House of Debt Jubilee:

Iceland After the 2008 Financial Crisis

Tiger Gao

Advised by Professor Atif Mian

Princeton University

Submitted to Princeton University
Department of Economics
In Partial Fulfillment of the Requirement for the A.B. Degree

April 2021

Abstract

Large-scale debt write-downs are understood as helpful for economies to recover more quickly from financial shocks, but very few have been implemented in practice. Iceland was particularly hard hit by the 2008 financial crisis due to its high banking sector concentration and dramatic pre-crisis credit boom. In response, Iceland implemented a generous debt jubilee in 2011 amounting to over 10% of its GDP. I use synthetic controls method (SCM) to investigate the impact of this program. I find it to result in substantial improvements for macroeconomic and financial conditions. Within 5 years of the policy treatment, Iceland's unemployment rate, non-performing loan ratio, private credit to GDP ratio, and international debt to GDP ratio were each reduced by around 2, 15, 50, and 200 more percentage points than what they would have been in the absence of the debt jubilee. I demonstrate the significance of my estimates through a variety of inferential methods and robustness checks, such as placebo studies, the cross-validation technique, and synthetic difference-in-differences (SDID).

Keywords: Debt Forgiveness, Risk Sharing, Financial Crises, Macrofinance, Synthetic Controls, Synthetic Difference in Differences

JEL Codes: E60, E65, G18, G21, G51, I38

Acknowledgement

I would like to thank my advisor Professor Atif Mian for patiently guiding me through this project. His deep expertise in macrofinancial issues led me to areas that I otherwise would not explore, and his support and words of wisdom in my moments of struggle have pushed me to be the best version of myself. The title of this thesis pays tribute to his book *House of Debt* and the important line of work of macrofinance research.

Karsten Müller and Lukas Althoff acted as my informal advisors with their detailed guidance over the months of research. My econometrics professors Chris Sims and Mikkel Plagborg-Moller provided valuable feedback that broadened my thinking on the econometric frameworks for this paper. My summer research advisor Owen Zidar and junior independent research advisor Iqbal Zaidi both played pivotal roles in helping me develop research skills and a passion for economics.

Princeton University data scientist Oscar Torres-Reyna and librarian Bobray Bordelon were instrumental in helping me with statistical applications and data collection. This thesis would not have come together without their gracious help.

I also express my sincere gratitude to the following scholars for their communications with me to provide feedback, guidance, and inspiration: Patrizia Baudino, Tom Bearpark, Matt Cocci, Jack Edmondson, Greg Lewis, Eugenia Menaguale, Sungho Park, Albrecht Ritschl, and Sebastian Toft Roelsgaard.

My passion for economics research cannot be separated from my extracurricular involvement with my podcast *Policy Punchline*. My guests have often inspired me to think more broadly and deeply about the urgent issues in our world, and this thesis topic in part stems from some of these discussions. My colleagues on the podcast team and my close friends have always acted as a crucial source inspiration of inspiration.

I thank everyone who has graciously supported me through my undergraduate career. I am deeply grateful.

To my parents, for everything

Contents

1 Introduction **1**

2 Background **2**
2.1 The 2011 Icelandic Debt Jubilee . . . 2
2.2 The Rise and Fall of the Icelandic Banking Sector . . . 3
2.3 Debt Restructuring and the Resurrection of Household Finance . . . 4
2.4 Criticism for the Debt Jubilee . . . 4

3 Literature Review **6**
3.1 Debt Forgiveness and Risk Sharing . . . 6
3.2 *Ex-ante* vs. *Ex-post* Risk Sharing . . . 7
3.3 Trade-offs between Micro- and Macro-level Studies on Debt Relief . . . 8

4 Data **9**
4.1 Overview and Sources . . . 9
4.2 Summary Statistics and Aggregate Trends . . . 10

5 Methodology **12**
5.1 Synthetic Controls Methods (SCM) . . . 12
5.1.1 Motivation and Background . . . 12
5.1.2 Econometric Setup for SCM . . . 13
5.1.3 Estimation . . . 14
5.1.4 Inference and Placebo Tests . . . 15
5.2 Cross-validation in SCM . . . 17
5.2.1 How to choose the V matrix? . . . 17
5.3 Synthetic Difference in Differences (SDID) . . . 19
5.3.1 Motivation: Comparison between DiD, SCM, and SDID . . . 19
5.3.2 SDID Methodology . . . 21
5.3.3 SDID as Robustness Check . . . 23

6 Empirical Strategy **23**
6.1 Variables of Interest . . . 23
6.2 Optimization Period Selection . . . 25
6.3 Relationship between Main Results and Robustness Checks . . . 26

7 Results **26**
7.1 Summary of Results . . . 26
7.2 Unemployment . . . 28
7.3 GDP per capita . . . 30
7.4 Non-Performing Loans (NPL) . . . 32
7.5 Private Credit . . . 34
7.6 International Debt . . . 36
7.7 Income Inequality . . . 38

8 Conclusion **40**

Appendices **45**

A. Summary Statistics **45**
1. Data Definition and Sources . . . 45
2. Summary Statistics Table . . . 48
3. Aggregate Trends Graphs . . . 49
   i. Macroeconomic outcomes . . . 49
   ii. Financial sector outcomes . . . 54
   iii. Systemic crisis outcomes . . . 55
4. Extension into Micro-level Data . . . 56

B. Robustness Checks with Cross-Validation and SDID **58**
1. Unemployment . . . 60
   i. p-values for Synthetic Controls Estimation . . . 60
   ii. Synthetic Controls with Cross-Validation . . . 60
   iii. Synthetic Difference-in-Differences (SDID) . . . 62
2. GDP per capita . . . 64
   i. p-values for Synthetic Controls Estimation . . . 64
   ii. Synthetic Controls with Cross-Validation . . . 65
   iii. Synthetic Difference-in-Differences (SDID) . . . 67
3. NPL . . . 69
   i. p-values for Synthetic Controls Estimation . . . 69
   ii. Synthetic Controls with Cross-Validation . . . 69
   iii. Synthetic Difference-in-Differences (SDID) . . . 71
4. Private credit . . . 73
   i. p-values for Synthetic Controls Estimation . . . 73
   ii. Synthetic Controls with Cross-Validation . . . 73
   iii. Synthetic Difference-in-Differences (SDID) . . . 75
5. International debt to GDP ratio . . . 77
   i. p-values for Synthetic Controls Estimation . . . 77
   ii. Synthetic Controls with Cross-Validation . . . 77
   iii. Synthetic Difference-in-Differences (SDID) . . . 79
6. Income inequality . . . 81
   i. p-values for Synthetic Controls Estimation . . . 81
   ii. Synthetic Controls with Cross-Validation . . . 81
   iii. Synthetic Difference-in-Differences (SDID) . . . 83

C. Discussion and Extensions **87**
1. Challenges of Applying Synthetic Controls to Iceland . . . 87
2. Why do the W weights have to sum up to 1 and be non-negative? . . . 88
3. Other Frameworks Under Synthetic Controls Methods . . . 89
4. Was the Icelandic Debt Jubilee an MMT Application? . . . 90

List of Figures

1 Synthetic controls trajectory for unemployment rate: Iceland vs. synthetic Iceland 28
2 Synthetic controls trajectory for GDP per capita: Iceland vs. synthetic Iceland 30
3 Synthetic controls trajectory for NPL: Iceland vs. synthetic Iceland 32
4 Synthetic controls trajectory for private credit: Iceland vs. synthetic Iceland 34
5 Synthetic controls trajectory for international debt to GDP ratio: Iceland vs. synthetic Iceland 36
6 Synthetic controls trajectory for top 10% income share ratio: Iceland vs. synthetic Iceland . . 38
7 Synthetic controls trajectory for middle 40% income share ratio: Iceland vs. synthetic Iceland 38
8 Synthetic controls trajectory for bottom 50% income share ratio: Iceland vs. synthetic Iceland 39
A1 GDP per capita trend: Iceland vs. OECD 49
A2 Unemployment rate trend: Iceland vs. OECD 49
A3 Labor market outcomes: Iceland vs. OECD average 50
A4 NPL ratio trend: Iceland vs. OECD 50
A5 Private credit trend: Iceland vs. OECD 50
A6 International private debt trend: Iceland vs. OECD 51
A7 Domestic debt to GDP ratio trend: Iceland vs. OECD average 51
A8 Household and corporate debt for Iceland and selected OECD countries 51
A9 Household and corporate debt to GDP ratio (%) for Iceland 2005Q1 - 2018Q4 52
A10 Iceland household debt write-offs after 2008 52
A11 Government and household consumption expenditure (% of GDP) 53
A12 Trends of Icelandic housing indices 2000-2020 53
A13 Bank concentration and profitability trend: Iceland vs. OECD average 54
A14 Bank deposit and regulatory capital trend: Iceland vs. OECD average 54
A15 Nonfinancial asset investments and net loans: Iceland vs. OECD average 54
A16 Financial cycles for Iceland and its trading partners 2000-2020 55
A17 NPL ratio by systemic banking crisis and IMF intervention 55
A18 NPL ratio by deposit freeze and insurance level 55
A19 NPL ratio by liquidity support level and loss treatment 56
A20 NPL ratio by nationalization and recapitalization of banks 56
B21 *p*-values for synthetic controls estimation of unemployment rate 60
B22 Cross-validation results for unemployment rate 61
B23 Estimation trajectories of SDID vs. DiD vs. SCM: unemployment rate 63
B24 *p*-values for synthetic controls estimation of GDP per capita 64
B25 Cross-validation results for GDP per capita 66
B26 Estimation trajectories of SDID vs. DiD vs. SCM: GDP per capita 68
B27 *p*-values for synthetic controls estimation of NPL ratio 69
B28 Cross-validation results for NPL ratio 70
B29 Estimation trajectories of SDID vs. DiD vs. SCM: NPL ratio 72
B30 *p*-values for synthetic controls estimation of private credit 73
B31 Cross-validation results for private credit 74
B32 Estimation trajectories of SDID vs. DiD vs. SCM: private credit 76
B33 *p*-values for synthetic controls estimation of international debt 77
B34 Cross-validation results for international debt 78
B35 Estimation trajectories of SDID vs. DiD vs. SCM: international debt 80
B36 *p*-values for synthetic controls estimation of income shares 81
B37 Cross-validation results for bottom 50% income share 82
B38 Estimation trajectories of SDID vs. DiD vs. SCM: bottom 50% income share 84
B39 Estimation trajectories of SDID vs. DiD vs. SCM: middle 40% income share 85
B40 Estimation trajectories of SDID vs. DiD vs. SCM: top 10% income share 86

1 Introduction

Relaxing debt payments and constructing better risk-sharing mechanisms between financial institutions and the public help the economy recover more quickly (Mian and Sufi 2014a; Agarwal et al. 2017; Auclert et al. 2019, etc.). Risk-sharing mechanisms determine how effectively an economy could absorb financial shocks, especially when monetary policy alone is unable to offset the aggregate demand shock like in the case of the United States during the 2008 crisis (Eggertsson and Krugman 2012; Mian and Sufi 2014b). Although large-scale debt write-downs are understood as being helpful for economies to recover more quickly from financial crises, very few have been implemented in practice. Iceland stands out as a pioneer in experimenting with this policy when it forgave household debt amounting to 10% of its GDP in 2011 in response to the 2008 financial crisis.

Shortly before 2008, Iceland appeared to be just like any other Nordic small open economy with a tiny population of 315,000, but it was in fact a financial "behemoth." Its banking sector was highly concentrated as the assets of its three largest commercial banks occupied around 100% of the total commercial banking assets. Its private credit level reached 200% of GDP; household debt level was at 120% of GDP; and international private debt level stood at nearly 300% of GDP. These metrics were all amongst the highest level across 37 OECD member states.

As a result, the Icelandic banks and households were particularly hard hit by the 2008 financial crisis due to their dramatic buildup in debt, and the government took a series of unconventional debt restructuring measures to combat the crisis. It first imposed a moratorium on foreclosures and suspended debt service on foreign-denominated and inflation-linked loans. Then in 2011, it rolled out the "110% option" debt restructuring plan that capped mortgage loan principal at 110% of the collateral. The total mortgage write-down has been estimated at $16 billion – over 10% of Iceland's GDP. The loan losses were borne by the banks, which were nationalized and capitalized, and then defaulted on foreign obligations. One could consider the Icelandic experience as the most extreme application of big debt write-down followed by big money creation at the country level after the 2008 crisis.

This thesis seeks to investigate the macroeconomic effects of Iceland's 2011 debt jubilee program. Namely, did Iceland's debt jubilee program in 2011 help the country's macroeconomic and financial conditions recover more quickly than had this extreme version of debt forgiveness policy not been enacted? The main contribution of this paper is to offer credible evidence on the macroeconomic effects of debt relief, since most debt forgiveness programs studies have been cross-sectional, concentrated on regional policies, and with the aim of identifying specific channels of policy impacts at the micro level (e.g. Agarwal et al. 2017; Auclert et al. 2019).

To do so, I use the synthetic controls methods (SCM), which is one of the most popular tools to evaluate counterfactual policy outcomes in cross-country contexts. Iceland acts as my singular treatment unit, and the other 36 OECD member countries are used to construct a "synthetic" Iceland as the control unit. I examine both aggregate outcomes for the macroeconomy like unemployment rate and GDP per capita, as well as financial variables such as non-performing loan (NPL) ratio, international private debt level, and private credit flow.

I find the debt jubilee, which was deemed somewhat radical in contrast to the conventional economic wisdom, had a positive impact on Iceland's overall recovery. Within 3-5 years of the program implementation, compared to a counterfactual scenario where the jubilee policy had not been in place, Iceland's unemployment rate was 2 percentage points lower, GDP per capita was $2,500 higher, non-performing loan (NPL) ratio was 20 percentage points lower, private credit to GDP ratio was 50 percentage points lower, and international private debt to GDP ratio was nearly 300 percentage points lower. The policy not only stabilized the overall economy, it also brought the financial sector back to levels before Iceland's financialization boom that preceded the 2008 crisis.

I also explore the drawbacks of classical SCM and implement cross-validation and synthetic difference-in-differences (SDID) as robustness checks. Cross-validation provides an alternative way to assign weights for countries and predictors used to construct the "synthetic" Iceland, and SDID is a less restrictive version of SCM as it allows for systematic differences between synthetic and actual units to exist as long as the parallel trend assumption holds. Section 6 details my rationale behind using these different frameworks and how they relate to each other to present a holistic picture on Iceland.

The paper is organized as follows. Section 2 provides background information on the 2008 financial crisis and 2011 debt jubilee in Iceland. Section 3 discusses the recent literature on the economics of risk-sharing and debt forgiveness and how this paper seeks to contribute to it. Section 4 describes the data and their aggregate trends. Section 5 details the methodologies used in this paper and applications of synthetic controls in cross-country contexts. Section 6 explains my specific estimation strategy and procedure for robustness checks. Section 7 describes the results for each of the variables of interest. Section 8 is my conclusion. Appendix A displays data definitions, sources, the summary statistics table, and aggregate trends graphs. Appendix B presents results from robustness checks and discusses the cross-validation and synthetic differences-in-differences (SDID) frameworks. Appendix C offers some additional discussions for those interested in learning more about the econometric discussions around the synthetic controls method.

2 Background

2.1 The 2011 Icelandic Debt Jubilee

After the 2008 financial crisis hit, banks and households in Iceland experienced a dramatic buildup in debt. A large portion of the household debts was either inflation-linked or denominated by foreign currency, and the sharp depreciation in Icelandic currency króna resulted in the sharp buildup in debt burdens. The government enacted unconventional responses and reduced household debt by a combination of government measures and court decisions.

As detailed in Dalio (2019), the Iceland government first imposed a moratorium on foreclosures and suspended debt service on foreign-denominated and inflation-linked loans. In 2011, the government imposed a "110% option" debt restructuring plan that capped mortgage loan principal at 110% of the collateral. The total mortgage write-down has been estimated at $16 billion – over 10% of Iceland's GDP. The loan losses were borne by the banks, which were nationalized and capitalized, and then defaulted on foreign obligations. Additional measures such as interest subsidies and further voluntary restructurings were later rolled out to help households. Some might consider this as a extreme application of big debt write-down followed by big money creation.

The "110% option" was one of Iceland's longer-term debt relief measures. As explained by Baudino et al. (2020), the program applied to borrowers with over-mortgaged housing. It allowed for the write-off of outstanding balances on housing debt that exceeded 110% of the value of the property, subject to limits. A second measure of payment mitigation period of one to three years for individuals was also enacted, allowing borrowers to postpone their payments and giving them time to adapt their debts to their payment capacity. The Supreme Court went further in 2010 and 2012 as well, first declaring currency-linking illegal, therefore reducing the outstanding principal amounts for household debt burdens.

When Iceland's jubilee policy was first enacted, it did not seem to have gained much traction in the U.S. media or policymaking realm, likely because Iceland is such a small economy that its experience wouldn't seem to be an appropriate lesson for larger, more complex economic depressions. The event was briefly reported on NPR's *Planet Money* radio.[1] It was also briefly praised by Paul Krugman in a short New York Times column that argued the jubilee likely helped protect the people and social safety nets.[2] The jubilee was largely perceived as a distant event that only a tiny small open economy like Iceland could pull off, and few fundamental studies about the policy measure have come out since then.

2.2 The Rise and Fall of the Icelandic Banking Sector

The economic literature on the Icelandic financial crisis is quite sparse, and most writings about it are either historical overviews or overarching assessment reports by international financial institutions. At a high level, there are several pieces to the puzzle.

The Icelandic financial crisis started as an unprecedented banking crisis in October 2008. The Icelandic banking system grew excessively from 100% of GDP in 1998 to 900% in 2008 (Benediktsdóttir et al. 2017). The three banks that made up over 80% of the financial system and had experienced break-neck growth – Kaupthing, Landsbanki and Glitnir – collapsed within a few days of each other – a speed unmatched elsewhere.

When the 2008 financial crisis hit, the banking sector's outsized share of both foreign assets and liabilities became the root cause for its swift collapse. Not even the Central Bank of Iceland (CBI) could act as the lender of last resort in foreign currency as its FX reserves and foreign credit lines were no match for the banks' needs (Baudino et al. 2020). More specifically, the larger banks suffered from the systemic issue of being too interconnected through credit links while having no credible lender of last resort in foreign currency.

As the global financial crisis wrecked havoc across the global, confidence in the Icelandic banks dwindled. Funding strains also increased because of a number of reasons: wholesale funding could no longer be renewed, customer deposits were withdrawn, and the central bank collateralized lending had reached its limits. By mid-2008, the banks had essentially exhausted all their funding options. The funding stress in global financial markets at large also simultaneously became critical after mid-September 2008, following the collapse of Lehman Brothers, when repo and money markets dried up. A government bailout for the Icelandic banks was also not a realistic option because the state's resources were dwarfed by the size of the problem, and a bailout would have risked a sovereign default.

[1] NPR, (2014). "Iceland Experiments With A Jubilee Of Debt Forgiveness," December 11, 2014. `https://www.npr.org/2014/12/11/370156273/iceland-experiments-with-a-jubilee-of-debt-forgiveness`.

[2] Krugman, Paul (2011). "The Path Not Taken." *New York Times*, October 27, 2011. `https://www.nytimes.com/2011/10/28/opinion/krugman-the-path-not-taken.html`.

2.3 Debt Restructuring and the Resurrection of Household Finance

The price tag of the 2008-09 crisis was high for Iceland. Its GDP contracted by 10% between 2008 and 2009; unemployment rate rose from 1% in 2008 to more than 9% two years later.[3] The Icelandic króna lost close to half its value during 2008 against the dollar and euro. Prices rose by 30% between 2008 and 2011 on aggregate (Baudino et al. 2020).

A combination of government measures and court decisions were implemented to reduce household debt. One set of long-term relief measures were enacted in 2011, such as the "110% option" that wrote off outstanding balances on housing debt that exceeded 110% of the value of the property. After a Parliamentary Resolution passed in June 2013, the government took additional debt relief measures for inflation-indexed housing mortgages to reduce the principal of housing mortgages. The first measure in 2013 allowed the write-down of part of the outstanding principal of inflation-indexed housing mortgages used to purchase real property for personal use. The maximum write-down of principal was limited to ISK 4 million per household (around EUR 27,000 at the time), and the total amount written down corresponded to a 13% adjustment vis-à-vis the CPI used for indexation (Baudino et al. 2020). The second measure was a tax exemption that allowed households with housing mortgages to use payments that would otherwise go to a private pension fund to pay down their housing mortgages without being subject to income tax.

Benediktsdóttir et al. (2017) analyzed some of the recovery policies like emergency legislation, capital controls, alleviation of balance of payments risks, and preservation of financial stability, which in hindsight seem to have worked well in the authors' conclusion. As shown in figure A8, the aggressive debt restructuring initiatives – from household debt write-downs and giving individuals access to pension savings to pay down loan principals, to write-offs and loan paybacks for corporations – led to a dramatic decline in debt following the crisis in comparison with a selected number of other countries.

2.4 Criticism for the Debt Jubilee

While Benediktsdóttir et al. (2017) focused more on the broader economic recovery and banking sector restructuring, Ólafsson and Vignisdóttir (2012) used a unique nationwide household-level database to analyse how households' financial position evolved in the run-up to and aftermath of the financial crisis in Iceland. They assessed how the share of indebted households in financial distress evolved and how it was affected by debt restructuring measures and court decisions. Their conclusion seemed to be much less optimistic.

The authors found that forbearance efforts provided temporary breathing space, but the share in distress is estimated to have peaked at 27.5% in autumn 2009. Around 21 thousand households were in financial distress at the year end of 2010 (20% of all households) after policy and legal interventions had enabled roughly five thousand households to escape from distress. At that time, roughly 38 per cent of distressed households, or roughly eight thousand households, were in acute distress, which is defined to be their negative financial margin exceeded 100,000 kr. a month. Furthermore, a quarter of distressed households, or roughly 5,220 households, had a negative margin between 50,000 and 100,000 kr., while 37 per cent (7,750 households) had a negative margin less than 50,000 kr. Financial distress is found to be inversely related to income and age, as well as being higher among families with children and those with foreign-denominated debt than among childless couples and those with ISK-denominated loans only.

[3] According to the Icelandic Directorate of Labour, registered unemployment was at its lowest (0.8%) in the last quarter of 2007 and peaked (9.3%) in February and March 2010. Cited in Baudino et al. (2020), p.28.

