}, \Lambda, \Psi, \Sigma, \beta)$,
4. Generate Σ from $\pi(\Sigma \mid \mathbf{Y}, \Lambda, \Psi, \mathbf{F}^*, \beta)$,
5. Generate Ψ from $\pi(\Psi \mid \mathbf{Y}, \Lambda, \mathbf{F}^*, \Sigma, \beta)$.

To determine the full conditionals, note that equation (3.2) can be rewritten by moving the treatment effect term $\tau_{it}D_{it}$ and the covariate component $\mathbf{x}'_{it}\beta$ to the left-hand side, yielding y_{it}^* . Hence, following Ghosh and Dunson (2009), the full conditional distributions of β , λ_i , $\mathbf{f}_t^*$, σ_t^2 , and ψ_l are derived using standard conjugate updating; see Appendix C for details.

First, for the coefficients β :

$$\beta \mid \mathbf{Y}, \Sigma \sim \mathcal{N}(\mathbf{M}_\beta, \mathbf{V}_\beta)$$

where

$$\begin{aligned} \mathbf{V}_\beta &= \left(\sum_{i=1}^N \mathbf{X}_i' (\Omega^*)^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right)^{-1}, \\ \mathbf{M}_\beta &= \mathbf{V}_\beta \sum_{i=1}^N \mathbf{X}_i' (\Omega^*)^{-1} (\mathbf{y}_i - \boldsymbol{\tau}_i \odot \mathbf{D}_i). \end{aligned}$$

For each $i = 1, \dots, N$, the full conditional posterior of λ_i is

$$\lambda_i \mid \mathbf{F}^*, \beta, \Sigma, \mathbf{Y} \sim \mathcal{N}(\mathbf{M}_{\lambda,i}, \mathbf{V}_{\lambda,i}),$$

where

$$\begin{aligned} \mathbf{V}_{\lambda,i} &= \left(\mathbf{F}^{*'} \Sigma^{-1} \mathbf{F}^* + \Psi^{-1} \right)^{-1}, \\ \mathbf{M}_{\lambda,i} &= \mathbf{V}_{\lambda,i} \mathbf{F}^{*'} \Sigma^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta - \boldsymbol{\tau}_i \odot \mathbf{D}_i). \end{aligned}$$

The full conditional posterior of $\mathbf{f}_t^*$ is

$$\mathbf{f}_t^* \mid \boldsymbol{\Lambda}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{Y} \sim \mathcal{N}(\mathbf{M}_{f,t}, \mathbf{V}_{f,t}),$$

where

$$\begin{aligned} \mathbf{V}_{f,t} &= (\sigma_t^{-2} \boldsymbol{\Lambda}_{k_t}' \boldsymbol{\Lambda}_{k_t} + \mathbf{I}_{k_t})^{-1}, \\ \mathbf{M}_{f,t} &= \mathbf{V}_{f,t} \sigma_t^{-2} \boldsymbol{\Lambda}_{k_t}' \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\tau}_t \odot \mathbf{D}_t \right). \end{aligned}$$

where $\tilde{\mathbf{y}}_t = (y_{1t}, \dots, y_{Nt})$ and $\tilde{\mathbf{X}}_t = (\mathbf{x}_{1t}, \dots, \mathbf{x}_{Nt})'$, for $t = 1, \dots, T$, $k_t = \min(t, r)$, and $\boldsymbol{\Lambda}_{k_t}$ is an $N \times k_t$ matrix of loadings.

Finally, the variance parameters follow Inverse-Gamma distributions:

$$\begin{aligned} \sigma_t^2 \mid \dots &\sim IG \left(\frac{a_\sigma + N}{2}, \frac{b_\sigma + \mathbf{e}_t' \mathbf{e}_t}{2} \right), \quad l = 1, \dots, T \\ \psi_l \mid \dots &\sim IG \left(\frac{a_\psi + N}{2}, \frac{b_\psi + \sum_{i=1}^N \lambda_{il}^2}{2} \right), \quad l = 1, \dots, r \end{aligned}$$

where $\mathbf{e}_t = \tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\Lambda} \mathbf{f}_t^* - \boldsymbol{\tau}_t \odot \mathbf{D}_t$ represents the vector of residuals at time t .

Step 2 is to recover the inferential factors $\mathbf{F}$ from the working factors $\mathbf{F}^*$. Given the PX-factor model, the sampled working factors $\mathbf{F}^*$ are transformed into the inferential factors $\mathbf{F}$ using the transformation defined in Section 3.1. This step ensures that the factor structure satisfies the identification restrictions required for the inferential model.

The remaining procedure for attaining Step 3 is as follows. To derive posterior distribution of $\boldsymbol{\tau}_1$ with prior distribution $\theta^2 \sim IG \left(\frac{a_\theta}{2}, \frac{b_\theta}{2} \right)$, we have the full conditional distribution:

$$\boldsymbol{\tau}_1 \mid \text{rest} \sim \mathcal{N} \left(\left(\mathbf{I} + \frac{\sigma_t^2}{\theta^2} \right)^{-1} \left(\mathbf{y}_{1t} - \mathbf{X}_{1t} \boldsymbol{\beta} - \boldsymbol{\lambda}_1' \mathbf{f}_t + \frac{\sigma_t^2}{\theta^2} \mathbf{B} \boldsymbol{\alpha} \right), \sigma_t^2 \left(\mathbf{I} + \frac{\sigma_t^2}{\theta^2} \right)^{-1} \right),$$

where “rest” denotes all remaining model parameters and observed data. To eventually determine the distribution of $\boldsymbol{\tau}_1$, we must first derive the full conditionals of $\boldsymbol{\alpha}$ and θ^2 .

These can be found as follows.

$$\begin{aligned}\boldsymbol{\alpha} \mid \boldsymbol{\tau}_1, \mathbf{B}, \theta^2 &\sim \mathcal{N}((\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\boldsymbol{\tau}_1, \theta^2(\mathbf{B}'\mathbf{B})^{-1}) \\ \theta^2 \mid \boldsymbol{\tau}_1, \boldsymbol{\alpha}, \mathbf{B} &\sim \text{IG}\left(\frac{T_1 + a_\theta}{2}, \frac{\|\boldsymbol{\tau}_1 - \mathbf{B}\boldsymbol{\alpha}\|^2 + b_\theta}{2}\right)\end{aligned}$$

The results follow from standard conjugate updating; see Appendix C for the detailed derivations.

3.3 Simulation study

In this section, we conduct a Monte Carlo simulation study to examine the performance of the proposed generalized synthetic control method. Given $N=40$, $T=100$, $T_0=60$, and $p=4$, we simulate the data $\{(Y_{it}, \mathbf{Z}_{it}) : i = 1, \dots, N, t = 1, \dots, T\}$ from the following linear factor model:

$$Y_{it} = \alpha_i + \gamma_t + \boldsymbol{\lambda}_i' \mathbf{f}_t + \mathbf{c}' \mathbf{Z}_{it} + \log(t - 60) D_{it} + \epsilon_{it},$$

where $D_{it}=1(i=1, t > T_0)$, $\alpha_i \stackrel{\text{iid}}{\sim} N(0, \sigma_{\alpha i}^2)$ with $\sigma_{\alpha i}^2 = 0.5$, $\gamma_t \stackrel{\text{iid}}{\sim} N(\gamma_{t-1}, \sigma_\gamma^2)$ with $\sigma_\gamma^2 = 0.1$ and $\gamma_0 = 0$, $\mathbf{f}_t = (f_{1t}, f_{2t})'$ denote the latent factor vector, where $\mathbf{f}_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, $\boldsymbol{\lambda}_i = (\lambda_{i1}, \lambda_{i2})'$ denote the corresponding factor loading vector. The factor loadings follows $\boldsymbol{\lambda}_i \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$, $\mathbf{c} = (1, 1, 1, 1)'$, $\mathbf{Z}$ is a three-dimensional $(N \times p \times T)$ matrix with $\mathbf{Z}_{it} \sim N_p(\mathbf{1}_p, 2\mathbf{I}_p)$, and $\epsilon_{it} \stackrel{\text{iid}}{\sim} N(0, 0.1)$.

We simulate the data $\{(\mathbf{Y}_i, \mathbf{X}_i) : i = 1, \dots, N\}$ from the data $\{(Y_{it}, \mathbf{Z}_{it}) : i = 1, \dots, N, t = 1, \dots, T\}$. $\mathbf{X}$ is a three-dimensional $(T \times p \times N)$ matrix, $\mathbf{X}_i$ is a $(T \times p)$ matrix with $\mathbf{X}_i = \mathbf{Z}_i'$. $\mathbf{Y}_i = (Y_{i1}', \dots, Y_{iT}')'$ is a $(T \times N)$ matrix. We use hyperparameters $v_0 = 100$, $a_\sigma = 1$, $b_\sigma = 1$, $a_\psi = 1$, $b_\psi = 1$, $a_\theta = 100$, and $b_\theta = 1$.

In the data generating process, the treatment effect of unit $i=1$ at time $t > T_0$ is given by:

$$\tau_{1t} = \log(t - 60).$$

For posterior computation, we run the Gibbs sampler for 10,000 iterations, with the first 2,000 iterations discarded as burn-in. The remaining samples are used for posterior inference. For the Bayesian GSCM specification, we model the treatment effect τ_{1t} using the B-spline functional form for the mean $\mu(t)$: We employ a cubic B-spline (degree = 3) with df degrees of freedom:

$$\mu(t) = \sum_{j=1}^{df} \alpha_j B_j(t)$$

where $B_j(t)$ represents the B-spline basis functions. In this simulation, we set $df = 4$. Figure 3.1 illustrates a representative simulation result for the B-spline treatment effect specification. The black line represents the true treatment effect trajectory, while the blue solid line shows the posterior mean estimate obtained from the Bayesian B-spline model. The dashed lines represent the 95% credible interval. The estimated treatment effect closely tracks the true effect across the post-treatment period, indicating that the proposed method successfully recovers the underlying treatment dynamics.

Table 3.1: *Summary of simulation study over 100 runs with a lower-noise setting.*

Method	MSE	S.E.	Avg. 95% CI Width
SCM	9.751	0.249	10.62
GSCM with CV rank	0.132	0.004	1.61
GSCM with true rank 4	0.120	0.003	1.44
B-spline Bayes GSCM	0.096	0.045	0.63

Table 3.1 summarizes the results of simulations that are conducted to compare the performance of four different methods: SCM, GSCM with cross-validation rank selection, GSCM with the factor true rank, and B-spline Bayesian GSCM in estimating the average treatment effects, the performance of various estimation methods is evaluated based on their Mean Squared Error and Standard Error. SCM exhibits the highest MSE among all methods evaluated, this indicates a relatively high level of error and less precision in estimation. The SCM shows a very high MSE, which might be explained by it cannot properly capture latent factor dynamics. The performance of GSCM is significantly better than that of SCM, as evidenced by its considerably reduced MSE of 0.132 and standard error of 0.004. This indi-

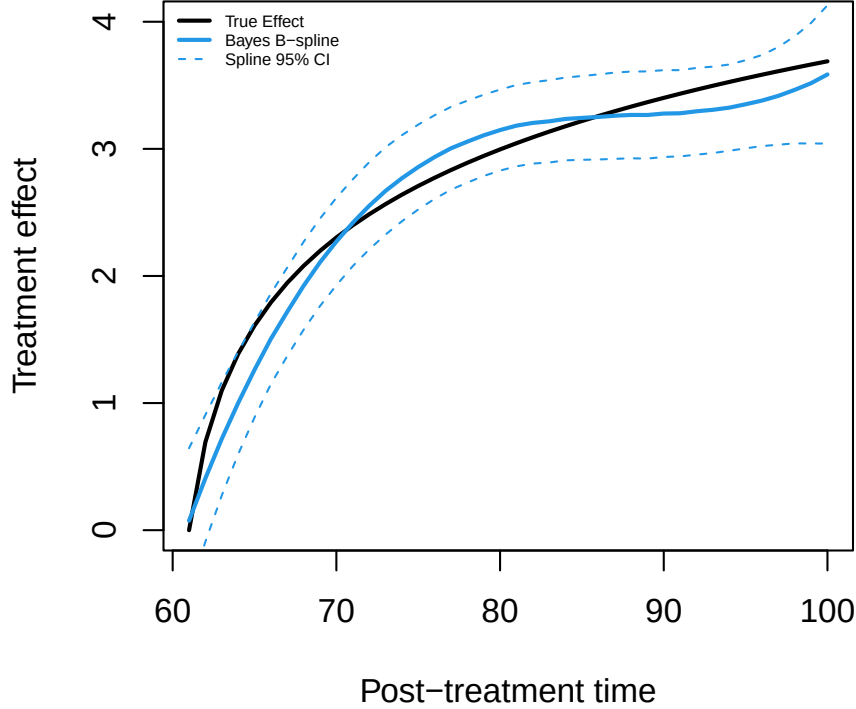


Figure 3.1: *True treatment effect, Bayesian B-spline estimate, and 95% credible interval from the simulated example.*

cates improved precision and accuracy in comparison to the baseline method. The B-spline method achieves the smallest MSE of 0.096 among all methods.

