Three Essays on Individual Causal Inference

Author:
Tian, Wei

Publication Date:
2021

DOI:
https://doi.org/10.26190/unsworks/1622

License:
https://creativecommons.org/licenses/by/4.0/
Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/100022 in https://unsworks.unsw.edu.au on 2024-04-16

Three Essays on Individual Causal Inference

Wei Tian

Supervisors: Valentyn Panchenko, Seojeong (Jay) Lee

A thesis in fulfillment of the requirements for
the degree of Doctor of Philosophy

School of Economics

UNSW Business School

August, 2021

1. Thesis Title and Abstract

Thesis Title

Three Essays on Individual Causal Inference

Thesis Abstract

In Chapter 1, we provide conditions for the synthetic control estimator to be asymptotically unbiased when the outcome is nonlinear, and propose a flexible and data-drive n method to choose the synthetic control weights. In the empirical application, we illustrate the method by estimating the impact of the 2019 anti-extradition law amendmen ts bill protests on Hong Kong's economy, and find that the year-long protests reduced real GDP per capita in Hong Kong by 11.27% in the first quarter of 2020, which was l arger in magnitude than the economic decline during the 1997 Asian financial crisis or the 2008 global financial crisis.

In Chapter 2, we generalise the conventional synthetic control method to a multiple-outcome framework, where the time dimension is supplemented with the extra dimensi on of related outcomes. As a result, the synthetic control method can now be used even if only a small number of pretreatment periods are available or if we worry about st ructural breaks over a longer time span. We show that the bound on the bias of the multiple-outcome synthetic control estimator is of a smaller stochastic order than that o f the single-outcome synthetic control estimator, provided that the unit of interest can be closely approximated by the synthetic control in terms of the observed predictors a nd the multiple related outcomes before the treatment. In the empirical application, we illustrate our method by estimating the effects of non-pharmaceutical interventions o n various outcomes in Sweden in the first 3 quarters of 2020. Our results suggest that if Sweden had implemented stricter NPIs like the other European countries by Marc h, then (1) there would have been about 70% fewer cumulative COVID-19 infection cases and deaths by July, and 20% fewer weekly deaths from all causes in early Ma y; (2) temporary absence from work would increase by 76% and total hours worked would decrease by 12% among the employed in the second quarter, but the impact wo uld vanish in the third quarter, and there would be no discernable effect on the employment rate throughout; (3) the volume of retail sales would shrink by 5%-13% from M arch to May, while the other economic outcomes including GDP, import, export, industrial production, and CPI would not be affected.

In Chapter 3, we propose a method based on the interactive fixed effects model to estimate treatment effects at the individual level, which allows both the treatment assign ment and the potential outcomes to be correlated with the unobserved individual characteristics. This method is suitable for panel datasets where multiple related outcome s are observed for a large number of individuals over a small number of time periods. To illustrate our method, we provide an example of estimating the effect of health ins urance coverage on individual usage of hospital emergency departments using the Oregon Health Insurance Experiment data.

2. Originality, Copyright and Authenticity Statements

ORIGINALITY STATEMENT

☑ I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

COPYRIGHT STATEMENT

☑ I hereby grant the University of New South Wales or its agents a non-exclusive licence to archive and to make available (including to members of the public) my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known. I acknowledge that I retain all intellectual property rights which subsist in my thesis or dissertation, such as copyright and patent rights, subject to applicable law. I also retain the right to use all or part of my thesis or dissertation in future works (such as articles or books).

For any substantial portions of copyright material used in this thesis, written permission for use has been obtained, or the copyright material is removed from the final public version of the thesis.

AUTHENTICITY STATEMENT

☑ I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.

3. Inclusion of Publications Statement

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in the candidate's thesis in lieu of a Chapter provided:

- The candidate contributed **greater than 50%** of the content in the publication and are the "primary author", i.e. they were responsible primarily for the planning, execution and preparation of the work for publication.
- The candidate has obtained approval to include the publication in their thesis in lieu of a Chapter from their Supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis.

☑ The candidate has declared that **their thesis contains no publications, either published or submitted for publication**.

Candidate's Declaration

I declare that I have complied with the Thesis Examination Procedure.

Acknowledgments

This journey has been challenging, but rewarding. Not only have I learnt more about academic research, but also about life and about myself.

I would like to express my sincere gratitude to my supervisors, Valentyn Panchenko and Seojeong (Jay) Lee, for their guidance and support during my PhD study. They have been very patient and understanding, and are willing to share their precious time for discussions. They are my role models not only for their excellent research, but also for their dedication and for being supportive of others.

I would also like to thank Paul Patterson for being a mentor, Artem Prokhorov for his encouragement, all faculty members for their help, my fellow PhD students for companionship and discussions, and my friends in Sydney and back in China for the warm friendship.

I acknowledge the support of the Australian Government Research Training Program Scholarship, and that this research includes computations using the computational cluster Katana supported by Research Technology Services at UNSW Sydney.

Above all, I would like to thank my family. I thank my parents for their unconditional support. I thank my wife, Fan, for her love and understanding, and for being a life-long friend. Her sacrifice to the family dwarfs mine. Finally, I thank my daughter, Alice, for bringing so much joy to the family.

Contents

Thesis Title and Abstract Sheet i

Originality, Copyright and Authenticity Statements i

Inclusion of Publications Statement i

Acknowledgments ii

Contents v

List of Figures vi

List of Tables viii

Introduction 1

1 The Synthetic Control Method with Nonlinear Outcomes: Estimating the Impact of the 2019 Anti-Extradition Law Amendments Bill Protests on Hong Kong's Economy 4

1.1 Introduction 4

1.2 The Synthetic Control Method 7

1.2.1 Overview 7

1.2.2 Discussion 10

1.3 The Synthetic Control Method with Nonlinearity 14

1.4 Monte Carlo Simulations 18

1.5 Empirical Applications 22

1.5.1 California's Tobacco Control Program 22

1.5.2 1990 German Reunification 25

1.5.3 The Economic Impact of the 2019 Hong Kong Protests 27

1.6 Conclusion 35

Appendix A 36

A.1 Data Sources 36
A.2 Proofs 36

2 The Synthetic Control Method with Multiple Outcomes: Estimating the Effects of Non-Pharmaceutical Interventions in the COVID-19 Pandemic 40
2.1 Introduction 40
2.2 Theoretical Framework 44
2.2.1 Multiple outcomes framework 45
2.2.2 Empirical illustration 48
2.2.3 Adjusting for differences in levels 50
2.2.4 Treatment effect on the untreated 52
2.3 Monte Carlo Simulations 53
2.4 Empirical Application 56
2.4.1 Data 57
2.4.2 Estimation 61
2.4.3 Inference 65
2.4.4 Robustness Checks 68
2.4.5 Further Analysis 73
2.5 Conclusion 75

Appendix B 77
B.1 Proofs 77
B.2 Additional Results 81

3 Individual Causal Inference Using Panel Data With Multiple Outcomes 96
3.1 Introduction 96
3.2 Theory 99
3.2.1 Set Up 99
3.2.2 Estimation 103
3.2.3 Model Selection 105
3.2.4 Related methods 106
3.2.5 Inference 109
3.3 Monte Carlo Simulations 110
3.4 Empirical Application 117
3.5 Conclusion 125

Appendix C 126
C.1 Proofs 126

Conclusion **129**

Bibliography **131**

List of Figures

1.1 Example . . . 11
1.2 Simulated Example . . . 13
1.3 Simulated Example . . . 20
1.4 States Used for Constructing the Synthetic California . . . 24
1.5 Revisiting the Example of the California Tobacco Law Program . . . 24
1.6 Countries Used for Constructing the Synthetic West Germany . . . 26
1.7 Revisiting the Example of German Reunification . . . 26
1.8 Trajectories of GDP per capita . . . 29
1.9 Economies Used for Constructing the Synthetic Hong Kong . . . 31
1.10 Effect of Anti-Extradition Law Protests in Hong Kong . . . 31
1.11 Backdating the Treatment to 2017 . . . 33
1.12 Additional Robustness Checks . . . 34

2.1 Re-analysis of the Economic Cost of the 1990 German Reunification 50
2.2 A Simulated Example . . . 54
2.3 Stringency Index . . . 57
2.4 Descriptive Graphs . . . 60
2.5 Outcomes for Sweden (solid line) and the Synthetic Sweden (dashed line) . . . 63
2.6 Aggregate Treatment Effects . . . 65
2.7 P-Values . . . 67
2.8 Aggregate P-Values . . . 68
2.9 Radar Chart of Outcomes (Benchmark) . . . 69
2.10 Radar Chart of Outcomes (Backdating) . . . 70
2.11 Radar Chart of Outcomes (Leave-One-Unit-Out) . . . 71
2.12 Radar Chart of Outcomes (Leave-One-Outcome-Out) . . . 72
2.13 Radar Chart of Outcomes (Single Outcome) . . . 72
2.14 Radar Chart of Outcomes (No Demeaning) . . . 73
2.15 Deaths by Gender and Age . . . 73
2.16 Reasons of Absence From Work . . . 74

2.17 Retail Sales 74

B.2.1 Placebo Gaps in Outcomes 82
B.2.2 Posttreatment/Pretreatment RMSPE Ratios 83
B.2.3 Aggregate Posttreatment/Pretreatment RMSPE Ratios 84
B.2.4 Backdating the Treatment 85
B.2.5 Leave-One-Unit-Out Distribution of the Synthetic Control for Sweden 86
B.2.6 Leave-One-Outcome-Out Distribution of the Synthetic Control for Sweden 87
B.2.7 Single Outcomes 88
B.2.8 No Demeaning 89
B.2.9 Deaths by Gender and Age 90
B.2.10 Reasons of Absence From Work 91
B.2.11 Retail 91
B.2.12 Labour Market Outcomes by Gender 92
B.2.13 Employment by Occupation 93
B.2.14 Employment by Education 93
B.2.15 Employment by Age 93
B.2.16 GDP 94
B.2.17 CPI 95

3.1 Distribution of the estimated individual treatment effects 119

List of Tables

1.1 Monte Carlo Results 21
1.2 Comparison of Synthetic Control Weights 23
1.3 Comparison of Synthetic Control Weights 25
1.4 Synthetic Control Weights 30

2.1 Balance on Economic Outcomes in 1989 49
2.2 Simulation 55
2.3 Outcomes 59
2.4 Synthetic Control Weights 62

3.1 Simulation Results on Model Selection 112
3.2 Simulation Results for the GMM estimator 113
3.3 Simulation 114
3.4 Simulation 115
3.5 Simulation 116
3.6 Sample Selection 118
3.7 Comparison of Characteristics 122
3.8 Comparison of Characteristics 123
3.9 Comparison of Characteristics 124

Introduction

This thesis consists of three self-contained essays on individual causal inference using panel data. The first two chapters are based on the synthetic control method (Abadie et al., 2010), which is a popular method for estimating the effect of a treatment (e.g., policy, event, etc.) on a single aggregate unit (e.g., country, city, etc.) in a setting where the outcome of interest is observed for a few units over many time periods (small N, large T). The third chapter proposes a method for estimating the individual treatment effects in a large N and small T setting.

The synthetic control estimator (Abadie et al., 2010) is asymptotically unbiased assuming that the outcome is a linear function of the underlying predictors and that the treated unit can be well approximated by the synthetic control before the treatment. When the outcome is nonlinear, the bias of the synthetic control estimator can be severe. In Chapter 1, we provide conditions for the synthetic control estimator to be asymptotically unbiased when the outcome is nonlinear, and propose a flexible and data-driven method to choose the synthetic control weights. Monte Carlo simulations show that compared with the competing methods, the nonlinear synthetic control method has similar or better performance when the outcome is linear, and better performance when the outcome is nonlinear, and that the confidence intervals have good coverage probabilities across settings. In the empirical application, we illustrate the method by estimating the impact of the 2019 anti-extradition law amendments bill protests on Hong Kong's economy, and find that the year-long protests reduced real GDP per capita in Hong Kong by 11.27% in the first quarter of 2020, which was larger in magnitude than the economic decline during the 1997 Asian financial crisis or the 2008 global financial crisis.

In Chapter 2, we generalise the conventional synthetic control method to a multiple-outcome framework, where the time dimension is supplemented with the extra dimension of related outcomes. As a result, the synthetic control method can now be used even if only a small number of pretreatment periods are available or if we worry about structural breaks over a longer time span. We show that the bound on the bias of the multiple-outcome synthetic control estimator is of a smaller stochastic

order than that of the single-outcome synthetic control estimator, provided that the unit of interest can be closely approximated by the synthetic control in terms of the observed predictors and the multiple related outcomes before the treatment. In the empirical application, we illustrate our method by estimating the effects of non-pharmaceutical interventions (NPIs) on various public health, labour market and economic outcomes in Sweden in the first 3 quarters of 2020. Our results suggest that if Sweden had implemented stricter NPIs like the other European countries by March, then (1) there would have been about 70% fewer cumulative COVID-19 infection cases and deaths by July, and 20% fewer weekly deaths from all causes in early May; (2) temporary absence from work would increase by 76% and total hours worked would decrease by 12% among the employed in the second quarter, but the impact would vanish in the third quarter, and there would be no discernible effect on the employment rate throughout; (3) the volume of retail sales would shrink by 5%-13% from March to May, while the other economic outcomes including GDP, import, export, industrial production, and CPI would not be affected.

Chapter 3 changes focus to individual causal inference in the context of empirical microeconomics. Policy evaluation in empirical microeconomics has been focusing on estimating the average treatment effect and more recently the heterogeneous treatment effects, often relying on the unconfoundedness assumption. We propose a method based on the interactive fixed effects model to estimate treatment effects at the individual level, which allows both the treatment assignment and the potential outcomes to be correlated with the unobserved individual characteristics. This method is suitable for panel datasets where multiple related outcomes are observed for a large number of individuals over a small number of time periods. Monte Carlo simulations show that our method outperforms related methods. To illustrate our method, we provide an example of estimating the effect of health insurance coverage on individual usage of hospital emergency departments using the Oregon Health Insurance Experiment data. We find heterogeneous treatment effects in the sample. Comparisons between different groups show that the individuals who would have fewer emergency-department visits if covered by health insurance were younger and not in very bad physical conditions. However, their access to primary care were limited due to being in much more disadvantaged positions financially, which made them resort to using the emergency department as the usual place for medical care. Health insurance coverage might have decreased emergency-department use among this group by increasing access to primary care and possibly leading to improved health. In contrast, the individuals who would have more emergency-department visits if covered by health insurance were more likely to be older and in poor health. So even with access to primary care, they still used emergency departments more

often for severe conditions, although sometimes for primary care treatable and non-emergent conditions as well. Health insurance coverage might have increased their emergency-department use by reducing the out-of-pocket cost of the visits.

Chapter 1

The Synthetic Control Method with Nonlinear Outcomes: Estimating the Impact of the 2019 Anti-Extradition Law Amendments Bill Protests on Hong Kong's Economy

1.1 Introduction

The synthetic control method (Abadie et al., 2010) estimates the treatment effect on a treated unit by comparing its outcome with the outcome of the synthetic control, which is constructed using a convex combination of the control units such that the observed predictors and pretreatment outcomes of the synthetic control closely match those of the treated unit. Abadie et al. (2010) show that the bias of the synthetic control estimator is bounded by a function that goes to zero as the number of pretreatment periods increases, provided that the outcome is a linear function of the observed and unobserved predictors, and that the treated unit is well approximated by the synthetic control in the pretreatment periods. When the outcome is nonlinear, however, the bias of the synthetic control estimator can be severe, as a good fit on the observed predictors and pretreatment outcomes between the treated unit and the synthetic control does not necessarily imply a good fit on the unobserved predictors. This paper generalises the synthetic control method to cases where the outcome is a nonlinear function of the predictors.

We first relax the non-negativity restriction on the weights, which is imposed by Abadie et al. (2010) to prevent extrapolation biases. We show using an example that there is no extrapolation bias when the outcome is linear, and when the outcome is nonlinear, the interpolation bias or extrapolation bias tends to be smaller if we construct the synthetic control using control units that are closer to the treated unit, whereas the synthetic control method with the non-negativity restriction does not prefer or implement the use of closer neighbours. The non-negativity restriction also makes it less likely to obtain a good pretreatment fit, especially when the treated unit takes extreme values in the matching variables or when the sample size is small, and thus limits the applicability of the synthetic control method. In cases where the synthetic control method can be used, the non-negativity restriction may distort the size of the permutation test since the post/pretreatment RMSPE ratio in the permutation test is conditional on a good pretreatment fit for the treated unit but unconditional for the others, which would lead to over-rejection of the null hypothesis, as noted in Ferman and Pinto (2017). Although the non-negativity restriction acts as a regularisation method and often ensures the sparsity of the weights, which is an appealing feature in comparative case studies due to the ease of interpretation, this may come at the cost of a larger bias of the estimator. Relaxing the non-negativity restriction allows more flexible regularisation methods to be implemented so that the bias (and potentially the variance) of the estimator is smaller, while the sparsity of the weights can also be achieved when appropriate.

We then move on to provide the conditions for the synthetic control estimator to be asymptotically unbiased when the outcome is nonlinear, to complement the theoretical result for the linear case in Abadie et al. (2010). Furthermore, we show that there is a tradeoff between the aggregate matching discrepancy, i.e., the matching discrepancy between the treated unit and the synthetic control, and the pairwise matching discrepancies, i.e., the matching discrepancies between the treated unit and the control units used for constructing the synthetic control, depending on the degree of nonlinearity. Specifically, when the degree of nonlinearity is low, the bias of the synthetic control estimator tends to be smaller if we construct the synthetic control using more control units, so that the matching discrepancy between the treated unit and the synthetic control is smaller and the weights are more spread out. When the outcome is highly nonlinear, the bias of the synthetic control estimator tends to be smaller if we use only the nearest neighbours, so that the pairwise matching discrepancies are smaller.

To address this trade-off, we propose choosing the weights with elastic net type regularisation, where the L_1 penalty terms are weighted by pairwise matching discrepancies between the treated unit and the control units to penalise using control units

that are farther away from the treated unit, whereas the L_2 penalty term penalises concentrating the weights on a few control units, and the optimal tuning parameters for the penalty terms are selected using cross-validation. This method can be considered as a combination of the methods in Doudchenko and Imbens (2017) and Abadie and L'Hour (2020), where Doudchenko and Imbens (2017) propose choosing the weights with the elastic net regularisation while relaxing the non-negativity restriction and other restrictions on the weights, and Abadie and L'Hour (2020) propose choosing the weights with L_1 penalty terms weighted by pairwise matching discrepancies between the treated unit and the control units while maintaining the non-negativity restriction. The motivation of both these studies is to use regularisation methods to ensure that there is a unique set of weights that minimise the matching discrepancy between the treated unit and the synthetic control when the number of control units is large with no regard to nonlinearity, whereas this paper aims to provide a flexible and data-driven method to obtain a synthetic control estimator that has a smaller bias when the outcome is potentially nonlinear. Monte Carlo simulations comparing the nonlinear synthetic control method and these two methods as well as the original synthetic control method show that the nonlinear synthetic control method has similar performance with the method in Doudchenko and Imbens (2017) and better performance than the other two in linear settings, and has the best performance in nonlinear settings.

In the main empirical application, we estimate the impact of the 2019 anti-extradition law amendments bill protests on Hong Kong's economy. The results suggest that the protests had a detrimental effect on Hong Kong's economy from the second quarter of 2019. The magnitude of the impact grew rapidly and reached its peak in the first quarter of 2020, when real GDP per capita in Hong Kong was 11.27% lower than what it would be if there were no protests. This exceeds the peak-to-trough decline in quarterly real GDP per capita in Hong Kong during the 1997 Asian financial crisis and the 2008 global financial crisis. The effect became insignificant in the second and third quarters of 2020, when almost all economies were severely hit by the COVID-19 pandemic, and was significant again in the fourth quarter, with the quarterly GDP per capita 8.8% lower than its counterfactual level due to the slow recovery of the economy in Hong Kong.

The rest of the paper is organised as follows. Section 1.2 provides an overview of the original synthetic control method and a discussion on the non-negativity restriction. Section 1.3 provides the conditions for the synthetic control estimator to be asymptotically unbiased when the outcome is nonlinear, and proposes the nonlinear synthetic control method for choosing the weights. Section 1.4 conducts Monte Carlo simulations to compare the nonlinear synthetic control method and the

competing methods. Section 1.5 revisits the two applications in Abadie et al. (2010) and Abadie et al. (2015) to illustrate the nonlinear synthetic control method, and estimates the impact of the 2019 anti-extradition law amendments bill protests on Hong Kong's economy. Section 1.6 concludes. Appendix A.1 lists the data sources for the main application. Appendix A.2 collects the proofs.

1.2 The Synthetic Control Method

1.2.1 Overview

This section provides an overview of the original synthetic control method in Abadie et al. (2010). For a more detailed review, see Abadie (2021). Suppose that we observe N units over T time periods. Without loss of generality, we assume that the first unit receives treatment at period $T_0 + 1 \leq T$ and remains treated afterwards, while all the other $J = N - 1$ units are untreated throughout the window of observation. If we denote the indicator for the binary treatment status for unit i at time t as D_{it}, then $D_{it} = 1$ for $i = 1$ and $t > T_0$, and $D_{it} = 0$ otherwise.

The quantity of interest is the treatment effect on the treated unit at time $t > T_0$, which is given by the difference between its treated potential outcome and untreated potential outcome at time t (Rubin, 1974),

$$\tau_{1t} = Y_{1t}^1 - Y_{1t}^0,$$

where Y_{1t}^1 is the outcome that we would observe for unit 1 at time t if unit 1 is treated at the time, and Y_{1t}^0 is the outcome that we would observe otherwise. The observed outcome can be written as $Y_{1t} = D_{1t}Y_{1t}^1 + (1 - D_{1t})\, Y_{1t}^0$. Since we only observe Y_{1t}^1 for $t > T_0$, estimating τ_{1t} requires predicting the untreated potential outcome Y_{1t}^0. We assume the following functional form for the untreated potential outcome.

Assumption 1.1. *The untreated potential outcome for unit i at period t is given by an interactive fixed effects model*

$$Y_{it}^0 = \boldsymbol{X}_i'\boldsymbol{\beta}_t + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_t + \varepsilon_{it}, \tag{1.1}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{\mu}_i$ are the $k \times 1$ and $f \times 1$ vectors of observed and unobserved predictors of Y_{it}^0 with coefficients $\boldsymbol{\beta}_t$ and $\boldsymbol{\lambda}_t$, respectively, and ε_{it} is the individual transitory shock.

Remark 1.1. Alternatively, the interactive fixed effects term $\boldsymbol{\mu}_i'\boldsymbol{\lambda}_t$ can be interpreted as the product of common time factors $\boldsymbol{\lambda}_t$ and individual factor loadings $\boldsymbol{\mu}_i$.

Both interpretations are discussed in more details in Bai (2009).

Remark 1.2. The factor model presented in Abadie et al. (2010) and Abadie (2021) also includes a common time-varying intercept, representing the time trend in the outcome. This can be considered as a special case of (1.1), where the time-varying intercept is one element in $\boldsymbol{\lambda}_t$ with the corresponding element in $\boldsymbol{\mu}_i$ being 1.

The individual transitory shocks are assumed to satisfy the following assumptions.

Assumption 1.2.

1) ε_{it} *are independent across* i *and* t*;*

2) $\mathbb{E}\left(\varepsilon_{it} \mid \boldsymbol{X}_j, \boldsymbol{\mu}_j, D_{js}\right) = 0$ *for all* i*,* j*,* t *and* s*;*

3) $\mathbb{E}|\varepsilon_{it}|^p < \infty$ *for all* i*,* t *and some even integer* $p \geq 2$*.*

Remark 1.3. The first part of Assumption 1.2 assumes that the individual transitory shocks are independent across units and time. In a panel data setting, cross sectional and time serial correlations in the individual transitory shocks are to be expected. Here we are making a simplifying assumption that the cross sectional and time serial correlations are due to the unobserved individual and time fixed effects. Indeed, if we treat $u_{it} = \boldsymbol{\mu}_i'\boldsymbol{\lambda}_t + \varepsilon_{it}$ as the individual transitory shock, then u_{it} are correlated across units and time, while ε_{it} remain independent across units and time. The second part assumes that the individual transitory shocks have zero mean conditional on the observed and unobserved predictors and the treatment status. The third part ensures that the predictors are not dominated by the transitory shocks in determining the outcomes.

To estimate the treatment effect for unit 1 at time $t > T_0$, a synthetic control is constructed as a linear combination of the control units using weights w_j, $j = 2, \ldots, N$ such that

$$\sum_{j=2}^{N} w_j = 1, \qquad \text{(adding-up)}$$

$$w_j \geq 0 \text{ for } j = 2, \ldots, N, \qquad \text{(non-negativity)}$$

$$\sum_{j=2}^{N} w_j \boldsymbol{X}_j = \boldsymbol{X}_1 \text{ and } \sum_{j=2}^{N} w_j Y_{jt} = Y_{1t} \text{ for all } t \leq T_0. \qquad \text{(pretreatment-fit)}$$

Assumption 1.3. *There exists a set of weights* $(w_2^*, \ldots, w_N^*)$ *that satisfy the adding-up, non-negativity and pretreatment-fit restrictions.*

Remark 1.4. $(w_2^*, \ldots, w_N^*)$ are random quantities that depend on the sample. Assumption 1.3 is satisfied if the observed predictors and the pretreatment outcomes

of the treated unit fall inside the convex hull of those for the control units, in which case there is either a unique or infinitely many sets of weights that satisfy the restrictions, because if there exist two different sets of weights that satisfy the restrictions, then any convex combination of them also satisfy the restrictions.[1] To ensure there is a single solution in this case, Abadie and L'Hour (2020) propose a penalised synthetic control method, which adds penalty terms weighted by the pairwise matching discrepancies between the treated unit and the control units, to the problem of minimising the matching discrepancy between the treated unit and the synthetic control. Note that the solution of the penalised optimisation problem may not belong to the solution set of the original optimisation problem. For a method that pick the solution that minimises the pairwise matching discrepancies from the sets of weights that satisfy Assumption 1.3, see the bilevel optimisation estimator in Díaz et al. (2015).

The synthetic control estimator for τ_{1t} is constructed as

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{N} w_j^* Y_{jt}. \tag{1.2}$$

Before proceeding to the main theoretical result of Abadie et al. (2010), we also need the following assumption, which ensures that matching on the observed predictors and pretreatment outcomes implies matching on the unobserved predictors.

Assumption 1.4. *The smallest eigenvalue of* $\frac{1}{T_0}\sum_{t=1}^{T_0} \boldsymbol{\lambda}_t \boldsymbol{\lambda}_t'$ *is bounded from below by some positive number* $\underline{\xi}$.

The following theorem gives the main theoretical result in Abadie et al. (2010), which shows that the bias of the synthetic control estimator goes to zero as the number of pretreatment periods goes to infinity, under the stated assumptions.

Theorem 1.1. *Under Assumptions 1.1, 1.2, 1.3 and 1.4,* $\mathbb{E}\left(\hat{\tau}_{1t} - \tau_{1t}\right) \to 0$ *as* $T_0 \to \infty$.

Remark 1.5. If we construct the synthetic control by matching only on the observed predictors, then the estimation suffers from the omitted variable bias since the unobserved predictors are not included. Theorem 1.1 shows that by matching on the observed predictors and the pretreatment outcomes, the unobserved predictors are implicitly matched as well under the stated assumptions. The intuition is that if the treated unit and the synthetic control have very different underlying predictors,

[1] The treated unit usually does not fall inside the convex hull of the control units in terms of the matching variables, unless the number of control units is much larger than the number of matching variables, as discussed in Section 1.2.2.

then it is unlikely that they would match well on all the pretreatment outcomes as $T_0 \to \infty$ simply due to the random noises.

Remark 1.6. The synthetic control estimator constructed by matching only on the pretreatment outcomes can also be shown to be asymptotically unbiased, albeit with a larger bound on the bias, as shown by Botosaru and Ferman (2019).

In practice, there may not be a set of weights that satisfy the restrictions in Assumption 1.3 exactly, and the weights are chosen as

$$(\widetilde{w}_2, \ldots, \widetilde{w}_N) = \arg \min_{(w_2,\ldots,w_N)} \sum_{m=1}^{K} v_m \left(Z_{1m} - \sum_{j=2}^{N} Z_{jm} w_j \right)^2 \tag{1.3}$$
$$s.t. \ \sum_j w_j = 1 \ \text{ and } \ w_j \geq 0,$$

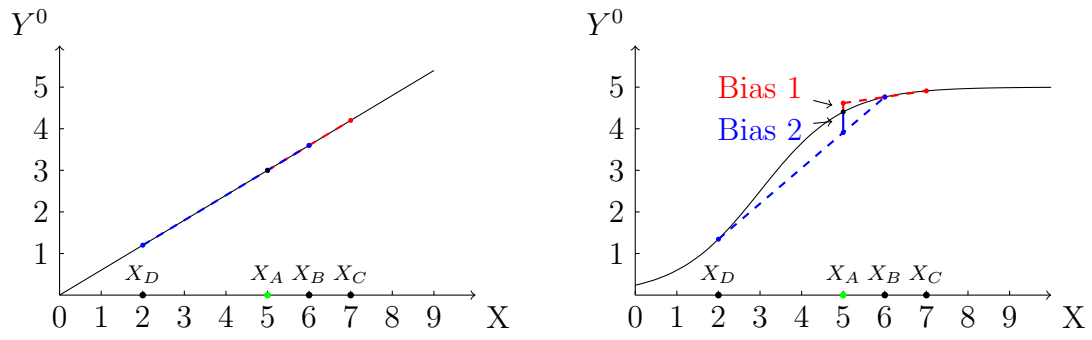
where $Z_{i1}, \ldots, Z_{iK}$ are the K pretreatment variables to match on, and $v_1, \ldots, v_K$ are the weights assigned to these variables, representing the importance of each variable in determining the outcomes. The pretreatment matching variables may include the observed predictors and the pretreatment outcomes, or some linear combinations of them, e.g., the mean of the observed predictors or the pretreatment outcomes across some pretreatment periods. $v_1, \ldots, v_K$ can be chosen by minimising the mean squared prediction errors in the pretreatment periods, with the option of using cross-validation by splitting the pretreatment periods into a training set and a validation set (see Abadie et al., 2015 and Abadie, 2021 for details).

Inference for the synthetic control method is based on the permutation test, where the treatment is recursively reassigned to each of the control units, and a synthetic control is constructed to predict the outcomes for the control unit using all the other units including the treated unit. The ratio between the posttreatment RMSPE (root mean squared prediction error) $\left[\frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \left(Y_{jt} - \hat{Y}_{jt} \right) \right]^{1/2}$ and the pretreatment RMSPE $\left[\frac{1}{T_0} \sum_{t=1}^{T_0} \left(Y_{jt} - \hat{Y}_{jt} \right) \right]^{1/2}$ is obtained for each unit, and the distribution of the post/pretreatment RMSPE ratios is used for inference, where a large ratio for the treated unit relative to the control units is considered evidence that the treatment effect is statistically significant.

1.2.2 Discussion

The non-negativity restriction is imposed by Abadie et al. (2010) to safeguard against extrapolation, which happens if the values of the predictors for the treated

unit fall outside of the convex hull of those of the control units.[2] However, being in the convex hull does not necessarily translate to nonnegative weights for all control units. According to the Carathéodory's theorem, it is possible for the treated unit in the convex hull of the control units to be represented by a linear combination of the control units, where some control units are assigned negative weights.[3] Furthermore, there is no extrapolation bias when the outcome is a linear function of the underlying predictors, whereas in the presence of nonlinearity, it is more important to use control units that are closer to the treated unit to reduce interpolation bias rather than restricting the weights to be non-negative. To illustrate this, we provide two simple examples in Figure 1.1.



(a) Linear outcome

(b) Nonlinear outcome

Figure 1.1: Example

In both examples, we assume a single observed predictor X of the untreated potential outcome Y^0. There is one treated unit A with $X_A = 5$ whose untreated potential outcome is not observed, and which we wish to estimate using the outcomes of the control units. Suppose that there are only two control units B and C with $X_B = 6$ and $X_C = 7$, then we are not able to construct a synthetic control that perfectly matches the treated unit with the non-negativity restriction imposed, even though $X_A = 5$ is just outside the convex hull of $X_B = 6$ and $X_C = 7$. If a third control unit D with $X_D = 1$ is available, we can construct a synthetic control using $X_D = 1$ and $X_B = 6$ under the non-negativity restriction, and compare it with the synthetic control constructed using $X_B = 6$ and $X_C = 7$ without the non-negativity restriction, both of which perfectly match the treated unit. In Figure 1.1a, the untreated potential outcome is a linear function of the predictor given by $Y^0 = 0.6X$. We see that there is no extrapolation bias with or without the non-negativity restriction, as both synthetic controls provide perfect estimates for the

[2]Interpolation happens when the values of the predictors for the treated unit fall in the convex hull.

[3]The Carathéodory's theorem states that if a point $Z \in \mathbb{R}^K$ lies in the convex hull of a set of points P, where $|P| > K + 1$, then Z is in the convex hull of some $K + 1$ points in P. In other words, Z can be expressed as an affine combination of the points in P, where some $K + 1$ points are assigned positive weights, while the other points can receive negative weights.

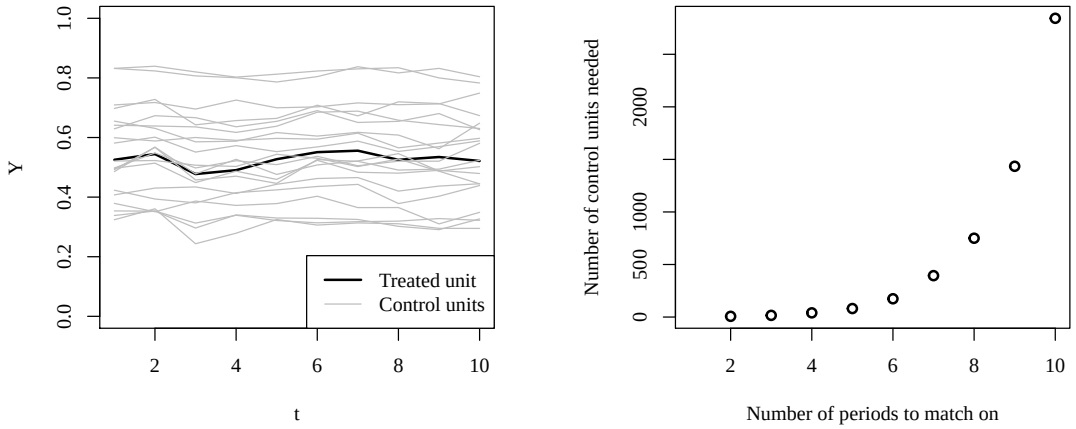
counterfactual outcome of the treated unit $Y_A^0 = 3$. In Figure 1.1b, the untreated potential outcome is a nonlinear function of the predictor given by the S-shaped logistic function $Y^0 = \frac{5}{1+e^{3-X}}$.[4] We see that the magnitude of the extrapolation bias ($|\text{Bias } 1| \approx 0.21$) for the synthetic control estimator constructed without the non-negativity restriction is smaller than that of the interpolation bias ($|\text{Bias } 2| \approx 0.50$) for the synthetic control estimator constructed with the non-negativity restriction, since the former uses closer neighbours. The moral of this example is that in the presence of nonlinearity, the interpolation bias or the extrapolation bias tends to be smaller if we construct the synthetic control using control units that are closer to the treated unit, whereas the original synthetic control method with the non-negativity restriction does not prefer or implement the use of closer neighbours. For example, suppose that there is an additional control unit E with $X_E = 3$, the original synthetic control method does not have a preference on $\frac{1}{3}E + \frac{2}{3}B$ over $\frac{1}{4}D + \frac{3}{4}B$ since both match X_A perfectly with positive weights, even though the bias of the former is smaller in the nonlinear case.

The non-negativity restriction also makes it less likely to obtain a set of weights that satisfy the adding-up, non-negativity and pretreatment-fit restrictions (Assumption 1.3). Note that without the non-negativity restriction, we need $J \geq L = 1+k+T_0$ to be able to find a set of weights that match the synthetic control and the treated unit perfectly, since otherwise the matching variables of the control units do not span $\mathbb{R}^L$. For the weights to satisfy the additional non-negativity restriction, it is likely that J needs to be much larger than L, especially when L is large due to the curse of dimensionality. As an example, we generate 1000 samples, with a single treated unit, 10,000 control units and 10 pretreatment periods in each sample. A typical sample is shown in Figure 1.2a, where the trajectory of the outcome for the treated unit is depicted in black and the trajectories for 20 of the control units are in gray.[5] The levels of the outcome for the units are relatively stable over time, which is similar to what we observe in real data. To make the treated unit even more likely to be in the convex hull of the control units, we calibrate the level of the outcome for the treated unit so that it is at the mean of the outcomes for the control units in period 1. We then use the method provided by King and Zeng (2006), where the convex hull membership check problem is characterised as a linear programming problem, to check whether the treated unit is in the convex hull of the control units. Figure 1.2b records the median sample size required for the treated unit to be in the convex hull of the control units in terms of the corresponding number of pretreatment outcomes. The result suggests that with the non-negativity restriction, the sample size needed for a perfect fit on only a few number of pretreatment periods already exceeds the

[4]The intuition applies to other nonlinear functions that satisfy Assumption 1.6.

[5]For the detailed data generating process, see Section 1.4.

sample sizes usually available for the synthetic control method.



(a) Typical sample (b) Sample size needed

Figure 1.2: Simulated Example

Even if we were to construct a synthetic control that matches the treated unit only approximately, the non-negativity restriction makes a good pretreatment fit less likely, especially when the treated unit takes extreme values in the matching variables or when the sample size is small, which limits the applicability of the synthetic control method.[6] Abadie et al. (2010, 2015) recommend using the synthetic control method only when the treated unit can be closely approximated by the synthetic control. However, since the units near the boundary of the distribution may not be well approximated by the synthetic controls constructed using the other units with the non-negativity restriction, this indicates that the post/pretreatment RMSPE ratio in the permutation test is conditional on a good pretreatment fit for the treated unit but unconditional for the others, which would lead to over-rejection of the null hypothesis, as noted in Ferman and Pinto (2017).

In light of the discussion, we relax the non-negativity restriction so that a good pretreatment fit is more likely to be obtained for the treated unit as well as for the other units when conducting inference. This not only expands the applicability of the synthetic control method, but also helps correct the size distortion for inference. In addition, with the non-negativity restriction lifted, we may be able to obtain a synthetic control estimator with a smaller bias using more flexible regularisation methods than if the restriction were imposed, since the solution space of the weights

[6]In cases where the treated unit can be closely approximated by the synthetic control, the non-negativity restriction often ensures that only a few control units receive positive weights, which is an appealing feature in comparative case studies, since it makes it easier to interpret the contribution of each control unit in the construction of the synthetic control.

in the former is a superset of that in the latter.

1.3 The Synthetic Control Method with Nonlinearity

The synthetic control estimator is shown to be asymptotically unbiased in Abadie et al. (2010), provided that the outcome is a linear function of the underlying predictors and that the pretreatment matching variables of the treated unit can be well approximated by those of the synthetic control. When the outcome is nonlinear, however, the bias of the synthetic control estimator may be severe, since a good fit on the pretreatment matching variables between the treated unit and the synthetic control does not necessarily imply a good fit on the unobserved predictors. In this section, we provide the conditions for the synthetic control estimator to be asymptotically unbiased when the outcome is nonlinear, and propose a flexible and data-driven method for choosing the synthetic control weights in practice.

We start by assuming the following conditions, which are adapted from the assumptions for the matching estimator in Abadie and Imbens (2006).

Assumption 1.5.

1) Let $\boldsymbol{H} = [\boldsymbol{X}' \ \boldsymbol{\mu}']'$ *be a* $(k+f)\times 1$ *random vector of continuous variables, with a version of the density* $c < f(\boldsymbol{H}) < d$ *for some* $c, d > 0$ *on its compact and convex support* $\mathbb{H} \in \mathbb{R}^{k+f}$*;*

2) $\{\boldsymbol{H}_i\}_{i=1}^N$ *are independent draws from the distribution of* $\boldsymbol{H}$*;*

3) For almost every $\boldsymbol{h} \in \mathbb{H}$*,* $Pr(D_{i,T_0+1} = 1 \mid \boldsymbol{H}_i = \boldsymbol{h}) < 1-\rho$ *for some* $\rho \in (0,1)$*.*

Remark 1.7. Assumption 1.5 ensures that the observed and unobserved predictors for all units can be drawn independently from almost any point in the support, and that for almost any values that the predictors of a treated unit take, it is possible to have a control unit whose predictors take those values. This assumption is stronger than Assumption 1.3, which restricts the application of the synthetic control method to samples where a synthetic untreated "twin" for the treated unit can be constructed as a linear combination of the control units. Assumption 1.5 assumes the existence of individual untreated near-identical twins in the population, and that if the random sample is large enough some of those twins will be in the sample. This stronger assumption is needed in the presence of nonlinearity.

The untreated potential outcome is assumed to be linked with the linear latent outcome through a strictly monotonic function, and that its expectation with respect

the individual transitory shock is a smooth function.

Assumption 1.6. *$Y_{it}^0 = F\left(\boldsymbol{X}_i'\boldsymbol{\beta}_t + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_t + \varepsilon_{it}\right)$, where $F(\cdot)$ is a strictly monotonic function. $\mathbb{E}_\varepsilon\left(Y_{it}^0\right) = G\left(\boldsymbol{X}_i'\boldsymbol{\beta}_t + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_t\right)$, where $\mathbb{E}_\varepsilon(\cdot)$ is the expectation conditional on $\boldsymbol{X}_i$ and $\boldsymbol{\mu}_i$, and $G(\cdot)$ is a smooth function.*

Remark 1.8. Assuming that $F(\cdot)$ is a strictly monotonic function excludes binary outcomes since they are usually modelled by discrete choice models like probit or logit, where different values of the latent outcomes can lead to the same observed outcome. Matching on these pretreatment outcomes only implies that the latent outcomes are in the same interval, and there is no guarantee that the unobserved predictors are matched. The unknown functions $F(\cdot)$ and $G(\cdot)$ need not to be estimated as long as the regularity conditions in Assumption 1.6 are satisfied, since our goal is to provide conditions for the synthetic control estimator to be asymptotically unbiased when the outcome has an unknown general nonlinear functional form.

Let $\mathbb{1}(\cdot)$ be the indicator function, $\|\cdot\|$ be the Euclidean norm, $\boldsymbol{Z}_i = [\boldsymbol{X}_i' \;\; Y_{i1} \;\; \cdots \;\; Y_{iT_0}]'$, and $\mathcal{J} = \{2, \ldots, N\}$, then the set of indices for the M closest neighbours of the treated unit in terms of the observed predictors and the pretreatment outcomes can be denoted as

$$\mathcal{J}_M = \left\{ j \in \mathcal{J} \,\middle|\, \sum_{l \in \mathcal{J}} \mathbb{1}\left(\|\boldsymbol{Z}_l - \boldsymbol{Z}_1\| < \|\boldsymbol{Z}_j - \boldsymbol{Z}_1\|\right) < M \right\}, \tag{1.4}$$

i.e., $\|\boldsymbol{Z}_l - \boldsymbol{Z}_1\| > \|\boldsymbol{Z}_j - \boldsymbol{Z}_1\|$ for any $l \in \mathcal{J}\backslash\mathcal{J}_M$ and $j \in \mathcal{J}_M$, where $\backslash$ takes the difference of two sets.

Assumption 1.7. *Only the nearest $M > k + T_0$ neighbours are used for constructing the synthetic control, i.e., $\sum_{j\in\mathcal{J}_M} w_j^* = 1$, $\sum_{j\in\mathcal{J}_M} w_j^* \boldsymbol{X}_j = \boldsymbol{X}_1$ and $\sum_{j\in\mathcal{J}_M} w_j^* Y_{jt} = Y_{1t}$ for all $t \leq T_0$, and $w_j^* = 0$ for all $j \notin \mathcal{J}_M$.*

Using the nearest M neighbours, we can construct the synthetic control estimator as $\tilde{\tau}_{1t} = Y_{1t} - \sum_{j\in\mathcal{J}_M} w_j^* Y_{jt}$. The following theorem provides the conditions for the synthetic control estimator to be asymptotically unbiased when the outcome is nonlinear.

Theorem 1.2. *Under Assumptions 1.2, 1.4, 1.5, 1.6 and 1.7, $\mathbb{E}\left(\tilde{\tau}_{1t} - \tau_{1t}\right) \to 0$ as $T_0 \to \infty$ if $J = O\left(T_0^{b(T_0)}\right)$ with $b(\cdot) \geq 1$ and $b'(\cdot) > 0$.*

Remark 1.9. Theorem 1.2 states that the synthetic control estimator is asymptotically unbiased if the outcome is a strictly monotonic function with smooth expectation, the synthetic control is constructed using the nearest M neighbours, and that the number of control units increases more than exponentially faster than the

number of pretreatment periods. The idea is that the bias of the synthetic control estimator will not vanish asymptotically when the outcome is an unknown nonlinear function, unless each control unit used for constructing the synthetic control converges to the treated unit in the underlying predictors, which happens if the synthetic control is constructed using the nearest neighbours, and that the number of control units increases much faster than the number of pretreatment periods so that the nearest neighbours become closer and closer to the treated unit. This result does not conflict with the fact that the sample sizes available for the synthetic control method are usually small, as it just provides the conditions for the existence of control units that are similar to the treated unit. The implication of this result in finite samples is that when the outcome is highly nonlinear, the bias of the synthetic control estimator is expected to be small if there exist control units that are similar to the treated unit and that we construct the synthetic control using only those control units.[7]

While Theorem 1.2 provides conditions for the synthetic control estimator to be asymptotically unbiased when the outcome is nonlinear, it is also important to examine the bias in finite samples. From the proof of Theorem 1.2, we have

$$\mathbb{E}_{\varepsilon}\left(\sum_{j} w_j^* Y_{jt} - Y_{1t}^0\right) = \sum_{n=1}^{\infty} \frac{G^{(n)}}{n!} \left\{ \sum_{j} w_j^* \left[\left(\boldsymbol{X}_j - \boldsymbol{X}_1\right)' \boldsymbol{\beta}_t + \left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_1\right)' \boldsymbol{\lambda}_t \right]^n \right\}, \tag{1.5}$$

which represents the weighted average of the distances between the treated unit and the control units in terms of the underlying predictors. Under the assumptions of Theorem 1.2, $\boldsymbol{Z}_j - \boldsymbol{Z}_1 \xrightarrow{p} 0$ implies that $\boldsymbol{\mu}_j - \boldsymbol{\mu}_1 \xrightarrow{p} 0$ for $j \in \mathcal{J}_M$. Therefore, when the outcome is highly nonlinear, the bias of the synthetic control estimator would be smaller if we use fewer and closer neighbours to construct the synthetic control so that the pairwise matching discrepancies in the pretreatment variables between the treated unit and its neighbours are smaller.

When the outcome is linear or if the degree of nonlinearity is low, the bias of the synthetic control estimator reduces to

$$\mathbb{E}_{\varepsilon}\left(\sum_{j} w_j^* Y_{jt} - Y_{1t}^0\right) = \left(\sum_{j} w_j^* \boldsymbol{X}_j - \boldsymbol{X}_1\right)' \boldsymbol{\beta}_t + \left(\sum_{j} w_j^* \boldsymbol{\mu}_j - \boldsymbol{\mu}_1\right)' \boldsymbol{\lambda}_t, \tag{1.6}$$

representing the distance between the treated unit and the synthetic control in the

[7] Note that this is different from the sparsity of the weights under the non-negativity restriction, which is achieved regardless of the degree of nonlinearity, as the few control units that receive positive weights are not necessarily close to the treated unit.

underlying predictors, and is smaller if we construct the synthetic control using more neighbours so that the matching discrepancy between the treated unit and the synthetic control is smaller. The bias is also shown in the proof of Theorem 1.1 to be bounded by a value that increases with $\bar{w} = \max_j |w_j^*|$ given a good pretreatment fit between the treated unit and the synthetic control constructed by J control units, and thus is smaller if the weights are assigned more evenly among the control units. The variance of the synthetic control estimator also tends to be smaller if the weights are more spread out, which is similar to the least square estimator, whose variance becomes smaller if the sample size is larger or if the explanatory variables are more spread out.