Ólafsson and Vignisdóttir (2012) also analyzed the distribution of estimated write-offs and found that only 650 households exit the distress state due to the 110% option, which is implies that less than a quarter of the write-offs were received by households in financial distress – a very low ratio. In other metrics, the share of indebted households in distress decreases by only 0.6 percentage points, to 19.4%. Zooming in on the distribution of the estimated write-offs by the 110% option, Ólafsson and Vignisdóttir (2012) showed that the two highest income quintiles received 57% of the write-offs, while the lowest ones received 22%. FX mortgagors receive 14% and ISK mortgagors 86%. More importantly, only 23.5% of the write-offs went to households in financial distress; households with a large positive financial margin receive 21.5%; and households in between these two groups received the rest (see 4.14a in Figure A10).

By 2012, roughly 74.6 billion króna had been paid out to roughly 60 thousand individuals since March 2009 through "the third-pillar pension fund payouts." Analysis of the payouts combined with tax return data shows that roughly two-thirds of the pension fund payouts to singles and single parents in 2009 and 2010 were received by the highest income quintile, a fifth was received by the second-highest quintile, and the remaining 11.5% was received by the three lowest quintiles (Ólafsson and Vignisdóttir 2012). Such an inequitable distribution of write-offs stands in stark contrast to a household's likelihood of being in financial distress, which was greatest in the lowest income quintiles and declined markedly as income level went up (see 4.14c in Figure A10).

The observations above are only two of the many findings by Ólafsson and Vignisdóttir (2012) that seemed to show that the write-offs were ineffective in alleviating households' financial burden. The authors wrote that the question remains whether these measures, particularly the 110% option and the special interest rebate, represent an efficient use of the limited resources available for debt restructuring. Perhaps, the government should have paid greater attention to fight debt overhang rather than the struggle against household financial distress?

It may appear that the 110% option was *de facto* a fiscal redistribution program to the households, using the limited tax revenue collected from struggling financial institutions, and the redistribution was indifferent to the actual financial standings and incomes of the households. In other words, because the 110% option simply focused on mortgage write-downs, the richer households with more than a 110% lever in mortgage would be getting an equivalent policy treatment as a poor household struggling at the same leverage ratio. Meanwhile, those in financial distress but didn't own a mortgage over 110% leverage ratio wouldn't receive any assistance under this program.

Because of the severity and heterogeneity of the household financial distress, Ólafsson and Vignisdóttir (2012) concluded that general debt restructuring measures were not sufficient and exhausted valuable capacity for future debt restructuring. They believed that the majority of acutely distressed households needed tailor-made solutions to escape from distress. In order for general across-the-board write-offs to effectively lift acutely distressed households out of distress, the scope of these write-offs would have to be so extensive such that they would likely seriously endanger public debt sustainability and financial stability.

I do not have access to the same kind of micro-level data in Ólafsson and Vignisdóttir (2012), but I will still be able to examine their claims on an aggregate level. Namely, was the Icelandic recovery was indeed as effective as praised in many studies cited above? Or did the debt jubilee program in fact lead to more long-term issues, such as those related to financial stability, debt sustainability, and overall decline of economic vitality as proxied by GDP per capita and labor participation rate?

3 Literature Review

"Severe recessions are special circumstances because macroeconomic failures prevent the economy from reacting to a severe drop in demand. [...] When such failures prevent the economy from adjusting to such a large decline in consumption, government policy should do what it can to boost household spending. Debt forgiveness is exactly one such policy, and arguably the most effective" (Mian and Sufi 2014a).[4]

3.1 Debt Forgiveness and Risk Sharing

Risk-sharing matters in how effectively the economy absorbs financial shocks (e.g. Mian and Sufi 2014; Agarwal et al. 2017; Auclert et al. 2019). Starting with the book *House of Debt*, there have been a series of studies that show how relaxing debt payments and constructing better risk-sharing mechanisms for the public were effective in bringing the economy back to speed after a financial shock or recession. It was also of a widespread belief that unemployment level rose high after 2008 because monetary policy was unable to offset the aggregate demand shock (Eggertsson and Krugman 2012; Mian and Sufi 2014b), which would have been better alleviated through better risk sharing mechanisms. We know from this line of research that it helps to relax debt service requirements in times when aggregate demand is falling and when the economy is running into constraints like the zero lower bound.

However, there are still relatively few studies that offer credible evidence on the macroeconomic effects of debt relief, and those that study debt forgiveness programs, and most such studies are cross-sectional and concentrated on local and regional policies within a specific country (e.g. Agarwal et al. 2017; Auclert et al. 2019). My thesis is more interested in understanding debt forgiveness at the macro level; namely, what is the macroeconomic consequence of the more extreme version of the debt forgiveness program, which was enacted by Iceland? Therefore, my thesis fits into the rest of the literature not in terms of examining the micro-level impacts of the debt jubilee program on the Icelandic people, but rather from a more macro level.

The motivation behind debt forgiveness is intuitive: people respond differently to a given shock, and the heterogeneity of spending and saving becomes even more salient and extreme in times of financial distress. This is when the government is needed to pick up the slack of all of the agents, whether households or firms, who are being affected. Meanwhile, studies have shown that those who need debt forgiveness, such as households benefiting from the consumer bankruptcy system, are in greater financial distress and have higher marginal propensities to consume (MPC) than the general population (Dobbie et al. 2015; Gross et al. 2016), so forgiving debt for this subset of population could potentially boost aggregate demand and improve macroeconomic conditions for the overall population (Auclert et al. 2019).

Therefore, the justification for risk sharing is rooted in improving the well-being for both a specific subset of disadvantaged people and the society as a whole. The idea of risk sharing becomes important in contexts of financial distress because it is not only unwise to leave a portion of the population to suffer more severe consequences than they deserve, but also helping them would in fact allow the entire population to benefit more from the policy measure. In other words, debt forgiveness policies have the potential to facilitate better risk sharing mechanisms that end up disproportionately lifting up those who are harder hit while more effectively improving macroeconomic outcome for the entire economy.

[4] This quote also appeared in Auclert et al. (2019).

In *House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again*, Mian and Sufi (2014a) consider a scenario in which housing losses from 2006 to 2009 in the U.S. had been automatically more eventually distributed between creditors and home owners, and they showed quantitatively that the Great Recession would have only been a mild recession under such a scenario. They propose "Shared-Responsibility Mortgages" (SRMs), which would provide downside protection to the mortgage owner in case the value of the home goes down from the value when the owner purchased it, and the mortgage-payment schedule would be linked to the local house-price index. More research later have indicated that mortgage cram-downs would have had major benefits for the economy (Posner and Zingales 2009), and principal write-down is the most effective way of curing default permanently (Geanakopolos 2010). Mian and Sufi (2014a) also show that SRMs would have prevented most of the 5.1% of houses that went into foreclosure and would have reduced the fall in house prices by 9.7% between 2007 and 2009. By preventing foreclosures, SRMs could possibly have saved a staggering 46% of total housing-wealth loss - or $2.5 trillion.

Another line of research by Agarwal et al. (2017) examines the Home Affordable Modification Program (HAMP), an ex-post risk sharing program that the Obama Administration put in place for 3-4 millions homeowners who were close to foreclosure. The program was introduced in March 2009 with the stated goal of helping 3-4 million homeowners avoid foreclosure. The authors find that HAMP did not crowd out private modifications significantly. In fact, even when some borrowers were not eligible for the government modification, the banks were actually more willing to do these modifications privately. In sum, the modification program had a positive impact on house prices and durable consumption, which lends more direct evidence that risk sharing helps the local economy by boosting overall house prices and overall consumption, and presumably that will help reduce unemployment.

3.2 *Ex-ante* vs. *Ex-post* Risk Sharing

Another important policy takeaway from Mian and Sufi (2014a) and Agarwal et al. (2017) is the importance of *ex-ante* policy action and how frictions in pass-through could affect the effectiveness of government assistance programs. HAMP program was *voluntary*, so the creditor could say no to accepting this help. This legal clause became important in practice because a creditor can make the claim in court that a formal legal contract had already been drafted before, and it would not be justified for the government to step in later to alter the contract ex-post. This important friction may become important since any policy action ex-post may require new Congressional legislation and cause legal battles when people say they did not give consent ex-ante, and it becomes much harder to implement these programs.

The channels explored in Agarwal et al. (2017) were most clearly shown in Auclert et al. (2019), who exploit variation in ex-ante regulation across states in the U.S. through bankruptcy regulations. The authors demonstrate that states where it was relatively easier to declare bankruptcy recovered better from the 2008 financial crisis. They measure the easiness of bankruptcy declaration by the metric of how much protection one would receive for one's existing assets. They found that a household lost less when it was easier to declare bankruptcy, and one would do better precisely because of this built-in risk sharing feature to bankruptcy regulation.

As a result, the nature of a country's bankruptcy regime, and hence the underlying risk sharing mechanism (ex-ante vs. ex-post), is important in determining macroeconomic outcomes. This can also explain the difference between Europe and the U.S., where Europe tends to do significantly worse than the U.S. after a financial shock. One reason for that is the existence of the monetary union and the lack of a unified fiscal response, but another related reason is that European bankruptcy regulations are more stringent than that in the U.S. Ultimately, if ex-ante measures for risk sharing are more effective than ex-post ones, and if we hope to improve risk sharing and internalize the externalities at the macroeconomic level, the government should put in regulations that may want to mandate this policy and reduce any kind of frictions that would get in the way of benefiting those that are suffering.

The same debate about ex-ante and ex-post policies is also reflected in the Icelandic debt jubilee program, when many were concerned with the legality of the government imposing debt forgiveness *ex-post*. The Icelandic Supreme Court had to rule in 2010 and 2012 to declare currency-linking illegal, hence reducing the outstanding principal amounts for household debt burdens. The controversy and legal troubles would not have taken place had Iceland designed *ex-ante* debt forgiveness policies. Whether the Icelandic debt jubilee should have been ex-ante vs. ex-post is not a point of focus for this thesis, but it could be an interesting extension into the literature if any reader is interested.

3.3 Trade-offs between Micro- and Macro-level Studies on Debt Relief

The outcomes of interest for my thesis are quite closely linked with those explored in Auclert et al. (2019), which examines the impact of debt relief on a much more micro level than me. As briefly mentioned above, the authors exploit cross-state variations in debtor protections to measure debt forgiveness. For example, homestead exemptions influence the amount of value a household can save from being expropriated by the creditors.[5] The authors find that households benefited more in states with high homestead exemptions, and this helped the households spend more and boost local employment levels. Specifically, the consumption shock from households with less taken away was localized and then translated into employment consequences. The researchers explored the asymmetry of the potential real effects of this consumption shock on the non-tradable versus tradable sectors. They found a significant, positive effect on the non-tradable employment but not much on the tradable employment, which testifies that debt forgiveness and better risk sharing indeed ended up translating to boosting employment in the economy.

Although I am also interested in the employment effect of debt relief, what I am bringing to the table is fundamentally different from Auclert et al. (2019). To put it simply, they estimate a difference-in-differences model at the local level, so any macro-level statement they make would need to rely on additional modeling assumptions. The current structural assumptions they make cannot allow them to convert their diff-in-diff estimates into macro-level statements. At the theoretical and empirical level, their study still does not tell us the macro-level, all-inclusive effect of debt forgiveness. For example, banks may have forgiven person A's loan and allowed A to spend more and boost employment in A's town, but because the banks lost money when dealing with A, they may have simply decided to cut down their lending to person B and hurt others in the process. In other words, we are not sure what the aggregate effect of debt forgiveness is at the country level. The only way to estimate the macro-level impact is through what I do in this paper or a similar methodology designed for macro-level studies.

The trade-offs and constraints faced by Auclert et al. (2019) and me are hence different. The empirical

[5] In the local level context, debt forgiveness is a much more nuanced practice as a wide range of assets can be protected. One canonical example is homestead exemption: a household can save up to a dollar amount of value that they have in their house from being expropriated (taken away) by the creditors, and that number amount varies across states depending on the law. Auclert et al. (2019) precisely exploit this variation across states to arrive at their difference-in-differences estimates.

strategy of Auclert et al. (2019) is tighter than mine and allows them to identify specific channels that may influence unemployment, but if they want to make macro-level statements, they would have to appeal to another structural model. Using a synthetic controls approach, I do not need to impose additional structures of a theoretical model if I hope to examine macro-level impacts of the debt jubilee, so I do not suffer from the issue of mis-specification of structural model that they suffer from. However, it is much more challenging for me to make empirical identifications of the specific factors that may have caused the employment effects. Namely, as the later results of my paper will show, I have a much tougher time isolating out the specific causal factors for the treatment effects.

In sum, Iceland's debt jubilee was arguably the closest macro-level implementation of Mian and Sufi's vision, so it would be meaningful to examine the large-scale impact of such a program. Meanwhile, I hope to contribute to the literature by offering credible evidence on the macroeconomic effects of debt relief, since most debt forgiveness programs studies have been cross-sectional, concentrated on regional policies, and with the aim of identifying specific channels of policy impacts at the micro level like ones discussed above.

4 Data

4.1 Overview and Sources

I have constructed a cross-country, macro-level data set from various databases. The panel data is with annual frequency spanning 37 OECD countries from 1999 to 2017. I limit the data range to start from 1999 because that is when the euro was created and the European Union began to adopt a single currency, which I considered to be a regime change. A complete list of variables and their definitions and sources are provided in Appendix A. I do not use any micro-level data due to the nature of my study, but I do list a number of sources in the appendix that I found helpful when I searched for data for Icelandic households and businesses.

For macroeconomic data, I rely on the World Bank's World Development Indicators (WDI). This is one of the most common data bases that researchers use to collect data on GDP, inflation, unemployment and labor market outcomes, exchange rate, taxation, trade openness, household and government consumption, business sector compositions and so on.

I rely most of my financial data on the World Bank's Global Financial Development Database (GFDD), which collects an extensive data set of 109 financial system characteristics for 214 economies starting from 1960. The variables I focus the most on include banking sector concentration and profitability, non-performing loan ratio, credit boom intensity, health of non-financial institutions, and other business activity metrics. Laevan and Valencia (2018) also provide a helpful data set on systemic banking crises with variables such as peak NPL ratio, liquidity support amount, recapitalization cost for banks etc. for a select number of countries.

Unfortunately, neither the GFDD nor WDI provide comprehensive data on housing prices across countries, understandably so given the regional heterogeneity of housing markets. To capture the boom and bust of the housing markets and cycles, I rely on the BIS and the Dallas Fed's global housing price indices. In addition, the World Inequality Database (WID) provides comprehensive data on income and wealth inequality, primarily measured by metrics such as the income shares of bottom 50%, middle 40%, top 10%, and top 1% of the population.

4.2 Summary Statistics and Aggregate Trends

Table A1 summarizes all the variables used in the actual synthetic controls estimation process. I list Iceland's summary statistics next to those for all 37 OECD countries to provide better context and further demonstrate Iceland's unique financial conditions. We can observe that Iceland stands around or above the 75% percentile of OECD for variables such as domestic private credit, bank concentration, bank non-performing loan ratio, housing price cumulative growth rate from 2003 to 2008, and international private debt to GDP ratio.

When estimating my model, I start with the more aggregate outcomes (GDP per capita and unemployment rate) and gradually zoom in on the financial characteristics that are at an abnormal level relative to the rest of the OECD median (NPL ratio, private credit supply, and international debt level). Meanwhile, a wide range of additional banking sector and labor market variables will be used to help construct synthetic controls, such as bank concentration, gross capital formation (proxy for investment rate), domestic savings rate, labor productivity growth rate, household and government consumption, trade openness, and so on.

To get a more complete picture of Iceland's general macroeconomic trends and better visualize the summary statistics, I plot out the aggregate trends from Figure A1 to A20. The most important macroeconomic and financial variables for this study are plotted with Iceland comparing to the OECD average or the rest of the OECD peers. I now describe some of the major findings from these time series trends, from which we can clearly observe that Iceland is a unique economy such that it is quite far from the OECD average in most of the important metrics I plot.

We observe that Iceland's GDP per capita level has followed roughly the same rise and fall over the years as the OECD average but has remained consistently higher with the same growth rates (Figure A1). Iceland's unemployment rate (mostly fluctuating between 2-4% before 2008) is structurally lower than the OECD average and in fact one of the lowest amongst all OECD countries (Figure A2). The 2008 financial crisis seems to have dampened Iceland's employment to population ratio from a consistently high 75% to 70%, but it did not cause any material changes to the labor participation rate for its overall population or females in particular, which could testify to the relative strength of Iceland's labor market (Figure A3).

Iceland's non-performing loan (NPL) ratio rose from less than 5% in 2006 to around 20% of gross loans in 2010, while the OECD average and most of its OECD peers remained around 5%. This suggests that the 2006-2008 dramatic rise in NPL likely served as a preface for an eventual crash after the 2008 financial crisis hit hit (Figure A4). The same situation happened with private credit, which rose from around 100% to more than 300% of GDP between 2002 to 2006, while the rest of OECD saw a more moderate expansion from 50% to 100% by 2008 (Figure A5).

Iceland's international private debt grew from around 100% of GDP in 2000 to 300% in 2008, while the rest of OECD countries mostly remained below 100% and had an average of below 50%. It is not an exaggeration to say that Iceland's international private debt level had started being the highest amongst all OECD countries since around 2006 and remained so until 2012 (Figure A6). Iceland's domestic public debt level, on the other hand, remained below 20% of GDP compared to the 35% OECD average over the years, though Iceland's domestic public debt ballooned between 2008 and 2011, eventually reaching 50% (Figure A7). Its domestic private debt level, which grew from around 40% in 2000 to more than 100% of GDP by

2007, had also been consistently higher than the OECD average of 30% (Figure A7). All the above results signal that Iceland's debt problem around 2008 was mostly an international one and a private one. The public sector has been more prudent with debt management, while the private sector dramatically expanded its balance sheets and bonded closely with the rest of the world.

Iceland's households and companies have also been heavily indebted, with household debt standing at more than 80% of GDP in 2008 and corporate debt growing from 50% in 2005 to almost 150% by 2008 (Figure A9). While government consumption expenditure has remained relatively steady over the years, household consumption expenditure had more volatile fluctuations and declined from around 56% of GDP in 2007 to 50% in 2009; it only recovered more after 2011 (Figure A11). Adding on top of all this was the dramatic boom and bust of the housing market: Iceland's normalized housing price index went from around 100 in 2005 to 150 by 2008 before crashing down to 100 in 2009. In other words, the steep rise and fall all took place with the span of three to four years (Figure A12). As a result, the heavy indebtedness of households, the decline in household consumption, and the dramatic falling of housing prices were likely a main driver of the 2011 debt forgiveness program, though surprisingly we are not observing a sharp decline in household debt level after 2011 as opposed to corporate debt dramatically coming down after 2011.

As for the financial sector, Figure A13 and A14 showcase some important banking sector metrics. We see that Iceland's bank concentration has consistently remained around 100% from 1999 to 2017. The OECD average is around 70%; the U.S. has stayed around 30%; and other notable countries that suffered from a dramatic banking crisis in 2008 like Ireland and Latvia only hovered between 50% and 80%. In other words, Iceland's banking sector has always been heavily concentrated and significantly above the level of other OECD countries. Iceland's banking sector profitability has been on par with its OECD peers and remained fairly steady, though it dipped more significantly than its OECD peers during the 2008 downturn to around -20% return on assets before tax. Iceland's bank deposits to GDP ratio has risen quite dramatically over the years from 40% in 2000 to around 90% by 2008, indicating an increase in domestic savings. The bank regulatory capital to risk-weighted assets ratio for Iceland has remained steady over the years and on par with the OECD average before rising to significantly higher levels after 2008 , likely due to an increase in regulatory oversight in light of the financial crisis. Other metrics that could signify Iceland's growing financialization over the years are such as the net investments in non-financial assets as % to GDP, for which Iceland has trailed below the OECD average, and this consistent underinvestment in the non-financial sector could exacerbate the economy's dependence on the swings of the financial sector. Net loans from nonresident banks as a % of GDP also rose dramatically from around 0% in 2002 to 40% to 2006, possibly indicating the growing connections with foreign capital (Figure A15).

Lastly, Figure A17 to A20 showcase some of the main data from Laeven and Valencia (2018) and the consequences of systemic banking crises. I plot the average level of NPL ratio by the range of outcomes and policy actions adopted after a systemic crisis took place. In these binscatter plots,[6] we see that NPL ratio usually rises to be dramatically higher for countries that have suffered from a systemic banking crisis, did not receive IMF program support, did not enact a deposit freeze program, had lower levels of deposit insurance,

[6] I used the *binscatter* program in Stata to produce a binned scatterplot by desired categories. Binned scatterplot is a non-parametric way to plot the conditional expectation function, thus describing the average y-value for each x-value. To generate a binned scatterplot, binscatter groups the x-axis variable into equal-sized bins, computes the mean of the x-axis and y-axis variables within each bin, then creates a scatterplot of these data points. By default, binscatter also plots a linear fit line using OLS, which represents the best linear approximation to the conditional expectation function. More information on binscatter can be found here: https://michaelstepner.com/binscatter/binscatter-StataConference2014.pdf

had lower levels of peak liquidity support, where losses were not imposed on depositors, and where banks were nationalized and recapitalized. As we know, Iceland suffered a systemic banking crisis, but it enacted a series of policies that would be considered as "on the right side of history" that stabilized and brought down the NPL ratio, such as receiving IMF support, enacting deposit freeze, imposing losses on depositors, and eventually nationalizing and recapitalizing the banks. Such a series of actions meant that Iceland stood out amongst its OECD peers in responding to the crisis, and it is thus worth further exploring whether its financial system indeed recovered better on a macro level with my estimations later.

5 Methodology

Section 5.1 explains the synthetic controls methods (SCM), which is the main methodology I use for my estimation. Sections 5.2 and 5.3 outline the cross-validation technique and synthetic difference-in-differences (SDID), which I use for robustness checks.

5.1 Synthetic Controls Methods (SCM)

> *The synthetic control approach "is arguably the most important innovation in the policy evaluation literature in the last 15 years. This method builds on difference-in-differences estimation, but uses systematically more attractive comparisons" (Athey and Imbens 2017).*

5.1.1 Motivation and Background

In order to estimate the effect of the 2011 debt jubilee policy, one would need to investigate the unobserved counterfactuals, namely what would have happened to Iceland's economy had the country not enacted the interventions. However, one problem of comparative case studies is that the selection of control group is often ambiguous, and the standard errors do not reflect the uncertainty about the ability of the control group to reproduce the counterfactual of interest. Hence, cross-country regressions are often criticized because they use countries of very different characteristics in the same estimation procedure without properly adjusting for their heterogeneity. In addition, it is often difficult to isolate out the treatment effect in contexts where only one treatment unit exists alongside a handful of control units. Nevertheless, we usually do not want to give up country-level data because many micro-level data are not readily available, like in this study.

