



THE UNIVERSITY OF
WESTERN
AUSTRALIA

Western Australian Economic History Summit 2020

Synthetic Growth Analysis and Western Australian Economic History



Rok Spruk

February 6th, 2020

University of Western Australia



CENTRE FOR
Western Australian History

Agenda

- Introductory remarks
- Synthetic control method
- Applications of synthetic control method in economics and economic history
- Synthetic control method and Western Australian economic history
- Concluding thoughts

Introduction

Arguably the most important innovation in the evaluation literature in the last fifteen years is the synthetic control approach developed by Abadie et al. [2010, 2014b] and Abadie and Gardeazabal [2003]. This method builds on difference-in-differences estimation, but uses arguably more attractive comparisons to identify causal effects.

Susan Athey and Guido Imbens, “The State of Applied Econometrics – Causality and Policy Evaluation“, *Journal of Economic Perspectives*, Vol. 31, No. 2., p.9

Introduction

- Synthetic control method has been designed to study the effect of a particular intervention (i.e. treatment) or event on the outcome of interest
- The method builds on a difference-in-difference approach and exploits pre-intervention trends in outcomes and covariates to simulate the counterfactual scenario in the hypothetical absence of the intervention itself
- The idea of the synthetic control method is to compute a gap in the given outcome of interest assuming that a certain event never happened
- By exploiting pre-intervention trends between treatment and control group, one can obtain a meaningful and plausible representation of the treated set based on the outcome and covariate comparison with the control group

Introduction

- In the original paper, [Abadie and Gardeazabal \(2003\)](#) study the effects of terrorism on the economic growth of Basque Country in Spain
- Their approach is to simulate the growth trajectory of the Basque Country in the absence of ETA terrorism
- To obtain the counterfactual growth trajectory, they construct a synthetic Basque Country as a convex combination of per capita income and its covariates' characteristics of other Spanish autonomous communities that never experienced such terrorism
- The counterfactual trajectory is constructed from the weighted average of outcomes and covariates of the control group that minimize pre-terrorism discrepancy between the Basque country and the rest of Spain
- The synthetic Basque Country (i.e. artificial control group) consists of Madrid (0.85) and Catalunya (0.15)

Introduction

The Economic Costs of Conflict: A Case Study of the Basque Country

By ALBERTO ABADIE AND JAVIER GARDEAZABAL*

This article investigates the economic effects of conflict, using the terrorist conflict in the Basque Country as a case study. We find that, after the outbreak of terrorism in the late 1960's, per capita GDP in the Basque Country declined about 10 percentage points relative to a synthetic control region without terrorism. In addition, we use the 1998–1999 truce as a natural experiment. We find that stocks of firms with a significant part of their business in the Basque Country showed a positive relative performance when truce became credible, and a negative relative performance at the end of the cease-fire. (JEL D74, G14, P16)

Political instability is believed to have strong adverse effects on economic prosperity. However, to date, the evidence on this matter is scarce, probably because it is difficult to know how economies would have evolved in absence of political conflicts.

This article investigates the economic impact of conflict, using the terrorist conflict in the Basque Country as a case study. The Basque conflict is especially interesting from an eco-

of terrorist and political conflict, the Basque Country had dropped to the sixth position in per capita GDP.¹ During that period, terrorist activity by the Basque terrorist organization ETA resulted in almost 800 deaths. Basque entrepreneurs and corporations had been specific targets of violence and extortion (including assassinations, robberies, and kidnappings-for-ransom). Not surprisingly, the economic downturn suffered by the Basque economy during those

- Is the effect of ETA terrorism specific to Basque Country?
- A simple placebo analysis where ETA terrorist activity shock is assigned to a region that did not experience it (i.e. Catalunya)
- The counterfactual growth trajectory of Catalunya before and after the reassigned ETA terrorist activity is the same
- Effect appears to be unique to Basque Country



Introduction

- The effect of terrorist activity appears to be specific to Basque Country

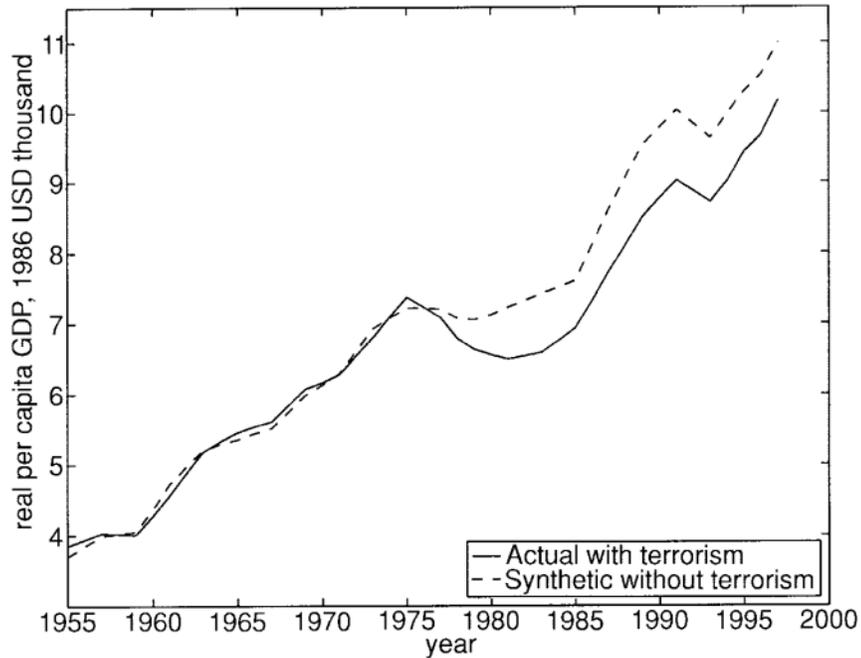


FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

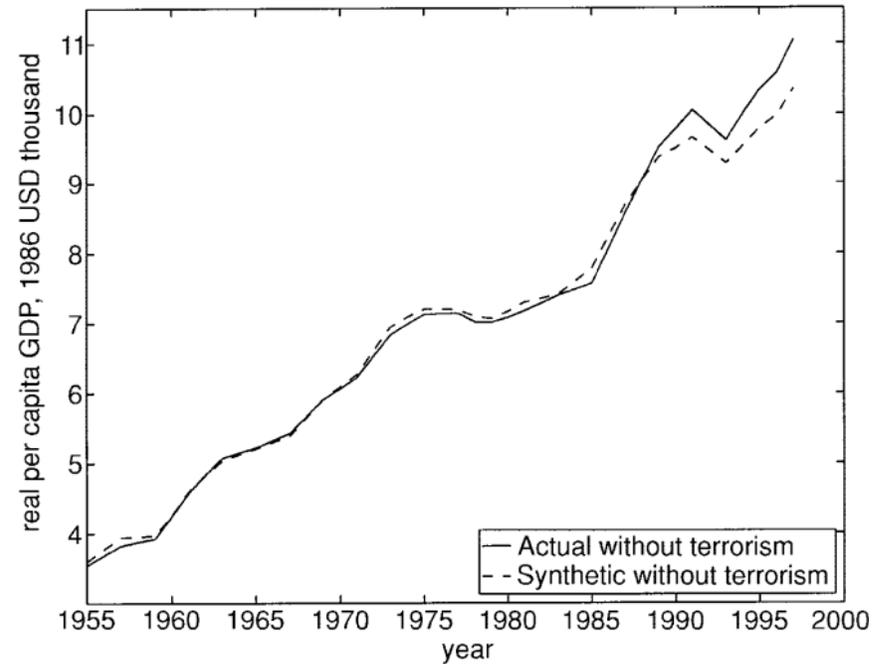
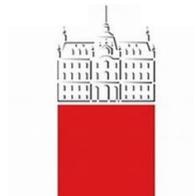


FIGURE 4. A "PLACEBO STUDY," PER CAPITA GDP FOR CATALONIA

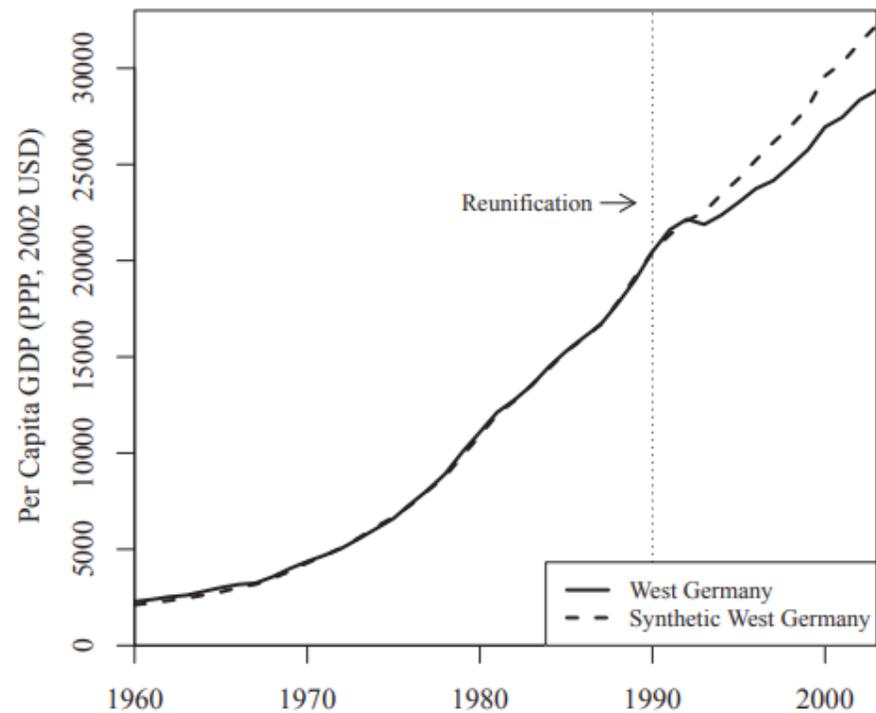


Introduction

- The method has been used in a wide range of applications such as impact evaluation of terrorism, civil wars, political risk (Abadie and Gardeazabal 2003, Bove et. al. 2014, Montalvo 2011, Yu and Wang 2013), natural disasters (Barone and Mocetti 2014, Cavallo et. al. 2013, Smith 2015), economic and trade liberalization (Billmeier and Nannicini 2013, Gathani et. al. 2013, Hosny 2012), health policy (Abadie et. al. 2010, Bauhoff 2014, Kreif et. al. 2015, Spruk and Kovac 2020), political reforms and regime changes (Billmeier and Nannicini 2009, Garcia Ribeiro et. al. 2013, Carrasco et. al. 2013, Abadie et. al. 2015, Spruk 2019, Campos et. al. 2019), social and political connections (Acemoglu et. al. 2013), labor (Bohn et. al. 2014, Calderon 2014, Peri and Yasenov 2019), crime (Pinotti 2012, Saunders et. al. 2014), taxation (Kleven et. al. 2013, de Souza 2014), and local development policies (Kirkpatrick and Benneer 2014, Gobillon and Magnac 2016, Possebom 2017, Ando 2015) to name a few
- The method is designed for comparative case studies in small and moderate sample sizes as a transparent and data-driven way of choosing comparison units
- Data-driven process to select weights that build the weighted outcome and covariate average of control group best reproduce the outcome trajectory of the treated unit before treatment

Introduction

FIGURE 2 Trends in per Capita GDP: West Germany versus Synthetic West Germany



- Economic cost of German unification (i.e. Wiedervereinigung) in 1990 studied by [Abadie, Diamond and Hainmüller \(2015, AJPS\)](#)
- The synthetic Germany without the unification prior to 1990 is best reproduced as a convex combination of Austria (42%), United States (22%), Japan (16%), Switzerland (11%), and Netherlands (9%)
- Training period and validation period
- Using the predictors in pre-unification period, weights are selected in such a way that minimizes root mean square prediction error (RMSE) over the validation period
- Such cross-validation technique yields the weights that minimize out-of-sample prediction error
- In-time placebo analysis: shift of the unification to a deliberately false date

Introduction

FIGURE 3 Per Capita GDP Gap between West Germany and Synthetic West Germany

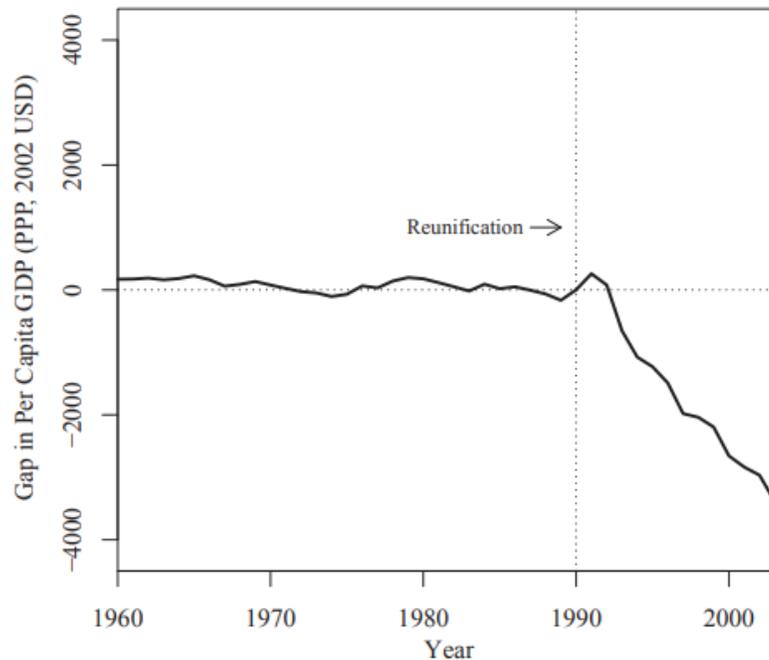
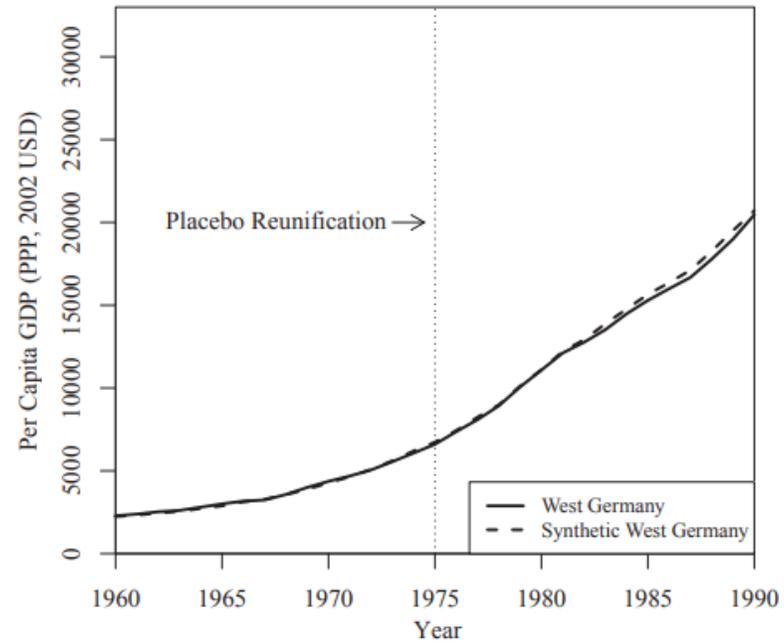


FIGURE 4 Placebo Reunification 1975–Trends in per Capita GDP: West Germany versus Synthetic West Germany



- The counterfactual growth trajectory unaffected by the placebo intervention in the middle of the pre-unification period

Introduction

- Economic growth and development effects of Hugo Chavez on the Venezuelan economy (Grier and Maynard 2016, JEBO)
- A controversial figure: won 3 elections, rewrote Venezuelan constitution, restructured its Supreme Court, survived coup, assassination and recall attempts
- A massive drop in poverty, inequality, a rapid expansion of access to health and education for the poor
- But... a rampant deterioration of institutional quality, rise of crony capitalism with numerous nationalizations of industries, food and price controls
- A hero or the precursor of Venezuela's current disaster?



Journal of Economic Behavior & Organization 125 (2016) 1–21

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

The economic consequences of Hugo Chavez: A synthetic control analysis

Kevin Grier^{a,*}, Norman Maynard^b

^a Department of Economics, University of Oklahoma, United States
^b Department of Economics, College of Charleston, United States

ARTICLE INFO

Article history:
 Received 20 January 2015
 Received in revised form 9 November 2015
 Accepted 30 December 2015
 Available online 14 January 2016

Keywords:
 Economic growth
 Inequality
 Population
 Venezuela
 Synthetic control
 Leaders

ABSTRACT

We use the synthetic control method to perform a case study of the impact of Hugo Chavez on the Venezuelan economy. We compare outcomes under Chavez's leadership and policies against a counterfactual of "business as usual" in similar countries. We find that, relative to our control, per capita income fell dramatically. While poverty, health, and inequality outcomes all improved during the Chavez administration, these outcomes also improved in each of the corresponding control cases and thus we cannot attribute the improvements to Chavismo. We conclude that the overall economic consequences of the Chavez administration were bleak.

© 2016 Elsevier B.V. All rights reserved.



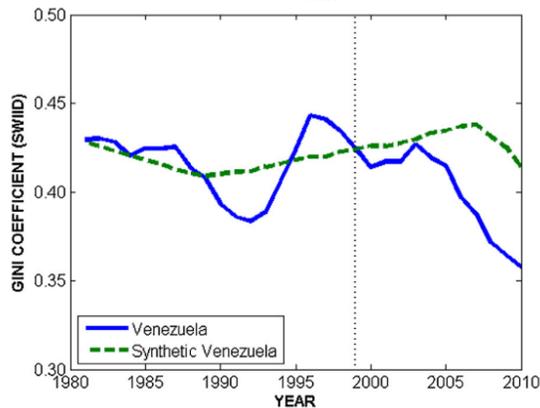
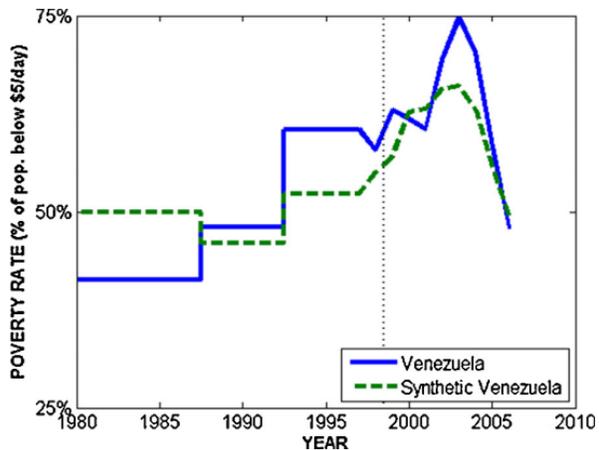
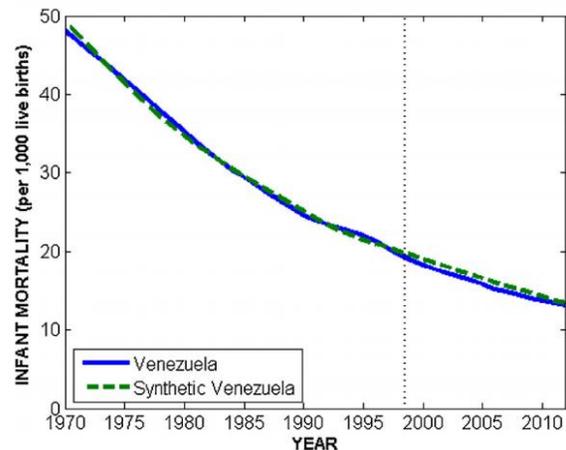
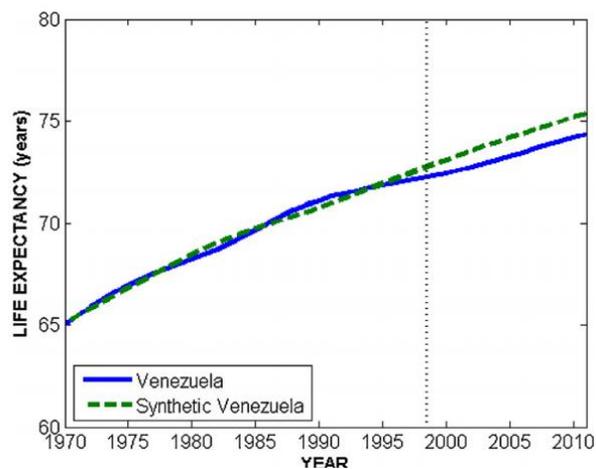
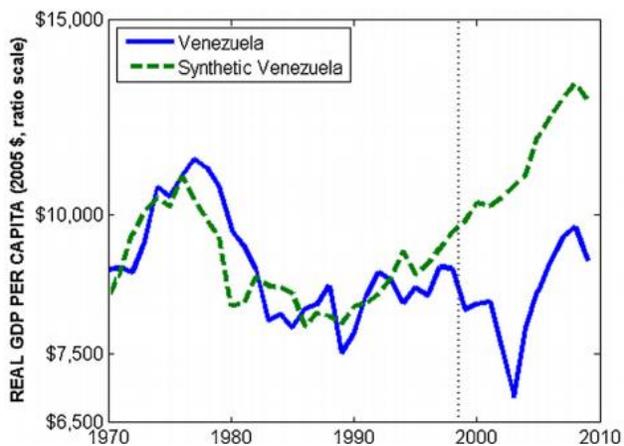
Introduction

“You know that I have ruled for twenty years, and from these I have derived only a few certainties: (1) [South] America is ungovernable, for us; (2) Those who serve a revolution plough the sea; (3) The only thing you can do in [South] America is emigrate; (4) This country will fall inevitably in the hands of the unbridled masses and then pass almost imperceptibly into the hands of petty tyrants, of all colours and races; (5) Once we have been devoured by every crime and extinguished by utter ferocity, the Europeans will not even regard us as worth conquering; (6) If it were possible for any part of the world to revert to primitive chaos, it would be America her final hour.”

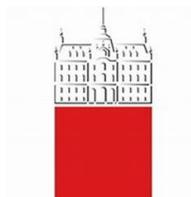
Simon Bolivar, Letter to General Juan José Flores, Baranquilla, November 25, 1830



Introduction



- No perceivable disproportionate improvement in poverty, health and inequality outcomes during Chavez administration since these outcomes also improved in the artificial control group for Venezuela



Introduction

Table 2
Estimated synthetic control weights for each outcome variable.