This presents a trade-off between the aggregate matching discrepancy and the pairwise matching discrepancies, depending on the degree of nonlinearity of the outcome function, similar to the bias-variance tradeoff in non-parametric methods, e.g., choosing bin width in kernel density estimation. To address this trade-off, we choose the set of weights by solving the following minimisation problem with elastic net type penalties,

$$\begin{aligned} \min_{\{w_j\}_j} & \left\| \boldsymbol{Z}_1 - \sum_j w_j \boldsymbol{Z}_j \right\|^2 + a \sum_j |w_j| \, \|\boldsymbol{Z}_1 - \boldsymbol{Z}_j\| + b \sum_j |w_j|^2 , \\ & \text{s.t.} \sum_j w_j = 1. \end{aligned} \tag{1.7}$$

The L_1 penalty terms are weighted by pairwise matching discrepancies between the treated unit and the control units, and penalise assigning weights to control units that are farther away from the treated unit. The L_2 penalty term penalises concentrating weights on a few control units and controls the scale of $\bar{w}$. The level of penalisation is adjusted through the nonnegative tuning parameters, a and b. When $a = 0$ and $b = 0$, the weights are chosen solely to minimise the aggregate matching discrepancy. When a becomes larger, the weights are more concentrated on control units that are closer to the treated unit, thus achieving sparsity of the weights. As $a \to \infty$, the estimator becomes the nearest neighbours matching estimator using only the nearest neighbour, as noted in Abadie and L'Hour (2020). When b becomes larger, the weights are assigned more evenly among the control units. As $b \to \infty$, all the control units are assigned equal weights and the estimator becomes the difference-in-differences estimator. When both a and b are large, the weights are spread out among a number of control units that are close to the treated unit, and the estimator becomes close to the nearest neighbours matching estimator using multiple neighbours.

Ultimately choosing the optimal tuning parameters is an empirical problem in finite samples, which can be done using cross-validation, as proposed in Doudchenko and Imbens (2017) and Abadie and L'Hour (2020). One way to conduct cross-validation is to predict the posttreatment outcomes for each control unit using the synthetic control constructed from the other control units, and the optimal set of tuning parameters is the one that minimises the mean squared prediction error. This method of conducting cross-validation is used for the Monte Carlo simulations and the applications in this paper. Alternatively, we can predict the outcome of the treated unit in each pretreatment period using the synthetic control constructed from the control units, and select the set of tuning parameters that minimises the mean squared prediction error. To make the selection of the optimal tuning parameters tractable in practice, the tuning parameters that enter the minimisation problem, a and b, are scaled by the nonzero eigenvalues of $\boldsymbol{Z}_0\boldsymbol{Z}_0'$, where $\boldsymbol{Z}_0$ is the $J \times (k+T_0)$ matrix of matching variables of the control units, so that the optimal tuning parameters a^* and b^* can be chosen from $[0,1]$. Specifically, $\boldsymbol{Z}_0\boldsymbol{Z}_0'$ has $n = \min(J,\ k+T_0)$ nonzero eigenvalues, denoted as $\lambda_1, \ldots, \lambda_n$, in ascending order. For $b^* \in [0,1]$, we set $b = b^*\lambda_{\lceil nb^* \rceil}$, where $\lceil \cdot \rceil$ is the ceiling function, so that when $a^* = 0$ and $b^* = 1$, the weights are roughly evenly assigned to the control units. a is similarly scaled by the nonzero eigenvalues of $\boldsymbol{Z}_0\boldsymbol{Z}_0' + \text{diag}(b)$, so that when $a^* = 1$, the weight will only be assigned to the nearest neighbour. To select the optimal tuning parameters, we start with the initial value $b^* = 0$, and then choose a^* from $[0,1]$ with a set grid size, e.g., 0.1, to minimise the mean squared prediction error from either cross-validation construction. Given the selected a^*, we then update b^* by minimising the mean squared prediction error. a^* and b^* are then updated iteratively until convergence.

1.4 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to compare the nonlinear synthetic control estimator (NSC) with the original synthetic control estimator (OSC) from Abadie et al. (2010), the synthetic control estimator with elastic net regularisation (ESC) from Doudchenko and Imbens (2017), and the penalised synthetic control estimator (PSC) from Abadie and L'Hour (2020). For the purpose of comparison, these other methods are modified so that OSC differs from NSC in that it imposes the non-negativity restriction and does not have L_1 and L_2 penalties, ESC differs from NSC in that the L_1 penalty terms are not weighted by pairwise matching discrepancies, and PSC differs from NSC in that it does not have L_2 penalty. The number of treated unit is fixed at 1, and the number of posttreatment period is fixed at 10 across settings. The data generating process is as follows.

First, the latent outcomes are generated from the interactive fixed effects model as

$$Y_{it}^{*} = \boldsymbol{X}_i'\boldsymbol{\beta}_t + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_t + \varepsilon_{it}, \tag{1.8}$$

where the vector of observed predictors $\boldsymbol{X}_i$ has dimension 2, and the vector of unobserved predictors $\boldsymbol{\mu}_i$ has dimension 4. The observed and unobserved predictors are independently and identically drawn from the uniform distribution $U\left[0, 2\sqrt{3}\right]$ for each unit, the coefficients are i.i.d. $N(10, 1)$, and the individual transitory shocks are i.i.d. $N(0, 1)$.

The untreated potential outcomes are then generated as

$$Y_{it}^{0} = \left(\frac{Y_{it}^{*} - Y_{\min}^{*}}{Y_{\max}^{*} - Y_{\min}^{*}}\right)^{r}, \tag{1.9}$$

where $Y_{\min}^{*}$ and $Y_{\max}^{*}$ are the smallest and largest values of Y_{it}^{*} respectively, so that $\frac{Y_{it}^{*} - Y_{\min}^{*}}{Y_{\max}^{*} - Y_{\min}^{*}}$ is between 0 and 1.[8] The degree of nonlinearity is adjusted by $r \in \{1, 2\}$. Y_{it}^{0} is a linear function of the predictors and the individual transitory shock when $r = 1$, and is nonlinear when $r = 2$.[9]

The treatment effects τ_{it} are set to $[0.02, 0.04, \ldots, 0.2]$ in the 10 posttreatment periods, and 0 in the pretreatment periods. And the observed outcomes are generated as

$$Y_{it} = Y_{it}^{0} + D_{it}\tau_{it}, \tag{1.10}$$

where

$$D_{it} = \begin{cases} 1, & \text{if } i = 1 \text{ and } t > T_0, \\ 0, & \text{otherwise.} \end{cases}$$

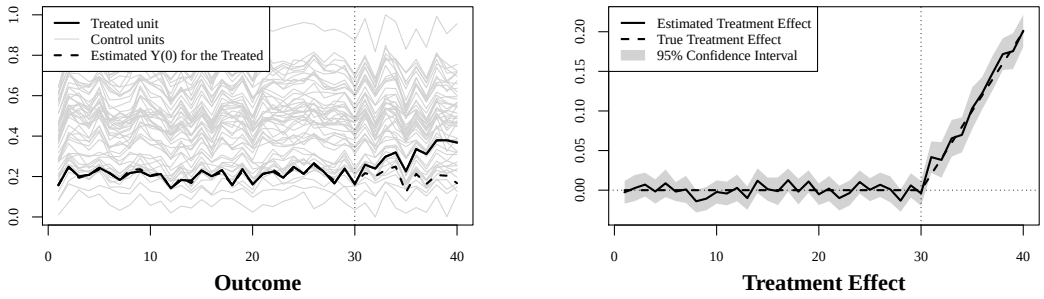
By varying the number of control units $J \in \{25, 50\}$, the number of pretreatment periods $T_0 \in \{15, 30\}$ and the degree of nonlinearity $r \in \{1, 2\}$, we have 8 settings. The observed and unobserved predictors and their coefficients are drawn 20 times for each setting, and the individual transitory shocks are drawn 250 times for each set of $(\boldsymbol{X}_i, \boldsymbol{\beta}_t, \boldsymbol{\lambda}_t, \boldsymbol{\mu}_i)$, so that we generate 5000 samples for each setting.[10]

Figure 1.3 illustrates the estimation of the treatment effects using the nonlinear synthetic control method in a typical sample with $J = 50$, $T_0 = 30$ in the linear

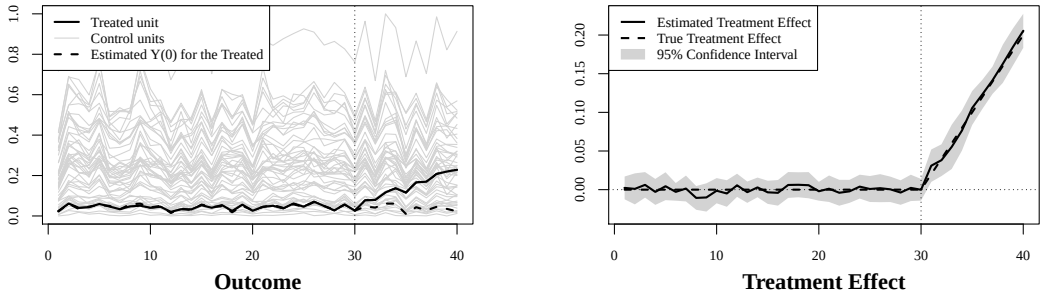
[8] The transformation in (1.9) is random since $Y_{\min}^{*}$ and $Y_{\max}^{*}$ depend on the sample. The purpose of rescaling the outcomes to be within $[0, 1]$ is to make results in different settings more comparable, and the findings do not fundamentally change if we use some non-random transformation.

[9] Using a larger r or adopting other functional forms such as the logistic function considered in Figure 1.1b does not fundamentally change the conclusion.

[10] To save computation time, for each set of $(\boldsymbol{X}_i, \boldsymbol{\beta}_t, \boldsymbol{\lambda}_t, \boldsymbol{\mu}_i)$, the tuning parameters are chosen once from $[0, 1]$ with grid size 0.1 using cross-validation, and are then fixed across the 250 simulations.



(a) Linear outcome



(b) Nonlinear outcome

Figure 1.3: Simulated Example

case ($r = 1$, upper panel) and nonlinear case ($r = 2$, lower panel), respectively. The graphs on the left visualise the trajectories of the outcome for the treated unit (black solid line), the control units (gray solid lines) and the synthetic control (black dashed line). The observed outcomes are more concentrated towards the bottom in the nonlinear case due to the nonlinear transformation. In both examples, we see that the trajectory of the outcome for the synthetic control, which is used to estimate the untreated potential outcome for the treated unit, is able to follow the trajectory of the treated unit closely before the treatment, and diverges after the treatment. The gap between the trajectories of the treated unit and the synthetic control is then used to estimate the treatment effect, as depicted in the graphs on the right, where the black solid line is the estimated treatment effect, the black dashed line is the true treatment effect, and the 95% confidence interval is in gray.[11] We see that the estimated effects are very close to the true treatment effects, and the

[11]To construct the confidence interval, we follow Doudchenko and Imbens (2017) and estimate the variance of the estimator in each period using the mean squared error obtained from predicting the outcome for each control unit in that period using the other control units. The confidence interval is then constructed using the estimated variance, assuming normal distribution for the estimator.

confidence intervals also accurately reveal that the treatment effect is not statistically significantly different from 0 before the treatment, and becomes significant after the treatment.

Table 1.1: Monte Carlo Results

			OSC		ESC		PSC		NSC		
J	T_0	r	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Coverage
Panel A: Linear Outcome											
25	15	1	2.17	1.07	0.99	1.17	1.11	1.32	0.99	1.18	0.936
50	15	1	1.19	0.94	0.77	0.93	0.87	1.06	0.77	0.94	0.944
25	30	1	1.79	1.03	0.90	1.09	0.98	1.19	0.91	1.10	0.936
50	30	1	1.13	0.92	0.75	0.92	0.83	1.02	0.75	0.92	0.944
Panel B: Nonlinear Outcome											
25	15	2	1.99	0.91	0.94	1.02	0.97	1.10	0.92	1.02	0.938
50	15	2	1.11	0.81	0.77	0.84	0.80	0.94	0.74	0.83	0.947
25	30	2	1.40	0.85	0.87	0.96	0.91	1.03	0.87	0.97	0.935
50	30	2	1.01	0.78	0.69	0.80	0.73	0.86	0.68	0.79	0.950

Note: This table compares the bias and SD of the nonlinear SC estimator and three other SC estimators from Abadie et al. (2010), Doudchenko and Imbens (2017) and Abadie and L'Hour (2020) respectively, and reports the coverage probability of the 95% confidence interval produced by the nonlinear SC method, in different settings that vary in the number of control units J, the number of pretreatment periods T_0 and the degree of nonlinearity r, based on 5000 simulations for each setting.

Table 1.1 reports the bias and the standard deviation (SD) for each estimator, as well as the coverage probability of the 95% confidence interval produced by the nonlinear synthetic control estimator in different settings.[12] With a larger J or T_0, the bias and SD become smaller for all the estimators, and the coverage probability of the 95% confidence interval also improves, in both the linear and nonlinear cases. Compared with the original estimator, the other estimators have smaller biases across different settings, showing the advantage of more flexible regularisation methods over the non-negativity restriction in reducing the bias of the estimators, whereas the non-negativity restriction has an edge on keeping the SD low. The bias and SD are on similar levels for ESC and NSC in the linear cases, and are smaller than those of PSC, as ESC and NSC use the L_2 regularisation to control the scale of the weights. The advantage of ESC over PSC becomes smaller in the nonlinear cases, since PSC

[12]The bias is measured using the average of the absolute biases across the posttreatment periods and the 5000 simulations for each setting. The standard deviation is calculated for each posttreatment period and each set of $(\boldsymbol{X}_i, \boldsymbol{\beta}_t, \boldsymbol{\lambda}_t, \boldsymbol{\mu}_i)$ and then averaged over the posttreatment periods and the 20 sets for each setting. Both the bias and the SD are multiplied by 100 for better presentation.

now constructs the synthetic control using closer neighbours so that its bias becomes smaller. NSC has the smallest bias among the estimators in the nonlinear cases, since it employs both the L_1 penalty terms weighted by pairwise matching discrepancies to select closer neighbours, and the L_2 regularisation to control the scale of the weights.

1.5 Empirical Applications

In this section, we first revisit the two empirical applications in Abadie et al. (2010) and Abadie et al. (2015) to illustrate the nonlinear synthetic control method, and compare it with the original synthetic control method. The synthetic controls in both methods are constructed by matching only on the outcomes in all pretreatment periods, since the observed predictors are not essential as long as there is a good fit on the pretreatment outcomes over an extended period of time (Botosaru and Ferman, 2019). We then move on to the main empirical application of this paper, where we estimate the impact of the 2019 anti-extradition law amendments bill protests in Hong Kong on the city's economy using the nonlinear synthetic control method.

1.5.1 California's Tobacco Control Program

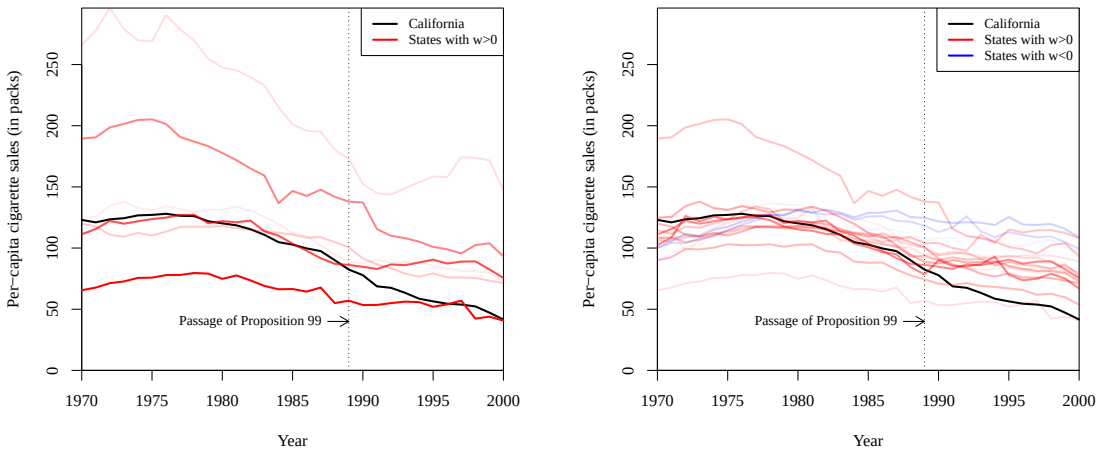
In the first example, we revisit the empirical application in Abadie et al. (2010), who examine the effect of a large-scale tobacco control program implemented in California in 1988 on the annual per-capita cigarette sales.

Table 1.2: Comparison of Synthetic Control Weights

State	OSC Weight	NSC Weight	State	OSC Weight	NSC Weight
Alabama	0	-0.015	Nevada	0.186	0.091
Arkansas	0	-0.057	New Hampshire	0.049	0
Colorado	0.03	0.119	New Mexico	0	0.103
Connecticut	0.08	0.112	North Carolina	0	0
Delaware	0	0	North Dakota	0	0
Georgia	0	0	Ohio	0	0
Idaho	0	0.183	Oklahoma	0	0
Illinois	0	0.02	Pennsylvania	0	0
Indiana	0	0	Rhode Island	0	0
Iowa	0	0.039	South Carolina	0	-0.003
Kansas	0	0	South Dakota	0	0
Kentucky	0	0	Tennessee	0	-0.071
Louisiana	0	0	Texas	0	0
Maine	0	0	Utah	0.385	0.045
Minnesota	0	0.027	Vermont	0	0
Mississippi	0	-0.007	Virginia	0	0
Missouri	0	0	West Virginia	0	0.083
Montana	0.271	0.176	Wisconsin	0	0.06
Nebraska	0	0.094	Wyoming	0	0

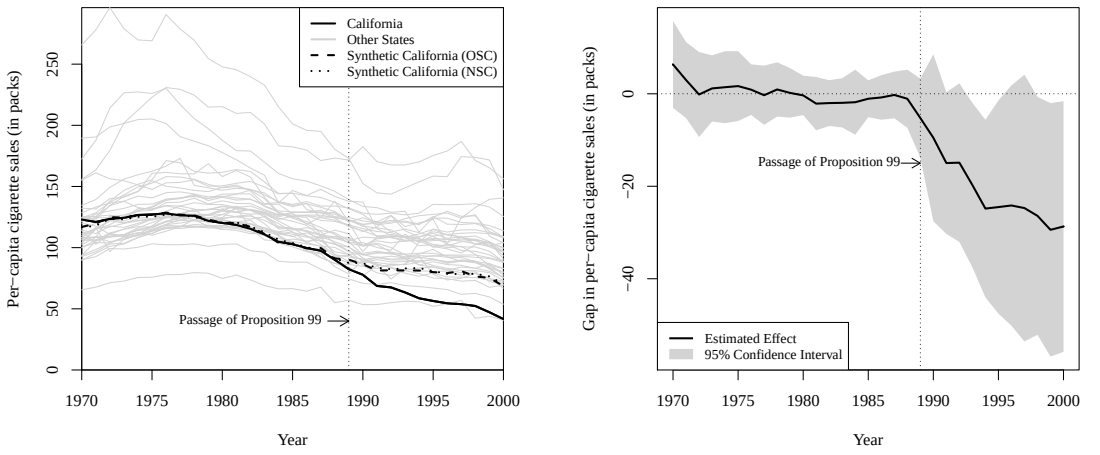
Table 1.2 compares the weights assigned to the other states by the original synthetic control method and the nonlinear synthetic control method ($a^* = 0.3$, $b^* = 0.7$), respectively. We see that the weights assigned by the original synthetic control method are concentrated on Utah (0.385), Montana (0.271) and Nevada (0.186), with relatively minor weights on Connecticut, New Hampshire and Colorado. In comparison, the weights in the nonlinear synthetic control method spread out among more states, but are still sparse.

Figure 1.4 displays the trajectories of the per-capita cigarette sales from 1970 to 2000 for California (black), the states that are assigned positive weights (red), and the states that are assigned negative weights (blue) in the two methods. For comparison, the depth of the colour for the trajectory is scaled by the magnitude of the weight assigned to the state. We see that although the weights in the original synthetic control method are sparse, they may be assigned to control units that are far away from the treated unit, as long as the constructed synthetic control approximates the treated unit well and the weights stay non-negative. In comparison, the nonlinear synthetic control method constructs the synthetic control using control units that are close to the treated unit in the presence of nonlinearity.



(a) Original SC Method (b) Nonlinear SC Method

Figure 1.4: States Used for Constructing the Synthetic California



(a) Trajectories of the Outcome (b) Estimated Treatment Effect

Figure 1.5: Revisiting the Example of the California Tobacco Law Program

Figure 1.5a depicts the trajectories of the per-capita cigarette sales for California (black solid line), the other states (gray solid line), the synthetic California constructed using the original synthetic control method (black dashed line), and the synthetic California constructed using the nonlinear synthetic control method (black dotted line). We see that the trajectories for the synthetic California from both methods are quite similar, both of which closely follow the trajectory for the real California before 1988, and diverge after the passage of the tobacco control legislation in 1988. Figure 1.5b shows the gap between the trajectories for California and the synthetic California in the nonlinear synthetic control method, which is

used to estimate the effect of the tobacco control program on cigarette sales. The result suggests that the tobacco control program reduced per-capita cigarette sales by 9.5 packs in 1990, 24.5 packs in 1995 and 28.7 packs in 2000. The confidence intervals imply that the effect became significant from 1993 onwards (except in 1996 and 1997).

1.5.2 1990 German Reunification

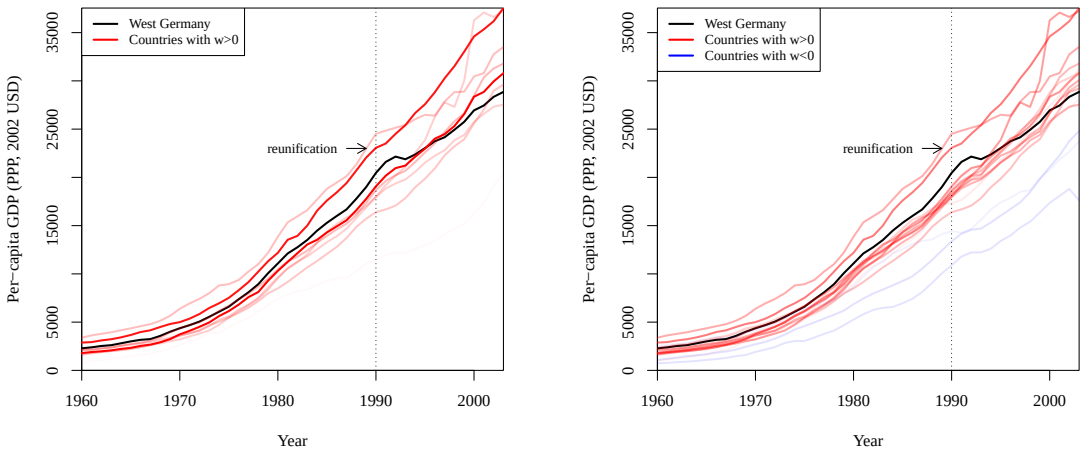
In the second example, we revisit the empirical application in Abadie et al. (2015), which analyse the effect of the 1990 German reunification on West Germany's per-capita GDP.

Table 1.3: Comparison of Synthetic Control Weights

Country	OSC Weight	NSC Weight	Country	OSC Weight	NSC Weight
Australia	0	0.027	Netherlands	0.091	0.087
Austria	0.325	0.134	New Zealand	0	-0.017
Belgium	0	0.101	Norway	0.062	0.123
Denmark	0	0.058	Portugal	0	-0.034
France	0	0.092	Spain	0	-0.037
Greece	0.008	0.003	Switzerland	0.082	0.106
Italy	0.062	0.096	UK	0.072	0.079
Japan	0	0.016	USA	0.299	0.168

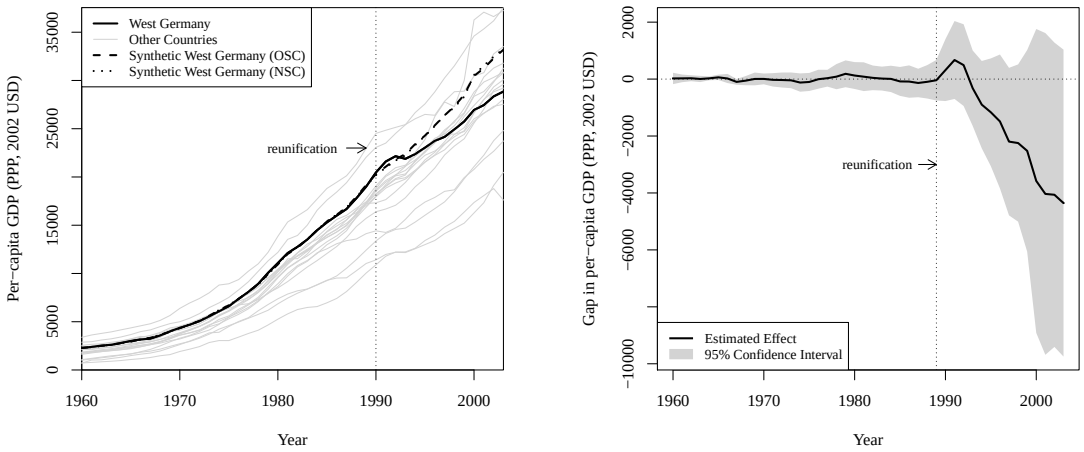
Table 1.3 compares the weights on the control countries in the original method and the nonlinear synthetic control method ($a^* = 0$, $b^* = 0.7$). The weights assigned by the original synthetic control method are concentrated on Austria, USA, the Netherlands, Switzerland, UK, Norway, Italy and Greece, with the weights in descending order. In comparison, the weights assigned by the nonlinear synthetic control method spread out among all countries, with positive weights on countries that are close to West Germany in terms of the outcomes, and negative weights on countries that are farther away. This can also be seen from Figure 1.6, which depicts the trajectories of GDP per capita for countries used for constructing the synthetic West Germany in the two methods.

Figure 1.7a depicts the trajectories of GDP per capita for West Germany (black solid line), the other countries (gray solid lines), the synthetic West Germany constructed using the original synthetic control method (black dashed line), and the synthetic West Germany constructed using the nonlinear synthetic control method (black dotted line). Despite differences in the weights, the trajectories for the two synthetic West Germanies are virtually the same, both of which follow the trajec-



(a) Original SC Method (b) Nonlinear SC Method

Figure 1.6: Countries Used for Constructing the Synthetic West Germany



(a) Trajectories of the Outcome (b) Estimated Effect

Figure 1.7: Revisiting the Example of German Reunification

tory for the real West Germany closely before the German reunification in 1990, and diverge afterwards. The gap between the trajectories for West Germany and the synthetic West Germany in the nonlinear synthetic control method indicates that the reunification reduced the per-capita GDP in West Germany by 1166 USD in 1995, 2520 USD in 1999, and 4356 USD in 2003. However, we fail to reject the null hypothesis that German reunification had no effect on West Germany's economy in any particular period.

1.5.3 The Economic Impact of the 2019 Hong Kong Protests

Background

In 2018, a young couple from Hong Kong, the 20-year-old woman, Poon Hiu-wing, and her 19-year-old boyfriend, Chan Tong-kai, travelled to Taiwan as tourists. Following a quarrel in the hotel room, Chan strangled Poon, took her valuables and fled back to Hong Kong (BBC, 2019a). Since the murder took place in Taiwan where the Hong Kong authorities had no jurisdiction, they could only charge Chan with money laundering but not homicide. Nor could they surrender Chan to Taiwan, as there was no extradition treaty or one-off surrender agreement between Hong Kong and Taiwan.[13]

To fill this legal loophole, the Hong Kong government proposed amendments to the existing extradition law (formally, the Fugitive Offenders Ordinance, Cap. 503) in February 2019, to allow case-based surrenders of fugitive offenders to jurisdictions apart from the twenty with which the city already had extradition treaties (Legislative Council, 2019).[14,15] While the proposed amendments would enable Hong Kong to surrender Chan to Taiwan, the inclusion of mainland China raised concerns among residents from different walks of life in the city that civil liberties would be infringed upon, given previous incidents of Hong Kong residents being abducted to the mainland for trial (AP, 2016). The open support of the amendments from several central government officials contributed to the rising anxieties (Reuters, 2019).

The amendments bill sparked a series of protests, which started as peaceful demonstrations in March and April, and eventually escalated into violence from June, when the government pushed for a speedy second reading of the bill to ensure its passage before the release of Chan from prison on money laundering charges. Hundreds of thousands joined the protests, and clashes broke out between the protesters and the police, with radical protesters throwing bricks dug up from the pavement, iron bars disassembled from the roadside railings, and later petrol bombs at the police,

[13]The term 'surrender' is used formally in place of 'extradition' to reflect that Hong Kong is part of China and does not have sovereign status.

[14]Amendments were also proposed to the Mutual Legal Assistance in Criminal Matters Ordinance, Cap. 525, which were less controversial.

[15]Some do not consider the limitation that excludes the rest of China from both ordinances a loophole, but rather a deliberate restriction to protect Hong Kong's legal system (Hong Kong Bar Association, 2019). However, since Hong Kong does not have the authority to enter into an extradition treaty or one-off surrender agreement only with Taiwan but not the other parts of China, this allows fugitives as Chan in the Taiwan homicide case to evade prosecution, and thus is effectively a loophole in this regard. Some lawmakers also advocated adding a sunset clause to the amendments, which would see the amendments expire after the resolution of the Taiwan case (The Standard, 2019). This was rejected by the government, who reiterated that the purpose of the amendments was not only to resolve the Taiwan case, but also to improve the existing arrangement for the surrender of fugitive offenders (HKSARG, 2019).

who responded with pepper spray, tear gas, and rubber bullets (SCMP, 2019; Ming Pao, 2019). Unlike the protests in previous years, the moderate protesters refused to split with the radical protesters this time, as most of them believed that "peaceful assembly should combine with confrontational actions to maximise the impact of protests", and that "radical tactics were understandable when the government refuses to listen", despite other peaceful avenues such as strikes (Yuen, 2019).

The suspension of the bill by the government on 15 June did not quiet, but rather boosted the morale of the protesters, who raised more demands including the full withdrawal of the bill, the retraction of the characterisation of the protests as "riots", the release and exoneration of the arrested protesters, the establishment of an independent commission of inquiry into police brutality, the resignation of Carrie Lam as chief executive, and the universal suffrage for the Legislative Council and the chief executive elections, pushing the protests to a non-resolvable end. Led by pro-independence activists who had been at the forefront of the protests from the beginning, the protests also evolved from aiming against the amendments bill to be against China or the Chinese government, challenging the "one country, two systems" principle (HKFP, 2020). The oftentimes violent protests persisted through the next few months, which saw the police headquarters besieged, the Legislative Council stormed, the Liaison Office of the Central People's Government attacked, the national emblem and flag of China desecrated, the international airport occupied, railway stations and shops vandalised, and several universities sieged by the protesters (HKFP, 2019b; CNN, 2019a; HKFP, 2019a; HuffPost, 2019; CNN, 2019b; ABC, 2019; BBC, 2019b; HKFP, 2021). Up till 14 April 2020, 8001 protesters were arrested, among whom 41% were students, with 60% being university students and 40% being secondary school students (China News, 2020).

On 30 June 2020, the Standing Committee of the National People's Congress of China passed the Hong Kong national security law, which was to be enacted by the city on its own, but which the city had failed to accomplish since its return to China in 1997.[16] Within hours, several pro-independence organisations announced the decision to disband and cease all operations (Al Jazeera, 2020). Subsequently, a dozen Legislative Council candidates were disqualified and several pro-independence activists were arrested under the national security law, bringing the year-long unrest in the city to a halt.

[16]Article 23 of Hong Kong's Basic Law: "The Hong Kong Special Administrative Region shall enact laws on its own to prohibit any act of treason, secession, sedition, subversion against the Central People's Government, or theft of state secrets, to prohibit foreign political organizations or bodies from conducting political activities in the Region, and to prohibit political organizations or bodies of the Region from establishing ties with foreign political organizations or bodies."(Basic Law, 1997)

Data

To estimate the economic impact of the 2019 anti-extradition law amendments bill protests in Hong Kong, we compare the quarterly GDP per capita of Hong Kong and the synthetic Hong Kong constructed using 48 other major economies listed in Table 1.4. The quarterly GDP per capita is measured in chained (2015) U.S. dollars and is Purchasing Power Parity (PPP) and seasonally adjusted. The treatment assignment period is the first quarter of 2019, during which the government proposed the amendments bill to the existing extradition law and the first round of protests was triggered. The window of observation is from the first quarter of 2011 to the fourth quarter of 2020, which is chosen to provide enough pretreatment periods, and in the meantime, to avoid potential structural breaks, e.g., due to the 2008 global financial crisis, over longer periods of time (Abadie, 2021).

The data for Hong Kong, Taiwan and Singapore are obtained from the respective government statistics websites. Purchasing power parities and the data for the other economies are obtained from OECD Stat, among which the seasonally adjusted GDP series for China is not available and is computed using the growth rate of the seasonally adjusted GDP. And the population for all economies is obtained from the United Nations. More detailed data sources are provided in Appendix A.1.

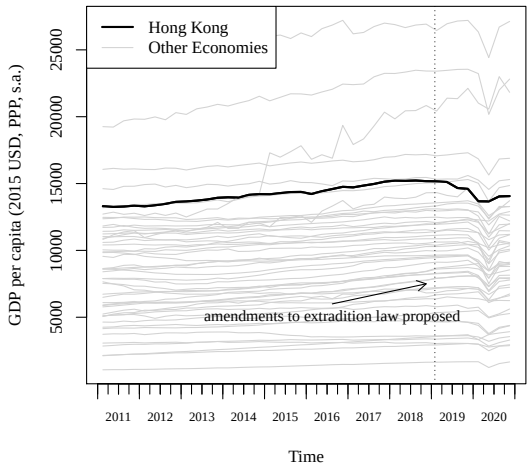


Figure 1.8: Trajectories of GDP per capita

Figure 1.8 visualises the trajectories of the quarterly GDP per capita for Hong Kong (black line) and the other economies in the sample (gray lines). We see that Hong Kong had one of the highest GDP per capita in the sample, which would be better approximated by the other economies without the non-negativity restriction. Most economies in the sample enjoyed steady growth in GDP per capita from 2011 until the first or second quarter of 2020, when almost all economies were severely hit by

the COVID-19 pandemic. Most economies then had strong rebounds in the third quarter. In contrast, the decline of the GDP per capita in Hong Kong started from early 2019, coinciding with the onset of the protests, and persisted through the outbreak of the pandemic, while the recovery from the third quarter of 2020 was only mild.

This preliminary comparison clearly points to a detrimental effect of the protests on Hong Kong's economy. To estimate the effect more accurately, we construct a synthetic Hong Kong that closely tracks the GDP per capita of Hong Kong before the protests in 2019 using the other economies in the sample, which presumably is also close to Hong Kong in the underlying predictors and thus can be used to predict the counterfactual outcome of Hong Kong. We can then estimate the economic impact of the 2019 protests using the difference in GDP per capita between the synthetic Hong Kong and the real Hong Kong. The weights assigned to the other economies for constructing the synthetic Hong Kong are determined by the nonlinear synthetic control method to ensure a small bias of the estimator given the data.

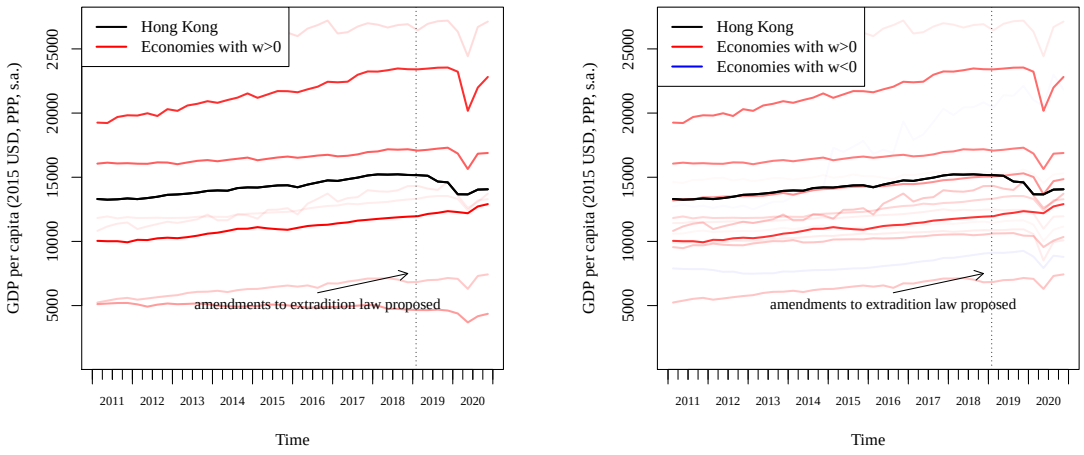
Results

Table 1.4: Synthetic Control Weights

Location	OSC Weight	NSC Weight	Location	OSC Weight	NSC Weight	Location	OSC Weight	NSC Weight
Argentina	0.12	0	Greece	0	0	Norway	0	0.01
Australia	0	0	Hungary	0	0	Poland	0	0
Austria	0	0	Iceland	0.03	0.06	Portugal	0	0
Belgium	0	0	India	0	0	Romania	0	0
Brazil	0	0	Indonesia	0	0	Russia	0	0
Bulgaria	0	0	Ireland	0	-0.01	Singapore	0.24	0.17
Canada	0	0.01	Israel	0	0	Slovakia	0	0
Chile	0	0	Italy	0	0	Slovenia	0	-0.02
China (Mainland)	0	0	Japan	0	0.07	South Africa	0	0
Colombia	0	0	South Korea	0	0	Spain	0	0
Czechia	0	0	Latvia	0	0	Sweden	0	0
Denmark	0.02	0.05	Lithuania	0	0	Switzerland	0.19	0.15
Estonia	0	0	Luxembourg	0.04	0.03	Taiwan	0.3	0.24
Finland	0	0	Mexico	0	0	Turkey	0.06	0.07
France	0	0	Netherlands	0	0	United Kingdom	0	0.02
Germany	0	0.02	New Zealand	0	0	United States	0	0.13

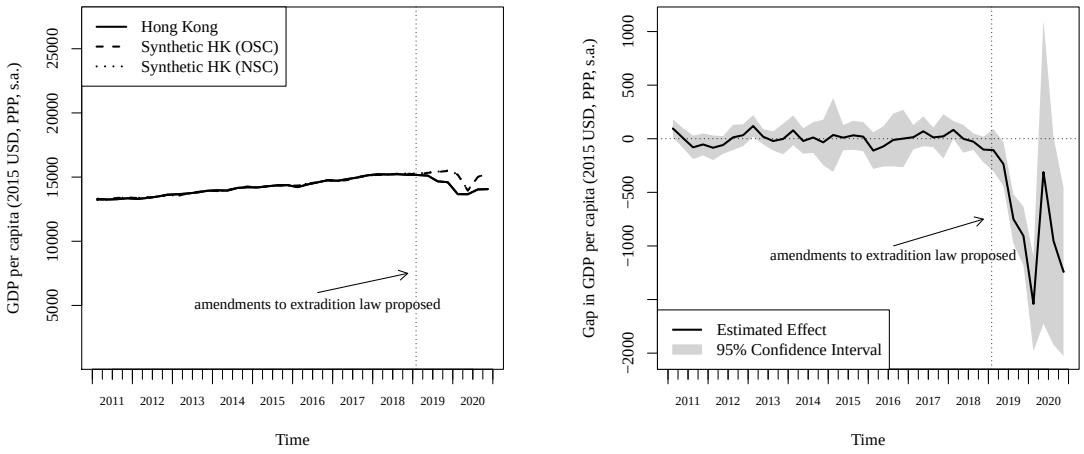
Table 1.4 displays the weights assigned to the other economies by the original method and the nonlinear synthetic control method ($a^* = 0.6$ and $b^* = 0.6$). The synthetic Hong Kong in the original method is constructed as a weighted average of Taiwan, Singapore, Switzerland, Argentina, Turkey, Luxembourg, Iceland and Denmark,

with the weights in descending order. All the other economies receive zero weight. The weights in the nonlinear synthetic control method are similar to those in the original method, with the noticeable difference that instead of assigning significant weight to Argentina, the nonlinear synthetic control method assigns weight to the United States, which is more similar to Hong Kong in terms of the outcome. This is also reflected in Figure 1.9. Although the weights in the original method are more sparse, the weights in the nonlinear synthetic control method are more concentrated on economies that are closer to Hong Kong.



(a) Original SC Method (b) Nonlinear SC Method

Figure 1.9: Economies Used for Constructing the Synthetic Hong Kong



(a) Trajectories for HK and synthetic HK (b) Estimated effect of protests

Figure 1.10: Effect of Anti-Extradition Law Protests in Hong Kong

Figure 1.10a displays the trajectories of quarterly GDP per capita for Hong Kong

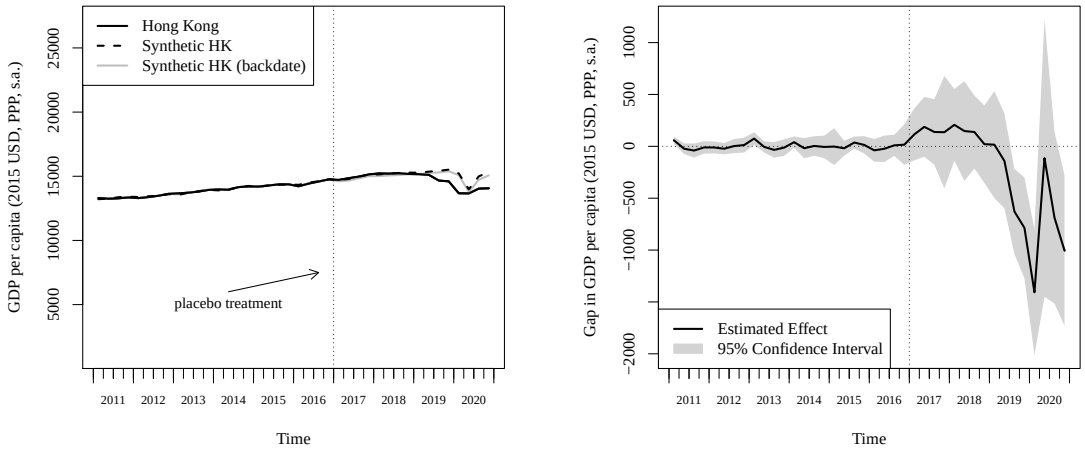
(black solid line), the synthetic Hong Kong constructed using the original synthetic control method (black dashed line), and the synthetic Hong Kong constructed using the nonlinear synthetic control method (black dotted line). Despite the differences in the weights, the trajectories of the synthetic Hong Kong using the two methods are very similar and track the trajectory of Hong Kong very closely before the proposal of the amendments bill in the first quarter of 2019, and begin to diverge immediately afterwards. This also indicates that the results are not sensitive to the choice of the tuning parameters. Figure 1.10b depicts the gap between the trajectories for Hong Kong and the synthetic Hong Kong constructed using the nonlinear synthetic control method, which is used to estimate the economic impact of the anti-extradition law amendments bill protests, as well as the 95% confidence intervals.[17] The results suggest that the protests had a negative impact on Hong Kong's economy from the second quarter of 2019. The magnitude of the impact grew rapidly and reached its peak in the first quarter of 2020, when the GDP per capita in Hong Kong was 1540.76 USD or 11.27% lower than what it would be if there were no protests. To put it into perspective, this magnitude exceeds the peak-to-trough decline in quarterly GDP per capita in Hong Kong during the previous two financial crises, which was 11.14% in the 1997 Asian financial crisis, and 8.08% in the 2008 global financial crisis, calculated using the same data series. The impact was no longer significant in the second and third quarters of 2020, but became significant again in the fourth quarter, with the quarterly GDP per capita 8.8% lower than its counterfactual level due to the slow recovery of the economy in Hong Kong.

Note that the above results rely on the assumption that there is no spill-over effect. If the protests in Hong Kong benefited competing economies such as Taiwan and Singapore by driving capital and labour to those economies, then the economic impact of the protests would be overestimated. On the other hand, if the protests in Hong Kong had negative spill-over effects on the economies that receive nonnegative weights, e.g., by damaging the economic cooperation, then the impact of the protests would be underestimated. Our results may also be confounded by the COVID-19 pandemic, which is a separate treatment from the protests. Although this treatment affected all economies, the magnitudes of the effects may be different. For example, there was virtually no further decline in the battered economy of Hong Kong in 2020, while the COVID-19 pandemic devastated almost all the other economies. The pandemic may even have benefitted Hong Kong's economy by restricting the protests. Thus, the results in 2020 should be interpreted with this in mind.

[17]Note that the confidence intervals in 2019 seem narrow only due to the slope. The average width of the confidence intervals is 274 in the pretreatment periods, and 452 in 2019.

Robustness Checks

The synthetic Hong Kong is constructed by matching on the quarterly GDP per capita of Hong Kong from 2011 to 2019. Although unlikely, people might have anticipated a proposal to amend the extradition law and the turbulence that might follow, after the Taiwan homicide case took place in 2018. To get rid of the potential anticipatory effect, we backdate the treatment to the first quarter of 2017, two years before the real treatment took place and one year before the Taiwan homicide case that triggered the government proposal of the amendments bill, and construct the synthetic Hong Kong by matching on the quarterly GDP per capita before 2017, to see if the results are sensitive to the choice of the treatment date. This exercise also allows us to examine the ability of the synthetic Hong Kong to replicate the quarterly GDP per capita of Hong Kong in the absence of the treatment. If we were to find a large gap between the trajectories of Hong Kong and the synthetic Hong Kong after the placebo treatment in 2017 and before the real treatment in 2019, then it would undermine the credibility of the previous results.

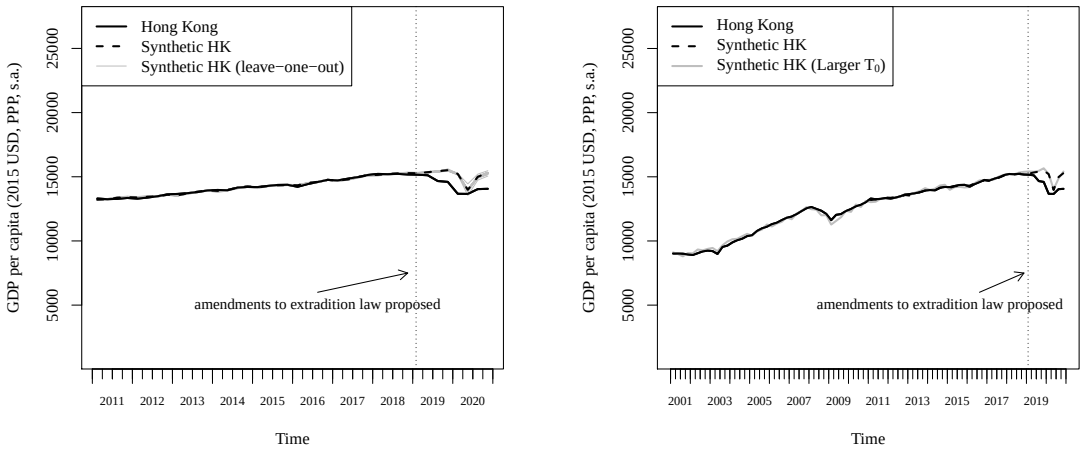


(a) Trajectories for HK and synthetic HK (b) Estimated effect of protests

Figure 1.11: Backdating the Treatment to 2017

The results of the backdating exercise are presented in Figure 1.11, which turn out to be very similar to the previous results. We find no significant placebo treatment effect as the trajectory of the newly constructed synthetic Hong Kong (gray solid line) follows that of Hong Kong closely not only before the placebo treatment in the first quarter of 2017, but also all the way through 2017 and 2018, and begins to diverge immediately after the real treatment took place in early 2019. The magnitude of the estimated treatment effect is close to the previous result, and the 95% confidence intervals in the backdating exercise similarly suggest that the treatment effect becomes significant from the third quarter of 2019 onwards, except in the

second and third quarter of 2020. The fact the the synthetic Hong Kong is able to reproduce the GDP per capita of Hong Kong in the absence of the real treatment shows the credibility of the synthetic control estimator. And the emergence of the estimated effect shortly after the real treatment provides confidence that the results are driven by a true detrimental impact of the protests.



(a) Leave-one-out Distribution (b) Longer Pretreatment Period

Figure 1.12: Additional Robustness Checks

Apart from the backdate exercise, we conduct two additional robustness checks. In Figure 1.12a, we conduct the leave-one-out exercise, where we exclude one economy at a time from the construction of the synthetic Hong Kong. We see that the trajectories for the synthetic Hong Kong constructed in the leave-one-out iterations are all very similar to the previous results. This shows that the estimated effect of the protests is robust to the exclusion of any particular economy.

We include observations from 2011 in our main analysis to avoid potential structural breaks over a longer timespan. This provides us with 36 pretreatment periods, which should be sufficient to produce credible results. Nevertheless, we check whether our results are robust to the inclusion of more pretreatment periods. In Figure 1.12b, we double the total time periods by further including outcomes observed from 2001 to 2010.[18] The trajectory for the synthetic Hong Kong constructed by matching on outcomes from 2001 to 2018 follows closely the trajectory for the real Hong Kong, and the results are very similar with the benchmark results after the treatment. Thus our results are not sensitive to the inclusion of more pretreatment periods.

[18]This excludes Mainland China from the analysis due to missing data.

1.6 Conclusion

In this paper, we generalise the synthetic control method to the case where the outcome is a nonlinear function of the underlying predictors. Specifically, we provide conditions for the asymptotic unbiasedness of the synthetic control estimator to complement the theoretical result for the linear case in Abadie et al. (2010), and propose a flexible and data-driven method for choosing the synthetic control weights. Monte Carlo simulations show that the nonlinear synthetic control method has similar or better performance in the linear case and better performance in the nonlinear case compared with competing methods, and that the confidence intervals have good coverage probabilities across settings. In the empirical application, we illustrate the method by estimating the impact of the 2019 anti-extradition law amendments bill protests on Hong Kong's economy, and find that the year-long protests reduced the real GDP per capita by 11.27% in the first quarter of 2020, which is larger in magnitude than the economic decline in the 1997 Asian financial crisis and the 2008 global financial crisis.