Meanwhile, as my aggregate plots for most of the outcome variables of interest (GDP per capita, unemployment rate, NPL ratio, etc.) for Iceland vs. the OECD sample have shown, there are no pre-treatment parallel trends, so a simple difference-in-differences model would not have worked well to isolate out the treatment effect of the 2011 debt jubilee. I also only have one effective treatment unit alongside 36 other control units, and not only does the parallel trend assumption not hold, the standard errors for a diff-in-diff estimation would simply be too high to yield any meaningful result. I thus would have to find an alternative in this cross-country, single-treatment-unit context.

Synthetic control methods (SCM) were originally proposed in Abadie and Gardeazabal (2003) and Abadie et al. (2010) with the aim to estimate the effects of aggregate interventions – interventions that are implemented at an aggregate level affecting a small number of large units (such as a cities, regions, or countries) on some aggregate outcome of interest. SCM is based on the idea that a combination of unaffected units often provides a more appropriate comparison than any single unaffected unit alone, and SCM seeks to formalize the selection of these comparisons units through a data-driven procedure (Abadie 2020).

For example, Abadie et al. (2015) estimates the effect of the 1990 German reunification on per-capita GDP in West Germany, which is a cross-country study where only one treatment unit exists just like my case. In their paper, the policy intervention is the 1990 German reunification and the treated unit is former West Germany. The donor pool is restricted to 16 OECD countries. A population weighted sample that includes all OECD countries is very far from the real happenings in West Germany, but by using an accurate set of comparisons, the authors successfully constructed a "synthetic" West Germany sample (including Austria with a weight of 0.42, Japan with weight of 0.16, the Netherlands with 0.10, Switzerland with 0.11, and the U.S. with 0.22) that actually fits West Germany's pre-treatment GDP per capita trajectory almost perfectly. This allowed them to isolate out the effect of the 1990 reunification. My application of SCM on Iceland largely follows their logic and general framework.

5.1.2 Econometric Setup for SCM

I now explain the detailed econometric framework of SCM.[7] Suppose that we observe $J+1$ units $j = 1, 2, \ldots, J+1$ in periods $1, 2, \ldots, T$. Without loss of generality, we assume that the first unit ($j = 1$) is the treated unit – the unit affected by the policy intervention of interest, ad the remaining J units are the "donor pool" – the set of potential comparisons where "$j = 2, ..., J+1$" is a collection of untreated units, not affected by the intervention. We assume that the data span T periods, with the first T_0 periods being before the intervention, so unit $j = 1$ is exposed to the intervention during periods $t = T_0 + 1, \ldots, T$.

For each unit, j, and time, t, we observe the outcome of interest, Y_{jt}. For each unit, j, we also observe a set of k predictors of the outcome, $X_{1j}, \ldots, X_{kj}$, which may include pre-intervention values of Y_{jt} and which are themselves unaffected by the intervention.

There would be two potential outcomes: Y_{it}^N is the outcome that would be observed for unit i at time t in the absence of the intervention; Y_{it}^I is the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods between $t = T_0 + 1, \ldots, T$.

We aim to estimate the effect of the intervention on the treated unit $\alpha_{1T_0+1}, \ldots, \alpha_{1T}$, where

$$\alpha_{1t} = Y_{it}^I - Y_{it}^N = Y_{1t} - Y_{1t}^N \tag{1}$$

for $t > T_0$, and Y_{1t} is the observed outcome for unit one at time t. In plain words, we're interested in the treatment unit *with* the intervention in the post-treatment period minus the outcome that we would've observed *without* the intervention in the post-treatment period.

The purpose of policy evaluation and the problem we need to solve is to estimate Y_{1t}^N for $t > T_0$, i.e. how the outcome of interest would have evolved for the treated unit *in the absence of the intervention*. We say that Y_{1t}^N is a *counterfactual outcome* for $t > T_0$ because the treated unit would have been exposed to the intervention of interest after $t = T_0$. We observe Y_{1t}^I and want to estimate Y_{1t}^N that is equivalent to the effect of the policy intervention of interest.

[7] The methodological write-up below largely consists of my personal explanations and direct quotations from Abadie (2020) and a sequence of video lectures done by Abadie to explain the methodology. For video lecture, see "IDSS Special Seminar: Alberto Abadie - 11/23/15." *MIT Institute for Data, Systems, and Society*. `https://www.youtube.com/watch?v=2jzL0DZfr_Y`. The methods used to compute the SCM estimators is available on the authors' webpages: `http://web.stanford.edu/~jhain/synthpage.html`

To do so, we find the combination of untreated units that best resembles the treated unit before the intervention in terms of the values of k relevant covariates (predictors of the outcome of interest measured before the intervention). These covariates $X_{1j}, \ldots, X_{kj}$ may include pre-intervention values of Y_{jt} and which are themselves unaffected by the intervention. The $k \times 1$ vectors $\boldsymbol{X_1, \ldots, X_{J+1}}$ contain the values of the predictors for units $j = 1, \ldots, J+1$, respectively. The $k \times J$ matrix $\boldsymbol{X_0} = [\boldsymbol{X_2, \ldots, X_{J+1}}]$ collects the values of the predictors for the J untreated units. For each unit j and time t, we may then estimate Y^N_{jt}, Y^I_{jt}, and so on as explained above. Notice that equation (1) allows the effect of the intervention to change over time. This is crucial because intervention effects may not be instantaneous and may accumulate or dissipate as time after the intervention passes.

5.1.3 Estimation

The synthetic control method is based on the observation that a combination of units in the donor pool may approximate the characteristics of the affected unit substantially better than any unaffected unit alone. A synthetic control is defined as a weighted average of the units in the donor pool. Formally, a synthetic control can be represented by a $J \times 1$ vector of weights, $\boldsymbol{W} = (w_2, ...w_{J+1})'$. Given a set of weights, $\boldsymbol{W}$, the synthetic control estimators of Y^N_{1t} and α_{1t} are:

$$\widehat{Y}^N_{1t} = \sum_{j=2}^{J+1} w_j Y_{jt} \tag{2}$$

$$\widehat{\alpha}_{1t} = Y_{1t} - \widehat{Y}^N_{1t}. \tag{3}$$

The weights are restricted to be non-negative and to sum to one, so synthetic controls are weighted averages of the units in the donor pool. Abadie (2020) explained that the requirement that weights should be non-negative and no greater than one can be relaxed at the cost of allowing extrapolation. For example, Abadie et al. (2015) show that, in the context of estimating the effect of a policy intervention, there is a regression estimator that can be represented as a synthetic control with weights that are unrestricted except for that the sum of the weights is equal to one. By not restricting the weights to be in [0, 1], regression allows extrapolation.[8]

As an example, a synthetic control that assigns equal weights, $w_j = 1/J$, to each of the units in the control group results in the following estimator for α_{1t}:

$$\widehat{\alpha}_{1t} = Y_{1t} - \frac{1}{J}\sum_{j=2}^{J+1} Y_{jt}. \tag{4}$$

Expressing the comparison unit as a synthetic control motivates the question of how the weights, $\boldsymbol{W} = (w_2, ...w_{J+1})'$, should be chosen in practice. Abadie and Gardeazabal (2003) and Abadie et al. (2010) propose to choose them so that the resulting synthetic control best resembles the pre-intervention values for the treated unit of predictors of the outcome variable. That is, given a set of non-negative constants $v_1, ...v_k$, we should choose the synthetic controls $\boldsymbol{W^*} = (w^*_2, ...w^*_{J+1})'$ that minimize:

[8] Many disagree that the weights have to sum to one, and this could cause potential trouble in the context of applying SCM to a unique small open economy like Iceland. I address this econometrics debate in Appendix C of my thesis.

$$||\boldsymbol{X_1} - \boldsymbol{X_0 W}|| = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)} \tag{5}$$

$$= \left(\sum_{h=1}^{k} v_h (X_{h1} - w_2 X_{h2} - ... - w_{J+1} X_{hJ+1})^2 \right)^{1/2} \tag{6}$$

subject to the same restriction that $w_2, \ldots, w_{J+1}$ are non-negative and sum to one. Then, the estimated treatment effect for the treated unit at time $t = T_0 + 1, \ldots, T$ is:

$$\widehat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}. \tag{7}$$

This metric $||\boldsymbol{X_1} - \boldsymbol{X_0 W}||$ depends on some $k \times k$, symmetric, and positive semidefinite matrix $\boldsymbol{V}$. Such a choice of $\boldsymbol{V}$ essentially produces a synthetic control, $\boldsymbol{W}(\boldsymbol{V}) = (w_2(\boldsymbol{V}), ...w_{J+1}(\boldsymbol{V}))'$ that can be determined by minimizing the above equations subjects to these weights $\boldsymbol{W}(\boldsymbol{V})$. This should a constrained quadratic optimization that is fairly solvable. And in turn, $\boldsymbol{W}(\boldsymbol{V})$ is meant to reproduce the behavior of the outcome variable for the treated unit in the absence of the treatment. Therefore, we may essentially see the weights $v_1, ...v_k$ as reflecting the predictive value of the covariates.

5.1.4 Inference and Placebo Tests

Following the synthetic control study, one would typically extend it with placebo tests for robustness checks. They produce quantitative inference in comparative case studies and are similar to the classic framework for permutation inference, where the distribution of a test statistic is computed under random permutations of the sample units' assignments to the intervention and non-intervention groups. As we apply SCM estimation to every potential control and every combination of placebo test groups, looping through that process would produce a list of outcomes that allow us to examine whether the counterfactual effect estimated for our original treatment unit is indeed significant relative to the effects estimated for a unit chosen at random (Abadie et al. 2015). The advantage of placebo tests is that we can always arrive at the exact distribution of the estimated effect of the placebo interventions and judge from that whether our original SCM estimation was robust.

There are two kinds of placebo tests that one could conduct for SCM: "in-time" and "in-place" placebo tests. Abadie et al. (2010) and Abadie et al. (2015) apply in-time and in-place estimates and explain the methodology in detail. The intuition is fairly straightforward. Suppose that the SCM estimates for the intervention of interest are statistically significant, but SCM also returns sizable results for dates when or units where the intervention did not occur, then our confidence in the validity of our SCM estimates would diminish significantly (Heckman and Hotz 1989, cited in Abadie et al. 2015). In other words, the purpose of placebo tests is to apply the counterfactual on another unit or date that we know did not undergo the policy intervention of interest, and if those that did not experience the structural shocks also exhibit similar effects as the unit that did experience the shocks, then it testifies that the effects are likely not a result of these shocks.

Both in-time and in-place placebo tests are possible to conduct when the data set has enough number of time periods and variables. In Abadie et al. (2015), the authors had data from 1960 and tested whether the SCM method produces large estimated effects when applied to 1975, a date before when the actual reunification took place in 1990. They found no significant results for 1975 but did so for 1990, so they thus concluded from this in-time placebo test that synthetic controls do provide a good counterfactual and predictions of the trajectory of the outcome in West Germany after the actual reunification.

An in-place placebo test was applied in Abadie et al. (2010), where the authors iteratively applied the synthetic method to each state in the donor pool (like Tennessee and then so on) and obtained a distribution of placebo effects. They then compared the gap (RMSPE) for California (which is the unit that actually received the policy intervention) to the distribution of the placebo gaps. They compared the root mean square prediction error (RMSPE) for each state using the formula below:

$$RMSPE = \left(\frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \left(Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt} \right)^2 \right)^{1/2}. \tag{8}$$

One crucial function of placebo tests is to determine the statistical significance of our inference. By producing a distribution of in-place placebo effects, I essentially arrive at a non-parametric, exact test for statistical significance that has the advantage of not imposing any distribution on the errors (Galiani and Quistorff 2017). More specifically, suppose that the estimated effect for a particular post-treatment period is $\hat{\alpha}_{1t}$ and the distribution of corresponding in-place placebos is $\hat{\alpha}_{1t}^{PL} = \{\hat{\alpha}_{jt} : j \neq 1\}$. I would then calculate the two-sided p-values with the formulas below:

$$\begin{aligned} p\text{-value} &= \Pr(|\hat{\alpha}_{1t}^{PL}| \geq |\hat{\alpha}_{1t}|) \\ &= \frac{\sum_{j \neq 1} 1(|\hat{\alpha}_{1t}^{PL}| \geq |\hat{\alpha}_{1t}|)}{J}, \end{aligned}$$

and the one-sided p-values (for positive effects) are

$$p\text{-value} = \Pr(\hat{\alpha}_{1t}^{PL}| \geq |\hat{\alpha}_{1t}).$$

When treatment is randomized, the setup above becomes classical randomization inference.[9] And if the treatment is not randomly assigned, the p-value can still be interpreted as the proportion of control units that have an estimated effect at least as large as that of the treated unit. Therefore, the p-values shown in the Appendix B robustness check section (such as in Figure B21 and the figures alike) are essentially a vector of the proportions of placebo effects that are at least as large as the main effect for each post-treatment period.[10]

[9] Galiani and Quistorff (2017) note here that one may want to include $\hat{\alpha}_{1t}$ in the comparison distribution as is common in randomization inference, which would 1 to the numerator and denominator of the p-value fraction above. They also note that there are several additional approaches as how to add the effects on the treated to the comparison distribution. Abadie et al. (2014) and Cavallo et al. (2013) do not take this approach and examine approaches that allow for more than one nit to experience treatment and at possibly different times. For more, see http://econweb.umd.edu/~galiani/files/synth_runner.pdf.

[10] One may also construct p-values with "psuedo t-stats," which would be a vector of the proportions of placebo pseudo t-statistics (unit's effect divided by its pre-treatment RMSPE) that are at least as large as the main pseudo t-statistic for each post-treatment period. For more information on this, see Galiani and Quistorff (2017).

5.2 Cross-validation in SCM

The cross-validation technique, which is also referred by some as out-of-sample testing, has become more popular over the years given the rise of machine learning in fields outside of economics like data science and machine learning. It assesses how the results of a statistical model could generalize to an independent data set and may add more structure to the current statistical analysis. Abadie et al. (2010) investigated California Proposition 99 and can be considered as the preeminent representation of the classical, baseline specification for SCM. Then, Abadie et al. (2015) studied the economic effect of the German reunification and is considered as an improvement of the baseline specification by using the cross-validation methodology. I do not consider cross-validation as a drastically different methodology because it is nested within the classical SCM framework initially proposed by Abadie and other pioneers in SCM, and I use cross-validation in this paper for the purpose of robustness checks for my main SCM estimations. The results for cross-validation can be found in Appendix B.

5.2.1 How to choose the V matrix?

One of the central challenges of SCM is to determine the optimal $\boldsymbol{V}$ matrix, which then subsequently determines the relative importance between the characteristic variables. $v_h = 1$ if characteristic h matters the most, and $v_h = 0$ if the least. Abadie (2020) explains that a simple selector of v_h is the inverse of the variance of $X_{h1}, \ldots, X_{hJ+1}$, which in effect rescales all rows of $[X_1 : X_0]$ to have unit variance. Alternatively, Abadie and Gardeazabal (2003) and Abadie et al. (2010) choose $\boldsymbol{V}$, such that the synthetic control $\boldsymbol{W}(\boldsymbol{V})$ minimizes the mean squared prediction error (MSPE) of this synthetic control with respect to Y_{1t}^N:

$$\sum_{t \in \Psi_0} \left(Y_{1t} - w_2(V)Y_{2t} - \cdots - w_{J+1}(V)Y_{J+1t}\right)^2, \tag{9}$$

for some set $\Psi \subseteq \{1, 2, \ldots, T_0\}$ of pre-intervention periods. This would be the default method used by Stata and R when executing synthetic controls commands.

The issue with the above methods of determining $\boldsymbol{V}$ and $\boldsymbol{W}$ is that it is still done on a more ad-hoc than formulaic basis. Even after being able to construct a good pre-trend for their treated unit, the question still remains: How robust are the synthetic controls to the choices of characteristics? There is a long footnote of characteristics that Abadie et al. (2015) tried out in order to construct the best comparison sets; they also reasoned that perhaps all the results depend on one particular country, so it is a really iterative and trial-and-error process to test out what countries and characteristics fit the best. As a result, constructing the synthetic control group is more of an art than strict science. There is no formula as to how exactly one should go about choosing the set of characteristics variables (predictors used for the MSPE optimization to determine the weights for the synthetic control group) and their time range.

Saying that SCM will always has to be done on a trial-and-error basis would leave many unsatisfied and confused, and the key contribution of Abadie et al. (2015) is using the cross-validation technique to determine the $\boldsymbol{V}$ matrix. Cross-validation also helps restrict the size of the donor pool and ends up forcing the program to considering only units similar to the treated unit, and this reduces the possibility of overfitting. As explained by Abadie et al. (2015), overfitting occurs when characteristics of the unit affected by the intervention or event of interest are artificially matched by combining idiosyncratic variations in a large sample of unaffected units. The risk of overfitting thus also motivates the adoption of the cross-validation techniques.

The formal logic of cross-validation goes as the following: the purpose of synthetic control method is to select a set of weights $\boldsymbol{W}$ such that the resulting synthetic control resembles the affected unit before the intervention along the values of the variables $X_{11}, \ldots, X_{k1}$. The question of choosing $\boldsymbol{V} = (v_1, \ldots, v_k)$ boils down to assessing the relative importance of each of $X_{11}, \ldots, X_{k1}$ as a predictor of Y_{1t}^N. That is, the value v_h aims to reflect the relative importance of approximating the value of X_{h1} for predicting Y_{1t}^N in the post-intervention period, $t = T_0 + 1, \ldots, T$. Because Y_{1t}^N is not observed for $t = T_0 + 1, \ldots, T$, we cannot directly evaluate the relative importance of fitting each predictor to approximate Y_{1t}^N in the post-intervention period. However, Y_{1t}^N is observed for the pre-intervention periods $t = 1, 2, \ldots, T_0$, so it is possible to use pre-intervention data to assess the predictive power on Y_{1t}^N of the variables $X_{1j}, \ldots, X_{kj}$. This can be accomplished in the following manner, as detailed in Abadie (2019):

1. Divide the pre-intervention periods into a initial *training* period and a subsequent *validation* period. For simplicity and concreteness, we will assume that T_0 is even and the training and validation periods span $t = 1, \ldots, t_0$ and $t = t_0 + 1, \ldots, T_0$, respectively, with $t_0 = T_0/2$. In practice, the lengths of the training and validation periods may depend on application-specific factors, such as the extent of data availability on outcomes in the pre-intervention and post-intervention periods, and the specific times when the predictors are measured in the data.

2. For every value $\boldsymbol{V}$, let $\tilde{w}_2(\boldsymbol{V}), \ldots, \tilde{w}_{J+1}(\boldsymbol{V})$ be the synthetic control weights computed with training period data on the predictors. The mean squared prediction error of this synthetic control with respect to Y_{1t}^N in the validation period is:

$$\sum_{t=t_0+1} \left(Y_{1t} - \tilde{w}_2(V)Y_{2t} - \cdots - \tilde{w}_{J+1}(V)Y_{J+1t}\right)^2 . \tag{10}$$

3. Minimize the mean squared prediction error in the previous equation with respect to $\boldsymbol{V}$.

4. Use the resulting $\boldsymbol{V}^*$ and data on the predictors for the last t_0 periods before in the intervention, $t = T_0 - t_0 + 1, \ldots, T_0$, to calculate $\boldsymbol{W}^* = \boldsymbol{W}(\boldsymbol{V}^*)$.

However, the cross-validation methodology is still constrained in certain other ways, such as we still have to rely on the economic literature and intuition to determine which characteristic outcomes to pick in the first place for the cross-validation methodology. In other words, we may make appropriate adjustments for optimizing $\boldsymbol{V}$ and $\boldsymbol{W}$, but the SCM approach forces the user to initially narrow in on a set of characteristic outcomes regardless. Therefore, one could say that if the classical SCM framework does not yield robust results, it is somewhat unlikely that using cross-validation alone would help yield more statistically significant results. My results later also testify to this. I thus consider cross-validation to be a robustness check to my results in addition to the placebo tests.

5.3 Synthetic Difference in Differences (SDID)

5.3.1 Motivation: Comparison between DiD, SCM, and SDID

The fundamental purpose of using SDID or SCM is to design good-quality simulation studies. While SCM is often considered as the preeminent methodology in cross-country policy evaluations, it suffers from a variety of estimation issues that are hard to address as shown in the context of this thesis. Therefore, I supplement my SCM estimations with SDID in an attempt to further isolate out certain effects of interest. The methodological write-up largely follow Arkhangelsky et al. (2020) and a sequence of video lectures done by Susan Athey to explain the methodology.[11]

A general simulation studies setup is as follows: we have a block assignment of the treatment $\boldsymbol{W}$ and an $(N \times T)$ matrix of observed outcomes $\boldsymbol{Y}$. The goal would be to estimate the average treatment effect, which means imputing potential outcomes for the states and time periods that are treated:

$$Y = \begin{pmatrix} Y_{11} & Y_{12} & Y_{13} & \dots & Y_{1T} \\ Y_{21} & Y_{22} & Y_{23} & \dots & Y_{2T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & Y_{N2} & Y_{N3} & \dots & Y_{NT} \end{pmatrix} \quad \text{(realized outcome)}$$

$$W = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 1 & 1 \end{pmatrix} \quad \text{(binary treatment with block assignment)}$$

Here, the rows of Y and W correspond to units and the columns correspond to time periods.

What has become one of the most popular approaches to solve the problem above is SCM, in which case we would predict

$$\hat{Y}_{NT}(0) = \sum_{i=1}^{N-1} w_i Y_{iT},$$

where w is chosen so that for all $t = 1, \dots, T-1$:

$$\begin{aligned} Y_{NT} &\approx \sum_{i=1}^{N-1} w_i Y_{iT}, \\ w_i &\geq 0, \\ \sum_{i=1}^{N-1} w_i &= 1. \end{aligned}$$

But as explained in sections above, when practitioners do classical SCM, there is some "eyeballing" element to it. One would have to make choices for optimization periods, characteristic variables and so on, and not even the cross-validation technique would spare one the need to make some trial-and-error tests in order to

[11] For video lecture, see "Susan Athey: Synthetic Difference in Differences." *Online Causal Inference Seminar*. https://youtu.be/r2DzGAigTl4.

improve accuracy. Scholars have made a conscious effort to make the procedure more systematic, and SDID can be seen as one of these attempts.

Doudchenko and Imbens (2016) propose that we can also view the SCM problem as a vertical regression, meaning that we are effectively regressing a row (outcomes of the treated unit) on the other rows (outcomes of the other units). Instead of trying to solve a minimization problem with the constraints of $\sum w_i = 1$, we can do a regularized regression to find proper weights, estimating:

$$\begin{aligned} \hat{\tau} &= Y_{NT} - \hat{Y}_{NT}(0), \\ \hat{Y}_{NT}(0) &= \sum_{i=1}^{N-1} \hat{w}_i Y_{iT}, \\ \text{where} \quad \hat{w} &= \min_{w \geq 0, \sum_i w_i = 1} \sum_{i=1}^{T-1} \left(Y_{NT} - \sum_{i=1}^{N-1} w_i Y_{it} \right)^2 \quad (\text{"vertical" regression}) \end{aligned}$$

If there are more units than time periods, and especially if we hope some of the w_i values can be negative, then this method could potentially be much more appealing.