Table 3.2 presents the average lengths of 95% interval for the average treatment effect across 100 Monte Carlo replications for different methods. The classical synthetic control method (SCM) yields the largest average 95% interval length, at 10.62. The classical SCM constructs the counterfactual using a sparse convex combination of donor units. Because the estimator places nonzero weight on only a limited subset of donors, the resulting synthetic control can be highly sensitive to random variation in those selected units. In a simulation setting, such sensitivity amplifies sampling variability across replications. Consequently, the estimated treatment effects exhibit substantial dispersion, leading to a comparatively wide 95% interval. This indicates that, while SCM can recover point estimates, its inferential

precision is relatively low under the simulated data-generating process.

The GSCM produces an average 95% interval length of 1.61, representing approximately an 85% reduction in interval width relative to the classical SCM. Unlike SCM, which relies on a sparse convex combination of donor units, GSCM is based on an interactive fixed effects factor model. The estimated treatment effects exhibit much lower dispersion, yielding significantly tighter confidence intervals. These findings indicate that, under the simulated setting, the factor-based structure of GSCM leads to markedly improved inferential precision compared to the classical SCM.

The B-spline Bayesian generalized synthetic control method achieves the smallest average 95% interval length (0.63) among all competing methods. This corresponds to an interval approximately 17 times narrower than that of the classical SCM and about 2.5 times narrower than the standard GSCM.

The superior performance of the proposed method reflects two key features of the model. First, the Bayesian framework provides full posterior inference through MCMC sampling, yielding a comprehensive characterization of parameter uncertainty. By integrating over the joint posterior distribution rather than relying solely on asymptotic approximations, the Bayesian approach can deliver more efficient and stable inference in finite samples. Second, the incorporation of B-spline basis functions introduces a flexible smoothing structure for modeling the treatment effect dynamics. The spline representation reduces short run noise and volatility while maintaining the underlying trend in the data.

Table 3.2: *Summary of simulation study over 100 runs with a higher-noise setting.*

Method	MSE	S.E.	Avg. 95% CI Width
SCM	10.94	0.26	11.32
GSCM with CV rank	1.33	0.03	5.01
GSCM with true rank 4	1.24	0.03	4.73
B-spline Bayes GSCM	1.06	0.30	1.37

To assess the robustness of the proposed method, we consider an additional simulation setting with a higher noise level. Specifically, we generate the error term as $\epsilon_{it} \sim \mathcal{N}(0, 1)$.

Table 3.2 reports the simulation results under the higher-noise setting. As expected, the

increased variance leads to larger mean squared errors and wider uncertainty intervals for all methods. However, the proposed B-spline Bayesian GSCM continues to outperform GSCM, achieving lower MSE and substantially shorter interval lengths.

These results suggest that the proposed method remains more stable and provides more precise uncertainty quantification under higher noise levels. Similar results are observed in both settings. Using the true factor rank leads to improved performance compared to the cross-validation-selected rank, indicating that accurate rank specification remains crucial even when incorporating flexible treatment effect modeling. This finding further motivates the development of a Bayesian rank selection approach, which is presented in Chapter 4.

3.4 Empirical example

To demonstrate the applicability of the Bayesian Generalized Synthetic Control Method with treatment modeling, we revisit the case study of (Andersson, 2019), which estimates the effect of a carbon tax (Howard and Sylvan, 2015; Sumner et al., 2011; Tol, 2009) and a value-added tax (VAT) on its per capita carbon dioxide (CO_2) emissions from transport in Sweden. We use the same data setting as Andersson (2019) and the data is obtained from the World Bank(2015). Derived from an initial list of 25 nations belonging to the Organisation for Economic Cooperation and Development (OECD). Excluded certain countries due to their implementation of carbon price policies or to prevent causing interpolation bias. In the end, we estimate CO_2 emissions from transport in Sweden using the donor pool consisting of 14 countries. In Andersson (2019), per capita CO_2 emissions from transport during 1960-2005 is used as the outcome and 4 variables (GDP per capita, number of motor vehicles, gasoline consumption per capita, and percentage of urban population) associated with CO_2 emissions reduction potential are considered as the time-invariant covariates. The period from 1960 to 1989 is used to create a synthetic control group, while the period from 1990 to 2005 is used to analyze the impact of the carbon tax reform on emissions in Sweden compared to the synthetic Sweden. Thus, the sample period consists of 30 years of data before the treatment and 16 years of data after the treatment. In this real data application, the Gibbs sampler is

run for 15,000 iterations, with the first 30% of iterations discarded as burn-in.

Figure 3.2, the SCM fitted path plot, shows that the synthetic control closely follows Sweden’s outcome in the pre-treatment period, suggesting a good fit. After the intervention, the observed path falls below the synthetic one, indicating a negative treatment effect. These findings are highlighted in gap plot Fig. 3.3, which fluctuates around zero before treatment, then becomes negative afterward. This suggests that CO₂ emissions decreased relative to the counterfactual, although the gap shows some variability over time.

The GSCM path plot Fig. 3.4 also fits the pre-treatment data well and produces a smoother counterfactual trajectory. After treatment, Sweden’s observed outcome declines relative to the estimated counterfactual, again indicating a negative effect. In the gap plot Fig. 3.5, the pre-treatment values remain close to zero, while the post-treatment gap is more stable and consistently negative. Compared to SCM, the GSCM results are smoother, reflecting its ability to capture underlying latent trends.

Figure 3.6 displays per capita CO₂ emissions from transport for Sweden and its synthetic counterpart using the Bayesian Generalized Synthetic Control Method with B-spline specification during the years 1960 to 2005. Following the introduction of the carbon tax at period 30, the observed emissions remain slightly below the counterfactual mean throughout most of the post-treatment period.

Figure 3.7 estimated treatment effect for Sweden under the Bayesian Generalized Synthetic Control Method with B-spline parameterization. The solid black line represents the posterior mean of the gap between Sweden’s observed and counterfactual CO₂ emissions from transport (measured in metric tons per capita). The grey dotted lines indicate the 95% posterior credible interval. In the post-treatment period, the gap turns negative, with an estimated effect of approximately -0.2 to -0.3 metric tons per capita, indicating a reduction in transport-related CO₂ emissions relative to the counterfactual trajectory. The 95% credible interval becomes noticeably wider in the post-treatment period and consistently includes zero. This suggests that, under the Bayesian GSCM framework, the estimated reduction in emissions associated with the carbon tax cannot be clearly distinguished from zero at conventional significance levels.

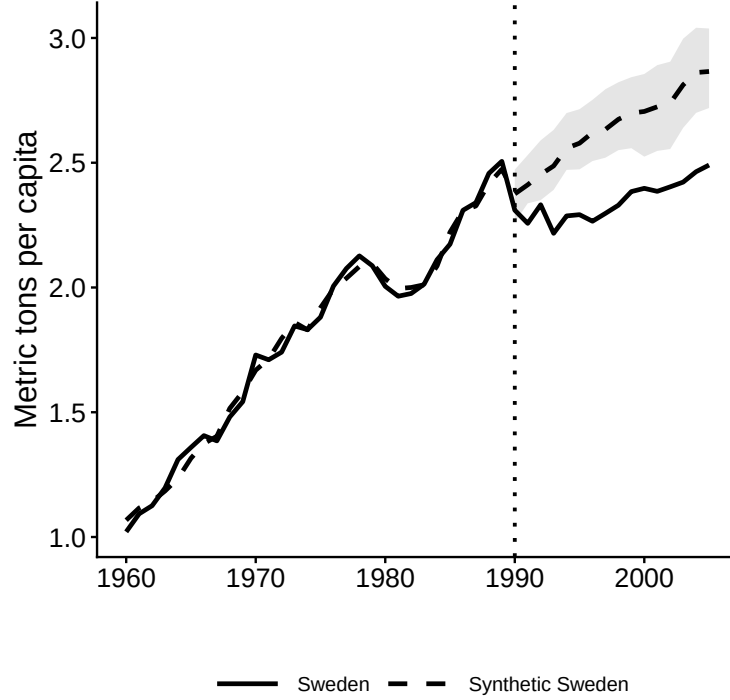


Figure 3.2: *Path Plot of Per Capita CO₂ Emissions from Transport during 1960–2005: Sweden versus Synthetic Sweden by the SCM*

Table 3.3 reports the estimated treatment effects obtained from SCM, GSCM, and Bayesian GSCM. All three methods indicate negative estimates throughout most of the post-treatment period, suggesting that emissions decreased relative to the counterfactual. The SCM tends to produce larger negative effects, indicating a stronger estimated impact. In contrast, the GSCM results are generally closer to zero and show less variation over time. The Bayesian GSCM lies between the two, with estimates that are smoother and gradually move toward zero, reflecting its ability to capture more stable and time-varying treatment effects.

Table 3.4 summarizes the 95% intervals and their lengths for SCM, GSCM, and Bayesian GSCM. SCM produces the narrowest intervals, which mostly exclude zero, suggesting a significant effect. GSCM yields much wider intervals that often include zero, indicating substantial uncertainty. The Bayesian GSCM provides intervals of intermediate width, offering a more balanced assessment of uncertainty and generally more stable inference.

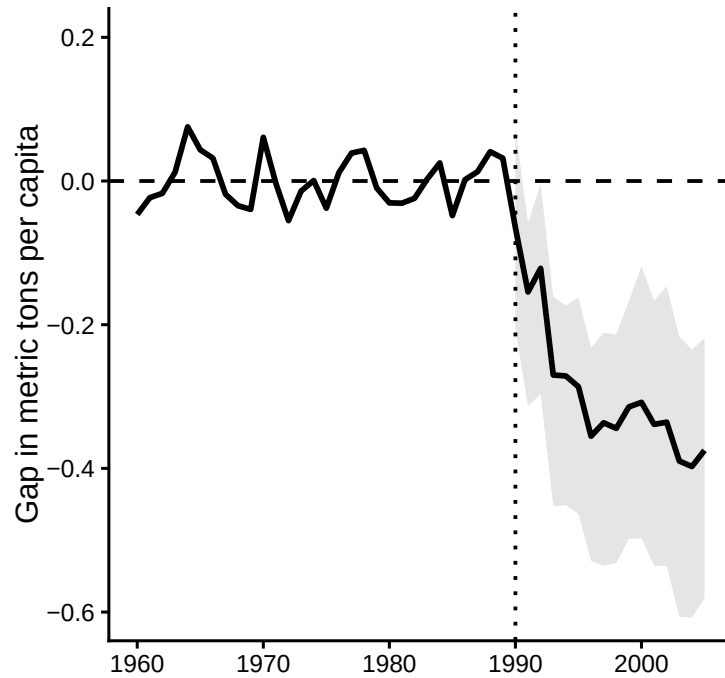


Figure 3.3: Gap in per capita CO₂ emissions from transport between Sweden and synthetic Sweden by the SCM

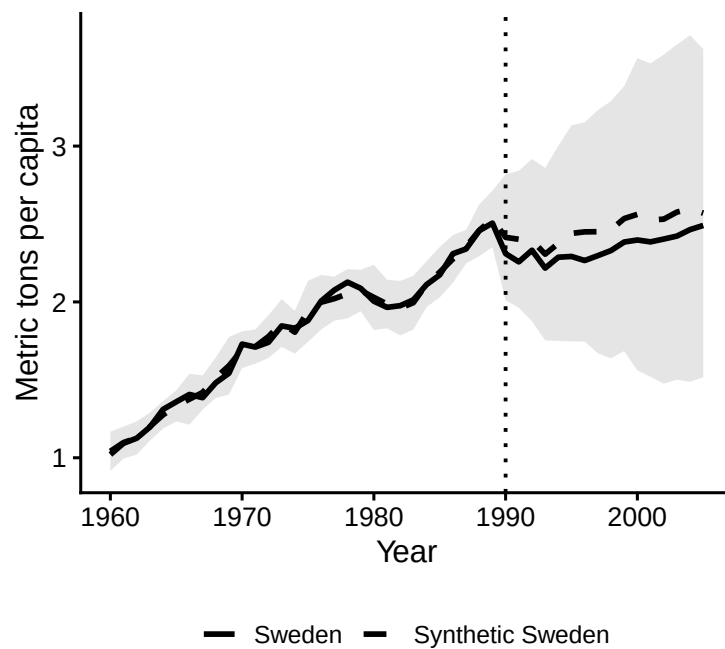


Figure 3.4: Path Plot of Per Capita CO₂ Emissions from Transport during 1960–2005: Sweden versus Synthetic Sweden by the GSCM