Appendix A

A.1 Data Sources

- Quarterly GDP for Hong Kong. Sources: Seasonally adjusted GDP in real terms and implicit price deflator (IPD) of GDP, Table E200-6, https://data.gov.hk/en-data/dataset/hk-censtatd-tablechart-gdp.
- Quarterly GDP for Taiwan. Sources: GDP by Expenditures - Seasonally Adjusted Series, Implicit Price Deflators, Principal Figures, https://eng.stat.gov.tw/ct.asp?xItem=37408&CtNode=5347&mp=5. Purchasing Power Parity/Exchange Rate, https://fred.stlouisfed.org/series/PLGDPOTWA670NRUG.
- Quarterly GDP for Singapore. Sources: Gross Domestic Product In Chained (2015) Dollars, By Industry, Quarterly, Seasonally Adjusted, https://www.tablebuilder.singstat.gov.sg/publicfacing/createDataTable.action?refId=16062.
- Quarterly GDP for the other economies. Sources: GDP expenditure approach, Quarterly National Accounts, OECD Stat, https://stats.oecd.org/.
- Purchasing power parities. Sources: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm.
- Population. Sources: Total Population - Both Sexes, Population Dynamics, Department of Economic and Social Affairs, United Nations, https://population.un.org/wpp/Download/Standard/Population/.

A.2 Proofs

Proof of Theorem 1.1. The proof follows closely the proof in Appendix B of Abadie et al. (2010). We thus omit many details. For more details, see Abadie et al. (2010) or Botosaru and Ferman (2019).

Under the assumptions, we have

$$\begin{aligned} e_{1t} \equiv & Y_{1t}^0 - \sum_j w_j^* Y_{jt} \\ = & \boldsymbol{\lambda}_t' \left(\boldsymbol{\lambda}^{T_0\prime} \boldsymbol{\lambda}^{T_0} \right)^{-1} \boldsymbol{\lambda}^{T_0\prime} \sum_j w_j^* \varepsilon_j^{T_0} \\ & - \boldsymbol{\lambda}_t' \left(\boldsymbol{\lambda}^{T_0\prime} \boldsymbol{\lambda}^{T_0} \right)^{-1} \boldsymbol{\lambda}^{T_0\prime} \varepsilon_1^{T_0} + \varepsilon_{1t} - \sum_j w_j^* \varepsilon_{jt}. \end{aligned} \tag{A.1}$$

The terms on the last line has zero conditional mean given Assumption 1.2, however, the term on the penultimate line does not have zero mean because w_j^* is correlated with $\varepsilon_j^{T_0}$.

Denote the first term as R_{1t}. Suppose that the elements of $|\boldsymbol{\lambda}_t|$ are bounded from above by $\bar{\boldsymbol{\lambda}}$ for $t = 1, \ldots, T$. Under Assumption 1.4 and using the Cauchy–Schwarz Inequality, we have

$$\left(\boldsymbol{\lambda}_t' \left(\sum_{n=1}^{T_0} \boldsymbol{\lambda}_n \boldsymbol{\lambda}_n' \right)^{-1} \boldsymbol{\lambda}_s \right) \leq \left(\frac{\bar{\boldsymbol{\lambda}}^2 f}{T_0 \underline{\xi}} \right).$$

Denote $\bar{w} = \max_j |w_j^*|$ ($\bar{w} \leq 1$ given the adding-up and non-negativity assumptions), and $\bar{\varepsilon}_j = \sum_{s=1}^{T_0} \boldsymbol{\lambda}_t' \left(\sum_{n=1}^{T_0} \boldsymbol{\lambda}_n \boldsymbol{\lambda}_n' \right)^{-1} \boldsymbol{\lambda}_s \varepsilon_{js}$. Then using Hölder's Inequality, we have

$$|R_{1t}| \leq \bar{w} \sum_j |\bar{\varepsilon}_j| \leq \bar{w} J^{1-\frac{1}{p}} \left(\sum_j |\bar{\varepsilon}_j|^p \right)^{1/p}$$

for some positive integer p.

Using Hölder's Inequality again and using Rosenthal's Inequality, we have

$$\mathbb{E}|\bar{\varepsilon}_j|^p \leq C(p) \left(\frac{\bar{\boldsymbol{\lambda}}^2 f}{T_0 \underline{\xi}} \right)^p \max \left\{ \sum_{s=1}^{T_0} \mathbb{E}|\varepsilon_{js}|^p, \left(\sum_{s=1}^{T_0} \mathbb{E}|\varepsilon_{js}|^2 \right)^{p/2} \right\},$$

where the constant $C(p) = \mathbb{E}(\theta - 1)^p$ with θ being a Poisson random variable with parameter 1.

Denote $\bar{m}_p(T_0) = \max_j (1/T_0) \sum_{s=1}^{T_0} \mathbb{E}|\varepsilon_{js}|^p$, then we have

$$\mathbb{E}|R_{1t}| < \bar{w} J C(p)^{1/p} \left(\frac{\bar{\boldsymbol{\lambda}}^2 f}{\underline{\xi}} \right) \max \left\{ \frac{\bar{m}_p(T_0)^{1/p}}{T_0^{1-1/p}}, \frac{\bar{m}_2(T_0)^{1/2}}{T_0{}^{1/2}} \right\}. \tag{A.2}$$

Thus, the bias is bounded by a value that goes to zero when the number of pretreatment periods goes to infinity. Since $\tau_{it} = Y_{it} - Y_{it}^0$, this implies that

$$\mathbb{E}\left(\hat{\tau}_{1t}^{SC} - \tau_{1t}\right) \to 0 \text{ as } T_0 \to \infty.$$

□

Proof of Theorem 1.2. Under Assumption 1.6 that $G(\cdot)$ is a smooth function with $G^{(n)} < \infty$ for all n, we can expand $G\left(\boldsymbol{X}_j'\boldsymbol{\beta}_t + \boldsymbol{\mu}_j'\boldsymbol{\lambda}_t\right)$ for $j \in \mathcal{J}_M$ at $\boldsymbol{X}_1'\boldsymbol{\beta}_t + \boldsymbol{\mu}_1'\boldsymbol{\lambda}_t$ using Taylor's rule:

$$\begin{aligned}
&G\left(\boldsymbol{X}_j'\boldsymbol{\beta}_t + \boldsymbol{\mu}_j'\boldsymbol{\lambda}_t\right) \\
=&\sum_{n=0}^{\infty} \frac{\partial^n G\left(\boldsymbol{X}_1'\boldsymbol{\beta}_t + \boldsymbol{\mu}_1'\boldsymbol{\lambda}_t\right)}{n!\partial\left(\boldsymbol{X}_1'\boldsymbol{\beta}_t + \boldsymbol{\mu}_1'\boldsymbol{\lambda}_t\right)^n} \\
=&G\left(\boldsymbol{X}_1'\boldsymbol{\beta}_t + \boldsymbol{\mu}_1'\boldsymbol{\lambda}_t\right) + \sum_{n=1}^{\infty} \frac{G^{(n)}}{n!} e_{jt}^n,
\end{aligned}$$

where $e_{jt} = \left(\boldsymbol{X}_j - \boldsymbol{X}_1\right)'\boldsymbol{\beta}_t + \left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_1\right)'\boldsymbol{\lambda}_t$.

Since $\sum_{j\in\mathcal{J}_M} w_j^* = 1$, taking the expectation of $Y_{1t}^0 - \sum_j w_j^* Y_{jt}$ w.r.t. ε_{1t} and ε_{jt}, $j \in \mathcal{J}_M$, we have

$$\begin{aligned}
&\mathbb{E}_\varepsilon\left(Y_{1t}^0 - \sum_j w_j^* Y_{jt}\right) \qquad \text{(A.3)} \\
=&\left[\left(\sum_j w_j^* \boldsymbol{X}_j - \boldsymbol{X}_1\right)'\boldsymbol{\beta}_t + \left(\sum_j w_j^* \boldsymbol{\mu}_j - \boldsymbol{\mu}_1\right)'\boldsymbol{\lambda}_t\right] G' \\
&- \sum_{n=2}^{\infty} \frac{G^{(n)}}{n!}\left(\sum_j w_j^* e_{jt}^n\right).
\end{aligned}$$

If the outcomes are linear functions of the predictors, then the bias goes to zero when $T_0 \to \infty$ according to Theorem 1.1. If the outcomes are nonlinear functions, then the first term on the RHS of equation (A.3) may not go to zero since matching on the observed predictors and the pretreatment outcomes does not guarantee matching on the unobserved predictors, and the second term will not vanish since $\boldsymbol{H}_j - \boldsymbol{H}_1$ will not go to zero when T_0 and J increase at the same rate.

We now develop conditions under which the bias given by (A.3) converges to zero, using results from Abadie and Imbens (2006).

Equation (A.3) can be written as $\mathbb{E}_\varepsilon\left(Y_{1t}^0 - \sum_j w_j^* Y_{jt}\right) = -\sum_{n=1}^{\infty} \frac{G^{(n)}}{n!}\left(\sum w_j^* e_{jt}^n\right)$.

This bias goes to zero if e_{jt} goes to zero.

Denote $\boldsymbol{U}_j = \boldsymbol{Z}_1 - \boldsymbol{Z}_j$, $j \in \mathcal{J}_M$. Without loss of generality, let $M = 1 + k + T_0$. Lemma 1 in Abadie and Imbens (2006) shows that $\boldsymbol{U}_j = O_p\left(J^{-\frac{1}{1+k+T_0}}\right)$. Thus for fixed T_0, $\boldsymbol{Z}_j - \boldsymbol{Z}_1 \xrightarrow{p} 0$ when $J \to \infty$.

Since $F(\cdot)$ is a strictly monotonic function, $\boldsymbol{Z}_j - \boldsymbol{Z}_1 \xrightarrow{p} 0$ implies that $\boldsymbol{\lambda}^{T_0}\left(\boldsymbol{\mu}_j - \boldsymbol{\mu}_1\right) + \varepsilon_j^{T_0} - \varepsilon_1^{T_0} \xrightarrow{p} 0$. With Assumption 1.4, we have $\boldsymbol{\mu}_j - \boldsymbol{\mu}_1 + \boldsymbol{\lambda}_t'\left(\boldsymbol{\lambda}^{T_0\prime}\boldsymbol{\lambda}^{T_0}\right)^{-1}\boldsymbol{\lambda}^{T_0\prime}\left(\varepsilon_j^{T_0} - \varepsilon_1^{T_0}\right) \xrightarrow{p} 0$, which implies that $\boldsymbol{\mu}_j - \boldsymbol{\mu}_1 \xrightarrow{p} 0$ as $T_0 \to \infty$. However, $\boldsymbol{U}_j = O_p\left(J^{-\frac{1}{1+k+T_0}}\right)$ may not hold when $T_0 \to \infty$.

Suppose $T_0^{b(T_0)}/J = O(1)$ for some $b(T_0) \geq 1$. For $\boldsymbol{U}_j$ to converge to 0 when T_0 goes to infinity, we need $\lim_{T_0\to\infty} J^{\frac{1}{1+k+T_0}} \to \infty$, i.e., $\lim_{T_0\to\infty} T_0^{\frac{b(T_0)}{1+k+T_0}} = \lim_{T_0\to\infty} e^{\frac{b(T_0)}{1+k+T_0}\ln T_0} = \lim_{T_0\to\infty} e^{b'(T_0)\ln T_0 + \frac{b(T_0)}{T_0}} \to \infty$ using L'Hôpital's rule. This requires $b'(T_0) > 0$ since $b(T_0) \geq 1$.

Given the regularity assumptions, $\mathbb{E}_\varepsilon\left(Y_{1t}^0 - \sum_j w_j^* Y_{jt}\right)$ is uniformly integrable. Thus $\mathbb{E}\left(\tilde{\tau}_{1t} - \tau_{1t}\right) = \mathbb{E}\left[\mathbb{E}_\varepsilon\left(Y_{1t}^0 - \sum_j w_j^* Y_{jt}\right)\right] \to 0$ when $T_0 \to \infty$ if $J = O\left(T_0^{b(T_0)}\right)$ with $b(T_0) \geq 1$ and $b'(T_0) > 0$.

□

Chapter 2

The Synthetic Control Method with Multiple Outcomes: Estimating the Effects of Non-Pharmaceutical Interventions in the COVID-19 Pandemic

2.1 Introduction

The synthetic control method (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015; Abadie, 2021) is a popular method for estimating the effect of a policy or intervention on an aggregate unit, such as a country or a city. The procedure consists of constructing a synthetic control unit using a convex combination of the control units such that the distance between the outcomes of the treated unit and the synthetic control before the treatment is minimised, and then estimating the treatment effects using the difference between the outcomes after the treatment. Formally, Abadie et al. (2010) show that the bias of the synthetic control estimator is bounded by a function that is inversely proportional to the number of pretreatment periods, provided that the treated unit is well approximated by the synthetic control in the pretreatment periods. Intuitively, if the synthetic control constructed using only a handful of control units can closely track the trajectory of the outcome for the treated unit in the absence of the treatment, then the longer the pretreatment time span is, the more confident we are that the close to perfect pretreatment fit is due to the similarity between the synthetic control and the treated unit in terms of the underlying predictors, rather than due to overfitting or coincidence.

Our first contribution is methodological. On the one hand, Abadie et al. (2015) point out that "the applicability of the method requires a sizable number of preintervention periods" and that "we do not recommend using this method when the pretreatment fit is poor or the number of pretreatment periods is small". On the other hand, there may be structural breaks in the relationship between the outcome of interest and the underlying predictors, when the number of pretreatment periods becomes large (Abadie, 2021). We generalise the conventional single-outcome synthetic control method to a multiple-outcome framework, where the time dimension is supplemented with the extra dimension of related outcomes in the same domain, so that the method can be used even when there are only a few pretreatment periods, or if we worry about structural breaks due to technological advances or demographic changes over a longer time span. As we demonstrate in the replication exercise in Section 2.2.2 and the treatment backdating exercise in Section 2.4.4, our method may work well even when we only observe the outcomes in a single pretreatment period, or in the extreme case, when we do not observe some of the outcomes before the treatment at all, as long as we have sufficient pretreatment information provided by the related outcomes. The multiple-outcome synthetic control method is also useful when there are an abundance of pretreatment periods and no structural breaks, as we show that the bound on the bias of the estimator is of a smaller stochastic order than that of the conventional single-outcome synthetic control estimator when the unit of interest can be closely approximated by the synthetic control in terms of the observed predictors and the multiple related outcomes before the treatment.

Apart from the methodological contribution, our paper also contributes empirically to understanding the impacts of non-pharmaceutical interventions (NPIs) in the COVID-19 pandemic. Examples of NPIs include closing school and workplaces, restricting travels and gatherings, public information campaigns, and so on. Since the outbreak of the pandemic, most countries have resorted to NPIs to reduce the spread of the virus and to prevent the health system from being overwhelmed, in the absence of vaccines. Although a variety of vaccines have become available since late 2020, the limited production capacity, the strict storage and transportation requirements, as well as the lack of knowledge on their effectiveness against mutant strains of the coronavirus indicate that NPIs will continue to play a major role in dealing with the COVID-19 pandemic in the near future.

The literature on the effects of NPIs on public health and the economy is rapidly growing. One set of studies focus on the public health effects of NPIs (Fang et al., 2020; Friedson et al., 2020; Flaxman et al., 2020; Hsiang et al., 2020; Born et al., 2020; Cho, 2020; Conyon et al., 2020; Mitze et al., 2020; Chernozhukov et al., 2021). Most of these papers point to the effectiveness of NPIs in reducing COVID-19 infection

cases and deaths. For example, exploiting the contrasting levels of stringency in NPIs between Sweden and the other European countries, Cho (2020) constructs a synthetic Sweden and finds that COVID-19 infection cases and excess mortality in Sweden would have been significantly reduced had Sweden initially adopted stricter containment measures.[1]

Another set of studies examine the impacts of NPIs on labour market or economic outcomes. Focusing on Scandinavia, where Sweden implemented much lighter NPIs than its neighbors, Juranek et al. (2020) show that the Swedish labour market was less severely hit, and Sheridan et al. (2020) find that most of the contraction in consumer spending was due to the pandemic rather than the NPIs. A number of studies exploiting the differences in the timing of the state stay-at-home policies in the US (Forsythe et al., 2020; Baek et al., 2020; Rojas et al., 2020; Murray and Olivares, 2020; Kong and Prinz, 2020) provide evidence that stay-at-home orders accounted for a relatively small share of the increase in unemployment insurance claims. Using cellphone records and consumer spending data, Goolsbee and Syverson (2020) find that legal shutdown orders accounted for only a modest share of the decline in consumer visits, and Alexander and Karger (2020) show that stay-at-home orders reduced spending on retail businesses, but increased spending on food delivery services. In contrast, Gupta et al. (2020) and Coibion et al. (2020) find that much of the decline in employment or consumer spending was driven by state social distancing policies or lockdowns, using survey data.

There are also studies addressing the "trade-off" between public health and the economy. Comparing countries and cities around the world, Fernández-Villaverde et al. (2020) document a positive correlation between fatality rates and GDP losses. Using high-frequency proxies like electricity consumption as economic indicators, Demirguc-Kunt et al. (2020) and Fezzi and Fanghella (2020) show that earlier NPIs led to better health and economic outcomes. Using historical data for the US during the 1918 flu, Correia et al. (2020) find that NPIs could reduce disease transmission without necessarily further depressing economic activities. Several studies using structural modelling (Aum et al., 2020; Arnon et al., 2020; Chen and Qiu, 2020; Acemoglu et al., 2020; Baqaee et al., 2020; Favero et al., 2020) show that NPIs may not necessarily induce a trade-off between public health and the economic performance: well-designed policies targeted at different age groups and risk sectors could save lives with limited economic costs.

In the empirical application, we illustrate the multiple-outcome synthetic control method by extending the analysis of Born et al. (2020) and Cho (2020) to multiple

[1] Cho (2020)'s empirical strategy follows Born et al. (2020), who did not find statistically significant effect of NPIs on COVID-19 infection cases, based on a shorter span of observation.

public health, labour market, and economic outcomes. Our results suggest that had Sweden implemented stricter NPIs as in the other European countries by March, the cumulative numbers of COVID-19 infection cases and deaths would have been reduced by 70% and 68% respectively by July, and weekly deaths from all causes would have been 20% fewer in early May. The impact on mortality was larger for males and most visible for people older than 60. As for the labour market, we find that stricter NPIs would increase absence from work by almost 76% mainly through temporary layoffs, and reduce total hours worked by about 12%, for the employed in the second quarter of 2020. The impacts would quickly vanish in the third quarter, and there would be no discernible effect on the employment rate throughout. In terms of the economy, we find that stricter NPIs would shrink the volume of retail sales by 5%-13% from March to May, almost exclusively due to reduced sales in non-food products. However, we do not find any statistically significant effects of stricter NPIs on the other economic outcomes including GDP, import, export, industrial production, and CPI, which indicates that almost all of the contraction in the economy was due to the pandemic itself rather than the NPIs.

Our empirical results are in line with the existing studies, which largely find that the NPIs had statistically and economically significant effects on the public health outcomes, but a much smaller role in the downturn of the labour market and the economy. There are several interesting new findings as well. First, we find that the potential reduction in the cumulative number of COVID-19 deaths (per million population) by July is 390, while the reduction in the cumulative number of deaths from all causes between April and July is 364. One probable explanation for the slightly smaller reduction in the latter is that people with existing conditions or weaker immune systems are more susceptible to COVID-19, and thus mortality that would have been attributed to other causes may have been counted as COVID-19 deaths instead. The NPIs may also have limited the spread of other transmissible diseases through promoting good hygiene behaviours and reducing face-to-face interactions. In either case, the estimated reduction in deaths from all causes due to the NPIs, which amounts to a 64% drop from the realised level of COVID-19 deaths, may serve as a lower bound for the reduction in COVID-19 deaths. Second, the NPIs would have significant effects on temporary absence from work and total hours worked among the employed, but no effect on the employment rate. This is likely due to the various employment support policies that were carried out across European countries, with the aim of preserving jobs through income support and wage subsidies (OECD, 2020). Third, the finding that the NPIs would only have significant impact on retail sales in the early stages, but no effects on the other economic outcomes may be somewhat counterintuitive. Further investigation into

how the economy quickly absorbed the shocks brought by the pandemic and the NPIs would be interesting for future research.

The rest of the paper is arranged as follows. Section 2.2 describes the theoretical framework for the the multiple-outcome synthetic control method. Section 2.3 compares the multiple-outcome synthetic control method with the conventional single-outcome synthetic control method using Monte Carlo simulations. Section 2.4 presents the empirical results and robustness checks. Section 2.5 discusses the limitations and concludes.

2.2 Theoretical Framework

In this section, we generalise the conventional single-outcome synthetic control method to a multiple-outcome framework, where the synthetic control is constructed using a convex combination of the control units, with weights selected to match the synthetic control and the treated unit in terms of the observed predictors and the pretreatment values of multiple related outcomes in the same domain. It can be shown that the bias of the multiple-outcome synthetic control estimator is bounded by a function that shrinks to zero as the number of pretreatment periods or the number of related outcomes increases. Moreover, the bound is of a smaller stochastic order than that of the single-outcome synthetic control estimator when the treated unit can be closely approximated by the synthetic control in multiple related outcomes before the treatment.

To illustrate the applicability of our method, we then conduct a compact re-analysis of the economic cost of the 1990 German reunification, which shows that the synthetic West Germany constructed by matching on multiple related outcomes in 1989 alone can closely track the trajectory of GDP per capita in West Germany from 1960 to 1990. This real-life example demonstrates that similarly with matching on the single outcome in many pretreatment periods, the synthetic control constructed by matching on multiple related outcomes in a small number of pretreatment periods can closely approximate the treated unit in terms of the underlying predictors of the outcomes as well, providing credibility of our method even when the number of pretreatment periods is small.

We also discuss two extensions that will be useful in the empirical application. One is to use demeaned outcomes, i.e., outcomes in differences with respect to their pretreatment averages, to allow better pretreatment fits when the differences in the level of the outcomes for different units are relatively stable over time. This may be especially helpful in the multiple-outcome framework, where the relative positions of the units may vary across different outcomes. The other extension is to include

an additional assumption on the functional form of the treatment effect, which is necessary if we want to estimate the treatment effect on the untreated using the synthetic control method in cases with many treated units.

To avoid confusion, our notations largely follow those in Abadie (2021). The proofs are collected in the appendix.

2.2.1 Multiple outcomes framework

Suppose that we observe K outcomes in domain $\mathbb{K} = \{1, 2, \ldots, K\}$ for $J + 1$ units over T time periods, where a domain refers to a collection of related outcomes driven by the same set of observed and unobserved predictors. For example, the economic domain contains different measures of the economic performance, such as GDP, industrial production, retail sales, and CPI, which can be assumed to depend on the same set of underlying predictors such as infrastructure, technology, natural resources, demographic composition, work ethic, etc.

Without loss of generality, we assume that the first unit ($i = 1$) receives the treatment at period $T_0 + 1 \leq T$ and remains treated afterwards, while all the other J units ($i = 2, \ldots, J + 1$) are untreated throughout the window of observation. Denoting the binary treatment status for unit i at time t as D_{it}, we have $D_{it} = 1$ for $i = 1$ and $t > T_0$, and $D_{it} = 0$ otherwise.

We are interested in the effect of the treatment on a single or multiple outcomes in domain $\mathbb{K}$ for the treated unit after the treatment:

$$\tau_{1t,k} = Y^1_{1t,k} - Y^0_{1t,k}, \;\; t > T_0, \;\; k \in \mathbb{K}, \tag{2.1}$$

where $Y^1_{1t,k}$ is the potential outcome under the treatment, and $Y^0_{1t,k}$ is the potential outcome without the treatment, so that the observed outcome can be written as $Y_{1t,k} = D_{1t}Y^1_{1t,k} + (1 - D_{1t})\, Y^0_{1t,k}$. Since we only observe the treated potential outcome but not the untreated potential outcome for unit 1 at $t > T_0$, we need to predict the counterfactual outcome $Y^0_{1t,k}$.

Suppose that the untreated potential outcome $k \in \mathbb{K}$ for unit i at time t is given by an interactive fixed effects model

$$Y^0_{it,k} = \delta_{t,k} + \boldsymbol{Z}_i'\boldsymbol{\theta}_{t,k} + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_{t,k} + \varepsilon_{it,k}, \tag{2.2}$$

where $\delta_{t,k}$ is the time trend in outcome k, $\boldsymbol{Z}_i$ and $\boldsymbol{\mu}_i$ are the $r \times 1$ and $f \times 1$ vectors of observed and unobserved predictors of $Y^0_{it,k}$ with outcome-specific coefficients $\boldsymbol{\theta}_{t,k}$ and $\boldsymbol{\lambda}_{t,k}$, respectively, and $\varepsilon_{it,k}$ is the individual transitory shock.

Remark 2.1. Note that the above functional form of the untreated potential outcomes does not exclude the possibility that some outcomes in the domain depend on other predictors that are independent from the included predictors and the treatment status, which can thus be treated as part of the transitory shocks. In addition, the coefficients may contain zero so that the corresponding predictors may affect some outcomes in some periods, but not all outcomes in all periods, as long as there is enough variation in the coefficients across different pretreatment periods or outcomes, as specified in Assumption 2.3.

Remark 2.2. Following Abadie (2021), we interpret the interactive fixed effects term as the product of the unobserved predictors and the corresponding coefficients. There is an alternative interpretation in the time series literature, where the interactive fixed effects term is the product of common time factors and unit-specific factor loadings. This interpretation does not apply to the model in (2.2), as it does not make much sense to have common time factors that vary across the outcomes, while the factor loadings stay the same. To accommodate this interpretation, we need a setup with common time factors $\boldsymbol{\lambda}_t$ that are invariant across the outcomes, and outcome-specific factor loadings $\boldsymbol{\mu}_{i,k}$. Since both interpretations are accepted in the literature, it would be important for future research to have a general model that incorporates both setups in the multiple-outcome framework. A more general model would be

$$Y_{it,k}^0 = \delta_{t,k} + \boldsymbol{Z}_i'\boldsymbol{\theta}_{t,k} + \boldsymbol{\mu}_{i,k}'\boldsymbol{\lambda}_{t,k} + \varepsilon_{it,k}, \tag{2.3}$$

where the first interpretation applies if we assume $\boldsymbol{\mu}_{i,k} = \boldsymbol{\mu}_i$, and the alternative interpretation applies if we assume $\boldsymbol{\lambda}_{t,k} = \boldsymbol{\lambda}_t$. If we assume $\boldsymbol{\mu}_{i,k} = \boldsymbol{\mu}_i + \boldsymbol{u}_k$ or $\boldsymbol{\mu}_{i,k} = \boldsymbol{\mu}_i \circ \boldsymbol{u}_k$, where $\boldsymbol{u}_k$ is the outcome-specific factor loadings and $\circ$ represents the element-wise product, then (2.3) reduces to (2.2). Allowing for a non-restrictive structure of $\boldsymbol{\mu}_{i,k}$ comes at the cost of the order of the bias, as the outcomes are no longer related and we lose the benefit of matching on multiple related outcomes.

The individual transitory shocks are assumed to be independent across units, time and outcomes, which is not as restrictive as it seems since the unobserved interactive fixed effects that account for the correlations along those dimensions have been separated out. They are also assumed to have zero mean conditional on the predictors and the treatment status, and that they do not dominate the predictors in determining the outcomes. These assumptions are given in Assumption 2.1.

Assumption 2.1.

1) $\varepsilon_{it,k}$ *are independent across* i*,* t*,* k*;*

2) $\mathbb{E}(\varepsilon_{it,k} \mid \boldsymbol{Z}_j, \boldsymbol{\mu}_j, D_{js}) = 0$ *for all* i*,* j*,* t*,* s *and* k*;*

3) $\mathbb{E}|\varepsilon_{it,k}|^p < \infty$ *for all* i*,* t*,* k *and some even integer* $p \geq 2$*.*

A synthetic control is constructed using a convex combination of the control units such that the synthetic control matches the treated unit in terms of the observed predictors and the pretreatment values of the K related outcomes. This can be achieved if the matching variables of the treated unit is in the convex hull of those of the control units.[2]

Assumption 2.2. *There exists a set of weights* $\left(w_2^*, \ldots, w_{J+1}^*\right)$ *such that* $w_j^* \geq 0$ *for* $j = 2, \ldots, J+1$*,* $\sum_{j=2}^{J+1} w_j^* = 1$*,* $\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j = \boldsymbol{Z}_1$ *and* $\sum_{j=2}^{J+1} w_j^* Y_{jt,k} = Y_{1t,k}$ *for all* $t \leq T_0$ *and* $k \in \mathbb{K}$*.*

The multiple-outcome synthetic control estimator for $\tau_{1t,k}$ is then constructed as

$$\widehat{\tau}_{1t,k} = Y_{1t,k} - \sum_{j=2}^{J+1} w_j^* Y_{jt,k}. \tag{2.4}$$

Note that to predict the counterfactual outcomes for the treated unit at time $t > T_0$, we need the synthetic control to match the treated unit in terms of the underlying predictors. However, matching on the observed predictors and the pretreatment outcomes does not necessarily translate to matching on the unobserved predictors, since different combinations of the unobserved predictors may produce the same outcomes if the coefficients are linearly dependent. This happens, for example, if $KT_0 < f$ or if the coefficients are fixed over time. We thus need to ensure that there is enough variation in the effects of the unobserved predictors across the T_0 pretreatment periods or the K outcomes.

Assumption 2.3. *The smallest eigenvalue of* $\frac{1}{KT_0} \sum_{k=1}^{K} \sum_{t=1}^{T_0} \boldsymbol{\lambda}_{t,k} \boldsymbol{\lambda}_{t,k}'$ *is bounded from below by some positive number* $\underline{\xi}$*.*

To facilitate a straightforward comparison between the multiple-outcome synthetic control estimator and the conventional synthetic control estimator based only on outcome k, let $\{\widetilde{w}_j^{(k)}\}_{j=2}^{J+1}$ be the single-outcome synthetic control weights such that $\widetilde{w}_j^{(k)} \geq 0$ for $j = 2, \ldots, J+1$, $\sum_{j=2}^{J+1} \widetilde{w}_j^{(k)} = 1$, $\sum_{j=2}^{J+1} \widetilde{w}_j^{(k)} \boldsymbol{Z}_j = \boldsymbol{Z}_1$ and $\sum_{j=2}^{J+1} \widetilde{w}_j^{(k)} Y_{jt,k} =$

[2] In practice, there may not be a set of weights that satisfy the restrictions in Assumption 2.2 exactly. If the treated unit is close to being in the convex hull of the control units, then we can choose the weights by minimising the distance of the synthetic control and the treated unit in the matching variables, so that the restrictions in Assumption 2.2 hold approximately. In Section 2.2.3, we discuss demeaning the outcomes when there are stable differences in the level of the outcomes for different units over time, so that the convex hull restriction is more likely to be satisfied.

$Y_{1t,k}$ for all $t \leq T_0$. Without loss of generality, we can write

$$\widetilde{w}_j^{(k)} = w_j^* + \widetilde{v}_j^{(k)},\ j = 2, \ldots, J+1, \tag{2.5}$$

where $-1 \leq \widetilde{v}_j^{(k)} \leq 1$.

The following result provides a unified framework for the biases of the two synthetic control estimators, where we show that the bias of the conventional single-outcome synthetic control method is usually $O\left(\frac{1}{\sqrt{T_0}}\right)$, but if the single-outcome synthetic control weights coincide with the multiple-outcome synthetic control weights, in which case $\widetilde{v}_j^{(k)} = 0$ and $\widetilde{w}_j^{(k)} = w_j^*$, then the order becomes $O\left(\frac{1}{\sqrt{KT_0}}\right)$.

Proposition 2.1. *Under Assumptions 2.1, 2.2 and 2.3,*

$$Y_{1t,k}^0 - \sum_{j=2}^{J+1} \widetilde{w}_j^{(k)} Y_{jt,k} = B_{1t,k} + B_{2t,k} + B_{3t,k} + B_{4t,k},$$
$$Y_{1t,k}^0 - \sum_{j=2}^{J+1} w_j^* Y_{jt,k} = B_{1t,k} + B_{2t,k},$$

where $\mathbb{E}\left|B_{1t,k}\right| = O\left(\frac{1}{\sqrt{KT_0}}\right)$, $\mathbb{E}\left[B_{2t,k}\right] = \mathbb{E}\left[B_{3t,k}\right] = 0$, *and* $\mathbb{E}\left|B_{4t,k}\right| = O\left(\frac{1}{\sqrt{T_0}}\right)$.

Remark 2.3. There are two implications from Proposition 2.1. First, since the bias of the multiple-outcome synthetic control estimator shrinks to zero as either the number of pretreatment periods or the number of related outcomes increases, we can now use the synthetic control method to credibly estimate the treatment effect even when the number of pretreatment periods is small, if multiple related outcomes are available and that the treated unit can be closely approximated by the synthetic control in these outcomes in the pretreatment periods. Second, since the stochastic order of the bound on the bias of the multiple-outcome synthetic control estimator (RHS of B.1.7 in the proof) decreases in K, it follows that the bound for the multiple-outcome synthetic control estimator constructed by matching on $K \geq 2$ related outcomes has a smaller order of $(KT_0)^{-1/2}$ than the order of ${T_0}^{-1/2}$ for the conventional synthetic control estimator, which only uses a single outcome. Therefore, the multiple-outcome synthetic control method is also useful in cases with an abundance of pretreatment periods, especially if we worry about structural breaks over a long time span.

2.2.2 Empirical illustration

The multiple-outcome synthetic control method extends the applicability of the conventional single-outcome synthetic control method to cases where only a small

number of pretreatment periods are available. It would be reassuring if we could show, using real data, that the results produced by matching on multiple related outcomes in a few pretreatment periods are similar to those produced by matching on a single outcome in many pretreatment periods.

To illustrate the applicability of our method, we re-analyse the effect of the 1990 German reunification on West Germany's GDP per capita (Abadie et al., 2015). This example is ideal, because not only is the outcome of interest observed in many pretreatment periods, but also numerous related outcomes in the economic domain are available from OECD stat. For comparison, we construct two synthetic controls for West Germany, one by matching on the annual GDP per capita from 1960 to 1990, the other by matching on multiple related outcomes only in 1989.[3] The list of outcomes in 1989, as well as their values for West Germany, the two synthetic controls, and the simple average of the countries in the comparison group are summarised in Table 2.1. We see that both synthetic controls are generally much closer to West Germany in the outcomes, compared with the simple average of the other countries in the sample.

Table 2.1: Balance on Economic Outcomes in 1989

	West Germany	Synthetic Control (multiple outcomes)	Synthetic Control (single outcome)	Sample Mean
Private social expenditure	3.4	3.5	3.5	2.0
Energy supply per GDP	0.2	0.1	0.1	0.1
Electricity generation	9.0	8.7	9.8	7.6
Triadic patent families	0.1	0.1	0.0	0.0
Real GDP growth	3.9	4.1	3.5	3.5
CPI	2.8	3.1	4.1	5.5
Trade openness	57.7	59.1	57.6	60.4
Total tax revenue	36.2	34.1	33.0	33.7
GDP per capita	18 994.0	19 028.5	19 071.6	16 483

Note: This table compares the list of economic outcomes in 1989 for West Germany, the synthetic West Germany constructed using the listed outcomes in 1989, the synthetic West Germany constructed using GDP per capita from 1960 to 1990, and the simple average of the countries in the comparison group.

Figure 2.1 compares the trajectories of GDP per capita for West Germany and the two synthetic controls. As we would hope for, both synthetic controls are able to track West Germany's trajectory closely over a span of 30 years prior to the treatment. This is not surprising for the single outcome synthetic control, which is constructed with the aim to match the values of the outcome observed over the 30 years as closely as possible. The fact that the synthetic West Germany constructed using only outcomes in 1989 can track the trajectory of GDP per capita in West Ger-

[3]The results produced by matching on multiple related outcomes in any single year from 1985-1989 are very similar.

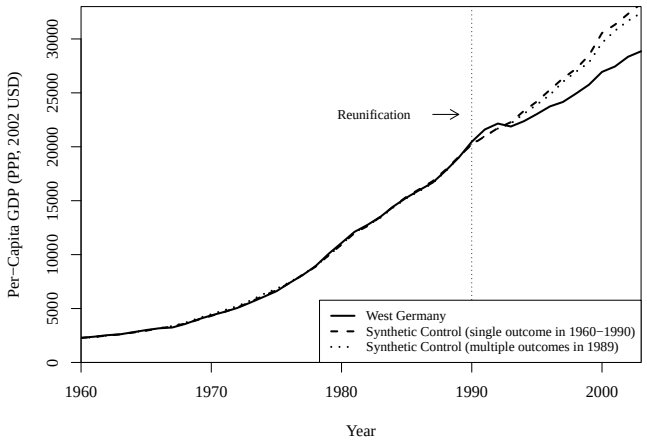


Figure 2.1: Re-analysis of the Economic Cost of the 1990 German Reunification

many so well for so long demonstrates the ability of the multiple-outcome synthetic control method to produce a synthetic control that closely approximates the treated unit in terms of the underlying predictors, even when the number of pretreatment periods is as small as 1. The estimated treatment effects, represented by the gaps between the trajectories of the realised outcome and the counterfactual outcomes, are also similar for the two synthetic control estimators after 1990, further providing credibility of our method.

2.2.3 Adjusting for differences in levels

The conventional single-outcome synthetic control method requires the treated unit to be or close to be in the convex hull of the control units in terms of the pretreatment matching variables. However, there are cases where the treated unit is extreme in the values of the matching variables, such that no convex combination of the control units can closely approximate the outcome of the treated unit in the pretreatment periods. In particular, it is often the case in practice that there are relatively stable differences in the level of the outcome across units before the treatment. In such cases, Ferman and Pinto (2019) and Abadie (2021) suggest constructing the synthetic control using demeaned outcomes, i.e., outcomes measured in differences with respect to their pretreatment means. This enables the synthetic control to track the dynamics in the outcome of the treated unit over time, while allowing the levels to differ by a constant amount, which is similar to the "parallel trends" assumption in the difference-in-differences method. Using demeaned outcomes is also similar to a proposal in Doudchenko and Imbens (2017), which includes an intercept when minimising the difference between the synthetic control and the treated unit in the matching variables. While their approach only accepts matching variables of the

same scale, there is no such restriction using demeaned outcomes.

Apart from allowing a better pretreatment fit for the treated unit, using demeaned outcomes has the additional merit that it helps correct the size distortion of the permutation test. Inference in the synthetic control method is based on the post-to-pretreatment RMSPE (root mean squared prediction error) ratios obtained from permuting the treatment status among all units.[4] As observed in Ferman and Pinto (2017), since the synthetic control method is only recommended when the treated unit can be well approximated by the synthetic control (Abadie et al., 2010, 2015), the RMSPE ratio in the permutation test is conditional on a good pretreatment fit for the treated unit while unconditional for the others, which would lead to over-rejection of the null hypothesis. Demeaning the outcomes alleviates the distortion in the size of the test, as it allows better pretreatment fits for all units in the permutation test, so that the asymptotic distributions of the RMSPE ratio for the treated unit and the control units are closer. Note that adjusting for differences in the levels may not fully correct the size distortion, since the dynamics in the outcomes over time are not guaranteed to be well approximated. In those cases, the RMSPE ratio may still tend to be conditional on a better pretreatment fit for the treated unit, and there may still be some degree of over-rejection.

Note that the interactive fixed effects model in equation (2.2) can be rewritten as

$$Y_{it,k}^0 = \rho_{i,k} + \dot{\delta}_{t,k} + \boldsymbol{Z}_i'\dot{\boldsymbol{\theta}}_{t,k} + \boldsymbol{\mu}_i'\dot{\boldsymbol{\lambda}}_{t,k} + \dot{\varepsilon}_{it,k}, \tag{2.6}$$

where $\rho_{i,k} = \frac{1}{T_0}\sum_{s=1}^{T_0}\delta_{s,k} + \boldsymbol{Z}_i'\frac{1}{T_0}\sum_{s=1}^{T_0}\boldsymbol{\theta}_{s,k} + \boldsymbol{\mu}_i'\frac{1}{T_0}\sum_{s=1}^{T_0}\boldsymbol{\lambda}_{s,k} + \frac{1}{T_0}\sum_{s=1}^{T_0}\varepsilon_{s,k}$, $\dot{\delta}_{t,k} = \delta_{t,k} - \frac{1}{T_0}\sum_{s=1}^{T_0}\delta_{s,k}$, $\dot{\boldsymbol{\theta}}_{t,k} = \boldsymbol{\theta}_{t,k} - \frac{1}{T_0}\sum_{s=1}^{T_0}\boldsymbol{\theta}_{s,k}$, $\dot{\boldsymbol{\lambda}}_{t,k} = \boldsymbol{\lambda}_{t,k} - \frac{1}{T_0}\sum_{s=1}^{T_0}\boldsymbol{\lambda}_{s,k}$, and $\dot{\varepsilon}_{it,k} = \varepsilon_{it,k} - \frac{1}{T_0}\sum_{s=1}^{T_0}\varepsilon_{is,k}$. In this model, $\rho_{i,k}$ represents the pretreatment level of outcome k for unit i, and the other components on the right-hand-side account for the dynamics in the outcome over time. Since the level of the outcome is determined by the underlying predictors, the bias in the synthetic control estimator and the size distortion in the permutation test, as a result of poor pretreatment fits due to the differences in the levels, can be removed by allowing negative synthetic control weights.[5] While the synthetic control estimator using demeaned outcomes is still asymptotically unbiased, it is less efficient in this case since the information on the level of the outcomes is lost. On the other hand, if $\rho_{i,k}$ were a free variable that does not depend on the underlying predictors, then the synthetic control estimator

[4]For more details about the inference procedure in the synthetic control method, see Abadie (2021).

[5]Relaxing the nonnegativity restriction on the synthetic control weights has been discussed in numerous studies since Doudchenko and Imbens (2017), but is beyond the focus of this paper.

using the original outcomes would be biased, whereas the process of demeaning the outcomes removes this bias.

In the multiple-outcome framework, the relative position of the units can vary across different outcomes, making it difficult to match on multiple outcomes simultaneously. The demeaning process would be helpful in these circumstances by improving the pretreatment fits. Denote the demeaned outcome as $\dot{Y}_{it,k} = Y_{it,k} - \frac{1}{T_0}\sum_{s=1}^{T_0} Y_{is,k}$.

Assumption 2.2′. *There exists a set of weights* $\left(w_2^*, \ldots, w_{J+1}^*\right)$ *such that* $w_j^* \geq 0$ *for* $j = 2, \ldots, J+1$, $\sum_{j=2}^{J+1} w_j^* = 1$, $\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j = \boldsymbol{Z}_1$ *and* $\sum_{j=2}^{J+1} w_j^* \dot{Y}_{jt,k} = \dot{Y}_{1t,k}$ *for all* $t \leq T_0$ *and* $k \in \mathbb{K}$.

We can then construct the multiple-outcome synthetic control estimator for $\tau_{1t,k}$ as

$$\tilde{\tau}_{1t,k} = \dot{Y}_{1t,k} - \sum_{j=2}^{J+1} w_j^* \dot{Y}_{jt,k} \tag{2.7}$$

to account for the differences in the level of the outcomes as long as $T_0 \geq 2$. Similarly with Proposition 2.1, the bias of this estimator can be shown to go to zero as the number of related outcomes or pretreatment periods goes to infinity.

Corollary 2.1. *Under Assumptions 2.1, 2.2′ and 2.3,* $\mathbb{E}\left(\tilde{\tau}_{1t,k} - \tau_{1t,k}\right) \to 0$ *as* $KT_0 \to \infty$.

2.2.4 Treatment effect on the untreated

We have been focusing on the setting with a single treated unit and many control units, where we estimate the treatment effects on the treated. However, there are cases with many treated units, and we may wish to estimate the treatment effects on the untreated.

Without loss of generality, suppose that unit 1 remains untreated within the window of observation, while all the other units are treated from $t = T_0 + 1$ onwards. Recall that the treated potential outcome is $Y_{it,k}^1 = Y_{it,k}^0 + \tau_{it,k}$. Since we have not imposed any assumption on the treatment effects except treating them as fixed given the sample, the treated potential outcomes may not have the interactive fixed effects functional forms or depend on the same predictors as the untreated potential outcomes. As a consequence, a synthetic unit that matches the untreated unit in the pretreatment matching variables may not credibly reproduce the counterfactual outcomes for the untreated unit after the treatment, even if it is similar to the untreated unit in the underlying predictors of the untreated potential outcomes. Therefore, in

order to estimate the treatment effects on the untreated, we include an additional assumption that the treatment effects are determined by the same predictors of the untreated potential outcomes. This assumption is more general than assuming that the treatment effects are constant.[6]

Assumption 2.4.

$$\tau_{it,k} = \alpha_{t,k} + \boldsymbol{Z}_i'\boldsymbol{\beta}_{t,k} + \boldsymbol{\mu}_i'\boldsymbol{\gamma}_{t,k}, \ \ \forall i, t, k,$$

where $\alpha_{t,k}$ is the time trend in $\tau_{it,k}$, and $\boldsymbol{\beta}_{t,k}$ and $\boldsymbol{\gamma}_{t,k}$ are the outcome-specific coefficients of the observed and unobserved predictors respectively.

Together with equations (2.1) and (2.2), this implies that the treated potential outcome k for unit i at time t is given by

$$Y_{it,k}^1 = (\delta_{t,k} + \alpha_{t,k}) + \boldsymbol{Z}_i' \left(\boldsymbol{\theta}_{t,k} + \boldsymbol{\beta}_{t,k}\right) + \boldsymbol{\mu}_i' \left(\boldsymbol{\lambda}_{t,k} + \boldsymbol{\gamma}_{t,k}\right) + \varepsilon_{it,k}. \tag{2.8}$$

The multiple-outcome synthetic control estimator for $\tau_{1t,k}$, the treatment effect on the untreated unit, can then be constructed as

$$\check{\tau}_{1t,k} = \sum_{j=2}^{J+1} w_j^* Y_{jt,k} - Y_{1t,k}. \tag{2.9}$$

Since the treated potential outcome $Y_{it,k}^1$ has an interactive fixed effects structure, we can similarly show that the bias of $\check{\tau}_{1t,k}$ vanishes when we have more and more pretreatment periods or related outcomes.

Corollary 2.2. *Under Assumptions 2.1, 2.2, 2.3 and 2.4,* $\mathbb{E}\left(\check{\tau}_{1t} - \tau_{1t}\right) \to 0$ *as* $KT_0 \to \infty$.

2.3 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to compare the multiple-outcome synthetic control method and the conventional single-outcome synthetic control method.

We fix the number of posttreatment periods at 1, and the sample size at 30, with a single treated unit and 29 control units. The treatment effect is set to 0, so that the treated potential outcomes are the same with the untreated potential outcomes. The potential outcomes are generated from the model in equation (2.2) with $r = 2$ and

[6] A similar assumption is discussed in Athey et al. (2021), where the treatment effect is assumed to have a low-rank pattern.

$f = 4$. The observed and unobserved predictors are drawn independently from the uniform distribution $U[-1, 1]$ for the control units, and $U[-d, d]$ with $d \in [0, 1]$ for the treated unit. When $d = 1$, the treated unit is equally likely to obtain extreme values in the outcomes as the control units. When d is smaller, the treated unit is more likely to be in the convex hull of the control units, and thus have better pretreatment fits.

Recall from (2.6) that the level of the outcome is given by $\gamma_{i,k} = \frac{1}{T_0}\sum_{s=1}^{T_0}\delta_{s,k} + \boldsymbol{Z}_i'\frac{1}{T_0}\sum_{s=1}^{T_0}\boldsymbol{\theta}_{s,k} + \boldsymbol{\mu}_i'\frac{1}{T_0}\sum_{s=1}^{T_0}\boldsymbol{\lambda}_{s,k} + \frac{1}{T_0}\sum_{s=1}^{T_0}\varepsilon_{is,k}$, which is determined by the mean time trend, coefficients and individual transitory shocks given the underlying predictors, while the dynamics over time is retained in the demeaned outcomes. To generate outcomes that are closer to real data, where there are often clear differences in the level of the outcomes across units and that the differences are relatively stable over time, we set the variance of the mean of the coefficients to be large relative to the variance of the coefficients and the transitory shocks. As such, the time trend and the coefficients are drawn independently from the normal distribution $N(\omega_k, 1)$ with $\omega_k \sim N(0, 10)$, and the transitory shocks are drawn independently from the standard normal distribution.[7]

In each simulation, the observed and unobserved predictors are drawn only once so that the outcomes share the same underlying predictors, while the time trend, the coefficients and the transitory shocks are drawn independently for each outcome. Figure 2.2 displays the trajectories of two related outcomes from a typical simulated sample with $d = 1$ and $T_0 = 10$. The trajectories for the treated unit are in black, and the control units in gray. As intended, there are visible and stable differences in the level of the outcomes, and the levels for the treated unit are different across the outcomes.