Note that there are $T-1$ observations and $N-1$ regressors. We call it vertical regression because the units are time periods are symmetric mathematically; so we can turn this problem around and say: instead of trying to express the treated unit as a weighted sum of other units, we can just think of this setup as expressing the final treatment time period as a weighted sum of other time periods for each unit, essentially transposing the matrix.

We can also view the SCM problem as a form of horizontal regression. This is more common when people take unconfoundedness type approaches or typical cross-sectional causal inference approaches to the panel data setting. Now I am trying to find weights such that for each unit i, the final time period is close to a time weighted average of that unit's previous outcomes, estimating:

$$\begin{aligned} \hat{\tau} &= Y_{NT} - \hat{Y}_{NT}(0), \\ \hat{Y}_{NT}(0) &= \lambda_0 + \sum_{i=1}^{T-1} \hat{\lambda} Y_{NT}, \\ \text{where} \quad \hat{\lambda} &= \min_{\lambda} \sum_{i=1}^{N-1} \left(Y_{iT} - \sum_{i=1}^{T-1} \lambda_t Y_{it} \right)^2 \quad (\text{"horizontal" regression}) \end{aligned}$$

Intuitively, we can think of each unit's previous outcomes as control/pre-treatment variables. Conditional on them, the treatment assignment should be as good as random, so we can use doubly robust methods to flexibly control for the pre-treatment variables.

Therefore, there is a fair amount of symmetry between SCM and horizontal regressions. People often use SCM when there are a small number of units but a large number of time periods, and unconfoundedness type regression when are a small number of time periods but a large number of units.

Lastly, we can think of the difference-in-differences (DiD) regression in a two-way fixed effect regression context. After capturing the differences between units and time periods, we may use outcome modeling, essentially claiming that the potential outcome is a function of a constant, time fixed effect, unit fixed effect, and the unobservables. We eventually put together the counterfactuals by adding up the treated effects of the treated units and time periods.

To sum up the explanation above, the classic SCM (vertical regression) assumes a relation between different units that is stable over time; unconfoundedness regression (horizontal regression) assumes a relation between outcomes in treated period and pre-treatment periods that is the same for all units; and DiD regressions assume an additive outcome model that captures the differences between units and time periods. Synthetic diff-in-diff (SDID) is essentially an attempt to reconcile all three approaches above.

5.3.2 SDID Methodology

The data generating process in SDID is assumed to be:

$$\boldsymbol{Y} = \boldsymbol{L} + \tau \boldsymbol{W} + \epsilon, \tag{11}$$

where the matrix of outcome $\boldsymbol{Y}$ is represented as the sum between three components: some latent matrix $\boldsymbol{L}$ that could include two-way fixed effects; the product between treatment effect τ and a matrix of treatment indicators $\boldsymbol{W}$; and noise ϵ.

Here, our model essentially proclaims that the outcomes have *some* structure to it, but not necessarily an *additive* structure. Likewise, we allow treatment assignments to depend on the latent vectors, i.e. $\boldsymbol{L} \not\perp \boldsymbol{W}$. In other words, the treatment assignments have *some* structure, but we do not have to explicitly observe the factors that determine this structure. Meanwhile, our estimator would not require estimating the true factor matrix $\boldsymbol{L}$,. Surely, we can do explicitly estimate $\boldsymbol{L}$, but we ultimately want to target the counterfactual interest, so the problem at hand is not to simply estimate the whole $\boldsymbol{L}$. In fact, the SDID method would be better giving us the actual counterfactual effects of our interest better than a regular model that generically estimates $\boldsymbol{L}$ (Arkhangelsky et al. 2020).

To begin estimating SDID, we would first construct a modified version of synthetic control weights $\hat{w}$. To do so, we use penalized least squares to find a weighted average of control units with a pre-treatment trend parallel to the treated unit average:

$$(\hat{w}_0, \hat{w}) = \underset{\substack{w_0 \in \mathbb{R};\\ w_1,\ldots,w_{N_0} \geq 0;\\ \sum_{i \leq N_0} w_i = 1}}{\arg\min} \frac{1}{T_0} \sum_{t \leq T_0} \left(\bar{Y}_{N_0+1:N,t} - w_{\mathbb{Q}} - \sum_{i \leq N_0} w_i Y_{it} \right)^2 + \zeta^2 \left\| w \right\|^2$$

This equation solves the minimization problem. Instead of simple regularization, we use a Ridge regression penalty. One of the important changes compared to SCM is that we add in a constant term. In classical SCM, we try to construct each state as a weighted average of other states, which would work well if the state of interest lies in the convex hull of the other states. But otherwise the traditional SCM doesn't work as well, and this equation allows us to improve upon the traditional framework.

Then, we would apply the similar approach for time weights:

$$(\hat{\lambda}_0, \hat{\lambda}) = \underset{\substack{\lambda_0 \in \mathbb{R};\\ \lambda_1,\ldots,\lambda_{T_0} \geq 0;\\ \sum_{t \leq T_0} \lambda_t = 1}}{\arg\min} \frac{1}{N_0} \sum_{i \leq N_0} \left(\bar{Y}_{i,T_0+1:T} - \lambda_0 - \sum_{t \leq T_0} \lambda_t Y_{it} \right)^2$$

Note that this minimization equation is different from the one for the unit weights in the sense that we will not regularize the time weights. The units may be exchangeable, but the time periods may not be. We thus may have an incentive to have the model prioritizing some of the more recent time periods rather than the older ones.

Finally, we can compose a 2×2 diff-in-diff table with synthetic control and synthetic pre-treatment period to clearly showcase the differences of the estimators we're interested in:

	synthetic pre-treatment	average post-treatment
synthetic control	$\sum_{i \le N_0} \sum_{t \le T_0} \hat{w}_i \hat{\lambda}_t Y_{it}$	$\sum_{i \le N_0} \sum_{t > T_0} \hat{w}_i T_1^{-1} Y_{it}$
average treated	$\sum_{i > N_0} \sum_{t \le T_0} N_1^{-1} \hat{\lambda}_t Y_{it}$	$\sum_{i > N_0} \sum_{t > T_0} N_1^{-1} T_1^{-1} Y_{it}$

$$\hat{\tau}^{SC} = \underset{\tau, \gamma}{\arg\min} \sum_{i,t} (Y_{it} - \gamma_t - \tau W_{it})^2 \times w_i^{SC} \tag{12}$$

$$\hat{\tau}^{DID} = \underset{\tau, \gamma, \alpha}{\arg\min} \sum_{i,t} (Y_{it} - \gamma_t - \alpha_i - \tau W_{it})^2 \tag{13}$$

$$\hat{\tau}^{SDID} = \underset{\tau, \gamma, \alpha}{\arg\min} \sum_{i,t} (Y_{it} - \gamma_t - \alpha_i - \tau W_{it})^2 \times w_i \times \lambda_t \tag{14}$$

I have only one treated unit (i.e. Iceland), so the only implemented method to estimate the standard errors for the above coefficients is the "placebo" method described in Section 5 of Arkhangelsky et al. (2020). It is not the objective of this thesis to explore the econometrics of SDID, so I do not detail the logic for inference here.[12]

To sum up: the classic SCM framework is solving a regression problem where we have a fixed effects for time and weighting units constructed with SC weights, and we regress outcomes on time effects and treatment dummy. There is time effect but not unit effect, so time and unit are treated *asymmetrically*. DiD has both time and unit fixed effects but is an unweighted regression. SDID is weighted regression with both time and unit fixed effects.

SDID is double robust in the sense that we are estimating both an outcome model and weighting. If we think of SDID as a local fixed effect regression with time and unit, then it wouldn't matter as much whether we get the outcome model wrong if we can get the weights correctly. Meanwhile, if we get the outcome model right, then it wouldn't matter as much if the weights are correct. In the words of Susan Athey: "If a factor model holds, but the weights are good (e.g., ADH weights), SDID is consistent. If the DID model holds, but we use arbitrary weights, SDID is consistent."[13] Therefore, if the true data-generating process is a latent factor model, we're essentially estimating a simpler model while the true model should be richer. But even if we mis-specified our outcome model by just estimating a fixed effects model, we're still consistent. In practice, we might want to estimate a factor model, but we might not be able to get it exactly right. This allows us to be mis-specified in our factor model. This gives us good bias properties and theoretical properties in general.

[12] Both the coefficient and the standard errors can be easily calculated in R using the "synthdid" package provided by Arkhangelsky et al. (2020). For more information on the estimation process and theoretical backing, please see `https://synth-inference.github.io/synthdid/articles/synthdid.html`.

[13] Susan Athey's presentation titled "Synthetic Differences in Differences," page 24. Download link for the slides is: `https://www.aeaweb.org/conference/2020/preliminary/powerpoint/fnQr9QDs`.

5.3.3 SDID as Robustness Check

The motivation for using SDID as a robustness check for the classical synthetic controls framework should be even clearer after the methodological explanation. In short, the pre-trend match is often not good under the SCM framework regardless with or without cross-validation, and this inhibits us from isolating out the "diff-in-diff" effect, but we can often observe somewhat of an effect whether Iceland performed better or worse for that specific metric relative to its peers after 2011. SDID should be able to help us better isolate out this effect when the other estimators do not work well.

For example, suppose we do not observe any significant effect of the 2011 policy on Iceland' GDP per capita. However, suppose we do observe that if we were just to take out the part of the data after 2011 and compute a DiD estimate, most of the leave-one-out estimates would suggest that Iceland indeed performed better in terms of GDP per capita growth rate, which is a result that should signify that the 2011 policy did have a statistically significant effect. However, because of the limitations of SCM, it is simply too hard to construct an accurate pre-trend, and we would not accurately identify the policy effect. Therefore, the use of SDID is to try to isolate out that effect. If the SDID and SCM estimates differ from each other significantly, then it likely implies that the estimation is inconclusive. SDID thus serves as an important robustness check in my thesis.

Since I only use SDID for robustness checks, I believe that the information above should be sufficient in getting my reader up-to-date about the overall logic behind SDID, and I do not go into more details about the framework.

6 Empirical Strategy

In this section, I discuss the specific steps I take to estimate my models, the choices I make in the process, and the assumptions behind them. Section 6.1 provides an overview of how I decided on the predictor and outcome variables used for estimation. Section 6.2 gives my reasoning for choosing a certain estimation time period for my models. Section 6.3 explains the relationship between my main estimation results from SCM and subsequent robustness checks using cross-validation and SDID.

6.1 Variables of Interest

I now explain the rationale behind my variable choices. For the main outcomes of interest, because the research question I am after is about the trajectory of overall economy after Iceland enacted the debt forgiveness program, it seems most important to first zoom in on the major macroeconomic outcomes. Unemployment rate is arguably the most important measure of slack in the aggregate economy, since it is updated timely and the measurement methodology for unemployment rate is fairly consistent and accurate across economies. GDP per capita, on the other hand, could suffer from measurement issues, so I will prioritize unemployment rate over GDP per capita even though the latter variable appears more often in SCM studies.

Then, I will move to examine the financial sector variables, namely non-performing loan (NPL) ratio, credit to private sector, and international debt to GDP ratio. As shown in Section 4.2 and Table A1, these measures make Iceland stand out in comparison with the OECD average, so they should somewhat accurately capture a snapshot of Iceland's financial conditions after 2011.

As quick background, a non-performing loan (NPL) is a loan in which the borrower is considered to be in default status, which could be a result of them not having made the scheduled payments for a specified period of time or is unlikely to be paid back in full. NPL has been an important metric in measuring the health of the banking sector. Caballero et al. (2008) studied the role of misdirected bank lending in prolonging the economic stagnation. Banks lent cheap credit to insolvent borrowers and zombie firms that should be failing. The authors particularly zoomed in on NPL, and one of their explanations for the misdirected bank lending is that banks did not want to see rising NPL, so they decided to "ever-green" the bad loans and counted on the government to bail out these zombie firms eventually. By keeping "zombie" firms alive through cheap credit, the banks distorted competition throughout the rest of the economy. This paper is one of the numerous examples in the macrofinancial literature of how NPL is an important metric in judging financial health.

Likewise, international debt is especially noteworthy in the Icelandic context because the Icelandic banking system was intimately connected with foreign creditors. As explained in Section 2, a key component of Iceland's 2011 programs was to allow the banks default on its foreign obligations. For this thesis, international private debt matters more than public debt because the build-up of foreign credit mainly concentrated amongst the private counterparties, and it was actually only after 2008 that Iceland began to gradually accrue more public debt.

I do include income inequality as part of my study, but it is likely not a first-order outcome that fiscal and monetary policymakers consider when they design policy responses to crises, so it is hard to extrapolate the debt jubilee's *immediate* effect on income inequality. I use the income share data from the World Inequality Databse to proxy income inequality, which could be an inaccurate measurement given how it is a more macro-level variable. It also does not differentiate between the households that received the "110% option" assistance from those that did not, so it would not clearly identify the channels of income effects. So, to draw any definitive conclusion about income inequality, I would likely need to change my empirical strategy and collect more micro-level data to make my model more fitting for identifying specific channels that influence inequality. In sum, I would caveat here that though I include income share estimations in my results, it is not a central mission of my thesis to examine the debt jubilee's impact on income inequality. I feel relatively comfortable leaving this as a question that needs further research by me and other scholars.

As for the characteristic variables I use to estimate the outcomes above, I use six sets of predictors. In the context of my research, Iceland is the treatment unit, and the other OECD countries are used to construct the synthetic Iceland. The challenge, however, is that Iceland is a unique small open economy that does not seem to yield great pre-trend fits based on most of the existing models and literature, as later results would demonstrate. One way I attempted to address the imperfect pre-trend match is by trying out multiple constructions of the predictor variable groups. After either replicating the various papers above or improving upon their constructions based on my context, I ended up with six groups of predictors that I try out to construct my synthetic Iceland:

- Macroeconomic characteristic variables: GDP per capita, imports as % of GDP, exports as % of GDP, industry share, domestic credit to private sector as % of GDP, inflation rate.
- Macrofinancial characteristic variables: GDP per capita, imports as % of GDP, exports as % of GDP, industry share of value added (including construction), credit to private sector, bank concentration, gross domestic savings as % of GDP, international debt issues as % of GDP.

- Replicating Born et al. (2019): imports as % of GDP, exports as % of GDP, employment to population ratio, labor productivity growth, household consumption as % of GDP, gross capital formation (% of GDP).
- Replicating Abadie et al. (2015): GDP per capita, trade as % of GDP, inflation, industry share, gross domestic savings as % of GDP, population share that have received post-secondary education.
- Banking and housing related characteristic variables: credit to private sector, bank concentration, gross domestic savings as % of GDP, housing price growth from 2003 to 2008.
- Labor market related characteristic variables: labor force participation rate, total (% of total population ages 15-64), labor productivity growth rate, industry share, trade as % of GDP, population share that have received post-secondary education, employment in agriculture (% of total employment).

Six sets of predictors should be sufficient in showing that I have gone through the proper "trial-and-error" process of choosing SCM predictors so that my estimation results below are robust in the sense that changing predictors alone would not significantly alter my conclusions. I do not present the results from all six predictor groups for all six of my outcomes, and I simply present the one construction that give the best estimation results. For more information on robustness checks, see Appendix B.

6.2 Optimization Period Selection

All characteristic predictors are averaged over the years of 1999 to 2006. The consideration is that 1999 is the year that Euro was adopted; though Iceland does not use the Euro but its own currency of króna, 1999 is a structural break for most other OECD countries, so I start my estimation period with it. 2006 marks the "precursor period" that will soon lead to the 2008 financial crisis. Including the 2007-2008 years in the optimization period could potentially yield even more inaccurate pre-trends, and I believe that 1999-2006 would be a fair selection for the pre-treatment period, even though Iceland technically only started enacting its debt forgiveness program starting in 2011. So, the entire results period will be 1999-2017, and 1999-2006 will be my pre-trend period.

In my studies, I do not differentiate between the pre-trend period and the period over which RMSPE is minimized – I see no reason why they would need to be different in the context of my study, and changing them did not seem to yield more accurate or dramatically different results.

To further check for the robustness of my SCM estimates, I implement the cross-validation technique, whose results are showcased in Appendix B. For all my cross-validation estimations, I divide the pre-treatment years into a training period from 1999 to 2005 and a validation period from 2006 to 2011. Then, with predictors measured in the training period, I select the weights v_m such that the resulting synthetic control minimizes the RMSPE over the validation period. This is largely in line with how Abadie et al. (2015) had decided on their training and validation periods, though their sample time periods span across a longer range of years than mine. As previously explained, the cross-validation technique selects the weights v_m that would minimize out-of-sample prediction errors. And such a set of v_m weights and predictor data measured 2006 and 2011 can help construct the final synthetic Iceland.

6.3 Relationship between Main Results and Robustness Checks

One may have guessed from the three subsections of the methodology chapter that I would employ three different identification and estimation strategies for this study. That is indeed what I have done during the process of my research, but I will not showcase all the results in Section 7. I primarily focus on classical SCM as my main empirical methodology, and then use SCM with cross-validation and SDID as robustness checks for it.

Specifically, for every outcome variable, I first outline my estimation results from SCM and their implications. This includes plotting out the SCM trajectory graphs with synthetic Iceland and actual Iceland overlaid and the graph with the distribution of in-place placebo estimates. After estimating the SCM effects, I also determine their statistical significance by running placebo tests and generating p-values, whose specific results are shown in Appendix B.

For robustness checks, I would first estimate the same synthetic controls model using the same set of predictor outcomes shown in Section 7 but this time with the cross-validation technique to choose the $\boldsymbol{V}$ and $\boldsymbol{W}$ weights. Then, I conduct a SDID estimation. The motivation for using SDID has been previously explained.

In sum, I derive my results primarily from the classical SCM framework and use cross-validation and SDID as robustness checks to make sure my thesis is relatively focused around one central methodology. Nevertheless, the total content of this thesis ends up providing a broad overview of the three estimation frameworks and their applications for readers that are more interested in surveying the different applications and econometric alternatives related to synthetic controls.

7 Results

In Section 7.1, I give a summary of my main results from Section 7.2-7.7, in which I present more detailed findings for each of my six outcomes of interest. The results in this section also occasionally reference the robustness checks in Appendix B.

7.1 Summary of Results

Figures 1 to 6 display the SCM trajectory of Iceland vs. its synthetic counterpart and the distribution of in-place placebo estimates for the six different outcome variables. Tables 1 to 12 showcase the predictor means and synthetic weights for respective SCM estimations shown in the figures. These graphs and tables capture the main results of this thesis. Before I dive into the specifics in each subsection, I want to outline some of my overarching conclusions:

- Unemployment: Iceland's unemployment rate was 2 percentage points lower than its synthetic counterpart within 2 years of the 2011 debt jubilee and 1 percentage point lower within 5 years. In other words, unemployment fell more quickly after the jubilee than had the policy not been in place, even though the pre-trend counterfactual does not match perfectly for years before the 2008 financial crisis because Iceland had structurally lower unemployment rate compared to its OECD peers.
- GDP per capita: Iceland's GDP per capita fell around \$5,000 in the aftermath of the 2008 crisis but rapidly recovered after 2011, eventually rising to be \$2,500 higher than its synthetic peer by 2016.

The jubilee program likely kicked off the recovery process two years sooner for Iceland than had the policy not been in place. However, the estimation results are not statistically significant at the 90% level. There may simply be too many other confounding factors that contribute to Iceland's recovery, but overall it seems reasonable to conclude that Iceland's 2011 policies indeed helped with the GDP recovery.

- Non-performing loan (NPL): Iceland's NPL ratio declined 15 percentage points within around 3 years after the debt jubilee, while its synthetic counterpart saw NPL ratio raising by almost 10 percentage points in the same time period. So, the total treatment effect was more than 20 percentage points within 3-5 years of the policy implementation. The results are statistically significant at the 90% level and validated by additional robustness checks. The OECD countries can be quite evenly divided between those that had and did not have a dramatic run-up in NPL ratios around 2008. Amongst those that did, we observe that Iceland was particularly successfully in bringing down its NPL ratio compared to its peers like Ireland and Latvia that did not enact similar bank restructuring and debt forgiveness policies. It is also noteworthy that Iceland was different from many other OECD countries in the sense that it started building up higher levels of NPL ratios around 2006 and before the actual financial crisis took place. So, the fact that Iceland's NPL ratio fell back down to not just pre-2008, but pre-2006 levels after the 2011 policies also testifies to the efficacy of the programs.

- Private credit: Iceland's private credit to GDP ratio was reduced by 50 more percentage points than its synthetic counterpart within 4 years of the debt jubilee. Iceland had the single most dramatic credit boom amongst all OECD countries between 2003 and 2008. Even though it was difficult to construct an accurate pre-trend match given the high credit boom level, the robust results of our estimates suggest that the 2011 programs did not just stabilize Iceland's private credit level; it actively reduced it further at a rapid pace.

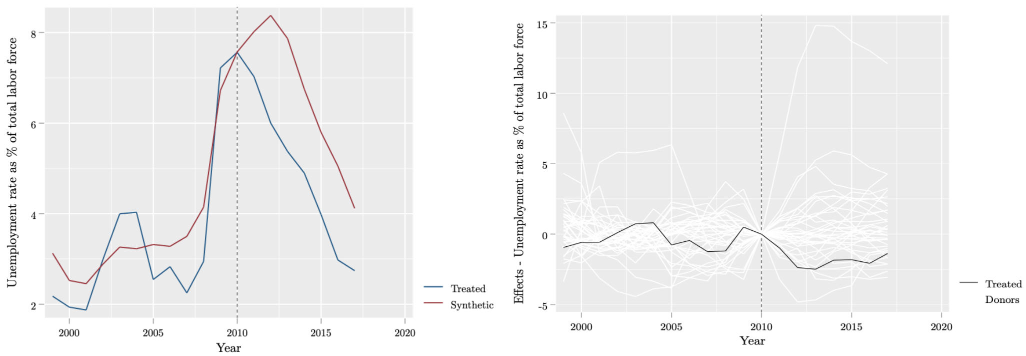
- International debt: Iceland's debt build-up was just as dramatic as that for NPL and private credit. The international private debt to GDP ratio ballooned from around 100% in the early 2000s to more than 350% by 2009, which was the highest amongst all OECD countries and around 7 times the median OECD level of around 50%. After the 2011 debt jubilee, Iceland's international private debt to GDP ratio declined from 300% to 100% within around 3 years and eventually to 50% by 2017. Synthetic Iceland, on the other hand, saw its debt staying around the 300% level and eventually rising to 400% by 2017. In other words, the total treatment effect was more than 200 percentage points within 3 years of the policy implementation and more than 300 percentage points after 6 years.