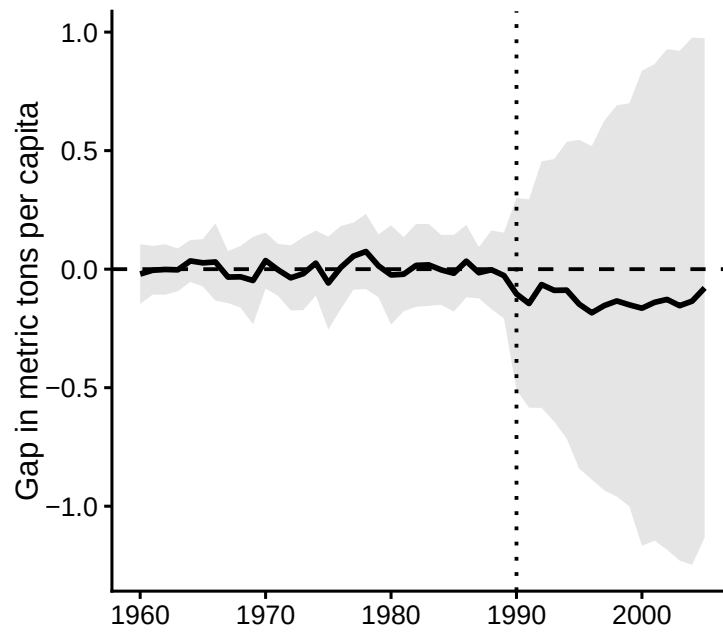


Figure 3.5: Gap in per capita CO₂ emissions from transport between Sweden and Synthetic Sweden by the GSCM

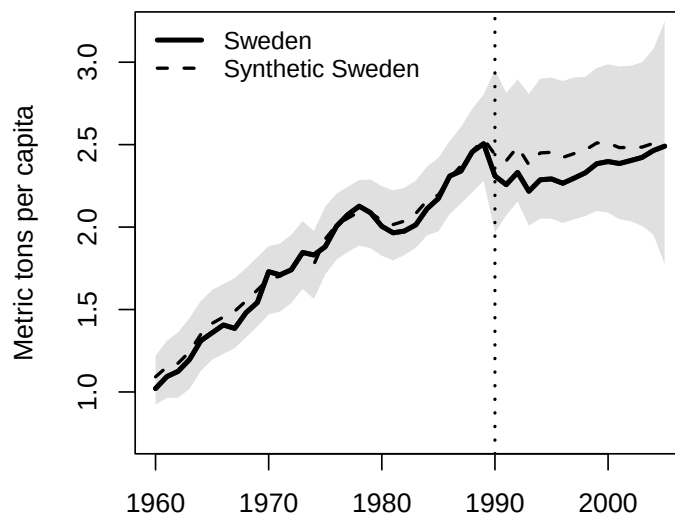


Figure 3.6: Gap in per capita CO₂ emissions from transport between Sweden and Synthetic Sweden by the B-spline GSCM

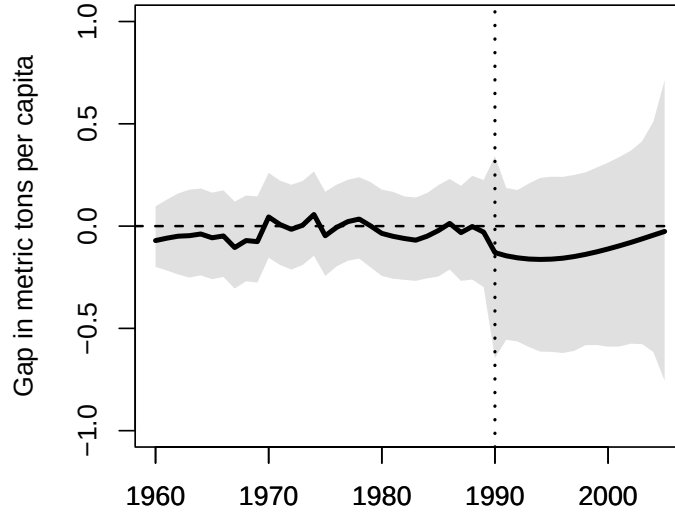


Figure 3.7: *Path Plot of Per Capita CO₂ Emissions from Transport during 1960–2005: Sweden versus Synthetic Sweden by the B-spline GSCM*

Table 3.3: *Comparison of Estimated Treatment Effects: SCM, GSCM, and BGSCM. Cell values represent the estimated treatment effect (Sweden - Synthetic Sweden)*

Year	SCM	GSCM	BGSCM
1991	-0.154	-0.105	-0.132
1992	-0.122	-0.145	-0.147
1993	-0.270	-0.065	-0.155
1994	-0.271	-0.089	-0.158
1995	-0.286	-0.089	-0.157
1996	-0.355	-0.148	-0.151
1997	-0.337	-0.184	-0.143
1998	-0.344	-0.154	-0.130
1999	-0.314	-0.134	-0.115
2000	-0.308	-0.151	-0.098
2001	-0.339	-0.165	-0.079
2002	-0.336	-0.140	-0.059
2003	-0.390	-0.128	-0.038
2004	-0.397	-0.154	-0.016
2005	-0.375	-0.135	0.006

Table 3.4: *95% CI of Treatment Effects (Sweden - Synthetic Sweden) and Interval Length*

Yr	SCM			GSCM			BGSCM		
	L	U	Len	L	U	Len	L	U	Len
1991	-0.31	-0.06	0.25	-0.41	0.20	0.61	-0.6434	0.3460	0.9894
1992	-0.29	-0.00	0.29	-0.43	0.14	0.56	-0.5637	0.1901	0.7538
1993	-0.45	-0.16	0.29	-0.42	0.29	0.71	-0.5807	0.1762	0.7569
1994	-0.45	-0.17	0.28	-0.45	0.27	0.73	-0.6072	0.2117	0.8189
1995	-0.46	-0.16	0.30	-0.53	0.35	0.88	-0.6236	0.2354	0.8590
1996	-0.53	-0.23	0.30	-0.64	0.34	0.98	-0.6169	0.2407	0.8576
1997	-0.53	-0.21	0.32	-0.67	0.31	0.98	-0.6133	0.2486	0.8619
1998	-0.53	-0.21	0.32	-0.72	0.41	1.12	-0.5889	0.2615	0.8504
1999	-0.50	-0.16	0.33	-0.72	0.45	1.17	-0.5556	0.2852	0.8408
2000	-0.50	-0.12	0.37	-0.75	0.45	1.21	-0.5440	0.3313	0.8753
2001	-0.53	-0.17	0.37	-0.91	0.58	1.49	-0.5531	0.3833	0.9364
2002	-0.54	-0.15	0.38	-0.88	0.60	1.47	-0.5452	0.4181	0.9633
2003	-0.60	-0.22	0.38	-0.90	0.65	1.55	-0.5367	0.4582	0.9949
2004	-0.60	-0.24	0.37	-0.92	0.61	1.54	-0.5290	0.5043	1.0333
2005	-0.58	-0.22	0.35	-0.93	0.66	1.60	-0.5550	0.5844	1.1394
AVG	—	—	0.33	—	—	1.07	—	—	0.90

3.5 Discussion

The results in this chapter highlight the clear benefits of explicitly incorporating treatment modeling within the Bayesian GSCM framework. Rather than relying solely on imputing missing counterfactual outcomes, as emphasized in Chapter 2, this extension offers a more systematic and transparent way to capture and interpret dynamic treatment effects over time.

The SCM displayed the largest Mean Squared Error, indicating a fundamental failure to appropriately capture the latent factor dynamics of the data-generating process. In contrast, all Bayesian techniques achieved much reduced error rates. While the frequentist GSCM showed significant improvements over the SCM, the B-spline Bayesian GSCM attained the highest overall accuracy, effectively decreasing the distance between the estimated and true treatment effects. For reference, linear and quadratic parametric specifications are also available. The detailed derivations and full conditional posterior distributions for these polynomial models are provided in Appendix D.

A key improvement of this chapter is the introduction of the parameter expansion tech-

nique into the Gibbs sampler. This strategy efficiently reduces the computational problems and slow convergence that often arise in standard Bayesian factor models. By increasing the mixing behavior of the Markov chain, the PX method leads to more stable and reliable posterior inference.

This chapter develops a flexible approach to modeling treatment effects and improves computation through parameter expansion. However, the performance of the factor model also depends on choosing the number of latent factors. This issue is taken up in the next chapter, where we propose a Bayesian approach to rank selection.

Chapter 4

Rank selection for the Bayesian Generalized Synthetic Control Method

In this chapter, we develop a Bayesian procedure for selecting the number of latent factors in the generalized synthetic control model. The number of factors is a key modeling choice in the interactive fixed effects framework because it determines the dimension of the latent factor structure in the model. An inappropriate choice of rank may lead to underfitting or overfitting problems, both of which can negatively affect treatment effect estimation and inferential accuracy. While several procedures exist for selecting the number of factors, many of them are based on informal guidelines and are not easily incorporated into a Bayesian framework. To address this issue, we propose a Bayesian rank selection approach based on posterior model probabilities. The method evaluates candidate ranks through their marginal likelihoods.

The remainder of this chapter is organized as follows. Section [4.1](#) reviews the motivation and background for rank selection. Section [4.2](#) presents the model setup and the pretreatment data used for rank estimation. Section [4.3](#) introduces the Bayesian rank selection criterion. Section [4.4](#) presents simulation results that illustrate the performance of the

proposed method. The chapter concludes with a brief summary.

4.1 Background and motivation

The estimation and choice of the number of common factors play a crucial role in factor analysis, as they impact the model’s ability to accurately represent the fundamental structure of the observed variables (Bai and Ng, 2002). Identifying the correct number of common factors helps in reducing the dimensionality of the data. Choosing an appropriate number of factors enhances the accuracy of the model. In general, the factor rank controls the degree of dimensionality reduction and determines how much unobserved variation can be captured through the common factor structure (Stock and Watson, 2002). If too few factors are included, the model may fail to account for important patterns in the data, resulting in underfitting and biased estimation. On the other hand, if too many factors are included, the model may fit noise rather than signal, leading to overfitting and unstable estimation.

This issue is particularly important in the generalized synthetic control method, where the untreated potential outcomes are modeled through an interactive fixed effects structure. In this framework, the number of latent factors determines the flexibility of the model used to approximate the counterfactual trajectory of the treated unit. Because the factor structure is used to approximate the untreated outcomes, selecting the wrong number of factors can have a substantial impact on treatment effect estimation. Although some early studies, such as Holtz-Eakin et al. (1988) and Bai (2009), proceed under the assumption that the number of factors is known, this assumption is rarely realistic in practice. In empirical applications, the true factor rank is unknown and must be estimated from the data before the factor model can be reliably implemented.

A variety of methods have been proposed in the literature to determine the number of factors. Classical approaches include the Kaiser-Guttman criterion (Kaiser, 1991), which retains factors associated with eigenvalues greater than one, and the scree test (Cattell, 1966), which identifies the rank visually from the pattern of ordered eigenvalues. Parallel analysis (Humphreys and Montanelli Jr, 1975) compares eigenvalues from the observed data with

those obtained from randomly generated data, and [Ledesma and Valero-Mora \(2019\)](#) provide practical guidance for implementing this method. Other approaches include maximum likelihood estimation, as used by [Moon and Weidner \(2017\)](#), and information-criterion-based procedures such as those proposed by [Bai and Ng \(2002\)](#) and [Hirose et al. \(2011\)](#). In the generalized synthetic control, [Xu \(2017\)](#) employ cross-validation to select the factor rank by minimizing prediction error over candidate models. The simulation results in tables [2.1](#), [3.1](#), and [3.2](#) demonstrate that the performance of GSCM is sensitive to the choice of the factor rank. In these settings, using the true rank leads to improved estimation accuracy and narrower interval widths compared to the rank selection with cross-validation. These findings indicate that misspecification of the factor dimension can result in suboptimal performance. However, since the true rank is unknown in practice, and cross-validation may not reliably identify the correct factor dimension, a more principled approach to rank selection is needed.

Although these approaches are widely used, they have several limitations when applied to the generalized synthetic control framework. First, many classical methods were originally developed for standard factor analysis and do not fully account for the panel data structure and the missing counterfactual outcomes that arise in synthetic control applications. Second, procedures such as the scree test and the Kaiser-Guttman rule are based on simple empirical guidelines and can be sensitive to sampling variation. Third, cross-validation methods used to select the rank may yield unstable results, particularly in small samples or in situations with high noise levels, short pretreatment periods, or limited donor pools. More importantly, these approaches are not naturally integrated into a Bayesian modeling framework and therefore do not provide a probabilistic measure of uncertainty for the selected rank. Consequently, the choice of rank is often treated as a preliminary tuning step rather than as a core component of the inferential procedure.

To address these limitations, this chapter proposes a Bayesian approach for selecting the number of factors in the generalized synthetic control model. The proposed method treats the rank as an unknown model index and estimates it through posterior model probabilities. For each candidate rank, the marginal likelihood is approximated using a Laplace-based criterion, and the selected rank is the one that maximizes the resulting posterior probability.

This approach provides a principled alternative to cross-validation by balancing model fit and complexity within a coherent Bayesian framework. In addition, it aligns rank selection with the broader posterior inference procedure developed for the Bayesian generalized synthetic control model, thereby offering a unified and probabilistically interpretable method for model determination.