	Outcome variables				
	Income	Infant Mort	Life Exp	Poverty	Inequality
Algeria	0.00	0.00	0.00	-	-
Argentina	0.00	0.00	0.99	20.38	0.40
Brazil	7.25	0.00	0.00	0.00	0.00
Canada	20.35	0.00	0.00	-	0.00
Chile	0.00	0.00	0.00	0.00	0.00
Colombia	0.00	0.00	0.00	79.62	6.85
Costa Rica	0.00	0.00	0.00	0.00	0.00
El Salvador	0.00	10.00	6.27	0.00	0.00
Guatemala	0.00	2.25	0.00	0.00	-
Honduras	0.00	0.00	0.00	0.00	-
Indonesia	0.00	-	-	-	0.00
Iran	42.24	0.00	0.00	-	-
Iraq	0.00	-	-	-	-
Mexico	12.67	0.00	0.00	0.00	0.00
Nigeria	-	3.10	8.46	-	0.01
Norway	0.00	33.98	18.86	-	5.19
Panama	0.00	41.69	44.53	0.00	0.00
Paraguay	0.00	-	-	0.00	0.00
Peru	17.49	0.00	0.00	0.00	0.00
Uruguay	0.00	8.97	20.89	0.00	87.55
<i>Model fit pre-intervention</i>					
RMSPE	0.068	0.32	0.16	6.23	1.48
APE-to-mean ratio	0.35%	0.28%	0.002%	2.16%	0.17%
<i>SCM inference: permutation test</i>					
RMSPE ratio	4.27	1.36	4.76	0.80	2.44
p-value: RMSPE	0.00	0.35	0.12	0.79	0.21

Note: Columns show the weight assigned to each country in the synthetic controls for Venezuela. Each column includes a synthetic control for a different outcome variable. A dash (-) indicates that the country is not available in the dataset for the given comparison. Weights are in percentage points. Rounding errors may prevent columns from summing to 100. APE-to-mean ratio indicates the average pre-intervention prediction error divided by the average pre-intervention outcome value.

Table 3
Indicator fits & V-matrix, GDP per capita.

Variables	Actual Ven	Synth. Ven	V-matrix	LA average	OPEC average
Avg. GDP per capita	9,071	9040	-	5573	4376
Pop. growth rate	2.67	2.35	1817.27	2.10	2.66
Openness 1970	46.64	50.51	11,026.64	40.18	76.57
Openness 1990	46.65	48.50	9224.12	43.87	93.86
Total ed. 15+, 1970	3.74	4.16	1019.36	3.98	1.93
Primary ed. 15+, 1970	2.91	2.80	25,821.79	3.11	1.47
Total ed. 15+, 1990	5.11	6.57	0.40	6.17	4.50
Primary ed. 15+, 1990	3.76	3.92	11,145.97	4.29	2.80

Note: Table shows the values of indicator variables and the average pre-Chavez outcome variable for actual and synthetic Venezuela. Average GDP is an average over the pre-Chavez period, 1970-1998. GDP per capita is measured in 2005 dollars. Population growth is in percentage points. Government share is a fraction of GDP, in percentage points. Polity and Ex. Const. follow the polity scores scale. Durability and the education variables are measured in years. The V-matrix column includes the diagonal entries in the V matrix. They determine how much weight each indicator variable receives in the minimization problem for the weights, $W = \arg \min(X_1 - X_0W)'V(X_1 - X_0W)$. The values are normalized relative to the entry for the first variable in V (1970 GDP per capita), which equals one. Although they resemble regression coefficients, they are formulated differently and cannot be interpreted as either partial correlations or marginal effects.

Table 4
Indicator fits & V-matrices, health, poverty, & inequality.

Variables	Actual Ven		Synth. Ven		Life Expectancy		Poverty		Inequality	
	Actual Ven	Synth. Ven	Actual Ven	Synth. Ven	Actual Ven	Synth. Ven	Actual Ven	Synth. Ven	Actual Ven	Synth. Ven
avg. Poverty	51.93	-	-	-	-	-	50.8	-	-	-
avg. Mortality	31.90	31.82	-	-	-	-	-	-	-	-
avg. Life Exp.	69.22	-	69.22	-	-	-	-	-	-	-
avg. Inequality	41.68	-	-	-	-	-	-	-	41.75	-
Pop. Growth Rate	2.67	-	-	-	-	-	-	-	0.96	4420.71
log GDP per capita 1970	8.10	8.72	30,975.58	8.47	351.86	-	-	8.53	58,994.32	-
1970 Investment share	31.80	26.11	118,315.94	23.77	4797.64	22.11	108.74	17.37	6439.86	-
1990 Investment share	10.14	15.40	532,586.6	13.68	79,215	16.07	63.60	12.90	76,943.7	-
1970 Government share	5.87	12.86	2.69	12.43	7.73	4.51	15.61	5.77	35,044.67	-
1990 Government share	4.97	12.59	22,127.8	11.92	1965.29	4.64	320.95	7.10	16,765.62	-
Polity score 1990	9	8.14	716,595.77	7.56	42,596.46	3.74	8.77	9.85	39,317.12	-
Durable score 1990	21	17.03	600.95	10.93	17,551.96	11.17	199.58	-	-	-
Economic score 1990	6	6.11	498,814.68	5.9	62,328.3	4.96	346.60	6.02	4.39	-
Polity score 1970	9	1.00	72,178.92	-0.24	13,410.62	7.80	9.51	7.97	99,601.63	-
Durable score 1970	1	11.76	848,731.76	10.12	126,632.05	27.70	8.44	-	-	-
Economic score 1970	6	3.74	108,702.97	3.16	31	5.80	13.72	5.16	17,358.28	-
Inflation variation	26.12	8.40	3886.26	19.44	188,732.86	137.58	9.92	34.09	69,903.87	-
GNI wealth per capita	20,189.65	11,545.27	77,935.44	6627.96	7295.11	-	-	1824.26	4501.4	-
Subsid asset wealth	22,775.54	-	-	-	-	-	-	2325.03	2694.5	-
Maternal mortality, 1990	94	100.99	250,656.77	157.59	786.78	149.82	4.23	-	-	-
Improved sanitation, 1990	82	76.64	1192.94	73.92	69.38	71.69	137.74	-	-	-

Note: Table shows the values of indicator variables and the average pre-Chavez outcome variable for actual and synthetic Venezuela. Population growth is in percentage points. Government share is a fraction of GDP, in percentage points. Polity and Ex. Const. follow the polity scores scale. Durability and the education variables are measured in years. The V-matrix column includes the diagonal entries in the V matrix. They determine how much weight each indicator variable receives in the minimization problem for the weights, $W = \arg \min(X_1 - X_0W)'V(X_1 - X_0W)$. The values are normalized relative to the entry for the first variable in V (generally the earliest available value of the outcome variable), which equals one. Although they resemble regression coefficients, they are formulated differently and cannot be interpreted as either partial correlations or marginal effects.

• A plausible composition of synthetic control groups for Venezuela



Introduction

“The [oil] sector’s decline began in 2003, following a strike by the employees and managers of Petróleos de Venezuela (PDVSA), the state oil company, in protest against Mr Chávez’s leftist policies. When the conflict ended, the president had almost 20,000 workers sacked. Since then, PDVSA’s chairman, Rafael Ramírez, has steadily replaced them with loyal chavistas: he has made it an explicit company policy to employ only supporters of the president. He has also allowed Mr Chávez to use the company as a piggy bank for his “socialist revolution”: last year, PDVSA spent twice as much on off-budget government programmes as it did on taxes, royalties and dividends...”

The Economist, 2012

Introduction

- Synthetic control method is a widely-used and popular tool for estimating the impact of the certain intervention on a single unit or multiple treated units ([Abaide and Gardeazabal 2003](#), [Abadie et. al. 2010, 2015](#), [Ben Michael et. al. 2018](#))
- The idea: construct a weighted average of control units that matches the treated unit's outcomes and auxiliary covariates before the intervention
- The difference in post-treatment outcomes between the treated and control unit captures the impact of the policy or intervention

Taxonomy of Institutional and Policy Shocks

- For a given intervention, one might be able to distinguish between three different types of shocks triggered by the intervention itself ([Garoupa and Spruk 2019](#))
- 1) **Gradual change** implies no major events but a set of continuous changes which help improve the outcome path without a major deviation
- 2) **Change imposed by a shock** is a major external event or intervention that imposes institutional change with short-run/temporary impact but does not derail long-run equilibrium
- 3) **Structural break** implies a major event that changes the long-run outcome path and depends on the nature of the shock:
 - *Breakup*: permanent upward change of long-run equilibrium
 - *Breakdown*: permanent downward change of long-run growth

Inference

- Is the effect of the intervention/event/shock on the outcome of interest statistically significant or not?
- How likely it is the effect is obtained by chance and is driven by the lack of fit?
- We compute the statistical significance of the counterfactual outcome in the hypothetical absence of the shock and obtain the distribution of placebo effects
- If the distribution of placebo effects of institutional shocks yields many effects as large as the core estimate for the quasi-treated unit, then the estimated impact of the intervention is observed by chance
- A non-parametric test without imposing any distribution on the random error term ([Galiani and Quirtstorff 2018](#))

Inference

- When the institutional shock is randomly distributed across the sample, the placebo distribution serves as a typical randomization inference
- But since it may not satisfy the strict exogeneity assumption, the obtained probabilities are then interpreted as the proportion of countries that have an estimated effect of the institutional shock in the same year at least as large as the effect of the affected country
- Placebo effects may be relatively large if the treatment and control units are not well matched in the period preceding the institutional shock

Inference

- Abadie et. al. (2010) study the effect of Proposition 99 (i.e. a large-scale tobacco control program) implemented by California in 1988 on tobacco consumption
- They implement synthetic control method and simulate the counterfactual trajectory of California's tobacco consumption in the hypothetical absence of the Proposition 99
- A large drop in tobacco consumption in the post-proposition period
- The gap in tobacco consumption between the actual and synthetic California is perceivable and appears to be specific to California
- A battery of inferential technique based on the elimination of poorly fit placebo distributions

Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program

Alberto ABADIE, Alexis DIAMOND, and Jens HAINMUELLER

Building on an idea in Abadie and Gardeazabal (2003), this article investigates the application of synthetic control methods to comparative case studies. We discuss the advantages of these methods and apply them to study the effects of Proposition 99, a large-scale tobacco control program that California implemented in 1988. We demonstrate that, following Proposition 99, tobacco consumption fell markedly in California relative to a comparable synthetic control region. We estimate that by the year 2000 annual per-capita cigarette sales in California were about 26 packs lower than what they would have been in the absence of Proposition 99. Using new inferential methods proposed in this article, we demonstrate the significance of our estimates. Given that many policy interventions and events of interest in social sciences take place at an aggregate level (countries, regions, cities, etc.) and affect a small number of aggregate units, the potential applicability of synthetic control methods to comparative case studies is very large, especially in situations where traditional regression methods are not appropriate.

KEY WORDS: Observational studies; Proposition 99; Tobacco control legislation; Treatment effects.

1. INTRODUCTION

Social scientists are often interested in the effects of events or policy interventions that take place at an aggregate level and affect aggregate entities, such as firms, schools, or geographic or administrative areas (countries, regions, cities, etc.). To estimate the effects of these events or interventions, researchers often use comparative case studies. In comparative case studies, researchers estimate the evolution of aggregate outcomes (such as mortality rates, average income, crime rates, etc.) for a unit affected by a particular occurrence of the event or intervention of interest and compare it to the evolution of the same ag-

Comparing the evolution of an aggregate outcome (e.g., state-level crime rate) between a unit affected by the event or intervention of interest and a set of unaffected units requires only aggregate data, which are often available. However, when data are not available at the same level of aggregation as the outcome of interest, information on a sample of disaggregated units can sometimes be used to estimate the aggregate outcomes of interest (like in Card 1990 and Card and Krueger 1994).

Given the widespread availability of aggregate/macro data (e.g., at the school, city, or region level), and the fact that many policy interventions and events of interest in the social sciences take place at an aggregate level, comparative case study re-

Inference

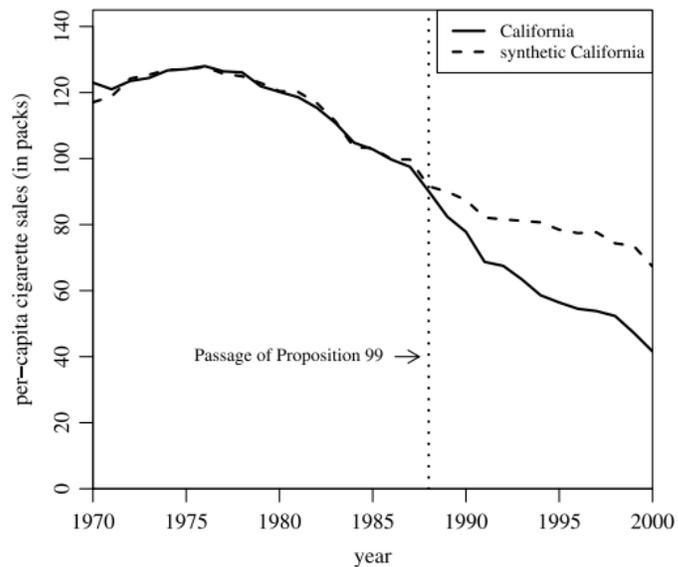


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Table 2. State weights in the synthetic California

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	-	Nebraska	0
Arizona	-	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	-
Connecticut	0.069	New Mexico	0
Delaware	0	New York	-
District of Columbia	-	North Carolina	0
Florida	-	North Dakota	0
Georgia	0	Ohio	0
Hawaii	-	Oklahoma	0
Idaho	0	Oregon	-
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	-	Vermont	0
Massachusetts	-	Virginia	0
Michigan	-	Washington	-
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

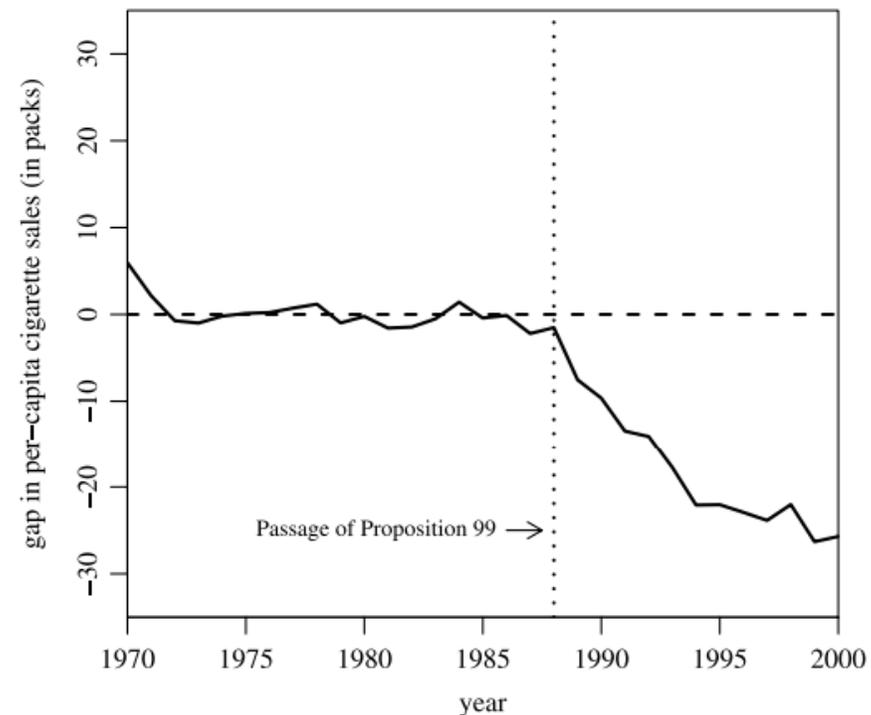


Figure 3. Per-capita cigarette sales gap between California and synthetic California.

Inference

- What if the results are entirely driven by chance alone?
- How often would a gap of this magnitude be obtained if a researcher chose a state at random for the study in place of California?
- Estimate placebo studies by applying the synthetic control algorithm to states that did not implement a large-scale tobacco control program during the evaluation period
- If the magnitude of placebo gaps is similar to the one for the treated state/country, the analysis does not provide evidence of significant impact of intervention
- If the magnitude of the gap for the treated state is unusually large relative to placebo gaps, the evidence of significant impact of the intervention is very likely

Inference

- The negative tobacco consumption gap for California is unusually large even when poorly fit placebos are discarded from the analysis
- A crude way to test the significance of the gap is to compute quasi p-values and post/pre-intervention RMSE ratio

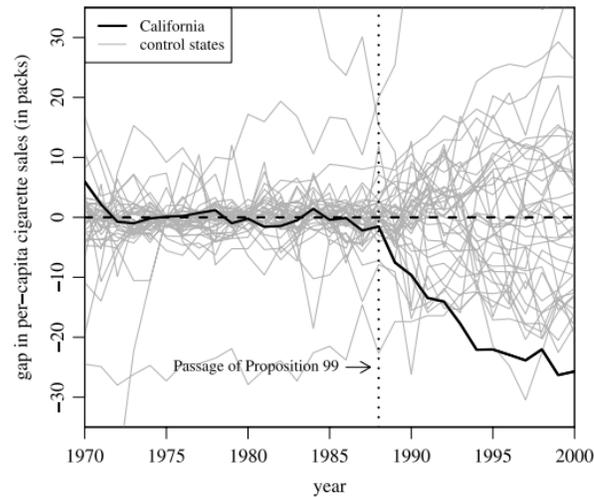


Figure 4. Per-capita cigarette sales gaps in California and placebo gaps in all 38 control states.

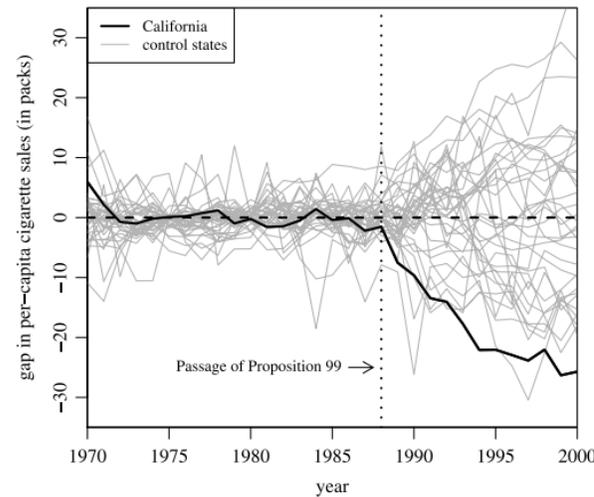


Figure 5. Per-capita cigarette sales gaps in California and placebo gaps in 34 control states (discards states with pre-Proposition 99 MSPE twenty times higher than California's).

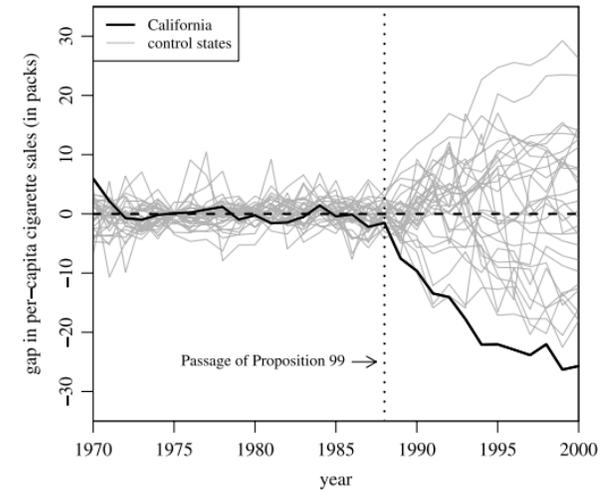


Figure 6. Per-capita cigarette sales gaps in California and placebo gaps in 29 control states (discards states with pre-Proposition 99 MSPE five times higher than California's).

Inference

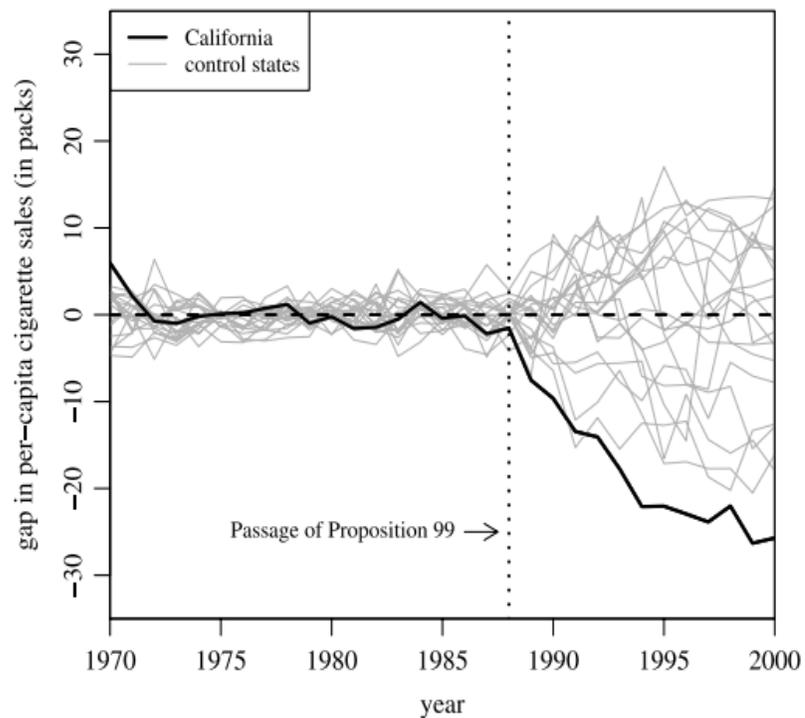


Figure 7. Per-capita cigarette sales gaps in California and placebo gaps in 19 control states (discards states with pre-Proposition 99 MSPE two times higher than California's).

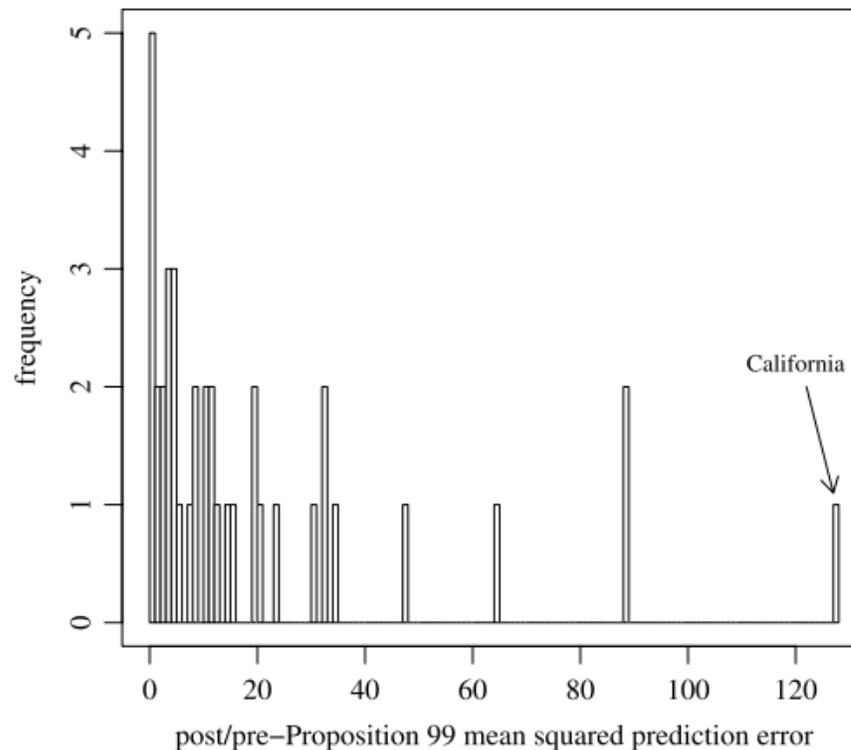
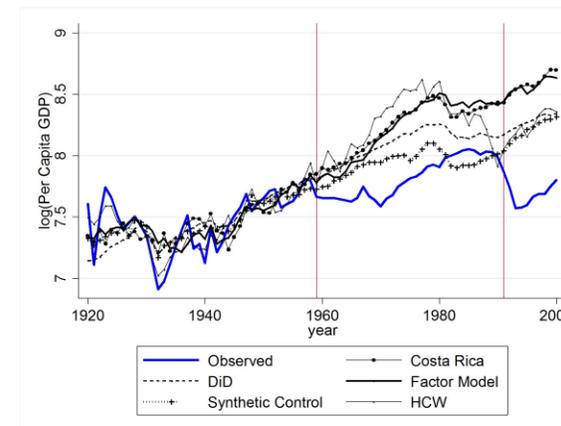
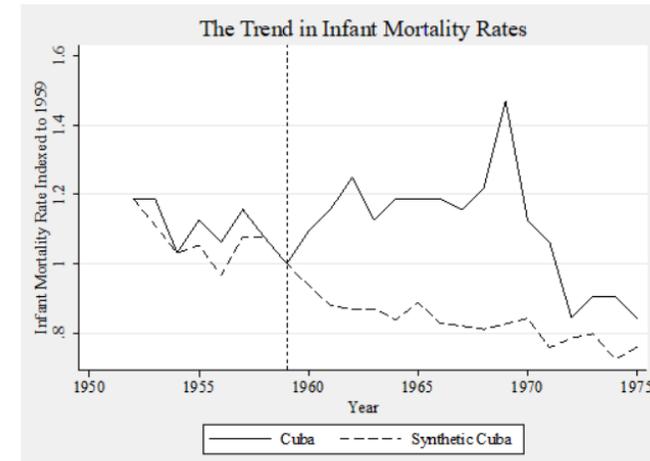


Figure 8. Ratio of post-Proposition 99 MSPE and pre-Proposition 99 MSPE: California and 38 control states.