- Income inequality: After 2008, Iceland's top 10% income share rose 20 percentage points to be amongst of the highest in OECD, and its middle 40% and bottom 50% income share each fell 10 percentage points to be amongst the lowest. The income share gap soon normalized after 2011 and fell back to pre-2008 levels within a year or two, but the results are not robust enough to attribute this income inequality reduction to the 2011 debt jubilee.

In sum, Iceland's macroeconomic conditions proxied by unemployment rate and GDP per capita both improved after the debt jubilee. It also did especially well in improving metrics related to the financial sector such as bringing down NPL ratio, reducing international debt, and restoring private credit to normalcy. In the subsections below, I now dive into the more specific estimation results and logic that led me to the headline conclusions above.

7.2 Unemployment

Figure 1: Synthetic controls trajectory for unemployment rate: Iceland vs. synthetic Iceland



(a) SCM trajectory

(b) In-place placebo estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland vs. synthetic Iceland. Figure (b) plots the gap between Iceland and synthetic Iceland in a "leave-one-out" in-place placebo setting, where I iteratively apply the synthetic method to each country in the donor pool to obtain a distribution of placebo effects.

Table 1: Predictor means before treatment for unemployment rate as outcome variable

Variable	**Treated**	**Synthetic**
Unemployment rate	2.797625	3.1385
Labor participation rate	86.93287	62.52675
Labor productivity growth rate	11.11063	10.03526
Industry share	22.21006	33.16032
Trade as % of GDP	71.8698	51.32967
Employment in agriculture (%)	7.3575	16.71475
Share of post-secondary education completion	31.00324	10.58858

Notes: The "treated" column indicates Iceland's actual data, while the "synthetic" column indicates the weighted average calculated through the synthetic control method. All predictors are averaged for the 1999-2006 period, except the share of post-secondary education completion is only considered for year 2006. All variables are in % unit. See Section 6.1 for more information on the logic behind variable selection and Appendix A for more on variable definitions.

Table 2: Synthetic weights for Iceland for GDP per capita as outcome variable

Country	**Unit weights**
Luxembourg	0.267
Mexico	0.733

Notes: The synthetic weight is the country weight assigned by the synthetic controls method. See Section 5.1 for details on the SCM methodology.

Figure 1a displays the main result that Iceland's unemployment rate was 2 percentage points lower than its synthetic counterpart within 2 years of the debt jubilee and 1 percentage point lower within 5 years. In other words, unemployment fell more quickly after the jubilee than had the policy not been in place, even though the pre-trend counterfactual does not match perfectly for years before the 2008 financial crisis. The effects gap shown in the placebo estimation graph in Figure 1b demonstrates that Iceland performed much better in bringing down the unemployment level after 2011 than its placebo peers after the treatment. I would consider the effect of the debt jubilee on Iceland to be significant since the estimate effect for Iceland is unusually large relative to the distribution of placebo effects.

The *p*-values for this estimation are between 0.2 to 0.3, i.e. the probability that the improvement happened by chance is around 20-30% (Figure B21). The other robustness checks with cross-validation (Figure B22) and SDID (Figure B23) confirm the observation that Iceland's unemployment rate dropped faster than synthetic Iceland. In sum, Iceland's post-2011 unemployment was indeed one of the lowest compared to its OECD peers, but the statistical significance of the policy effect estimation is not as strong as we ideally wish it to be, such as at the 95% level.

Nevertheless, I do acknowledge that while the 2011 debt jubilee program likely have substantively helped Iceland's unemployment rate reduction, another factor contributing to the reduction could be that Iceland was simply returning to its structurally lower level of unemployment before the 2008 financial crisis (Figure A2). The structural reasons such as Iceland historically enjoying much lower unemployment rate and much faster run-up of unemployment during the financial crisis could also help explain why it is difficult to find a good pre-trend match for unemployment rate. To dig further, I would make the observation that Iceland unemployment rate experienced three different stages between 1999 and 2017:

1. Pre-2008: Iceland had structurally low, though somewhat volatile, unemployment rates before 2008, mostly varying between 2-4% and experiencing historically low levels of unemployment rate right before 2008.

2. 2008-2011: When the 2008 financial crisis hit, Iceland's unemployment increased dramatically and arguably more dramatically than most of its OECD counterparts, shooting up to near 8%.

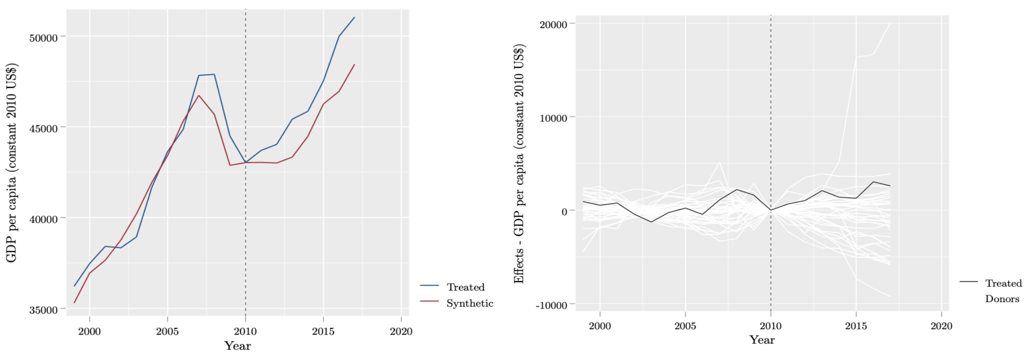
3. Post-2011: Starting around 2011, as the economy gradually recovered, the unemployment rate started to rapidly fall, eventually coming back to pre-2008 levels around 2016 and later.

The nature of SCM is to construct a synthetic Iceland that resembled this path, but it is in fact quite difficult to find a combination of countries that had experienced similar trajectories. The wealthier OECD countries like the U.S., Germany, U.K., and France enjoyed relatively lower unemployment rates before the 2008, but their unemployment rate never increased so dramatically like Iceland when 2008 hit. The poorer countries like Ireland, Latvia, Greece, Italy and the like had dramatic rises in unemployment after 2008, but they did not enjoy the structurally low unemployment rate and good management of the economy like Iceland did before 2008. Therefore, what we observe amongst our SCM trajectories is that the pre-trend match is either good before 2008 (fundamentally wealthier and better managed economies) or after 2008 (i.e. a good fit for the run-up of unemployment as the dramatic fallout of the crisis), but it's hard to construct a synthetic Iceland that resembled both periods.

In sum, I would conclude that Iceland experienced a more dramatic decline in unemployment rate after 2011 than synthetic Iceland, likely because of the effectiveness of the debt jubilee, but its low unemployment rate level before 2008 likely also contributed to the decline, and such structural factors also make it more difficult to construct a synthetic Iceland that would perfectly capture the pre-trend of Iceland's unemployment rate before the 2011 treatment.

7.3 GDP per capita

Figure 2: Synthetic controls trajectory for GDP per capita: Iceland vs. synthetic Iceland



(a) SCM trajectory

(b) In-place placebo estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland vs. synthetic Iceland. Figure (b) plots the gap between Iceland and synthetic Iceland in a "leave-one-out" in-place placebo setting, where I iteratively apply the synthetic method to each country in the donor pool to obtain a distribution of placebo effects.

Figure 2a displays the per capita GDP trajectory of Iceland vs. its synthetic counterpart from 1999 to 2017. The estimates show that Iceland's GDP per capita fell around \$5,000 in the aftermath of the 2008 crisis but rapidly recovered after 2011, eventually rising to be \$2,500 higher than its synthetic peer by 2016. Here, we see that a decent pre-trend fit and that Iceland's GDP per capita did immediately react to the 2011 policy intervention while its synthetic counterpart flattened in the first 1-2 years. From 2013 and onwards, the two trend lines began to recover at roughly the same pace, but because Iceland's economy started recovering earlier, it eventually arrived at a higher GDP per capita level years later. Another way to interpret this is that the 2011 jubilee program likely kicked off the recovery process two years sooner for Iceland than had the policy not been in place.

Figure 2b shows the gaps between the treated and control units and the distribution of in-place placebo estimates. The results here indicate that Iceland performs around the median level when other countries replaced its place. Meanwhile, Figure B24 display the p-value estimated for this regression is between 0.4 to 0.7 for the coming years. The result here give a much more pessimistic outlook of my estimation. The probability that my estimation would happen by chance, as indicated by the p-values, could be as high as 40% even 4 years after the event. In contrast, running the same placebo inference on the Abadie et al. (2010) study about cigarette sales shows a probability of below 0.1 consistently across the years (Galiani and Quistorff 2017). In sum, the results for GDP per capita are not statistically significant.

Table 3: Predictor means before treatment for GDP per capita as outcome variable

Variable	**Treated**	**Synthetic**
GDP per capita	39941.79	39959.67
Trade as % of GDP	71.8698	71.90666
Industry share	22.21006	22.25233
Inflation for consumers	4.482299	4.470365
Capital formation	25.38841	25.37863
Share of post-secondary education completion in 2006	31.00324	31.001

Notes: The "treated" column indicates Iceland's actual data, while the "synthetic" column indicates the weighted average calculated through the synthetic control method. All predictors are averaged for the 1999-2006 period, except the share of post-secondary education completion is only considered for year 2006. GDP per capita is measured in real dollar terms and expressed in absolute levels; all other variables are in % unit. See Section 6.1 for more information on the logic behind variable selection and Appendix A for more on variable definitions.

Table 4: Synthetic weights for Iceland for GDP per capita as outcome variable

Country	**Unit weights**
Latvia	0.341
Luxembourg	0.064
Switzerland	0.146
United States	0.449

Notes: The synthetic weight is the country weight assigned by the synthetic controls method. See Section 5.1 for details on the SCM methodology.

The results in Table 3 and 4 correspond to the predictor match and synthetic weights for Figure 2. This set of variable selection is quite representative of the synthetic matches using other predictors. We see a fairly good match of the set of characteristic predictors between Iceland and synthetic Iceland, especially in GDP per capita, imports and exports ratios, and industry share. Measures for inflation and private credit do not exhibit good matches, likely because Iceland experienced structurally higher inflation and credit boom. It should be well expected and reasonable, however, that these variables cannot be perfectly fitted using the weighted averages – we see from later results that credit boom does not fit well as an outcome variable either, which implies that there could be structural factors prohibiting good pre-trend matches, and it can not be easily improved by changing out predictor selections.

Meanwhile, synthetic Iceland is a combination of Latvia, Luxembourg, Switzerland, and the United States. All other countries in the donor pool do not contribute to the construction. This makes sense because Latvia also experienced a dramatic credit boom before 2008, and the financial sectors occupy a large part of Luxembourg and Switzerland's economies; the U.S. is included likely because of the similarity in GDP per capita and industry share levels.

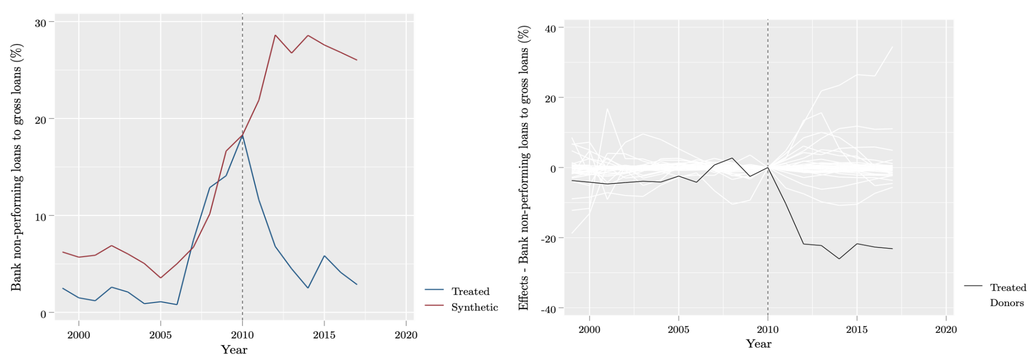
Cross-validation does not seem to improve the statistical significance but do validate the results of my classical SCM estimation. Results of the cross-validation procedure are showcased in Figure B25 and Table B5 to B6. Cross-validation and "in-time" placebo estimates seem to suggest that 2011 indeed had an effect on Iceland's GDP per capita in the sense that Iceland had a more rapidly increasing per-capita GDP before the 2008 crisis, followed by a sharp drop and a quick bounce-back to be growing faster than its peers again. In other words, Iceland may have suffered more than its synthetic peers immediately after 2008, but it did recover and catch up more quickly after 2011.

Lastly, Figure B26 displays the results from the SDID estimation for GDP per capita. We may observe that SDID has successfully captured the parallel trend much better than SCM, but the post-treatment effects are not statistically significant. It seems that SDID is simply making the previous SCM estimations more accurate and the effects more pronounced. In other words, SDID successfully captured the "diff-in-diff" relationship better than the other estimators; it's just that the effect of GDP per capita is not significant.

All in all, the results suggest that the various predictor constructions all give relatively decent but still imperfect pre-trend SCM fits, with or without cross-validation. The post-intervention trends and placebo estimates may suggest that Iceland did better in terms of GDP per capita after the 2011 debt forgiveness program relative to those that did not enact a similar program, but the standard errors are high. In other words, Iceland's GDP per capita performed better compared to its synthetic peer, but the likelihood of this happening by chance or due to factors other than the debt jubilee *per se* is still as high as 40%, and the SDID analysis confirms that the estimates are not statistically significant at the 95% level.

7.4 Non-Performing Loans (NPL)

Figure 3: Synthetic controls trajectory for NPL: Iceland vs. synthetic Iceland



(a) SCM trajectory

(b) In-place placebo estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland vs. synthetic Iceland. Figure (b) plots the gap between Iceland and synthetic Iceland in a "leave-one-out" in-place placebo setting, where I iteratively apply the synthetic method to each country in the donor pool to obtain a distribution of placebo effects.

Table 5: Predictor means before treatment for NPL as outcome variable

Variable	**Treated**	**Synthetic**
NPL ratio	1.5875	1.649137
GDP per capita	39941.79	44243.56
Imports as % of GDP	38.9257	38.93082
Exports as % of GDP	32.9441	41.1478
Industry share	22.21006	22.07893
Credit to private sector as % GDP	149.1479	114.6257
Bank concentration (%)	99.76052	68.19336
Domestic savings as % GDP	19.40681	23.82372
International debt to GDP ratio	170.4484	59.16293

Notes: The "treated" column indicates Iceland's actual data, while the "synthetic" column indicates the weighted average calculated through the synthetic control method. All predictors are averaged for the 1999-2006 period. GDP per capita is measured in real dollar terms and expressed in absolute levels; all other variables are in % unit. See Section 6.1 for more information on the logic behind variable selection and Appendix A for more on variable definitions.

Table 6: Synthetic weights for Iceland for NPL as outcome variable

Country	**Unit weights**
Australia	0.243
Netherlands	0.461
United Kingdom	0.296

Notes: The synthetic weight is the country weight assigned by the synthetic controls method. See Section 5.1 for details on the SCM methodology.

Figure 3 display the main result that Iceland's NPL ratio declined 15 percentage points within around 3 years after the debt jubilee, while its synthetic counterpart saw NPL ratio raising by almost 10 percentage points in the same time period. In other words, the total treatment effect was more than 20 percentage points within 3-5 years of the policy implementation. The placebo estimation graph also demonstrates that Iceland's decline in NPL ratio was in fact the largest amongst all its placebo peers and significantly larger relative to the others, which further convinces us that the policy effect was significant.

The p-values for this estimation is less than 0.1 throughout the years, i.e. the probability that the improvement happened by chance is less than 10% (Figure B27). The other robustness checks with cross-validation (Figure B28) and SDID (Figure B29) confirm that Iceland's NPL ratio dropped faster than synthetic Iceland. In sum, Iceland's post-2011 NPL ratio reduction more significant compared to its OECD peers.

I recognize that it is particularly hard to construct a good pre-trend match for NPL because of Iceland's dramatic NPL boom before 2008. Indeed, Iceland's NPL growth trajectory between 2006 and 2010 was literally amongst the most dramatic amongst all OECD countries. It had a very unusual banking system before 2008 especially given its leverage with foreign debt, and it also got very large relative to its economy in comparison to its OECD peers.

However, these structural factors should not hinder us from confidently conclude that Iceland outperformed its peers. From the trajectory graphs in Figure A4, we can observe that even amongst all the countries that had dramatic run-up in NPL ratios around 2008, Iceland was particularly successful in bringing down its

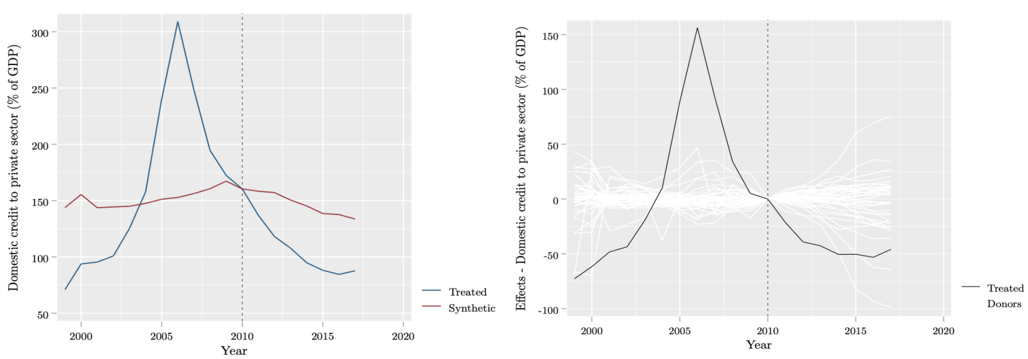
NPL ratio. In other words, amongst the countries that suffered from high NPL ratios, Iceland did much better compared to its peers like Ireland and Latvia that did not enact similar bank restructuring and debt forgiveness policies. Most other OECD countries did not suffer from the same problem of high NPL and thus would not serve as a good synthetic peer in the first place, but compared to those that did, Iceland still performed notably better.

Additionally, another important factor to consider for NPL is that the rise of NPL ratio started around 2005 and much earlier than the 2008 financial crisis. This distinguishes NPL from other outcome variables like GDP per capita and unemployment, which did not worsen until 2008. The banking sector, as represented by NPL in our simplified portrayal, likely started having structural buildups of bad loans and systemic issues much earlier than 2008. Such a credit boom and mismanagement of the financial sector, however, would not be immediately reflected in the real economy until the blowup. Therefore, while some might say that Iceland's macroeconomy simply returned to its pre-2008 level after 2011 and not solely due to the efficacy of the 2011 programs, it is much more unlikely for that to be true for NPL and credit boom, which started as early as 2005, so the 2011 policies not only had to bring down these financial metrics back to their pre-2008 levels, but rather actually back to their pre-2005 "normal" levels that are in line with the rest of the OECD countries. In other words, the fact that the NPL ratio and credit boom subdued all the way back to levels even lower than right before the 2008 financial crisis likely shows that the 2011 policies indeed worked, or else the dramatic decline would not have happened.

All in all, it seems that we should be able to confidently conclude that Iceland's 2011 debt forgiveness program was particularly effective for dramatically bringing down the NPL ratio, outperforming its synthetic peer by more than 20 percentage points in 3-5 years following the treatment.

7.5 Private Credit

Figure 4: Synthetic controls trajectory for private credit: Iceland vs. synthetic Iceland



(a) SCM trajectory

(b) In-place placebo estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland vs. synthetic Iceland. Figure (b) plots the gap between Iceland and synthetic Iceland in a "leave-one-out" in-place placebo setting, where I iteratively apply the synthetic method to each country in the donor pool to obtain a distribution of placebo effects.

Table 7: Predictor means before treatment for private credit as outcome variable

Variable	**Treated**	**Synthetic**
Credit to private sector as % GDP	149.1479	146.631
NPL ratio	1.5875	1.169062
GDP per capita	39941.79	40952.75
Imports as % of GDP	38.9257	37.89874
Exports as % of GDP	32.9441	41.28095
Industry share	22.21006	27.78365
Bank concentration (%)	99.76052	71.67848
Domestic savings as % GDP	19.40681	24.85396
International debt to GDP ratio	170.4484	33.13433

Notes: The "treated" column indicates Iceland's actual data, while the "synthetic" column indicates the weighted average calculated through the synthetic control method. All predictors are averaged for the 1999-2006 period. GDP per capita is measured in real dollar terms and expressed in absolute levels; all other variables are in % unit. See Section 6.1 for more information on the logic behind variable selection and Appendix A for more on variable definitions.

Table 8: Synthetic weights for Iceland for private credit as outcome variable

Country	**Unit weights**
Canada	0.844
Luxembourg	0.028
Portugal	0.118
Switzerland	0.01

Notes: The synthetic weight is the country weight assigned by the synthetic controls method. See Section 5.1 for details on the SCM methodology.

Results for private credit as the outcome variable are shown in Figure 4. Iceland's private credit to GDP ratio was reduced by 50 more percentage points than its synthetic counterpart within 4 years of the debt jubilee. The placebo estimation graph also demonstrates that Iceland's decline in private credit was in fact the largest amongst all its placebo peers and significantly larger relative to the others. The *p*-values are less than 0.1 for the first 4 years after the shock and then remains 0.2 for the later years, i.e. the probability that the improvement happened by chance is less than 10% (Figure B30). The other robustness checks with cross-validation (Figure B31) and SDID (Figure B32) confirm that Iceland's private credit ratio decline was statistically significant, though the SDID estimator gives a smaller point estimate than SCM.

One may question the robustness of my conclusion given how we see arguably the most inaccurate pre-trend match for private credit amongst all outcomes throughout this paper. The lack of good pre-trend match is mostly because of the dramatic rise and fall of the credit boom before and after 2008. It is not an exaggeration to say that Iceland had the single most dramatic credit boom amongst all OECD countries between 2003 and 2008, reaching unprecedented levels unmatched by any of its peers (Figure A5). Under these structural factors, when the absolute level of its private credit fluctuated between 0.5-3 times of most of the other OECD countries, it is simply impossible to construct an accurate pre-trend for Iceland.