4.2 Model setup for rank selection

Following the interactive fixed effects framework introduced in equation 2.1, the model can be expressed in the matrix form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{\Lambda}\mathbf{F}' + \mathbf{E},$$

where $\mathbf{Y}$ is the $N \times T$ outcome matrix, with N denoting the number of units and T the number of time periods. $\mathbf{X}$ represents the matrix of observed covariates, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients associated with the k covariates. $\mathbf{\Lambda}$ is an $N \times r$ matrix of factor loadings, where r denotes the number of latent factors. $\mathbf{F}$ is a $T \times r$ matrix containing the common latent factors that evolve over time. Finally, $\mathbf{E}$ is an $N \times T$ matrix of idiosyncratic error terms.

In the synthetic control framework, the outcome matrix can be partitioned according to treatment status and time periods as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{1,1:T_0} & \mathbf{Y}_{\text{miss}} \\ \mathbf{Y}_0 & \mathbf{Y}_1 \end{bmatrix},$$

where $\mathbf{Y}_{\text{miss}}$ denotes the unobserved counterfactual outcomes for the treated unit in the post-treatment period, $\mathbf{Y}_{1,1:T_0}$ is a $(T_0 \times 1)$ vector, $\mathbf{Y}_0$ is a $(N_{co} \times T_0)$ matrix, and $\mathbf{Y}_1$ is a $(N_{co} \times (T - T_0))$ matrix.

Due to the structure of our dataset, we estimate r using the observed part during the

pretreatment periods of the dataset, then define:

$$\mathbf{Y}_{\text{pre}} = \begin{bmatrix} \mathbf{Y}_{1,1:T_0} \\ \mathbf{Y}_0 \end{bmatrix}.$$

Then, we can write the reduced factor model:

$$\mathbf{Y}_{\text{pre}} = \mathbf{X}_{\text{pre}}\boldsymbol{\beta}_{T_0} + \boldsymbol{\Lambda}_{T_0}\mathbf{F}'_{T_0} + \mathbf{E}_{T_0},$$

where $\mathbf{Y}_{\text{pre}}$ is a $(N \times T_0)$ matrix, $\mathbf{X}_{\text{pre}}$ is a $(T_0 \times k \times N)$ matrix, $\boldsymbol{\beta}_{T_0}$ is a $(k \times 1)$ vector, $\boldsymbol{\Lambda}_{T_0}$ is a $(N \times r)$ matrix, $\mathbf{E}_{T_0}$ is an $N \times T_0$ matrix of idiosyncratic error terms, and $\mathbf{F}_{T_0}$ is a $(T_0 \times r)$ matrix with the following form to ensure identifiability ([Geweke, 1996](#))

$$\mathbf{F}_{T_0} = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{F}_{T_0}^* \end{bmatrix},$$

where $\mathbf{F}_{T_0}^*$ is a $(T_0 - r) \times r$ matrix.

4.3 Bayesian rank selection method

From a Bayesian perspective, we can use the posterior probability of r to estimate it. By Bayes' theorem, the posterior can be obtained by:

$$\pi(r \mid \mathbf{Y}_{\text{pre}}) = \frac{f(\mathbf{Y}_{\text{pre}} \mid r)\pi(r)}{\sum_{r=1}^{\min(T, N_{co})} f(\mathbf{Y}_{\text{pre}} \mid r)\pi(r)},$$

where

$$f(\mathbf{Y}_{\text{pre}} \mid r) = \int f(\mathbf{Y}_{\text{pre}} \mid \mathbf{F}_{T_0}, \boldsymbol{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \boldsymbol{\Sigma}_{T_0}, r) p(\mathbf{F}_{T_0}, \boldsymbol{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \boldsymbol{\Sigma}_{T_0} \mid r) d\mathbf{F}_{T_0} d\boldsymbol{\Lambda}_{T_0} d\boldsymbol{\beta}_{T_0} d\boldsymbol{\Sigma}_{T_0}.$$

Hence, we can estimate r by maximizing $\pi(r \mid \mathbf{Y}_{\text{pre}})$. However, the computation of $f(\mathbf{Y}_{\text{pre}} \mid r)$ is complicated since $f(\mathbf{Y}_{\text{pre}} \mid r)$ cannot be expressed in a closed form. We employ the Laplace method (Tierney and Kadane, 1986) to address this issue. Let $\phi = (\mathbf{F}_{T_0}, \mathbf{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \boldsymbol{\Sigma}_{T_0})$ and $\hat{\phi} = \underset{\phi}{\operatorname{argmax}} p(\mathbf{Y}_{\text{pre}} \mid \phi, r)$. $J(\hat{\phi})$ be the Hessian matrix of $\log(\mathbf{Y}_{\text{pre}} \mid \phi, r)$ evaluated at $\hat{\phi}$. Then using the Laplace method, the marginal likelihood can be computed by

$$\begin{aligned} f(\mathbf{Y}_{\text{pre}} \mid r) &= \int p(\mathbf{Y}_{\text{pre}} \mid \phi, r) p(\phi \mid r) d\phi \\ &\approx (2\pi)^{\frac{(T_0-r)r+Nr+k+1}{2}} \left[\det(\tilde{\boldsymbol{\Sigma}}) \right]^{\frac{1}{2}} p(\mathbf{Y}_{\text{pre}} \mid \hat{\phi}, r) p(\hat{\phi} \mid r), \end{aligned} \quad (4.1)$$

where $\tilde{\boldsymbol{\Sigma}} = \left[-J(\hat{\phi}) \right]^{-1}$.

We can further simplify the computation of the marginal likelihood. Take the logarithm of both sides and apply Taylor expansion to equation (4.1) to improve the speed of calculating the marginal likelihood by avoiding the computation of the Hessian matrix, we have

$$\log f(\mathbf{Y}_{\text{pre}} \mid r) \approx \log f(\mathbf{Y}_{\text{pre}} \mid \hat{\mathbf{F}}_{T_0}, \hat{\mathbf{\Lambda}}_{T_0}, \hat{\boldsymbol{\beta}}_{T_0}, \hat{\boldsymbol{\Sigma}}_{T_0}, r) - \frac{1}{2} [(T_0 - r)r + Nr + k + 1] \log(N),$$

where

$$\log f(\mathbf{Y}_{\text{pre}} \mid \hat{\mathbf{F}}_{T_0}, \hat{\mathbf{\Lambda}}_{T_0}, \hat{\boldsymbol{\beta}}_{T_0}, \hat{\boldsymbol{\Sigma}}_{T_0}, r) = \log \prod_{i=1}^N \prod_{j=1}^{T_0} f(\mathbf{Y}_{ij} \mid \mathbf{x}_{ij} \boldsymbol{\beta} + \mathbf{c}_{ij}, \sigma^2),$$

where c_{ij} is the (i, j) th element of $\mathbf{C} = \mathbf{\Lambda}_{T_0} \mathbf{F}'_{T_0}$ and $\hat{\mathbf{F}}_{T_0}, \hat{\mathbf{\Lambda}}_{T_0}, \hat{\boldsymbol{\beta}}_{T_0}, \hat{\boldsymbol{\Sigma}}_{T_0}$ are the maximizers of the posteriors.

The maximizers of the posteriors can be obtained from the joint posterior

$$\left(\hat{\mathbf{F}}_{T_0}, \hat{\mathbf{\Lambda}}_{T_0}, \hat{\boldsymbol{\beta}}_{T_0}, \hat{\boldsymbol{\Sigma}}_{T_0} \right) = \arg \max_{\mathbf{F}_{T_0}, \mathbf{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \boldsymbol{\Sigma}_{T_0}} \pi(\mathbf{F}_{T_0}, \mathbf{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \boldsymbol{\Sigma}_{T_0} \mid \mathbf{Y}_{\text{pre}}) \quad (4.2)$$

However, the direct maximization is difficult to calculate. As an alternative, we propose iterating the full conditional posteriors listed below until convergence. Steps 1-4 are available:

$$\hat{\boldsymbol{\beta}}_{T_0} = \arg \max_{\boldsymbol{\beta}_{T_0}} \pi(\boldsymbol{\beta}_{T_0} \mid \mathbf{Y}_{\text{pre}}, \mathbf{F}_{T_0}, \boldsymbol{\Sigma}_{T_0})$$

$$\hat{\Lambda}_{T_0} = \arg \max_{\Lambda_{T_0}} \pi(\Lambda_{T_0} \mid \mathbf{Y}_{\text{pre}}, \mathbf{F}_{T_0}, \Sigma_{T_0}, \beta_{T_0})$$

$$\hat{\mathbf{F}}_{T_0} = \arg \max_{\mathbf{F}_{T_0}} \pi(\mathbf{F}_{T_0} \mid \mathbf{Y}_{\text{pre}}, \Lambda_{T_0}, \Sigma_{T_0}, \beta_{T_0})$$

$$\hat{\Sigma}_{T_0} = \arg \max_{\Sigma_{T_0}} \pi(\Sigma_{T_0} \mid \mathbf{Y}_{\text{pre}}, \Lambda_{T_0}, \mathbf{F}_{T_0}, \beta_{T_0})$$

The result of equation (4.2) and the answer to steps 1-4 are the same once we reach convergence. Because the full conditionals of steps 1-4 under the conjugated model previously specified have a closed form, the maximums can be updated by the following straightforward procedure.

Step 1. Update $\hat{\beta}_{T_0}$. Assume the prior

$$\beta_{T_0} \sim \mathcal{N}(\mathbf{0}, v_0 \mathbf{I}_k).$$

Then

$$\begin{aligned} & \pi(\beta_{T_0} \mid \mathbf{Y}_{\text{pre}}, \mathbf{F}_{T_0}, \Omega) \\ & \propto p(\mathbf{Y}_{\text{pre}} \mid \beta_{T_0}, \mathbf{F}_{T_0}, \Omega) \pi(\beta_{T_0}) \\ & \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\mathbf{Y}_{\text{pre}_i} - \mathbf{X}_{\text{pre}_i} \beta_{T_0})' \Omega^{-1} (\mathbf{Y}_{\text{pre}_i} - \mathbf{X}_{\text{pre}_i} \beta_{T_0}) \right\} \exp \left\{ -\frac{1}{2} \beta_{T_0}' (v_0 \mathbf{I}_k)^{-1} \beta_{T_0} \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[\beta_{T_0}' \left(\sum_{i=1}^N \mathbf{X}_{\text{pre}_i}' \Omega^{-1} \mathbf{X}_{\text{pre}_i} + \frac{1}{v_0} \mathbf{I}_k \right) \beta_{T_0} - 2 \beta_{T_0}' \sum_{i=1}^N \mathbf{X}_{\text{pre}_i}' \Omega^{-1} \mathbf{Y}_{\text{pre}_i} \right] \right\}. \end{aligned}$$

Thus,

$$\beta_{T_0} \mid \mathbf{Y}_{\text{pre}}, \mathbf{F}_{T_0}, \Omega \sim \mathcal{N}(\mathbf{B}_\beta \mathbf{b}_\beta, \mathbf{B}_\beta),$$

where

$$\mathbf{B}_\beta = \left(\sum_{i=1}^N \mathbf{X}_{\text{pre}_i}' \Omega^{-1} \mathbf{X}_{\text{pre}_i} + \frac{1}{v_0} \mathbf{I}_k \right)^{-1}, \quad \mathbf{b}_\beta = \sum_{i=1}^N \mathbf{X}_{\text{pre}_i}' \Omega^{-1} \mathbf{Y}_{\text{pre}_i}.$$

Step 2. Update $\hat{\lambda}_{i,T_0}$. For each $i = 1, \dots, N$, assume

$$\lambda_{i,T_0} \sim \mathcal{N}_r(\mathbf{0}, \xi_0 \mathbf{I}_r).$$

Then

$$\begin{aligned}
& \pi(\boldsymbol{\lambda}_{i,T_0} \mid \mathbf{F}_{T_0}, \boldsymbol{\beta}_{T_0}, \boldsymbol{\Sigma}_{T_0}, \mathbf{Y}_{\text{pre}}) \\
& \propto p(\mathbf{Y}_{\text{pre}_i} \mid \mathbf{F}_{T_0}, \boldsymbol{\lambda}_{i,T_0}, \boldsymbol{\beta}_{T_0}, \boldsymbol{\Sigma}_{T_0}) \pi(\boldsymbol{\lambda}_{i,T_0}) \\
& \propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\lambda}'_{i,T_0} \left(\mathbf{F}'_{T_0} \boldsymbol{\Sigma}_{T_0}^{-1} \mathbf{F}_{T_0} + \frac{1}{\xi_0} \mathbf{I}_r \right) \boldsymbol{\lambda}_{i,T_0} - 2 \boldsymbol{\lambda}'_{i,T_0} \mathbf{F}'_{T_0} \boldsymbol{\Sigma}_{T_0}^{-1} (\mathbf{Y}_{\text{pre}_i} - \mathbf{X}_{\text{pre}_i} \boldsymbol{\beta}_{T_0}) \right] \right\}.
\end{aligned}$$