Synthetic Control Method: Applications in Economic History

- The Cuban Experiment
- A large-scale socialist revolution with a long-lasting Communist Party rule
- A major experiment to study the effects of socialist economic and political institutions on a range of economic outcomes
- Bologna Pavlik and Geloso (2018) employ SCM to study the effect of Castro regime on infant mortality in post-1959 period and find that mortality rate increased relative to the counterfactual
- Jales et. al. (1959, WE): use SCM to study the effect of Cuban Revolution on the trajectory of growth, and find a negative impact

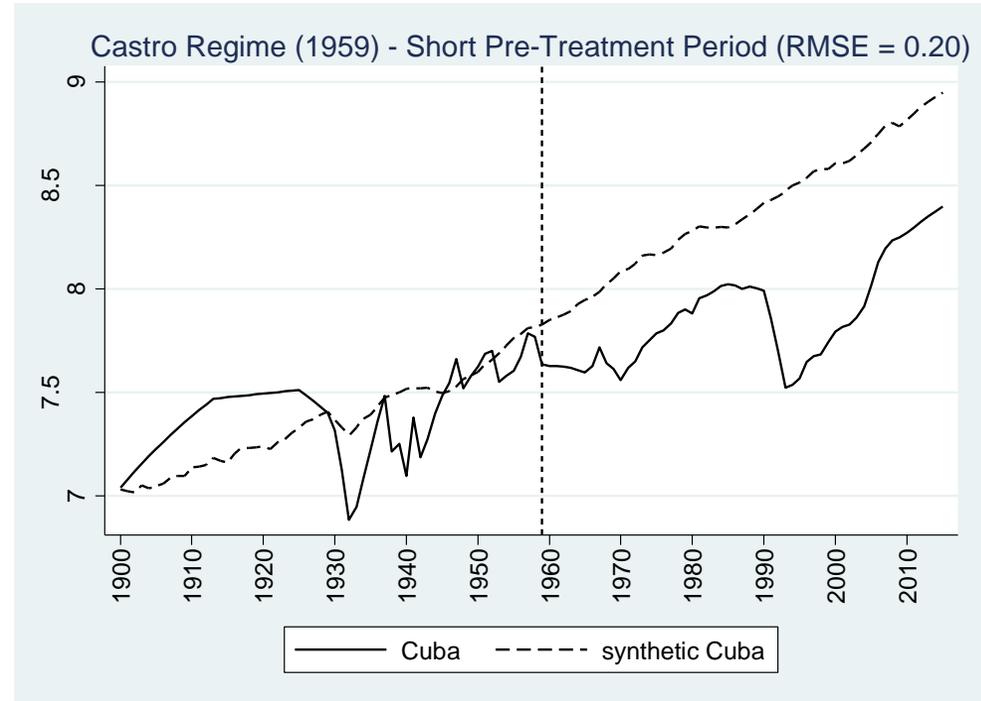
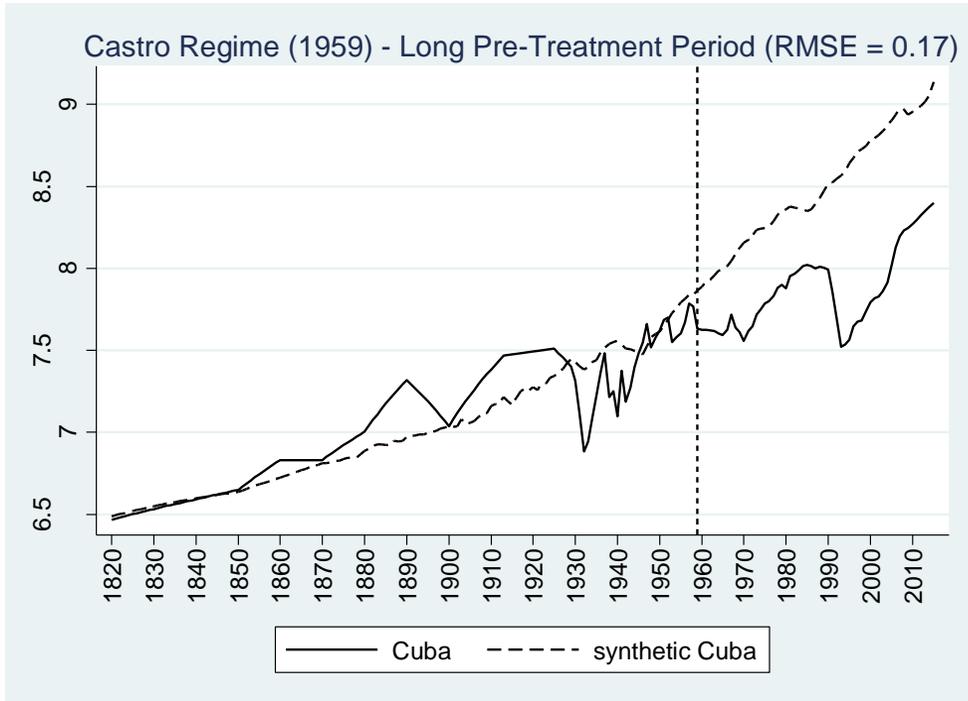
Figure 7: IMR (indexed to 1959) in Cuba compared to synthetic Cuba with pre-treatment period restrained to 1952–1959



Synthetic Control Method: Applications in Economic History

- The dependent variable is per capita GDP (1990 Geary-Khamis international dollar with a single benchmark) from [Bolt and Van Zanden \(2014\)](#)
- A rich set of covariates related to lagged outcomes, physical geography characteristics, legal history, demographic structure, genetic diversity, and institutional framework before the intervention
- 58 countries for the period 1820-2015 with a total of 10,976 country/year observation pairs

Synthetic Control Method: Applications in Economic History

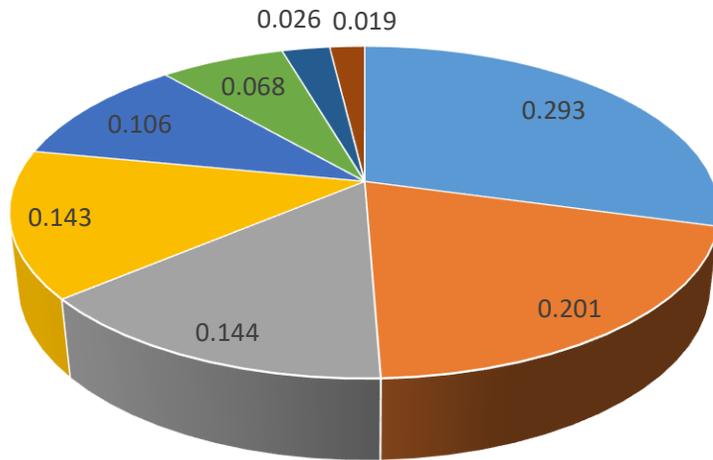


- In the long term (i.e. in 2015), the GDP per capita of Cuban economy w/o Castro regime is 42 percent higher

Synthetic Control Method: Applications in Economic History

Castro Regime (1959) – Long Pre-Treatment Period
(1820-1958)

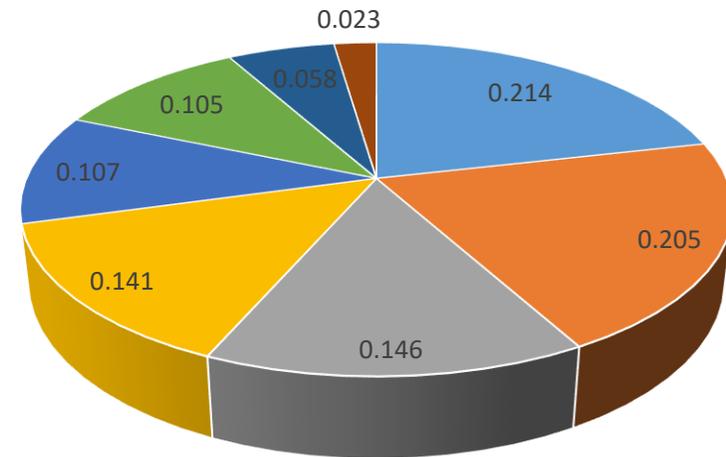
RMSE = 0.17



■ Ireland ■ Philippines ■ Colombia ■ Jamaica ■ Sri Lanka ■ Chile ■ Venezuela ■ Mexico

Castro Regime (1959) - Short Pre-Treatment Period

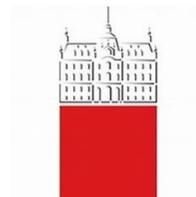
RMSE = 0.20



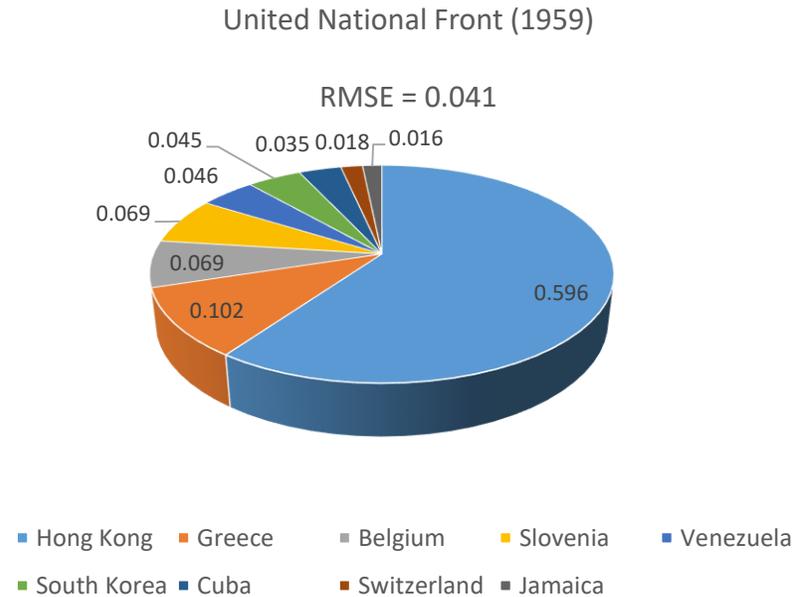
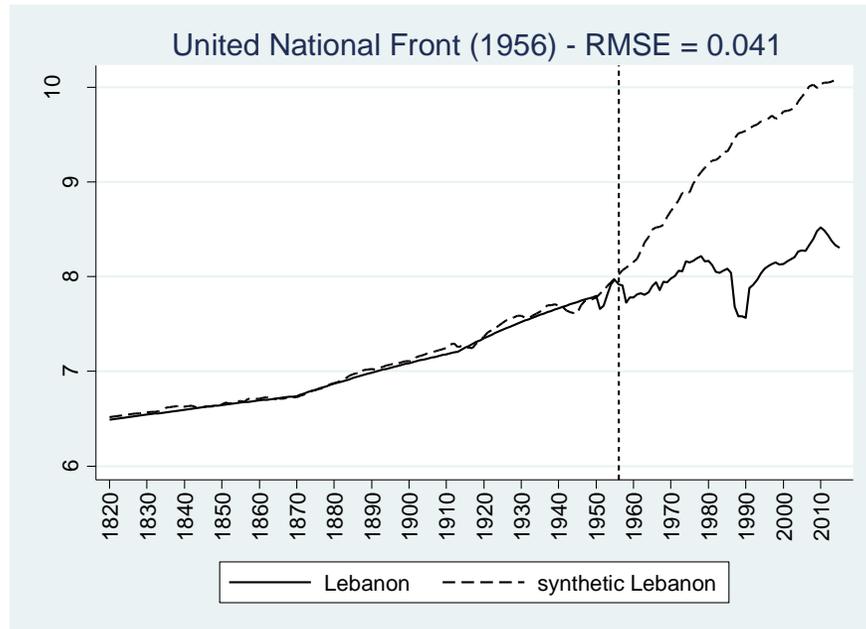
■ Sri Lanka ■ Colombia ■ Mexico ■ Jamaica
 ■ Philippines ■ United Kingdom ■ Chile ■ Netherlands

Synthetic Control Method: Applications in Economic History

- Lebanon formally achieved its independence from League of Nations (LON) mandate under French administration on November 22, 1943
- In 1958, Lebanon plunged into a short-lived civil war that set the stage for the subsequent political instability and another armed civil conflict in 1975
- The civil war represented a conflict between the pro-Western government of Camille Chamoun and opposition-led United National Front consisting of Al-Mourabitoun and a coalition of communist and socialist parties wanting to join the United Arab Republic
- An important and long-lasting institutional shock



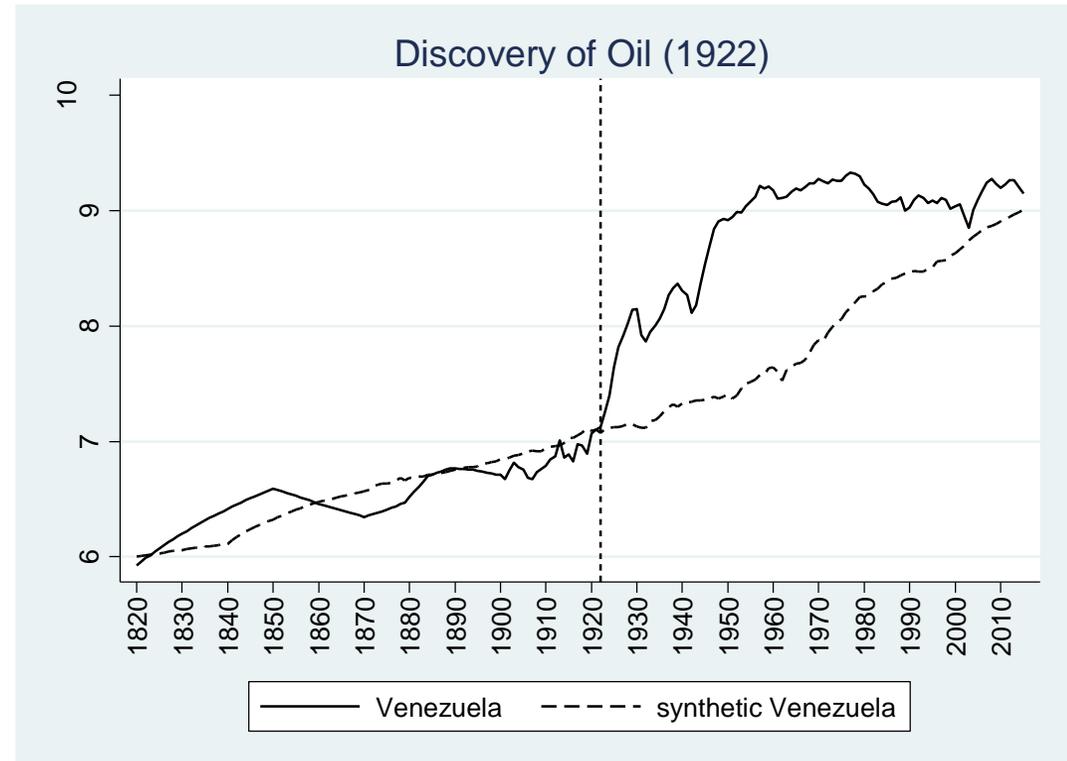
Synthetic Control Method: Applications in Economic History



- A synthetic Lebanon w/o the 1958 political crisis/UNF has more than 4-fold higher per capita down to the present day

Synthetic Control Method: Applications in Economic History

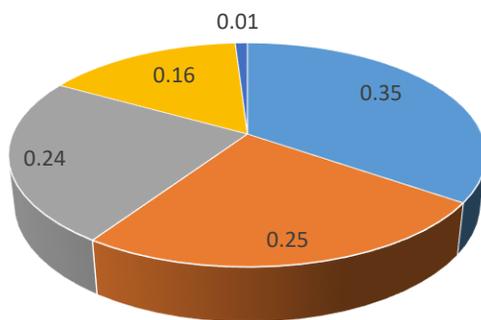
- The blowout of Barosso II oil well in Cabimas (Zulia State) in 1922 marks the beginning of Venezuela's major oil producer status
- Enormous economic growth and structural change potential posited by oil discovery
- One of the worst instances of Dutch disease
- Distortionary economic policies and institutional instability gradually eliminated all economic advantages of oil discovery
- The counterfactual dynamics is not driven by (i) democratization in 1950s, (ii) oil nationalization in 1973, and (iii) rise of Chavez
- Adjustment back to long-run equilibrium without oil discovery



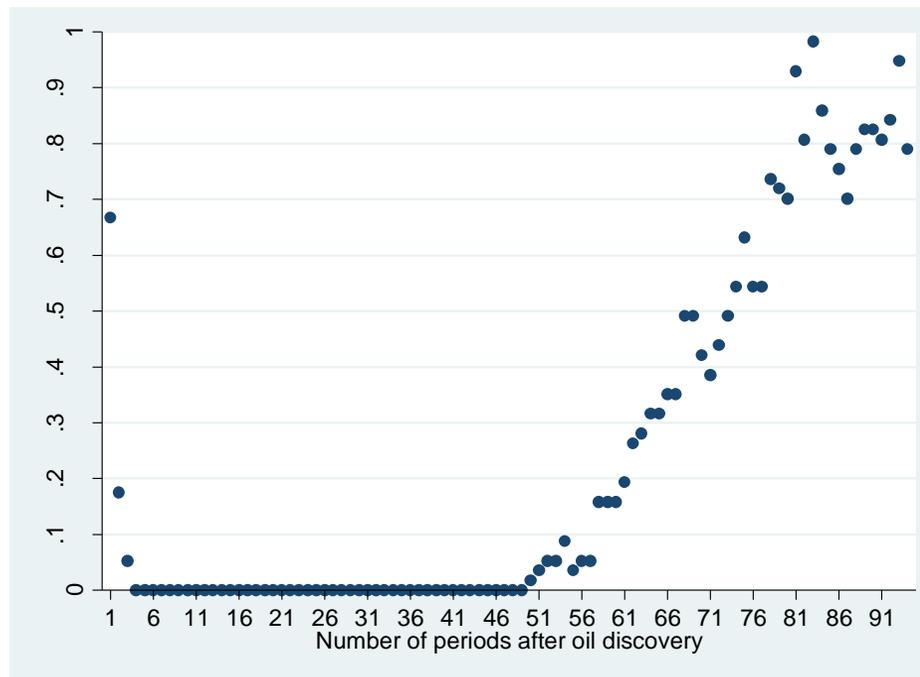
Synthetic Control Method: Applications in Economic History

synthetic Venezuela that would have left oil in the ground in 1922

RMSE = 0.16



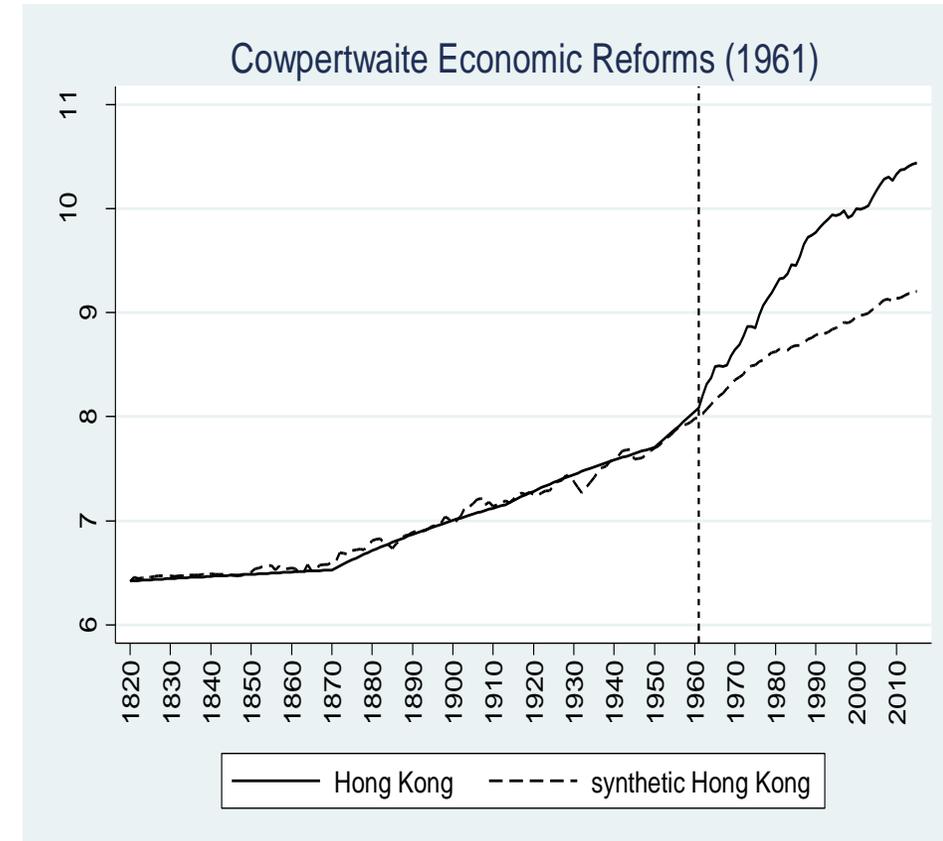
■ Algeria ■ Morocco ■ South Korea ■ New Zealand ■ Indonesia



- The growth significance of oil discovery begins to disappear in the period right after the nationalization of oil production in 1973 (!)