However, I do not think the imperfect pre-trend match is a major issue that can bring down my conclusion. I would counter the above concerns by digging deeper into the economic intuition behind the rise and fall of Iceland's private credit: While most OECD countries maintained a relatively stable level of private credit to

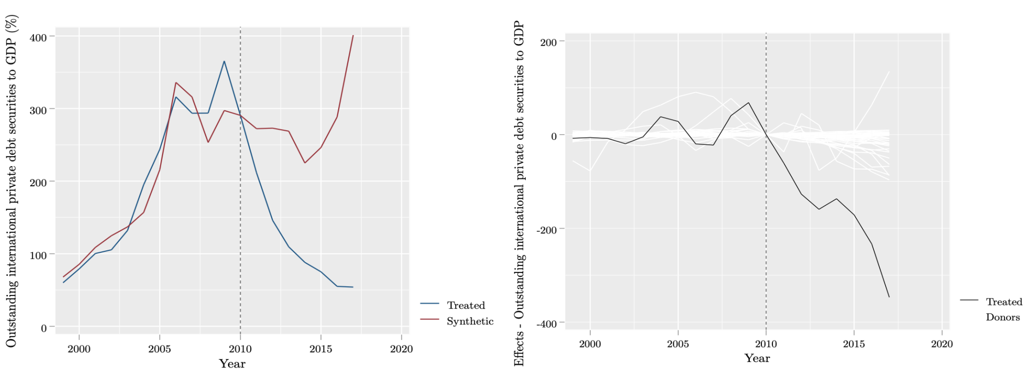
GDP ratio throughout the years and only experienced a 25-50% expansion around the 2008 years, Iceland actually experienced much lower credit flow compared to the rest of the OECD (around 50% of the median) before 2003. But starting around 2003, Iceland dialed up its private credit flow, eventually pushing it up by 4 times within a span of 4-5 years by 2008. When the financial crisis hit, the credit flow actually reduced dramatically, and by 2011 it was back to around the same level experienced by the rest of the OECD countries.

What is especially noteworthy, nevertheless, is that after the implementation of the 2011 programs, the private credit flow *continued* to rapidly decline, eventually reaching to about half of its 2011 level by 2015 and back to its early 2000s level. That is how by 2015, Iceland actually had one of the lowest level of credit flows compared to its OECD peers. In other words, by 2011 Iceland's private credit level was already much lower than the 2008 peak; the consequence of the debt jubilee was not to just keep Iceland's private credit at a stable level; it actively reduced it further at a rapid pace. If the jubilee did not work, we would not be observing a further reduction in Iceland's private credit level after 2011.

It is another discussion whether such a low level of credit would be healthy for the growth of the economy, but it seems that Iceland's 2011 policies indeed changed the fundamental structure of the economy to be less reliant on credit booms. The dramatic boom before 2008 led to a dramatic bust after 2008, and in some sense the bust and subsequent measure to further reduce it was arguably equally interesting to observe as its initial boom. I am confident to conclude that the 2011 debt restructuring program had a substantive impact on further reducing private debt to GDP ratio by around 50 percentage points.

7.6 International Debt

Figure 5: Synthetic controls trajectory for international debt to GDP ratio: Iceland vs. synthetic Iceland



(a) SCM trajectory

(b) In-place placebo estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland vs. synthetic Iceland. Figure (b) plots the gap between Iceland and synthetic Iceland in a "leave-one-out" in-place placebo setting, where I iteratively apply the synthetic method to each country in the donor pool to obtain a distribution of placebo effects.

Table 9: Predictor means before treatment for international private debt as outcome variable

Variable	**Treated**	**Synthetic**
International private debt as % GDP	149.1479	146.631
GDP per capita	39941.79	39454.88
Imports as % of GDP	38.9257	38.17483
Exports as % of GDP	32.9441	34.26579
Inflation rate	4.482299	2.695202
Industry share	22.21006	22.17661
Gross capital formation as % GDP	25.38841	25.43427
Share of post-secondary education completion	31.00324	30.62283

Notes: The "treated" column indicates Iceland's actual data, while the "synthetic" column indicates the weighted average calculated through the synthetic control method. All predictors are averaged for the 1999-2006 period, except the share of post-secondary education completion is only considered for year 2006. GDP per capita is measured in real dollar terms and expressed in absolute levels; all other variables are in % unit. See Section 6.1 for more information on the logic behind variable selection and Appendix A for more on variable definitions.

Table 10: Synthetic weights for Iceland for international private debt as outcome variable

Country	**Unit weights**
Greece	0.219
Latvia	0.254
Luxembourg	0.032
Switzerland	0.25
United States	0.245

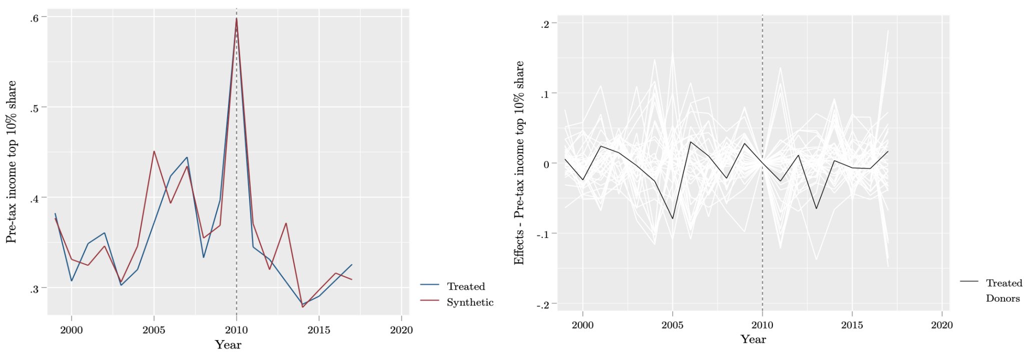
Notes: The synthetic weight is the country weight assigned by the synthetic controls method. See Section 5.1 for details on the SCM methodology.

Figure 5 display the main result that Iceland's international debt to GDP ratio declined from 300% to 100% within around 3 years after the debt jubilee and eventually to 50% by 2017. Synthetic Iceland, on the other hand, saw its debt staying around the 300% level and eventually rising to 400% by 2017. In other words, the total treatment effect was more than 200 percentage points within 3 years of the policy implementation and more than 300 percentage points after 6 years. The placebo estimation graph also demonstrates that Iceland's decline in debt ratio was in fact the largest amongst all its placebo peers and significantly quicker relative to the others, which further convinces us that the policy effect was significant. The p-values for this estimation is around 0.2 throughout the years, i.e. the probability that the improvement happened by chance is only around 20% (Figure B33). The other robustness checks with cross-validation (Figure B34) and SDID (Figure B35) confirm that Iceland's debt ratio dropped significantly faster than synthetic Iceland.

We previously saw from some of the aggregate trends graphs that Iceland's debt build-up was just as dramatic as that for NPL and private credit (Figure A6 and A9). The international private debt to GDP ratio ballooned from around 100% in the early 2000s (which was already high in comparison to most other OECD countries) to more than 300% by 2008, eventually peaking at more than 350% by 2009. This level of international debt was the highest amongst all OECD countries and around 7 times the median OECD level around 50%, so it was again impossible to construct a nearly accurate synthetic Iceland for the pre-trend. However, just like the case for private credit, I am still confident to conclude from the various estimations and robustness checks that the debt level did come down dramatically after 2011.

7.7 Income Inequality

Figure 6: Synthetic controls trajectory for top 10% income share ratio: Iceland vs. synthetic Iceland

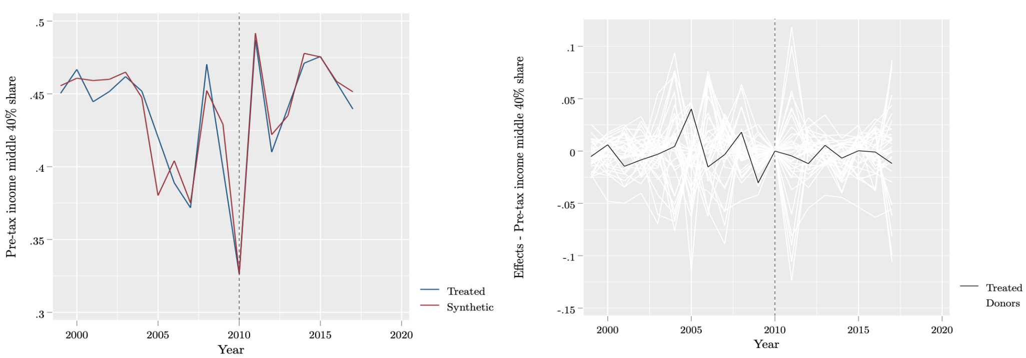


(a) SCM trajectory

(b) In-place placebo estimates

Notes: The y-axis scale is normalized to 1, so 0.5 means 50% income share. Figure (a) shows the synthetic controls estimation trajectory for Iceland vs. synthetic Iceland. Figure (b) plots the gap between Iceland and synthetic Iceland in a "leave-one-out" in-place placebo setting, where I iteratively apply the synthetic method to each country in the donor pool to obtain a distribution of placebo effects.

Figure 7: Synthetic controls trajectory for middle 40% income share ratio: Iceland vs. synthetic Iceland

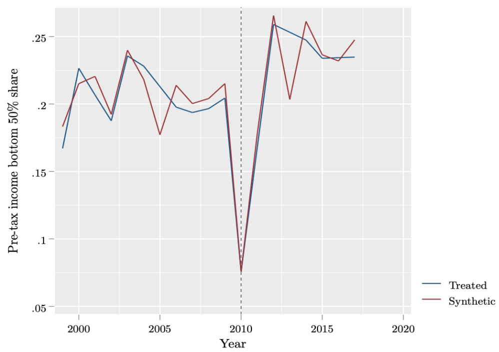


(a) SCM trajectory

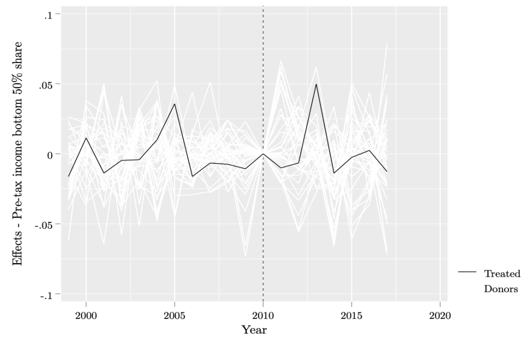
(b) In-place placebo estimates

Notes: The y-axis scale is normalized to 1, so 0.5 means 50% income share. Figure (a) shows the synthetic controls estimation trajectory for Iceland vs. synthetic Iceland. Figure (b) plots the gap between Iceland and synthetic Iceland in a "leave-one-out" in-place placebo setting, where I iteratively apply the synthetic method to each country in the donor pool to obtain a distribution of placebo effects.

Figure 8: Synthetic controls trajectory for bottom 50% income share ratio: Iceland vs. synthetic Iceland



(a) SCM trajectory



(b) In-place placebo estimates

Notes: The y-axis scale is normalized to 1, so 0.5 means 50% income share. Figure (a) shows the synthetic controls estimation trajectory for Iceland vs. synthetic Iceland. Figure (b) plots the gap between Iceland and synthetic Iceland in a "leave-one-out" in-place placebo setting, where I iteratively apply the synthetic method to each country in the donor pool to obtain a distribution of placebo effects.

Table 11: Predictor means before treatment for bottom 50% income share as outcome variable

Variable	**Treated**	**Synthetic**
Income share of middle 40%	44.20625	44.07053
GDP per capita	39941.79	30681.91
Imports as % of GDP	38.9257	38.96324
Exports as % of GDP	32.9441	33.05761
Inflation rate	4.482299	4.494707
Industry share	22.21006	22.23235

Notes: The "treated" column indicates Iceland's actual data, while the "synthetic" column indicates the weighted average calculated through the synthetic control method. All predictors are averaged for the 1999-2006 period. GDP per capita is measured in real dollar terms and expressed in absolute levels; all other variables are in % unit. See Section 6.1 for more information on the logic behind variable selection and Appendix A for more on variable definitions.

Table 12: Synthetic weights for Iceland for bottom 50% income share as outcome variable

Country	**Unit weights**
Australia	0.001
Greece	0.536
Latvia	0.261
Luxembourg	0.038
Norway	0.124
Turkey	0.041

Notes: The synthetic weight is the country weight assigned by the synthetic controls method. See Section 5.1 for details on the SCM methodology.

Figure 6 to 8 display the results for top 10%, middle 40% and the bottom 50% income share ratio as percentage of the total pre-tax income over the years. Because the predictor means and synthetic weights are fairly similar across all three income groups, I only include the results from the bottom 50% income group in Table 11 and 12 for concision purposes.

The results seem to suggest that the Icelandic public (bottom 90% of the population) really suffered a lot during the 2008 financial crisis and recovered quickly, but whether the 2011 program was the reason of inequality reduction is uncertain. After 2008, Iceland's top 10% income ratio rose 20 percentage points to be amongst of the highest in OECD, and its middle 40% and bottom 50% income ratio each fell 10 percentage points to be one of the lowest. The income share gap soon normalized after 2011 and fell back to pre-2008 levels within a year or two. However, we see from the SCM trajectory graphs that both the pre- and post-treatment trends match quite well between Iceland and synthetic Iceland, so the treated unit did not perform especially better than its controlled peers after the treatment (Figure 6a, 7a, and 8a).

The placebo estimations also show a lack of significant treatment effect since Iceland's income share changes after 2011 were largely in line with its placebo peers (Figure 6b, 7b, and 8b). The *p*-values are between 0.6 and 1 throughout the years for the top 10% income share estimation, between 0.5 and 1 for the middle 40%, and between 0.3 and 0.8 for the bottom 50%, which means the probability that the improvement happened by chance is quite high (Figure B36). The other robustness checks with cross-validation (Figure B37) and SDID (Figure B38 to B40) confirm that Iceland's income share changes did not differ significantly from synthetic Iceland. Therefore, the most likely conclusion seems to be that Iceland's income inequality normalized after 2011, but it is likely more due to other factors such as the general recovery of the economy rather than the debt jubilee program *per se*.

8 Conclusion

"Every creditor shall release what he has lent to his neighbor" - Deuteronomy 15:2[14]

While debt reliefs are understood as effective measures to help struggling economies and communities recover from financial shocks, very few have been done in practice, especially not at a country-wide scale. Iceland's 2011 debt jubilee amounted to 10% of its GDP and was arguably the world's most radical debt forgiveness program in the aftermath of the 2008 financial crisis. Economists like Mian and Sufi have long proposed better risk sharing mechanisms and debt forgiveness programs to improve societal welfare, and Iceland's macro-level jubilee came closest to their vision.

In this thesis, I investigate the macroeconomic impacts of the Icelandic jubilee using the synthetic controls method (SCM). My main results show that compared to its synthetic counterpart constructed using OECD countries, Iceland's economy and financial system recovered more quickly following its 2011 debt jubilee policy. Within 3-5 years, compared to a counterfactual scenario where the jubilee policy had not been in place, Iceland's unemployment rate was 2 percentage points lower, GDP per capita was $2,500 higher, non-performing loan (NPL) ratio was 20 percentage points lower, private credit to GDP ratio was 50 percentage points lower, and international private debt to GDP ratio was nearly 300 percentage points lower. The policy not only stabilized the economy, it also successfully reigned in the financial sector back to levels before the financialization boom that started as early as 2005.

My study contributes to the broader literature on debt forgiveness by offering additional evidence on the macroeconomic efficacy of debt relief. Namely, most other debt forgiveness studies have been cross-sectional,

[14] I learned this quote from my "Empirical Macro-Finance" course co-taught by Professors Atif Mian and Ernest Liu.

concentrated on regional policies, and with the aim of identifying specific channels of policy impacts at the micro level like ones discussed above. Using a synthetic controls approach, I am not constrained by the need to impose additional structures of a theoretical model in order to make statements about the macroeconomic outcomes like traditional difference-in-differences studies are.

I release this paper at a time when the global economy stands at a challenging crossroad. The most dominant macro-financial trends in the past 40 years have culminated to more social unrest as the Covid-19 pandemic ravaged the world in the past year. These trends include the dramatic fall in real interest rates, build-up in both household and government debt, greater financialization of the real economy, higher market concentration, decline in productivity growth and real investments, secular stagnation, and rising inequality.

These were global trends but were particularly dominant in Iceland shortly before the 2008 financial crisis. As I introduced throughout this paper, Iceland's dramatic credit boom and financialization in the run-up to 2008 eventually led to its fatal crash. It is a cautionary tale of "big boom followed by big bust," but Iceland ventured into the unthinkable and enacted the unprecedented debt jubilee. The "big debt write-down followed by big money creation" lifted Iceland from the abyss and in fact enabled it to outperform its OECD peers in the recovery.

Iceland thus serves as both a cautionary tale and a message of hope. On one hand, we witness how the currently dominant macro-financial trends could have such a destructive effect on an economy. On the other hand, we see how bold policy actions and more egalitarian risk sharing mechanisms could vastly improve the collective well-being. Meanwhile, Iceland's debt jubilee was also a case of what I might term "safe radicalism." The program initially appeared to be radical, but 10 years after its implementation, it neither shook up the society's fundamental structures nor resulted in a mismanaged economy as some may have feared.

Debt reliefs of all forms have garnered increasingly more attention in today's public discourse, from progressive American politicians calling on student loan forgiveness to IMF officials discussing large-scale debt write-downs for developing nations that are hard hit by the Covid-19 crisis. The punchline from this thesis is that we ought not to be afraid of these bold policy proposals as long as they are grounded in sound economics, and Iceland showed us that debt jubilees could indeed be a feasible and even preferable way forward.

References

[1] Abadie, Alberto (2020). "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature*, forthcoming.

[2] Abadie, A., A. Diamond, and J. Hainmueller (2010). "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association 105* (490), 493–505.

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

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

[5] Abadie, A. and J. L'Hour (2019). "A penalized synthetic control estimator for dissagregated data." Working paper. Available at https://sites.google.com/site/jeremylhour/research.

[6] Agarwal, S., G. Amromin, I. Ben-David, S. Chomsisengphet, T. Piskorski, and A. Seru (2017), "Policy Intervention in Debt Renegotiation: Evidence from the Home Affordable Modification Program." *Journal of Political Economy*, 125(3): 654–712.

[7] Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2020), "Synthetic difference in differences." Originally released as *National Bureau of Economic Research* Working Paper 25532 in 2019. The 2020 version is available at https://arxiv.org/abs/1812.09970.

[8] Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2020), "Matrix Completion Methods for Causal Panel Data Models." *National Bureau of Economic Research*, Working Paper 25132.

[9] Athey, S., and G. W. Imbens (2017), "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives*, 31 (2): 3-32.

[10] Auclert, A., W. Dobbie, and P. Goldsmith-Pinkham (2019), "Macroeconomic Effects of Debt Relief: Consumer Bankruptcy in the Great Recession." Forthcoming in *American Economic Review*. The most recent version was available was released on March 15, 2019 and accessible via http://web.stanford.edu/~aauclert/macrodr.pdf.

[11] Baudino P., J. T. Sturluson, and J. Svoronos (2020). "The banking crisis in Iceland." *Financial Stability Institute Crisis Management Series* No. 1, Bank for International Settlements (March 2020).

[12] Benediktsdottir, S., G. Eggertsson and E. Thorarinsson (2017). "The rise, the fall, and the resurrection of Iceland: A postmortem analysis of the 2008 financial crisis." *Brookings Papers on Economic Activity*, autumn.

[13] Born, B., G. Müller, M. Schularick, and P. Sedláček (2019), "The Costs of Economic Nationalism: Evidence from the Brexit Experiment," *The Economic Journal*, Volume 129, Issue 623, October 2019, Pages 2722–2744, https://doi.org/10.1093/ej/uez020.

[14] Caballero, Ricardo J., Takeo Hoshi, and Anil K. Kashyap (2008), "Zombie Lending and Depressed Restructuring in Japan." *American Economic Review*, 98 (5): 1943-77.

[15] Dalio, Ray (2018). *Principles for Navigating Big Debt Crises*. Published by Bridgewater Associates. ISBN: 978-1-7326898-0-0.

[16] Dobbie, W., P. Goldsmith-Pinkham, and C. Yang (2015), "Consumer Bankruptcy and Financial Health." *National Bureau of Economic Research*, Working Paper 21032.

[17] Doudchenko, Nikolay, and Guido W. Imbens (2016), "Balancing, regression, difference-in-differences and synthetic control methods: A synthesis." *National Bureau of Economic Research*, Working Paper 22791.

[18] Dalio, Ray (2018). "It's Time to Look More Carefully at 'Monetary Policy 3 (MP3)' and 'Modern Monetary Theory (MMT)'". Bridgewater Associates Daily Observations 4/30/2019. Available at https://economicprinciples.org/downloads/MMT_%20MP3_MK.pdf.

[19] Eggertsson, G. B., and P. Krugman (2012), "Debt, Deleveraging, and the Liquidity Trap: A Fisher-Minsky-Koo Approach." *Quarterly Journal of Economics*, 127(3): 1469–1513.

[20] Ferman, B., and C. Pinto (2021), "Synthetic Controls with Imperfect Pre-Treatment Fit." Forthcoming in *Quantitative Economics*. Available at https://arxiv.org/abs/1911.08521.

[21] Funke, M., M. Schularick, and C. Trebesch (2020), "Populist Leaders and the Economy" *CEPR Discussion Paper No. DP15405*, Available at SSRN: https://ssrn.com/abstract=3723597.

[22] Galiani, S. and B. Quistorff (2017) "The synth runner package: Utilities to automate synthetic control estimation using synth." *The Stata Journal* (2017) 17, Number 4, pp. 834–849. https://journals.sagepub.com/doi/pdf/10.1177/1536867X1801700404.

[23] Geanakoplos, J. D., (2010) "Solving the Present Crisis and Managing the Leverage Cycle." *FRBNY Economic Policy Review* 16m no. 1 (August 2010).

[24] Gross, T., M. J. Notowidigdo, and J. Wang (2016), "The Marginal Propensity to Consumer Over the Business Cycle." *National Bureau of Economic Research*, Working Paper 22518.

[25] Heckman, J., and V. Hotz (1989), "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association*, 84(408), 862-874. doi:10.2307/2290059.

[26] Jordà, Òscar (2005), "Estimation and inference of impulse responses by local projections." *American Economic Review*, 95 (1): 161-182.

[27] Kellogg, M., M. Mogstad, G. Pouliot, and A. Torgovitsky (2020). "Combining Matching and Synthetic Control to Trade off Biases from Extrapolation and Interpolation." *National Bureau of Economic Research*, Working Paper 26624.

[28] Kelton, Stephanie (2020), *The Deficit Myth: Modern Monetary Theory and the Birth of the People's Economy.* Published by Public Affairs.

[29] Li, K., and V. Shankar (2020), "Estimating the Causal Effect of A Digitally Native Retailer Opening a New Store: A New Two-Step Synthetic Control Method." Available at SSRN: https://ssrn.com/abstract=3628589 or http://dx.doi.org/10.2139/ssrn.3628589.

[30] Mian A. and A. Sufi (2014a), *House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again.* The University of Chicago Press, 2014.