Hence,

$$\boldsymbol{\lambda}_{i,T_0} \mid \mathbf{F}_{T_0}, \boldsymbol{\beta}_{T_0}, \boldsymbol{\Sigma}_{T_0}, \mathbf{Y}_{\text{pre}} \sim \mathcal{N}(\mathbf{B}_\lambda \mathbf{b}_{\lambda,i}, \mathbf{B}_\lambda),$$

where

$$\mathbf{B}_\lambda = \left(\mathbf{F}'_{T_0} \boldsymbol{\Sigma}_{T_0}^{-1} \mathbf{F}_{T_0} + \frac{1}{\xi_0} \mathbf{I}_r \right)^{-1}, \quad \mathbf{b}_{\lambda,i} = \mathbf{F}'_{T_0} \boldsymbol{\Sigma}_{T_0}^{-1} (\hat{\mathbf{Y}}_i - \hat{\mathbf{X}}_i \boldsymbol{\beta}_{T_0}).$$

Step 3. Update for $\hat{\mathbf{f}}_{t,T_0}$. For each $t = 1, \dots, T_0$, assume

$$\mathbf{f}_{t,T_0} \sim \mathcal{N}_r(\mathbf{0}, C_0 \mathbf{I}_r).$$

Using

$$\tilde{\mathbf{y}}_t = \tilde{\mathbf{X}}_t \boldsymbol{\beta}_{T_0} + \boldsymbol{\Lambda}_{T_0} \mathbf{f}_{t,T_0} + \tilde{\mathbf{e}}_t,$$

we have

$$\begin{aligned}
& \pi(\mathbf{f}_{t,T_0} \mid \boldsymbol{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \sigma_{T_0}^2) \\
& \propto p(\tilde{\mathbf{y}}_t \mid \mathbf{f}_{t,T_0}, \boldsymbol{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \sigma_{T_0}^2) \pi(\mathbf{f}_{t,T_0}) \\
& \propto \exp \left\{ -\frac{1}{2} \left[\mathbf{f}'_{t,T_0} \left(\sigma_{T_0}^{-2} \boldsymbol{\Lambda}'_{T_0} \boldsymbol{\Lambda}_{T_0} + C_0^{-1} \mathbf{I}_r \right) \mathbf{f}_{t,T_0} - 2 \mathbf{f}'_{t,T_0} \sigma_{T_0}^{-2} \boldsymbol{\Lambda}'_{T_0} (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}_{T_0}) \right] \right\}.
\end{aligned}$$

Thus,

$$\mathbf{f}_{t,T_0} \mid \boldsymbol{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \sigma_{T_0}^2 \sim \mathcal{N}(\mathbf{B}_f \mathbf{b}_{f,t}, \mathbf{B}_f),$$

where

$$\mathbf{B}_f = \left(\sigma_{T_0}^{-2} \boldsymbol{\Lambda}'_{T_0} \boldsymbol{\Lambda}_{T_0} + C_0^{-1} \mathbf{I}_r \right)^{-1}, \quad \mathbf{b}_{f,t} = \sigma_{T_0}^{-2} \boldsymbol{\Lambda}'_{T_0} (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}_{T_0}).$$

Step 4. Update $\hat{\sigma}_{T_0}^2$. Assume

$$\sigma_{T_0}^2 \sim \text{IG}\left(\frac{a_0}{2}, \frac{b_0}{2}\right),$$

then

$$\begin{aligned} & \pi(\sigma_{T_0}^2 \mid \mathbf{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \mathbf{F}_{T_0}) \\ & \propto \prod_{t=1}^{T_0} p(\tilde{\mathbf{y}}_t \mid \mathbf{f}_{t,T_0}, \sigma_{T_0}^2, \mathbf{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}) \pi(\sigma_{T_0}^2) \\ & \propto (\sigma_{T_0}^2)^{-(T_0 N + a_0)/2 - 1} \\ & \quad \times \exp \left\{ -\frac{1}{2\sigma_{T_0}^2} \left[\sum_{t=1}^{T_0} \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}_{T_0} - \mathbf{\Lambda}_{T_0} \mathbf{f}_{t,T_0} \right)' \right. \right. \\ & \quad \left. \left. \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}_{T_0} - \mathbf{\Lambda}_{T_0} \mathbf{f}_{t,T_0} \right) + b_0 \right] \right\}. \end{aligned}$$

Therefore,

$$\sigma_{T_0}^2 \mid \mathbf{\Lambda}_{T_0}, \boldsymbol{\beta}_{T_0}, \mathbf{F}_{T_0} \sim \text{IG}\left(\frac{a_0 + T_0 N}{2}, \frac{S_{T_0} + b_0}{2}\right),$$

where

$$S_{T_0} = \sum_{t=1}^{T_0} \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}_{T_0} - \mathbf{\Lambda}_{T_0} \mathbf{f}_{t,T_0} \right)' \left(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}_{T_0} - \mathbf{\Lambda}_{T_0} \mathbf{f}_{t,T_0} \right).$$

Finally, we let

$$\begin{aligned} \tilde{f}(\mathbf{Y}_{\text{pre}} \mid r) = \exp & \left\{ \log f(\mathbf{Y}_{\text{pre}} \mid \hat{\mathbf{F}}_{T_0}, \hat{\mathbf{\Lambda}}_{T_0}, \hat{\boldsymbol{\beta}}_{T_0}, \hat{\boldsymbol{\Sigma}}_{T_0}, r) \right. \\ & \left. - \frac{1}{2} [(T_0 - r)r + Nr + k + 1] \log(N) \right\}, \end{aligned}$$

and

$$\tilde{\pi}(r \mid \mathbf{Y}_{\text{pre}}) \propto \tilde{f}(\mathbf{Y}_{\text{pre}} \mid r) \pi(r)$$

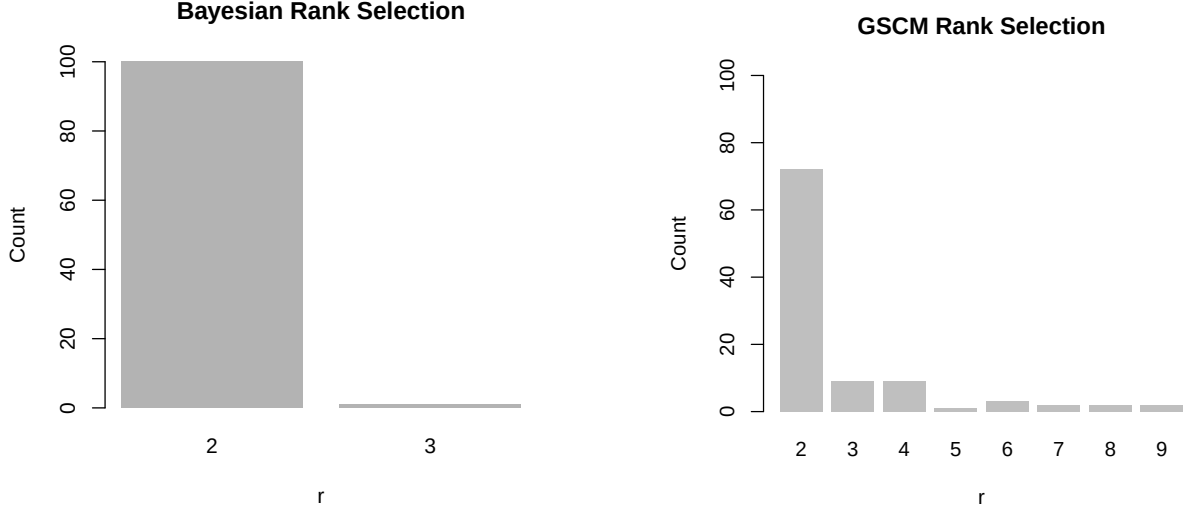


Figure 4.1: Bar plots of the selected factor rank for the Bayesian GSCM (left) and GSCM (right) for the 100 simulated datasets described in Section 2.3.

As a result, the rank of Bayesian generalized synthetic control method is defined as

$$\hat{r} = \operatorname{argmax}_r \tilde{\pi}(r \mid \mathbf{Y}_{\text{pre}}),$$

which can be equivalent to deriving the following

$$\hat{r} = \operatorname{argmax}_r \tilde{f}(\mathbf{Y}_{\text{pre}} \mid r)\pi(r),$$

with $\pi(r)$ is assumed uniform distribution.

4.4 Simulation study

To evaluate the performance of the proposed rank selection method, we conduct simulation studies under different data-generating processes. The bar plots shown in Figure 4.1 are obtained from the simulations in Chapter 2, while those in Figure 4.2 are from Chapter 3.

In Figure 4.1, we display the results obtained by applying the GSCM and Bayes GSCM on a simulated dataset, where the true factor rank is set to $r = 2$. The main objective of this

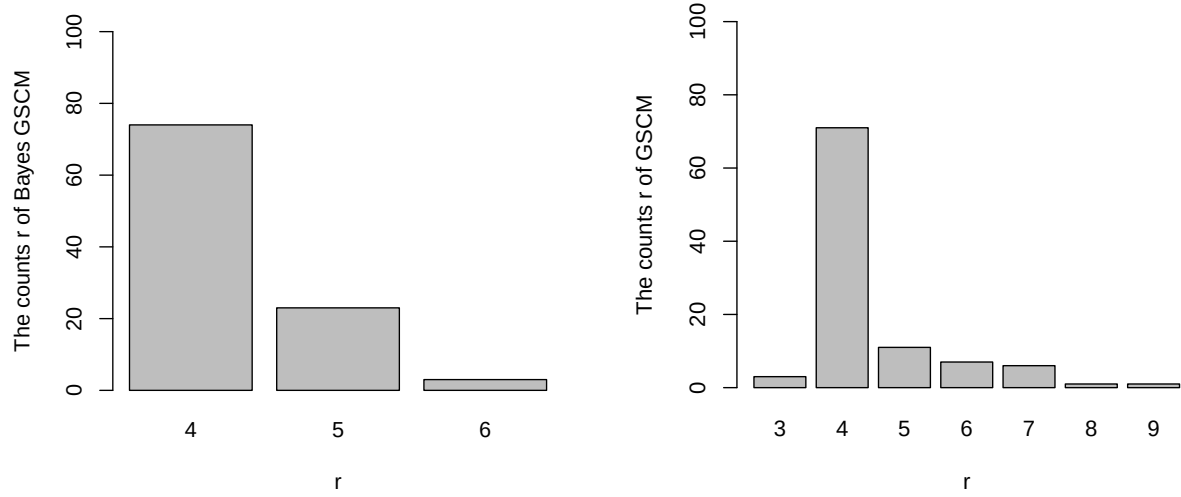


Figure 4.2: Bar plots of the rank for the Bayesian GSCM and GSCM for the 100 simulated datasets described in Section 3.3.

analysis is to evaluate the strength and reliability of the model rank, which is a crucial aspect of model performance. The bar plots represent the rank values recorded over 100 iterations for each approach. It is easy to see the rank of the Bayes GSCM remains consistent and stable, since nearly all runs consistently selected a rank of 2. This suggests that the Bayes GSCM consistently and accurately identifies the actual ranking of the dataset. Conversely, the rank of the GSCM model shows instability, displaying a diverse range of values ranging from 2 to 9 in various iterations.

Figure 4.2 shows the outcomes of applying the GSCM and Bayesian GSCM to a simulated data set with true rank 4. Two bar plots show the rank values obtained across 100 runs of each method. The left panel shows that the Bayesian GSCM consistently identifies the true rank. Across the 100 simulations, the posterior model probabilities overwhelmingly select $r=4$, with very little dispersion. This trend indicates that the Bayesian method is robust to sampling variability and random noise. In contrast, the rank is unstable for the GSCM model, as it takes on a wide range of values from 2 to 9 across different runs.

4.5 Discussion

This chapter proposed a Bayesian procedure for selecting the number of latent factors in the generalized synthetic control framework. The number of factors plays a crucial role in the interactive fixed effects model, as it determines the dimension of the latent structure used to capture unobserved heterogeneity across units and over time. An inappropriate choice of rank may lead to either underfitting or overfitting, both of which can negatively affect the estimation of treatment effects.

To address this issue, we developed a Bayesian rank selection method based on posterior model probabilities. The proposed approach evaluates candidate ranks through their marginal likelihoods, which are approximated using a Laplace-based criterion. This formulation provides a principled way to balance model fit and model complexity while remaining consistent with the Bayesian framework used for posterior inference.

Simulation results demonstrate that the Bayesian rank selection method consistently identifies the true factor rank and produces more stable results compared with the frequentist GSCM approach. These findings suggest that incorporating rank selection into the Bayesian modeling framework can improve both the reliability and interpretability of generalized synthetic control models. Simulation results from both chapters [2](#) and [3](#) illustrate that Bayesian rank selection is sharply concentrated on the true rank, while GSCM cross-validation is much more dispersed.

Chapter 5

Conclusion

In this dissertation, we have developed the generalized synthetic control method from a Bayesian perspective. Through various simulation studies and real data analyses, we have demonstrated the advantages of our proposed method.

Chapter 2 focuses on developing a fully Bayesian approach to the generalized synthetic control method. The proposed framework models latent factors and counterfactual outcomes within a probabilistic setting, providing coherent uncertainty quantification and improving inference compared to existing approaches.