Synthetic Control Method: Applications in Economic History

- Long-run development of Hong Kong in response to the 1961 economic reforms introduced by Governor Sir Cowperthwaite
- The evidence clearly suggests that Hong Kong case study is an example of the structural breakup
- He implemented a broad range of economic reforms. These reforms promoted free trade, low taxation, budget surpluses, limited state intervention in the economy, a distrust of industrial planning, and sound money (Monnerly 2017)
- The real Hong Kong vastly outperforms the synthetic Hong Kong without the Cowperthwaite economic reforms in 1961
- In 2015, the real Hong Kong's GDP per capita is 109% higher (=33,860/16,155) compared to the counterfactual scenario with no implementation of economic reforms in 1961.



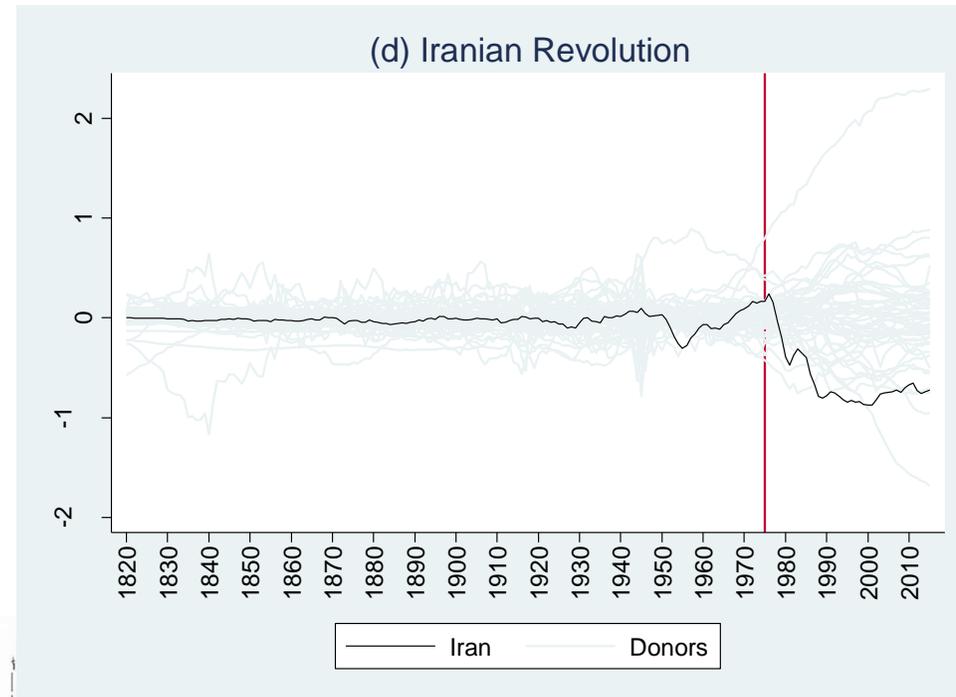
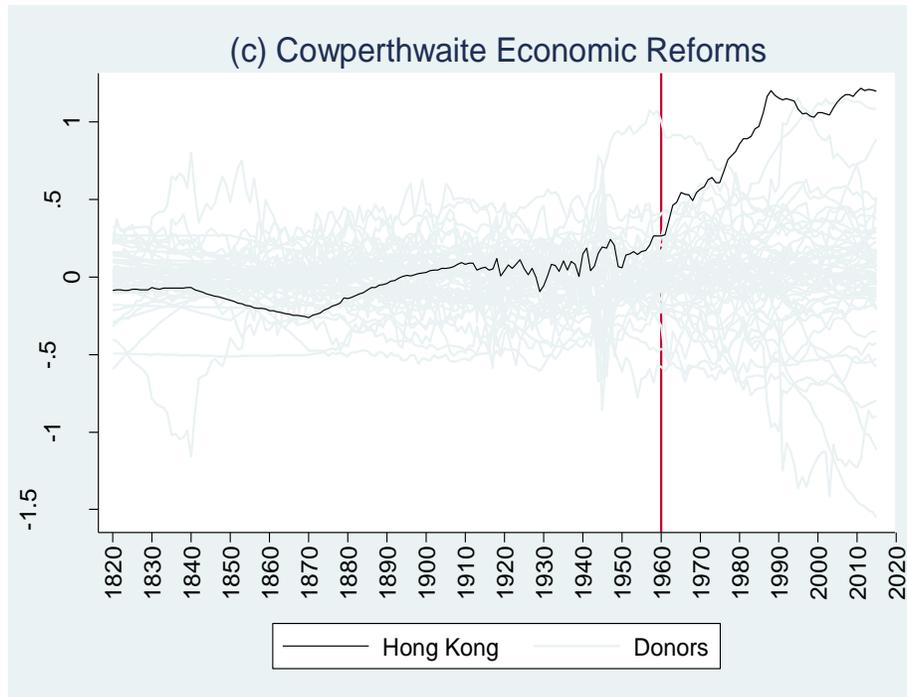
Synthetic Control Method: Applications in Economic History

- Following the Iranian Revolution, the economic policies in the post-revolutionary period emphasized significant state ownership of enterprises, a large public sector, price controls, subsidies, administrative controls and rationing of goods such as gasoline (Garoupa and Spruk 2019)
- Institutional change towards the Islamic republic is associated with a pervasive structural breakdown
- Long-run growth path of synthetic Iran outperforms its real counterpart and indicates a significant long-term growth shortfall
- Since the post/pre-1977 prediction error is low (8.6%), it is unlikely that shocks in the pre-1977 period affect the direction of the counterfactual scenario in the post-intervention period
- In 2015, Iran's per capita income is about 46% lower (=6,247/11,498) than per capita income of Iran's synthetic control group



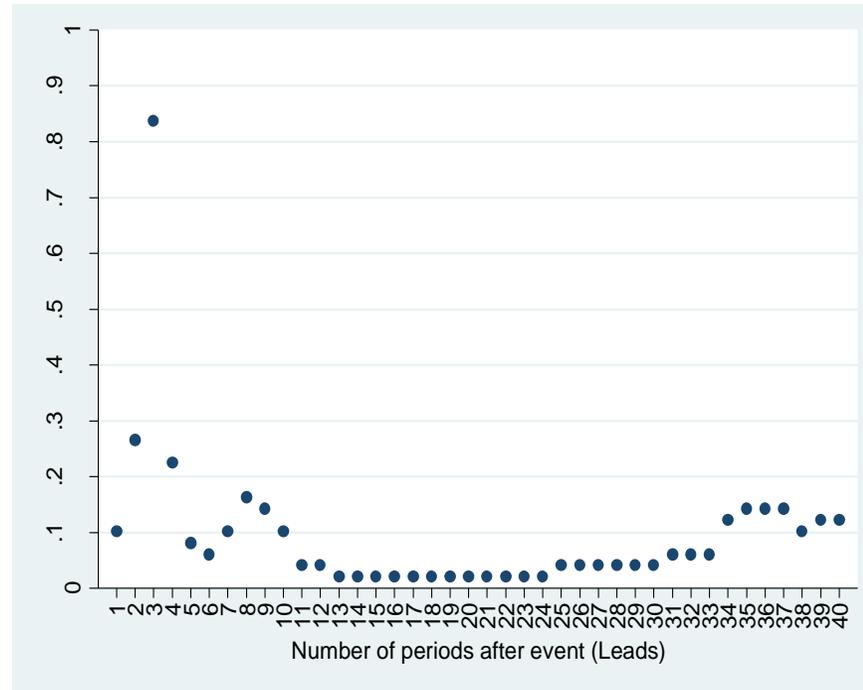
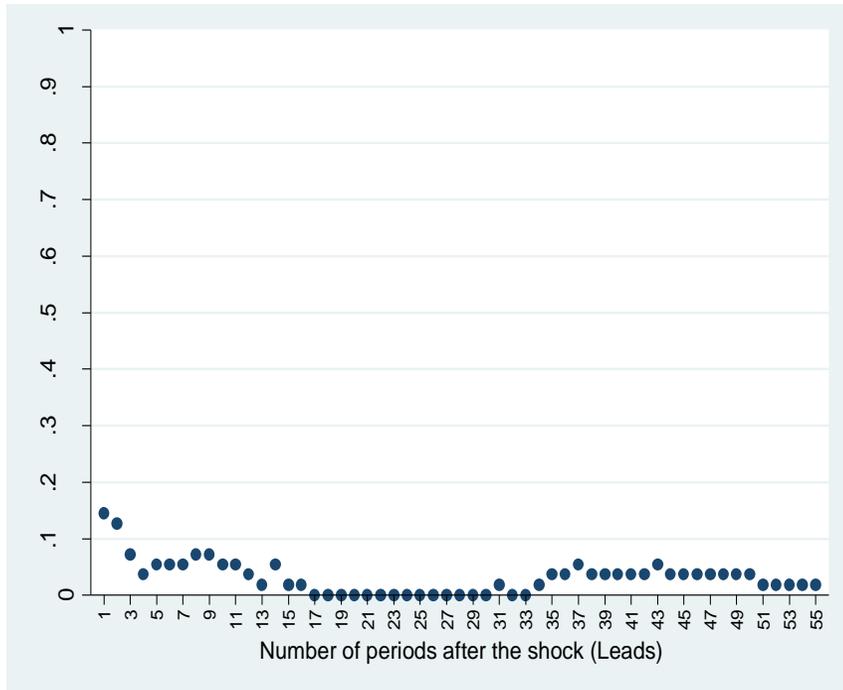
Synthetic Control Method: Applications in Economic History

- In-space placebo studies indicate unusually large per capita income gaps that do not appear to be perceivable elsewhere



Synthetic Control Method: Applications in Economic History

Permutation-based p-values on the null hypothesis of growth effects of Cowperthwaite reforms and Iranian revolution within the convention 5%-10% bound



Synthetic Control Method: Applications in Economic History



Synthetic Control Method: Applications in Economic History

- Exploit the 1947 partition of Trieste territory to obtain a plausible counterfactual scenario of Friuli Venezia Giulia's economic growth trajectory had not adopted the institutional framework of postwar Italian Republic
- Long-run effect of rejecting socialist beliefs and avoiding democratic socialism
- A synthetic Friuli Venezia Giulia without the Italian unification in 1947 has 41 percent lower per capita income than its actual counterpart
- The set of countries that best reproduces Friuli's pre-unification growth trajectory is dominated by Argentina (37%)
- Almost all countries in the synthetic control group never experienced a similar type of shock (i.e. a permanent transition to checks and balances without institutional instability, civil unrest and political turmoil)

Synthetic Control Method: Applications in Economic History

Pre-unification descriptive statistics

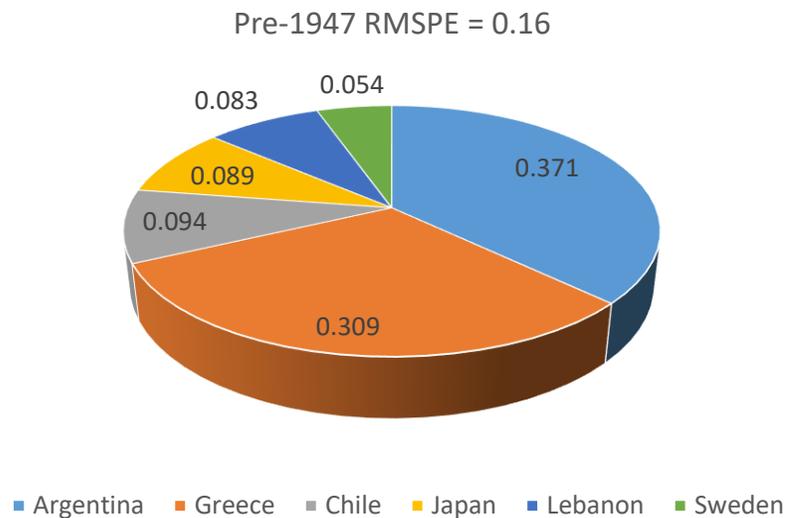
	Friuli Venezia Giulia		Country-Level Control Sample	
	Mean	Std	Mean	Std
Panel A: Past GDP Per Capita Dynamics Covariates				
log GDP per capita in 1871	7.46	-	7.18	0.61
log GDP per capita (t-1)	8.93	1.09	8.20	1.10
log GDP per capita (t-2)	8.91	1.08	8.19	1.10
Average growth rate (1872-1900)	0.013	0.021	0.010	0.045
Average growth rate (1900-1946)	0.009	0.108	0.009	0.068
Panel B: Physical Geography Covariates				
Latitude	46.22	0	32.59	16.61
Longitude	13.1	0	25.77	64.98
Desert	0	0	0.23	0.42
Land area	3035	0	413854	834132
Landlocked	0	0	0.11	0.31
Island	0	0	0.14	0.34
Panel C: Legal and Institutional History Covariates				
French civil law	1	0	0.25	0.43
British common law	0	0	0.50	0.49
German civil law	0	0	0.15	0.36
Scandinavian civil law	0	0	0.06	0.24

Pre-unification covariate balance

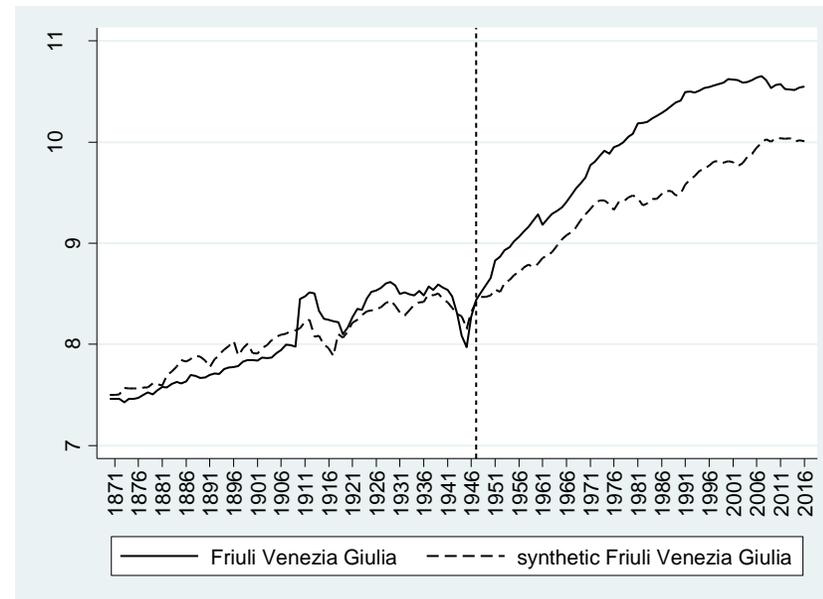
	Friuli Venezia Giulia	Synthetic Control Unit
Log gdp per capita in 1871	7.46	7.49
Log gdp per capita(t-1)	8.03	8.03
Log gdp per capita(t-2)	8.03	8.03
Average growth rate (1872-1900)	0.013	0.014
Average growth rate (1900-1946)	0.009	0.006
Latitude	46.22	38.30
Longitude	13.1	-7.62
Desert	0	0.46
Land area	3035	464350
Landlocked	0	0
French civil law	1	0.86
British common law	0	0
German civil law	0	0.09
Scandinavian civil law	0	0.05

Synthetic Control Method: Applications in Economic History

Synthetic Friuli Venezia Giulia

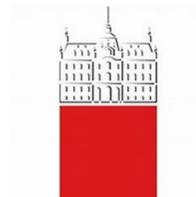


Long-Run Growth Impact of Avoiding Socialist Economic and Political Institutions



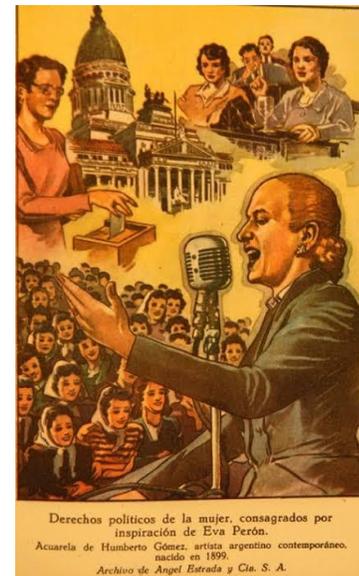
Synthetic Control Method: Applications in Economic History

- The long-term economic growth effect of populist political movements ([Marzetti and Spruk 2019](#))
- Exploit the rise of Peronist labor and social legislation in 1943 to estimate the long-run growth impact of populist political movements
- An idea: construct an artificial control group for Argentina that mimics its growth trajectory that never had such a large-scale social justice movement in de jure and de facto political power
- Argentina is the only developing country that once had higher GDP per capita than most of today's developed nations



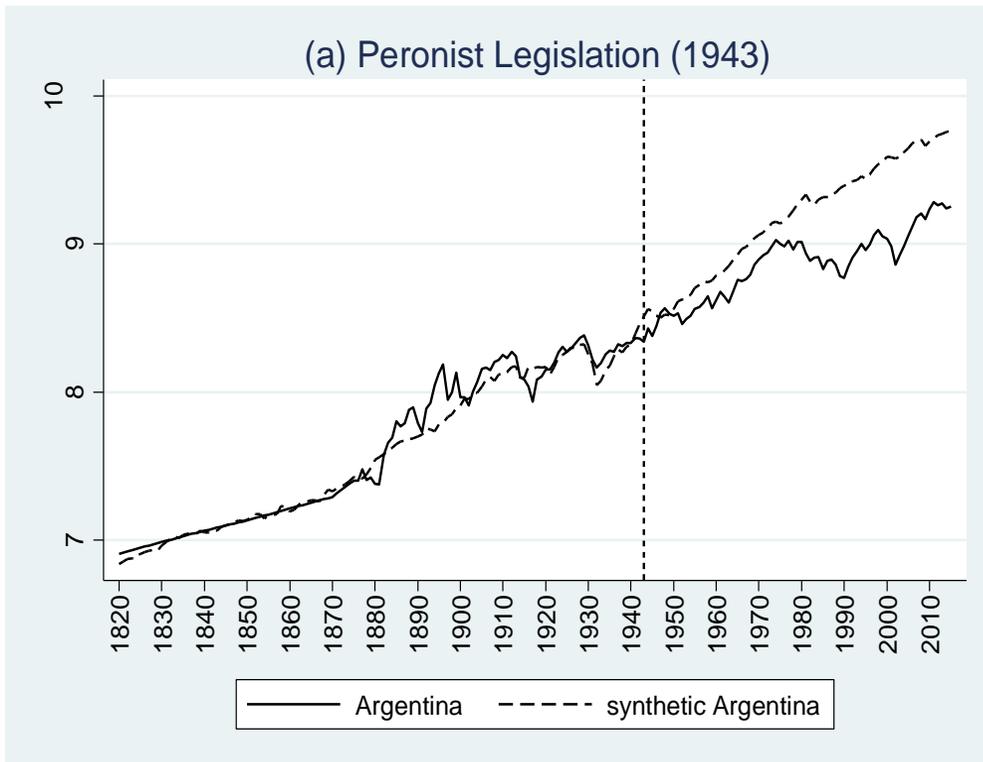
Synthetic Control Method: Applications in Economic History

- 1943 populist political movement espousing social justice, massive expansion of the welfare state and political clientelism



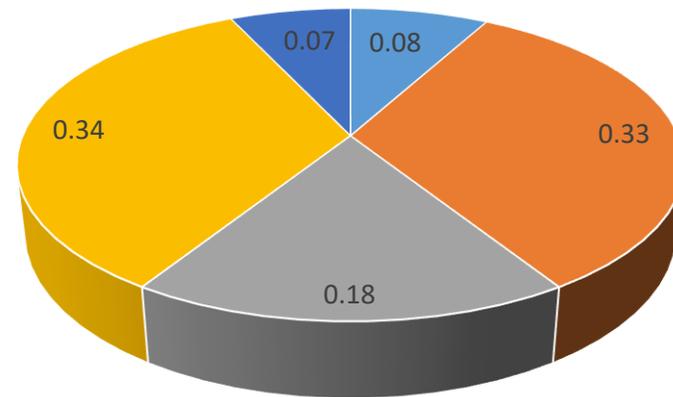
Synthetic Control Method: Applications in Economic History

The long-run growth effect of Peronism



synthetic Argentina w/o Péron

RMSE = 0.08

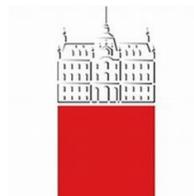
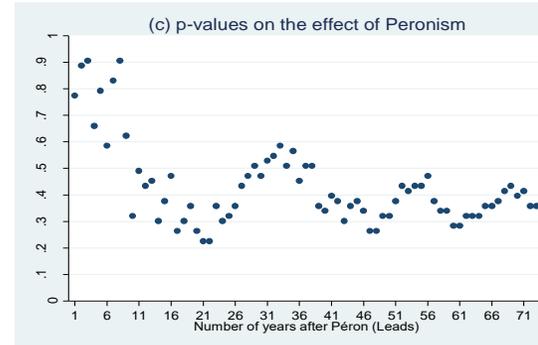
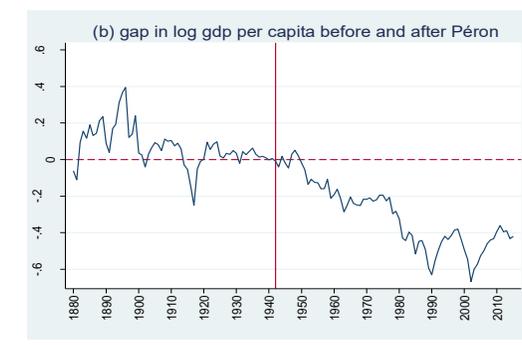
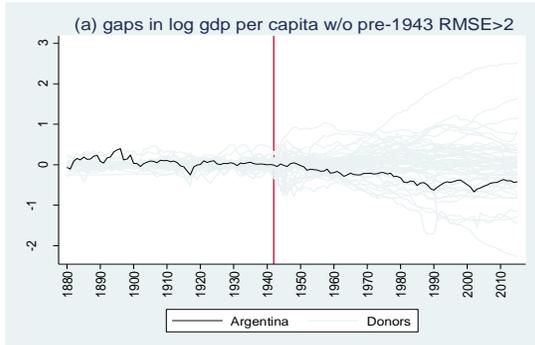
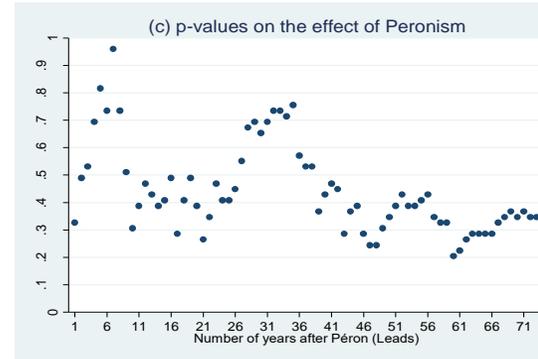
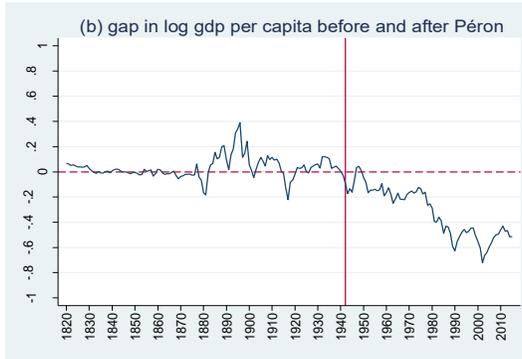
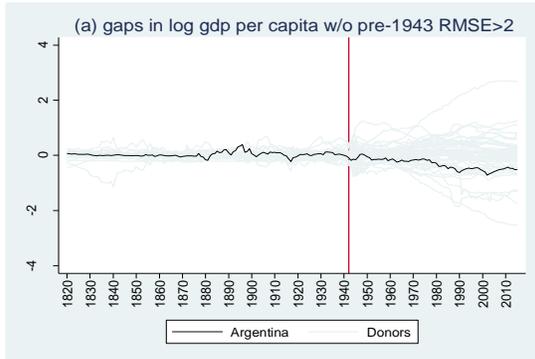