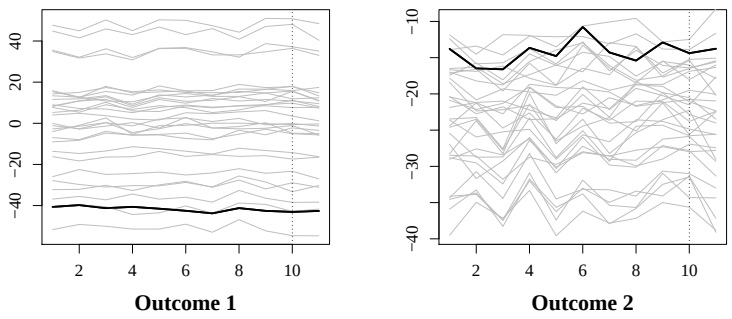


Figure 2.2: A Simulated Example

[7]When $\omega_k = 0$, there would be no distinguishable levels in the outcomes for different units, and demeaning the outcomes would not be useful in this case.

We compare the conventional single-outcome synthetic control estimator and the multiple-outcome synthetic control estimators constructed using $K = 1$, 2 and 4 demeaned outcomes, respectively. The only difference between the single-outcome synthetic control method and the multiple-outcome synthetic control method when $K = 1$ is the use of the demeaned outcomes. To measure their performances, we estimate the treatment effect on outcome 1 for the treated unit at $t = T_0 + 1$, and compute the average absolute bias and standard deviation of the estimators as well as the average rejection rate of the 10% test in 5000 simulations. The null hypothesis of zero treatment effect is rejected in each simulation, if the RMSPE ratio for the treated unit is ranked among the largest 10%, i.e., top 3 in our sample, in the permutation test. When the estimation improves with better pretreatment fits and larger numbers of pretreatment periods and related outcomes, we expect the average absolute bias and standard deviation of the estimators to be closer to $\sqrt{\frac{2}{\pi}} \approx 0.8$ and 1, which are the mean and standard deviation of the standard normal distribution folded at the mean (half-normal distribution).[8] When the distributions of the RMSPE ratio for the treated unit and the control units are close, so that there is little size distortion in the permutation test, we expect the average rejection rate of the 10% test to be close to the nominal rejection rate at 10%.

Table 2.2: Simulation

		Conventional SC			Multi-Outcome SC ($K = 1$)			Multi-Outcome SC ($K = 2$)			Multi-Outcome SC ($K = 4$)		
d	T_0	Bias	SD	Rej.	Bias	SD	Rej.	Bias	SD	Rej.	Bias	SD	Rej.
1	5	1.86	2.76	0.108	1.41	1.78	0.104	1.34	1.68	0.102	1.29	1.62	0.107
1	10	1.63	2.45	0.098	1.26	1.59	0.097	1.19	1.50	0.095	1.14	1.45	0.096
1	20	1.52	2.28	0.098	1.17	1.48	0.102	1.13	1.42	0.105	1.09	1.38	0.103
0.5	5	1.11	1.39	0.362	1.16	1.46	0.315	1.11	1.40	0.184	1.08	1.35	0.140
0.5	10	1.02	1.29	0.232	1.07	1.34	0.181	1.01	1.27	0.144	0.98	1.22	0.122
0.5	20	0.95	1.20	0.175	0.98	1.23	0.144	0.94	1.19	0.126	0.92	1.16	0.112
0	5	1.06	1.32	0.584	1.10	1.38	0.489	1.08	1.35	0.264	1.04	1.30	0.176
0	10	0.97	1.22	0.343	1.02	1.28	0.244	0.96	1.21	0.163	0.93	1.16	0.134
0	20	0.92	1.15	0.235	0.96	1.20	0.181	0.92	1.15	0.139	0.90	1.12	0.131

Note: This table compares the absolute bias, standard deviation, and rejection rate of the 10% test for the single-outcome SC estimator, and the multiple-outcome SC estimators constructed using 1, 2 and 4 demeaned outcomes respectively, with varying d and T_0, based on 5000 simulations for each setting.

Several findings emerge from the results of the simulations, which are reported in Table 2.2. First, when the number of pretreatment periods increases, the bias

[8]Note that the terms from the bias decomposition in the proof of Proposition 2.1 are all close to 0, except the posttreatment transitory shock, which follows a standard normal distribution in our simulation.

decreases and becomes closer to the expected value for all estimators as expected. The bias of the multiple-outcome synthetic control estimator also becomes smaller when the number of related outcomes is larger, albeit at a slower rate due to the loss of information in the level of the outcomes. Similar patterns are observed for the standard deviation of the estimators.

Second, when the support of the predictors for the treated unit is the same with that for the control units ($d = 1$), demeaning substantially improves estimation in terms of both bias and standard deviation, as the conventional single-outcome synthetic control method is likely to perform poorly when the treated unit is far from the convex hull of the control units, whereas demeaning adjusts for the differences in the levels and improves the pretreatment fit. Meanwhile, the rejection rate of the 10% test is close to the nominal size, with or without demeaning in this case, since the RMSPE ratio for the treated unit is not conditional on a good pretreatment fit. When d is smaller, the probability of obtaining a good pretreatment fit increases for the treated unit while staying unchanged for the other units. As a result, the bias and standard deviation for both the conventional synthetic control estimator and the demeaned synthetic control estimator are smaller, and the improvement in estimation by demeaning the outcomes becomes less pronounced. In contrast, the distortion in the size of the test increases drastically, and demeaning alleviates the size distortion by improving the pretreatment fits for all units.

Third, a larger number of pretreatment periods or related outcomes also reduces the size distortion, since the pretreatment RMSPE for the treated unit is less likely to be very close to 0. Overall, the results show that the multiple-outcome synthetic control method outperforms the conventional single-outcome method in terms of both estimation and inference, when there are multiple related outcomes with stable differences in the level of the outcomes.

2.4 Empirical Application

In order to curb the spread of COVID-19, most countries in Europe had implemented strict non-pharmaceutical interventions (NPIs), such as requiring residents to stay at home, restricting social gatherings and travels, as well as shutting down schools and workplaces, by late March. As an exception, Sweden opted against a general lockdown and implemented much lighter NPIs. For example, social distancing and working from home were advised but not mandated, bars, restaurants, and schools for children under 16 were kept open, quarantines for infected cases were not enforced, and facemasks were not recommended outside health care (Ludvigsson, 2020). Building upon the empirical strategy in Born et al. (2020) and Cho (2020),

we exploit this natural experiment to estimate the impacts of the NPIs on various public health, labour market and economic outcomes, using the multiple-outcome synthetic control method. To minimise the risk of structural breaks, we start the pretreatment periods from 2019. And to examine the dynamics of the impacts of the NPIs over time, we estimate the treatment effects in the first 3 quarters of 2020, before the second wave of COVID-19 cases in Europe from October.

2.4.1 Data

There are 26 countries in the treatment group for Sweden, including the other European Union members (excluding Cyprus, Luxembourg, and Malta due to their small sizes) as well as Norway, Switzerland and the United Kingdom.[9] The NPIs implemented in each country usually consist of a bundle of individual policies with varying duration and magnitudes. To compare the strictness of the NPIs across countries, we employ the Government Stringency Index, which is obtained from Our World in Data (Roser et al., 2020). This index is a composite measure of the strictness of government responses based on several individual metrics, and ranges from 0 to 100, with 100 representing the strictest response (Hale et al., 2020).

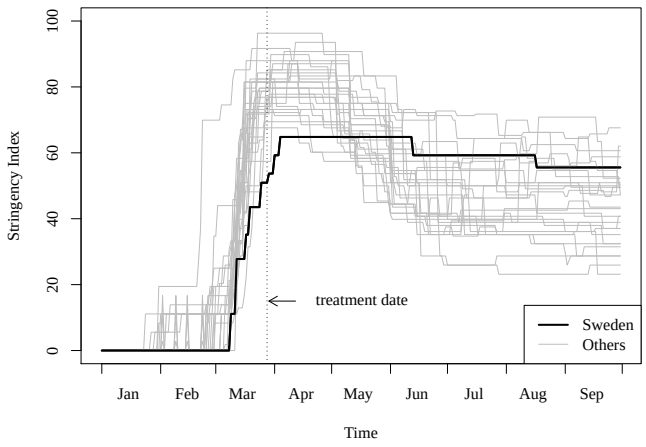


Figure 2.3: Stringency Index

Figure 2.3 depicts the stringency index for each country in our sample in the first 3 quarters of 2020, with Sweden shown in black and the others in gray. We see that most countries rapidly tightened their intervention policies in March, and kept

[9]The full list of countries in our sample are: Austria (AUT), Belgium (BEL), Bulgaria (BGR), Croatia (HRV), Czech Republic (CZE), Denmark (DNK), Estonia (EST), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Hungary (HUN), Ireland (IRL), Italy (ITA), Latvia (LVA), Lithuania (LTU), Netherlands (NLD), Norway (NOR), Poland (POL), Portugal (PRT), Romania (ROU), Slovakia (SVK), Slovenia (SVN), Spain (ESP), Sweden (SWE), Switzerland (CHE), and the UK (GBR).

them in place until May, when the strict NPIs began to be gradually eased. The stringency index of Sweden almost always stayed as the lowest in the sample during this period, consistent with the earlier observation that Sweden implemented much lighter NPIs than the other European countries. To derive the binary treatment status from the stringency index, we pick the average date that the stringency index peaked for each country in the treatment group, March 28th, as the treatment date, which is denoted by the vertical dotted line in Figure 2.3. One concern about this choice is that the NPIs have already started to be implemented prior to this date, albeit at lower levels compared with the peaks, and matching on outcomes that were realised after some NPIs have been in place may attenuate our estimates of the treatment effects. However, as we show in Section 2.4.4, the results of our analysis remain virtually the same even if we backdate the treatment to October 1, 2019, before the first case of COVID-19 in the world was reported.

We are interested in the impacts of the NPIs in three domains, namely, public health, the labour market, and the economy. In the public health domain, we examine 3 outcomes: the cumulative numbers of COVID-19 infection cases and deaths, and the number of weekly deaths from all causes. The numbers of COVID-19 cases and deaths may suffer from reporting issues such as measurement errors and time lags, due to differences in reporting standards in different countries and time constraints, whereas the number of deaths from all causes is more accurately recorded, but does not specify the exact causes. The data on COVID-19 cases and deaths are obtained from Our World in Data and are available on a daily basis, and the data on weekly deaths from all causes is obtained from Eurostat, the statistical office of the European Union.[10] We consider 3 labour market outcomes: employment rate, absence from work, and total hours worked. The employment rate is the percentage of employed persons in the total population. Absence from work is the percentage of employed persons that are temporarily absent from work, where persons absent from work are considered as employed if there is a formal attachment to the job in the form of continued receipt of wage or salary, or an assurance of return to work on an agreed date. And total hours worked are computed for all persons employed in their main occupation and are indexed to be equal to 100 in 2006 for purpose of comparability across countries.[11] The data on the labour market outcomes are seasonally adjusted and are available from Eurostat on a quarterly basis. Finally, we look at 6 outcomes in the economic domain: GDP, import, export, industrial production, retail sales, and CPI. The data on the economic outcomes are seasonally and calendar adjusted

[10]Since the numbers of COVID-19 cases and deaths are very small in the early stage of the pandemic, we use the number of COVID-19 cases only on days when its sample mean is larger than or equal to 1 per million population, and use the number of COVID-19 deaths only on days when its sample mean is larger than or equal to 0.1 per million population.

[11]See `https://ec.europa.eu/eurostat/cache/metadata/en/lfsi_esms.htm` for more details.

except for CPI (not adjusted), and are also available from Eurostat.[12] GDP is available on a quarterly basis, while the other economic outcomes are available on a monthly basis. As the outcomes are observed in different frequencies, the numbers of periods in which they are observed are also different.[13] Table 2.3 summarises the unit of measurement, the frequency of observation, the seasonal and calendar adjustment, as well as the numbers of pretreatment and posttreatment periods available for the outcomes of interest. Since we have an abundance of pretreatment variables to match on, we do not include any additional covariates.[14]

Table 2.3: Outcomes

Domain	Outcome	Unit	Frequency	Adjustment	Pretreatment	Posttreatment
public health	COVID-19 cases	per million population	daily	NA	28	213
	COVID-19 deaths	per million population	daily	NA	22	207
	all deaths	per million population	weekly	NA	64	27
labour market	employment	percent of population	quarterly	SA	4	3
	absence from work	percent of employment	quarterly	SA	4	2
	total hours worked	index (2006=100)	quarterly	SA	4	2
economy	GDP	2015 Euro per capita	quarterly	SCA	4	3
	import	2015 Euro per capita	monthly	SCA	14	7
	export	2015 Euro per capita	monthly	SCA	14	7
	industrial production	index (2015=100)	monthly	SCA	14	7
	retail	index (2015=100)	monthly	SCA	14	7
	CPI	index (2015=100)	monthly	NA	14	7

Note: This table summarises the outcomes of interest in our empirical application. 'NA' is short for 'Not Adjusted', 'SA' is short for 'Seasonally Adjusted', and 'SCA' is short for 'Seasonally and Calendar Adjusted'.

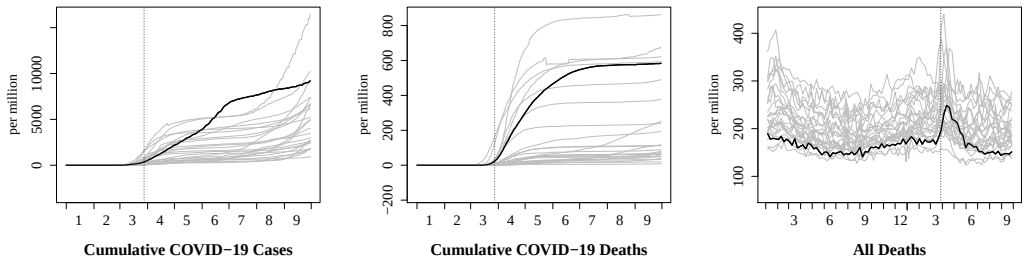
Figure 2.4 visualises the outcomes over time for all countries in our sample, with the trajectories for Sweden in black and the other countries in gray. Since the outcomes are observed in different frequencies, the vertical dotted lines are used as delimiters to visually separate the pretreatment and posttreatment periods for the outcomes, and may not sit on the treatment date exactly.[15] In the public health domain, we see that the numbers of COVID-19 cases and deaths began to rise quickly for countries in the treatment group from mid-March, before flattening out in early May. In comparison, although starting at lower levels and rates, the numbers of COVID-19

[12]The data on import and export for Norway, Switzerland and the United Kingdom are missing on Eurostat, and are thus obtained from OECD stat, where they are seasonally adjusted. GDP for Slovakia is also only seasonally adjusted.

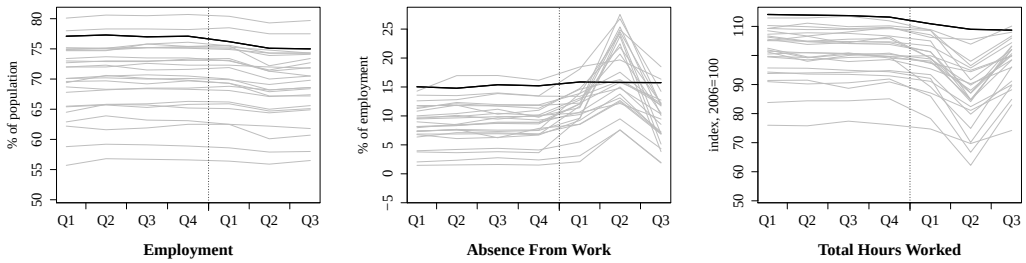
[13]Although the frequencies are different for different outcomes, they are similar for outcomes in the same domain. Using outcomes of different frequencies is similar to using linear combinations of the outcomes as in Abadie et al. (2010), e.g., quarterly data can be considered as averages of monthly data.

[14]See Botosaru and Ferman (2019) for a discussion on the role of the covariates in the synthetic control method.

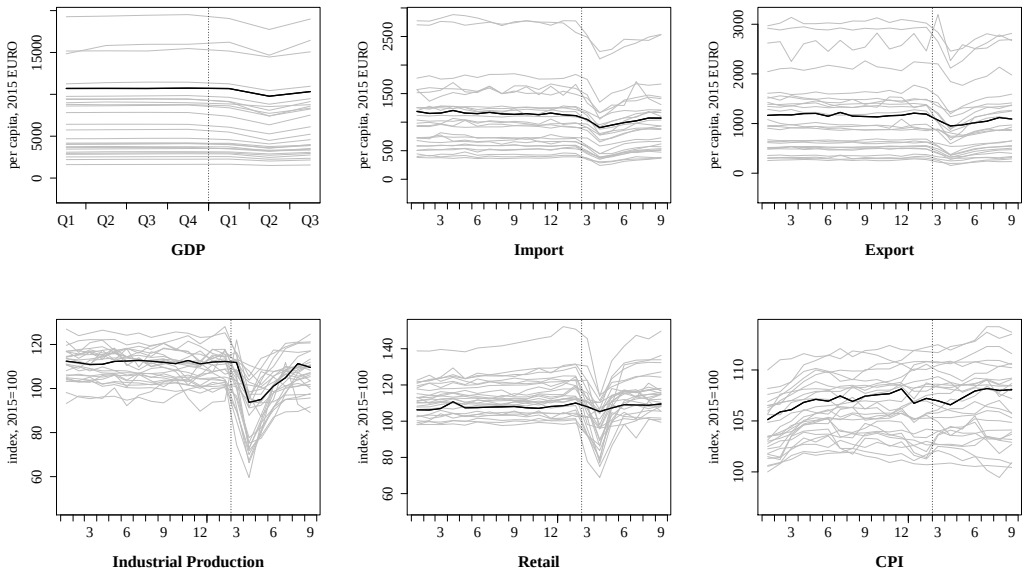
[15]Specifically, we only include the daily and weekly outcomes observed on and before Mar 28th, 2020, the monthly outcomes observed before March 2020, and the quarterly outcomes observed in 2019, in the pretreatment matching variables, none of which contain observations after the treatment date. The vertical dotted lines are positioned to reflect these choices.



(a) Public Health Outcomes



(b) Labour Market Outcomes



(c) Economic Outcomes

Figure 2.4: Descriptive Graphs

cases and deaths in Sweden continued growing to become among the highest in the sample, before slowing down only from June. Similarly, the number of weekly deaths from all causes in Sweden stayed near the bottom throughout 2019, but came close to the sample mean with a sharp spike in April, despite rises in other countries as well in this period, and only fell back to its usual level after June. The labour market outcomes present a more varied picture. Due to various protective measures to contain employment losses, the employment rate only experienced modest dips in the second and third quarters of 2020 for all countries in the sample. However, there were much more visible changes in absence from work and total hours worked among the employed. Compared with the 2019 levels, the percentage of the employed who were temporarily absent from work more than doubled, and the total hours worked dropped by about 10-20% in the second quarter for countries in the treatment group, and quickly returned to their normal levels in the third quarter when the NPIs were relaxed. In contrast, there were only very mild changes in these two outcomes for Sweden. When it comes to the economic domain, all outcomes except CPI were adversely impacted by the pandemic in the second quarter and somewhat recovered in the third. The changes were relatively modest in GDP, import and export, and more dramatic in industrial production and retail sales for countries in the treatment group. The economic outcomes in Sweden experienced similar changes, with the noticeable exception that the drop in the volume of retail sales was much smaller. It is also worth noting that the relative position for Sweden differs across the outcomes, highlighting the necessity of adjusting for the differences in the level of the outcomes through demeaning.

2.4.2 Estimation

As outcomes in different domains may depend on different sets of predictors, we construct a synthetic Sweden in each of the three domains, using a convex combination of the countries in the comparison group, to closely approximate the dynamics of the outcomes in the domain for Sweden before the treatment. Specifically, the synthetic control weights in domain $\mathbb{K}$ are chosen to minimise the distance between the synthetic Sweden and the actual Sweden in terms of the demeaned pretreatment outcomes in the domain as follows,

$$\left(w_2^*, \ldots, w_{J+1}^*\right)^{\mathbb{K}} = \arg \min_{(w_2,\ldots,w_{J+1})} \sum_{k=1}^{K} \frac{1}{\#_k^{\text{pre}}} \sum_{t=1}^{T_0} \left(\frac{\dot{Y}_{1t,k}^{\text{ob}} - \sum_{j=2}^{J+1} w_j \dot{Y}_{jt,k}^{\text{ob}}}{\sigma(\dot{Y}_{it,k}^{\text{ob}})} \right)^2$$
$$s.t. \ \sum_{j=2}^{J+1} w_j = 1 \ \text{ and } \ w_j \geq 0,$$

where Sweden is assigned as unit 1 and other countries as unit $2, \ldots, J+1$, $\#_k^{\text{pre}}$ is the number of pretreatment periods in which outcome k is observed, the upper dot in $\dot{Y}_{it,k}^{\text{ob}}$ represents that the outcome is demeaned, the superscript "ob" indicates that the outcome is used if observed in period t, and is replaced by 0 otherwise, and $\sigma(\cdot)$ computes the cross-sectional standard deviation of the variable. Hence, the outcomes in the domain are equally weighted, the values of the outcomes observed in different pretreatment periods are equally weighted within each outcome, and the matching variables are demeaned and standardised. Once the synthetic control weights $\left(w_2^*, \ldots, w_{J+1}^*\right)^{\mathbb{K}}$ for domain $\mathbb{K}$ are obtained, we can then estimate the treatment effect on outcome $k \in \mathbb{K}$ for Sweden at time t using $\widehat{\tau}_{1t,k} = \sum_{j=2}^{J+1} w_j^{*\mathbb{K}} \dot{Y}_{jt,k} - \dot{Y}_{1t,k}$.

Table 2.4: Synthetic Control Weights

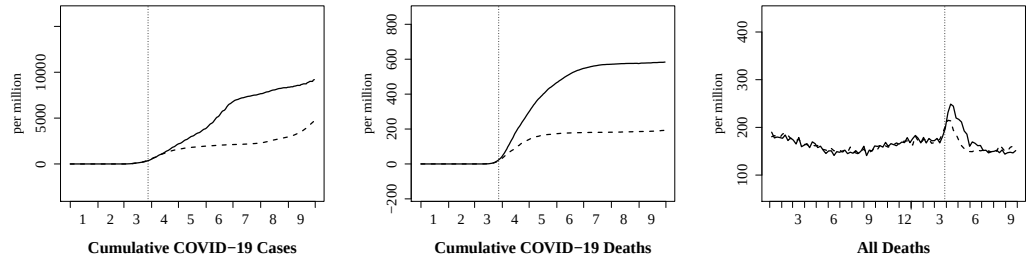
Country	Health	Labour	Economic	Country	Health	Labour	Economic
Austria	0	0	0.06	Italy	0.02	0	0.1
Belgium	0	0	0.08	Latvia	0	0	0.05
Bulgaria	0	0	0.09	Lithuania	0	0.18	0.07
Croatia	0	0	0.05	Netherlands	0.31	0.12	0
Czech Republic	0	0.03	0	Norway	0.07	0	0.1
Denmark	0.26	0	0	Poland	0.09	0	0
Estonia	0	0	0.01	Portugal	0	0	0
Finland	0.2	0.02	0.09	Romania	0	0	0
France	0.03	0.17	0	Slovakia	0	0.18	0
Germany	0	–	0	Slovenia	0	0	0
Greece	0.03	0	0	Spain	0	0.27	0
Hungary	0	0	0.06	Switzerland	0	0	0
Ireland	–	0.04	0.03	United Kingdom	0	0	0.21

Note: This table shows the synthetic control weights in each domain. – indicates that the country has missing data in the domain and is thus excluded from constructing the corresponding synthetic Sweden.

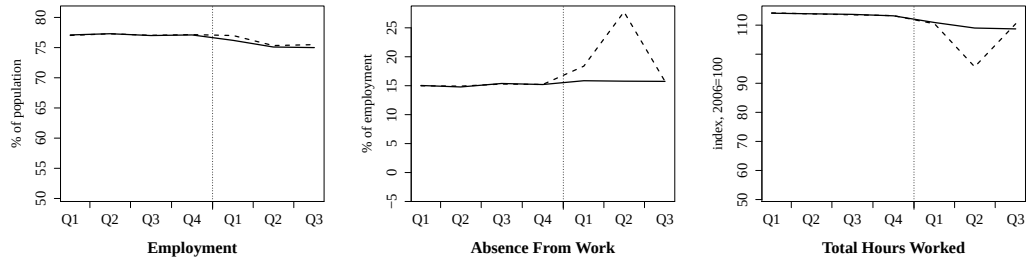
Table 2.4 displays the weights assigned to the countries in the treatment group for constructing the synthetic Sweden in each domain.[16] We see that the public health outcomes for Sweden are best approximated by a combination of the Netherlands, Denmark, Finland, Poland, Norway, France, Greece and Italy, with the weights in descending order. The labour market outcomes are best approximated by a combination of Spain, Lithuania, Slovakia, France, the Netherlands, Ireland, Czech Republic, and Finland. And the economic outcomes are best approximated by a combination of the UK, Italy, Norway, Bulgaria, Finland, Belgium, Lithuania, Austria, Hungary, Croatia, Latvia, Ireland, and Estonia.

Figure 2.5 compares the trajectories of each outcome for Sweden and the synthetic Sweden. From a visual inspection, the actual Sweden starts to accumulate more COVID-19 cases and deaths than the synthetic Sweden from April. The gaps be-

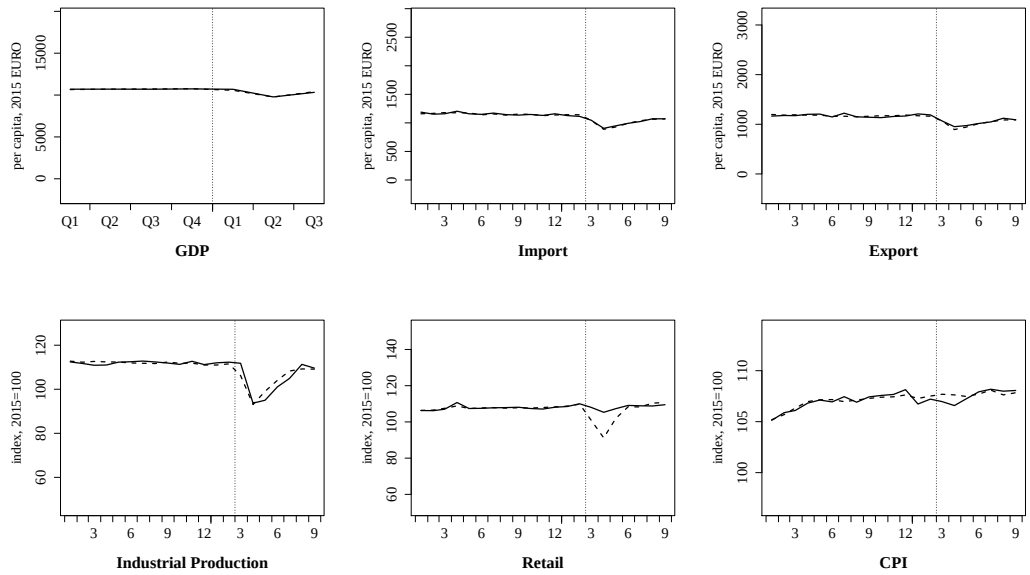
[16]Due to lack of data on weekly deaths, we exclude Ireland from the construction of synthetic control in the public health domain. Similarly, we exclude Germany from the labour market domain due to lack of data on absence from work and total hours worked in the posttreatment periods.



(a) Public Health Outcomes



(b) Labour Market Outcomes



(c) Economic Outcomes

Figure 2.5: Outcomes for Sweden (solid line) and the Synthetic Sweden (dashed line)

tween the two widen rapidly thereafter, before stabilising from July. Similarly, we observe more deaths from all causes in the actual Sweden than the synthetic Sweden from April to July, with the gap peaking in May. A back-of-the-envelope calculation suggests that had Sweden implemented stricter NPIs like the other European countries in March, the cumulative COVID-19 cases and deaths could have been reduced by about 5300 and 390 per million population respectively by July, amounting to 70% and 68% drops from the realised levels in Sweden, and there could have been 20% fewer weekly deaths from all causes in early May. For comparison, we also compute the cumulative deaths from all causes since April for both the synthetic Sweden and the actual Sweden, and find that the cumulative deaths from all causes could have been reduced by 364 per million population by July, representing an 11% drop from the realised level in Sweden.[17] Since COVID-19 is more easily contracted by people with existing conditions or weaker immune systems, and NPIs may have reduced the transmission of other diseases by promoting good hygiene behaviours and reducing face-to-face interactions, the estimated reduction in deaths from all causes, which amounts to a difference of 64% compared with the realised number of COVID-19 deaths, may serve as a lower bound for the reduction in COVID-19 deaths. As for the labour market outcomes, we find that stricter NPIs would increase absence from work among the employed by almost 76%, and reduce total hours worked by about 12%, in the second quarter of 2020. The impacts would disappear in the third quarter, and there would be no visible effect on the employment rate throughout the first 3 quarters in 2020. In terms of the economy, we find that stricter NPIs would shrink the volume of retail sales by 5%-13% from March to May, whereas the effects on the other economic outcomes including GDP, import, export, industrial production, and CPI, were all close to 0. This suggests that almost all of the contraction in the economy was due to the pandemic itself rather than the NPIs.

Since we are interested in the treatment effects on multiple outcomes in particular domains, we may follow Kling et al. (2007) and summarise the estimated treatment effects in domain $\mathbb{K}$ using the aggregate treatment effect, or the average of the standardised treatment effects as

$$\widehat{\tau}_i^{\mathbb{K}}(t_1, t_2) = \frac{1}{K} \sum_{k \in \mathbb{K}} \frac{\sum_{t=t_1}^{t_2} \widehat{\tau}_{it,k}^{\text{ob}}}{\#_k^{t_1,t_2} \sigma_k},$$

where $T_0 < t_1 \leq t_2 \leq T$, $\#_k^{t_1,t_2}$ is the number of periods that the estimated treatment effect on outcome k is available between t_1 and t_2, and σ_k is the average of the cross-

[17] Note that the percentage drop in deaths from all causes is much smaller than that in COVID-19 deaths, because the base number is much larger.

sectional standard deviations of outcome k over the posttreatment periods.

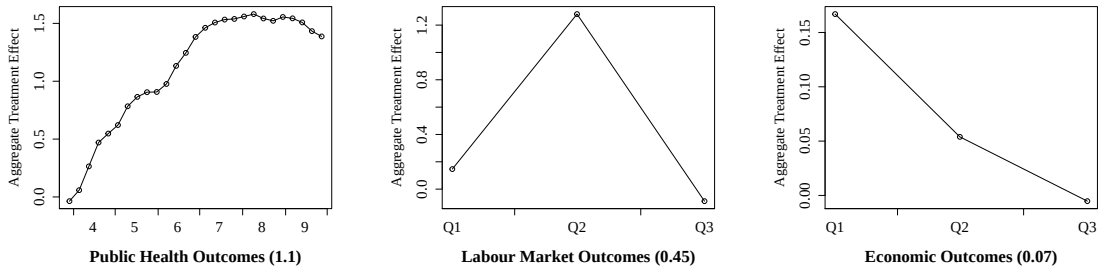


Figure 2.6: Aggregate Treatment Effects

Figure 2.6 summarises the aggregate treatment effects in different posttreatment periods in the three domains. The aggregate treatment effects in the first three quarters of 2020 overall are included in the parentheses. We see that the conclusions drawn for the three domains from the estimated treatment effects on each outcome are preserved using the aggregate treatment effects. The effect of the NPIs on the public health domain increased steadily from April to June, before leveling off from July, and the aggregate treatment effect in the posttreatment periods overall is also the largest on the public health domain. The effect on the labour market reached above 1.2 in the second quarter, but was close to zero in the first and third quarters. And the magnitude of the effect on the economy remained at very low levels throughout the first three quarters of 2020.

2.4.3 Inference

Given the common perception that the implementation of NPIs would benefit public health by reducing the spread of the virus, while harming the labour market and the economy, we test the null hypothesis for outcome k at $t > T_0$

$$H_0 : \tau_{1t,k} = 0,$$

against the alternative hypothesis

$$H_1 : \tau_{1t,k} < 0,$$

so that if the null hypothesis is rejected, then we can conclude that the implementation of strict NPIs has statistically significant negative effect on outcome k for Sweden.[18]

[18]We reverse the sign of the treatment effect on absence from work, so that it has the same expected direction with the other outcomes.

Following the one-sided inference procedure in Abadie (2021), we permute the treatment status among all the units, and compute the pretreatment RMSPE for unit i and outcome k as

$$R_{i,k}^{\text{pre}} = \left(\frac{1}{\#_k^{\text{pre}}} \sum_{t=1}^{T_0} \left(\widehat{Y}_{it,k}^{1,ob} - Y_{it,k}^{ob} \right)^2 \right)^{1/2},$$

and the posttreatment RMSPE as

$$R_{i,k}^{\text{post}} = \left(\frac{1}{\#_k^{\text{post}}} \sum_{t=T_0+1}^{T} \left[\left(\widehat{Y}_{it,k}^{1,ob} - Y_{it,k}^{ob} \right)^{-} \right]^2 \right)^{1/2},$$

where $\#_k^{\text{post}}$ is the number of posttreatment periods in which outcome k is observed, and $\left(\widehat{Y}_{it,k}^{1} - Y_{it,k} \right)^{-} = \widehat{Y}_{it,k}^{1,ob} - Y_{it,k}^{ob}$ if $\widehat{Y}_{it}^{1,ob} - Y_{it,k}^{ob} < 0$ and 0 otherwise.

The post-to-pretreatment RMSPE ratio is simply $r_{i,k} = R_{i,k}^{\text{post}} / R_{i,k}^{\text{pre}}$, and we can compute the p-value based on the ranking of $r_{i,k}$ as

$$p_k = \frac{1}{J+1} \sum_{i=1}^{J+1} I_+ \left(r_{i,k} - r_{1,k} \right),$$

where $I_+ ()$ is an indicator for nonnegative arguments. Replacing the posttreatment RMSPE, $R_{i,k}^{\text{post}}$, with $R_{it,k}^{\text{post}} = \left| \left(\widehat{Y}_{it,k}^{1} - Y_{it,k} \right)^{-} \right|$, $t > T_0$, allows us to compute the RMSPE ratio $r_{it,k}$ and the p-value $p_{t,k}$ in a single posttreatment period.

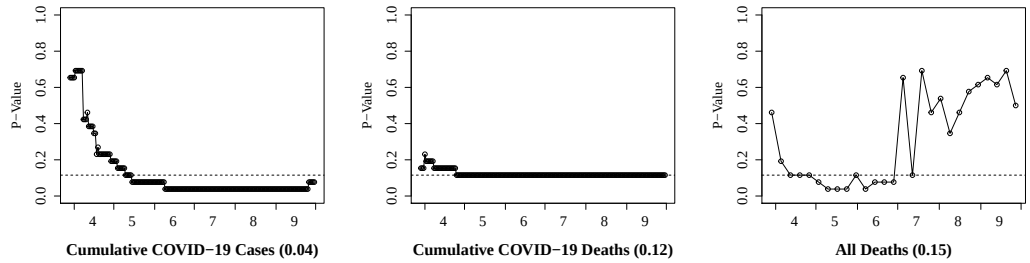
Note that when outcome k is observed in only a few periods before the treatment, the pretreatment RMSPE $R_{i,k}^{\text{pre}}$ may be very close to 0 and the RMSPE ratio may be extremely large for some units. For this reason, we can add a small value η on both the numerator and the denominator when computing the RMSPE ratio, so that the ratios are in reasonable ranges.[19] In the extreme case where we do not observe the outcome at all before the treatment, $R_{i,k}^{\text{pre}}$ would be 0 for all units, and ranking of the RMSPE ratios $\frac{R_{i,k}^{\text{post}}+\eta}{R_{i,k}^{\text{pre}}+\eta}$ becomes ranking of the gaps in the outcomes after the treatment.

Similarly to summarising the treatment effects in a domain using the aggregate treatment effect, we can summarise the statistical significance of the treatment effects in a domain in period t and in the posttreatment periods overall using

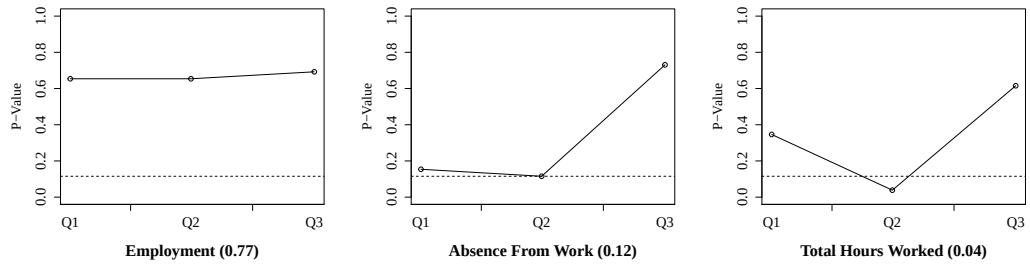
$$r_{it}^{\mathbb{K}} = \frac{\sum_{k \in \mathbb{K}} R_{it,k}^{\text{post}} / \sigma_k}{\sum_{k \in \mathbb{K}} R_{i,k}^{\text{pre}} / \sigma_k} \text{ and } r_{i}^{\mathbb{K}} = \frac{\sum_{k \in \mathbb{K}} R_{i,k}^{\text{post}} / \sigma_k}{\sum_{k \in \mathbb{K}} R_{i,k}^{\text{pre}} / \sigma_k},$$

[19] In this empirical application, we set $\eta = 0.01\sigma_k$, and our results are not sensitive to a larger η.

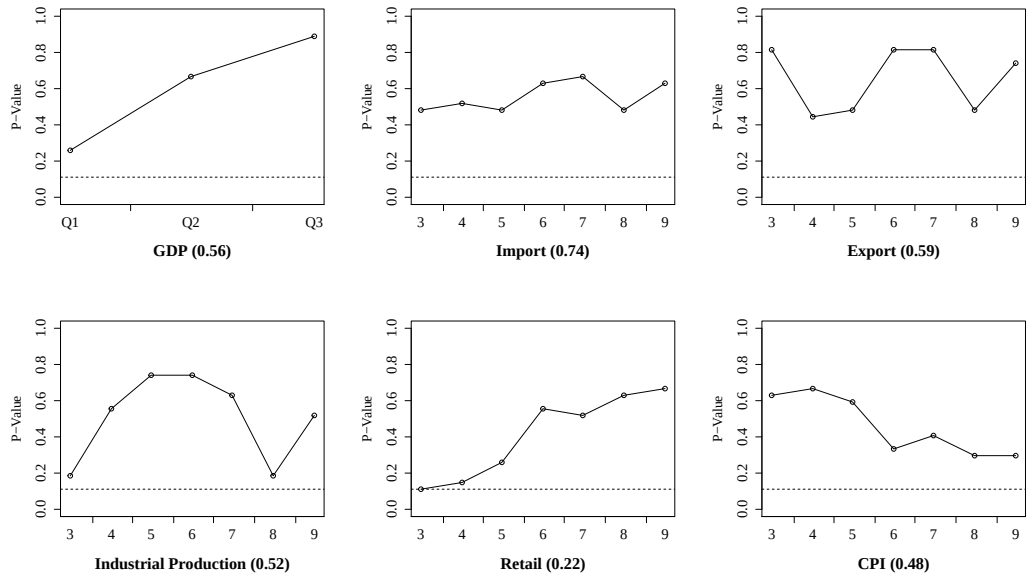
respectively. Aggregate p-values in period t and in the posttreatment periods overall can be computed accordingly based on these ratios.



(a) Public Health Outcomes



(b) Labour Market Outcomes



(c) Economic Outcomes

Figure 2.7: P-Values

Figure 2.7 shows the per-period p-values for each outcome, with the p-values in the posttreatment periods overall reported in the parentheses. The horizontal dotted line represents the significance level at $\alpha = 3/(J+1)$ (top 3 in the ranking of the post-to-pretreatment RMSPE ratios in the sample), with $J = 25$ ($\alpha \approx 0.12$) in the public health and labour market domains, and $J = 26$ ($\alpha \approx 0.11$) in the

economic domain. We find that the effects of the NPIs on COVID-19 cases and deaths were statistically significant at the 12% level from May, and the effect on deaths from all causes was significant from April to June. There were significant effects on absence from work and total hours worked in the second quarter, whereas the treatment effect on the employment rate remained insignificant throughout. As for the economic outcomes, we only find significant effect on retail sales in March, and we cannot reject the null hypotheses that the NPIs had no effects on the other economic outcomes in the first three quarters of 2020. The detailed graphs showing the gaps between each country and their synthetic counterparts in the permutation test are in Figure B.2.1 in the Appendix.

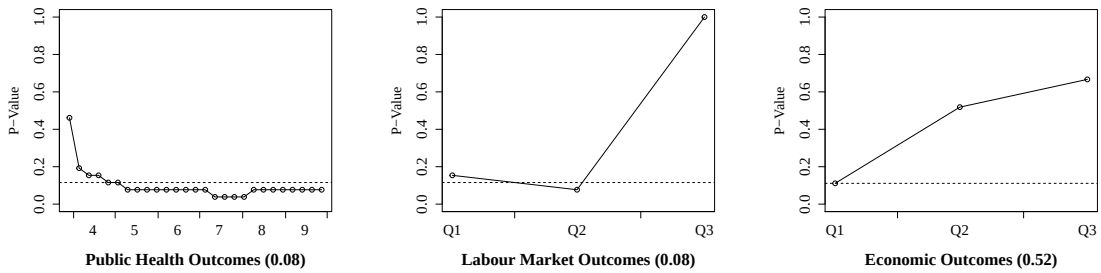


Figure 2.8: Aggregate P-Values

Aggregate p-values for the treatment effects on each domain in the posttreatment periods are presented in Figure 2.8, where the aggregate p-values in the posttreatment periods overall are reported in the parentheses. We find that the effect of the NPIs on public health was statistically significant from May onwards, the effect on the labour market was only significant in the second quarter, and the effect on the economy was only significant in the first quarter due to the effect on retail sales in March. Rankings of the post-to-pretreatment RMSPE ratios for the individual outcomes as well as the three domains are shown in Figures B.2.2 and B.2.3 in the Appendix.

2.4.4 Robustness Checks

In this section, we conduct several robustness checks to assess the sensitivity of our results to changes in the design of the study. Since we have multiple outcomes of interest, we can summarise the results succinctly using the radar charts to save space for comparing results from different specifications.[20]

As an example, Figure 2.9 summarises the results in the benchmark specification, where the outcomes for Sweden are connected with solid lines, and the outcomes

[20]The full detailed graphs showing the outcomes for Sweden and the synthetic Sweden in all periods are in Figures B.2.4, B.2.5, B.2.6, B.2.7 and B.2.8 in the Appendix.

for the synthetic Sweden are connected with dashed lines. The outermost polygon represents the maximum values for each outcome and the innermost polygon represents the minimums. The outcomes for Sweden and the synthetic Sweden at different stages are presented in the two charts, where the left chart displays the outcomes averaged over the pretreatment periods, and the right chart shows the outcomes averaged in the second quarter of 2020, when the impacts of the NPIs were most profound. We see that the average outcomes for Sweden and the synthetic Sweden overlap perfectly in the pretreatment periods. In comparison, there are visible differences between the two in the public health outcomes as well as total hours worked and absence from work, in the second quarter of 2020. By and large, the main results from Figure 2.5 are preserved in the radar charts.

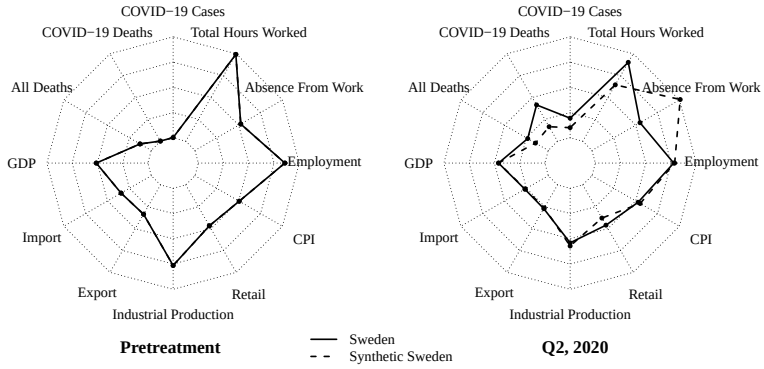


Figure 2.9: Radar Chart of Outcomes (Benchmark)

Backdating

We pick March 28, 2020, the average date on which the stringency index peaked for each country in the treatment group, as the treatment date. As mentioned earlier, this choice may lead to attenuated estimates of the treatment effects if the outcomes were affected by the treatment before the chosen date either because countries had started implementing the NPIs before this date, or because individuals had anticipated the implementation of the NPIs and had acted in advance.

In this exercise, we backdate the treatment to October 1, 2019, to see whether we can obtain results that are similar to those in the benchmark specification. This alternative date was before the first case of COVID-19 infection in the world was detected, and almost 4 months before the first case was reported in Europe, so that we can be confident that the outcomes observed before this date were by no means contaminated. In addition, the backdating exercise is also useful to evaluate the credibility of the synthetic control estimator by assessing whether the synthetic Sweden can reproduce the outcomes of the actual Sweden in the absence of the treatment.

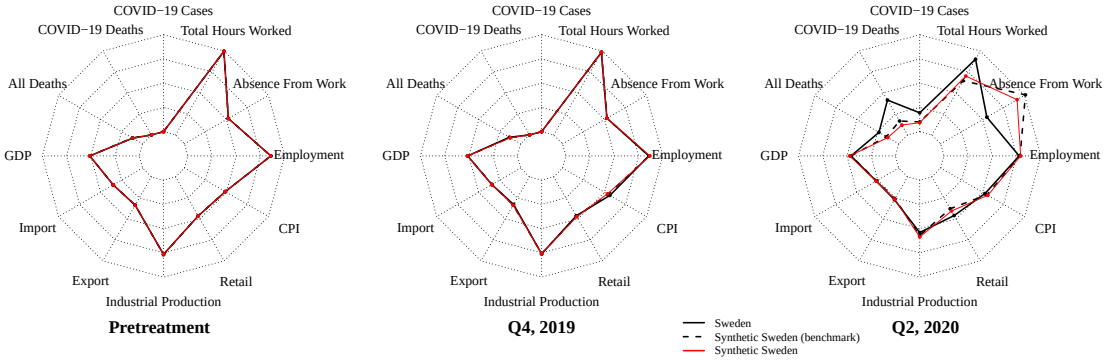


Figure 2.10: Radar Chart of Outcomes (Backdating)

Figure 2.10 compares the synthetic Sweden constructed by matching only on outcomes observed before the placebo treatment date, October 1, 2019, with the actual Sweden at three different stages. The left figure shows the outcomes averaged in the first three quarters of 2019, before the placebo treatment date, the middle figure shows the outcomes in the fourth quarter of 2019, which was after the placebo treatment date but before the actual treatment date, and the right figure shows the outcomes in the second quarter of 2020, after the actual treatment assignment. We see that the synthetic Sweden is very close to the actual Sweden not only in the first chart, but in the second chart as well. The absence of estimated effects before the actual treatment date shows that the synthetic Sweden can reliably reproduce the untreated potential outcomes of the actual Sweden. The differences between the two emerges in the third figure and the results are almost identical to our benchmark results. The close resemblance of our benchmark results and the results produced by backdating the treatment to almost 6 months earlier provides compelling evidence that our estimates for the treatment effects are credible. Note that we would not have been able to do this had we been focusing on estimating the effect of the NPIs on COVID-19 infection cases or deaths using the conventional single-outcome synthetic control method, like in Born et al. (2020) or Cho (2020), since COVID-19 cases and deaths were not observed before this placebo treatment date.[21]

This exercise, along with the replication exercise in Section 2.2.2, highlights one of the advantages of the multiple-outcome synthetic control method over the conventional single-outcome synthetic control method, i.e., our method allows the users to construct a credible synthetic control when we only observe the outcomes in very few pretreatment periods, or in the extreme case, when we do not observe the outcome of interest before the treatment at all, as long as the synthetic control can closely approximate the unit of interest in multiple related outcomes before the

[21] Cho (2020) backdates the treatment by 3 days in one of the robustness checks, which may not be sufficient to address the concern that the pretreatment outcome may have been affected by the treatment.

treatment.

Leave-one-unit-out

To check if our results are sensitive to the choice of countries in constructing the synthetic Sweden, we conduct the leave-one-unit-out re-analysis, where we iterate the estimation procedure, excluding one of the countries that received positive weights from the construction of the synthetic Sweden at a time.