Chapter 3 expands the Bayesian GSCM by changing from basic counterfactual imputation to explicit treatment effect modeling. To capture the non-linear and dynamic character of policy impacts, we parameterize the treatment effect using a flexible B-spline basis. This enables the model to track complicated temporal dynamics, such as those observed in long-term environmental interventions, while avoiding imposing restrictive functional forms.

In Chapter 4, we developed a Bayesian procedure for selecting the number of latent factors in the generalized synthetic control framework. The proposed method determines the factor rank based on posterior model probabilities, with the marginal likelihood approximated using a Laplace-based criterion. This approach provides a principled way to balance model fit and model complexity while remaining consistent with the Bayesian modeling framework. Simulation results demonstrate that the proposed method can accurately recover the true

factor rank and produce more stable rank estimates compared with existing approaches.

The current Bayesian GSCM approach uses a single rank selection process based on posterior model probabilities. For future study, exploring multiple rank selections using Bayesian Model Averaging (BMA) could provide a more flexible approach. BMA would enable the incorporation of model uncertainty by averaging over a variety of models with variable ranks, which could result in more robust estimates of treatment effects.

In the current model setup, all covariates are included in the analysis. However, not all covariates contribute significantly to the model. To eliminate irrelevant covariates by integrating a variable selection method into the Bayesian framework. Techniques such as Bayesian Lasso, Spike-and-Slab priors, or other shrinkage priors could be explored for variable selection.

The current framework focuses on a single treated unit. A natural extension is to allow multiple treated units, $i \in \mathcal{T}$ with $|\mathcal{T}| > 1$, under a common treatment time. In practice, treatment effects may vary across units due to heterogeneous responses. This can be addressed by modeling τ_{it} within a hierarchical framework, allowing for a shared trend with unit-specific variation.

Bibliography

- Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program. *Journal of the American statistical Association* 105(490), 493–505.
- Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative politics and the synthetic control method. *American Journal of Political Science* 59(2), 495–510.
- Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American economic review* 93(1), 113–132.
- Amjad, M., D. Shah, and D. Shen (2018). Robust synthetic control. *The Journal of Machine Learning Research* 19(1), 802–852.
- Andersson, J. J. (2019). Carbon taxes and co 2 emissions: Sweden as a case study. *American Economic Journal: Economic Policy* 11(4), 1–30.
- Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association* 116(536), 1716–1730.
- Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* 77(4), 1229–1279.
- Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70(1), 191–221.
- Ben-Michael, E., A. Feller, and J. Rothstein (2021). The augmented synthetic control method. *Journal of the American Statistical Association* 116(536), 1789–1803.

- Bonander, C., D. Humphreys, and M. Degli Esposti (2021). Synthetic control methods for the evaluation of single-unit interventions in epidemiology: a tutorial. *American journal of epidemiology* 190(12), 2700–2711.
- Card, D. and A. B. Krueger (1993). Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania.
- Casella, G. and E. I. George (1992). Explaining the gibbs sampler. *The American Statistician* 46(3), 167–174.
- Cattaneo, M., Y. Feng, F. Palomba, and R. Titiunik (2025). scpi: Uncertainty quantification for synthetic control methods. *Journal of Statistical Software* 113, 1–38.
- Cattaneo, M. D., Y. Feng, and R. Titiunik (2021). Prediction intervals for synthetic control methods. *Journal of the American Statistical Association* 116(536), 1865–1880.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research* 1(2), 245–276.
- Chandra, N. K., A. Sarkar, J. F. de Groot, Y. Yuan, and P. Müller (2023). Bayesian nonparametric common atoms regression for generating synthetic controls in clinical trials. *Journal of the American Statistical Association* 118(544), 2301–2314.
- Freire, D. (2018). Evaluating the effect of homicide prevention strategies in são paulo, brazil: A synthetic control approach. *Latin American Research Review* 53(2), 231–249.
- Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of econometrics* 75(1), 121–146.
- Ghosh, J. and D. B. Dunson (2009). Default prior distributions and efficient posterior computation in bayesian factor analysis. *Journal of Computational and Graphical Statistics* 18(2), 306–320.
- Gobillon, L. and T. Magnac (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics* 98(3), 535–551.

- Goh, G. and J. Yu (2022). Synthetic control method with convex hull restrictions: a bayesian maximum a posteriori approach. *The Econometrics Journal* 25(1), 215–232.
- Goin, D. E., T. Benmarhnia, S. Huang, F. Lurmann, A. Mukherjee, R. Morello-Frosch, and A. M. Padula (2025). The camp fire and perinatal health: an example of the generalized synthetic control method to identify susceptible windows of exposure. *American journal of epidemiology* 194(3), 722–729.
- Hirose, K., S. Kawano, S. Konishi, and M. Ichikawa (2011). Bayesian information criterion and selection of the number of factors in factor analysis models. *Journal of Data Science* 9(2), 243–259.
- Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica: Journal of the econometric society* ??(??), 1371–1395.
- Howard, P. H. and D. Sylvan (2015). The economic climate: Establishing expert consensus on the economics of climate change. *Institute for Policy Integrity* ??(??), 438–441.
- Huang, H.-N., Z. Yang, Y. Wang, C.-Q. Ou, and Y. Guan (2022). The use of generalized synthetic control method to evaluate air pollution control measures of g20 hangzhou summit. *Frontiers in Public Health* 10, 1021177.
- Humphreys, L. G. and R. G. Montanelli Jr (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research* 10(2), 193–205.
- Kaiser, H. F. (1991). Coefficient alpha for a principal component and the kaiser-guttman rule. *Psychological reports* 68(3), 855–858.
- Kim, S., C. Lee, and S. Gupta (2020). Bayesian synthetic control methods. *Journal of Marketing Research* 57(5), 831–852.
- Klinenberg, D. (2023). Synthetic control with time varying coefficients a state space approach with bayesian shrinkage. *Journal of Business & Economic Statistics* 41(4), 1065–1076.

- Kreif, N., R. Grieve, D. Hangartner, A. J. Turner, S. Nikolova, and M. Sutton (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics* 25(12), 1514–1528.
- Kurz, C. F., M. Rehm, R. Holle, C. Teuner, M. Laxy, and L. Schwarzkopf (2019). The effect of bariatric surgery on health care costs: A synthetic control approach using bayesian structural time series. *Health economics* 28(11), 1293–1307.
- Ledesma, R. D. and P. Valero-Mora (2019). Determining the number of factors to retain in efa: An easy-to-use computer program for carrying out parallel analysis. *Practical assessment, research, and evaluation* 12(1), 2.
- Letellier, N., M. Hale, K. U. Salim, Y. Ma, F. Rerolle, L. Schwarz, and T. Benmarhnia (2025). Applying a two-stage generalized synthetic control approach to quantify the heterogeneous health effects of extreme weather events: a 2018 large wildfire in california event as a case study. *Environmental Epidemiology* 9(1), e362.
- Liu, J. S. and Y. N. Wu (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association* 94(448), 1264–1274.
- Ma, X., Q. Gao, J. Wang, M. Wang, and C. Zhu (2026). A bayesian synthetic control method via horseshoe priors. *Economic Modelling* ??(??), 107502.
- Mader, S. and T. Rüttenauer (2022). The effects of non-pharmaceutical interventions on covid-19 mortality: A generalized synthetic control approach across 169 countries. *Frontiers in Public Health* 10, 820642.
- McClelland, R. and S. Gault (2017). The synthetic control method as a tool to understand state policy. *Washington, DC: Urban-Brookings Tax Policy Center* ??(??), ??
- Moon, H. R. and M. Weidner (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory* 33(1), 158–195.

- Pang, X., L. Liu, and Y. Xu (2022). A bayesian alternative to synthetic control for comparative case studies. *Political Analysis* 30(2), 269–288.
- Pinkney, S. (2021). An improved and extended bayesian synthetic control. *arXiv preprint arXiv:2103.16244* ??(??), ??
- Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55.
- Sheridan, P., S. McElroy, J. Casey, and T. Benmarhnia (2022). Using the generalized synthetic control method to estimate the impact of extreme weather events on population health. *Epidemiology* 33(6), 788–796.
- Silva, V. H. M. C., F. Z. Mariano, G. L. Campêlo, N. C. d. França, J. M. S. d. França, and M. R. Firmiano (2021). Evaluation of policies to combat poverty in brazil: application of the generalized synthetic control method for fecop in ceará. *Nova Economia* 31, 273–302.
- Smith, A. F. and G. O. Roberts (1993). Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)* 55(1), 3–23.
- Stigler, S. M. (1983). Who discovered bayes’s theorem? *The American Statistician* 37(4a), 290–296.
- Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association* 97(460), 1167–1179.
- Sumner, J., L. Bird, and H. Dobos (2011). Carbon taxes: a review of experience and policy design considerations. *Climate Policy* 11(2), 922–943.
- Tierney, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association* 81(393), 82–86.
- Tol, R. S. J. (2009). The economic effects of climate change. *Journal of economic perspectives* 23(2), 29–51.

- Wang, Q., L. Qiu, and W. Xu (2024). Informal payments and doctor engagement in an online health community: An empirical investigation using generalized synthetic control. *Information Systems Research* 35(2), 706–726.
- Wen, Y., J. Xiao, and J. Peng (2023). The effects of the “zero routine flaring by 2030” initiative: International comparisons based on generalized synthetic control method. *Environmental Impact Assessment Review* 100, 107095.
- Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis* 25(1), 57–76.
- Xu, Y. and Q. Zhou (2025). Bayesian synthetic control with a soft simplex constraint. *arXiv preprint arXiv:2503.06454* ??(??), ??

Appendix A

Calculation of the Normal Linear Factor Model

Consider the factor model

$$\mathbf{y}_i = \boldsymbol{\beta}\mathbf{f}_i + \boldsymbol{\epsilon}_i,$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$ is a p -dimensional observation vector for $i = 1, \dots, n$, $\boldsymbol{\beta}$ is a $p \times k$ matrix of factor loadings, $\mathbf{f}_i$ is the k -dimensional vector of common factors, and $\boldsymbol{\epsilon}_i$ is the vector of specific errors.

Assume

$$\mathbf{y}_i \mid \mathbf{f}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma} \sim N(\boldsymbol{\beta}\mathbf{f}_i, \boldsymbol{\Sigma}),$$

$$\mathbf{f}_i \mid \mathbf{H} \sim N(\mathbf{0}, \mathbf{H}),$$

where

$$\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2), \quad \mathbf{H} = \text{diag}(h_1, \dots, h_k).$$

We derive the marginal distribution of $\mathbf{y}_i$ by integrating out the latent factors:

$$p(\mathbf{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{H}) = \int p(\mathbf{y}_i \mid \mathbf{f}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma})p(\mathbf{f}_i \mid \mathbf{H})d\mathbf{f}_i.$$

Substituting the normal densities gives

$$p(\mathbf{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{H}) \propto \int \exp \left\{ -\frac{1}{2} [(\mathbf{y}_i - \boldsymbol{\beta}\mathbf{f}_i)' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\beta}\mathbf{f}_i) + \mathbf{f}_i' \mathbf{H}^{-1} \mathbf{f}_i] \right\} d\mathbf{f}_i.$$

Expanding the quadratic form yields

$$\propto \int \exp \left\{ -\frac{1}{2} [\mathbf{y}_i' \boldsymbol{\Sigma}^{-1} \mathbf{y}_i - 2\mathbf{f}_i' \boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \mathbf{y}_i + \mathbf{f}_i' (\boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta} + \mathbf{H}^{-1}) \mathbf{f}_i] \right\} d\mathbf{f}_i.$$

Let

$$\mathbf{D} = \boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta} + \mathbf{H}^{-1}.$$

Completing the square in $\mathbf{f}_i$ and integrating the resulting normal density gives

$$p(\mathbf{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{H}) \propto \exp \left\{ -\frac{1}{2} \mathbf{y}_i' (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta} \mathbf{D}^{-1} \boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1}) \mathbf{y}_i \right\}.$$

Using the matrix identity

$$\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \boldsymbol{\beta} \mathbf{D}^{-1} \boldsymbol{\beta}' \boldsymbol{\Sigma}^{-1} = (\boldsymbol{\beta} \mathbf{H} \boldsymbol{\beta}' + \boldsymbol{\Sigma})^{-1},$$

we obtain

$$p(\mathbf{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{H}) \propto \exp \left\{ -\frac{1}{2} \mathbf{y}_i' (\boldsymbol{\beta} \mathbf{H} \boldsymbol{\beta}' + \boldsymbol{\Sigma})^{-1} \mathbf{y}_i \right\}.$$

Therefore,

$$\mathbf{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{H} \sim N(\mathbf{0}, \boldsymbol{\beta} \mathbf{H} \boldsymbol{\beta}' + \boldsymbol{\Sigma}).$$

Appendix B

The full conditional distributions for Chapter 2

We consider the prior $\boldsymbol{\beta} \sim \mathcal{N}_k(\mathbf{0}, v_0 \mathbf{I}_k)$, and the likelihood function derives from equation (2.1). By Bayes' theorem ([Stigler, 1983](#)),

$$\begin{aligned} \pi(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\Omega}) & \propto p(\mathbf{y} \mid \boldsymbol{\Omega}, \boldsymbol{\beta}) \pi(\boldsymbol{\beta}) \\ & \propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} \boldsymbol{\beta}' (v_0 \mathbf{I}_k)^{-1} \boldsymbol{\beta} \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}' \sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i \boldsymbol{\beta} - 2 \boldsymbol{\beta}' \sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Y}_i + \boldsymbol{\beta}' (v_0 \mathbf{I}_k)^{-1} \boldsymbol{\beta} \right] \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta}' \left(\sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right) \boldsymbol{\beta} - 2 \boldsymbol{\beta}' \sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Y}_i \right] \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\beta} - \left(\sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Y}_i \right]' \right. \\ & \quad \left. \left(\sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right) \left[\boldsymbol{\beta} - \left(\sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Y}_i \right] \right\} \end{aligned}$$

Hence,

$$\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\Omega} \sim \mathcal{N} \left(\left(\sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{Y}_i, \left(\sum_{i=1}^N \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right)^{-1} \right),$$

where $\boldsymbol{\Omega} = \mathbf{F}\mathbf{F}' + \boldsymbol{\Sigma}$.