■ Chile ■ Mexico ■ Switzerland ■ United States ■ Uruguay



Synthetic Control Method: Applications in Economic History

Inference on the long-run growth impact of Peronism

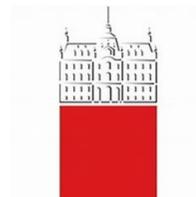


Synthetic Control Method: Applications in Economic History

- The difference in economic development between the north and south of Italy is well known in economic literature ([Putnam 1993](#), [Iuzzolino et al. 2011](#); [Ciccarelli and Fenoaltea 2013](#), [Dincecco et al. 2011](#); [de Oliveira and Guerriero 2018](#))
- In spite of 150 years of common economic, political and legal institutions, the divide between a rich north and poor south persists down to the present day
- Richest region (i.e. Lombardia) enjoys per capita income two times higher than the poorest region (i.e. Calabria)
- Regional economic development has depended on how institutions have performed locally, the latter aspect depending on the local level of social capital and cohesion

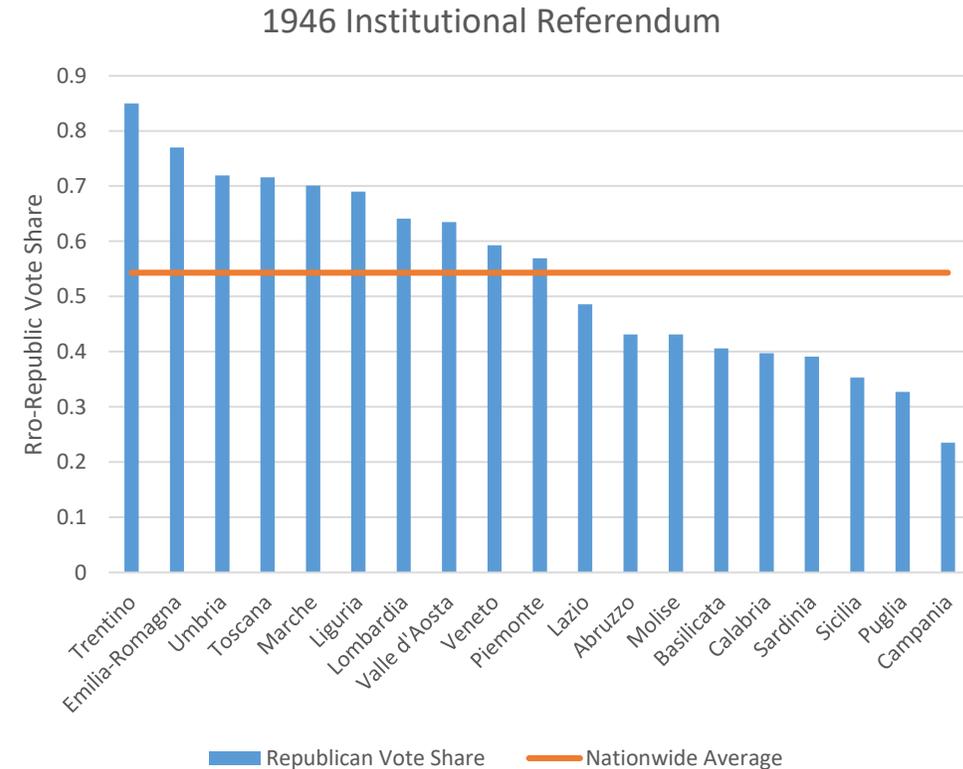
Synthetic Control Method: Applications in Economic History

- Exploit the 1946 institutional transition from monarchy to republic to estimate how institutional dependence on local customs and social capital determines the growth effect of the transition ([Melcarne and Spruk 2019](#))
- In the institutional referendum Republic won, but with significant variance across regions, with northern regions up to 80% in favor of the republic, as opposed to an average 30% for the south
- Exploit this historical shock to compute Italian regions' counterfactual development paths through synthetic control method
- Pro-republic northern regions are also the ones that most benefited from the institutional transition, as opposed to most southern ones, for which the counterfactuals are substantially smaller
- Non-democracies as a control sample

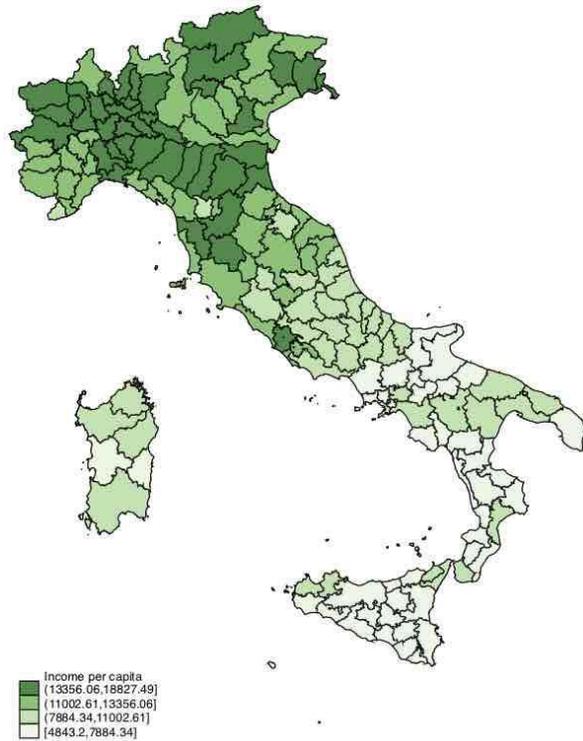


Synthetic Control Method: Applications in Economic History

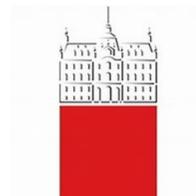
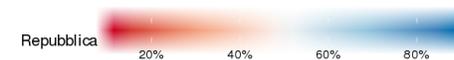
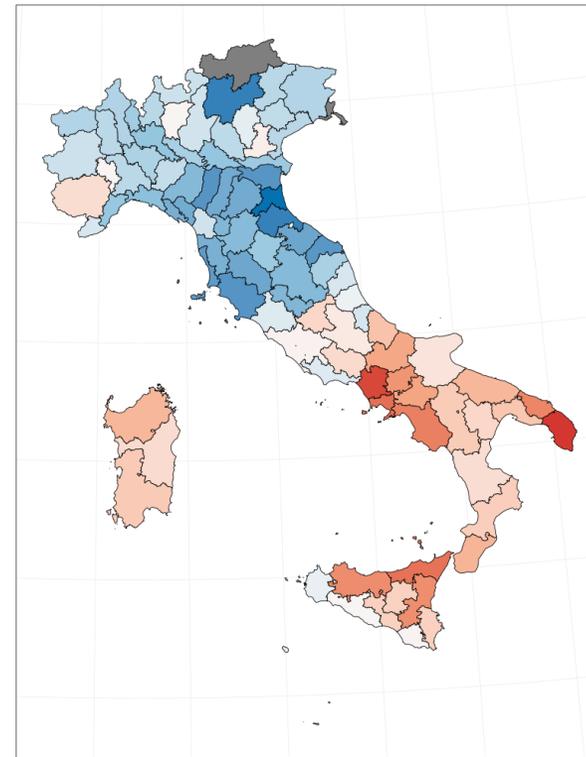
Regional Republic Vote Share in 1946 Institutional Referendum



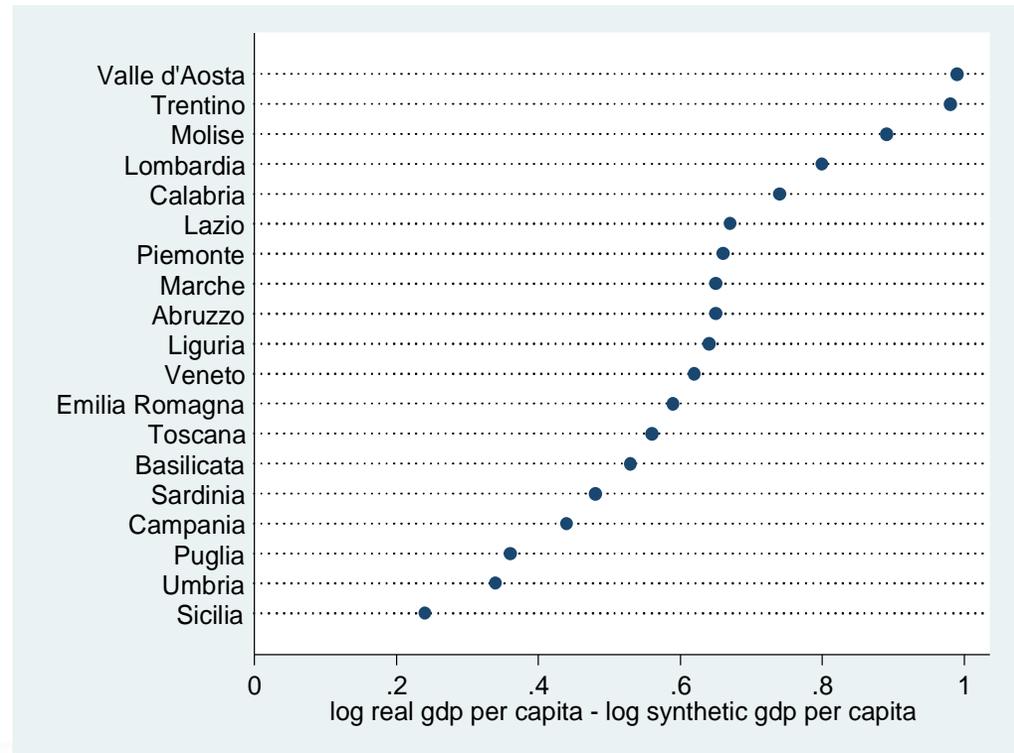
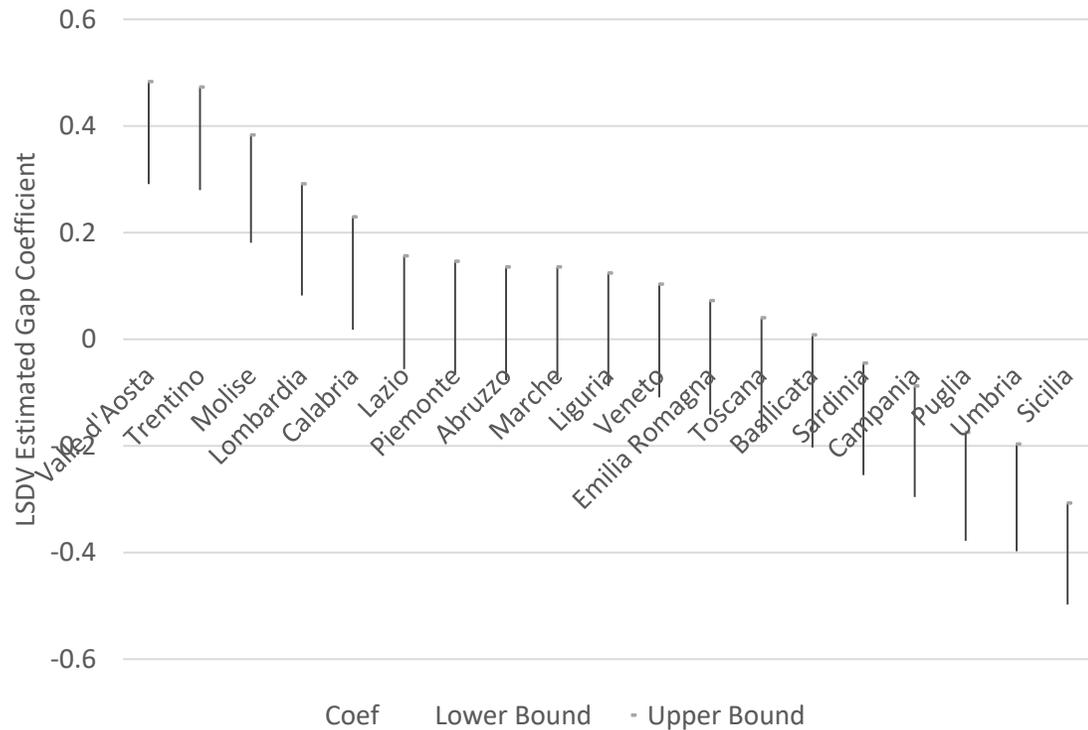
Synthetic Control Method: Applications in Economic History



Referendum sulla forma istituzionale dello Stato, 1946

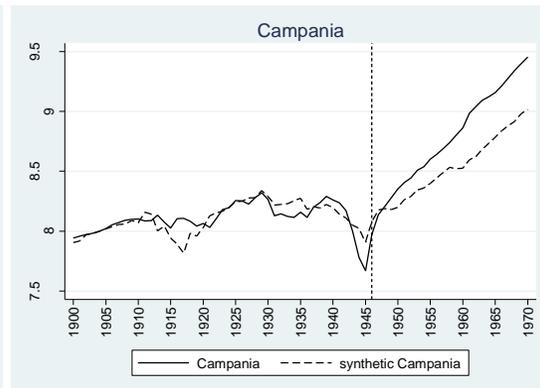
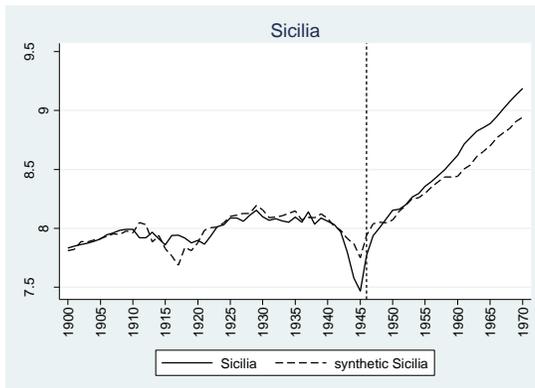
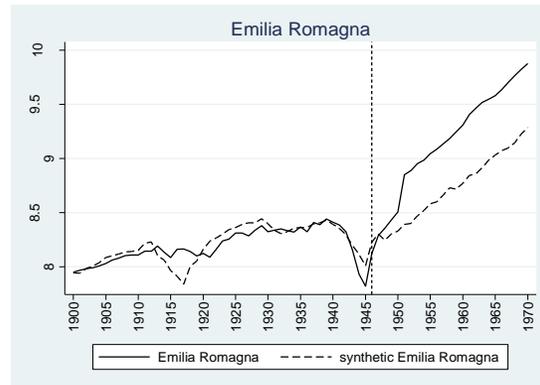
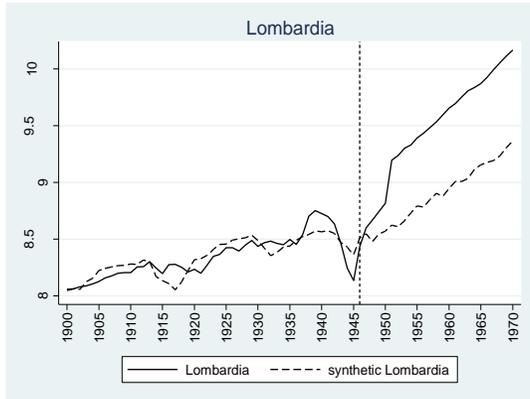


Synthetic Control Method: Applications in Economic History



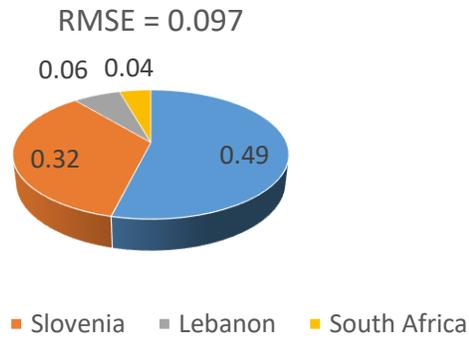
Synthetic Control Method: Applications in Economic History

- Quite large regional variance in the long-term growth effect of institutional transition

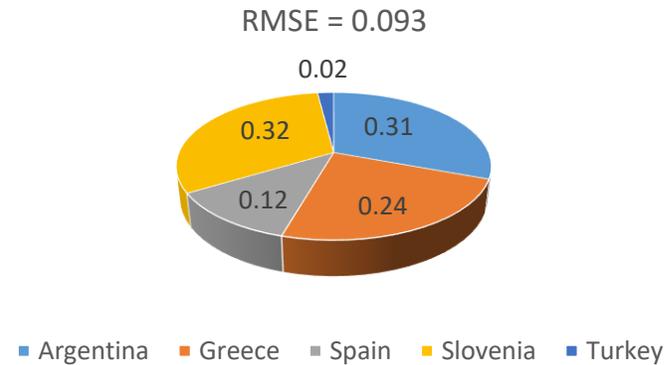


Synthetic Control Method: Applications in Economic History

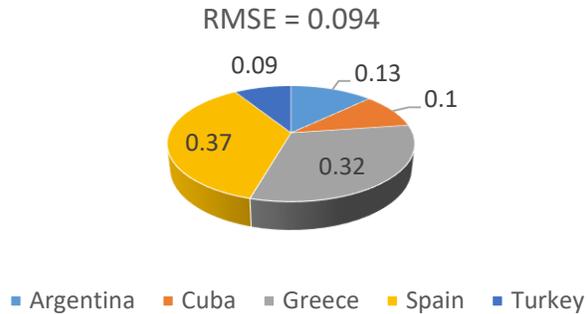
synthetic Lombardia



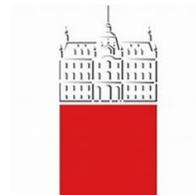
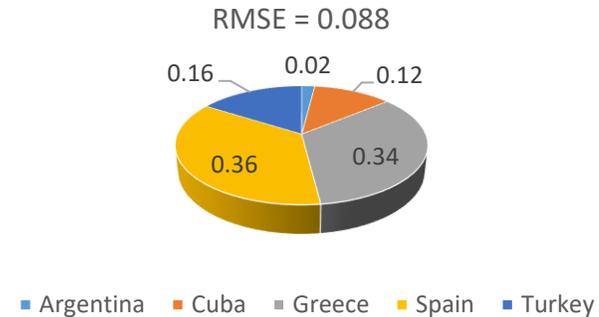
synthetic Emilia Romagna