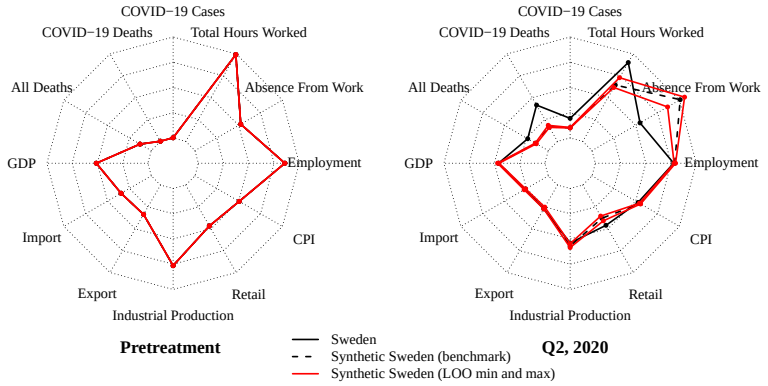


Figure 2.11: Radar Chart of Outcomes (Leave-One-Unit-Out)

The results are presented in Figure 2.11, where the gray lines represent the boundaries for the distributions of the leave-one-out estimates. We see that the leave-one-out estimates closely centre around the benchmark estimates represented by the dashed lines, albeit with slightly larger dispersions for retail sales and absence from work, showing that our results are robust to the exclusion of any particular country.

Leave-one-outcome-out

Since our method constructs the synthetic Sweden by matching on multiple outcomes in a domain, we also conduct a leave-one-outcome-out exercise to check if our results are sensitive to the exclusion of any particular outcome from the construction of the synthetic Sweden.

We see that the leave-one-outcome-out estimates closely centre around the benchmark, despite somewhat larger dispersions for the labour market outcomes due to the small number of pretreatment periods available. This shows that our results are robust to the exclusion of any particular outcome.

Single outcomes

Instead of matching on multiple related outcomes simultaneously, we can construct the synthetic controls by matching on each outcome separately, although the relia-

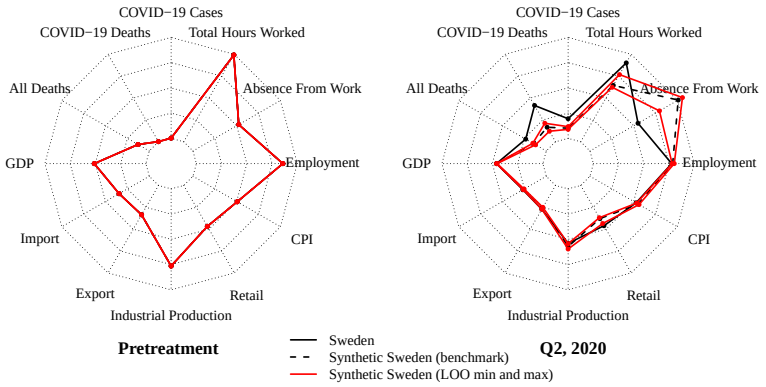


Figure 2.12: Radar Chart of Outcomes (Leave-One-Outcome-Out)

bility of the estimates may be questionable if the number of pretreatment periods is small. The results are presented in Figure 2.13, which shows that despite small changes in the magnitudes, constructing the synthetic controls by matching on single outcomes does not fundamentally change our conclusions.

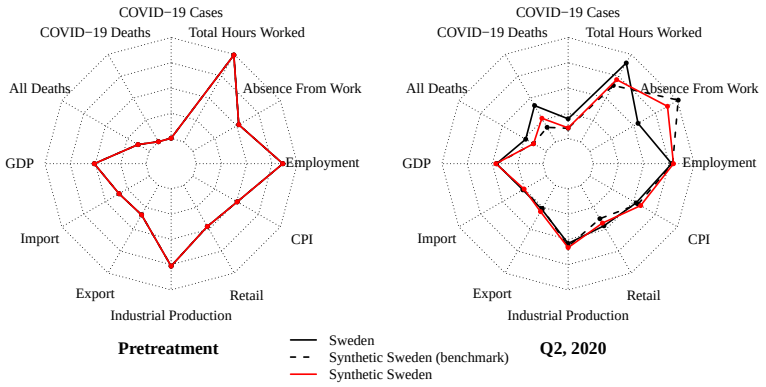


Figure 2.13: Radar Chart of Outcomes (Single Outcome)

No demeaning

Our estimates in the benchmark specification are obtained using demeaned outcomes to account for the differences in the level of the outcomes. In particular, the values of the outcomes are extreme for Sweden in the labour market domain, in which case the synthetic control method using the original outcomes would not be able to provide credible estimates of the treatment effects due to poor pretreatment fits. Matching on multiple outcomes may exacerbate the problem in this case as it is more difficult to obtain a good fit on multiple outcomes. Note that this does not invalidate the multiple-outcome synthetic control method, but rather highlights the importance of taking appropriate measures to obtain a good pretreatment fit. Bearing this in mind, we check whether the results produced by matching on the original outcomes differ dramatically from the benchmark results, especially in the

labour market domain.

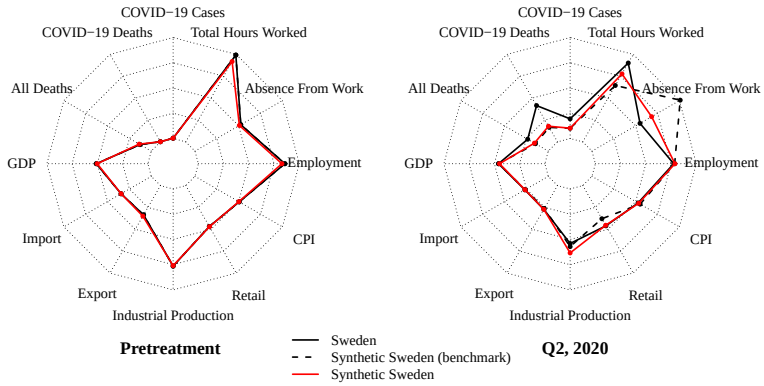


Figure 2.14: Radar Chart of Outcomes (No Demeaning)

Figure 2.14 compares the outcomes for Sweden and the synthetic Sweden constructed using the original outcomes. Note that the synthetic Sweden is not able to match the pretreatment means of the labour market outcomes for Sweden, indicating that the pretreatment fit is poor for those outcomes, which is more visible in Figure B.2.8. The other outcomes are matched reasonably well in the pretreatment periods. The estimates in the second quarter of 2020 are not dramatically different from the results in the benchmark specification (gray lines), except for absence from work, for which the estimated effect is not reliable due to the poor pretreatment fit.

2.4.5 Further Analysis

In this section, we conduct subgroup analysis when data permits, to investigate whether there is heterogeneity in the effects of the NPIs across different groups in the population or different industries in the economy.

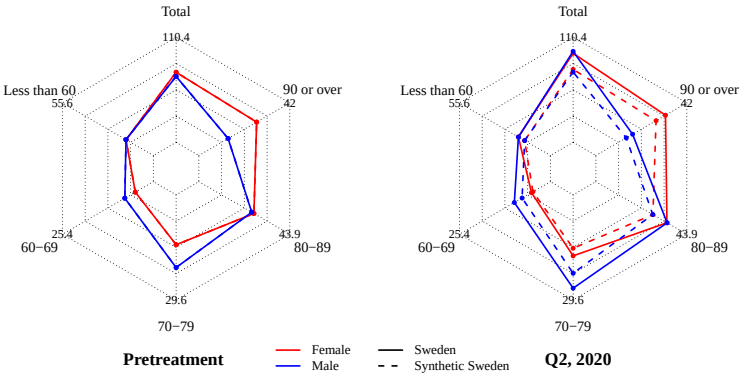


Figure 2.15: Deaths by Gender and Age

Figure 2.15 shows deaths from all causes in different gender and age groups for Sweden and the synthetic Sweden, where the outcomes for female are in red, male

in blue, and the numbers in the peripheral axis labels are the maximums in each age group. We see that the synthetic Sweden is very close to the actual Sweden in the number of deaths in different gender and age groups before the treatment, but there are differences between the two in all subgroups in the second quarter of 2020. The comparison suggests that the impact of the NPIs was larger for males in the 60-79 age group, and larger for females in the group with age 90 or over, possibly as a result of differences in lifestyles or risk preferences.

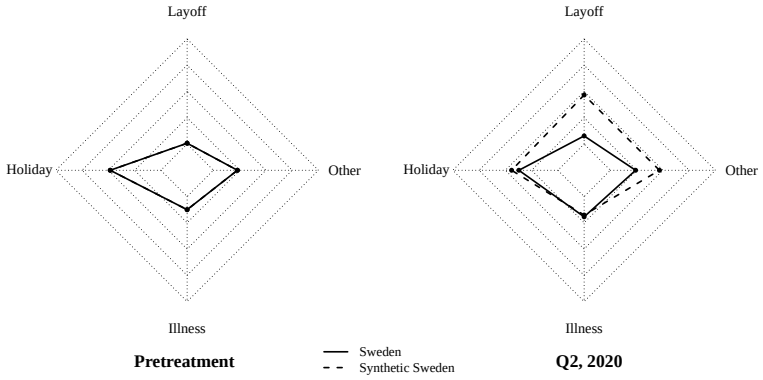


Figure 2.16: Reasons of Absence From Work

Figure 2.16 compares the reasons of absence from work for Sweden and the synthetic Sweden.[22] We see that absence from work was mainly due to holidays and very rarely due to temporary layoffs before the treatment. As a result of the NPIs, absence from work due to temporary layoffs would quintuple and absence from work due to other reasons would almost double in the second quarter of 2020, whereas absence from work due to holiday and illness would not be affected.

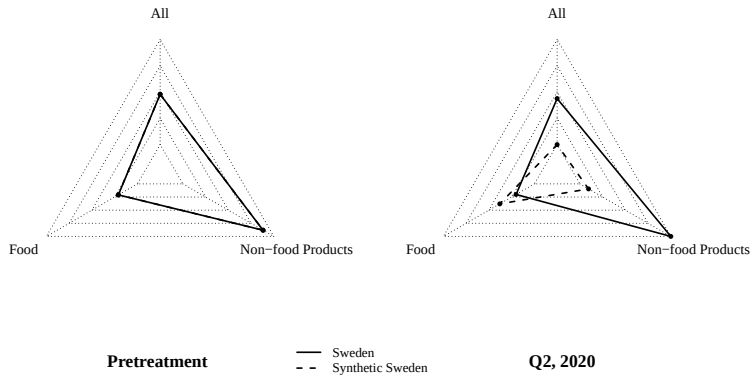


Figure 2.17: Retail Sales

Figure 2.17 shows the volume of retail sales in all products, food products (including

[22]There are missing values in 'temporary layoffs' and 'other reasons' for some countries that received positive weights in the construction of the synthetic Sweden in the labour market domain. These missing values are imputed using the difference between the total percentage of absence from work and the percentages due to the observed reasons.

food, beverages and tobacco) and non-food products (except fuel). We see that the reduction in retail sales due to the NPIs would be mainly through reduced sales of non-food products. Examining in more details from Figure B.2.11, we find a drop of 9%-26% in the sales of non-food products from March to May, and very mild changes in the sales of food products, resulting in a drop of about 5%-13% in the total volume of retail sales from March to May.

We also examine the effects of the NPIs on the labour market outcomes by gender, the effect on the employment rate by occupation, education and age, as well as the effects on the individual components of GDP and CPI. However, we do not find any heterogeneity in the treatment effects for these subgroups. These results are in Figures B.2.12-B.2.17 in the Appendix.

2.5 Conclusion

This paper generalises the conventional single-outcome synthetic control method to a multiple-outcome framework, where the number of pretreatment periods is supplemented with the number of related outcomes in the domain, making the method applicable even when the number of pretreatment periods is small or if we worry about structural breaks over a longer time span. Following Abadie et al. (2010), we show that the bound on the bias of the multiple-outcome synthetic control estimator is of a smaller stochastic order than that of the single-outcome synthetic control estimator, when the synthetic control can closely approximate the unit of interest in terms of the observed predictors and the multiple related outcomes. We also discuss the role of demeaning the outcomes before constructing the synthetic control, which is to account for the differences in the level of the outcomes for different units, and show in simulation that using demeaned outcomes can reduce both the bias and the variance of the synthetic control estimator and alleviate the size distortion of the permutation test, if there are relatively stable differences in the level of the outcomes.

We move on to evaluate the effects of the non-pharmaceutical interventions on various outcomes in the public health, labour market, and economic domains using the multiple-outcome synthetic control method, where we construct a synthetic Sweden in each domain using the other European countries that implemented much stricter NPIs. We find that the NPIs would significantly reduce the cumulative numbers of COVID-19 cases and deaths as well as deaths from all causes, increase temporary absence from work and reduce total hours worked among the employed, but would have limited impacts on the employment rate and the economy, other than shrinking the volume of retail sales in the early stage.

There are several limitations in our empirical analysis. First, our conclusions are limited to the outcomes that are available for our analysis, and may not apply to other outcomes in the domain. For example, our finding that the NPIs had beneficial effects on reducing COVID-19 cases and deaths does not necessarily imply that they had positive impacts on other outcomes in the public health domain, say mental health. Similarly, the effects of the NPIs are estimated for Sweden in a case study context, and may not directly apply to other countries. Second, the data on COVID-19 cases and deaths may suffer from reporting issues such as measurement errors or time lags, due to differences in the reporting standards in different countries and time constraints. This issue is alleviated to some extent in our analysis by matching on deaths from all causes, which is more accurately measured. Third, the complexities in the NPIs due to different timings and contents in different countries are not likely to be accurately or fully captured by the stringency index or the binary treatment indicator. It would be worthwhile to conduct more detailed analysis on the effects of the individual policies, as in Castex et al. (2021) and Chernozhukov et al. (2021). Fourth, there may be spill-over effects. For example, risk-loving individuals from other European countries might have fled to Sweden before the travel restrictions were officially implemented in their countries, which would drive the number of COVID-19 cases up in Sweden, and cause the public health effects of the NPIs to be over-estimated. Note that this is different from the voluntary social distancing behaviours, which does not invalidate our results. To see this, if all individuals voluntarily practice social distancing to the same degree as if there were NPIs, then we would find no effect of the NPIs, and this is not a biased estimate since the true effect is 0 in this extreme case. Similar to the spill-over effects across the countries, there may be other policies implemented at the same time that could confound the estimation. This would be less of a problem if these policies were implemented across different countries with similar contents and magnitudes, which is the case for the employment support programs that were carried out in many European countries.

Appendix B

B.1 Proofs

Proof of Proposition 2.1. The proof follows closely the proof in Appendix B of Abadie et al. (2010).

Under the restrictions $\sum_{j=2}^{J+1} w_j^* \boldsymbol{Z}_j = \boldsymbol{Z}_1$, we have

$$\begin{aligned} e_{1t,k} &\equiv Y_{1t,k}^0 - \sum_{j=2}^{J+1} w_j^* Y_{jt,k} \\ &= \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} w_j^* \boldsymbol{\mu}_j \right)' \boldsymbol{\lambda}_{t,k} + \varepsilon_{1t,k} - \sum_{j=2}^{J+1} w_j^* \varepsilon_{jt,k}. \end{aligned} \tag{B.1.1}$$

Stacking the pretreatment outcomes $\boldsymbol{Y}_{it,k}$ over the T_0 pretreatment periods, we have

$$\boldsymbol{Y}_{i,k} = \boldsymbol{\delta}_k + \boldsymbol{\theta}_k \boldsymbol{Z}_i + \boldsymbol{\lambda}_k \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{i,k}, \tag{B.1.2}$$

where $\boldsymbol{Y}_{i,k}$, $\boldsymbol{\delta}_k$ and $\boldsymbol{\varepsilon}_{i,k}$ are $T_0 \times 1$, and $\boldsymbol{\theta}_k$ and $\boldsymbol{\lambda}_k$ are $T_0 \times r$ and $T_0 \times f$, respectively. Since the K outcomes are determined by the same set of predictors in our multiple-outcome framework, we can further stack (B.1.2) over the K outcomes to get

$$\boldsymbol{Y}_i = \boldsymbol{\delta} + \boldsymbol{\theta} \boldsymbol{Z}_i + \boldsymbol{\lambda} \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i, \tag{B.1.3}$$

where $\boldsymbol{Y}_i$, $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}_i$ are $KT_0 \times 1$, and $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are $KT_0 \times r$ and $KT_0 \times f$, respectively. The restrictions $\sum_{j=2}^{J+1} w_j^* \boldsymbol{Y}_j = \boldsymbol{Y}_1$ can be simplified to

$$\boldsymbol{\lambda} \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} w_j^* \boldsymbol{\mu}_j \right) = \sum_{j=2}^{J+1} w_j^* \boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}_1. \tag{B.1.4}$$

Assumption 2.3 states that the $f \times f$ matrix $\boldsymbol{\lambda}'\boldsymbol{\lambda}$ has full rank, thus pre-multiplying

$(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'$ on both sides of (B.1.4), we have

$$\left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} w_j^* \boldsymbol{\mu}_j\right) = (\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'\left(\sum_{j=2}^{J+1} w_j^* \boldsymbol{\varepsilon}_j - \boldsymbol{\varepsilon}_1\right), \tag{B.1.5}$$

so that (B.1.1) can be written as

$$\begin{aligned} e_{1t,k} =& \boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'\sum_{j=2}^{J+1} w_j^*\boldsymbol{\varepsilon}_j && (B_{1t,k}) \\ & - \boldsymbol{\lambda}_{t,k}'(\boldsymbol{\lambda}'\boldsymbol{\lambda})^{-1}\boldsymbol{\lambda}'\boldsymbol{\varepsilon}_1 + \varepsilon_{1t,k} - \sum_{j=2}^{J+1} w_j^*\varepsilon_{jt,k}. && (B_{2t,k}) \end{aligned}$$

Whereas $B_{2t,k}$ has zero mean given Assumption 2.1, $B_{1t,k}$ does not because w_j^* is a function of $\boldsymbol{\varepsilon}_j$ (Botosaru and Ferman, 2019).

We can rewrite $B_{1t,k}$ as

$$B_{1t,k} = \sum_{j=2}^{J+1} w_j^* \sum_{q=1}^{K}\sum_{s=1}^{T_0} \boldsymbol{\lambda}_{t,k}'\left(\sum_{l=1}^{K}\sum_{n=1}^{T_0}\boldsymbol{\lambda}_{n,l}\boldsymbol{\lambda}_{n,l}'\right)^{-1}\boldsymbol{\lambda}_{s,q}\varepsilon_{js,q}. \tag{B.1.6}$$

Let the largest element of $|\boldsymbol{\lambda}_{t,k}|$ for $t = 1,\ldots,T$ and $k = 1,\ldots,K$ be bounded from above by $\bar{\lambda}$. Under Assumption 2.3 and using the Cauchy–Schwarz Inequality, we have

$$\begin{aligned} & \left(\boldsymbol{\lambda}_{t,k}'\left(\sum_{l=1}^{K}\sum_{n=1}^{T_0}\boldsymbol{\lambda}_{n,l}\boldsymbol{\lambda}_{n,l}'\right)^{-1}\boldsymbol{\lambda}_{s,q}\right) \\ \leq & \left(\boldsymbol{\lambda}_{t,k}'\left(\sum_{l=1}^{K}\sum_{n=1}^{T_0}\boldsymbol{\lambda}_{n,l}\boldsymbol{\lambda}_{n,l}'\right)^{-1}\boldsymbol{\lambda}_{t,k}\right)^{\frac{1}{2}}\left(\boldsymbol{\lambda}_{s,q}'\left(\sum_{l=1}^{K}\sum_{n=1}^{T_0}\boldsymbol{\lambda}_{n,l}\boldsymbol{\lambda}_{n,l}'\right)^{-1}\boldsymbol{\lambda}_{s,q}\right)^{\frac{1}{2}} \\ \leq & \left(\frac{\bar{\lambda}^2 f}{KT_0\underline{\xi}}\right). \end{aligned}$$

Let $\bar{\varepsilon}_j = \sum_{q=1}^{K}\sum_{s=1}^{T_0}\boldsymbol{\lambda}_{t,k}'\left(\sum_{l=1}^{K}\sum_{n=1}^{T_0}\boldsymbol{\lambda}_{n,l}\boldsymbol{\lambda}_{n,l}'\right)^{-1}\boldsymbol{\lambda}_{s,q}\varepsilon_{js,q}$. Then by Hölder's Inequality and the norm monotonicity, we have

$$|B_{1t,k}| \leq \sum_{j=2}^{J+1} w_j^*|\bar{\varepsilon}_j| \leq \left(\sum_{j=2}^{J+1}|w_j^*|^q\right)^{1/q}\left(\sum_{j=2}^{J+1}|\bar{\varepsilon}_j|^p\right)^{1/p} \leq \left(\sum_{j=2}^{J+1}|\bar{\varepsilon}_j|^p\right)^{1/p},$$

with $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$.

Using Hölder's Inequality again, we have

$$\mathbb{E}\left[\sum_{j=2}^{J+1}|\bar{\varepsilon}_j|\right] \leq \mathbb{E}\left[\left(\sum_{j=2}^{J+1}|\bar{\varepsilon}_j|^p\right)^{1/p}\right] \leq \left(\mathbb{E}\left[\sum_{j=2}^{J+1}|\bar{\varepsilon}_j|^p\right]\right)^{1/p} = \left(\sum_{j=2}^{J+1}\mathbb{E}|\bar{\varepsilon}_j|^p\right)^{1/p}.$$

Then using Rosenthal's Inequality, we have

$$\mathbb{E}|\bar{\varepsilon}_j|^p \leq C(p)\left(\frac{\bar{\lambda}^2 f}{KT_0\underline{\xi}}\right)^p \max\left\{\sum_{q=1}^{K}\sum_{s=1}^{T_0}\mathbb{E}|\varepsilon_{js,q}|^p, \left(\sum_{q=1}^{K}\sum_{s=1}^{T_0}\mathbb{E}|\varepsilon_{js,q}|^2\right)^{p/2}\right\},$$

where the constant $C(p) = \mathbb{E}(\phi - 1)^p$ with ϕ being a Poisson random variable with parameter 1.

Let $\bar{m}_p = \max_j \frac{1}{KT_0}\sum_{q=1}^{K}\sum_{s=1}^{T_0}\mathbb{E}|\varepsilon_{js,q}|^p$, then we have

$$\mathbb{E}|B_{1t,k}| \leq C(p)^{1/p}\left(\frac{\bar{\lambda}^2 f}{\underline{\xi}}\right) J^{1/p}\max\left\{\frac{\bar{m}_p^{1/p}}{(KT_0)^{1-1/p}}, \frac{\bar{m}_2^{1/2}}{(KT_0)^{1/2}}\right\}. \tag{B.1.7}$$

Therefore, $\mathbb{E}|B_{1t,k}| = O\left(\frac{1}{\sqrt{KT_0}}\right)$, and $\mathbb{E}(\widehat{\tau}_{1t,k} - \tau_{1t,k}) \to 0$ as $KT_0 \to \infty$, i.e., the bias of the multiple-outcome synthetic control estimator is bounded by a function that goes to zero when the number of outcomes in the domain or the pretreatment periods goes to infinity.

We now provide a unified framework for the biases of the multiple-outcome synthetic control estimator and the single-outcome synthetic control estimator. The bias of the single-outcome synthetic control estimator for outcome k and $t > T_0$ is the expectation of

$$\begin{aligned} Y_{1t,k} - \sum_{j=2}^{J+1}\widetilde{w}_j^{(k)}Y_{jt,k} &= \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1}\widetilde{w}_j^{(k)}\boldsymbol{\mu}_j\right)'\boldsymbol{\lambda}_{t,k} + \varepsilon_{1t,k} - \sum_{j=2}^{J+1}\widetilde{w}_j^{(k)}\varepsilon_{jt,k}, \\ &= \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1}w_j^*\boldsymbol{\mu}_j\right)'\boldsymbol{\lambda}_{t,k} - \sum_{j=2}^{J+1}\widetilde{v}_j^{(k)}\boldsymbol{\mu}_j'\boldsymbol{\lambda}_{t,k} + \varepsilon_{1t,k} - \sum_{j=2}^{J+1}\widetilde{w}_j^{(k)}\varepsilon_{jt,k}. \end{aligned} \tag{B.1.8}$$

Stacking the observations over the T_0 pretreatment periods, we have

$$\boldsymbol{Y}_{i,k} = \boldsymbol{\delta}_k + \boldsymbol{\theta}_k\boldsymbol{Z}_i + \boldsymbol{\lambda}_k\boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{i,k},$$

where $\boldsymbol{Y}_{i,k}$, $\boldsymbol{\delta}_k$, and $\boldsymbol{\varepsilon}_{i,k}$ are $T_0 \times 1$, and $\boldsymbol{\theta}_k$ and $\boldsymbol{\lambda}_k$ are $T_0 \times r$ and $T_0 \times F$, respectively.

The restrictions $\sum_{j=2}^{J+1} \widetilde{w}_j^{(k)} \boldsymbol{Y}_{j,k} = \boldsymbol{Y}_{1,k}$ can be simplified to

$$\boldsymbol{\lambda}_k \left(\boldsymbol{\mu}_1 - \sum_{j=2}^{J+1} \widetilde{w}_j^{(k)} \boldsymbol{\mu}_j \right) = \sum_{j=2}^{J+1} \widetilde{w}_j^{(k)} \boldsymbol{\varepsilon}_{j,k} - \boldsymbol{\varepsilon}_{1,k}. \tag{B.1.9}$$

Using (B.1.4), we can simplify (B.1.9) further to

$$-\boldsymbol{\lambda}_k \sum_{j=2}^{J+1} \widetilde{v}_j^{(k)} \boldsymbol{\mu}_j = \sum_{j=2}^{J+1} \widetilde{v}_j^{(k)} \boldsymbol{\varepsilon}_{j,k}. \tag{B.1.10}$$

Suppose that $\boldsymbol{\lambda}_k' \boldsymbol{\lambda}_k$ has full rank, pre-multiplying $\boldsymbol{\lambda}_{t,k}' (\boldsymbol{\lambda}_k' \boldsymbol{\lambda}_k)^{-1} \boldsymbol{\lambda}_k'$ to both sides of (B.1.10) gives

$$-\boldsymbol{\lambda}_{t,k}' \sum_{j=2}^{J+1} \widetilde{v}_j^{(k)} \boldsymbol{\mu}_j = \boldsymbol{\lambda}_{t,k}' (\boldsymbol{\lambda}_k' \boldsymbol{\lambda}_k)^{-1} \boldsymbol{\lambda}_k' \sum_{j=2}^{J+1} \widetilde{v}_j^{(k)} \boldsymbol{\varepsilon}_{j,k}. \tag{B.1.11}$$

Therefore, (B.1.8) can be written as

$$\begin{aligned}
& Y_{1t,k}^0 - \sum_{j=2}^{J+1} \widetilde{w}_j^{(k)} Y_{jt,k} \\
=& \boldsymbol{\lambda}_{t,k}' (\boldsymbol{\lambda}' \boldsymbol{\lambda})^{-1} \boldsymbol{\lambda}' \sum_{j=2}^{J+1} w_j^* \boldsymbol{\varepsilon}_j && (B_{1t,k}) \\
& - \boldsymbol{\lambda}_{t,k}' (\boldsymbol{\lambda}' \boldsymbol{\lambda})^{-1} \boldsymbol{\lambda}' \boldsymbol{\varepsilon}_1 + \varepsilon_{1t,k} - \sum_{j=2}^{J+1} w_j^* \varepsilon_{jt,k} && (B_{2t,k}) \\
& - \sum_{j=2}^{J+1} \widetilde{v}_j^{(k)} \varepsilon_{jt,k} && (B_{3t,k}) \\
& + \boldsymbol{\lambda}_{t,k}' (\boldsymbol{\lambda}_k' \boldsymbol{\lambda}_k)^{-1} \boldsymbol{\lambda}_k' \sum_{j=2}^{J+1} \widetilde{v}_j^{(k)} \boldsymbol{\varepsilon}_{j,k}. && (B_{4t,k})
\end{aligned}$$

We have shown that $\mathbb{E}\,|B_{1t,k}| = O\left(\frac{1}{\sqrt{KT_0}}\right)$ and $\mathbb{E}\,(B_{2t,k}) = 0$. It can be similarly shown that $\mathbb{E}\,(B_{3t,k}) = 0$ and $\mathbb{E}\,|B_{4t,k}| = O\left(\frac{1}{\sqrt{T_0}}\right)$.

□

Proof of Corollary 2.1. The bias for the demeaned synthetic control estimator is

$$\begin{aligned}\tilde{\tau}_{1t,k} - \tau_{1t,k} &= \dot{Y}_{1t,k} - \sum_{j=2}^{J+1} w_j^* \dot{Y}_{jt,k} - Y_{1t,k}^1 + Y_{1t,k}^0 \\ &= Y_{1t,k}^1 - \frac{1}{T_0}\sum_{s=1}^{T_0} Y_{1s,k} - \sum_{j=2}^{J+1} w_j^* \dot{Y}_{jt,k}^0 - Y_{1t,k}^1 + Y_{1t,k}^0 \\ &= \dot{Y}_{1t,k}^0 - \sum_{j=2}^{J+1} w_j^* \dot{Y}_{jt,k}^0.\end{aligned}$$

Notice that the demeaned equation for $Y_{it,k}^0$ retains the interactive fixed effects structure:

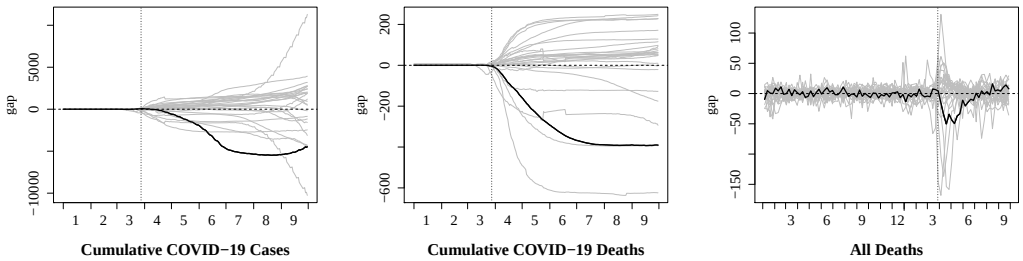
$$\begin{aligned}\dot{Y}_{it,k}^0 &= Y_{it,k}^0 - \frac{1}{T_0}\sum_{s=1}^{T_0} Y_{is,k}^0 \\ &= \dot{\delta}_{t,k} + \boldsymbol{Z}_i' \dot{\boldsymbol{\theta}}_{t,k} + \boldsymbol{\mu}_i' \dot{\boldsymbol{\lambda}}_{t,k} + \dot{\varepsilon}_{it,k}.\end{aligned}$$

We can thus follow similar steps to show that $\mathbb{E}\left(\tilde{\tau}_{1t,k} - \tau_{1t,k}\right) \to 0$ as $KT_0 \to \infty$ under 2.1, 2.2′ and 2.3.

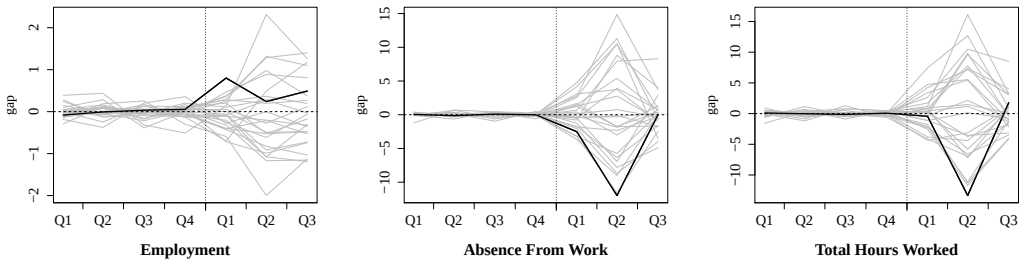
□

Proof of Corollary 2.2. The proof is similar to that of Proposition 2.1 and thus omitted. □

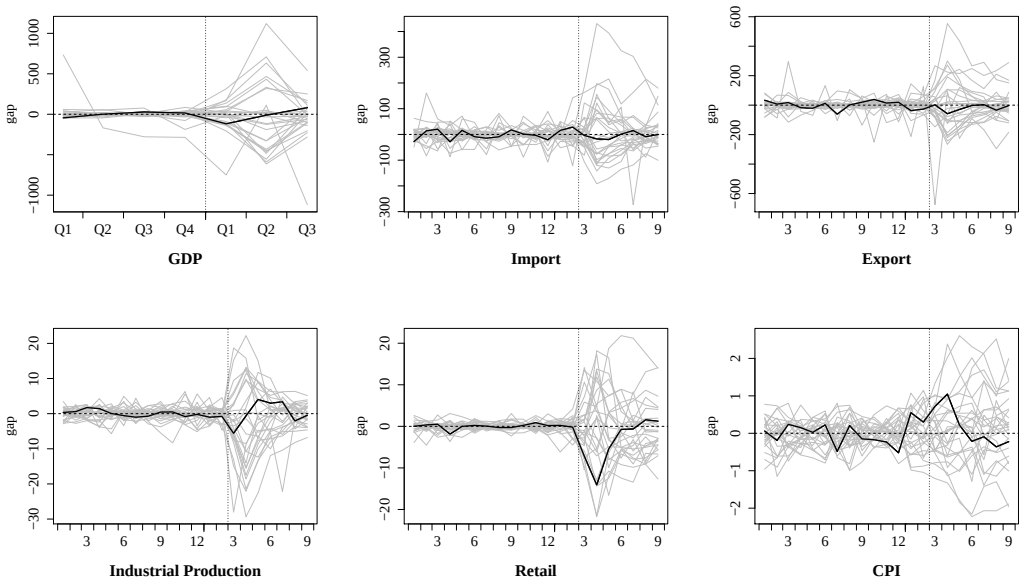
B.2 Additional Results



(a) Public Health Outcomes

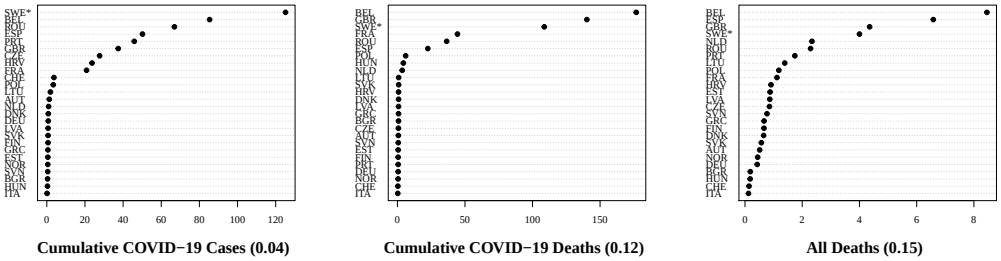


(b) Labour Market Outcomes

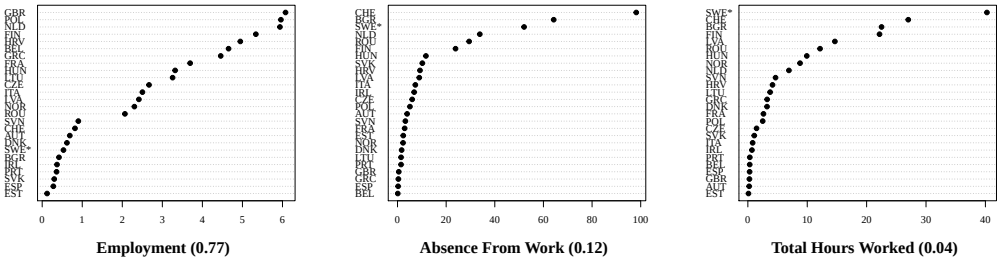


(c) Economic Outcomes

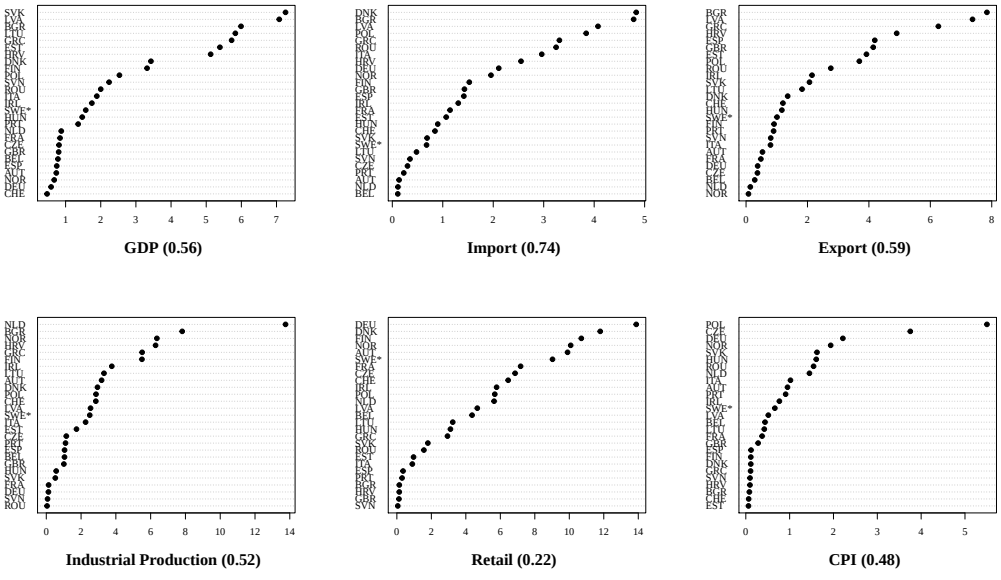
Figure B.2.1: Placebo Gaps in Outcomes



(a) Public Health Outcomes



(b) Labour Market Outcomes



(c) Economic Outcomes

Figure B.2.2: Posttreatment/Pretreatment RMSPE Ratios

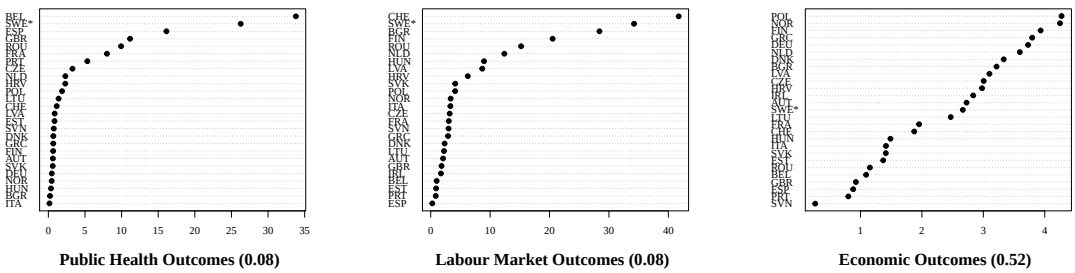
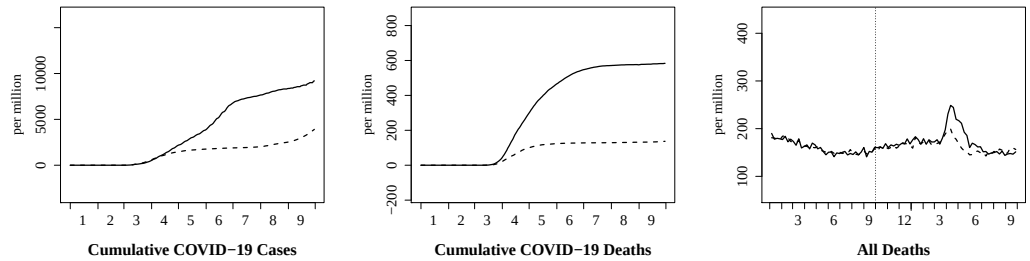
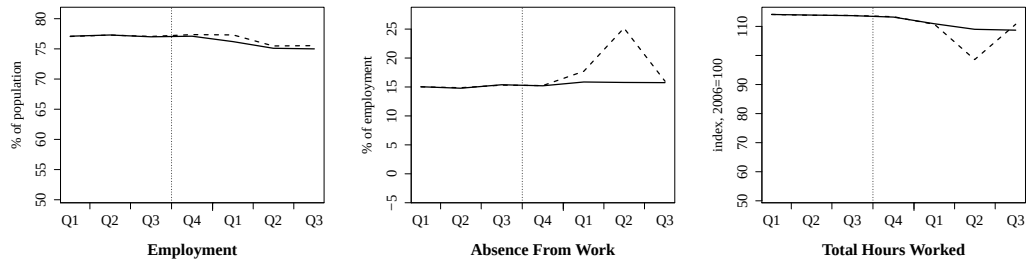


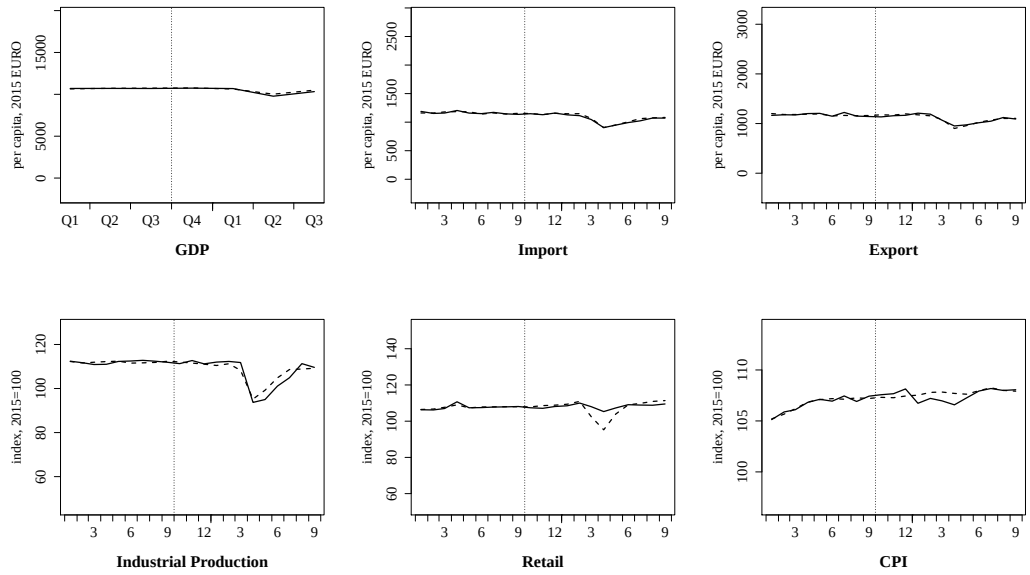
Figure B.2.3: Aggregate Posttreatment/Pretreatment RMSPE Ratios



(a) Public Health Outcomes

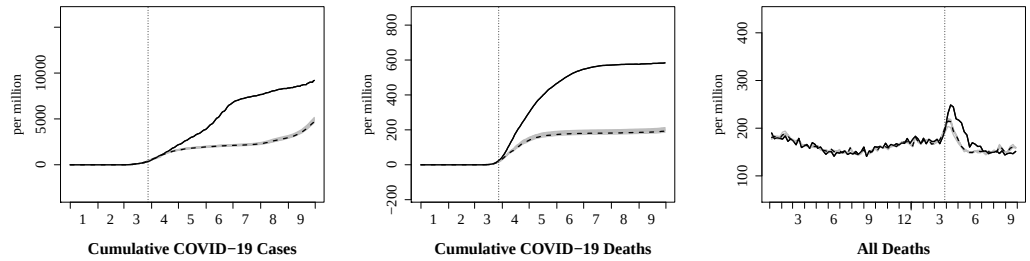


(b) Labour Market Outcomes

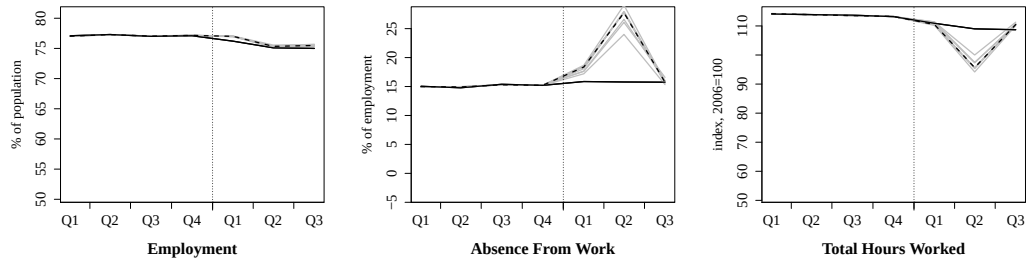


(c) Economic Outcomes

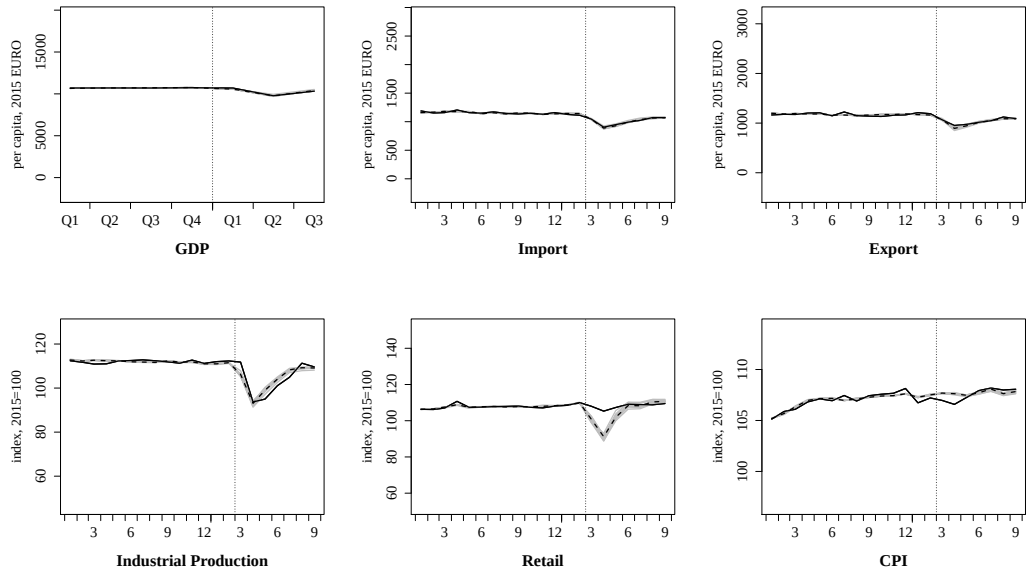
Figure B.2.4: Backdating the Treatment



(a) Public Health Outcomes



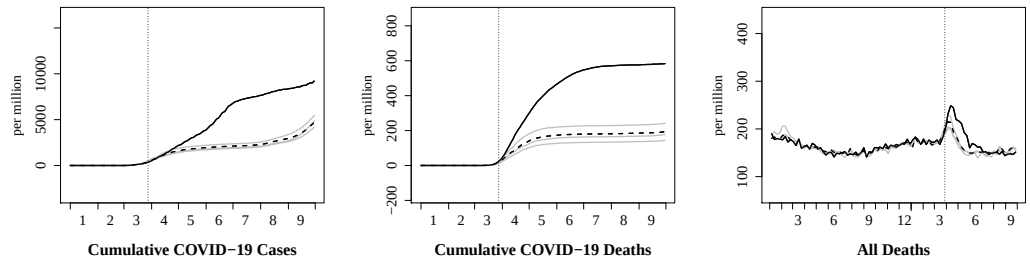
(b) Labour Market Outcomes



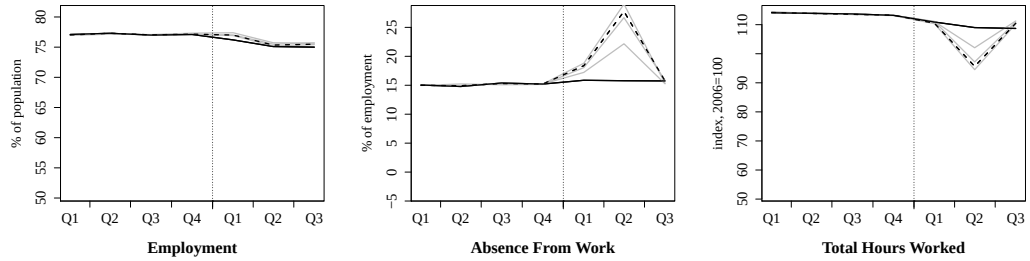
(c) Economic Outcomes

Figure B.2.5: Leave-One-Unit-Out Distribution of the Synthetic Control for Sweden

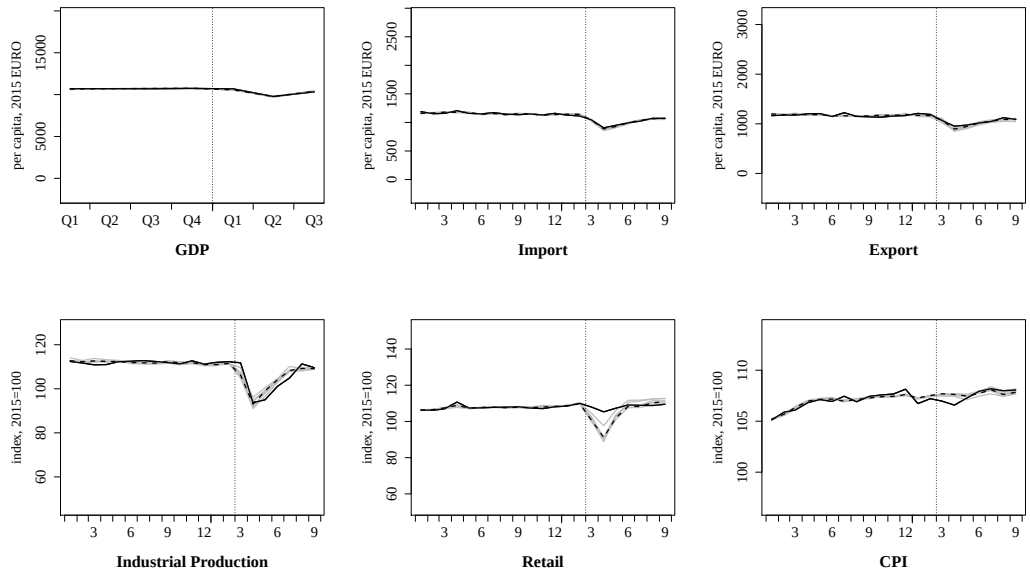
Note: The outcome trajectories for Sweden and the synthetic Sweden are in solid and dashed lines respectively. The trajectories for the leave-out-out synthetic controls are in gray. The vertical dotted line delimits the pre/posttreatment periods for each outcome.