To obtain $\pi(\boldsymbol{\lambda}_i \mid \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{y})$, we assume the prior $\boldsymbol{\lambda}_i \stackrel{iid}{\sim} \mathcal{N}_r(\mathbf{0}, \xi_0 \mathbf{I}_r)$, and the likelihood function come from equation (2.1). By Bayes' theorem,

$$\begin{aligned} & \pi(\boldsymbol{\lambda}_i \mid \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{y}) \\ & \propto p(\mathbf{Y}_i \mid \mathbf{F}, \boldsymbol{\lambda}_i, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \pi(\boldsymbol{\lambda}_i) \\ & \propto \exp \left\{ -\frac{1}{2} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{F} \boldsymbol{\lambda}_i)' \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{F} \boldsymbol{\lambda}_i) - \frac{1}{2\xi_0} \boldsymbol{\lambda}_i' \mathbf{I}_r^{-1} \boldsymbol{\lambda}_i \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\lambda}_i' \left(\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F} + \frac{1}{\xi_0} \mathbf{I}_r \right) \boldsymbol{\lambda}_i - 2 \boldsymbol{\lambda}_i' \mathbf{F}' \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \left[\boldsymbol{\lambda}_i - \left(\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F} + \frac{1}{\xi_0} \mathbf{I}_r \right)^{-1} \mathbf{F}' \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right]' \right. \\ & \quad \left. \left(\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F} + \frac{1}{\xi_0} \mathbf{I}_r \right) \left[\boldsymbol{\lambda}_i - \left(\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F} + \frac{1}{\xi_0} \mathbf{I}_r \right)^{-1} \mathbf{F}' \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] \right\} \end{aligned}$$

Hence,

$$\boldsymbol{\lambda}_i \mid \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{y} \sim \mathcal{N} \left(\left(\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F} + \frac{1}{\xi_0} \mathbf{I}_r \right)^{-1} \mathbf{F}' \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta}), \left(\mathbf{F}' \boldsymbol{\Sigma}^{-1} \mathbf{F} + \frac{1}{\xi_0} \mathbf{I}_r \right)^{-1} \right).$$

For $t = r + 1, \dots, T$, to obtain $\pi(\mathbf{f}_t \mid \boldsymbol{\Lambda}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{y})$, we consider the prior $\mathbf{f}_t \stackrel{iid}{\sim} N(\mathbf{0}, C_0 \mathbf{I}_r)$, and the likelihood function is derived from equation (2.2). By Bayes' theorem,

$$\begin{aligned} & \pi(\mathbf{f}_t \mid \boldsymbol{\Lambda}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{y}) \\ & \propto p(\tilde{\mathbf{Y}}_t \mid \mathbf{f}_t, \sigma^2, \boldsymbol{\Lambda}, \boldsymbol{\beta}) \pi(\mathbf{f}_t) \\ & \propto \exp \left\{ -\frac{1}{2} \left[\mathbf{f}_t - (\sigma^{-2} \boldsymbol{\Lambda}' \boldsymbol{\Lambda} + C_0^{-1} \mathbf{I}_r)^{-1} \sigma^{-2} \boldsymbol{\Lambda}' (\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}) \right]' \right. \\ & \quad \times (\sigma^{-2} \boldsymbol{\Lambda}' \boldsymbol{\Lambda} + C_0^{-1} \mathbf{I}_r) \\ & \quad \left. \times \left[\mathbf{f}_t - (\sigma^{-2} \boldsymbol{\Lambda}' \boldsymbol{\Lambda} + C_0^{-1} \mathbf{I}_r)^{-1} \sigma^{-2} \boldsymbol{\Lambda}' (\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta}) \right] \right\}, \end{aligned}$$

Thus,

$$\mathbf{f}_t \mid \mathbf{\Lambda}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{y} \sim N\left((\sigma^{-2}\mathbf{\Lambda}'\mathbf{\Lambda} + C_0^{-1}\mathbf{I}_r)^{-1}\sigma^{-2}\mathbf{\Lambda}'(\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta}), (\sigma^{-2}\mathbf{\Lambda}'\mathbf{\Lambda} + C_0^{-1}\mathbf{I}_r)^{-1}\right).$$

We consider the prior for σ^2 to be an inverse gamma density function such that $\sigma^2 \stackrel{iid}{\sim} \text{IG}\left(\frac{a_0}{2}, \frac{b_0}{2}\right)$. The likelihood function comes from equation (2.2) and is normally distributed. By Bayes' theorem,

$$\begin{aligned} & \pi(\sigma^2 \mid \mathbf{\Lambda}, \boldsymbol{\beta}, \mathbf{F}, \mathbf{y}) \\ & \propto \prod_{t=1}^T p(\tilde{\mathbf{Y}}_t \mid \mathbf{f}_t, \sigma^2, \mathbf{\Lambda}, \boldsymbol{\beta}) \pi(\sigma^2) \\ & \propto (\sigma^2)^{-TN/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=1}^T (\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta} - \mathbf{\Lambda}\mathbf{f}_t)' (\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta} - \mathbf{\Lambda}\mathbf{f}_t)\right\} \\ & \quad \times (\sigma^2)^{a_0/2+1} \exp\left\{-\frac{b_0/2}{2\sigma^2}\right\} \\ & \propto (\sigma^2)^{-(TN/2+a_0/2+1)} \exp\left\{-\frac{\sum_{t=1}^T (\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta} - \mathbf{\Lambda}\mathbf{f}_t)' (\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta} - \mathbf{\Lambda}\mathbf{f}_t)/2 + b_0/2}{\sigma^2}\right\} \end{aligned}$$

Hence,

$$\sigma^2 \mid \mathbf{\Lambda}, \boldsymbol{\beta}, \mathbf{F}, \mathbf{y} \sim \text{IG}\left(\frac{a_0 + TN}{2}, \frac{\sum_{t=1}^T (\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta} - \mathbf{\Lambda}\mathbf{f}_t)' (\tilde{\mathbf{Y}}_t - \tilde{\mathbf{X}}_t\boldsymbol{\beta} - \mathbf{\Lambda}\mathbf{f}_t) + b_0}{2}\right).$$

Appendix C

The full conditional distributions for Chapter 3

In the following sections we detail the derivation of the full conditionals needed at each step of the posterior inference for the model specified in Chapter 3.

C.1 Step 1. Estimation of β , Λ , $\mathbf{F}^*$, Σ , and Ψ

The Gibbs sampler under the PX-factor model that generates the joint posterior distributions by iterating the following steps until convergence:

1. Generate β from $\pi(\beta \mid \mathbf{Y}, \mathbf{F}^*, \Sigma)$
2. Generate Λ from $\pi(\Lambda \mid \mathbf{Y}, \Psi, \mathbf{F}^*, \Sigma, \beta)$
3. Generate $\mathbf{F}^*$ from $\pi(\mathbf{F}^* \mid \mathbf{Y}, \Lambda, \Psi, \Sigma, \beta)$
4. Generate Σ from $\pi(\Sigma \mid \mathbf{Y}, \Lambda, \Psi, \mathbf{F}^*, \beta)$
5. Generate Ψ from $\pi(\Psi \mid \mathbf{Y}, \Lambda, \mathbf{F}^*, \Sigma, \beta)$

C.1.1 Posterior distribution of β

We assume a prior $\beta \sim \mathcal{N}_k(\mathbf{0}, v_0 \mathbf{I}_k)$. The likelihood is defined by the integrated covariance $\Omega^* = \mathbf{F}^* \mathbf{F}^{*'} + \Sigma$. The likelihood function comes from equation (3.3). By Bayes' Theorem:

$$\begin{aligned} \pi(\beta \mid \mathbf{Y}, \Omega^*) &\propto p(\mathbf{Y} \mid \beta, \Omega^*) \pi(\beta) \\ &\propto \exp \left[-\frac{1}{2} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i \beta - \boldsymbol{\tau}_i \odot \mathbf{D}_i)' (\Omega^*)^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta - \boldsymbol{\tau}_i \odot \mathbf{D}_i) \right] \\ &\quad \times \exp \left[-\frac{1}{2v_0} \beta' \beta \right] \\ &\propto \exp \left\{ -\frac{1}{2} \left[\beta' \left(\sum_{i=1}^N \mathbf{X}_i' (\Omega^*)^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right) \beta \right. \right. \\ &\quad \left. \left. - 2\beta' \sum_{i=1}^N \mathbf{X}_i' (\Omega^*)^{-1} (\mathbf{y}_i - \boldsymbol{\tau}_i \odot \mathbf{D}_i) \right] \right\} \end{aligned}$$

By completing the square for the quadratic form $\beta' \mathbf{A} \beta - 2\beta' \mathbf{b}$, we identify:

$$\begin{aligned} \beta \mid \mathbf{Y}, \Omega^* &\sim \mathcal{N} \left(\left(\sum_{i=1}^N \mathbf{X}_i' (\Omega^*)^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' (\Omega^*)^{-1} (\mathbf{y}_i - \boldsymbol{\tau}_i \odot \mathbf{D}_i), \right. \\ &\quad \left. \left(\sum_{i=1}^N \mathbf{X}_i' (\Omega^*)^{-1} \mathbf{X}_i + \frac{1}{v_0} \mathbf{I}_k \right)^{-1} \right) \end{aligned}$$

C.1.2 Posterior distribution of λ_i

With prior $\lambda_i \sim \mathcal{N}_r(\mathbf{0}, \Psi)$, the likelihood function comes from equation (3.3). We isolate the terms containing λ_i the exponent:

$$\begin{aligned}
\pi(\lambda_i \mid \dots) &\propto p(\mathbf{y}_i \mid \mathbf{F}^*, \lambda_i, \beta, \Sigma) \pi(\lambda_i) \\
&\propto \exp \left[-\frac{1}{2} (\mathbf{y}_i - \mathbf{X}_i \beta - \tau_i \odot \mathbf{D}_i - \mathbf{F}^* \lambda_i)' \Sigma^{-1} \right. \\
&\quad \times (\mathbf{y}_i - \mathbf{X}_i \beta - \tau_i \odot \mathbf{D}_i - \mathbf{F}^* \lambda_i) - \frac{1}{2} \lambda_i' \Psi^{-1} \lambda_i \left. \right] \\
&\propto \exp \left[-\frac{1}{2} \left(\lambda_i' (\mathbf{F}^{*'} \Sigma^{-1} \mathbf{F}^* + \Psi^{-1}) \lambda_i \right. \right. \\
&\quad \left. \left. - 2 \lambda_i' \mathbf{F}^{*'} \Sigma^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta - \tau_i \odot \mathbf{D}_i) \right) \right]
\end{aligned}$$

Hence, the posterior distribution is:

$$\begin{aligned}
\lambda_i \mid \dots \sim \mathcal{N} \left(\left(\mathbf{F}^{*'} \Sigma^{-1} \mathbf{F}^* + \Psi^{-1} \right)^{-1} \mathbf{F}^{*'} \Sigma^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta - \tau_i \odot \mathbf{D}_i), \right. \\
\left. \left(\mathbf{F}^{*'} \Sigma^{-1} \mathbf{F}^* + \Psi^{-1} \right)^{-1} \right)
\end{aligned}$$

C.1.3 Posterior distribution of $\mathbf{f}_t^*$

Assuming $\mathbf{f}_t^* \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_t})$ where $k_t = \min(t, r)$, and the likelihood function comes from equation (3.4):

$$\begin{aligned}
\pi(\mathbf{f}_t^* \mid \dots) &\propto p(\tilde{\mathbf{y}}_t \mid \mathbf{f}_t^*, \sigma_t^2, \boldsymbol{\Lambda}, \boldsymbol{\beta}) \pi(\mathbf{f}_t^*) \\
&\propto \exp \left[-\frac{1}{2} (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\tau}_t \odot \mathbf{D}_t - \boldsymbol{\Lambda}_{k_t} \mathbf{f}_t^*)' \boldsymbol{\Sigma}_t^{-1} \right. \\
&\quad \left. \times (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\tau}_t \odot \mathbf{D}_t - \boldsymbol{\Lambda}_{k_t} \mathbf{f}_t^*) - \frac{1}{2} \mathbf{f}_t^{*'} \mathbf{f}_t^* \right] \\
&\propto \exp \left[-\frac{1}{2} \left(\mathbf{f}_t^{*'} (\sigma_t^{-2} \boldsymbol{\Lambda}_{k_t}' \boldsymbol{\Lambda}_{k_t} + \mathbf{I}_{k_t}) \mathbf{f}_t^* \right. \right. \\
&\quad \left. \left. - 2 \mathbf{f}_t^{*'} \sigma_t^{-2} \boldsymbol{\Lambda}_{k_t}' (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\tau}_t \odot \mathbf{D}_t) \right) \right]
\end{aligned}$$