[31] Mian A. and A. Sufi (2014b), "What Explains the 2007–2009 Drop in Employment?" *Econometrica*, 82(6): 2197–2223.

[32] Mühlbach, N. S., and M. S. Nielsen (2021), "Tree-based synthetic control methods: Consequences of relocating the US embassy," *Working Paper*, 2021.

[33] Ólafsson, T. and K. Á. Vignisdóttir (2012), "Households' position in the financial crisis in Iceland." *Central Bank of Iceland Working Paper* No. 59 (June 2012).

[34] Posner, E. and L. Zingales (2009), "A Loan Modification Approach to the Housing Crisis." *American Law and Economics Review,* Vol. 11, Issue 2, pp. 575-607, 2009, Available at SSRN: https://ssrn.com/abstract=1540381 or http://dx.doi.org/ahp019.

[35] Stricker L. and M. Baruffini (2019), "The effect of reduced unemployment duration on the unemployment rate: a Synthetic Control Approach," *European Journal of Government and Economics*, 9(1), 46-73. https://doi.org/10.17979/ejge.2020.9.1.5714.

Appendix

A. Summary Statistics

1. Data Definition and Sources

In this appendix section I define the core data used for my estimation and specify their sources. For purpose of concision, I do not define the additional data I used to construct aggregate trend graphs.

- GDP per capita: Gross domestic product divided by midyear population. Data is measured in current U.S. dollar. Retrieved from World Development Indicators.
- CPI (2010=100, December): Consumer price index reflects changes in the cost to the average consumer of acquiring a basket of goods and services that may be fixed or changed at specified intervals, such as yearly. Data are period averages. Retrieved from World Development Indicators.
- Unemployment rate (%): Unemployment refers to the share of the labor force that is without work but available for and seeking employment. Retrieved from World Development Indicators.
- Labor force participation rate, ages 15-64: Labor force participation rate is the proportion of the population between ages 15 and 64 that is economically active: all people who supply labor for the production of goods and services during a specified period. Data retrieved from World Development Indicators.
- Employment to population ratio, 15+ (%): The proportion of a country's population that is employed. Ages 15 and older are generally considered the working-age population. An alternative "employment share," which is calculated as the log difference between quarterly real GDP and quarterly total employment according to Born et al. (2019). Because the employment to population ratio data is readily available from World Development Indicators, I use WDI's series rather than calculating employment share with the formula by Born et al. (2019), but others may have different preferences when choosing between these variables as predictor for synthetic controls.
- Employment in agriculture (% total employment): Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit. The agriculture sector consists of activities in agriculture, hunting, forestry and fishing, in accordance with division 1 (ISIC 2) or categories A-B (ISIC 3) or category A (ISIC 4). I include this measure to take into consideration Iceland's unique economic and geographical characteristics, since it may have a large part of its economy being in the fishing industry. Data retrieved from World Development Indicators.
- Labor productivity growth: Calculated as the log difference between quarterly real GDP and quarterly total employment. Formula provided by Born et al. (2019). Calculated based on raw labor market data retrieved from World Development Indicators.
- Industry, value added (% of GDP): Value added denotes the net output of a sector after adding up all outputs and subtracting intermediate inputs. Industry corresponds to ISIC divisions 10-45 and includes manufacturing (ISIC divisions 15-37). It comprises value added in mining, manufacturing

(also reported as a separate subgroup), construction, electricity, water, and gas. Data retrieved from World Development Indicators.

- Trade (% of GDP): The sum of exports and imports of goods and services measured as a share of gross domestic product. Data retrieved from World Development Indicators.

- Imports of goods and services (% of GDP): Imports of goods and services represent the value of all goods and other market services received from the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. Data retrieved from World Development Indicators.

- Exports of goods and services (% of GDP): Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. Data retrieved from World Development Indicators.

- Domestic credit to private sector (% of GDP): Defined as financial resources provided to the private sector, such as through loans, purchases of nonequity securities, and trade credits and other accounts receivable, that establish a claim for repayment. Data retrieved from the Global Financial Database.

- Gross capital formation (% of GDP): Gross capital formation (formerly gross domestic investment) consists of outlays on additions to the fixed assets of the economy plus net changes in the level of inventories. Fixed assets include land improvements (fences, ditches, drains, and so on); plant, machinery, and equipment purchases; and the construction of roads, railways, and the like, including schools, offices, hospitals, private residential dwellings, and commercial and industrial buildings. Abadie et al. (2015) use "Investment Rate" as a predictor variable, which is defined as the ratio of real domestic investment (private plus public) to real GDP. I consider gross capital formation to GDP ratio as my measure for investment rate. Data retrieved from World Development Indicators.

- Gross domestic savings (% of GDP): Gross domestic savings are calculated as GDP less final consumption expenditure (total consumption). Data retrieved from World Development Indicators.

- Government final consumption expenditure (% of GDP): General government final consumption expenditure (formerly general government consumption) includes all government current expenditures for purchases of goods and services (including compensation of employees). It also includes most expenditures on national defense and security, but excludes government military expenditures that are part of government capital formation. Data retrieved from World Development Indicators.

- Households final consumption expenditure (% of GDP): Household final consumption expenditure (formerly private consumption) is the market value of all goods and services, including durable products (such as cars, washing machines, and home computers), purchased by households. Data retrieved from World Development Indicators.

- Bank concentration (%): Assets of three largest commercial banks as a share of total commercial banking assets. Total assets include total earning assets, cash and due from banks, foreclosed real estate, fixed assets, goodwill, other intangibles, current tax assets, deferred tax assets, discontinued operations and other assets. Data retrieved from the Global Financial Database.

- Bank non-performing loans to gross loans (%): Ratio of defaulting loans (payments of interest and principal past due by 90 days or more) to total gross loans (total value of loan portfolio). The loan amount recorded as nonperforming includes the gross value of the loan as recorded on the balance sheet, not just the amount that is overdue.

- BIS housing price cumulative growth rate 2003-08: The Bank of International Settlements releases data on nominal and real residential property prices. I calculate the cumulative growth rate of real residential property prices between 2003 and 2008. Link to data: `https://www.bis.org/statistics/pp_selected.htm`. An alternative to BIS is the Dallas Federal Reserve International House Price Database, which comprises quarterly house price and personal disposable income (PDI) series for a number of countries. However, the Dallas Fed database does not have data for Iceland, so I opt for the BIS series for consistency. Link to the Dallas database: `https://www.dallasfed.org/institute/houseprice`.

- International private debt to GDP (%): Defined as the amount of private international debt securities (amount outstanding) as a share of GDP. It covers long-term bonds and notes and money market instruments placed on international markets. Data retrieved from the Global Financial Database.

- International public debt to GDP (%): Defined as the amount of public international debt securities (amount outstanding) as a share of GDP. It covers long-term bonds and notes and money market instruments placed on international markets. Data retrieved from the Global Financial Database.

- Post-secondary education share (% of population): The percentage of population ages 25 and over that attained or completed post-secondary non-tertiary education. Data retrieved from World Development Indicators.

- Pre-tax income bottom 50% share (%): Pre-tax national income share held by 0 to 50 percentile group of population. Pre-tax national income is the sum of all pre-tax personal income flows accruing to the owners of the production factors, labor and capital, before taking into account the operation of the tax/transfer system, but after taking into account the operation of pension system. The central difference between personal factor income and pre-tax income is the treatment of pensions, which are counted on a contribution basis by factor income and on a distribution basis by pre-tax income. The population is comprised of individuals over age 20. The base unit is the individual (rather than the household) but resources are split equally within couples. Data retrieved from World Inequality Database.

- Pre-tax income middle 40% share (%): Pre-tax national income share held by 50 to 90 percentile group of population. Data retrieved from World Inequality Database.

- Pre-tax income top 10% share (%): Pre-tax national income share held by 90 to 100 percentile group of population. Data retrieved from World Inequality Database.

2. Summary Statistics Table

Table A1: Summary Statistics – Iceland vs. 37 OECD Countries; 1999-2017

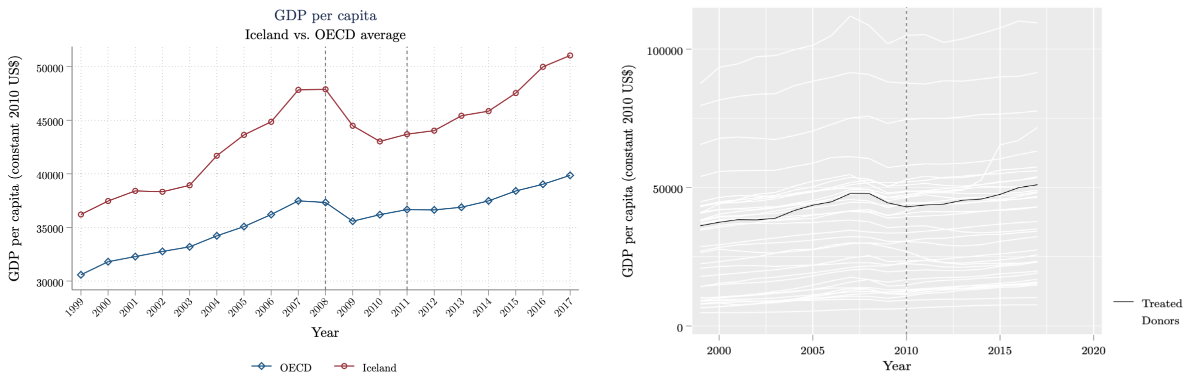
Variables	Iceland		OECD				
	Mean	SD	Mean	25p	50p	75p	SD
GDP per capita	43702.56	4291.53	35666.77	15940.16	36210.68	47423.27	22019.31
CPI (2010=100, December)	88.53	25.22	95.3	85.94	97.92	106.17	15.21
Unemployment rate (%)	3.97	1.86	7.86	4.84	6.99	9.71	4.17
Labor force participation rate, ages 15–64	86.48	1.52	71.76	67.78	72.08	76.25	6.62
Employment to population ratio, 15+	73.31	2.47	55.7	51.08	56.66	59.82	6.58
Employment in agriculture (% total employment)	5.88	1.54	6.62	2.64	4.53	8.7	5.87
Labor productivity growth (%)	11.19	0.09	10.98	10.40	11.21	11.44	0.66
Industry, value added (% of GDP)	21.38	1.34	25.42	21.22	25.46	29.2	5.37
Trade (% of GDP)	83.79	12.45	91.6	56.8	75.25	116.39	55.43
Imports of goods and services (% of GDP)	41.78	4.24	44.81	28.73	37.63	56.43	25.41
Exports of goods and services (% of GDP)	42.01	9.48	46.79	27.86	39.2	57.88	30.31
Domestic credit to private sector (% of GDP)	141.41	65.71	93.58	53.34	89.8	125.01	47.99
Gross capital formation (% of GDP)	21.86	5.91	23.33	20.69	22.93	25.33	4.25
Gross domestic savings (% of GDP)	22.09	2.94	25.32	20.98	24.58	28.48	7.47
Government final consumption expenditure (% of GDP)	23.75	.86	18.92	16.95	19.13	21.2	3.79
Households final consumption expenditure (% of GDP)	54.16	3.11	55.76	51.34	55.62	61.43	7.76
Bank concentration (%)	99.57	.52	68.16	55.1	67.65	83.62	18.97
Bank non-performing loans to gross loans (%)	4.9	5.1	4.3	1.2	2.69	4.9	5.37
BIS housing price cumulative growth rate 2003–08 (%)	66.75	0	36.09	15.05	30.76	50.41	39.78
International private debt to GDP (%)	169.19	102.72	38.65	5.91	21.19	51.76	50.48
International public debt to GDP (%)	17.2	5.2	6.55	.92	4.88	8.88	6.77
Post-secondary education share in 2006 (% of population)	31	0	28.18	18.74	28.78	35.73	11.42
Pre-tax income bottom 50% share (%)	20.39	4.35	21.32	19.24	21.75	23.85	3.86
Pre-tax income middle 40% share (%)	43.54	4.43	43.85	42.94	44.58	46.21	3.53
Pre-tax income top 10% share (%)	36.19	7.81	34.91	30.61	33.11	37.35	6.51

Notes: Columns 2-3 show mean and standard deviation for Iceland. Columns 4-8 show summary statistics for all 37 OECD countries, including mean, 25% percentile, 50% percentile (median), 75% percentile, and standard deviation. For a detailed list of definitions and sources for the above variables, please see the appendix.

3. Aggregate Trends Graphs

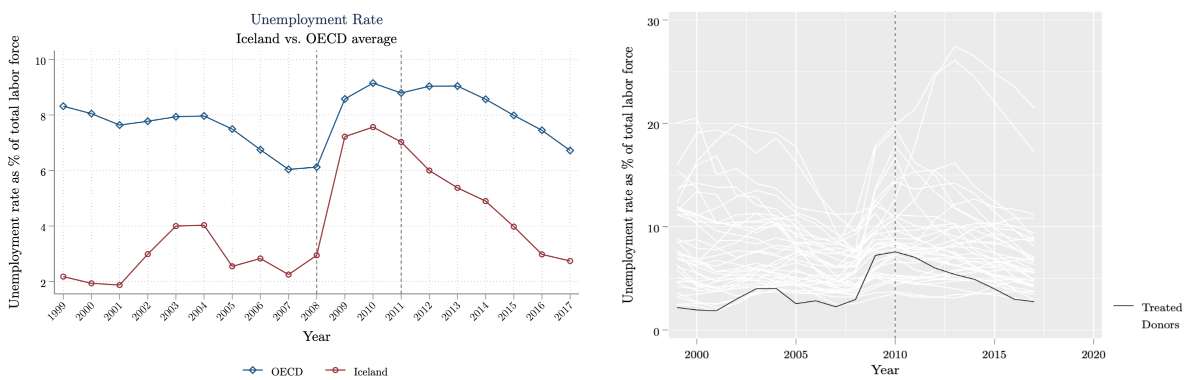
i. Macroeconomic outcomes

Figure A1: GDP per capita trend: Iceland vs. OECD



Left: Iceland vs. OECD average. Right: Iceland vs. rest of OECD.

Figure A2: Unemployment rate trend: Iceland vs. OECD



Left: Iceland vs. OECD average. Right: Iceland vs. rest of OECD.

Figure A3: Labor market outcomes: Iceland vs. OECD average

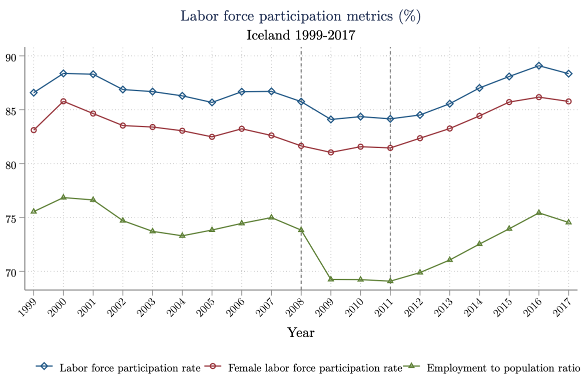
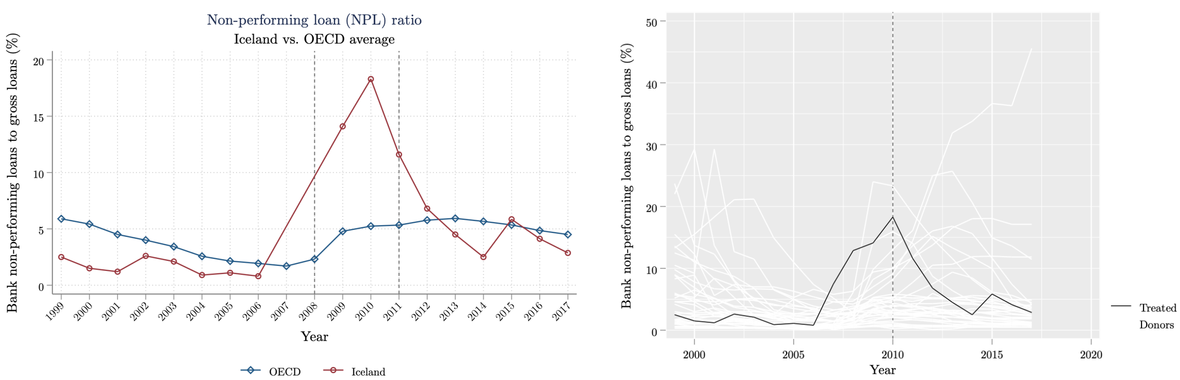
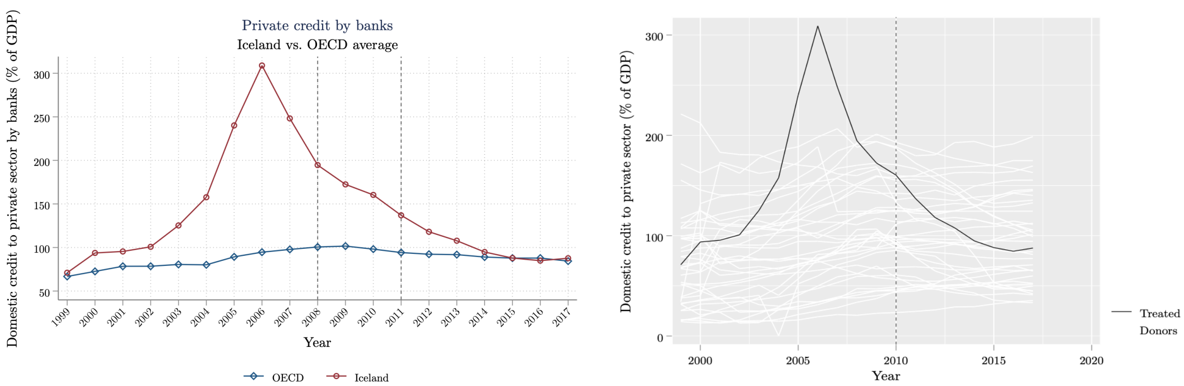


Figure A4: NPL ratio trend: Iceland vs. OECD



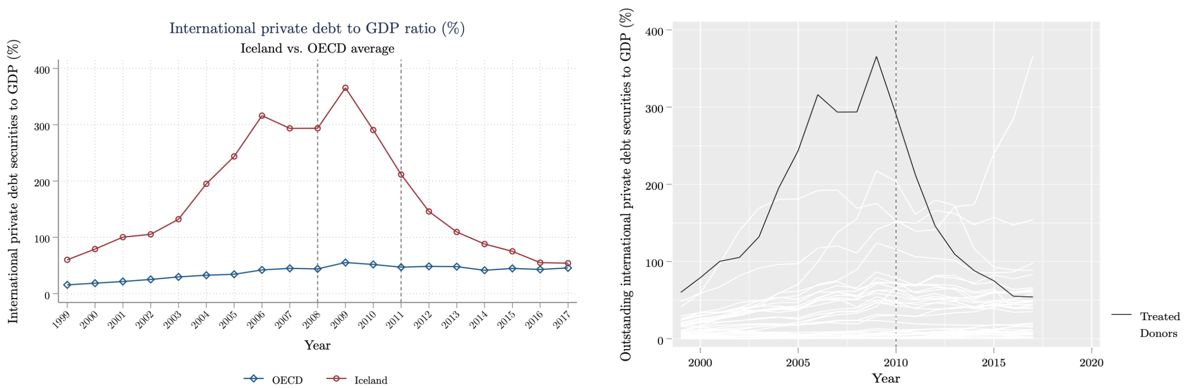
Left: Iceland vs. OECD average. Right: Iceland vs. rest of OECD.

Figure A5: Private credit trend: Iceland vs. OECD



Left: Iceland vs. OECD average. Right: Iceland vs. rest of OECD.

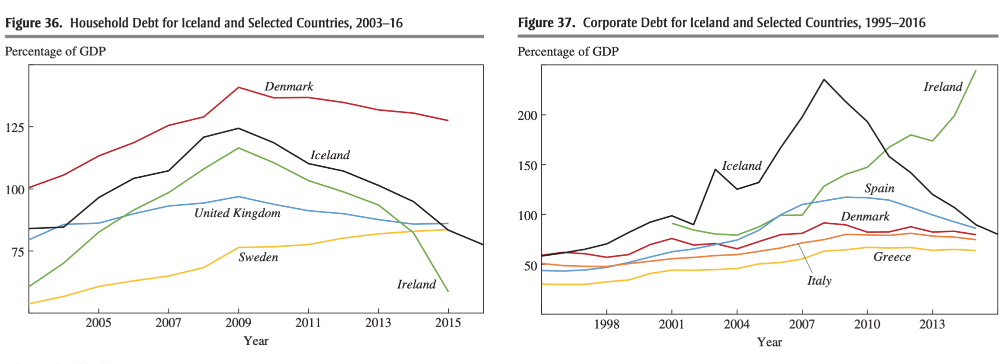
Figure A6: International private debt trend: Iceland vs. OECD



Left: Iceland vs. OECD average. Right: Iceland vs. rest of OECD.