(a) Public Health Outcomes



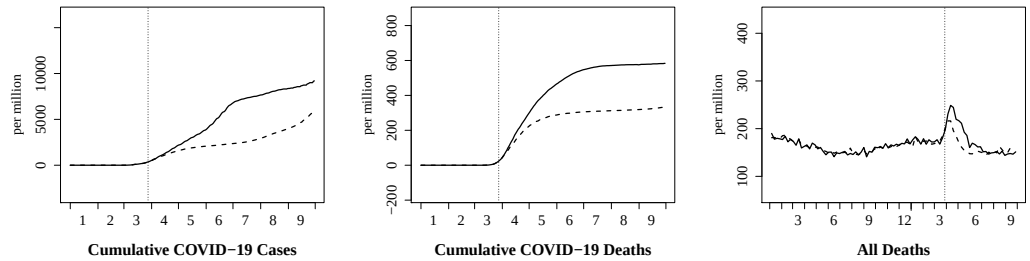
(b) Labour Market Outcomes



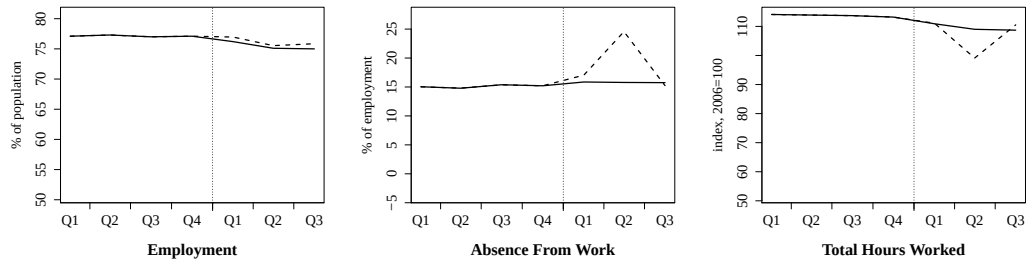
(c) Economic Outcomes

Figure B.2.6: Leave-One-Outcome-Out Distribution of the Synthetic Control for Sweden

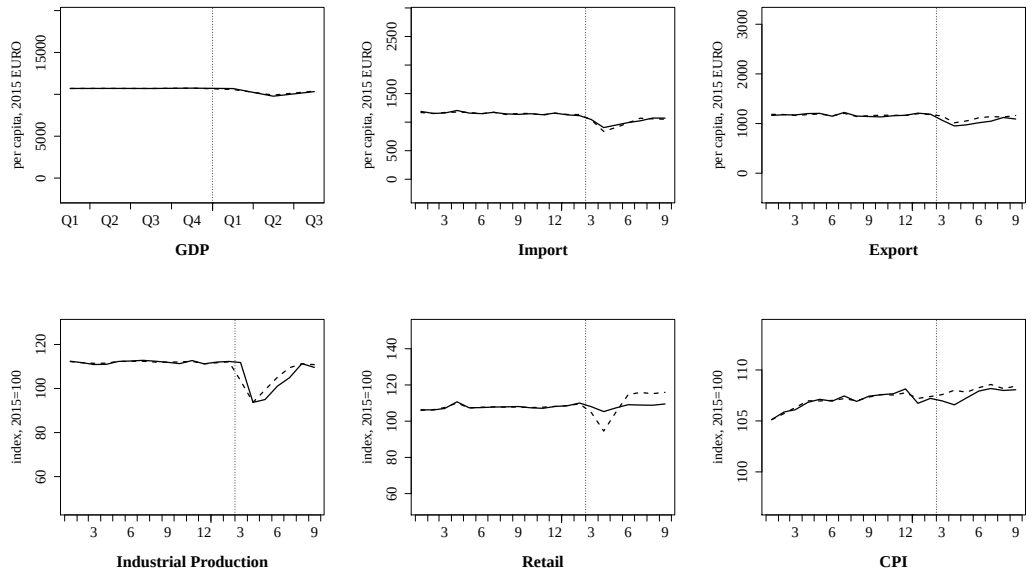
Note: The outcome trajectories for Sweden and the synthetic Sweden are in solid and dashed lines respectively. The trajectories for the leave-out-out synthetic controls are in gray. The vertical dotted line delimits the pre/posttreatment periods for each outcome.



(a) Public Health Outcomes

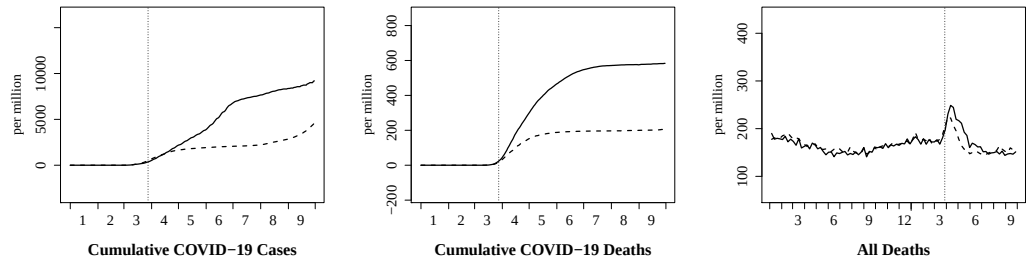


(b) Labour Market Outcomes

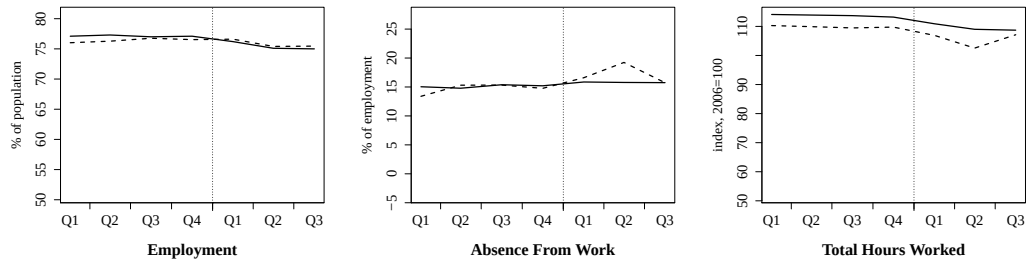


(c) Economic Outcomes

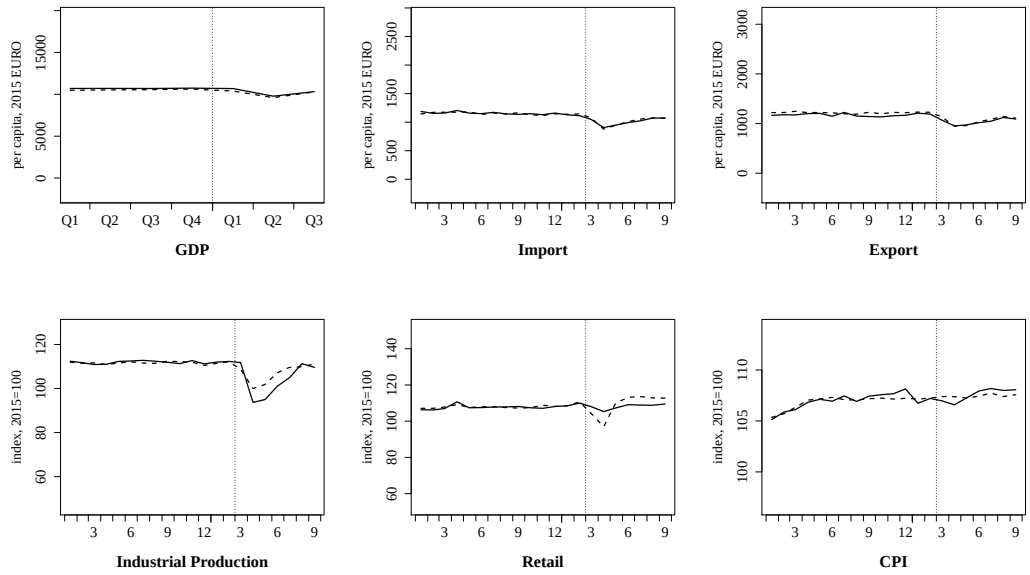
Figure B.2.7: Single Outcomes



(a) Public Health Outcomes

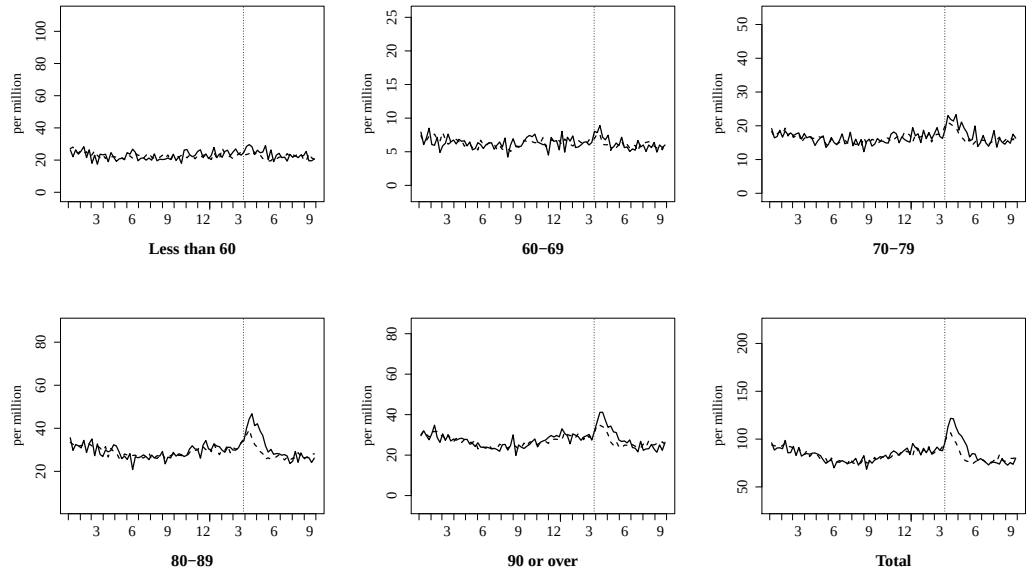


(b) Labour Market Outcomes

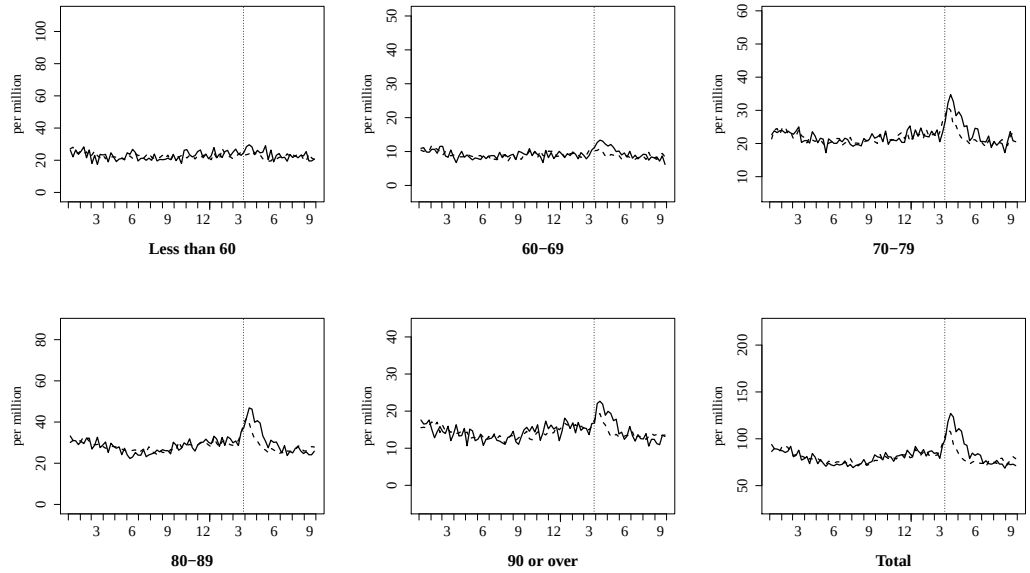


(c) Economic Outcomes

Figure B.2.8: No Demeaning



(a) Female



(b) Male

Figure B.2.9: Deaths by Gender and Age

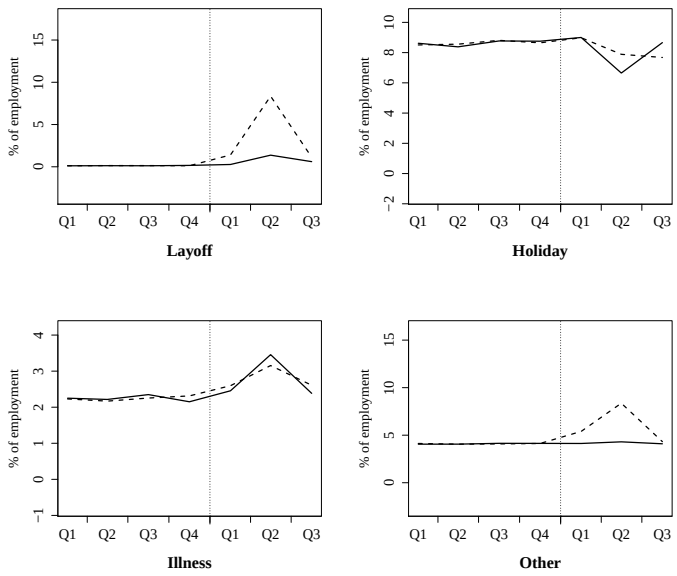


Figure B.2.10: Reasons of Absence From Work

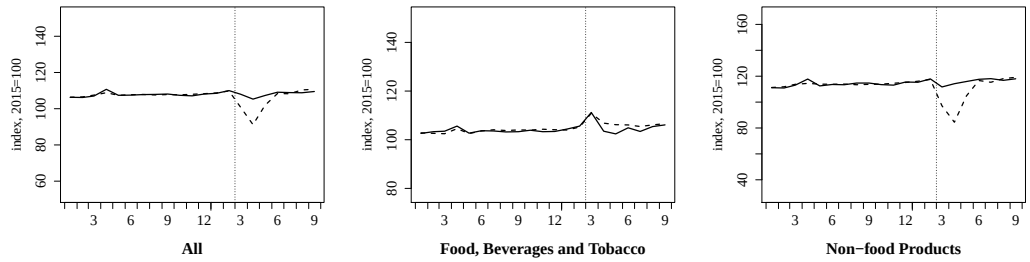
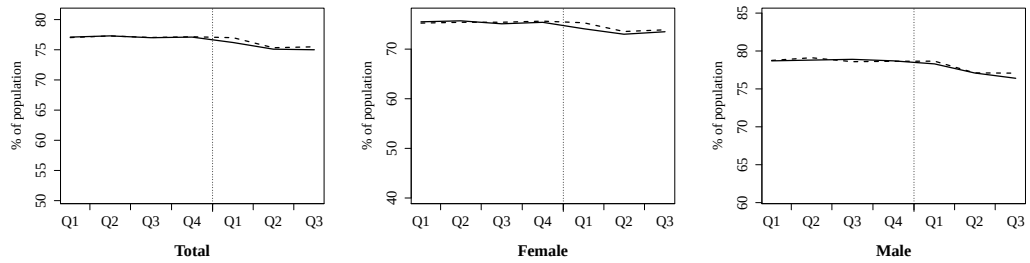
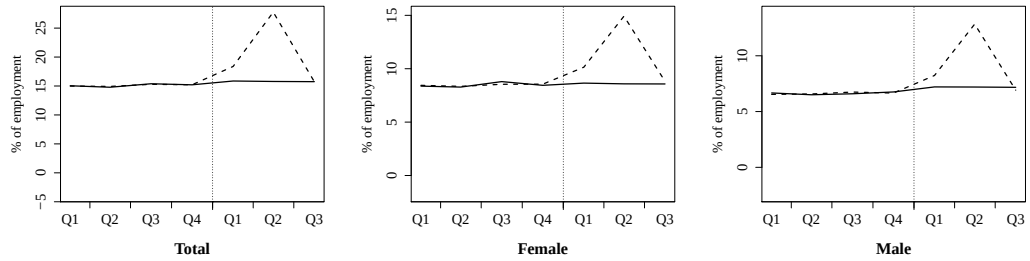


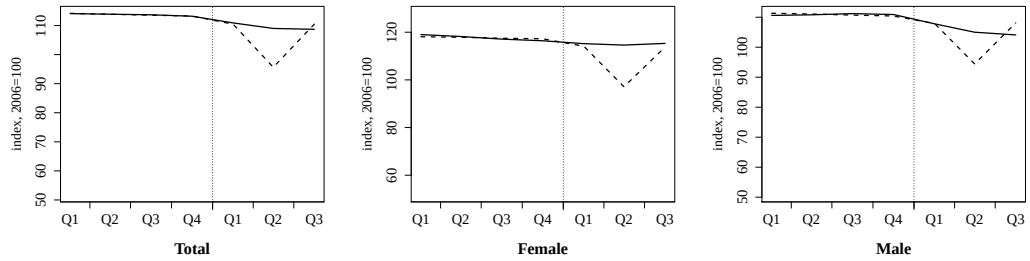
Figure B.2.11: Retail



(a) Employment



(b) Absence From Work



(c) Total Hours Worked

Figure B.2.12: Labour Market Outcomes by Gender

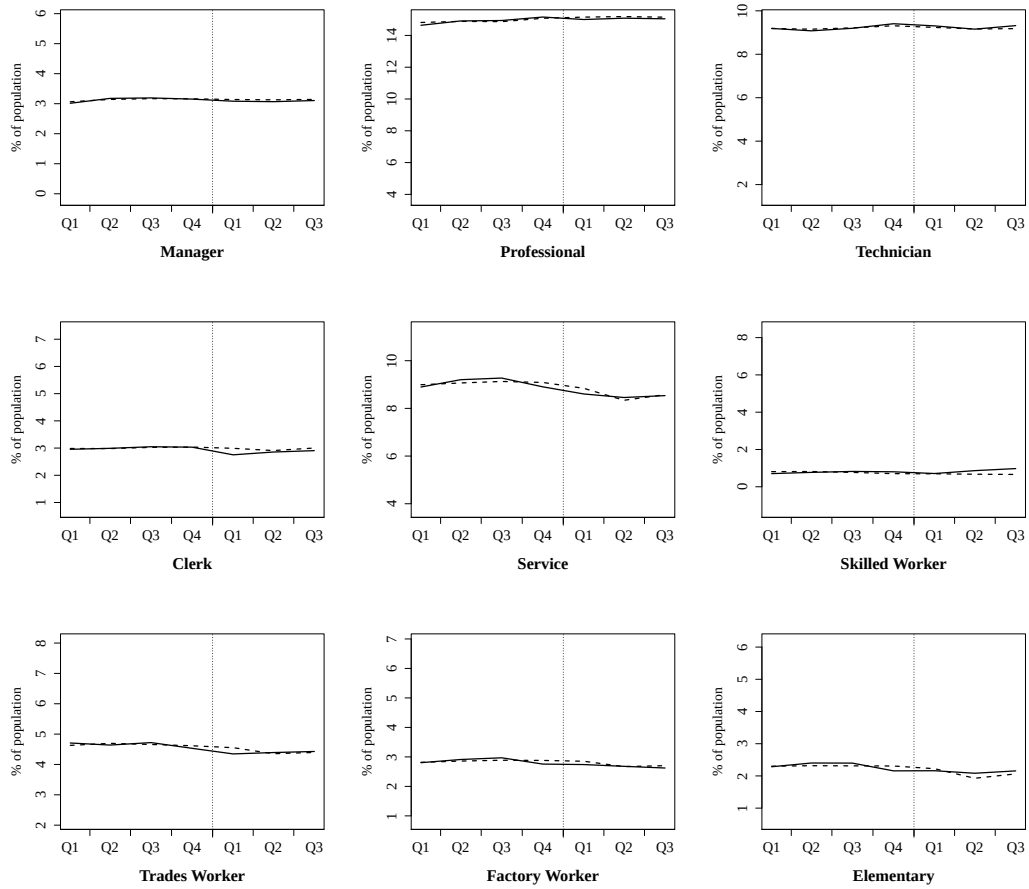


Figure B.2.13: Employment by Occupation

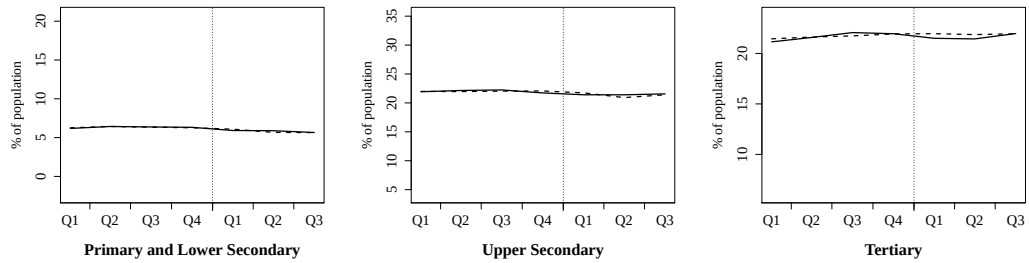


Figure B.2.14: Employment by Education

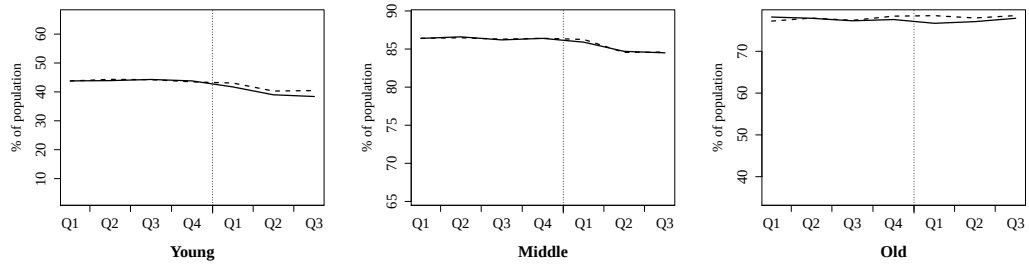


Figure B.2.15: Employment by Age

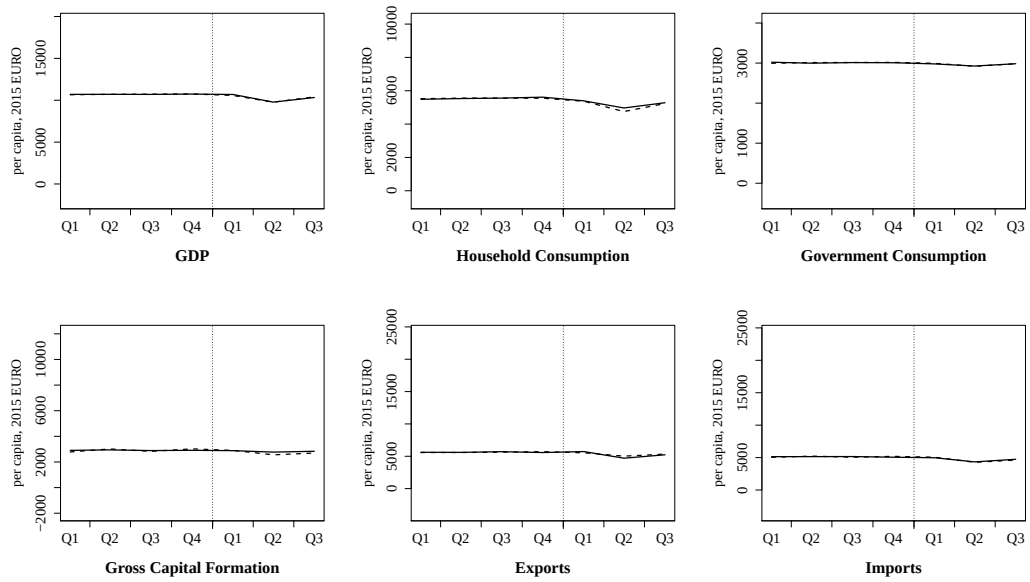


Figure B.2.16: GDP

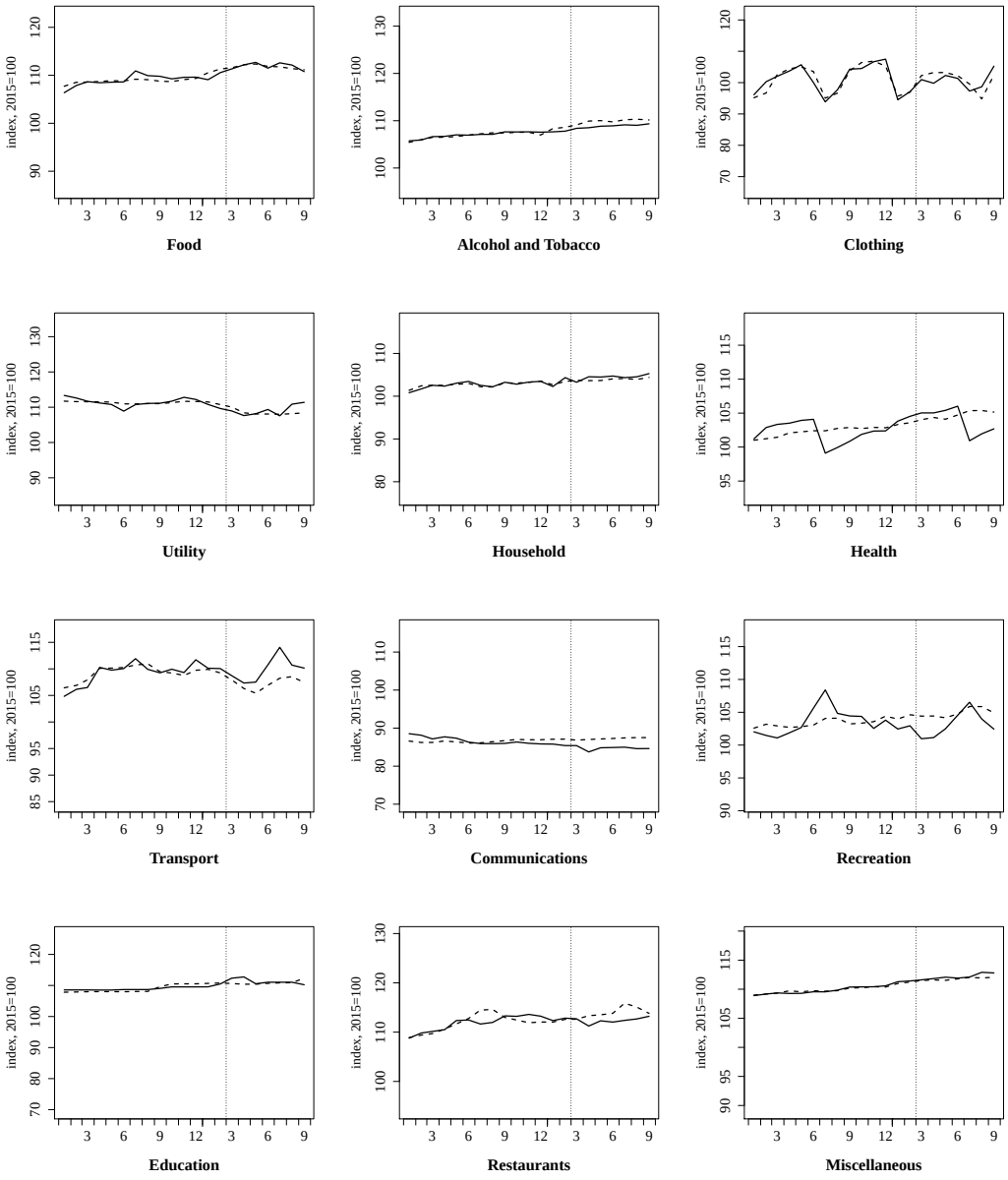


Figure B.2.17: CPI

Chapter 3

Individual Causal Inference Using Panel Data With Multiple Outcomes

3.1 Introduction

The main focus of the policy evaluation literature has been the average treatment effect and more recently the heterogeneous treatment effects or conditional average treatment effects, which are the average treatment effects for heterogeneous subgroups defined by the observed covariates (for reviews of these methods, see Athey and Imbens, 2017; Abadie and Cattaneo, 2018). Ubiquitous in these studies is the unconfoundedness assumption, or the strong ignorability assumption, which requires all the covariates correlated with both the potential outcomes and the treatment assignment to be observed (Rosenbaum and Rubin, 1983).[1] Under this assumption, the potential outcomes and the treatment status are independent conditional on the observed covariates, and the difference between the mean outcomes of the treated and the untreated groups with the same values of the observed covariates is an unbiased estimator of the average treatment effect for the units in the groups. The unconfoundedness assumption is satisfied in randomised controlled experiments, but may not be plausible otherwise even with a rich set of covariates, since the access to certain essential individual characteristics remains limited for the researchers due to privacy or ethical concerns, despite the explosive growth of data availability in the big data era.

One popular method to circumvent the unconfoundedness assumption is difference-in-differences (DID), which assumes that the effect of the unobserved confounder on

[1] This is also known as selection on observables or the conditional independence assumption.

the untreated potential outcome is constant over time, so that the average outcomes of the treated and untreated units would follow parallel trends in the absence of the treatment.[2] This is also a strong assumption, and in many cases is not supported by data. The interactive fixed effects model relaxes the "parallel trends" assumption and allows the unobserved confounders to have time-varying effects on the outcomes, by modeling them using an interactive fixed effects term, which incorporates the additive unit and time fixed effects model or difference-in-differences as a special case (Bai, 2009).

Several methods have been developed based on the interactive fixed effects model to estimate the treatment effect on a single or several treated units, where the units are observed over an extended period of time before the treatment (Abadie et al., 2010; Hsiao et al., 2012; Xu, 2017). These methods exploit the cross-sectional correlations attributed to the unobserved common factors to predict the counterfactual outcomes for the treated units, and are mainly used in macroeconomic settings with a large number of pretreatment periods, which is crucial for the results to be credible. For example, Abadie et al. (2015) point out that "the applicability of the method requires a sizable number of preintervention periods" and that "we do not recommend using this method when the pretreatment fit is poor or the number of pretreatment periods is small", while Xu (2017) states that users should be cautious when there are fewer than 10 pretreatment periods. As a consequence, despite the potential to estimate individual treatment effects without imposing the unconfoundedness assumption, these methods have not seen much use in empirical microeconomics, since the individuals are rarely tracked for more than a few periods that justify the use of these methods.

The main contribution of this paper is that we propose a method for estimating the individual treatment effects in applied microeconomic settings, characterised by multiple related outcomes being observed for a large number of individuals over a small number of time periods. The method is based on the interactive fixed effects model, which assumes that an outcome of interest can be well approximated by a linear combination of a small number of observed and unobserved individual characteristics. Analogous to Hsiao et al. (2012) who predict the posttreatment outcomes using pretreatment outcomes in lieu of the unobserved time factors, we use a subsample of the pretreatment outcomes to replace the unobserved individual characteristics in the models, and use the remaining pretreatment outcomes as instrumental variables. Although our method does not require a large number of

[2]Alternative methods that do not rely on the unconfoundedness assumption include the instrumental variables method and the regression discontinuity design, which estimate the average treatment effect for specific subpopulations (the compliers or those with values of the running variable near the cutoff).

pretreatment periods, the number of pretreatment outcomes needs to be at least as large as the number of unobserved individual characteristics, which may still be difficult to satisfy in microeconomic datasets if we use only a single outcome, especially if the treatment assignment took place in the early stages of the survey or if the study subjects are children or youths. Utilising multiple related outcomes allows our method to be applicable in cases where there is only a single period before the treatment. Under the assumption that these outcomes depend on roughly the same set of observed covariates and unobserved individual characteristics with time-varying and outcome-specific coefficients shared by all individuals, our method exploits the correlations across related outcomes and over time, which are induced by the unobserved individual characteristics, to predict the counterfactual outcomes and estimate the treatment effects for each individual in the posttreatment periods.

Our method has several advantages. First, with the assumption on the model specification, it relaxes the arguably much stronger unconfoundedness assumption, and allows the treatment assignment to be correlated with the unobserved individual characteristics. Second, it enables the estimation of treatment effects on the individual level, which may be helpful for designing more individualised policies to maximize social welfare, as well as for other fields such as precision medicine and individualised marketing. It also has the potential to be combined with more flexible machine learning methods to work with big datasets and more general nonlinear function forms. Third, it is intuitive. In real life, we may never know a person through and through, and a viable approach to predicting the outcome of a person is using his or her related outcomes in the past, assuming that the outcomes are affected by the underlying individual characteristics and that these characteristics are stable over time, at least within the study period. For example, past academic performance is an important consideration when recruiting a student into college, as it is believed that a student that excelled in the past is likely to continue to have outstanding performance. To the extent that we may never observe all the confounders, this is perhaps the only way to predict potential outcomes and estimate treatment effects on the individual level in social sciences without going deeper to the levels of neuroscience or biology. Fourth, our method has wide applicability, as it is common to have multiple related outcomes collected in microeconomics data. For example, we may observe several health related outcomes such as health facility usage, health related cost, general health, etc.

The rest of the study is organised as follows. Section 3.2 presents the theoretical framework. Section 3.3 examines the small sample performance of our method using Monte Carlo simulation, and compares it with related methods. Section 3.4 provides an empirical example of estimating the effect of health insurance coverage

on individual usage of hospital emergency departments using the Oregon Health Insurance Experiment data. Section 3.5 concludes and discusses potential directions for future research. The proofs are collected in the appendix.

3.2 Theory

3.2.1 Set Up

Suppose that we observe K outcomes in domain $\mathcal{K} = \{1, 2, \ldots, K\}$ for N individuals or units over $T \geq 2$ time periods, where a domain refers to a collection of related outcomes that depend on the same set of observed covariates and unobserved characteristics. For example, health-related outcomes may be affected by observed covariates such as age, education, occupation and income, as well as unobserved individual characteristics such as genetic inheritance, health habits and risk preferences. Assume that the N_1 individuals in the treated group $\mathcal{T}$ receive the treatment at period $T_0 + 1 \leq T$ and remain treated afterwards, while the $N_0 = N - N_1$ individuals in the control group $\mathcal{C}$ remain untreated throughout the T periods. Denoting the binary treatment status for individual i at time t as D_{it}, we have $D_{it} = 1$ for $i \in \mathcal{T}$ and $t > T_0$, and $D_{it} = 0$ otherwise.

Following the "Rubin Causal Model" (Rubin, 1974), the treatment effect on outcome $k \in \mathcal{K}$ for individual i at time t is given by the difference between the treated and untreated potential outcomes

$$\tau_{it,k} = Y^1_{it,k} - Y^0_{it,k}, \tag{3.1}$$

where $Y^1_{it,k}$ is the treated potential outcome, the outcome that we would observe for individual i at time t if $D_{it} = 1$, and $Y^0_{it,k}$ is the untreated potential outcome, the outcome that we would observe if $D_{it} = 0$. Instead of assuming the unconfoundedness condition, we characterise the two potential outcomes for individual i at time t and $k \in \mathcal{K}$ using the interactive fixed effects models:

$$Y^1_{it,k} = \boldsymbol{X}'_{it}\boldsymbol{\beta}^1_{t,k} + \boldsymbol{\mu}'_i\boldsymbol{\lambda}^1_{t,k} + \varepsilon^1_{it,k}, \tag{3.2}$$

$$Y^0_{it,k} = \boldsymbol{X}'_{it}\boldsymbol{\beta}^0_{t,k} + \boldsymbol{\mu}'_i\boldsymbol{\lambda}^0_{t,k} + \varepsilon^0_{it,k}, \tag{3.3}$$

where $\boldsymbol{X}_{it}$ is the $r \times 1$ vector of observed covariates unaffected by the treatment, $\boldsymbol{\mu}_i$ is the $f \times 1$ vector of unobserved individual characteristics, $\boldsymbol{\beta}^1_{t,k}$ and $\boldsymbol{\lambda}^1_{t,k}$ are the $r \times 1$ and $f \times 1$ vectors of coefficients of $\boldsymbol{X}_{it}$ and $\boldsymbol{\mu}_i$ respectively for the treated potential outcome, $\boldsymbol{\beta}^0_{t,k}$ and $\boldsymbol{\lambda}^0_{t,k}$ are the coefficients for the untreated potential outcome, and $\varepsilon^1_{it,k}$ and $\varepsilon^0_{it,k}$ are the idiosyncratic shocks.

Remark 3.1. Our models for the potential outcomes are quite general, and incorporate the models in Abadie et al. (2010), Hsiao et al. (2012) and Xu (2017), as well as the the additive fixed effects model for difference-in-differences as special cases.[3] Note that the related outcomes need not depend on exactly the same set of observed covariates and unobserved individual characteristics. The vectors of outcome-specific and time-varying coefficients may contain 0, so that outcome k may be affected by some of the observed covariates and unobserved individual characteristics in some periods, but not necessarily by all of them in all periods, as long as there is enough variation in the coefficients over time or across the outcomes. The potential outcomes are also allowed to depend on predictors not included in $\boldsymbol{X}_{it}$ or $\boldsymbol{\mu}_i$, as long as they are not correlated with the included predictors and the treatment status so that they can be treated as part of the idiosyncratic shock.

The regularity conditions on the observed covariates and the unobserved individual characteristics are stated in Assumption 3.1, and the assumptions on the idiosyncratic shocks are given in Assumption 3.2.

Assumption 3.1.

1) $\boldsymbol{X}_{it}$, $\boldsymbol{\mu}_i$ *are independent for all* i, *and are identically distributed for all* $i \in \mathcal{T}$ *and all* $i \in \mathcal{C}$ *respectively;*

2) There exists $M \in [0, \infty)$ *such that* $\mathbb{E}\|\boldsymbol{X}_{it}\|^4 < M$ *and* $\mathbb{E}\|\boldsymbol{\mu}_i\|^4 < M$.

Assumption 3.2. *For* $d \in \{0, 1\}$, *we have*

1) $\mathbb{E}\left(\varepsilon_{it,k}^d \mid \boldsymbol{X}_{js}, \boldsymbol{\mu}_j, D_{js}\right) = 0$ *for all* i, j, t, s *and* k;

2) $\varepsilon_{it,k}^d$ *are independent across* i *and* t;

3) $\mathbb{E}\left(\varepsilon_{it,k}^d, \varepsilon_{it,l}^d\right) = \sigma_{t,kl}^d$ *for all* i, t, k, l;

4) There exists $M \in [0, \infty)$ *such that* $\mathbb{E}|\varepsilon_{it,k}^d|^4 < M$ *for all* i, t, k.

Remark 3.2. The distributions of the observed covariates and the unobserved individual characteristics are allowed to differ for the treated and untreated individuals, i.e., selection on unobservables is allowed, which is a great advantage over the policy evaluation methods that rely on the unconfoundedness condition. The idiosyncratic shocks are assumed to have zero mean conditional on the observed covariates, unobserved individual characteristics and the treatment status. They are also assumed to be independent across individuals and time periods, as the unobserved interactive

[3]Specifically, if we assume $\boldsymbol{\beta}_{t,k}^0 = \boldsymbol{\beta}_k^0$, model (3.3) reduces to the model in Bai (2009) and Xu (2017); if we assume $\boldsymbol{X}_{it} = \boldsymbol{X}_i$ and the first element of $\boldsymbol{\mu}_i$ is 1, model (3.3) reduces to the model in Abadie et al. (2010); if we assume $\boldsymbol{X}_{it} = \boldsymbol{X}_i$ are unobserved and the first element of $\boldsymbol{\lambda}_{t,k}^0$ is 1, then model (3.3) reduces to the model in Hsiao et al. (2012); if we assume $\boldsymbol{\mu}_i = [1 \;\; a_i]'$ and $\boldsymbol{\lambda}_{t,k}^0 = [b_t \;\; 1]'$, then model (3.3) reduces to the additive fixed effects model for difference-in-differences.

fixed effects that account for the cross-sectional and time-serial correlations have been separated out.[4] Furthermore, they are assumed to be homoskedastic across individuals for our inference method to be valid. The last part in Assumption 3.2 is a regularity condition which, together with the conditions in Assumption 3.1, ensures the weak law of large numbers and the central limit theorem hold.

Given the models for $Y_{it,k}^1$ and $Y_{it,k}^0$ in (3.2) and (3.3), the individual treatment effect is identified by the observed covariates and the unobserved individual characteristics, i.e., two persons with the same values for these underlying predictors have identical individual treatment effect. Denote the set of observed covariates and unobserved individual characteristics as $\boldsymbol{H}_{it} = [\boldsymbol{X}_{it}' \ \boldsymbol{\mu}_i']'$, then the individual treatment effect for individual i with $\boldsymbol{H}_{it} = \boldsymbol{h}_{it}$ is given by

$$\bar{\tau}_{it,k} \equiv \mathbb{E}\left(Y_{it,k}^1 - Y_{it,k}^0 \mid \boldsymbol{H}_{it} = \boldsymbol{h}_{it}\right), \tag{3.4}$$

which may appear similar to the conditional average treatment effect, but is different by conditioning not only on the observed covariates, but also on the unobserved individual characteristics.[5]

Our goal is to estimate the individual treatment effects $\bar{\tau}_{it,k}$, $i = 1, \ldots, N$. Once we have estimated the individual treatment effects, the estimates of the average treatment effects for heterogeneous subgroups defined by some observed covariates, also known as the conditional average treatment effects, and the estimate for the average treatment effect for the sample or the population are also readily available using the average of the estimated individual treatment effects in the corresponding groups.

As $\boldsymbol{\mu}_i$ is not observed, a direct application of least squares estimation to estimate the models in (3.2) and (3.3) would suffer from omitted variables bias. Since we have multiple outcomes that depend on the same set of underlying predictors, and we observe the untreated potential outcomes for all individuals prior to the treatment, we can use these pretreatment outcomes to replace $\boldsymbol{\mu}_i$ in the models.[6] Stacking the

[4]The idiosyncratic shocks may be allowed to be correlated both over time and across outcomes, as long as they can be modelled parametrically and removed using a quasi-differencing approach. This is left for future research.

[5]As we assume the parametric models for the potential outcomes in (3.2) and (3.3) for all individuals, there is no need to impose additional assumptions on the propensity distribution for the individual treatment effect to be identified on its full support.

[6]This is analogous to the first step of the approach in Hsiao et al. (2012), who predict the posttreatment outcomes using pretreatment outcomes in lieu of the unobserved time factors in a small N, big T environment.

K outcomes observed in $t \leq T_0$, we have

$$\boldsymbol{Y}_{it}^0 = \boldsymbol{\beta}_t^0 \boldsymbol{X}_{it} + \boldsymbol{\lambda}_t^0 \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{it}^0, \tag{3.5}$$

where $\boldsymbol{Y}_{it}^0$ and $\boldsymbol{\varepsilon}_{it}^0$ are $K \times 1$, $\boldsymbol{\beta}_t^0$ is $K \times r$, and $\boldsymbol{\lambda}_t^0$ is $K \times f$. Let $\mathcal{P} \subseteq \{1, \cdots, T_0\}$ be a set of P pretreatment periods. We can further stack the outcomes over these periods to get

$$\boldsymbol{Y}_i^{\mathcal{P}} = \boldsymbol{\delta}_i^{\mathcal{P}} + \boldsymbol{\lambda}^{\mathcal{P}} \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i^{\mathcal{P}}, \tag{3.6}$$

where $\boldsymbol{\delta}_i^{\mathcal{P}} = \begin{bmatrix} \cdots & \left(\boldsymbol{\beta}_s^0 \boldsymbol{X}_{is}\right)' & \cdots \end{bmatrix}'$ with $s \in \mathcal{P}$ is $KP \times 1$, $\boldsymbol{\lambda}^{\mathcal{P}}$ is $KP \times f$, and $\boldsymbol{\varepsilon}_i^{\mathcal{P}}$ is $KP \times 1$.

To be able to recover $\boldsymbol{\mu}_i$ from the covariates and outcomes observed in $\mathcal{P}$, we need the following full rank condition, which ensures that there is enough variation in the effects of the unobserved individual characteristics over time or across different outcomes.

Assumption 3.3. *$\boldsymbol{\lambda}^{\mathcal{P}\prime}\boldsymbol{\lambda}^{\mathcal{P}}$ has rank f.*

Remark 3.3. Although we do not require the number of pretreatment outcomes to be large, Assumption 3.3 implies that KP needs to be at least as large as f. As T_0 (and thus P) is usually small in empirical microeconomics, this assumption is made more plausible by having $K > 1$, i.e., using multiple related outcomes.

Remark 3.4. The number of factors f is generally not observed. To determine f, one may use the method in Bai and Ng (2002) when both N and T are large. One may also adopt a cross-validation procedure to choose f that minimises the out-of-sample mean squared prediction error, as in Xu (2017). Although we do not estimate the interactive fixed effects term directly, we may choose the number of pretreatment outcomes that best accommodates f using cross-validation as well, which will be discussed in more details later.

Under Assumption 3.3, we can pre-multiply both sides of equation (3.6) by $(\boldsymbol{\lambda}^{\mathcal{P}\prime}\boldsymbol{\lambda}^{\mathcal{P}})^{-1}\boldsymbol{\lambda}^{\mathcal{P}\prime}$ to obtain

$$\boldsymbol{\mu}_i = (\boldsymbol{\lambda}^{\mathcal{P}\prime}\boldsymbol{\lambda}^{\mathcal{P}})^{-1}\boldsymbol{\lambda}^{\mathcal{P}\prime}(\boldsymbol{Y}_i^{\mathcal{P}} - \boldsymbol{\delta}_i^{\mathcal{P}} - \boldsymbol{\varepsilon}_i^{\mathcal{P}}). \tag{3.7}$$

Substituting (3.7) into $Y_{it,k}^0 = \boldsymbol{X}_{it}'\boldsymbol{\beta}_{t,k}^0 + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_{t,k}^0 + \varepsilon_{it,k}^0$, $t > T_0$, and with a little abuse on the notation by omitting the superscript $\mathcal{P}$ on the new coefficients and error

term, we have

$$Y^0_{it,k} = \boldsymbol{X}'_{it}\boldsymbol{\beta}^0_{t,k} \underbrace{- \cdots - \boldsymbol{X}'_{is}\boldsymbol{\alpha}^0_{st,k} - \cdots}_{P \text{ terms}} + \boldsymbol{Y}_i^{\mathcal{P}\prime}\boldsymbol{\gamma}^0_{t,k} + e^0_{it,k}, \tag{3.8}$$

where

$$\begin{aligned}
\boldsymbol{\alpha}^0_{st,k} &= \boldsymbol{\beta}_s^{0\prime}\boldsymbol{\lambda}^0_s(\boldsymbol{\lambda}^{\mathcal{P}\prime}\boldsymbol{\lambda}^{\mathcal{P}})^{-1}\boldsymbol{\lambda}^0_{t,k}, \quad s \in \mathcal{P}, \\
\boldsymbol{\gamma}^0_{t,k} &= \boldsymbol{\lambda}^{\mathcal{P}}(\boldsymbol{\lambda}^{\mathcal{P}\prime}\boldsymbol{\lambda}^{\mathcal{P}})^{-1}\boldsymbol{\lambda}^0_{t,k}, \\
e^0_{it,k} &= \varepsilon^0_{it,k} - \boldsymbol{\gamma}^0_{t,k}{}'\boldsymbol{\varepsilon}_i^{\mathcal{P}}.
\end{aligned}$$

Let $Z = r(P{+}1){+}KP$. If we denote the $Z{\times}1$ vector of observables $[\boldsymbol{X}'_{it} \cdots \boldsymbol{X}'_{is} \cdots \boldsymbol{Y}_i^{\mathcal{P}\prime}]'$ as $\boldsymbol{Z}_{it}$, and the $Z \times 1$ vector of coefficients $[\boldsymbol{\beta}^0_{t,k}{}' \cdots \boldsymbol{\alpha}^0_{st,k}{}' \cdots \boldsymbol{\gamma}^0_{t,k}{}']'$ as $\boldsymbol{\theta}^0_{t,k}$, then equation (3.8) can be abbreviated as

$$Y^0_{it,k} = \boldsymbol{Z}'_{it}\boldsymbol{\theta}^0_{t,k} + e^0_{it,k}. \tag{3.9}$$

Similarly, substituting (3.7) into $Y^1_{it,k} = \boldsymbol{X}'_{it}\boldsymbol{\beta}^1_{t,k} + \boldsymbol{\mu}'_i\boldsymbol{\lambda}^1_{t,k} + \varepsilon^1_{it,k}$, $t > T_0$, we have

$$Y^1_{it,k} = \boldsymbol{Z}'_{it}\boldsymbol{\theta}^1_{t,k} + e^1_{it,k}, \tag{3.10}$$

where

$$\begin{aligned}
\boldsymbol{\theta}^1_{t,k} &= [\boldsymbol{\beta}^1_{t,k}{}' \cdots \boldsymbol{\alpha}^1_{st,k}{}' \cdots \boldsymbol{\gamma}^1_{t,k}{}']', \\
\boldsymbol{\alpha}^1_{st,k} &= \boldsymbol{\beta}_s^{0\prime}\boldsymbol{\lambda}^0_s(\boldsymbol{\lambda}^{\mathcal{P}\prime}\boldsymbol{\lambda}^{\mathcal{P}})^{-1}\boldsymbol{\lambda}^1_{t,k}, \quad s \in \mathcal{P}, \\
\boldsymbol{\gamma}^1_{t,k} &= \boldsymbol{\lambda}^{\mathcal{P}}(\boldsymbol{\lambda}^{\mathcal{P}\prime}\boldsymbol{\lambda}^{\mathcal{P}})^{-1}\boldsymbol{\lambda}^1_{t,k}, \\
e^1_{it,k} &= \varepsilon^1_{it,k} - \boldsymbol{\gamma}^1_{t,k}{}'\boldsymbol{\varepsilon}_i^{\mathcal{P}}.
\end{aligned}$$

3.2.2 Estimation

Under Assumption 3.2, we have $\mathbb{E}(e^1_{it,k} \mid \boldsymbol{H}_{it}) = 0$ and $\mathbb{E}(e^0_{it,k} \mid \boldsymbol{H}_{it}) = 0$. This suggests using $\widehat{\tau}_{it,k} = \boldsymbol{Z}'_{it}(\widehat{\boldsymbol{\theta}}^1_{t,k} - \widehat{\boldsymbol{\theta}}^0_{t,k})$, where $\widehat{\boldsymbol{\theta}}^1_{t,k}$ and $\widehat{\boldsymbol{\theta}}^0_{t,k}$ are some estimators of $\boldsymbol{\theta}^1_{t,k}$ and $\boldsymbol{\theta}^0_{t,k}$, to estimate $\bar{\tau}_{it,k}$. Note, however, that the error terms $e^1_{it,k}$ and $e^0_{it,k}$ are correlated with the regressors, since $\boldsymbol{Z}_{it}$ contains $\boldsymbol{Y}_i^{\mathcal{P}}$ which is correlated with $\boldsymbol{\varepsilon}_i^{\mathcal{P}}$. This renders the OLS estimators biased and inconsistent, which can be seen as a classical measurement errors in variables problem.[7] We thus use the remaining

[7]This is noted in Ferman and Pinto (2019) as well, who also suggested using pre-treatment outcomes as instrumental variables to deal with the problem. Our method is also related to the quasi-differencing approach in Holtz-Eakin et al. (1988) and the GMM approach in Ahn et al. (2013). While these studies focus on estimating the coefficients on the observed covariates, our

outcomes as instrumental variables for $\boldsymbol{Y}_i^{\mathcal{P}}$ to consistently estimate $\boldsymbol{\theta}_{t,k}^1$ and $\boldsymbol{\theta}_{t,k}^0$ in each period, which would then allow us to obtain asymptotically unbiased estimates for the individual treatment effects.[8] Since the outcomes depend on about the same set of observed and unobserved individual characteristics, the remaining outcomes are strongly correlated with the outcomes included in $\boldsymbol{Y}_i^{\mathcal{P}}$. Additionally, given that the idiosyncratic shocks are independent across time, the remaining outcomes are not correlated with $e_{it,k}^1$ or $e_{it,k}^0$. Thus, both the relevance and exogeneity conditions are satisfied, and the remaining outcomes can serve as valid instrumental variables.