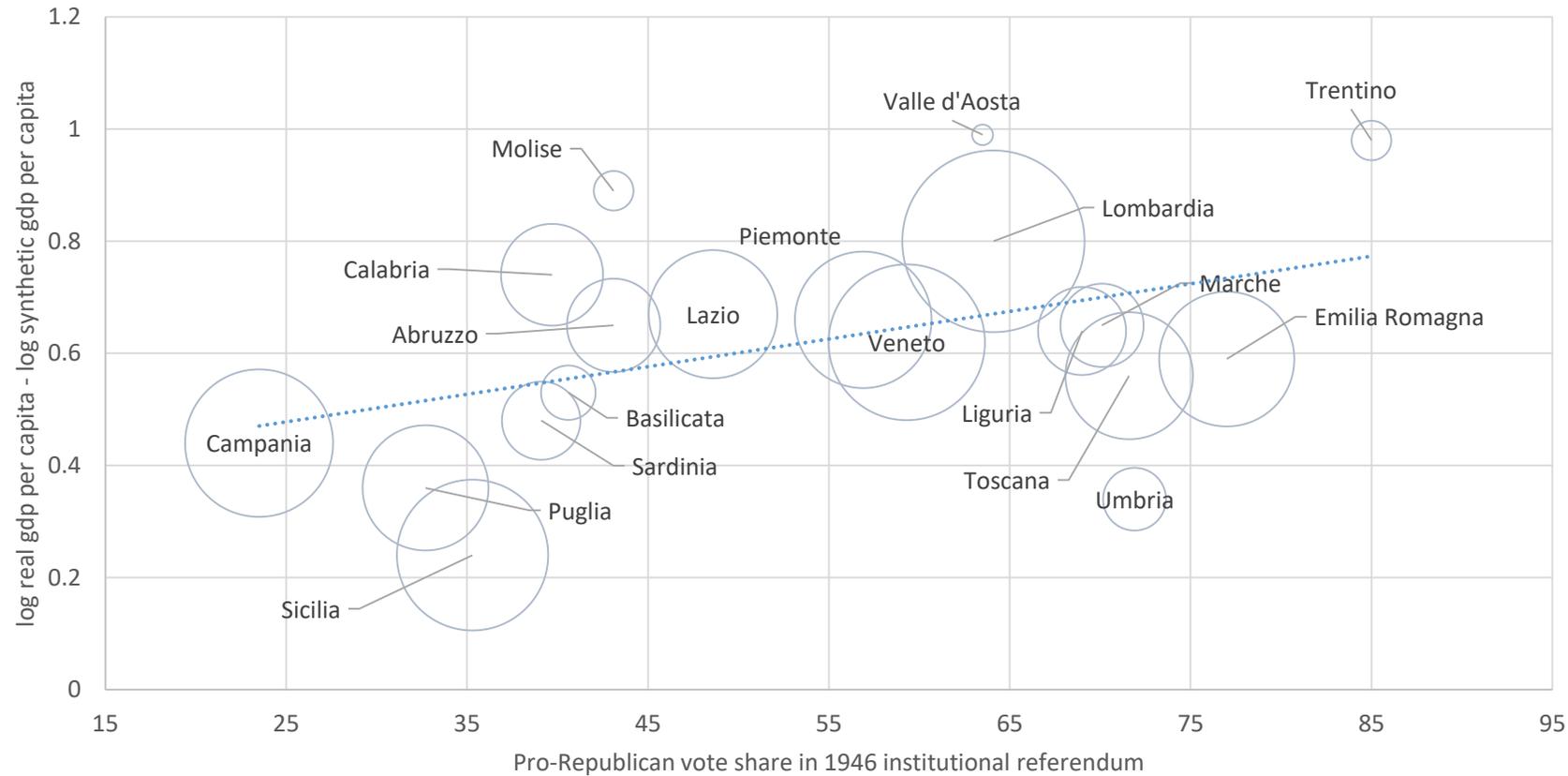
synthetic Campania



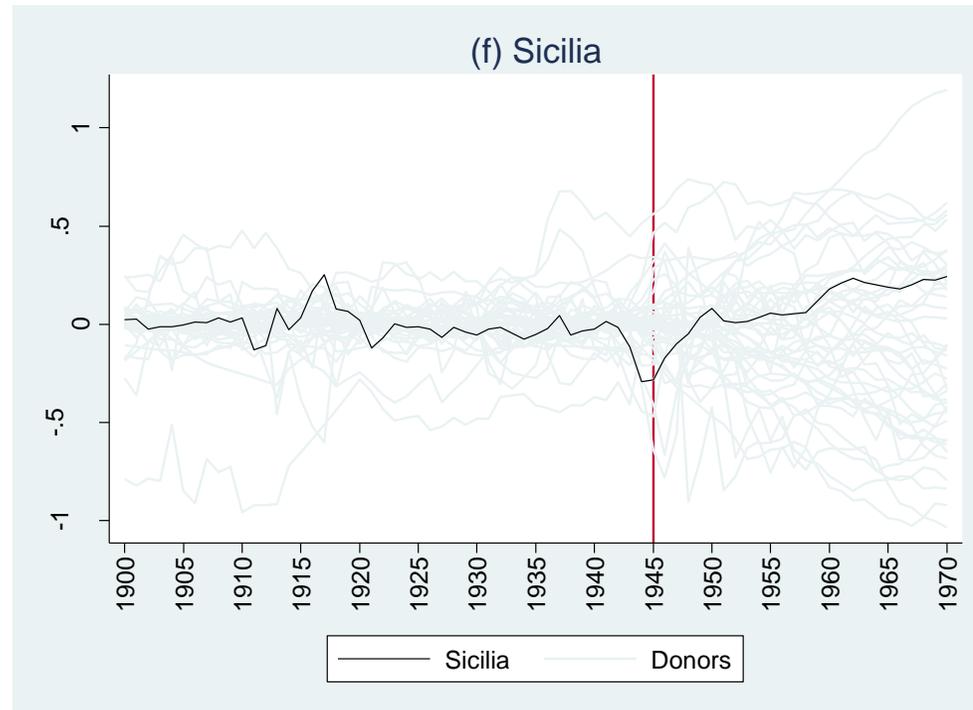
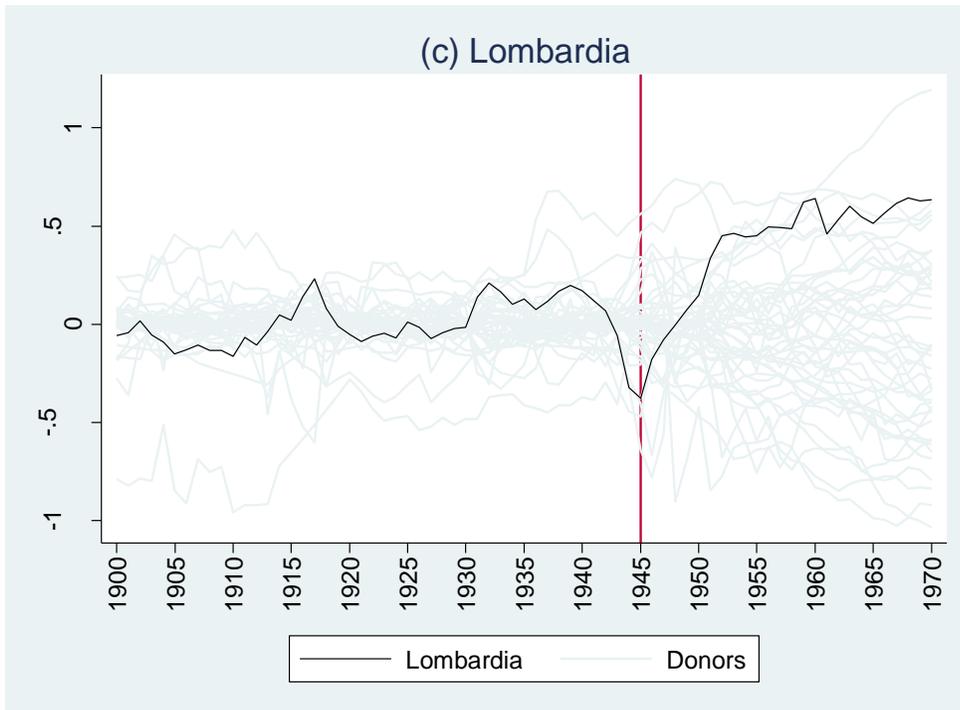
synthetic Sicilia



Synthetic Control Method: Applications in Economic History



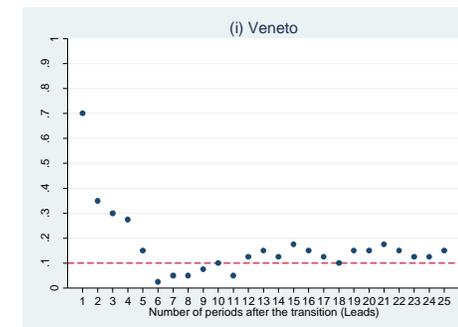
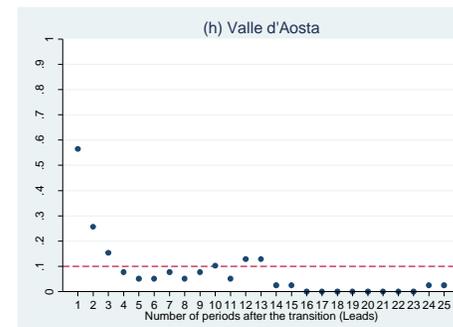
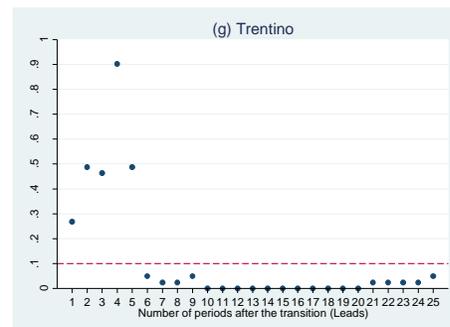
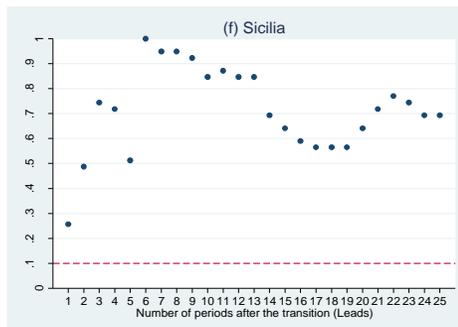
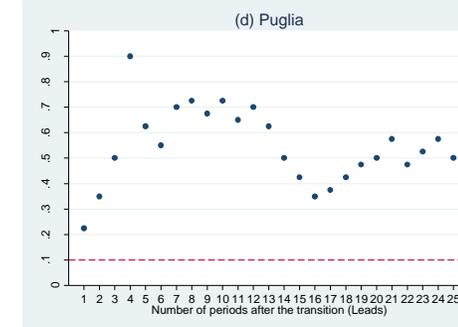
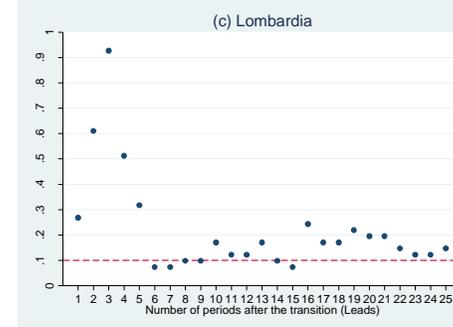
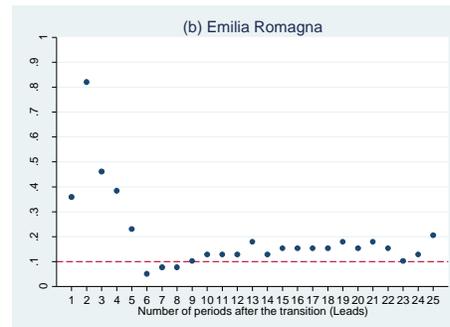
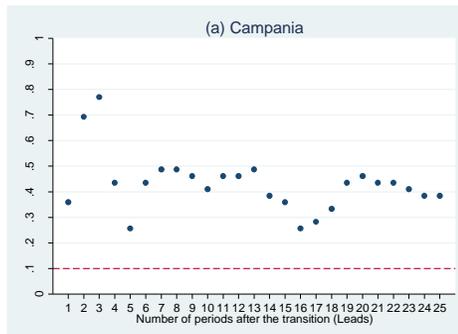
Synthetic Control Method: Applications in Economic History



- The theoretical argument on the social capital-moderated impact of institutional transition further supported by an extensive battery of placebo checks

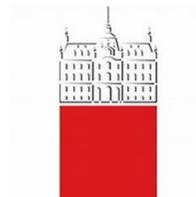
Synthetic Control Method: Applications in Economic History

Permutation-based randomization inference on the effect of institutional transition with 10,000 random donor sample size (as advocated by [Campos et. al. 2019](#))



Synthetic Control Method: Applications in Economic History

- Effect of institutional integration on the economic growth in Europe ([Campos, Coricelli and Moretti 2019, JME](#))
- Growth benefits of EU membership for non-founding states using synthetic control estimator ([Abadie et. al. 2015](#))
- On balance, they found heterogenous and positive effects of EU membership on growth which are robust across more than 10,000 random donor samples
- The only exception is Greece for which negative effects of EU membership are found



Synthetic Control Method: Applications in Economic History

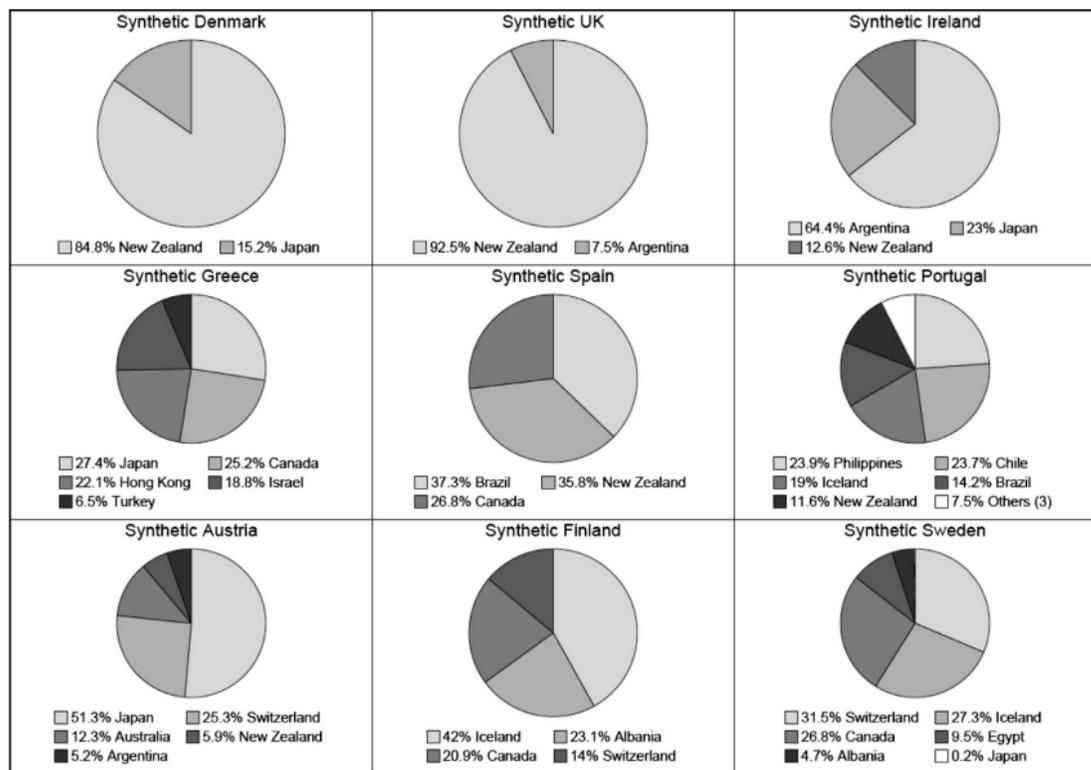


Fig. 1. Composition of the synthetic country – GDP per capita, northern and southern enlargements.

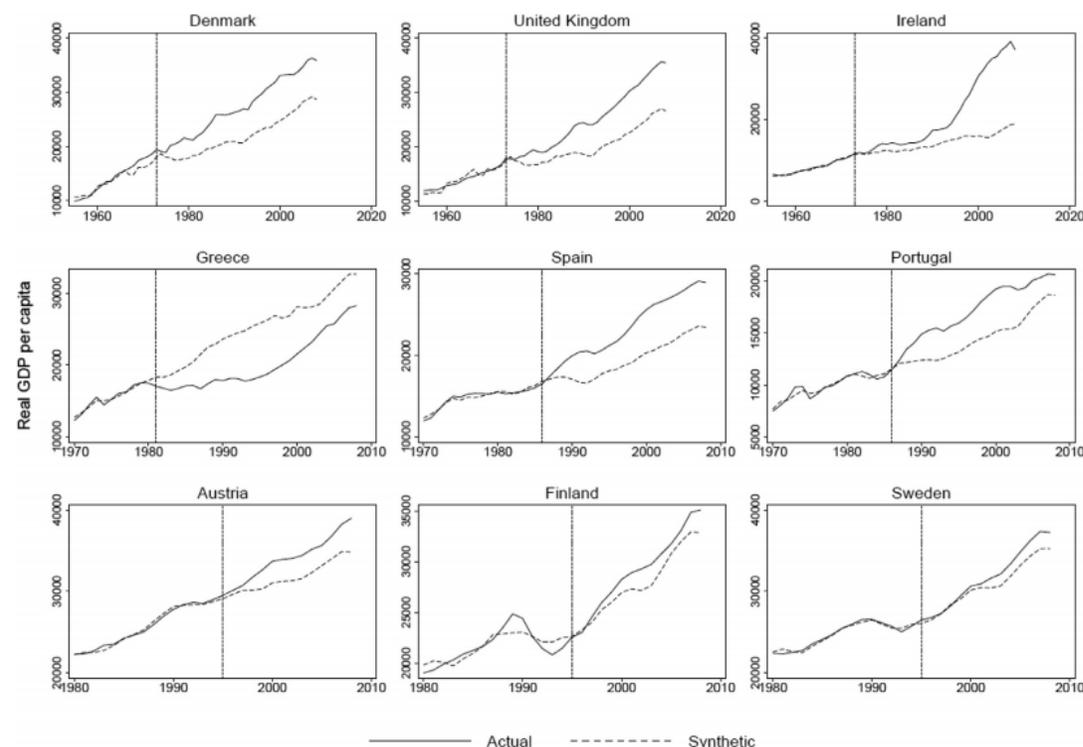


Fig. 2. Actual vs. synthetic – GDP per capita trends, northern and southern enlargements.



Synthetic Control Method: Applications in Economic History

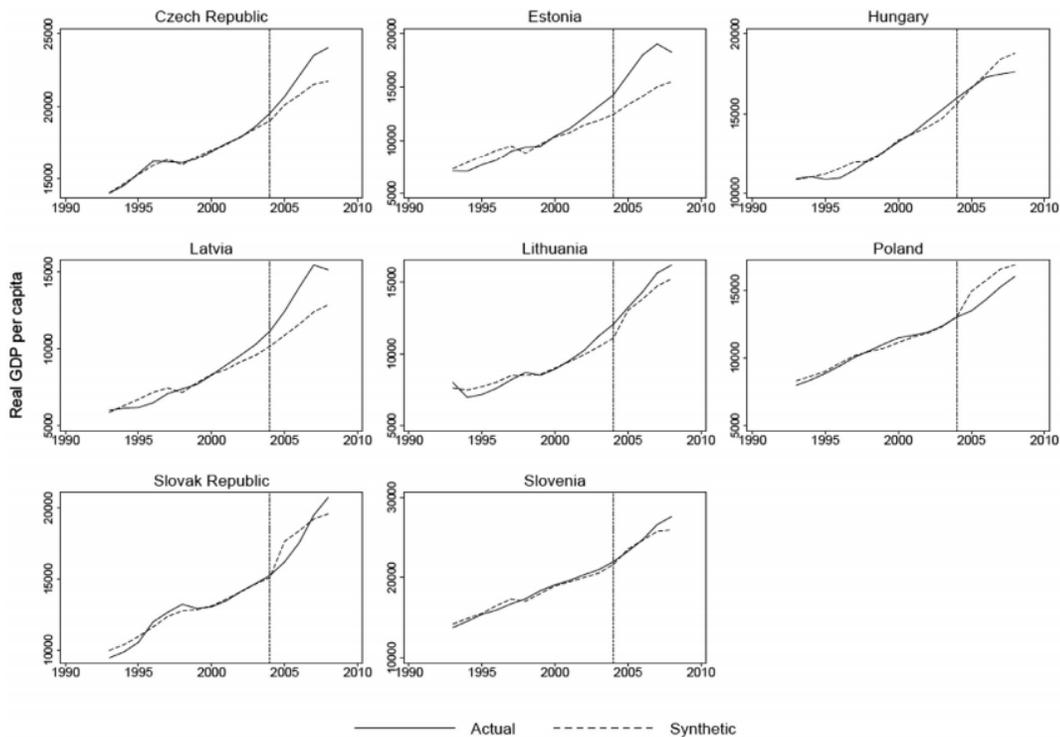


Fig. 3. Actual vs. synthetic country – GDP per capita trends, eastern enlargement (2004 accession).

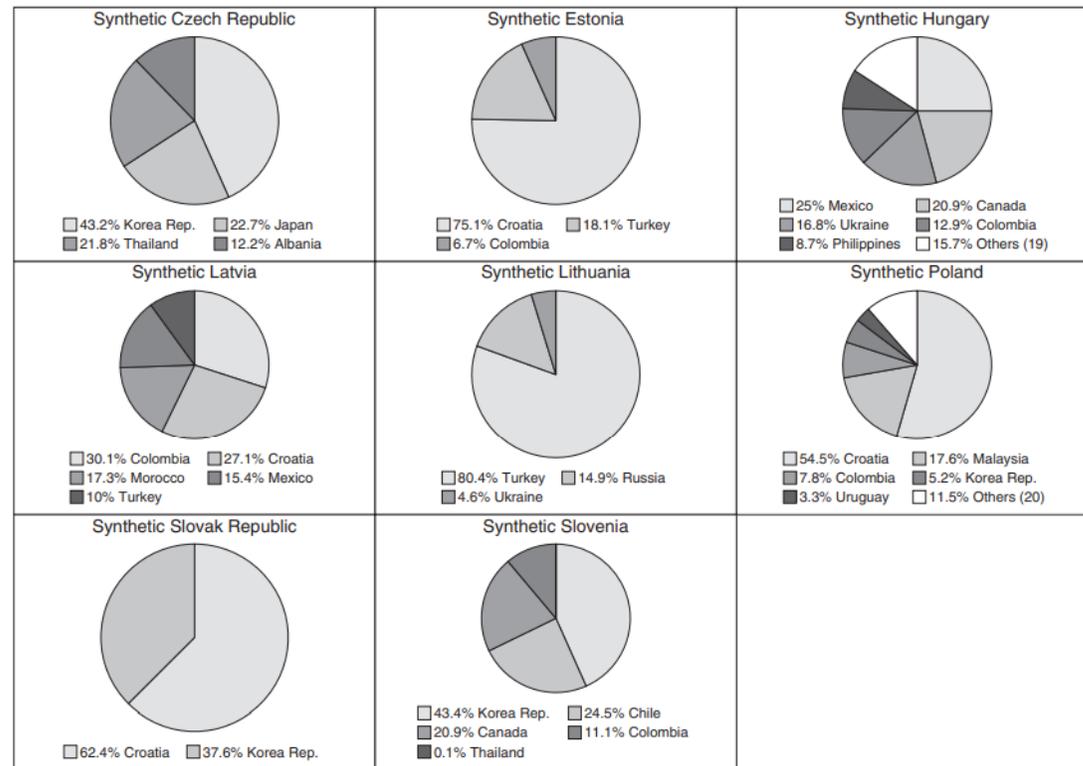


Fig. 4. Composition of the synthetic country – GDP per capita, eastern enlargement (1998 anticipation).



Synthetic Control Method: Applications in Economic History

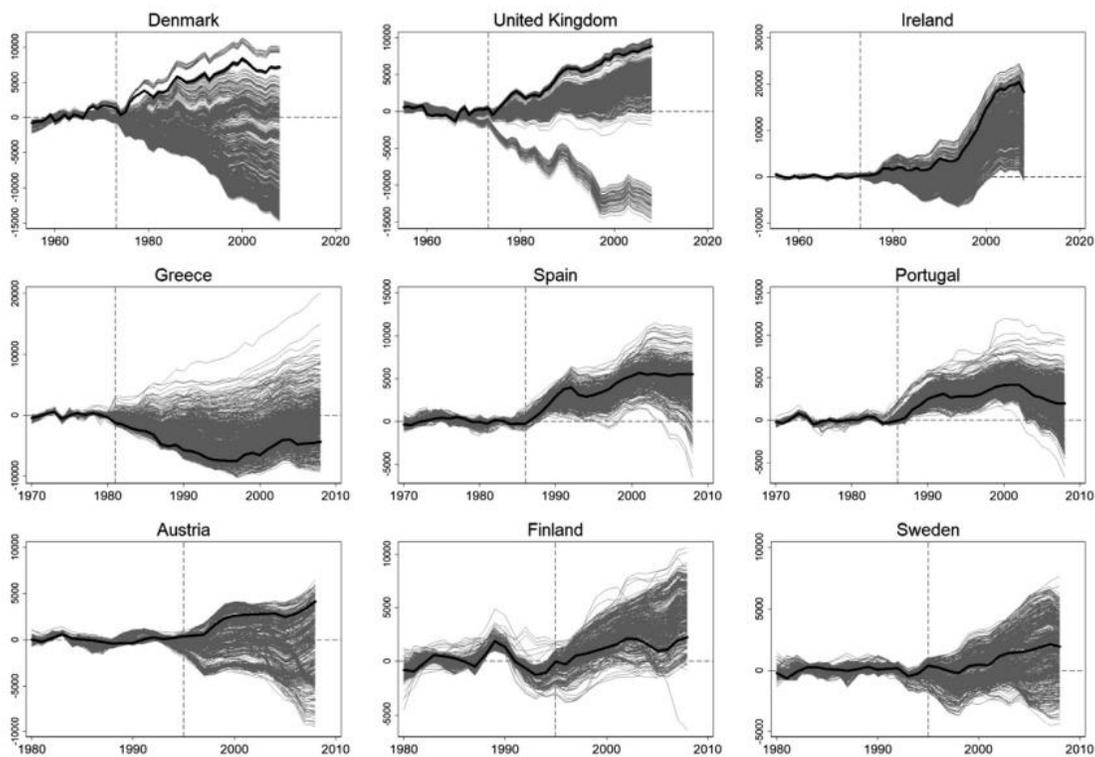


Fig. 6. Random donor samples (1,000 replications) – GDP per capita, northern and southern enlargements.