Hence, the posterior distribution is:

$$\begin{aligned}
\mathbf{f}_t^* \mid \dots &\sim \mathcal{N} \left((\sigma_t^{-2} \boldsymbol{\Lambda}_{k_t}' \boldsymbol{\Lambda}_{k_t} + \mathbf{I}_{k_t})^{-1} \sigma_t^{-2} \boldsymbol{\Lambda}_{k_t}' (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\tau}_t \odot \mathbf{D}_t), \right. \\
&\quad \left. (\sigma_t^{-2} \boldsymbol{\Lambda}_{k_t}' \boldsymbol{\Lambda}_{k_t} + \mathbf{I}_{k_t})^{-1} \right)
\end{aligned}$$

C.1.4 Posterior distribution of σ_t^2

Using the prior $\sigma_t^2 \sim IG(a_\sigma/2, b_\sigma/2)$, and with the likelihood function comes from equation (3.3):

$$\begin{aligned}
\pi(\sigma_t^2 \mid \dots) &\propto (\sigma_t^2)^{-N/2} \exp \left[-\frac{1}{2\sigma_t^2} (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\Lambda} \mathbf{f}_t^* - \boldsymbol{\tau}_t \odot \mathbf{D}_t)'(\dots) \right] \\
&\quad \times (\sigma_t^2)^{-(a_\sigma/2+1)} \exp \left[-\frac{b_\sigma}{2\sigma_t^2} \right] \\
&\propto (\sigma_t^2)^{-\left(\frac{a_\sigma+N}{2}+1\right)} \exp \left[-\frac{1}{2\sigma_t^2} \left(b_\sigma + (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\Lambda} \mathbf{f}_t^* - \boldsymbol{\tau}_t \odot \mathbf{D}_t)' \right. \right. \\
&\quad \left. \left. \times (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\Lambda} \mathbf{f}_t^* - \boldsymbol{\tau}_t \odot \mathbf{D}_t) \right) \right]
\end{aligned}$$

Hence, the posterior distribution is:

$$\sigma_t^2 \mid \dots \sim \text{IG} \left(\frac{a_\sigma + N}{2}, \frac{b_\sigma + (\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\Lambda} \mathbf{f}_t^* - \boldsymbol{\tau}_t \odot \mathbf{D}_t)'(\tilde{\mathbf{y}}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta} - \boldsymbol{\Lambda} \mathbf{f}_t^* - \boldsymbol{\tau}_t \odot \mathbf{D}_t)}{2} \right)$$

C.1.5 Posterior distribution of ψ_l

With prior $\psi_l \sim \text{IG}(a_\psi/2, b_\psi/2)$:

$$\begin{aligned}
\pi(\psi_l \mid \dots) &\propto \prod_{i=1}^N \mathcal{N}(\lambda_{il} \mid 0, \psi_l) \pi(\psi_l) \\
&\propto (\psi_l)^{-N/2} \exp \left[-\frac{\sum_{i=1}^N \lambda_{il}^2}{2\psi_l} \right] (\psi_l)^{-(a_\psi/2+1)} \exp \left[-\frac{b_\psi}{2\psi_l} \right] \\
&\propto (\psi_l)^{-\left(\frac{a_\psi+N}{2}+1\right)} \exp \left[-\frac{b_\psi + \sum_{i=1}^N \lambda_{il}^2}{2\psi_l} \right]
\end{aligned}$$

Hence, the posterior distribution is:

$$\psi_l \mid \dots \sim \text{IG} \left(\frac{a_\psi + N}{2}, \frac{b_\psi + \sum_{i=1}^N \lambda_{il}^2}{2} \right)$$

C.2 Step 3. Derivation of the posterior distribution for the treatment effect

In Step 3 of the Gibbs sampling algorithm, we derive the conditional posterior distribution of the treatment effect vector $\boldsymbol{\tau}_1 \in \mathbb{R}^{T_1}$. This is achieved by combining the Gaussian likelihood of the treated unit with a smooth B-spline prior.

C.2.1 The Full conditional for $\boldsymbol{\tau}_1$

The likelihood for the treated unit during the post-treatment period ($t > T_0$) is given by:

$$\mathbf{y}_{1t} | \boldsymbol{\beta}, \boldsymbol{\lambda}_1, \mathbf{f}_t, \boldsymbol{\tau}_1, \sigma_t^2 \sim \mathcal{N}(\mathbf{X}_{1t}\boldsymbol{\beta} + \boldsymbol{\lambda}'_1 \mathbf{f}_t + \boldsymbol{\tau}_1, \sigma_t^2 \mathbf{I})$$

We impose a B-spline prior on $\boldsymbol{\tau}_1$ to ensure temporal smoothness:

$$\boldsymbol{\tau}_1 | \mathbf{B}, \boldsymbol{\alpha}, \theta^2 \sim \mathcal{N}(\mathbf{B}\boldsymbol{\alpha}, \theta^2 \mathbf{I})$$

The posterior $\pi(\boldsymbol{\tau}_1 | \dots)$ is proportional to the product of the likelihood and the prior:

$$\begin{aligned} \pi(\boldsymbol{\tau}_1 | \dots) &\propto \exp\left(-\frac{1}{2\sigma_t^2} \|\mathbf{y}_{1t} - \mathbf{X}_{1t}\boldsymbol{\beta} - \boldsymbol{\lambda}'_1 \mathbf{f}_t - \boldsymbol{\tau}_1\|^2\right) \exp\left(-\frac{1}{2\theta^2} \|\boldsymbol{\tau}_1 - \mathbf{B}\boldsymbol{\alpha}\|^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma_t^2} \left[(\boldsymbol{\tau}_1 - \hat{\mathbf{y}})'(\boldsymbol{\tau}_1 - \hat{\mathbf{y}}) + \frac{\sigma_t^2}{\theta^2} (\boldsymbol{\tau}_1 - \mathbf{B}\boldsymbol{\alpha})'(\boldsymbol{\tau}_1 - \mathbf{B}\boldsymbol{\alpha})\right]\right) \end{aligned}$$

where $\hat{\mathbf{y}} = \mathbf{y}_{1t} - \mathbf{X}_{1t}\boldsymbol{\beta} - \boldsymbol{\lambda}'_1 \mathbf{f}_t$. By completing the square with respect to $\boldsymbol{\tau}_1$, we obtain:

$$\boldsymbol{\tau}_1 \sim \mathcal{N}\left(\mathbf{V}_\tau \left[\frac{1}{\sigma_t^2} \hat{\mathbf{y}} + \frac{1}{\theta^2} \mathbf{B}\boldsymbol{\alpha}\right], \mathbf{V}_\tau\right), \quad \text{where } \mathbf{V}_\tau = \sigma_t^2 \left(\mathbf{I} + \frac{\sigma_t^2}{\theta^2} \mathbf{I}\right)^{-1}$$

C.2.2 The full conditional for spline coefficients α

Given τ_1 and the smoothness parameter θ^2 , the posterior for the spline coefficients α is:

$$\pi(\alpha|\tau_1, \theta^2) \propto \exp\left(-\frac{1}{2\theta^2}(\tau_1 - \mathbf{B}\alpha)'(\tau_1 - \mathbf{B}\alpha)\right)$$

Applying standard Gaussian linear regression results:

$$\alpha \mid \tau_1, \mathbf{B}, \theta^2 \sim \mathcal{N}\left((\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\tau_1, \theta^2(\mathbf{B}'\mathbf{B})^{-1}\right)$$

C.2.3 The full conditional for smoothing parameter θ^2

Using a conjugate Inverse-Gamma prior $\theta^2 \sim \text{IG}(\frac{a_\theta}{2}, \frac{b_\theta}{2})$, the posterior is updated using the residuals of the spline fit:

$$\begin{aligned} \pi(\theta^2|\tau_1, \alpha) &\propto (\theta^2)^{-\frac{T_1}{2}} \exp\left(-\frac{\|\tau_1 - \mathbf{B}\alpha\|^2}{2\theta^2}\right) (\theta^2)^{-(\frac{a_\theta}{2}+1)} \exp\left(-\frac{b_\theta}{2\theta^2}\right) \\ &\propto (\theta^2)^{-(\frac{T_1+a_\theta}{2}+1)} \exp\left(-\frac{\|\tau_1 - \mathbf{B}\alpha\|^2 + b_\theta}{2\theta^2}\right) \end{aligned}$$

Thus, the smoothing parameter is distributed as:

$$\theta^2 \mid \tau_1, \alpha, \mathbf{B} \sim \text{IG}\left(\frac{T_1 + a_\theta}{2}, \frac{\|\tau_1 - \mathbf{B}\alpha\|^2 + b_\theta}{2}\right)$$

Appendix D

Derivations for parametric treatment effect specifications

In Chapter 3, the B-spline basis expansion is adopted as the primary specification because of its flexibility and ability to capture localized variation in the treatment trajectory. For reference, linear and quadratic parametric specifications are also available. The detailed derivations and full conditional posterior distributions for these polynomial models are provided here.

D.1 Polynomial treatment effect specification

For the treated unit ($i = 1$), suppose the post-treatment effect τ_{1t} follows a polynomial of degree p . The mean function is written as

$$\mu(t) = \mathbf{M}\boldsymbol{\alpha} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \cdots + \alpha_p t^p,$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_p)'$ denotes the vector of polynomial coefficients.

Let $t = T_0 + 1, \dots, T$ index the post-treatment periods, and define $T_{\text{post}} = T - T_0$. The corresponding design matrix $\mathbf{M}$ is

$$\mathbf{M} = \begin{pmatrix} 1 & T_0 + 1 & (T_0 + 1)^2 & \cdots & (T_0 + 1)^p \\ 1 & T_0 + 2 & (T_0 + 2)^2 & \cdots & (T_0 + 2)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & T & T^2 & \cdots & T^p \end{pmatrix}.$$

Model the treatment effect τ_{1t} using the following polynomial functional form for the mean $\mu(t)$:

- **Linear Specification:** We assume the treatment effect follows a first-order polynomial:

$$\mu(t) = \alpha_0 + \alpha_1 t$$

where α_0 and α_1 are the intercept and slope coefficients estimated via the MCMC process.

- **Quadratic Specification:** We assume a second-order polynomial to allow for simple curvature:

$$\mu(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

This model is designed to capture basic acceleration or deceleration in the treatment effect over the post-treatment period.

D.2 Gibbs sampling for the polynomial specification

Assume the hierarchical prior

$$\boldsymbol{\tau}_1 \mid \boldsymbol{\alpha}, \theta^2 \sim \mathcal{N}(\mathbf{M}\boldsymbol{\alpha}, \theta^2 \mathbf{I}), \quad \theta^2 \sim \text{IG}\left(\frac{a_\theta}{2}, \frac{b_\theta}{2}\right).$$

The Gibbs sampler proceeds by iteratively sampling from the following full conditional distributions.

Let

$$\mathbf{r}_{1t} = \mathbf{y}_{1t} - \mathbf{X}_{1t}\boldsymbol{\beta} - \boldsymbol{\lambda}'_1 \mathbf{f}_t$$

denote the residual for the treated unit after adjusting for observed covariates and latent factors. Then the conditional posterior of $\boldsymbol{\tau}_1$ is

$$\boldsymbol{\tau}_1 \sim \mathcal{N} \left(\left(\mathbf{I} + \frac{\sigma_t^2}{\theta^2} \mathbf{I} \right)^{-1} \left(\mathbf{r}_{1t} + \frac{\sigma_t^2}{\theta^2} \mathbf{M} \boldsymbol{\alpha} \right), \sigma_t^2 \left(\mathbf{I} + \frac{\sigma_t^2}{\theta^2} \mathbf{I} \right)^{-1} \right).$$

Given $\boldsymbol{\tau}_1$, the polynomial coefficients follow the standard Bayesian linear regression update:

$$\boldsymbol{\alpha} \sim \mathcal{N} \left((\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}'\boldsymbol{\tau}_1, \theta^2 (\mathbf{M}'\mathbf{M})^{-1} \right).$$

The variance parameter governing deviations from the polynomial mean is updated as

$$\theta^2 \sim \text{IG} \left(\frac{T_{\text{post}} + a_\theta}{2}, \frac{\|\boldsymbol{\tau}_1 - \mathbf{M} \boldsymbol{\alpha}\|^2 + b_\theta}{2} \right).$$