Let $\boldsymbol{R}_{it} = [\boldsymbol{X}_{it}' \cdots \boldsymbol{X}_{is}' \cdots \boldsymbol{Y}_i^{-\mathcal{P}\prime}]'$ be the $R \times 1$ vector of instruments, where the $(KT - KP - 1) \times 1$ vector $\boldsymbol{Y}_i^{-\mathcal{P}}$ comprises the remaining pretreatment outcomes as well as the posttreatment outcomes other than $Y_{it,k}$.[9] Stacking $\boldsymbol{Z}_{it}$, $\boldsymbol{R}_{it}$ and $\boldsymbol{Y}_{it,k}^0$ respectively over the N_0 untreated individuals, we obtain the $N_0 \times Z$ matrix of regressors $\boldsymbol{Z}_t^0$, the $N_0 \times R$ matrix of instruments $\boldsymbol{R}_t^0$ and the $N_0 \times 1$ matrix of outcomes $\boldsymbol{Y}_{t,k}^0$ for the untreated individuals. We can obtain $\boldsymbol{Z}_t^1$, $\boldsymbol{R}_t^1$ and $\boldsymbol{Y}_{t,k}^1$ similarly for the N_1 treated individuals. The GMM estimator for the individual treatment effect $\bar{\tau}_{it,k}$ can then be constructed as

$$\widehat{\tau}_{it,k} = \boldsymbol{Z}_{it}' \left(\widehat{\boldsymbol{\theta}}_{t,k}^1 - \widehat{\boldsymbol{\theta}}_{t,k}^0 \right), \tag{3.11}$$

where

$$\widehat{\boldsymbol{\theta}}_{t,k}^1 = \left(\boldsymbol{Z}_t^{1\prime} \boldsymbol{R}_t^1 \boldsymbol{W}^1 \boldsymbol{R}_t^{1\prime} \boldsymbol{Z}_t^1 \right)^{-1} \boldsymbol{Z}_t^{1\prime} \boldsymbol{R}_t^1 \boldsymbol{W}^1 \boldsymbol{R}_t^{1\prime} \boldsymbol{Y}_{t,k}^1, \tag{3.12}$$

$$\widehat{\boldsymbol{\theta}}_{t,k}^0 = \left(\boldsymbol{Z}_t^{0\prime} \boldsymbol{R}_t^0 \boldsymbol{W}^0 \boldsymbol{R}_t^{0\prime} \boldsymbol{Z}_t^0 \right)^{-1} \boldsymbol{Z}_t^{0\prime} \boldsymbol{R}_t^0 \boldsymbol{W}^0 \boldsymbol{R}_t^{0\prime} \boldsymbol{Y}_{t,k}^0, \tag{3.13}$$

with $\boldsymbol{W}^1$ and $\boldsymbol{W}^0$ being some $R \times R$ positive definite matrices.

Remark 3.5. Using the residuals $\widehat{\boldsymbol{e}}_{t,k}^1 = \boldsymbol{Y}_{t,k}^1 - \boldsymbol{Z}_t^1 \widehat{\boldsymbol{\theta}}_{t,k}^1$ and $\widehat{\boldsymbol{e}}_{t,k}^0 = \boldsymbol{Y}_{t,k}^0 - \boldsymbol{Z}_t^0 \widehat{\boldsymbol{\theta}}_{t,k}^0$, we can further construct the two-step efficient GMM estimator by replacing $\boldsymbol{W}^1$ and $\boldsymbol{W}^0$ in equations (3.12) and (3.13) with $N_1 \left(\boldsymbol{R}_t^{1\prime} \boldsymbol{U}_t^1 \boldsymbol{R}_t^1 \right)^{-1}$ and $N_0 \left(\boldsymbol{R}_t^{0\prime} \boldsymbol{U}_t^0 \boldsymbol{R}_t^0 \right)^{-1}$, where $\boldsymbol{U}_t^1$ and $\boldsymbol{U}_t^0$ are diagonal matrices with the squared elements of $\widehat{\boldsymbol{e}}_{t,k}^1$ and $\widehat{\boldsymbol{e}}_{t,k}^0$ on the diagonals.

---

focus is on estimating the individual treatment effects.

[8] We may construct the vectors of regressors and instruments differently under alternative assumptions on the dependence structure of the idiosyncratic shocks. For example, if the idiosyncratic shocks are correlated across time but are independent across outcomes, then we can split different outcomes into regressors and instruments. This would be similar to using the characteristics of similar products (Berry et al., 1995) or trading countries (see the Trade-weighted World Income instrument in Acemoglu et al., 2008) as instrumental variables. Incorporating more complex structures of the idiosyncratic shocks in the model is left for future research.

[9] In the special case of $T_1 = 1$ and $T_0 = 1$, we can include $K - 1$ pretreatment outcomes as regressors, and use the posttreatment outcomes other than $Y_{it,k}$ as instruments so that $R \geq Z$.

Remark 3.6. One may also construct the estimators for the individual treatment effects using authentic predicted outcomes obtained from a leave-one-out procedure, where $\boldsymbol{\theta}_{t,k}^1$ and $\boldsymbol{\theta}_{t,k}^0$ are estimated for each individual using the sample that excludes that individual. This procedure may be computationally expensive though, as there are no simple linear expressions for the leave-one-out coefficients estimates and residuals as for those in linear regression (Hansen, 2021).

The following result shows that the bias of the GMM estimator for the individual treatment effect in (3.11) goes away as both the number of treated individuals and the number of untreated individuals become larger.

Proposition 3.1. *Under Assumptions 3.1-3.3,* $\mathbb{E}\left(\widehat{\tau}_{it,k} - \tau_{it,k} \mid \boldsymbol{H}_{it} = \boldsymbol{h}_{it}\right) \to 0$ *as* $N_1, N_0 \to \infty$.

Once we have the estimates for the individual treatment effects, the average treatment effect $\tau_{t,k} = \mathbb{E}\left(\tau_{it,k}\right)$ can be conveniently estimated using the average of the estimated individual treatment effects $\widehat{\tau}_{t,k} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\tau}_{it,k}$, which can be shown to be consistent.

Proposition 3.2. *Under Assumptions 3.1-3.3,* $\widehat{\tau}_{t,k} - \tau_{t,k} \xrightarrow{p} 0$ *as* $N_0, N_1 \to \infty$, *and* $\widehat{\tau}_{t,k} - \tau_{t,k} = O_p\left(N_1^{-1/2}\right) + O_p\left(N_0^{-1/2}\right)$.

3.2.3 Model Selection

To satisfy Assumption 3.3, we need the number of pretreatment outcomes that we include as regressors in the model to be at least as large as f. Including more pretreatment outcomes may increase the variance of the estimator by increasing the variances of $\widehat{\boldsymbol{\theta}}_{t,k}^1$ and $\widehat{\boldsymbol{\theta}}_{t,k}^0$, but may also reduce the variance of the estimator when the sample is large and the variances of $\widehat{\boldsymbol{\theta}}_{t,k}^1$ and $\widehat{\boldsymbol{\theta}}_{t,k}^0$ are small, since

$$\left(\boldsymbol{\gamma}_{t,k}^1 - \boldsymbol{\gamma}_{t,k}^0\right)' \boldsymbol{\varepsilon}_i^{\mathcal{P}} = \frac{1}{KP}\sum_{q\in\mathcal{K}}\sum_{s\in\mathcal{P}}\left(\boldsymbol{\lambda}_{t,k}^1 - \boldsymbol{\lambda}_{t,k}^0\right)' \left(\frac{1}{KP}\sum_{l\in\mathcal{K}}\sum_{n\in\mathcal{P}}\boldsymbol{\lambda}_{n,l}^0 \boldsymbol{\lambda}_{n,l}^{0\prime}\right)^{-1} \boldsymbol{\lambda}_{s,q}^0 \varepsilon_{is,q}^0 \tag{3.14}$$

in the prediction error converges in probability to 0 as KP grows.[10]

To select the number of pretreatment outcomes to include in the model, we follow a model selection procedure similar to that in Hsiao et al. (2012), where for each usable

[10]Consistency of the individual treatment effect estimator may also be shown by allowing both N and KP to grow, with restrictions on the relative growth rate, e.g., $\frac{KP}{\min\left(\sqrt{N_1},\sqrt{N_0}\right)} \to 0$. We do not pursue this path in this study, as the number of pretreatment outcomes in empirical microeconomics that we focus on is usually not large.

number of pretreatment outcomes, we construct many different models by including a random subset of the pretreatment outcomes as regressors and the remaining outcomes as instruments. We then estimate the models using GMM and obtain the leave-one-out prediction errors for all or a subsample of the individuals. The best set of pretreatment outcomes is chosen as the one that minimises the mean squared leave-one-out prediction error.[11]

In addition to the models using only a subset of the pretreatment outcomes, we also consider averaging different models that use the same number of pretreatment outcomes. Since the estimators constructed using only a subset of the pretreatment outcomes are asymptotically unbiased, as long as the number of pretreatment outcomes is larger than f, this property is passed on to the averaged estimator. The averaged estimator may also be more efficient as it uses more information in the sample and reduces uncertainty caused by a small number of sample splits.[12] The leave-one-out prediction errors are also averaged over the models, and the best number of pretreatment outcomes to be used for the averaged estimator is similarly determined by minimising the mean squared leave-one-out prediction error.

3.2.4 Related methods

Linear conditional mean

An alternative approach to estimating the treatment effects is to follow Hsiao et al. (2012) and assume that

$$\mathbb{E}\left(\boldsymbol{\varepsilon}_i^{\mathcal{P}} \mid \boldsymbol{Z}_{it}\right) = \boldsymbol{C}'\boldsymbol{Z}_{it}, \tag{3.15}$$

where $\boldsymbol{C} = \mathbb{E}\left(\boldsymbol{Z}_{it}\boldsymbol{Z}_{it}'\right)^{-1}\mathbb{E}\left(\boldsymbol{Z}_{it}\boldsymbol{\varepsilon}_i^{\mathcal{P}\prime}\right)$ is $Z \times KT_0$.[13] We can then separate the error term into a part correlated with the regressors and a part that has zero conditional

[11] An alternative way to select the best set of pretreatment outcomes is to use information criteria such as GMM-BIC and GMM-AIC (Andrews, 1999). To avoid the potential problem of post-selection inference, we may also randomly split the sample into two parts, where we select the best model on one part, and conduct inference on the other.

[12] We stick with simple averaging in this paper. More flexible averaging scheme, e.g., with larger weights on those with smaller out of sample prediction errors, would be an interesting direction for future research.

[13] This assumption holds in special cases, e.g., when the unobserved predictors and the idiosyncratic shocks all follow the normal distribution (Li and Bell, 2017). In more general cases, this assumption may be considered to hold approximately.

mean, and rewrite the untreated potential outcome $Y^0_{it,k}$ as

$$\begin{aligned} Y^0_{it,k} &= \mathbb{E}\left(Y^0_{it,k} \mid \boldsymbol{Z}_{it}\right) + u^0_{it,k} \\ &= \left(\boldsymbol{\theta}^{0\,\prime}_{t,k} - \boldsymbol{\gamma}^{0\,\prime}_{t,k}\boldsymbol{C}'\right)\boldsymbol{Z}_{it} + u^0_{it,k} \\ &= \boldsymbol{Z}'_{it}\boldsymbol{\theta}^{*0}_{t,k} + u^0_{it,k}, \end{aligned} \tag{3.16}$$

where $u^0_{it,k} = \varepsilon^0_{it,k} - \boldsymbol{\gamma}^{0\,\prime}_{t,k}\boldsymbol{\varepsilon}^{\mathcal{P}}_i + \boldsymbol{\gamma}^{0\,\prime}_{t,k}\boldsymbol{C}'\boldsymbol{Z}_{it}$. Similarly, the treated potential outcome $Y^1_{it,k}$ can be rewritten as

$$Y^1_{it,k} = \boldsymbol{Z}'_{it}\boldsymbol{\theta}^{*1}_{t,k} + u^1_{it,k}, \tag{3.17}$$

where $\boldsymbol{\theta}^{*1}_{t,k} = \boldsymbol{\theta}^1_{t,k} - \boldsymbol{C}\boldsymbol{\gamma}^1_{t,k}$, and $u^1_{it,k} = \varepsilon^1_{it,k} - \boldsymbol{\gamma}^{1\,\prime}_{t,k}\boldsymbol{\varepsilon}^{\mathcal{P}}_i + \boldsymbol{\gamma}^{1\,\prime}_{t,k}\boldsymbol{C}'\boldsymbol{Z}_{it}$.

Since $\mathbb{E}(u^1_{it,k} \mid \boldsymbol{Z}_{it}) = \mathbb{E}[e^1_{it,k} - \mathbb{E}(e^1_{it,k} \mid \boldsymbol{Z}_{it}) \mid \boldsymbol{Z}_{it}] = 0$ and $\mathbb{E}(u^0_{it,k} \mid \boldsymbol{Z}_{it}) = 0$, it is straightforward to show that the least squares estimators $\widehat{\boldsymbol{\theta}}^{*1}_{t,k} = (\boldsymbol{Z}^{1\prime}_t\boldsymbol{Z}^1_t)^{-1}\boldsymbol{Z}^{1\prime}_t\boldsymbol{Y}^1_{t,k}$ and $\widehat{\boldsymbol{\theta}}^{*0}_{t,k} = (\boldsymbol{Z}^{0\prime}_t\boldsymbol{Z}^0_t)^{-1}\boldsymbol{Z}^{0\prime}_t\boldsymbol{Y}^0_{t,k}$ are the unbiased estimators of $\boldsymbol{\theta}^{*1}_{t,k}$ and $\boldsymbol{\theta}^{*0}_{t,k}$ respectively.[14]

We can then construct an estimator as

$$\tilde{\tau}_{it,k} = \boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^{*1}_{t,k} - \widehat{\boldsymbol{\theta}}^{*0}_{t,k}\right), \tag{3.18}$$

which is an unbiased estimator for the average treatment effect for individuals with the same values of $\boldsymbol{Z}_{it}$, or the conditional average treatment effect. It follows that the average of the conditional average treatment effects estimators $\tilde{\tau}_{t,k} = \frac{1}{N}\sum_{i=1}^N \tilde{\tau}_{it,k}$ is an unbiased estimator for the average treatment effect $\tau_{t,k}$. In addition, it can also be shown that $\tilde{\tau}_{t,k}$ is a consistent estimator without imposing the linear conditional mean assumption (Li and Bell, 2017).

Proposition 3.3. *Under Assumptions 3.1-3.3,*

(i) if $\mathbb{E}\left(\boldsymbol{\varepsilon}^{\mathcal{P}}_i \mid \boldsymbol{Z}_{it}\right) = \boldsymbol{C}'\boldsymbol{Z}_{it}$, *then* $\mathbb{E}\left(\tilde{\tau}_{it,k} - \tau_{it,k} \mid \boldsymbol{Z}_{it} = \boldsymbol{z}_{it}\right) = 0$ *and* $\mathbb{E}\left(\tilde{\tau}_{t,k} - \tau_{t,k}\right) = 0$;

(ii) $\tilde{\tau}_{t,k} - \tau_{t,k} = O_p\left(N_1^{-1/2}\right) + O_p\left(N_0^{-1/2}\right)$.

Remark 3.7. Note that $\mathbb{E}\left(\tau_{it,k} \mid \boldsymbol{Z}_{it} = \boldsymbol{z}_{it}\right)$ is the average treatment effect for individuals with $\boldsymbol{Z}_{it} = \boldsymbol{z}_{it}$, or the conditional average treatment effect, whereas the individual treatment effect is $\mathbb{E}\left(\tau_{it,k} \mid \boldsymbol{H}_{it} = \boldsymbol{h}_{it}\right)$ as given in (3.4). The two are generally not the same since $\boldsymbol{C}'\boldsymbol{Z}_{it} \neq 0$.

[14] The linear conditional mean assumption also implies that the unconfoundedness assumption is satisfied, as $\mathbb{E}(Y^0_{it,k} \mid \boldsymbol{Z}_{it}, D_{it} = 1) = \mathbb{E}(Y^0_{it,k} \mid \boldsymbol{Z}_{it}, D_{it} = 0)$ and $\mathbb{E}(Y^1_{it,k} \mid \boldsymbol{Z}_{it}, D_{it} = 1) = \mathbb{E}(Y^1_{it,k} \mid \boldsymbol{Z}_{it}, D_{it} = 0)$.

Interactive fixed effects model

Instead of replacing the unobserved confounders with the observed pretreatment outcomes, Bai (2009) models the unobserved fixed effects directly by iterating between estimating the coefficients on the observed covariates and estimating the unobserved factors and factor loadings using the principal component analysis, given some initial values. This approach allows more general structures in the error terms, but requires both N and T to be large, and is also more restrictive on the model specification: the observed covariates need to be time-varying, while the coefficients are assumed constant over time. Xu (2017) adapts this method to the potential outcomes framework to estimate the average treatment effects on the treated, assuming that the untreated potential outcomes for both the treated and untreated units follow the interactive fixed effects model, and proposes a cross-validation procedure to choose the number of unobserved factors and a parametric bootstrap procedure for inference.

This approach has the desired feature of being less computationally expensive compared with repeated pretreatment set splitting and averaging, and is potentially more efficient compared with using only the best set of pretreatment outcomes and discarding the remaining information when all outcomes are related. However, its potential to be adapted to our settings is limited by the restrictions discussed above. In particular, we may assume the coefficients to be constant over time, but it would be unrealistic to assume that they are the same across different outcomes, if we were to use multiple related outcomes.

Another closely related study is Athey et al. (2021), which generalises the results from the matrix completion literature in computer science to impute the missing elements of the untreated potential outcome matrix for the treated units in the posttreatment periods, where the matrix is assumed to have a low rank structure, similar to that of the interactive fixed effects model. The bias of the estimator is shown to have an upper bound that goes to 0 as both N and T grow. This method allows staggered adoption of the treatment, i.e., the treated units receive the treatment at different time periods.

Synthetic control method

Abadie et al. (2010) estimate the treatment effect on a treated unit by predicting its untreated potential outcome using a synthetic control constructed as a weighted average of the control units. The synthetic control method applies to cases where the pretreatment characteristics of the treated unit can be closely approximated by the synthetic control constructed using a small number of control units over an

extended period of time before the treatment, which may not generally hold. In terms of implementation, the objective function for the synthetic control method is similar to that of the linear regression approach in Hsiao et al. (2012). However, the weights on the control units in the synthetic control method are restricted to be nonnegative to avoid extrapolation. This reduces the risk of overfitting, but may also limit its applicability by making it difficult to find a set of weights that satisfy the restrictions.

3.2.5 Inference

To measure the conditional variance of the individual treatment effect estimator, $\text{Var}\left(\widehat{\tau}_{it,k} \mid \boldsymbol{H}, \boldsymbol{D}\right)$, where $\boldsymbol{H}$ is the matrix of observed covariates and unobserved individual characteristics and $\boldsymbol{D}$ is the matrix of the treatment status for all individuals and all time periods in the sample, we follow Xu (2017) and employ a parametric bootstrap procedure.

First, we apply our method to all outcomes in all periods to obtain $\widehat{Y}^1_{it,k}$ and $\widehat{e}^1_{it,k}$ for the treated individuals in the posttreatment periods, and $\widehat{Y}^0_{it,k}$ and $\widehat{e}^0_{it,k}$ for the untreated individuals in the posttreatment periods and for all individuals in the pretreatment periods. Note that the residuals $\widehat{e}^1_{it,k}$ and $\widehat{e}^0_{it,k}$ are estimates for $\varepsilon^1_{it,k} - \boldsymbol{\gamma}^1_{t,k}{}'\boldsymbol{\varepsilon}^{\mathcal{P}}_i$ and $\varepsilon^0_{it,k} - \boldsymbol{\gamma}^0_{t,k}{}'\boldsymbol{\varepsilon}^{\mathcal{P}}_i$, respectively, rather than the idiosyncratic shocks in the original model, $\varepsilon^1_{it,k}$ and $\varepsilon^0_{it,k}$. Thus, the variance of the individual treatment effect estimator tends to be overestimated using the parametric bootstrap by resampling these residuals, especially when the number of pretreatment outcomes is small.[15] Correcting for this bias would be a necessary step for future research.

These fitted values of the outcomes can be stacked into a $TK \times 1$ vector $\widehat{\boldsymbol{Y}}_i$ for each individual, where $\widehat{\boldsymbol{Y}}_i$ for $i \in \mathcal{T}$ contains $\widehat{Y}^1_{it,k}$ in the posttreatment periods and $\widehat{Y}^0_{it,k}$ in the pretreatment periods, and $\widehat{\boldsymbol{Y}}_i$ for $i \in \mathcal{C}$ contains $\widehat{Y}^0_{it,k}$ in all periods. The $TK \times 1$ vector of residuals $\widehat{\boldsymbol{e}}_i$ can be obtained similarly.

We then start bootstrapping for B rounds:

1. In round $b \in \{1, \ldots, B\}$, generate a bootstrapped sample as

$$\boldsymbol{Y}^{(b)}_i = \widehat{\boldsymbol{Y}}_i + \widehat{\boldsymbol{e}}^{(b)}_i, \text{ for all } i,$$

   where $\widehat{\boldsymbol{e}}^{(b)}_i$ is randomly drawn from $\{\widehat{\boldsymbol{e}}_i\}_{i \in \mathcal{T}}$ for $i \in \mathcal{T}$, and from $\{\widehat{\boldsymbol{e}}_i\}_{i \in \mathcal{C}}$ for $i \in \mathcal{C}$.[16]

[15] See the discussion on equation (3.14).

[16] Since the entire series of residuals over the T periods and K outcomes are resampled, correlation and heteroskedasticity across time and outcomes are preserved (Xu, 2017).

2. Construct $\widehat{\tau}_{it,k}^{(b)}$ for each i using the above bootstrapped sample.

The variance for the individual treatment effect estimator is computed using the bootstrap estimates as

$$\mathrm{Var}\left(\widehat{\tau}_{it,k} \mid \boldsymbol{H}, \boldsymbol{D}\right) = \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\tau}_{it,k}^{(b)} - \frac{1}{B}\sum_{a=1}^{B}\widehat{\tau}_{it,k}^{(a)}\right), \quad i = 1, \ldots, N,$$

and the $100(1-\alpha)\%$ confidence intervals for $\bar{\tau}_{it,k}$, $i = 1, \ldots, N$ can be constructed as

$$\left[\widehat{\tau}_{it,k}^{\left[\frac{\alpha}{2}B\right]}, \ \widehat{\tau}_{it,k}^{\left[(1-\frac{\alpha}{2})B\right]}\right],$$

where the superscript denotes the index of the bootstrap estimates in ascending order. Alternatively, we can use a normal approximation and construct the confidence intervals as

$$\left[\widehat{\tau}_{it,k} + \Phi^{-1}\left(\frac{\alpha}{2}\right)\widehat{\sigma}_{it,k}\,, \ \widehat{\tau}_{it,k} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\widehat{\sigma}_{it,k}\right],$$

where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution, and $\widehat{\sigma}_{it,k} = \sqrt{\mathrm{Var}\left(\widehat{\tau}_{it,k} \mid \boldsymbol{H}, \boldsymbol{D}\right)}$.

The variance for the average treatment effect estimator $\widehat{\tau}_{t,k} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\tau}_{it,k}$ and the confidence interval for the average treatment effect $\tau_{t,k}$ can be obtained in similar manners using the bootstrap estimates $\widehat{\tau}_{t,k}^{(b)} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\tau}_{it,k}^{(b)}$, $b = 1, \ldots, B$.

3.3 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to assess the performance of our estimator in small samples, and compare it with related methods in relevant settings. The number of posttreatment period T_1 is fixed at 1, and the number of related outcomes K is fixed at 5 in all settings.

The untreated potential outcomes are generated from

$$Y_{it,k}^{0} = \boldsymbol{X}_{it}'\boldsymbol{\beta}_{t,k}^{0} + \boldsymbol{\mu}_i'\boldsymbol{\lambda}_{t,k}^{0} + \varepsilon_{it,k}^{0}, \quad k \in \mathcal{K}, \tag{3.19}$$

where $\boldsymbol{X}_{it}$ contains 2 observed covariates, and $\boldsymbol{\mu}_i$ contains 2 unobserved individual characteristics as well as the constant 1. The 2 observed covariates are i.i.d. $N(0,1)$ in period 1, and then follow an AR(1) process, $\boldsymbol{X}_{it} = 0.9\boldsymbol{X}_{i,t-1} + \xi_{it}$, where ξ_{it} are i.i.d. $N(0, \sqrt{1-0.9^2})$, so that the observed covariates are correlated across time and the variances stay 1. The 2 unobserved individual characteristics are also i.i.d. $N(0,1)$. The coefficients $\boldsymbol{\beta}_{t,k}^{0}$ and $\boldsymbol{\lambda}_{t,k}^{0}$ are i.i.d. $N(\omega_k, 1)$ with $\omega_k \sim N(1,1)$, for $k \in \mathcal{K}$, so that the means of the coefficients differ across outcomes, and the

idiosyncratic shocks $\varepsilon^0_{it,k}$ are i.i.d. $N(0,1)$.

The individual treatment effect in the posttreatment period $\bar{\tau}_{iT_0+1,k}$ is a deterministic function of $\boldsymbol{X}_{it}$ and $\boldsymbol{\mu}_i$ with the coefficients being i.i.d. $N(0.5, 0.5)$, for $k \in \mathcal{K}$. And the observed outcomes $Y_{it,k}$, $k \in \mathcal{K}$ are equal to $Y^0_{it,k} - \varepsilon^0_{it,k} + \bar{\tau}_{iT_0+1,k} + \varepsilon^1_{it,k}$, where $\varepsilon^1_{it,k}$ are i.i.d. $N(0,1)$, for the treated individuals in the posttreatment period, and $Y^0_{it,k}$ otherwise. $\boldsymbol{X}_{it}$ and $\boldsymbol{\mu}_i$ as well as their coefficients for the untreated potential outcomes and the treatment effects are drawn 5 times, and for each set of $\{\boldsymbol{X}_{it}, \boldsymbol{\mu}_i\}$ and their coefficients drawn, $\varepsilon^0_{it,k}$ and $\varepsilon^1_{it,k}$ are drawn 1000 times, which allows us to compute the bias and variance of the estimator conditional on the observed covariates and the unobserved individual characteristics.

To measure the performances of the estimators, we compute the biases and standard deviations for the estimates of the individual treatment effects and the average treatment effect for outcome K in the posttreatment period. Specifically, the bias of the individual treatment effect estimator $\widehat{\tau}_{iT_0+1,K}$ is measured by $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{5}\sum_{d=1}^{5}\left|\mathbb{E}\left(\widehat{\tau}^{(d,s)}_{iT_0+1,K}\right) - \bar{\tau}^{(d)}_{iT_0+1,K}\right|$, where the superscript d denotes the dth draw of $\{\boldsymbol{X}_{it}, \boldsymbol{\mu}_i\}$ and s denotes the sth draw of $\varepsilon^0_{it,k}$ and $\varepsilon^1_{it,k}$, and the standard deviation is constructed as $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{5}\sum_{d=1}^{5}\sqrt{\mathbb{E}\left(\widehat{\tau}^{(d,s)}_{iT_0+1,K} - \mathbb{E}\widehat{\tau}^{(d,s)}_{iT_0+1,K}\right)^2}$. Similarly, the bias of the average treatment effect estimator $\widehat{\tau}_{T_0+1,K}$ is measured by $\frac{1}{5}\sum_{d=1}^{5}\left|\mathbb{E}\left(\widehat{\tau}^{(d,s)}_{T_0+1,K}\right) - \bar{\tau}^{(d)}_{T_0+1,K}\right|$, and the standard deviation is constructed as $\frac{1}{5}\sum_{d=1}^{5}\sqrt{\mathbb{E}\left(\widehat{\tau}^{(d,s)}_{T_0+1,K} - \mathbb{E}\widehat{\tau}^{(d,s)}_{T_0+1,K}\right)^2}$.[17]

[17]The performance of the estimators can also be measured using RMSE, which is computed as $\frac{1}{5}\sum_{d=1}^{5}\sqrt{\mathbb{E}\left(\widehat{\tau}^{(d,s)}_{iT_0+1,K} - \mathbb{E}\bar{\tau}^{(d,s)}_{iT_0+1,K}\right)^2}$ for $\widehat{\tau}_{iT_0+1,K}$ and $\frac{1}{5}\sum_{d=1}^{5}\sqrt{\mathbb{E}\left(\widehat{\tau}^{(d,s)}_{T_0+1,K} - \mathbb{E}\bar{\tau}^{(d,s)}_{T_0+1,K}\right)^2}$ for $\widehat{\tau}_{T_0+1,K}$. Since the biases of our estimators are small, these measures are quite similar to SD and are thus omitted from reporting.

Table 3.1: Simulation Results on Model Selection

			Best Set					Model Averaging				
				ITE		ATE			ITE		ATE	
N_1	N_0	T_0	P	Bias	SD	Bias	SD	P	Bias	SD	Bias	SD
50	50	1	2.2	0.151	1.225	0.082	0.384	2.3	0.231	1.163	0.120	0.336
100	100	1	2.2	0.076	0.764	0.032	0.212	2.4	0.065	0.836	0.024	0.205
200	200	1	2.1	0.038	0.476	0.004	0.127	2.6	0.046	0.712	0.011	0.133
50	50	2	2.6	0.062	0.875	0.014	0.253	2.9	0.150	0.758	0.040	0.232
100	100	2	2.7	0.035	0.685	0.003	0.165	2.9	0.073	0.563	0.014	0.159
200	200	2	3.2	0.038	0.729	0.003	0.137	4.0	0.031	0.702	0.003	0.131

Note: This table compares the estimator using only the best set of pretreatment outcomes and the estimator constructed from model averaging, in terms of the optimal number of pretreatment outcomes selected by LOO cross-validation, as well as the bias and SD for the ITE and ATE estimates, with varying sample size and number of pretreatment periods, based on 5000 simulations for each setting.

Table 3.1 compares the GMM estimator constructed using only the best set of pretreatment outcomes with that constructed by averaging estimators from different models with the same number of pretreatment outcomes. We see that the best number of pretreatment outcomes, P, is slightly larger than the number of unobserved individual characteristics ($f = 2$) for both estimators, and increases when the sample size is larger and when there are more pretreatment outcomes available, which is in line with our discussions in section 3.2.3. The estimators constructed by model averaging also tends to select a slightly larger P than the estimator using only the best set of pretreatment outcomes.

In almost all settings, the estimator using only the best set of pretreatment outcomes tends to have a smaller bias, whereas the estimator constructed from model averaging tends to have a smaller variance, except for estimating the individual treatment effects when the number of pretreatment outcomes is very small. The bias and SD also become smaller for both estimators when the sample size as well as the number of pretreatment outcomes grow.

In the following simulations, we fix P at 2 when $T_0 = 1$, and 3 when $T_0 = 2$, and construct the GMM estimator using only the best set of pretreatment outcomes, with the best set of pretreatment outcomes selected at the first simulation and used for the remaining simulations for each setting.[18]

[18]This is mainly to save computing time and does not fundamentally change the conclusions.

Table 3.2: Simulation Results for the GMM estimator

			ITE			ATE		
N_1	N_0	T_0	Bias	SD	Coverage	Bias	SD	Coverage
Panel A: $\varepsilon^0_{it,k}$ uncorrelated across t and k								
50	50	1	0.096	1.422	0.997	0.040	0.443	0.995
100	100	1	0.043	0.856	0.992	0.005	0.226	0.984
50	50	2	0.043	1.163	0.973	0.011	0.288	0.959
100	100	2	0.025	0.900	0.953	0.005	0.181	0.956
Panel B: $\varepsilon^0_{it,k}$ correlated across t and k								
50	50	1	0.134	1.419	0.996	0.045	0.431	0.993
100	100	1	0.065	0.851	0.991	0.018	0.230	0.982
50	50	2	0.063	1.162	0.976	0.008	0.294	0.961
100	100	2	0.037	0.906	0.964	0.009	0.183	0.967

Note: This table compares the bias and SD of the GMM estimator, as well as the coverage probability of the 95% confidence interval, with varying sample size and number of pretreatment periods, based on 5000 simulations for each setting.

Table 3.2 reports the bias and SD of the GMM estimator, as well as the coverage probability of the 95% confidence interval, for estimating the individual treatment effects and the average treatment effect. Panel A shows that the bias and SD for the estimators are small even with a small sample size and a small number of pretreatment outcomes. However, the 95% confidence intervals tend to have larger coverage probabilities, especially when the number of pretreatment outcomes is small. This distortion is alleviated as more pretreatment outcomes are available.

Since the validity of the GMM estimator relies on the assumption that $\varepsilon_{it,k}$ are uncorrelated across time or outcomes, we examine the performance of the estimator when this assumption is violated in Panel B, where the idiosyncratic shocks follow an AR(1) process over time with the autoregression coefficient being 0.1, and are correlated across outcomes by sharing a common component for different outcomes in the same period. This slightly increases the biases and SD's of the estimators, but the performance of the estimators are still quite good, especially in comparison with related methods as shown in the following tables.

Table 3.3: Simulation

			OLS				GMM			
			ITE		ATE		ITE		ATE	
N_1	N_0	T_0	Bias	SD	Bias	SD	Bias	SD	Bias	SD
Panel A: Linear conditional mean										
100	100	1	0.099	0.622	0.041	0.186	0.113	1.297	0.086	0.257
200	200	1	0.059	0.415	0.015	0.119	0.011	0.430	0.003	0.129
100	100	2	0.046	0.677	0.012	0.158	0.026	0.901	0.003	0.179
200	200	2	0.064	0.547	0.012	0.121	0.028	0.937	0.003	0.142
Panel B: Nonlinear conditional mean										
100	100	1	0.097	0.687	0.057	0.199	0.025	0.806	0.011	0.244
200	200	1	0.140	0.456	0.040	0.122	0.016	0.552	0.003	0.149
100	100	2	0.077	0.734	0.015	0.173	0.115	1.118	0.007	0.213
200	200	2	0.093	0.551	0.004	0.116	0.025	0.910	0.003	0.142

Note: This table compares the bias and SD for the OLS estimator and the GMM estimator, with varying sample size and number of pretreatment periods, based on 5000 simulations for each setting.

Table 3.3 compares our method with the OLS approach in Hsiao et al. (2012). In panel A, both the unobserved individual characteristics and the idiosyncratic shocks are normally distributed so that the linear conditional mean assumption is satisfied. The results show that the GMM estimator outperforms the OLS estimator by having a smaller bias in estimating both the individual treatment effects and the average treatment effect, although the variance of the GMM estimator is also larger.

In panel B, the unobserved individual characteristics are drawn from the uniform distribution, and the linear conditional mean assumption is no longer satisfied (Li and Bell, 2017). We see that the results are virtually unchanged for the GMM estimator, while the OLS estimator performs slightly worse by having larger biases and SD's, which is more pronounced in estimating the individual treatment effects. The results indicate that the linear conditional mean assumption is not a very strong one. Indeed, the distribution of the sum of several random variables would become more bell-shaped like the normal distribution under fairly general conditions, as a result of the central limit theorem. The simulation results are very similar when the unobserved individual characteristics are drawn from a mix of other distributions.

Table 3.4: Simulation

			IFE				GMM			
			ITT		ATT		ITT		ATT	
N_1	N_0	T_0	Bias	SD	Bias	SD	Bias	SD	Bias	SD
Panel A: $\boldsymbol{X}_{it}$ constant across t										
5	100	1	1.203	1.357	0.657	0.610	0.047	1.499	0.016	0.687
5	200	1	1.264	1.229	0.492	0.548	0.089	1.624	0.023	0.730
5	100	2	0.922	1.140	0.263	0.518	0.034	1.265	0.015	0.585
5	200	2	0.982	1.147	0.378	0.520	0.040	1.387	0.018	0.634
Panel B: $\boldsymbol{\beta}_{t,k}^0$ constant across t										
5	100	1	1.289	1.220	0.773	0.579	0.029	1.349	0.016	0.616
5	200	1	1.681	1.370	0.836	0.620	0.034	1.563	0.020	0.713
5	100	2	0.930	1.070	0.440	0.486	0.026	1.261	0.017	0.577
5	200	2	1.417	1.083	1.015	0.489	0.031	1.250	0.011	0.564

Note: This table compares the bias and SD for the IFE estimator and the GMM estimator, with varying sample size and number of pretreatment periods, based on 5000 simulations for each setting.

Table 3.4 compares our method with the method of estimating the interactive fixed effects model directly, which was first developed in Bai (2009) and then adapted into the potential outcomes framework by Xu (2017) to allow heterogeneous treatment effects. We fix the number of treated individuals at 5, and compare the performance of the two methods in estimating the individual treatment effect on the treated and the average treatment effect on the treated.

We consider two scenarios that are relevant in the context of empirical microeconomics. In panel A, the observed covariates are constant over time. This is plausible for covariates such as gender, race or education level, which are likely to be stable over time. Since the IFE method requires the observed covariates to be time-varying, the covariates that are constant over time are dropped from the estimation and become part of the unobserved individual characteristics, which makes the model equivalent to a pure factor model with 4 unobserved factors. As we have 5 related outcomes, this model should still be estimable by the IFE method. However, we see that IFE method perform poorly when there are only a small number of pretreatment outcomes to recover the unobserved individual characteristics. The bias and SD of the IFE estimator become smaller as more pretreatment outcomes are available, but are still quite large compared with our method.

To accommodate the restrictive model specification for the IFE method, we allow the covariates to be time-varying while keeping the coefficients constant over time in panel B, although the coefficients are allowed to vary across outcomes since it is unlikely that the coefficients for different outcomes would be the same in practice. We

see that the IFE estimator has poor performance since the model is still misspecified in their method, whereas the results for our method are virtually unchanged.

Table 3.5: Simulation

			SCM				GMM			
			ITT		ATT		ITT		ATT	
N_1	N_0	T_0	Bias	SD	Bias	SD	Bias	SD	Bias	SD
Panel A: distributions of $\boldsymbol{\mu}_i$ same for treated and control										
5	100	1	0.376	1.247	0.245	0.577	0.029	1.349	0.016	0.617
5	200	1	0.883	1.376	0.698	0.628	0.035	1.563	0.020	0.713
5	100	2	0.930	1.191	0.526	0.547	0.023	1.243	0.014	0.565
5	200	2	0.469	1.186	0.126	0.531	0.031	1.250	0.012	0.564
Panel B: distributions of $\boldsymbol{\mu}_i$ different for treated and control										
5	100	1	0.763	1.253	0.634	0.605	0.036	1.368	0.022	0.658
5	200	1	1.412	1.413	1.269	0.656	0.037	1.573	0.024	0.735
5	100	2	0.982	1.204	0.513	0.558	0.027	1.249	0.020	0.578
5	200	2	0.781	1.203	0.613	0.551	0.032	1.256	0.014	0.577

Note: This table compares the bias and SD for the SCM estimator and the GMM estimator, with varying sample size and number of pretreatment periods, based on 5000 simulations for each setting.

Table 3.5 compares our method with the synthetic control method (Abadie et al., 2010). In panel A, the unobserved individual characteristics for both the treated individuals and the untreated individuals are drawn from $N(0, 1)$, while in panel B, the unobserved individual characteristics for the treated individuals are drawn from $N(1, 1)$. Since the synthetic control method requires the treated units to be in the convex hull of the control units by restricting the weights assigned to the control units to be nonnegative, their method may perform poorly when the support of the unobserved individual characteristics are different for the treated and untreated individuals. While our method should be unaffected by the degree of overlapping in the distributions of the unobserved individual characteristics for the two treatment groups. The simulation results show that indeed the synthetic control estimator performs worse in panel B. Perhaps somewhat surprising is that its performance is also poor compared with our method in panel A. This is because the coefficients are outcome-specific, so that the levels of the outcomes are also likely to vary across outcomes, which makes it more difficult to obtain a good pretreatment fit under the nonnegativity restriction. In comparison, our method has good performance in both panels.

Overall, the simulation results show that our method has good performance in terms of the bias and SD in estimating the individual treatment effects and the average treatment effect under various settings, and has superior performance than related

methods. The shortcoming of our method is that the confidence intervals tend to be too wide, especially when the number of pretreatment outcomes is small.

3.4 Empirical Application

We illustrate our method by estimating the effect of health insurance coverage on the individual usage of hospital emergency departments.

Although the usage of emergency departments applies to only a small proportion of the population, it imposes great financial pressure on the health care system. In addition, it is not clear ex ante what the direction of the effect should be. E.g., Taubman et al. (2014) argues that health insurance coverage could either increase emergency-department use by reducing its cost for the patients, or decrease emergency-department use by encouraging primary care use or improving health.

The findings on emergency-department use have been mixed. Using survey data collected from the participants of the Oregon Health Insurance Experiment (OHIE) about a year after they were notified of the selection results, Finkelstein et al. (2012) find no discernible impact of health insurance coverage on emergency-department use.[19] While using the visit-level data for all emergency-department visits to twelve hospitals in the Portland area probabilistically matched to the OHIE study population on the basis of name, date of birth, and gender, Taubman et al. (2014) find that health insurance coverage significantly increases emergency-department use by 0.41 visits per person, from an average of 1.02 visits per person in the control group in the first 15 months of the experiment. They also examine whether the effect differs across heterogeneous groups, and find statistically significant increases in emergency-department use across most subgroups in terms of the number of pre-experiment emergency-department visits, hospital admission (inpatient or outpatient visits), timing (on-hours or off-hours visits), the type of visits (emergent and not preventable, emergent and preventable, primary care treatable, and non-emergent), as well as gender, age, and health condition.

In this application, we wish to estimate the effect of health insurance coverage on emergency-department use for each individual in the sample. This would poten-

[19]The Oregon Health Insurance Experiment (OHIE) was initiated in 2008, targeting at low-income adults in Oregon who had been without health insurance for at least 6 months. Among the 89,824 individuals who signed up, 35,169 individuals were randomly selected by the lottery and were eligible to apply for the Oregon Health Plan (OHP) Standard program, which provided relatively comprehensive medical benefits with no consumer cost sharing, and the monthly premiums was only between \$0 and \$20 depending on the income. As a randomised controlled experiment, the OHIE offers an opportunity for researchers to study the effect of health insurance coverage on various health outcomes without confounding factors.

tially help us better understand whether and how health insurance coverage affects emergency-department use, compared with using only the average treatment effect for the whole sample or for some preassigned subgroups (conditional average treatment effects).

Our data combines both the hospital emergency-department visit-level data and the survey data. There are two time periods, one before the randomisation and one after.[20] To estimate the individual treatment effects, we include 3 observed covariates including gender, birth year, and household income as a percentage of the federal poverty line, and 10 related outcomes including different types of emergency-department visits and medical charges. We also consider a rich list of variables on which we make comparisons for individuals with different estimated treatment effects. There are 2154 individuals with complete information on these variables.[21]

Table 3.6: Sample Selection

	Selected	Not-selected	Difference	Insured	Not-insured	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.59	0.60	-0.01	0.63	0.58	0.05*
Birth year	1966.24	1966.44	-0.19	1967.03	1966.09	0.95
Household income as percent of federal poverty line	79.77	75.67	4.10	53.73	86.57	-32.84***
# ED visits	0.32	0.43	-0.11**	0.48	0.33	0.15**
# outpatient ED visits	0.27	0.35	-0.09**	0.40	0.28	0.13**
# weekday daytime ED visits	0.18	0.24	-0.06**	0.27	0.19	0.08**
# emergent non-preventable ED visits	0.07	0.09	-0.02	0.11	0.07	0.04**
# emergent preventable ED visits	0.03	0.03	0.00	0.03	0.02	0.01
# primary care treatable ED visits	0.10	0.15	-0.05***	0.15	0.12	0.04
Total charges	859.71	1276.90	-417.19*	1379.58	947.54	432.04
Total ED charges	345.48	504.64	-159.16**	494.95	396.86	98.09
# ED visits to a high uninsured volume hospital	0.17	0.22	-0.05	0.25	0.17	0.08**
# ED visits (survey)	0.24	0.30	-0.06*	0.38	0.23	0.15***
N	1103	1051		577	1577	

1) This table compares the mean values of the covariates and related outcomes in the pretreatment period for individuals selected/not-selected by the lottery, and individuals insured/not-insured.
2) Significance levels of the two-sample t-test: * 10%, ** 5%, *** 1%.

The first 3 columns in Table 3.6 present the mean values of the covariates and outcomes in the pretreatment period for individuals selected by the lottery and for individuals not selected by the lottery, as well as the difference between the two groups. Since a considerable number of observations with incomplete information are

[20]The pre-randomisation period in the hospital visit-level data was from January 2007 to March 2008, and the post-randomisation period was from March 2008 to September 2009. The two surveys were collected shortly after the randomisation and about a year after randomisation, respectively, each covering a 6-month period before the survey.

[21]Note that our sample size is significantly smaller than the other studies using the OHIE data, due to the inclusion of the extensive list of variables. For example, the sample size in Finkelstein et al. (2012) is 74,922, and the sample size in Taubman et al. (2014) is 24,646. So our sample may not be representative of the OHIE sample and the results in different studies may not be directly comparable.

dropped, being selected by the lottery is negatively correlated with different types of emergency-department visits in the pretreatment period in our sample, which suggests that the lottery assignment is not likely to be a valid instrument for health insurance coverage. Table 3.6 also compares the mean pretreatment characteristics for individuals covered by health insurance and those not covered, which shows that individuals who were covered were poorer and used emergency-department more frequently in the pretreatment period than people who were not covered by health insurance.