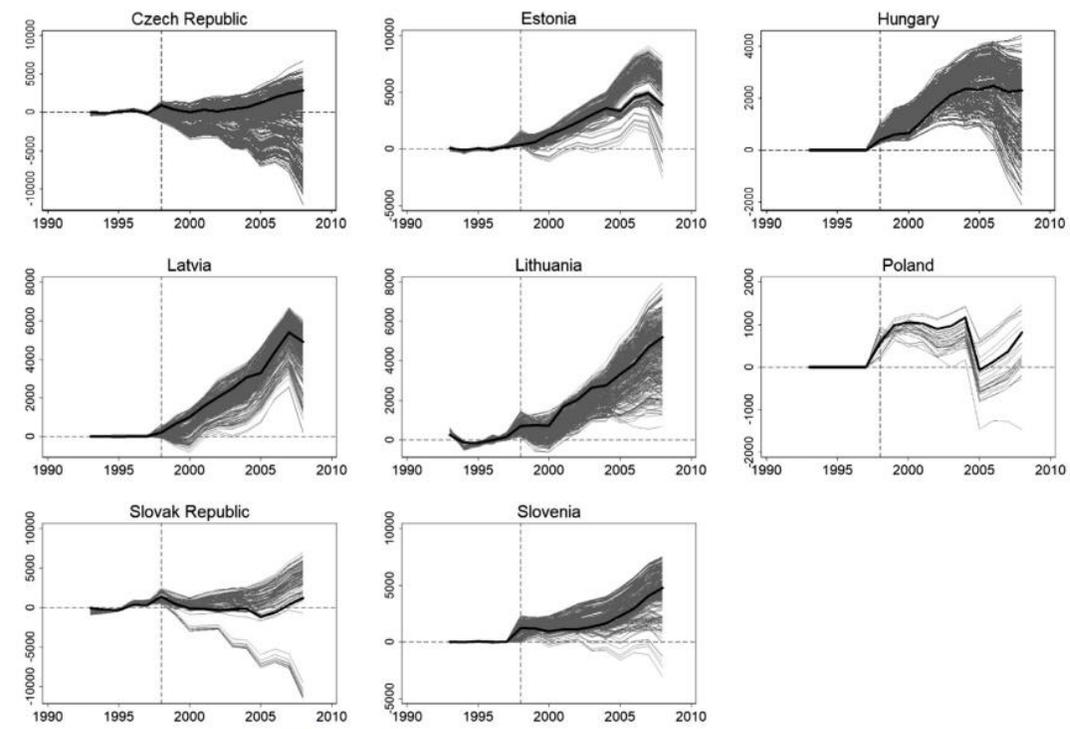
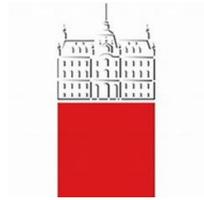


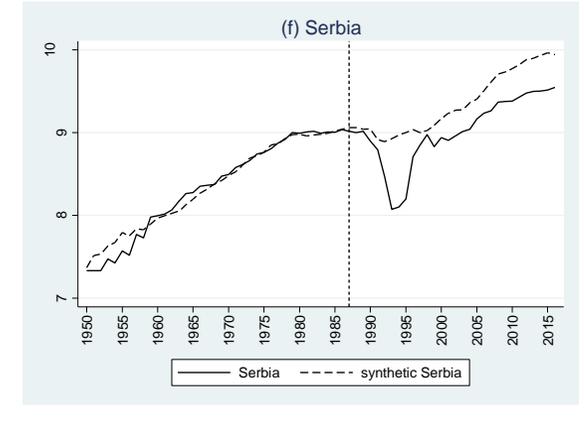
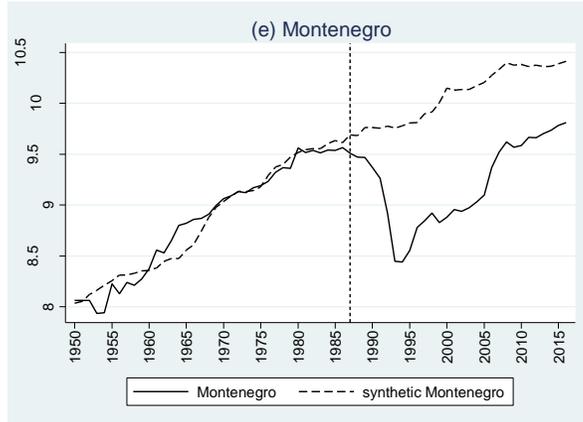
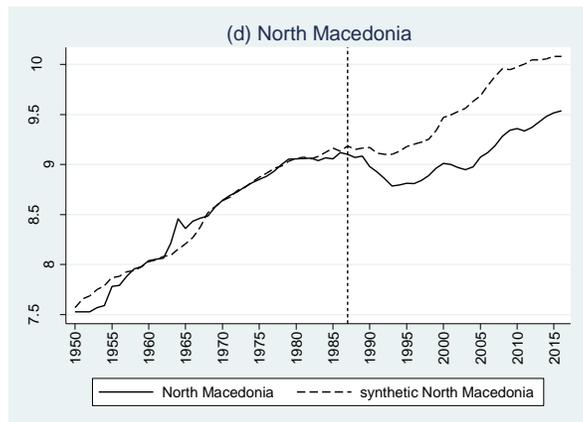
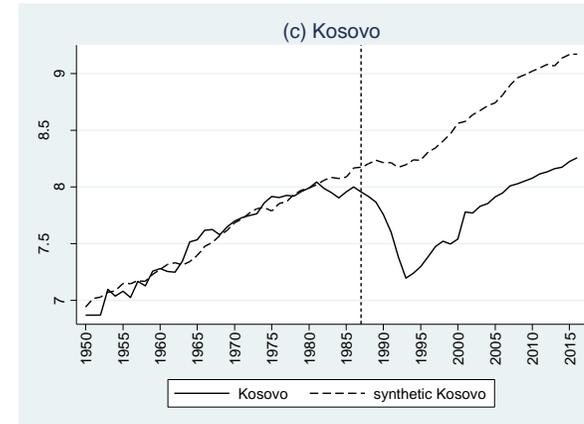
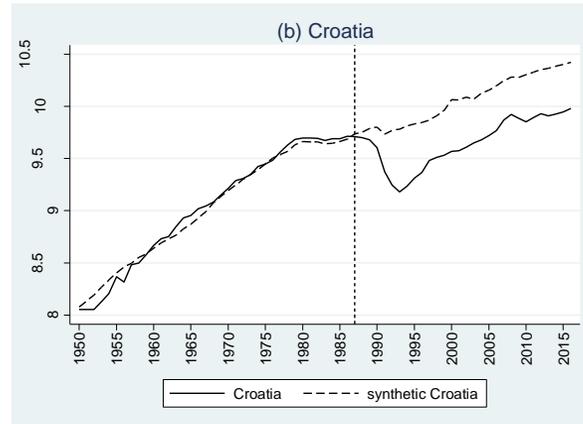
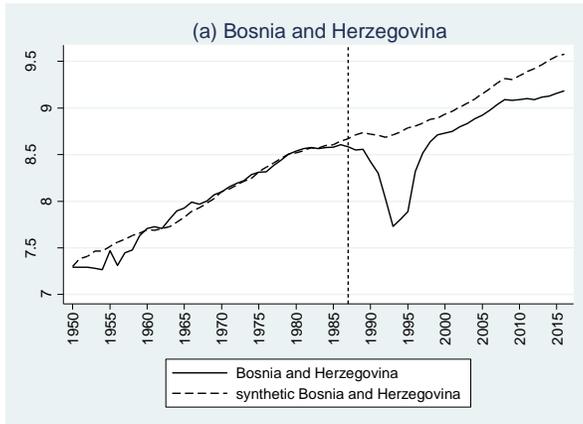
Fig. 7. Random donor samples (1000 replications) – GDP per capita, eastern enlargement (1998 anticipation).



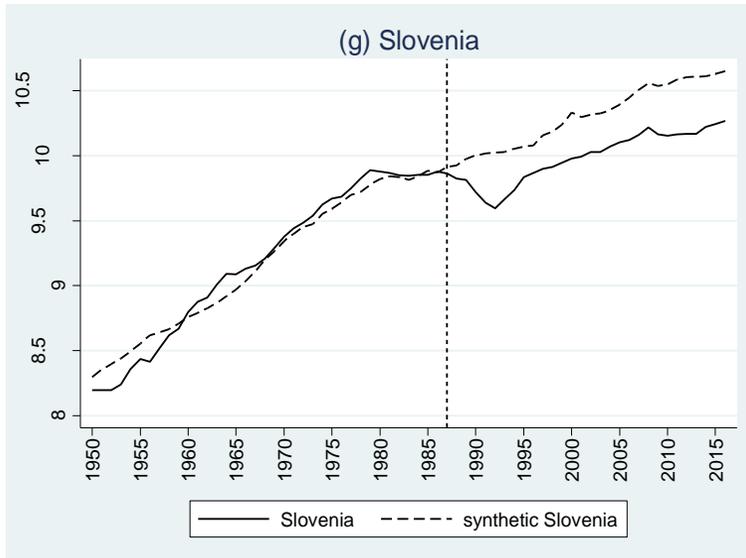
Synthetic Control Method: Applications in Economic History

- On the contrary, in late 1980s, former Yugoslavia degenerated into the state of institutional disintegration, ethnic violence and civil war
- Death toll around 130,000-140,000
- More than 4,000,000 people forcibly displaced
- Former Yugoslav republics became independent states by early 1990s
- The economic effects of civil war and the institutional disintegration remain unknown
- Synthetic control analysis of the growth effects of former Yugoslav civil war ([Spruk and Keseljevic 2020](#))

Synthetic Control Method: Applications in Economic History

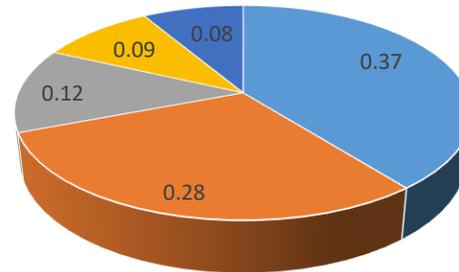


Synthetic Control Method: Applications in Economic History



synthetic Slovenia w/o Yugoslav war

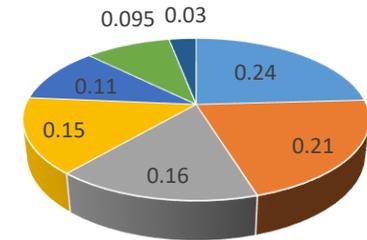
RMSE = 0.097



■ Switzerland ■ Malta ■ Iceland ■ Japan ■ Oman

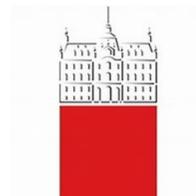
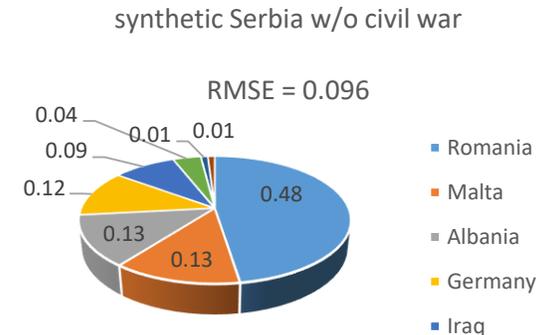
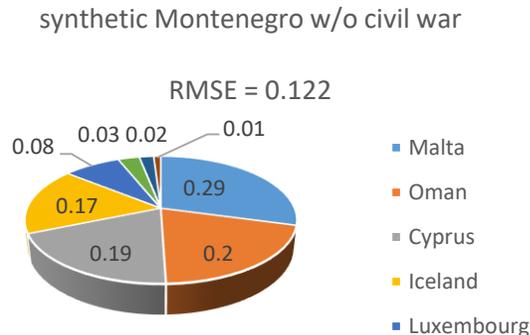
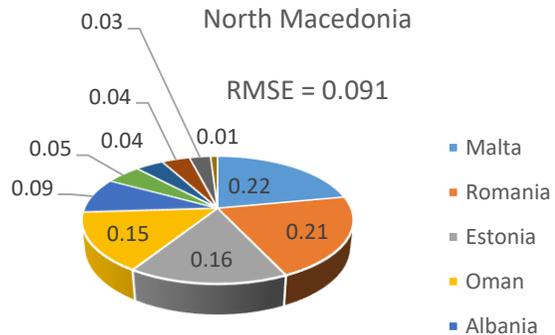
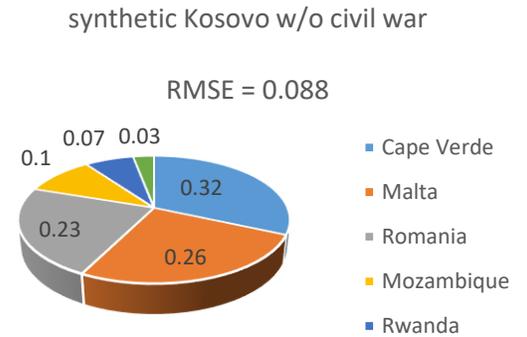
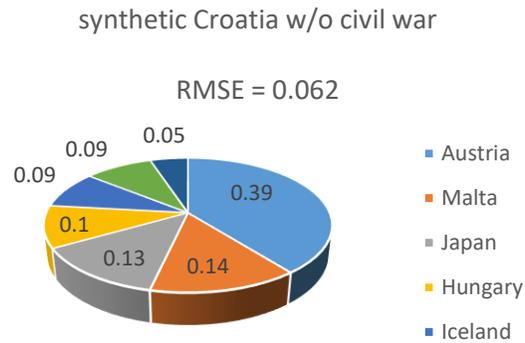
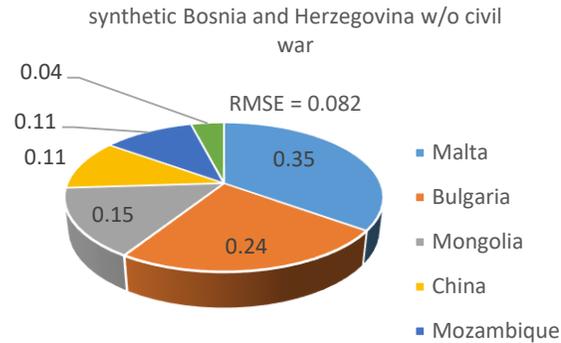
synthetic Slovenia w/o Yugoslav war
excluding Switzerland from control group

RMSE = 0.094



■ Malta ■ Japan ■ Luxembourg ■ Iceland
■ Saudi Arabia ■ Austria ■ Puerto Rico

Synthetic Control Method: Applications in Economic History



Synthetic Control Method: Applications in Economic History

- Selected works in progress:
 - Economic growth effects of joining the United States (with R. Maseland)
 - The effect of Franco's 1959 economic reforms on province-level economic growth in Spain (with A. Melcarne and J.S. Mora-Sanguinetti)
 - The impact of democratization on province-level growth in Spain (with A. Melcarne & J.S. Mora-Sanguinetti)
 - Subnational region-level analysis of growth effects of Yugoslav war (with A. Keseljevic)
 - Subnational region-level growth impact of EU membership (with N. Garoupa)

Synthetic Control Method and Western Australian Economic History

- An extensive potential to study the economic history of Western Australia with synthetic control techniques
- Joint research with [David Gilchrist](#), [Tom Emery](#) and [Nuno Garoupa](#)
- One such example is the election of Sir Charles Court as a state premier in 1975
- An activist industrial policy combined with the policymaking that upholds free-enterprise allocation of resources
- Several autocratic tendencies using executive discretion to overcome traditional government failures from public choice literature (i.e. inefficient policy-making, pork-barrel politics, unions' resistance to structural reforms) possibly led to lower transaction costs, greater international trade, better governance and higher growth

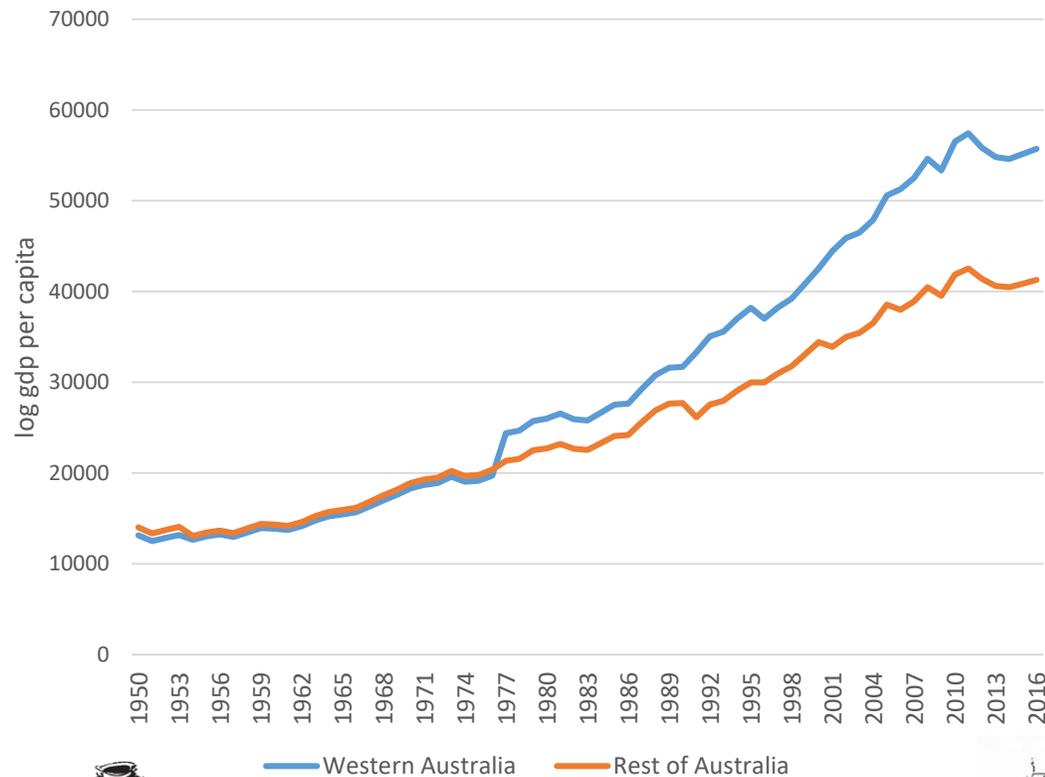
Synthetic Control Method and Western Australian Economic History

“Any nation that is allowed the requisite freedom will recover much more quickly than the one which is subject to severe controls.”

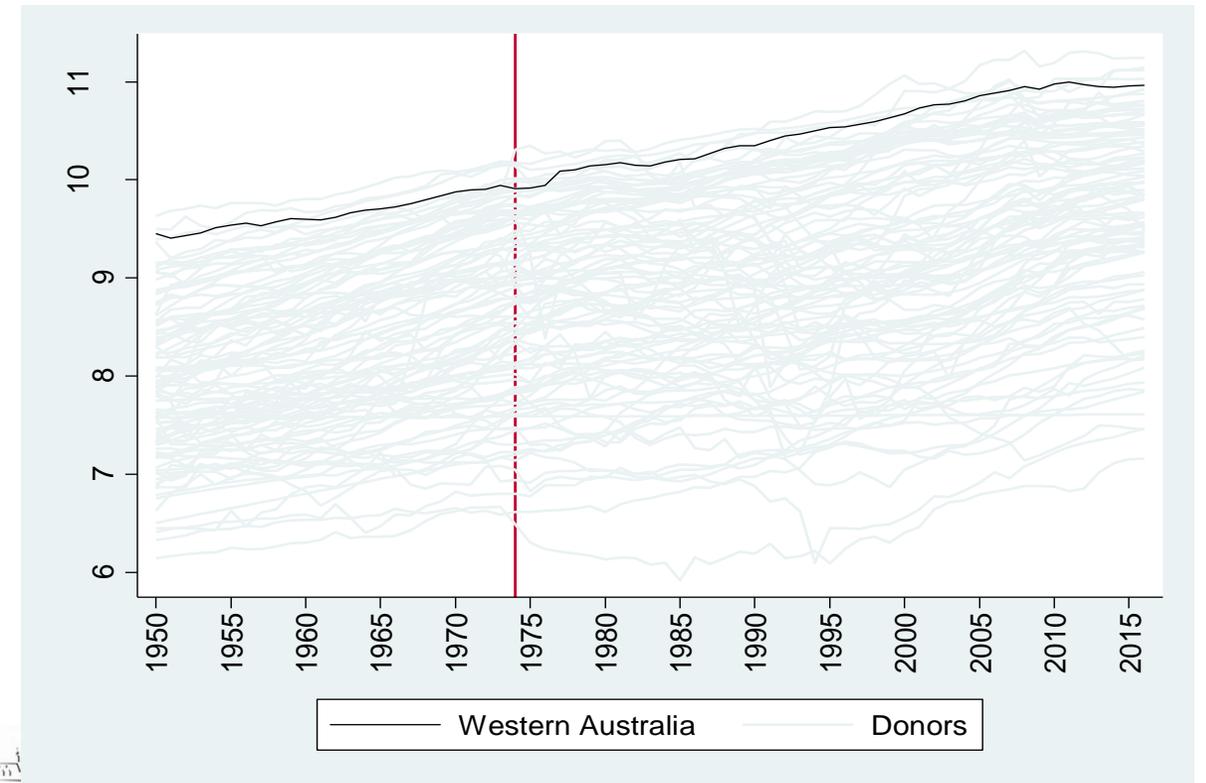
Sir Charles W.M. Court, Inaugural Speech

Synthetic Control Method and Western Australian Economic History

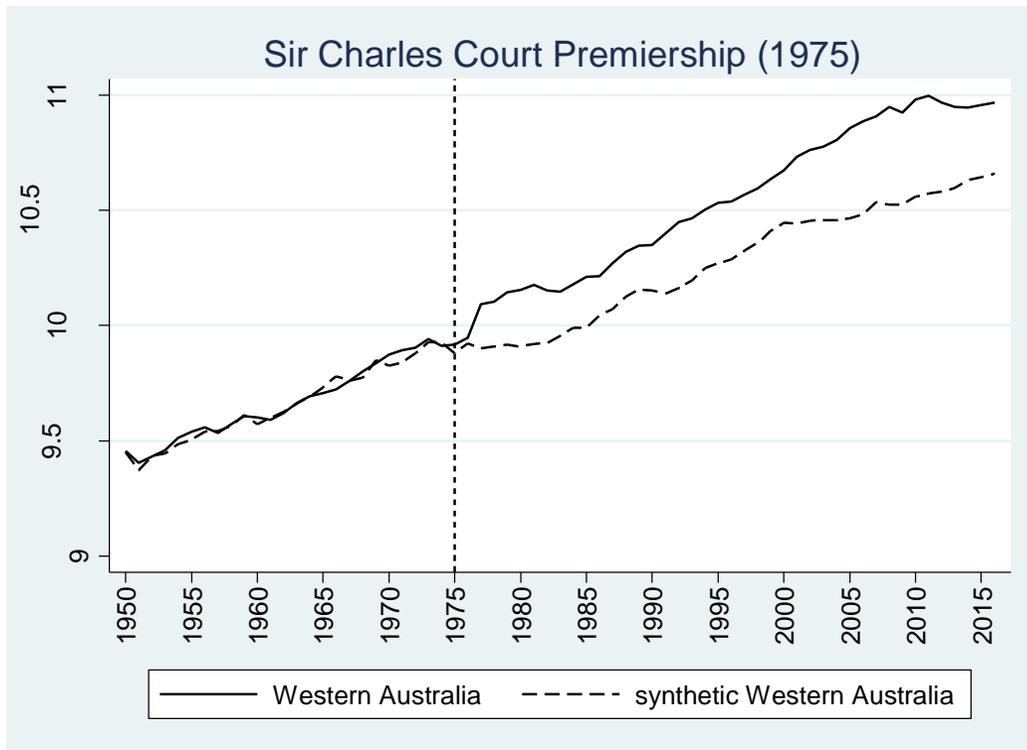
Western Australia vs. the rest of the country



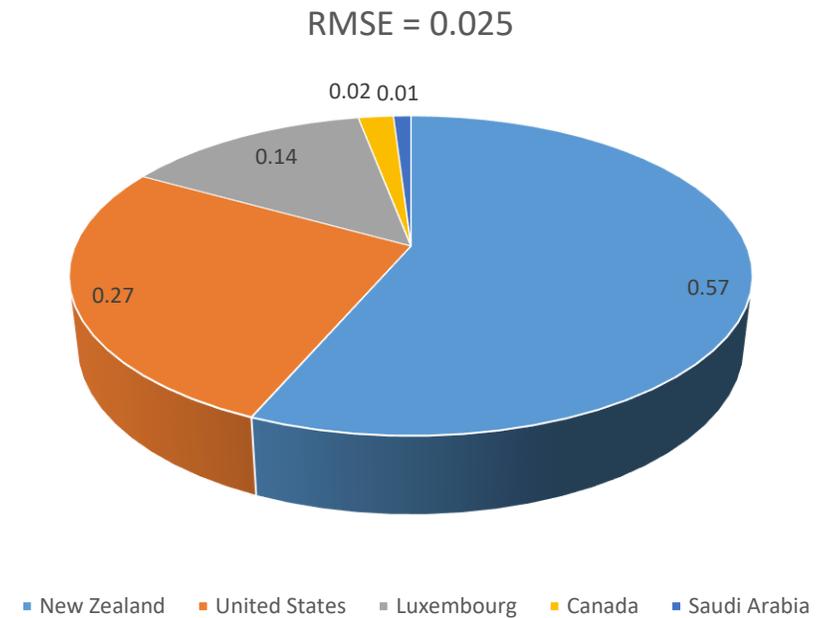
Western Australia vs. the rest of the world



Synthetic Control Method and Western Australian Economic History

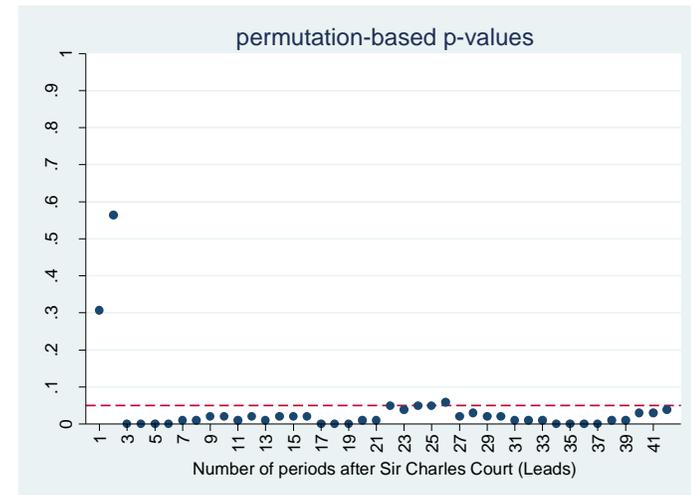
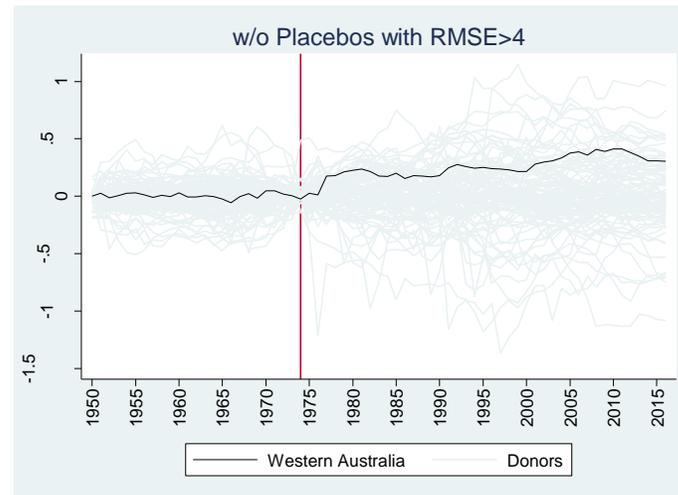
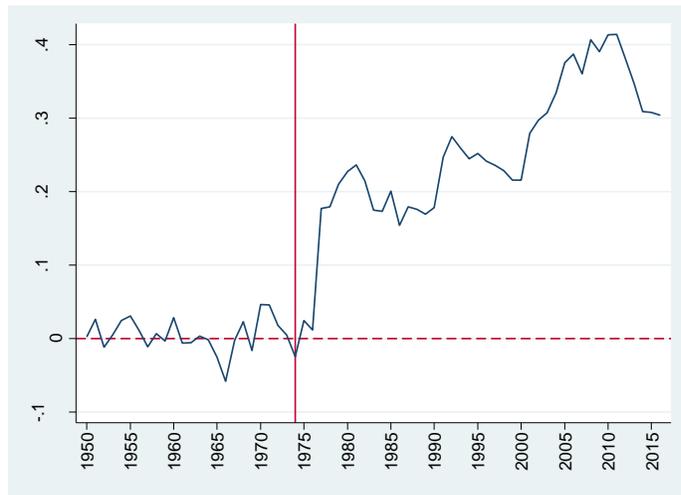


Synthetic Western Australia w/o Sir Charles Court Premiership



Synthetic Control Method and Western Australian Economic History

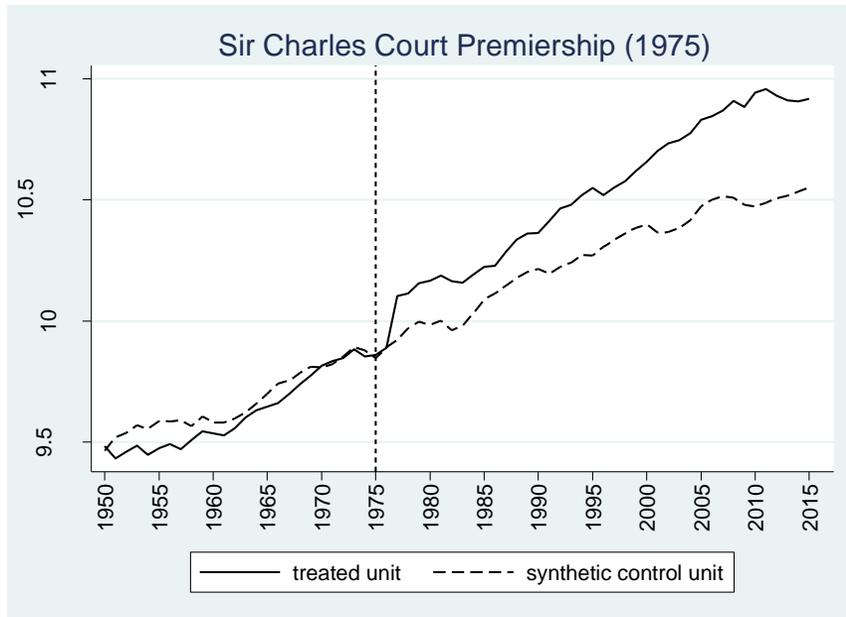
Inference on economic development impact of Sir Charles Court



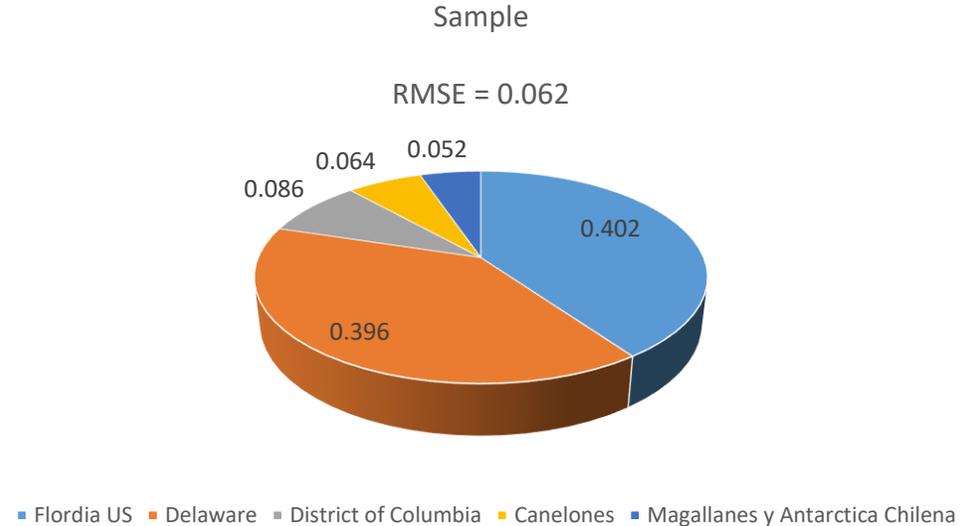
The per capita GDP gap between the actual Western Australia and its synthetic version w/o Sir Charles Court premiership is +36% which appears to be quite large and statistically significant at 5% down to the present day

Synthetic Control Method and Western Australian Economic History

- Using a control sample of 689 regions in other countries, the long-term growth and development impact of Sir Charles Court tends to be even larger



Sir Charles Court Premiership (1975) - Region-Level Control



The per capita GDP gap between the actual Western Australia and its synthetic version w/o Sir Charles Court premiership is +41% which appears to be quite large



Recent Developments in Synthetic Control Method

- Using inverted percentile ranks of the treatment effect to better understand the heterogeneous impact of policy changes of interest ([Dube and Zipperer 2015](#))
- Stationary models and extensive specification searching strategies to address covariate imbalance and avoid overfitting ([Pinto and Ferman 2018](#))
- Leveraging pre-treatment fit against the benchmark model ([Adhikari and Alm 2015](#))
- Augmented synthetic control with ridge regression to remove the imbalance in covariates that would otherwise lead to biases ([Ben Michael et. al. 2018](#))
- Strong forms of covariate choice to build synthetic control specifications using extreme bounds approach ([Garoupa and Spruk 2020](#))
- Differential trend assumption test in pre/post-treatment stage ([Spruk and Kovac 2020](#))
- Difference-in-difference parametric tests regarding the behavior of placebo gaps against the control sample ([Garoupa and Spruk 2020](#))

Conclusion

- Synthetic control analysis as a tool for estimating, detecting and understanding the counterfactual scenario in response to policy shocks, institutional changes or rare unexpected events
- Identification of the shock and the evolution of outcomes assuming the event of interest would have never happened
- Extensive in-space and in-time placebo inference with large samples of donors
- Few but a growing list of publications that use SCM
- A large potential to study Western Australian economic history with SCM

Contact Details

Name Rok Spruk

Email rok.spruk@ef.uni-lj.si

Telephone +386 31 596 033

Other (e.g. Twitter, LinkedIn, Facebook):

<https://twitter.com/rokspruk>

<https://sites.google.com/site/rokspruk/>

<https://www.facebook.com/rok.spruk>

Auxiliary Slides

Introduction

- Synthetic control method minimizes imbalance in pre-intervention outcomes and covariates
- The key limitation of synthetic control method: exact balancing between treated and control units almost impossible which may produce bias ([Ferman and Pinto 2018](#), [Ben Michael et. al. 2018](#))
- Bias correction for inexact matching ([Abadie and Imbens 2011](#), [Doudchenko and Imbens 2017](#))
- Outcome model to estimate the bias due to covariate imbalance ([Ben Michael et. al. 2018](#)) with inverse probability weighting (IPW) regularization of SCM using ridge penalty on propensity score coefficients → augmented SCM
- Augmented SCM: allow for the extrapolation outside the convex hull of the control units temporal observations to avoid zero probability of exact balancing when pre-intervention time period is large → negative weights obtained through ridge regression and IPW regularization

Introduction

- Use of outcome model with a large pre-intervention period (T0) to partially remove the bias arising from covariate imbalance and inexact matching ([Garoupa and Spruk 2019](#))
- Instead of imposing negative weights on the donor pool through calibrated propensity scores ([Hainmüller 2011](#), [Ben Michael et. al. 2018](#), [Wand and Zubizarreta 2018](#)), use of Newton-Raphson root-finding constrained optimization and interior point method (Vanderbei 1999) to remove the bias that arise from inexact matching
- Based on the lagged outcome model, the use of SCM on a large sample where pre-intervention period is large, is underexploited
- General estimates of the impact of institutional changes on long-run growth using SCM with randomization-based inference and empirical rejection probabilities for the null hypothesis ([Firpo and Possebom 2017](#))

Setup and Notation

- A finite set of units/countries/ regions $i = 1, 2, \dots, J$ which may be exposed to the set of independent interventions/policies/shocks
- Only the first unit is exposed to the shock while the remaining J unit in its neighborhood are the potential control units
- Let $\ln y_{i,t}^{No-Shock}$ be the economic outcome of interest for i -th unit in the absence of the intervention/shock at time t within the discrete time horizon $t=1, 2, \dots, T$ and let T_0 denote the number of years in the pre-shock period such that

$$1 \leq T_0 < T$$
- The economic outcome of interest in the presence of intervention/shock is $\ln y_{i,t}^{Shock}$
- The shock period lasts from T_0 to T



Setup and Notation

- **Assumption #1:** The intervention/shock in country i at time T_0 is independent of the economic outcome before the shock period so that for $t < T_0$ and $t \in \{1, \dots, T_0\}$ we have $i = \{1, 2, \dots, N\}$

$$\ln y_{i,t}^{\text{Shock}} \approx \ln y_{i,t}^{\text{No-Shock}}$$

- Shock-affected countries do not directly interfere and potentially interfering countries are excluded from the donor pool (Rosenbaum 2007, Abadie et. al. 2010)

- Let describe $\lambda_i = \ln y_{i,t}^{\text{Shock}} - \ln y_{i,t}^{\text{No-Shock}}$ the effect of the institutional shock for country i at time t , such that $\lambda_i = (\lambda_{i,T_0+1}, \dots, \lambda_{i,T})$ be a simple linear indicator function which takes the value of 1 if the i -th country is exposed to the institutional shock at time t and 0 otherwise

- The observed economic outcome for country i at time t is given by:

$$\ln y_{i,t} = \ln y_{i,t}^{\text{No-Shock}} + \lambda_{i,t} \cdot D_{i,t}$$

Setup and Notation

- **Goal:** estimate the effect of institutional shocks on the economic outcome of interest
- Counterfactual scenario of the absence of institutional shocks unobserved to the econometrician
- Only the first country from the finite set is affected by the institutional or policy shock strictly so with $1 \leq T_0 < T$, the distribution of institutional shock is described by the following set:

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

- Which implies that for $1 \leq t \leq T_0$, we estimate the sequence to construct the counterfactual level of the economic outcome without the institutional shock given by

$$\lambda_1 = (\lambda_{1,T_0+1}, \dots, \lambda_{1,T})$$

Setup and Notation

- Since $\ln y_{i,t}^{No-Shock}$ is unobserved, the counterfactual outcome has to be estimated ex-post
- The counterfactual outcome set is given by the latent factor model with unobserved components:

$$\ln y_{i,t}^{No-Shock} = \theta_t + \eta_t \cdot \mathbf{Z}_i + \delta_t \cdot \mathbf{M}_i + \varepsilon_{i,t}$$

- θ_t – unknown common factor of the quasi-treated country matched with control countries with constant linear functions
 - η_t is an $r \times 1$ vector of unobserved covariates unaffected by the institutional shock
 - \mathbf{Z}_i is an $1 \times r$ vector of unknown parameters
 - η_t is an $1 \times F$ vector of unobserved common factors
 - $\delta_t \in \mathbb{R}^{F \times 1}$ vector of unknown factor loadings
 - $\mathbf{M}_i \in \mathbb{R}^r$
- Zero conditional mean independence assumption on the behavior of transitory shocks such that they exhibit a zero covariance moment restriction

Setup and Notation

- Under Assumption #1, heterogeneous responses to multiple unobserved factors are possible by embedding time trends into the counterfactual economic outcome
- **The basic idea:** reweigh the control group in the donor pool so that a synthetic country (=artificial control unit) is set to match on a subset of pre-shock covariate-level characteristics of the affected country
- By default, $M_i \in \mathbb{R}$ is automatically matched with the synthetic counterfactual process
- Matching pre-shock characteristics of the affected country to its similar counterparts unaffected by the institutional shock at time T_0 ensures that the unobserved country-level heterogeneity bias is not projected out of the counterfactual model specification



Setup and Notation

- Suppose that the donor pool comprises $J-1$ countries excluding the affected ones
- Let $J \times 1$ be the vector of weights $\mathbf{W} = (w_2, w_3, \dots, w_{J+1})$ such that for $w_j \geq 0$ and $2, \dots, J+1$ $w_2 + \dots + w_{J+1} = 1$
- Each particular value of \mathbf{W} represents the potential synthetic control
- The synthetic control is a weighted average of control countries sharing similar pre-shock characteristics captured by $\mathbf{Z}_j \in \mathbb{R}^r$
- The reweighted economic outcome of interest from the artificial control unit, indexed by \mathbf{W} , is given by:

$$\sum_{j=2}^{J+1} w_j \cdot \ln y_{jt} = \theta_t \cdot \sum_{j=2}^{J+1} w_j \cdot \mathbf{Z}_j + \delta_t \cdot \sum_{j=2}^{J+1} w_j \cdot \mathbf{M}_j + \sum_{j=2}^{J+1} w_j \cdot \varepsilon_{jt}$$

where the vector of weights $(w_2^*, \dots, w_{J+1}^*)$ such that

$$\sum_{j=2}^{J+1} w_j^* \cdot \ln y_{j,1} = Y_{11}, \quad \sum_{j=2}^{J+1} w_j^* \cdot \ln y_{j,2} = Y_{12} \quad \text{and} \quad \sum_{j=2}^{J+1} w_j^* \cdot \ln y_{j,T_0} = Y_{1T_0}$$

$$\sum_{j=2}^{J+1} w_j^* \cdot \mathbf{Z}_j = \mathbf{Z}_1$$

Setup and Notation

- Theorem #1:** Suppose $\exists \mathbf{W}^*$ such that the artificial/synthetic country matches up the quasi-treated country in the pre-shock period, $\sum_{j=2}^{J+1} w_j^* \cdot \ln y_{j,t} = \ln y_{1,t}$ and $t \in \{1, \dots, T_0\}$ where $\sum_{j=2}^{J+1} w_j^* \neq 0$. For all $t > T_0$, the difference between the synthetic counterfactual and actual economic outcome, $\varepsilon_{i,t}$, disappears as $T_0 \rightarrow \infty$ or it is large relative to $\ln y_{1,t}^{No-Shock} - \sum_{j=2}^{J+1} w_j \ln y_{j,t}$.

- Under standard moment restrictions, the synthetic control unit with \mathbf{W} replicates the missing counterfactual

- The approximately unbiased estimator of $\lambda_{1,t}$ given by:

$$\hat{\lambda}_{1,t} = \ln y_{1,t} - \sum_{j=2}^{J+1} w_j^* \cdot \ln y_{j,t}$$



Setup and Notation

- For $t \in \{T_0 + 1, \dots, T\}$ as an estimator of τ_t , the set of counterfactual outcomes weighted by the additive weights from the donor pool can hold only if τ_t (inside the convex hull of $\{(\ln y_{11}, \dots, \ln y_{1,T_0}, \mathbf{z}'_1), \dots, (\ln y_{J+1,1}, \dots, \ln y_{J+1,T_0}, \mathbf{z}'_{J+1})\}$)
- If the standard conditions for the missing counterfactual hold, the synthetic control group (=artificial control unit) is constructed as a weighted combination of the unaffected countries such that Theorem #1 holds
- This allows us to match pre-shock covariate-level characteristics from the DGP for the outcome variable
- The fit of the synthetic counterfactual outcome may be poor if the interpolation biases of the constant linear model are large relative to the sample size
- Remove observations with pre-shock root MSE greater than $\sqrt{3}$ (Abadie et. al. 2010, Cavallo et. al. 2013, Acemoglu et. al. 2016)

Two Standard Assumptions

- Assumption #2 (Parallel Trends):** The economic outcome in the quasi-treated country prior to the institutional shock follows the parallel trend in the unaffected countries. The identification if the shock effect is possible as long as

$$E(\ln y_{i,t}^{t-T_0} - \ln y_{j,t}^{t-T_0} | D_{i,t} = 1) = E(\ln y_{i,t}^{t-T_0} - \ln y_{j,t}^{t-T_0} | D_{i,t} = 0)$$
 holds for any composition of the donor pool

- Assumption #3 (Independence of Institutional Shock Conditional on Past Outcomes):** Similar economic outcome is anticipated in the pre-institutional shock is anticipated so that the outcome without the treatment imposed by institutional shock is similar to the outcome path in the affected country is similar to the outcome path in the post-shock period which implies that

$$\ln y_{i,t < T_0}^{No\ Shock} \perp (\mathbf{X}_{i,t}, \ln y_{i,t > T_0}^{Shock})$$



Constructing the Synthetic Control Group

- A vector of non-negative and additive weights \mathbf{W} from the donor pool represents a weighted average of the control countries and serves as a potential synthetic control for the country affected by the institutional shock
- Covariate-level characteristic matching between the quasi-treated and control countries to minimize the covariate-implied treatment/control differences subject to 10% error bound restriction
- Using a semi-positive $r \times r$ V matrix, treatment/control distance minimization is based on the [Abadie et. al. \(2010\)](#) which implies that:

$$\|\mathbf{X}_{1,j} - \mathbf{X}_{0,i} \mathbf{W}\|_V = \sqrt{(\mathbf{X}_{1,j} - \mathbf{X}_{0,i} \mathbf{W})' \mathbf{V} (\mathbf{X}_{1,j} - \mathbf{X}_{0,i} \mathbf{W})}$$

Inference

- Assume that the effect of intervention in the full post-treatment period is described by $\hat{\lambda}_{1t}$
- The distribution of in-place placebo effects is $\hat{\lambda}_{1t}^{Placebo} = \{ \hat{\lambda}_{jt} : j \neq 1 \}$
- We compute the two-tailed p-value for the effect of the intervention as:

$$p\text{-value} = \Pr \left(\left| \hat{\lambda}_{1t}^{Placebo} \right| \geq \left| \hat{\lambda}_{1t} \right| \right) = \frac{\sum_{j \neq 1} 1 \cdot \mathbb{1} \left(\left| \hat{\lambda}_{jt} \right| \geq \left| \hat{\lambda}_{1t} \right| \right)}{J}$$

Inference

- We are able to adjust the set of placebo coefficients for the quality of the pre-shock match in two steps:
 - 1) Restrict the set of synthetic control effects to include those that match well. This implies that the five-fold multiples of placebo effects are removed ([Abadie et. al. 2010](#))
 - 2) The placebo effects are divided by the pre-shock match quality parameter to obtain the distribution of pseudo t-statistics and compute the relevant p-values