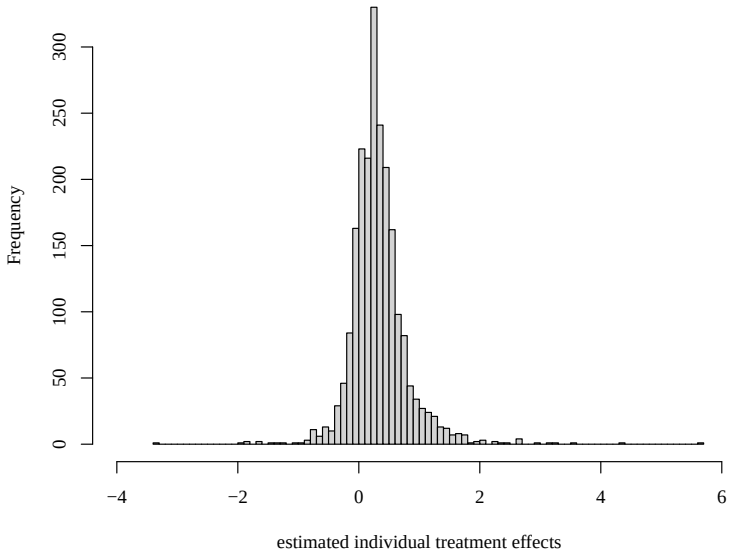


Figure 3.1: Distribution of the estimated individual treatment effects

Figure 3.1 shows the distribution of the estimated individual treatment effects using our method.[22] The mean of estimated individual treatment effects, or the estimated average treatment effect is 0.33, which is significant at 1% level. 114 individuals have treatment effects that are significant at 10% level, among which 23 are negative and 91 are positive.[23] We then move on to compare the characteristics of the individuals based on their estimated treatment effects, which are presented in Table 3.7-3.9.

[22]As mentioned earlier, since only individuals with complete information on the variables are selected into our sample, the distribution may not be representative of the OHIE participants. The distribution of the estimated individual treatment effects may be more spread out than the distribution of the true effects due to noise, or less spread out since the estimates are based on the parametric models for the potential outcomes, which may be over-simplifying compared with the true models.

[23]Note that if we were to adjust for multiple testing, e.g., using the Benjamini–Hochberg procedure to control the false discovery rate (FDR) at 10% level, then we would be left with only one individual whose treatment effect is significant. Although the small number of individuals with significant treatment effects may also be attributed to the overestimation of the variance of the individual treatment effect estimator.

Column (1) shows the mean characteristics of individuals whose treatment effects are not significant at 10% level, column (2) shows the mean characteristics of individuals whose treatment effects are significantly negative, column (3) shows the differences between column (2) and column (1), column (4) shows the mean characteristics of individuals whose treatment effects are significantly positive, and column (5) shows the differences between column (4) and column (1).

Compared with individuals who would not be significantly affected by the treatment, individuals who would significantly decrease or increase their emergency-department visits if covered by health insurance both had more emergency-department visits and more medical charges in the pretreatment period. However, these two groups were also distinct in some characteristics.

The individuals who would have fewer emergency-department visits if covered by health insurance were on average 7 years younger than individuals in the control group and 10 years younger than the positive group, more likely to be female with less education, and in particular, were much poorer than individuals in the other groups. They were more likely to be diagnosed with depression but not other conditions. Importantly, they were less likely to have any primary care visits, and more likely to use emergency department as the place for medical care. In terms of emergency-department use in the pretreatment period, they had fewer visits resulting in hospitalisation, more outpatient visits, more preventable and non-emergent visits, more visits to hospitals with a low fraction of uninsured patients, fewer visits for chronic conditions, and more visits for injury. Although their medical charges were not as high as those for individuals in the positive group, they owed more money for medical expenses.

In comparison, individuals who would have more emergency-department visits if covered by health insurance were more likely to be older, male, and with household income right above the federal poverty line, which means that they were not as poor as the individuals in the other groups. They were in worse health conditions, more likely to be diagnosed with diabetes and high blood pressure, and were taking more prescription medications. They also had more emergency-department visits of all types in the pretreatment period, including visits resulting in hospitalisation and visits for more severe conditions such as chronic conditions, chest pain and psychological conditions, and they incurred more medical charges.

Overall, these comparisons suggest that the individuals who would have fewer emergency-department visits if covered by health insurance were younger and not in very bad physical conditions. However, their access to primary care were limited due to being in much more disadvantaged positions financially, which made them

resort to using the emergency department as the usual place for medical care. In contrast, the individuals who would have more emergency-department visits if covered by health insurance were more likely to be older and in poor health. So even with access to primary care, they still used emergency departments more often for severe conditions, although sometimes for primary care treatable and non-emergent conditions as well.

All in all, it seems that both mechanisms discussed by Taubman et al. (2014) are playing a role. For people who used emergency department for medical care because they did not have access to primary care service, health insurance coverage decreases emergency-department use because it increases access to primary care and may also lead to improved health. Whereas for people who had access to primary care and still used emergency department due to worse physical conditions, health insurance coverage increases emergency-department use because it reduces the out-of-pocket cost of the visits.

This application shows the potential value of estimating individual treatment effects for policy evaluation. Our findings would not have been possible by only estimating conditional average treatment effects, as we would not be able to distinguish individuals with positive or negative treatment effects at the first place.

Table 3.7: Comparison of Characteristics

	Same	Fewer	Difference	More	Difference
	(1)	(2)	(3)	(4)	(5)
Birth year	1966.42	1973.90	7.48*	1964.09	-2.33**
Female	0.61	1.00	0.39***	0.37	-0.24***
Education	2.52	1.90	-0.62**	2.30	-0.22**
English	0.89	0.90	0.01	0.95	0.07***
Race:					
White	0.75	0.50	-0.25	0.79	0.04
Hispanic	0.10	0.10	0.00	0.12	0.02
Black	0.06	0.30	0.24	0.06	0.00
Asian	0.10	0.10	0.00	0.06	-0.05*
American Indian or Alaska Native	0.04	0.10	0.06	0.07	0.03
Native Hawaiian or Pacific Islander	0.01	0.00	-0.01***	0.01	0.00
Other races	0.07	0.10	0.03	0.06	-0.01
Employed	0.53	0.60	0.07	0.48	-0.05
Average hours worked per week	2.25	2.40	0.15	2.20	-0.05
Household income as percent of federal poverty line	76.12	28.02	-48.09***	114.98	38.86***
Household Size (adults and children)	2.99	3.50	0.51	2.61	-0.39**
Number of family members under 19 living in house	0.88	1.00	0.12	0.56	-0.33***
Overall health	3.01	2.40	-0.61	2.84	-0.17
Health change	-0.09	-0.10	-0.01	-0.08	0.01
# days physical health not good	6.97	7.60	0.63	9.49	2.52**
# days mental health not good	8.43	13.90	5.47	10.30	1.87
# days poor health impaired regular activities	6.09	7.90	1.81	8.36	2.27**
Diabetes	0.10	0.20	0.10	0.20	0.10**
Asthma	0.13	0.20	0.07	0.13	0.01
High blood pressure	0.24	0.20	-0.04	0.38	0.15***
Depression	0.37	0.70	0.33*	0.47	0.10**
Any primary care visits	0.55	0.20	-0.35**	0.57	0.02
# primary care visits	1.64	1.80	0.16	1.86	0.22
Any hospital visits	0.04	0.10	0.06	0.12	0.07**
# hospital visits	0.05	0.10	0.05	0.16	0.11**
Usual place for medical care:					
Private clinic	0.18	0.00	-0.18***	0.28	0.10**
Public clinic	0.17	0.10	-0.07	0.18	0.01
Hospital-based clinic	0.07	0.00	-0.07***	0.08	0.01
Hospital ER	0.03	0.50	0.47**	0.06	0.03
Urgent care clinic	0.03	0.10	0.07	0.02	-0.01
Other places	0.06	0.00	-0.06***	0.07	0.00
Don't have usual place	0.46	0.30	-0.16	0.32	-0.14***

1) This table shows the mean characteristics for individuals whose treatment effects are not statistically significant at 10% level (column 1), whose treatment effects are significantly negative (column 2), and whose treatment effects are significantly positive (column 4). Column (3) contains the differences between column (2) and column (1), and column (5) contains the differences between column (4) and column (1).

2) Significance levels of the two-sample t-test: * 10%, ** 5%, *** 1%.

Table 3.8: Comparison of Characteristics

	Same	Fewer	Difference	More	Difference
	(1)	(2)	(3)	(4)	(5)
Needed medical care	0.68	1.00	0.32***	0.76	0.08*
Got all needed medical care	0.58	0.40	-0.18	0.56	-0.02
Reason went without care:					
Cost too much	0.33	0.20	-0.13	0.31	-0.02
No insurance	0.36	0.50	0.14	0.35	-0.01
Doc wouldn't take insurance	0.01	0.00	-0.01***	0.00	-0.01***
Owed money to provider	0.05	0.00	-0.05***	0.10	0.05*
Couldn't get an appointment	0.03	0.00	-0.03***	0.01	-0.02*
Office wasn't open	0.01	0.00	-0.01***	0.00	-0.01***
Didn't have a doctor	0.11	0.20	0.09	0.10	-0.01
Other reasons	0.03	0.00	-0.03***	0.04	0.01
Don't know	0.00	0.00	0.00***	0.00	0.00***
Needed prescription medications	0.61	0.80	0.19	0.77	0.16***
Got all needed prescriptions	0.74	0.70	-0.04	0.64	-0.09*
Currently taking any prescription medications	0.44	0.30	-0.14	0.68	0.25***
# prescription medications taking	1.37	1.30	-0.07	2.56	1.18***
Reason went without prescription medication:					
Cost too much	0.21	0.20	-0.01	0.27	0.06
No insurance	0.20	0.10	-0.10	0.20	0.00
Didn't have doctor	0.08	0.10	0.02	0.10	0.01
Couldn't get prescription	0.08	0.10	0.02	0.07	-0.01
Couldn't get to pharmacy	0.01	0.10	0.09	0.00	-0.01***
Other reasons	0.02	0.00	-0.02***	0.04	0.02
Don't know	0.00	0.00	0.00	0.00	0.00
Needed dental care	0.70	1.00	0.30***	0.70	0.00
Got all needed dental care	0.41	0.30	-0.11	0.37	-0.05
Any ER visits	0.14	0.90	0.76***	0.26	0.12***
# of ER visits	0.25	2.40	2.15***	0.45	0.20**
Used emergency room for non-emergency care	0.02	0.10	0.08	0.06	0.03
Reason went to ER:					
Needed emergency care	0.05	0.80	0.75***	0.06	0.01
Clinics closed	0.01	0.30	0.29*	0.02	0.00
Couldn't get doctor's appointment	0.02	0.20	0.18	0.02	0.00
Didn't have personal doctor	0.02	0.30	0.28	0.03	0.01
Couldn't afford copay to see a doctor	0.01	0.20	0.19	0.02	0.00
Didn't know where else to go	0.02	0.20	0.18	0.04	0.02
Other reason	0.01	0.10	0.09	0.02	0.01
Needed prescription drug	0.01	0.10	0.09	0.00	-0.01***
Don't know	0.00	0.00	0.00	0.00	0.00
Any out of pocket costs for medical care	0.65	0.80	0.15	0.71	0.06
Total out of pocket costs for medical care	5195.07	1257.00	-3938.07	1136.44	-4058.62
Borrowed money/skipped bills to pay health care bills	0.34	0.50	0.16	0.47	0.13**
Currently owe money for medical expenses	0.46	0.80	0.34**	0.71	0.25***
Total amount currently owed for medical expenses	1559.40	7354.00	5794.60**	5694.17	4134.77***

1) This table shows the mean characteristics for individuals whose treatment effects are not statistically significant at 10% level (column 1), whose treatment effects are significantly negative (column 2), and whose treatment effects are significantly positive (column 4). Column (3) contains the differences between column (2) and column (1), and column (5) contains the differences between column (4) and column (1).

2) Significance levels of the two-sample t-test: * 10%, ** 5%, *** 1%.

Table 3.9: Comparison of Characteristics

	Same	Fewer	Difference	More	Difference
	(1)	(2)	(3)	(4)	(5)
Any ED visits	0.16	0.90	0.74***	0.84	0.68***
# ED visits	0.28	3.10	2.82***	1.94	1.67***
Any ED visits resulting in hospitalization	0.03	0.00	-0.03***	0.37	0.34***
# ED visits resulting in hospitalization	0.03	0.00	-0.03***	0.62	0.59***
Any outpatient ED visits	0.15	0.90	0.75***	0.67	0.52***
# outpatient ED visits	0.25	3.10	2.85***	1.32	1.07***
Any weekday daytime ED visits	0.10	0.60	0.50**	0.62	0.52***
# weekday daytime ED visits	0.15	1.50	1.35**	1.14	0.99***
Any off-time ED visits	0.09	0.90	0.81***	0.51	0.42***
# off-time ED visits	0.12	1.60	1.48***	0.83	0.71***
# emergent non-preventable ED visits	0.05	0.17	0.12	0.66	0.61***
# emergent preventable ED visits	0.02	0.14	0.12**	0.15	0.12***
# primary care treatable ED visits	0.10	1.37	1.27***	0.46	0.36***
# non-emergent ED visits	0.05	1.22	1.16***	0.35	0.29***
# unclassified ED visits	0.05	0.20	0.15	0.36	0.30***
Any ambulatory case sensitive ED visits	0.01	0.00	-0.01***	0.15	0.14***
# ambulatory case sensitive ED visits	0.02	0.00	-0.02***	0.15	0.13***
Any ED visits to a high uninsured volume hospital	0.08	0.20	0.12	0.76	0.68***
# ED visits to a high uninsured volume hospital	0.12	0.30	0.18	1.61	1.49***
Any ED visits to a low uninsured volume hospital	0.09	0.90	0.81***	0.19	0.10**
# ED visits to a low uninsured volume hospital	0.15	2.80	2.65***	0.33	0.17**
Any ED visits for chronic conditions	0.03	0.30	0.27	0.36	0.32***
# ED visits for chronic conditions	0.05	0.50	0.45	0.62	0.57***
Any ED visits for injury	0.06	0.40	0.34*	0.32	0.26***
# ED visits for injury	0.07	0.40	0.33*	0.43	0.37***
Any ED visits for skin conditions	0.01	0.10	0.09	0.02	0.01
# ED visits for skin conditions	0.02	0.10	0.08	0.03	0.01
Any ED visits for abdominal pain	0.01	0.10	0.09	0.06	0.05**
# ED visits for abdominal pain	0.01	0.10	0.09	0.10	0.09**
Any ED visits for back pain	0.01	0.30	0.29*	0.04	0.03
# ED visits for back pain	0.01	0.40	0.39	0.06	0.04
Any ED visits for chest pain	0.01	0.00	-0.01***	0.05	0.04*
# ED visits for chest pain	0.01	0.00	-0.01***	0.05	0.04*
Any ED visits for headache	0.01	0.00	-0.01***	0.01	0.00
# ED visits for headache	0.01	0.00	-0.01***	0.01	0.00
Any ED visits for mood disorders	0.00	0.00	0.00**	0.09	0.08***
# ED visits for mood disorders	0.00	0.00	0.00**	0.15	0.15**
Any ED visits for psych conditions/substance abuse	0.01	0.00	-0.01***	0.17	0.17***
# ED visits for psych conditions/substance abuse	0.01	0.00	-0.01***	0.36	0.34***
Total ED charges	274.98	1818.44	1543.46**	3195.22	2920.24***
Total charges	639.70	2223.85	1584.16**	9260.22	8620.52***

1) This table shows the mean characteristics for individuals whose treatment effects are not statistically significant at 10% level (column 1), whose treatment effects are significantly negative (column 2), and whose treatment effects are significantly positive (column 4). Column (3) contains the differences between column (2) and column (1), and column (5) contains the differences between column (4) and column (1).

2) Significance levels of the two-sample t-test: * 10%, ** 5%, *** 1%.

3.5 Conclusion

In this paper, we propose a method for estimating the individual treatment effects using panel data, where multiple related outcomes are observed for a large number of individuals over a small number of pretreatment periods. The method is based on the interactive fixed effects model, and allows both the treatment assignment and the potential outcomes to be correlated with the unobserved individual characteristics. Monte Carlo simulations show that our method outperforms related methods. We also provide an example of estimating the effect of health insurance coverage on individual usage of hospital emergency departments using the Oregon Health Insurance Experiment data.

There are several directions for future research. First, our method requires the idiosyncratic shocks in the pretreatment outcomes to be uncorrelated either over time or across outcomes. It would be a valuable addition to allow (or detect and adjust for) more general dependence structure in the idiosyncratic shocks. Second, since the residuals of the rearranged models are not estimates of the idiosyncratic shocks, the variance of our estimator may be over-estimated, especially when the number of pretreatment outcomes is small. A necessary step for future research is to correct for this bias. Third, the repeated pretreatment set splitting and averaging approach in our method is computationally expensive. It would be an interesting direction for future research to find better ways to select related outcomes or use more flexible averaging scheme. Fourth, the linear model specification may be restrictive. There is potential to extend our method, perhaps in combination with more flexible machine learning methods, to work with more general nonlinear outcomes.

Appendix C

C.1 Proofs

Proof of Proposition 3.1.

$$\begin{aligned}\widehat{\tau}_{it,k} - \tau_{it,k} =& \widehat{Y}^1_{it,k} - \widehat{Y}^0_{it,k} - \left(Y^1_{it,k} - Y^0_{it,k}\right) \\ =& \boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^1_{t,k} - \boldsymbol{\theta}^1_{t,k}\right) - \boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^0_{t,k} - \boldsymbol{\theta}^0_{t,k}\right) - e^1_{it,k} + e^0_{it,k}.\end{aligned}$$

Given the assumptions and our models in (3.2) and (3.3), we have that $Y_{it,k}$, $t \leq T_0$, $k \in \mathcal{K}$ are i.i.d. for all $i \in \mathcal{T}$ and all $i \in \mathcal{C}$, and $\mathbb{E}|Y_{it,k}|^2 < \infty$, so that $\boldsymbol{Z}_{it}$ and $\boldsymbol{R}_{it}$ are i.i.d. for all $i \in \mathcal{T}$ and all $i \in \mathcal{C}$, $\mathbb{E}\|\boldsymbol{Z}_{it}\|^2 < \infty$, and $\mathbb{E}\|\boldsymbol{R}_{it}\|^2 < \infty$. In addition, $\mathbb{E}\left(\boldsymbol{Z}_{it}\boldsymbol{R}'_{it}\right)$ has full rank due to the observed covariates and the unobserved individual characteristics, so that $\mathbb{E}\left(\boldsymbol{Z}_{it}\boldsymbol{R}'_{it}\right)\boldsymbol{W}^1\mathbb{E}\left(\boldsymbol{R}_{it}\boldsymbol{Z}'_{it}\right)$ is invertible. Therefore, the weak law of large numbers and the continuous mapping theorem hold, and

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^1_{t,k} - \boldsymbol{\theta}^1_{t,k} &= \left(\boldsymbol{Z}^{1\prime}_t\boldsymbol{R}^1_t\boldsymbol{W}^1\boldsymbol{R}^{1\prime}_t\boldsymbol{Z}^1_t\right)^{-1}\boldsymbol{Z}^{1\prime}_t\boldsymbol{R}^1_t\boldsymbol{W}^1\boldsymbol{R}^{1\prime}_t\boldsymbol{e}^1_{t,k} \\ &= \left(\frac{1}{N_1}\boldsymbol{Z}^{1\prime}_t\boldsymbol{R}^1_t\boldsymbol{W}^1\frac{1}{N_1}\boldsymbol{R}^{1\prime}_t\boldsymbol{Z}^1_t\right)^{-1}\frac{1}{N_1}\boldsymbol{Z}^{1\prime}_t\boldsymbol{R}^1_t\boldsymbol{W}^1\frac{1}{N_1}\boldsymbol{R}^{1\prime}_t\boldsymbol{e}^1_{t,k} \\ &\xrightarrow{p} \left[\mathbb{E}\left(\boldsymbol{Z}_{it}\boldsymbol{R}'_{it}\right)\boldsymbol{W}^1\mathbb{E}\left(\boldsymbol{R}_{it}\boldsymbol{Z}'_{it}\right)\right]^{-1}\mathbb{E}\left(\boldsymbol{Z}_{it}\boldsymbol{R}'_{it}\right)\boldsymbol{W}^1\mathbb{E}\left(\boldsymbol{R}_{it}e^1_{it,k}\right) \\ &= 0,\end{aligned}$$

as $N_1 \to \infty$. Similarly, it can be shown that $\widehat{\boldsymbol{\theta}}^0_{t,k} - \boldsymbol{\theta}^0_{t,k} \xrightarrow{p} 0$ as $N_0 \to \infty$.

Since $\boldsymbol{Z}_{it} = O_p(1)$, we have $\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^1_{t,k} - \boldsymbol{\theta}^1_{t,k}\right) = o_p(1)$ and $\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^0_{t,k} - \boldsymbol{\theta}^0_{t,k}\right) = o_p(1)$. We also have $\mathbb{E}\left(e^1_{it,k} \mid \boldsymbol{H}_{it} = \boldsymbol{h}_{it}\right) = 0$ and $\mathbb{E}\left(e^0_{it,k} \mid \boldsymbol{H}_{it} = \boldsymbol{h}_{it}\right) = 0$ under Assumption 3.2.

Under the assumptions and by the Cauchy-Schwarz inequality, there exists $M^* \in [0, \infty)$ such that $\mathbb{E}\left|\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^1_{t,k} - \boldsymbol{\theta}^1_{t,k}\right)\right| \leq \left(\mathbb{E}\|\boldsymbol{Z}_{it}\|^2\right)^{1/2}\left(\mathbb{E}\left\|\widehat{\boldsymbol{\theta}}^1_{t,k} - \boldsymbol{\theta}^1_{t,k}\right\|^2\right)^{1/2} < M^*$,

$\mathbb{E}\left|\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^0_{t,k}-\boldsymbol{\theta}^0_{t,k}\right)\right| < M^*$, and $\mathbb{E}\left|e^0_{it,k}-e^1_{it,k}\right| < M^*$. By the triangle inequality, $\mathbb{E}\left|\widehat{\tau}_{it,k}-\tau_{it,k}\right| \leq \mathbb{E}\left|\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^1_{t,k}-\boldsymbol{\theta}^1_{t,k}\right)\right| + \mathbb{E}\left|\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^0_{t,k}-\boldsymbol{\theta}^0_{t,k}\right)\right| + \mathbb{E}\left|e^0_{it,k}-e^1_{it,k}\right| < 3M^*$, which implies that $\widehat{\tau}_{it,k}-\tau_{it,k}$ is uniformly integrable. Then by Lebesgue's Dominated Convergence Theorem, convergence in probability implies convergence in means, i.e.,

$$\begin{aligned}
&\lim_{N_1,N_0\to\infty}\mathbb{E}\left(\widehat{\tau}_{it,k}-\tau_{it,k}\mid \boldsymbol{H}_{it}=\boldsymbol{h}_{it}\right)\\
=&\mathbb{E}\left[\operatorname*{plim}_{N_1,N_0\to\infty}\left(\widehat{\tau}_{it,k}-\tau_{it,k}\right)\mid \boldsymbol{H}_{it}=\boldsymbol{h}_{it}\right]\\
=&\mathbb{E}\left(e^0_{it,k}-e^1_{it,k}\mid \boldsymbol{H}_{it}=\boldsymbol{h}_{it}\right)\\
=&0.
\end{aligned}$$

□

Proof of Proposition 3.2. Under Assumptions 3.1-3.3, central limit theorem applies, and we have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\tau}_{it,k}-\tau_{it,k}\right)\\
=&\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^1_{t,k}-\boldsymbol{\theta}^1_{t,k}\right)-\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^0_{t,k}-\boldsymbol{\theta}^0_{t,k}\right)-\frac{1}{N}\sum_{i=1}^{N}\left(e^1_{it,k}-e^0_{it,k}\right)\\
=&O_p\left(N_1^{-1/2}\right)+O_p\left(N_0^{-1/2}\right)+O_p\left((N_1+N_0)^{-1/2}\right)\\
=&O_p\left(N_1^{-1/2}\right)+O_p\left(N_0^{-1/2}\right).
\end{aligned}$$

Since $\frac{1}{N}\sum_{i=1}^{N}\tau_{it,k}-\mathbb{E}\left(\tau_{it,k}\right)=O_p\left((N_1+N_0)^{-1/2}\right)$. We have that $\frac{1}{N}\sum_{i=1}^{N}\widehat{\tau}_{it,k}-\tau_{t,k}=\frac{1}{N}\sum_{i=1}^{N}\widehat{\tau}_{it,k}-\frac{1}{N}\sum_{i=1}^{N}\tau_{it,k}+\frac{1}{N}\sum_{i=1}^{N}\tau_{it,k}-\tau_{t,k}=O_p\left(N_1^{-1/2}\right)+O_p\left(N_0^{-1/2}\right)$.

□

Proof of Proposition 3.3. (i)

$$\tilde{\tau}_{it,k}-\tau_{it,k}=\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^{*1}_{t,k}-\boldsymbol{\theta}^{*1}_{t,k}\right)-\boldsymbol{Z}'_{it}\left(\widehat{\boldsymbol{\theta}}^{*0}_{t,k}-\boldsymbol{\theta}^{*0}_{t,k}\right)-\left(u^1_{it,k}-u^0_{it,k}\right).$$

Since $\mathbb{E}\left(u^1_{it,k}\mid \boldsymbol{Z}_{it}\right)=0$ and under Assumption 3.2, we have $\mathbb{E}\left(\widehat{\boldsymbol{\theta}}^{*1}_{t,k}-\boldsymbol{\theta}^{*1}_{t,k}\mid \boldsymbol{Z}_{t}\right)=0$, and $\mathbb{E}\left(\widehat{\boldsymbol{\theta}}^{*1}_{t,k}-\boldsymbol{\theta}^{*1}_{t,k}\mid \boldsymbol{Z}_{it}\right)=\mathbb{E}_{-i}\left[\mathbb{E}\left(\widehat{\boldsymbol{\theta}}^{*1}_{t,k}-\boldsymbol{\theta}^{*1}_{t,k}\mid \boldsymbol{Z}_{t}\right)\right]=0$, where $\mathbb{E}_{-i}(\cdot)$ denotes the expectation taken with respect to $\boldsymbol{Z}_{jt}$, $j\neq i$. Similarly, we have $\mathbb{E}\left(\widehat{\boldsymbol{\theta}}^{*0}_{t,k}-\boldsymbol{\theta}^{*0}_{t,k}\mid \boldsymbol{Z}_{it}\right)=0$.

Thus, $\mathbb{E}\left(\tilde{\tau}_{it,k} - \tau_{it,k} \mid \boldsymbol{Z}_{it} = \boldsymbol{z}_{it}\right) = 0$. It follows that $\mathbb{E}\left(\tilde{\tau}_{t,k} - \tau_{t,k}\right) = 0$ using the law of iterated expectations.

(ii)

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left(\tilde{\tau}_{it,k} - \tau_{it,k}\right) \\
=&\frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{Z}_{it}'\left(\widehat{\boldsymbol{\theta}}_{t,k}^{*1} - \boldsymbol{\theta}_{t,k}^{1}\right) + \boldsymbol{\gamma}_{t,k}^{1\,\prime}\boldsymbol{\varepsilon}_{i}^{\mathcal{P}} - \varepsilon_{it,k}^{1}\right] - \frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{Z}_{it}'\left(\widehat{\boldsymbol{\theta}}_{t,k}^{*0} - \boldsymbol{\theta}_{t,k}^{0}\right) + \boldsymbol{\gamma}_{t,k}^{0\,\prime}\boldsymbol{\varepsilon}_{i}^{\mathcal{P}} - \varepsilon_{it,k}^{0}\right] \\
=&\frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{Z}_{it}'\left(\sum_{j\in\mathcal{T}}\boldsymbol{Z}_{jt}\boldsymbol{Z}_{jt}'\right)^{-1}\sum_{j\in\mathcal{T}}\boldsymbol{Z}_{jt}\left(\varepsilon_{jt,k}^{1} - \boldsymbol{\gamma}_{t,k}^{1\,\prime}\boldsymbol{\varepsilon}_{j}^{\mathcal{P}}\right) + \boldsymbol{\gamma}_{t,k}^{1\,\prime}\boldsymbol{\varepsilon}_{i}^{\mathcal{P}} - \varepsilon_{it,k}^{1}\right] \\
&- \frac{1}{N}\sum_{i=1}^{N}\left[\boldsymbol{Z}_{it}'\left(\sum_{j\in\mathcal{C}}\boldsymbol{Z}_{jt}\boldsymbol{Z}_{jt}'\right)^{-1}\sum_{j\in\mathcal{C}}\boldsymbol{Z}_{jt}\left(\varepsilon_{jt,k}^{0} - \boldsymbol{\gamma}_{t,k}^{0\,\prime}\boldsymbol{\varepsilon}_{j}^{\mathcal{P}}\right) + \boldsymbol{\gamma}_{t,k}^{0\,\prime}\boldsymbol{\varepsilon}_{i}^{\mathcal{P}} - \varepsilon_{it,k}^{0}\right].
\end{aligned}
$$

The following two statements hold:

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_{it}'\left(\sum_{j\in\mathcal{T}}\boldsymbol{Z}_{jt}\boldsymbol{Z}_{jt}'\right)^{-1}\sum_{j\in\mathcal{T}}\boldsymbol{Z}_{jt}\varepsilon_{jt,k}^{1} &= O_p\left(N_1^{-1/2}\right), \\
\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_{it}'\left(\sum_{j\in\mathcal{C}}\boldsymbol{Z}_{jt}\boldsymbol{Z}_{jt}'\right)^{-1}\sum_{j\in\mathcal{C}}\boldsymbol{Z}_{jt}\varepsilon_{jt,k}^{0} &= O_p\left(N_0^{-1/2}\right).
\end{aligned}
$$

Following similar arguments as in Li and Bell (2017), we denote $\tilde{\Delta}_i^1 = \boldsymbol{\gamma}_{t,k}^{1\,\prime}\boldsymbol{\varepsilon}_i^{\mathcal{P}} - \varepsilon_{it,k}^1 - \boldsymbol{Z}_{it}'\left(\sum_{j\in\mathcal{T}}\boldsymbol{Z}_{jt}\boldsymbol{Z}_{jt}'\right)^{-1}\sum_{j\in\mathcal{T}}\boldsymbol{Z}_{jt}\boldsymbol{\gamma}_{t,k}^{1\,\prime}\boldsymbol{\varepsilon}_j^{\mathcal{P}}$, and $\Delta_i^1 = \boldsymbol{\gamma}_{t,k}^{1\,\prime}\boldsymbol{\varepsilon}_i^{\mathcal{P}} - \varepsilon_{it,k}^1 - \boldsymbol{Z}_{it}'\mathbb{E}\left(\boldsymbol{Z}_{it}\boldsymbol{Z}_{it}'\right)^{-1}\mathbb{E}\left(\boldsymbol{Z}_{it}\boldsymbol{\gamma}_{t,k}^{1\,\prime}\boldsymbol{\varepsilon}_i^{\mathcal{P}}\right)$. We have that $\mathbb{E}\left(\boldsymbol{Z}_{it}\Delta_i^1\right) = 0$. Since $\boldsymbol{Z}_{it}$ contains constant 1, it follows that $\mathbb{E}\left(\Delta_i^1\right) = 0$. Thus $\frac{1}{N}\sum_{i=1}^{N}\tilde{\Delta}_i^1 \xrightarrow{p} \mathbb{E}\left(\Delta_i^1\right) = 0$ as $N_1 \to \infty$, and $\frac{1}{N}\sum_{i=1}^{N}\tilde{\Delta}_i^1 = O_p\left(N_1^{-1/2}\right)$. Similarly, we have $\frac{1}{N}\sum_{i=1}^{N}\tilde{\Delta}_i^0 = O_p\left(N_0^{-1/2}\right)$.

Thus, $\frac{1}{N}\sum_{i=1}^{N}\tilde{\tau}_{it,k} - \frac{1}{N}\sum_{i=1}^{N}\tau_{it,k} = O_p\left(N_1^{-1/2}\right) + O_p\left(N_0^{-1/2}\right)$, and $\frac{1}{N}\sum_{i=1}^{N}\tilde{\tau}_{it,k} - \tau_{t,k} = O_p\left(N_1^{-1/2}\right) + O_p\left(N_0^{-1/2}\right)$.

□

Conclusion

This thesis contains three studies on individual causal inference using panel data. The first two focus on the small N and large T setting, whereas the third focuses on the large N and small T setting.

Chapter 1 generalises the synthetic control method to the case where the outcome is a nonlinear function of the underlying predictors. Specifically, we provide conditions for the asymptotic unbiasedness of the synthetic control estimator to complement the theoretical result for the linear case in Abadie et al. (2010), and propose a flexible and data-driven method for choosing the synthetic control weights. Monte Carlo simulations show that the nonlinear synthetic control method has similar or better performance in the linear case and better performance in the nonlinear case compared with competing methods, and that the confidence intervals have good coverage probabilities across settings. In the empirical application, we illustrate the method by estimating the impact of the 2019 anti-extradition law amendments bill protests on Hong Kong's economy, and find that the year-long protests reduced the real GDP per capita by 11.27% in the first quarter of 2020, which is larger in magnitude than the economic decline in the 1997 Asian financial crisis and the 2008 global financial crisis.

Chapter 2 generalises the conventional single-outcome synthetic control method to a multiple-outcome framework, where the number of pretreatment periods is supplemented with the number of related outcomes in the domain, making the method applicable even when the number of pretreatment periods is small or if we worry about structural breaks over a longer time span. Following Abadie et al. (2010), we show that the bound on the bias of the multiple-outcome synthetic control estimator is of a smaller stochastic order than that of the single-outcome synthetic control estimator, when the synthetic control can closely approximate the unit of interest in terms of the observed predictors and the multiple related outcomes. We also discuss the role of demeaning the outcomes before constructing the synthetic control, which is to account for the differences in the level of the outcomes for different units, and show in simulation that using demeaned outcomes can reduce both

the bias and the variance of the synthetic control estimator and alleviate the size distortion of the permutation test, if there are relatively stable differences in the level of the outcomes.

We move on to evaluate the effects of the non-pharmaceutical interventions on various outcomes in the public health, labour market, and economic domains using the multiple-outcome synthetic control method, where we construct a synthetic Sweden in each domain using the other European countries that implemented much stricter NPIs. We find that the NPIs would significantly reduce the cumulative numbers of COVID-19 cases and deaths as well as deaths from all causes, increase temporary absence from work and reduce total hours worked among the employed, but would have limited impacts on the employment rate and the economy, other than shrinking the volume of retail sales in the early stage.

Chapter 3 proposes a method for estimating the individual treatment effects using panel data, where multiple related outcomes are observed for a large number of individuals over a small number of pretreatment periods. The method is based on the interactive fixed effects model, and allows both the treatment assignment and the potential outcomes to be correlated with the unobserved individual characteristics. Monte Carlo simulations show that our method outperforms related methods. We also provide an example of estimating the effect of health insurance coverage on individual usage of hospital emergency departments using the Oregon Health Insurance Experiment data.

Bibliography

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

Abadie, A. and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 10:465–503.

Abadie, A., Diamond, A., and Hainmueller, J. (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., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1):113–132.

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Abadie, A. and L'Hour, J. (2020). A penalized synthetic control estimator for disaggregated data. *Working Paper*.

ABC (2019). Hong Kong university siege ends as officials are left with devastating aftermath. `https://www.abc.net.au/news/2019-11-27/hong-kong-university-siege-ends-as-cross-harbour-tunnel-reopens/11743146`. Retrieved April 25, 2021.

Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. (2020). Optimal targeted lockdowns in a multi-group SIR model. *NBER Working Paper*, 27102.

Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and democracy. *American Economic Review*, 98(3):808–42.

Ahn, S. C., Lee, Y. H., and Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174(1):1–14.

Al Jazeera (2020). China passes Hong Kong security law, deepening fears for future. https://www.aljazeera.com/news/2020/6/30/china-passes-hong-kong-security-law-deepening-fears-for-future. Retrieved April 25, 2021.

Alexander, D. and Karger, E. (2020). Do stay-at-home orders cause people to stay at home? Effects of stay-at-home orders on consumer behavior. *FRB of Chicago Working Paper*, No. WP-2020-12.

Andrews, D. W. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, 67(3):543–563.

AP (2016). Hong Kong unsettled by case of 5 missing booksellers. https://apnews.com/article/acf943c14a5049e99aa1ab61bb9965dd. Retrieved April 25, 2021.

Arnon, A., Ricco, J., and Smetters, K. (2020). Epidemiological and economic effects of lockdown. *Brookings Papers on Economic Activity*.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, pages 1–15.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Aum, S., Lee, S. Y. T., and Shin, Y. (2020). Inequality of fear and self-quarantine: Is there a trade-off between gdp and public health? *Journal of Public Economics (forthcoming)*.

Baek, C., McCrory, P. B., Messer, T., and Mui, P. (2020). Unemployment effects of stay-at-home orders: Evidence from high frequency claims data. *Review of Economics and Statistics*, pages 1–72.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Baqaee, D., Farhi, E., Mina, M. J., and Stock, J. H. (2020). Reopening scenarios. Working paper, National Bureau of Economic Research.

Basic Law (1997). Basic Law of the Hong Kong Special Administrative Region of the People's Republic of China. https://www.basiclaw.gov.hk/en/basiclawtext/chapter_2.html. Retrieved April 25, 2021.

BBC (2019a). The murder behind the Hong Kong protests: A case where no-one

wants the killer. https://www.bbc.com/news/world-asia-china-50148577. Retrieved April 25, 2021.

BBC (2019b). PolyU: Hong Kong police find almost 4,000 petrol bombs on campus. https://www.bbc.com/news/world-asia-china-50598185. Retrieved April 25, 2021.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890.

Born, B., Dietrich, A., and Müller, G. J. (2020). Do lockdowns work? a counterfactual for Sweden. *Covid Economics*, 16:1–22.

Botosaru, I. and Ferman, B. (2019). On the role of covariates in the synthetic control method. *The Econometrics Journal*, 22(2):117–130.

Castex, G., Dechter, E., and Lorca, M. (2021). COVID-19: The impact of social distancing policies, cross-country analysis. *Economics of disasters and climate change*, 5(1):135–159.

Chen, X. and Qiu, Z. (2020). Scenario analysis of non-pharmaceutical interventions on global COVID-19 transmissions. *Covid Economics*, 7:46–67.

Chernozhukov, V., Kasahara, H., and Schrimpf, P. (2021). Causal impact of masks, policies, behavior on early COVID-19 pandemic in the U.S. *Journal of Econometrics*, 220(1):23 – 62. Themed Issue: Pandemic Econometrics / Covid Pandemics.

China News (2020). More than 3200 students arrested during the Hong Kong amendments bill disturbance (in Chinese). https://www.chinanews.com/ga/shipin/cns/2020/05-05/news855499.shtml. Retrieved April 25, 2021.

Cho, S.-W. S. (2020). Quantifying the impact of non-pharmaceutical interventions during the COVID-19 outbreak–the case of Sweden. *The Econometrics Journal*.

CNN (2019a). Hong Kong protesters storm government building over China extradition bill. https://edition.cnn.com/asia/live-news/hong-kong-july-1-protests-intl-hnk/h_cc86e6e2cbbfd32446e41356400ac566. Retrieved April 25, 2021.

CNN (2019b). Violence and chaos spreads through Hong Kong as protests enter 16th week. https://edition.cnn.com/2019/09/22/asia/hong-kong-protest-violence-intl-hnk/index.html. Retrieved April 25, 2021.

Coibion, O., Gorodnichenko, Y., and Weber, M. (2020). The cost of the COVID-19 crisis: Lockdowns, macroeconomic expectations, and consumer spending. Working paper, National Bureau of Economic Research.

Conyon, M. J., He, L., and Thomsen, S. (2020). Lockdowns and COVID-19 deaths in Scandinavia. *Covid Economics*, 26:46–67.

Correia, S., Luck, S., and Verner, E. (2020). Public health interventions do not: Evidence from the 1918 flu. Working paper.

Demirguc-Kunt, A., Lokshin, M., and Torre, I. (2020). The sooner, the better: The early economic impact of non-pharmaceutical interventions during the COVID-19 pandemic. *World Bank Policy Research Working Paper*, (9257).

Díaz, J., Rau, T., and Rivera, J. (2015). A matching estimator based on a bilevel optimization problem. *Review of Economics and Statistics*, 97(4):803–812.

Doudchenko, N. and Imbens, G. W. (2017). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research.

Fang, H., Wang, L., and Yang, Y. (2020). Human mobility restrictions and the spread of the Novel Coronavirus (2019-nCoV) in China. *Journal of Public Economics*, 191:104272.

Favero, C. A., Ichino, A., and Rustichini, A. (2020). Restarting the economy while saving lives under COVID-19. *CEPR Discussion Paper*, (No. DP14664).

Ferman, B. and Pinto, C. (2017). Placebo tests for synthetic controls. *Working Paper*.

Ferman, B. and Pinto, C. (2019). Synthetic controls with imperfect pre-treatment fit. Working paper. arXiv:1911.08521.

Fernández-Villaverde, J., Jones, C. I., et al. (2020). Macroeconomic outcomes and COVID-19: a progress report. *NBER Work Paper Series*, 28004.

Fezzi, C. and Fanghella, V. (2020). Tracking GDP in real-time using electricity market data: insights from the first wave of COVID-19 across Europe. Working paper.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Oregon Health Study Group (2012). The Oregon Health Insurance Experiment: evidence from the first year. *The Quarterly Journal of Economics*, 127(3):1057–1106.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261.

Forsythe, E., Kahn, L. B., Lange, F., and Wiczer, D. (2020). Labor demand in the time of COVID-19: Evidence from vacancy postings and UI claims. *Journal of Public Economics*, 189:104238.

Friedson, A. I., McNichols, D., Sabia, J. J., and Dave, D. (2020). Did California's shelter-in-place order work? Early coronavirus-related public health effects. Working paper, National Bureau of Economic Research.

Goolsbee, A. and Syverson, C. (2020). Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. *Journal of Public Economics*, 193:104311.

Gupta, S., Montenovo, L., Nguyen, T. D., Rojas, F. L., Schmutte, I. M., Simon, K. I., Weinberg, B. A., and Wing, C. (2020). Effects of social distancing policy on labor market outcomes. Working paper, National Bureau of Economic Research.

Hale, T., Angrist, N., Cameron-Blake, E., Hallas, L., Kira, B., Majumdar, S., Petherick, A., Phillips, T., Tatlow, H., and Webster, S. (2020). Oxford COVID-19 government response tracker.

Hansen, B. E. (2021). Econometrics. *Manuscript*. https://www.ssc.wisc.edu/ bhansen/econometrics/.

HKFP (2019a). 'absolutely intolerable': Protesters at Beijing's Hong Kong office hurt the feelings of all Chinese people, top official says. `https://hongkongfp.com/2019/07/22/absolutely-intolerable-protesters-beijings-hong-kong-office-hurt-feelings-chinese-people-top-official-says/`. Retrieved April 25, 2021.

HKFP (2019b). In pictures: 'flow like water' – Hong Kong protesters converge on police HQ after day of wildcat road occupations. `https://hongkongfp.com/2019/06/21/pictures-flow-like-water-hong-kong-protesters-converge-police-hq-day-wildcat-road-occupations/`. Retrieved April 25, 2021.

HKFP (2020). Explainer: From 'five demands' to 'independence' – the evolution of Hong Kong's protest slogans. `https://hongkongfp.com/2020/06/25/explainer-from-five-demands-to-black-cops-to-independence-the-evolution-of-hong-kongs-protest-slogans/`. Retrieved April 25, 2021.

HKFP (2021). Hong Kong's 2019 protest movement in pictures. `https://hongkongfp.com/hong-kongs-2019-protest-movement-in-pictures/`. Retrieved April 25, 2021.

HKSARG (2019). Sunset clause unacceptable. `https://www.news.gov.hk/eng/`

2019/04/20190420/20190420_125635_148.html?type=ticker. Retrieved April 25, 2021.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica: Journal of the econometric society*, pages 1371–1395.

Hong Kong Bar Association (2019). A brief guide to issues arising from the fugitive offenders and mutual legal assistance in criminal matters legislation (amendment) bill 2019 ("the bill"). https://www.hkba.org/sites/default/files/A%20Brief%20Guide%20to%20issues%20arising%20from%20the%20Fugitive%20Offenders%20And%20Mutual%20Legal%20Assistance%20in%20Criminal%20Matters%20Legislation%20%28Amendment%29%20Bill%202019%20%28%E2%80%9CThe%20Bill%E2%80%9D%29_0.pdf. Retrieved April 25, 2021.

Hsiang, S., Allen, D., Annan-Phan, S., Bell, K., Bolliger, I., Chong, T., Druckenmiller, H., Huang, L. Y., Hultgren, A., Krasovich, E., et al. (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, 584(7820):262–267.

Hsiao, C., Steve Ching, H., and Ki Wan, S. (2012). A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics*, 27(5):705–740.

HuffPost (2019). Hong Kong protesters take suspected undercover police officer hostage during violent stand-off inside airport. https://www.huffingtonpost.co.uk/entry/hong-kong-airport-protests-videos_uk_5d52e26fe4b0c63bcbee09d5. Retrieved April 25, 2021.

Juranek, S., Paetzold, J., Winner, H., and Zoutman, F. (2020). Labor market effects of COVID-19 in Sweden and its neighbors: Evidence from novel administrative data. *NHH Dept. of Business and Management Science Discussion Paper*, (2020/8).

King, G. and Zeng, L. (2006). The dangers of extreme counterfactuals. *Political Analysis*, 14(2):131–159.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.

Kong, E. and Prinz, D. (2020). The impact of shutdown policies on unemployment during a pandemic. *Covid Economics*, 17:24–72.

Legislative Council (2019). Fugitive offenders and mutual legal assistance in criminal

matters legislation (amendment) bill 2019. https://www.legco.gov.hk/yr18-19/english/bills/b201903291.pdf. Retrieved April 25, 2021.

Li, K. T. and Bell, D. R. (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics*, 197(1):65–75.

Ludvigsson, J. F. (2020). The first eight months of Sweden's COVID-19 strategy and the key actions and actors that were involved. *Acta Paediatrica*, 109(12):2459–2471.

Ming Pao (2019). Live news, photos and videos of the protest on june 12, 2019 (in Chinese). https://news.mingpao.com/ins/%e6%b8%af%e8%81%9e/article/20190612/s00001/1560296192111/%e3%80%90%e9%80%83%e7%8a%af%e6%a2%9d%e4%be%8b-%e7%9f%ad%e7%89%87-%e5%a4%9a%e5%9c%96%e3%80%91%e7%89%b9%e9%a6%96%e9%81%b8%e5%a7%94%e8%81%af%e7%bd%b2%e4%bf%83%e6%9e%97%e9%84%ad%e4%b8%8b%e5%8f%b0-%e6%92%a4%e5%9b%9e%e4%bf%ae%e4%be%8b. Retrieved April 25, 2021.

Mitze, T., Kosfeld, R., Rode, J., and Wälde, K. (2020). Face masks considerably reduce COVID-19 cases in Germany: A synthetic control method approach.

Murray, S. and Olivares, E. (2020). Job losses during the onset of the COVID-19 pandemic: Stay-at-home orders, industry composition, and administrative capacity. Working paper.

OECD (2020). Country policy tracker: Employment and social measures.

Reuters (2019). How murder, kidnappings and miscalculation set off Hong Kong's revolt. https://www.reuters.com/investigates/special-report/hongkong-protests-extradition-narrative/. Retrieved April 25, 2021.

Rojas, F. L., Jiang, X., Montenovo, L., Simon, K. I., Weinberg, B. A., and Wing, C. (2020). Is the cure worse than the problem itself? Immediate labor market effects of COVID-19 case rates and school closures in the US. Working paper, National Bureau of Economic Research.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Roser, M., Ritchie, H., Ortiz-Ospina, E., and Hasell, J. (2020). Coronavirus pandemic (COVID-19). *Our World in Data*. https://ourworldindata.org/coronavirus.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688.

SCMP (2019). Violent clashes mar protest after 'more than a million' hongkongers of all ages and backgrounds march against controversial extradition bill. https://www.scmp.com/news/hong-kong/politics/article/3013757/violent-clashes-mar-protest-after-more-million-hongkongers. Retrieved April 25, 2021.

Sheridan, A., Andersen, A. L., Hansen, E. T., and Johannesen, N. (2020). Social distancing laws cause only small losses of economic activity during the COVID-19 pandemic in Scandinavia. *Proceedings of the National Academy of Sciences*, 117(34):20468–20473.

Taubman, S. L., Allen, H. L., Wright, B. J., Baicker, K., and Finkelstein, A. N. (2014). Medicaid increases emergency-department use: evidence from Oregon's Health Insurance Experiment. *Science*, 343(6168):263–268.

The Standard (2019). 24 pan-democrats propose fugitive law change to suit Taiwan. https://www.thestandard.com.hk/breaking-news/section/4/126337/24-pan-democrats-propose-fugitive-law-change-to-suit-Taiwan. Retrieved April 25, 2021.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.

Yuen, S. (2019). New research shows vast majority of Hong Kong protesters support more radical tactics. https://theconversation.com/new-research-shows-vast-majority-of-hong-kong-protesters-support-more-radical-tactics-122531. Retrieved April 25, 2021.