Figure A7: Domestic debt to GDP ratio trend: Iceland vs. OECD average

Figure A8: Household and corporate debt for Iceland and selected OECD countries



Source: Benediktsdóttir et al. (2017)

Figure A9: Household and corporate debt to GDP ratio (%) for Iceland 2005Q1 - 2018Q4

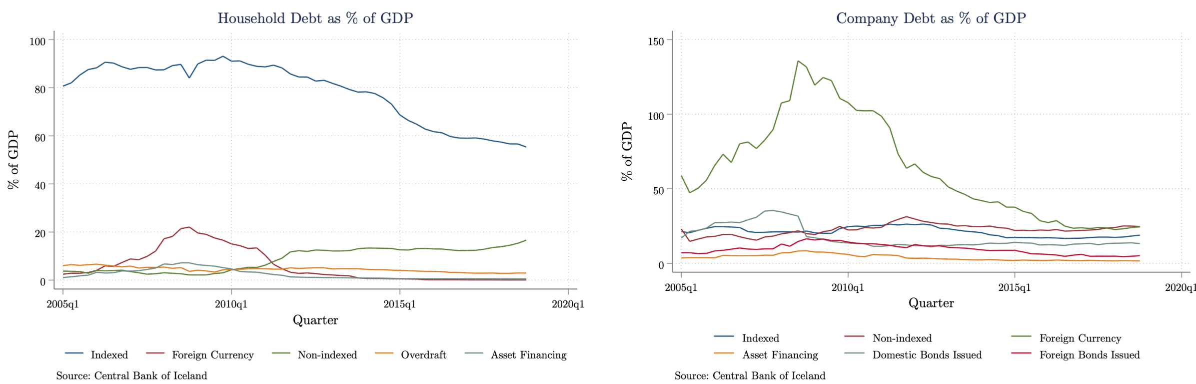
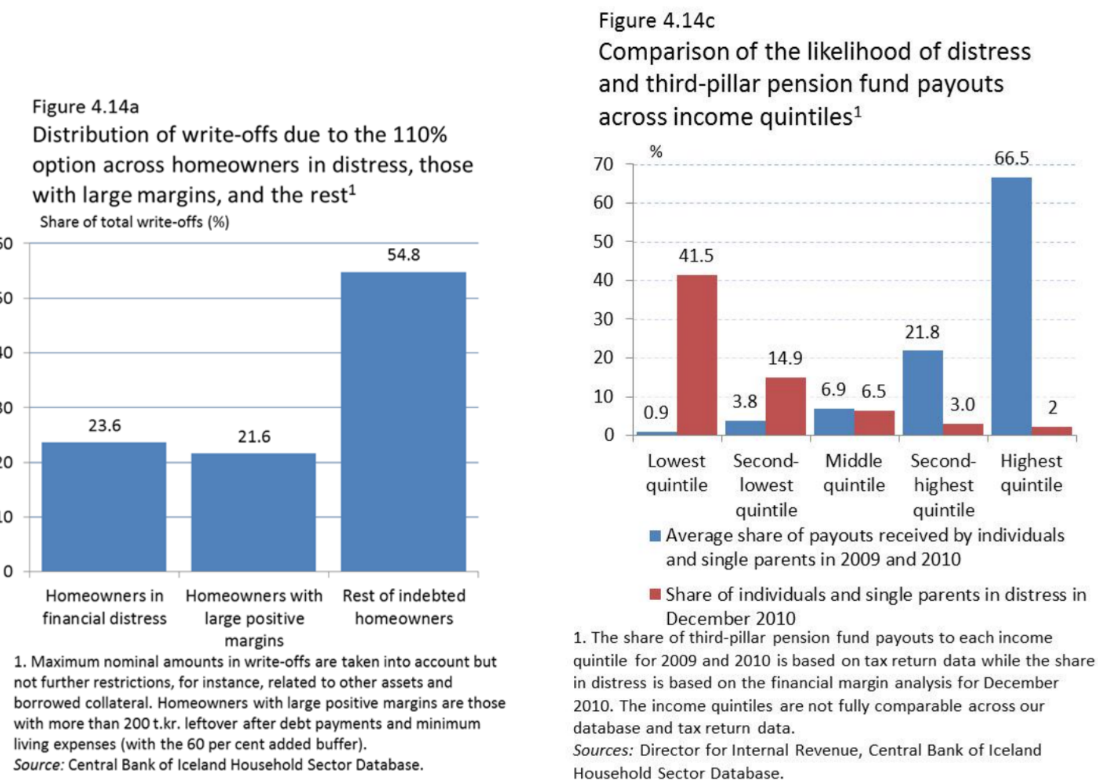


Figure A10: Iceland household debt write-offs after 2008



Source: Ólafsson and Vignisdóttir (2012)

Figure A11: Government and household consumption expenditure (% of GDP)

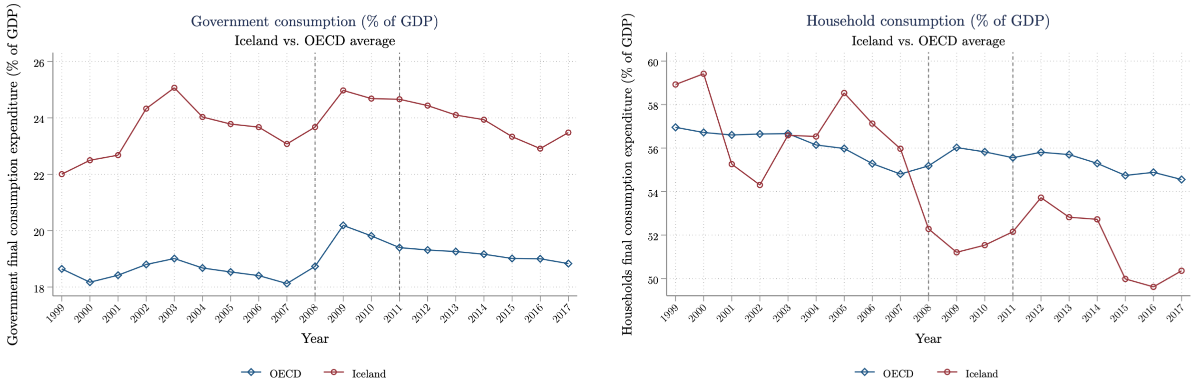
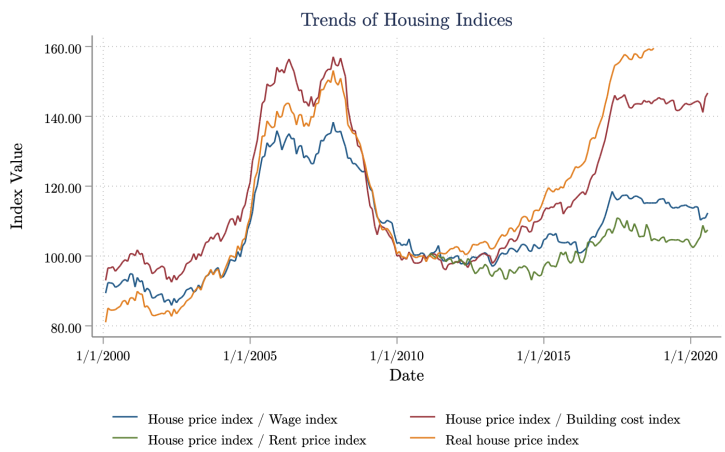


Figure A12: Trends of Icelandic housing indices 2000-2020



Source: Central Bank of Iceland

ii. Financial sector outcomes

Figure A13: Bank concentration and profitability trend: Iceland vs. OECD average

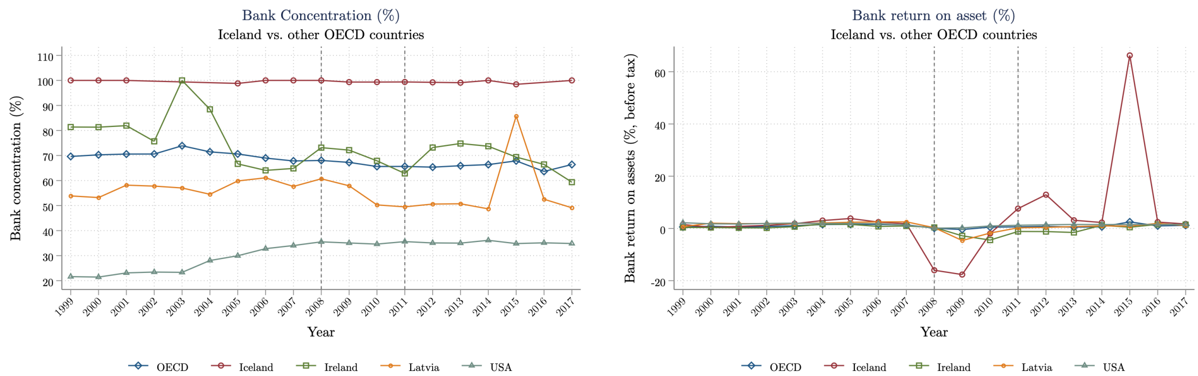


Figure A14: Bank deposit and regulatory capital trend: Iceland vs. OECD average

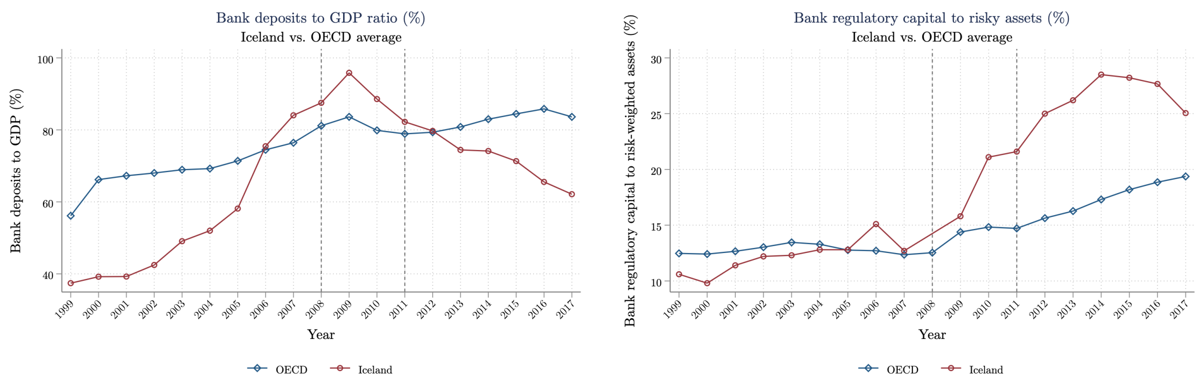
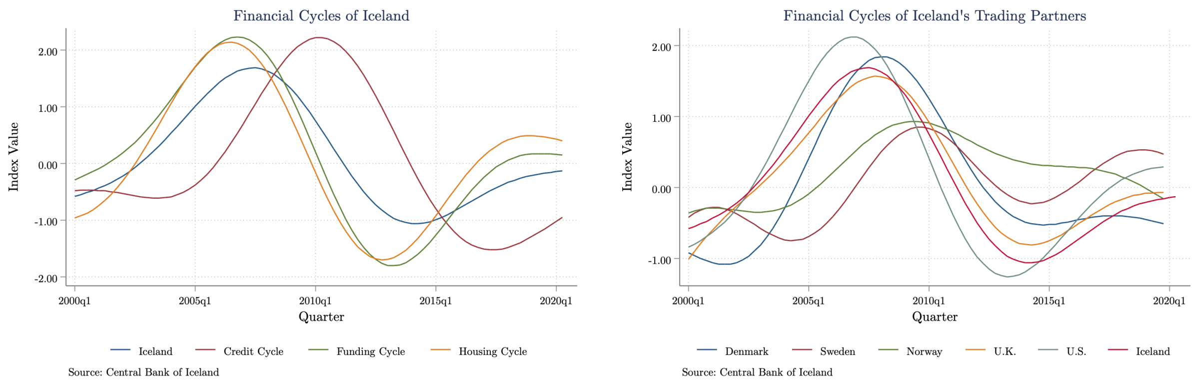


Figure A15: Nonfinancial asset investments and net loans: Iceland vs. OECD average

Figure A16: Financial cycles for Iceland and its trading partners 2000-2020



iii. Systemic crisis outcomes

Figure A17: NPL ratio by systemic banking crisis and IMF intervention

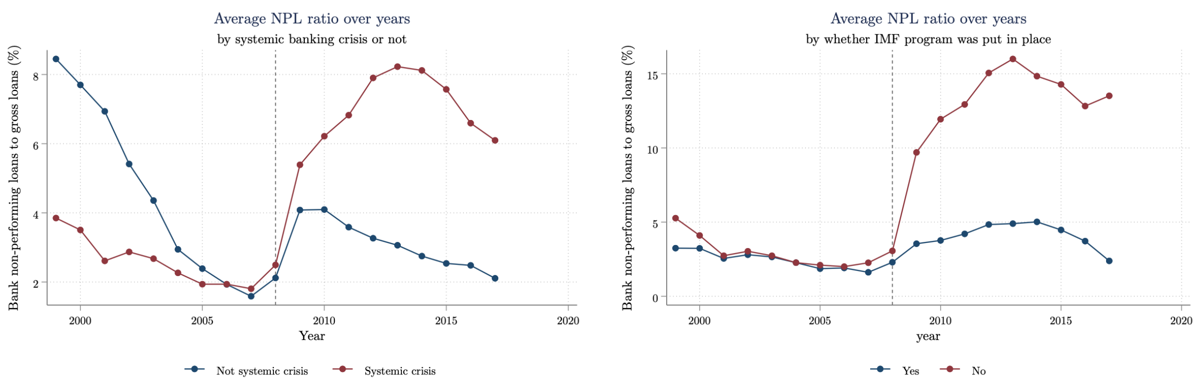


Figure A18: NPL ratio by deposit freeze and insurance level

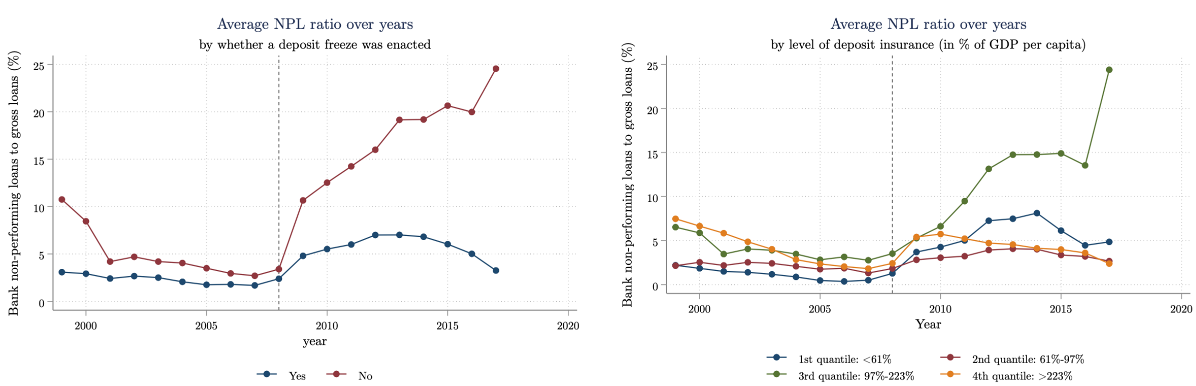


Figure A19: NPL ratio by liquidity support level and loss treatment

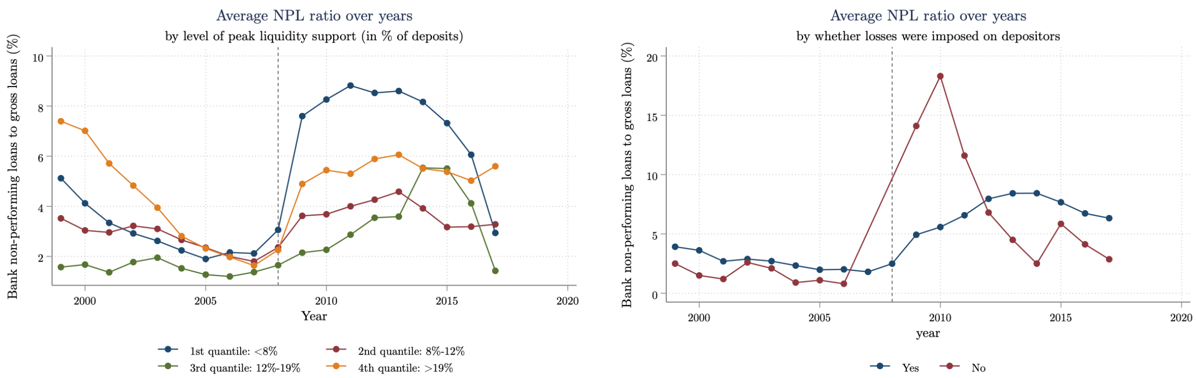
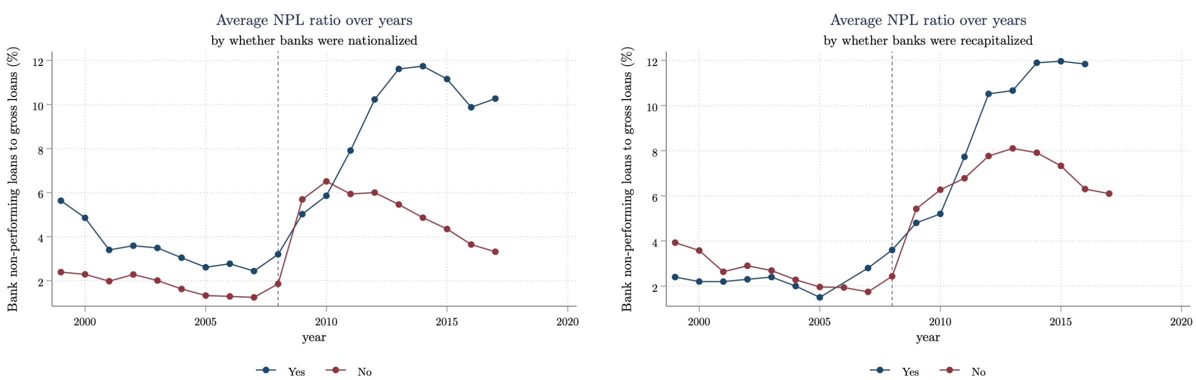


Figure A20: NPL ratio by nationalization and recapitalization of banks



4. Extension into Micro-level Data

It is not the mission of this thesis to investigate the micro-level impacts of the Icelandic debt jubilee. However, I hope that after reading my paper, one may be motivated to further investigate the issue on a micro level. During the initial phases of my research, I did look for micro-level data in Iceland and primarily relied on the following data sources:

- Central Bank of Iceland (CBI): the *Fiscal Stability Report* published every half an year come with data appendices that are used to produced the graphs in the report. These data sets are fairly comprehensive snapshots of Iceland's macroeconomy.
- Statistics Iceland: the official statistics agency of Iceland publishes data on business sectors, labor costs, agriculture, and so on. Many of these data have already been directly incorporated in the CBI *Fiscal Stability Report* data sets.
- Bankscope: Detailed international banking data from 1980-2016. I have collected all available banking sector data for around Icelandic 80 banks between 2008-2016 from the Bankscope data base. There

are a few hundred variables with many esoteric ratios. It seems like a very good data set but many values are missing, so it might require some extra work from the researcher.

- OECD iLibrary: it contains various aggregate data on overall macroeconomic situation such as inequality. The OECD Social and Welfare Statistics are especially helpful in informing me of household welfare.

I collected *quasi* micro-level data for the following variables. Most are available at least from 2012 and onwards. Most are quarterly and yearly data. I would be happy to share my data if anyone is interested:

- Housing: house prices; housing price indices; number of properties listed; composition of mortgage debt; changes in mortgage debt stock; HFF (Housing Financing Fund) prepayment of customer loans and new lending; real prices of capital area commercial real estate; housing status and household mortgage debt; new construction ratio and population growth; proportion of commercial property under construction relative to all constructed properties.

- Households: disposable income; household wealth; new lending to households; real growth in household debt; household debt relative to GDP ratio; household assets and liabilities relative to disposable income and real estate value ratios; debt/assets ratio of individuals with mortgages; number of individuals on default register; share of taxpayers owning more than 300% of disposable income.

- Businesses: lending to companies; companies on default register; bankruptcies and unsuccessful distraint actions; corporate debt to GDP ratio.

- Banking sector: D-SIB (domestic systemically important banks) assets, asset distribution, liquid assets, liquidity coverage ratio, registered new investments using foreign capital, lending classified by borrowers/industries, funding in foreign currency and average residual maturity; foreign-owned ISK deposits and Icelandic securities.

B. Robustness Checks with Cross-Validation and SDID

In this section of the appendix, I include more detailed explanations as to how I reasoned through my analysis, which would be too long to include in the main write-up. For each outcome of interest, I present three sets of results: the p-values for synthetic controls estimations detailed in Section 6; robustness check using the cross-validation technique; and robustness check using synthetic difference-in-differences (SDID). Section 4.1.5 details the methodology of placebo inference and the calculation for p-values in synthetic controls estimation. The details for cross-validation and SDID are detailed in Section 4.2 and 4.3 respectively.

The p-values shown in the graphs below are essentially a vector of the proportions of placebo effects that are at least as large as the main effect for each post-treatment period. So, in other words, if the p-value on a graph is 0.5, we may directly interpret it as that the "probability that the event would happen by chance" is 0.5. One might wonder why there are six different p-value graphs for each outcome. This is because I try out different combinations of predictors. As explained in section 6.3 on my empirical strategy, the main reason that I produced dozens of p-value graphs is because SCM is by construction a more "trial-and-error" process where the researchers have to try out many different combinations of predictor combinations to arrive at the "best" synthetic fit. Often, because the pre-trend fit was not perfect in the baseline estimate and because the corresponding p-values are still quite high, it seems that the result is inconclusive at best. For example, while a specific set of predictors consisting of macroeconomic variables like trade openness and industry might yield high p-values for NPL as the outcome and imply that the result is not statistically significant, trying out a different set of predictors with financial sector variables like bank concentration may yield a much lower p-value. Therefore, the different sets of p-values graphs below will give us a clearer view on whether the SCM estimation for a given outcome is indeed statistically significant.

I summarize the headline conclusions of this section to be:

- Unemployment: The robustness checks validate the SCM result that the 2011 debt jubilee policy helped reduce unemployment. Although the estimated effect is not as statistically significant or as large in magnitude as the effects on some of the other outcomes like NPL ratio and private credit, the likelihood of the unemployment rate improvement not having anything to do with the 2011 policy is quite low.

- GDP per capita: Neither the cross-validation procedure nor the SDID estimation dramatically improve the lack of statistical significance of the previous SCM estimate. Iceland's GDP per capita still trended upwards after 2011 and performed better than synthetic Iceland and its OECD peers under certain regression specifications, but it is hard to isolate out the effect of the 2011 policy. This is partly due to the limitation of the synthetic controls methodology, which puts very stringent assumptions on matching pre-trends, but also likely because GDP per capita is just such an all-encompassing variable that it would be hard to isolate out the effect on it. The robustness checks do not contradict the previous SCM conclusion in the main write-up.

- NPL ratio: The robustness checks confirm the previous conclusion that Iceland's 2011 debt jubilee policy substantially reduced the NPL ratio. The SCM results in the main write-up are statistically significant at the 95% level. The in-time and in-place placebo estimates with the cross-validation procedure confirm that Iceland performed better than its OECD peers. The SDID method estimates the NPL ratio reduction to be at around 10 percentage points, which are even larger in magnitude than the 7.5 percentage point reduction estimated by SCM.

- Private credit: The robustness checks confirm the previous conclusion that Iceland's 2011 debt jubilee policy substantially reduced the private credit to GDP ratio. The SCM results in the main write-up are statistically significant at the 95% level. The in-time and in-place placebo estimates with the cross-validation procedure confirm that Iceland performed better than its OECD peers. The SDID method estimates the private credit to GDP ratio reduction to be at around 18 percentage points, which is lower than the 65-percentage-point reduction estimated by SCM, but the SDID method gives tighter standard errors. Regardless, all results here confirms that the 2011 policy had a statistically significant effect on normalizing the private credit boom.

- International debt: The robustness checks confirm the previous conclusion that Iceland's 2011 debt jubilee policy substantially reduced the private international debt to GDP ratio. The SCM results are statistically significant at the 90% level using certain macroeconomic predictors for RMSPE optimization. The in-time and in-place placebo estimates with the cross-validation procedure confirm that Iceland performed better than its OECD peers, though it is particularly challenging to construct a good pre-trend given that Iceland's international debt level was unusually higher compared to any other OECD country. The SDID method estimates the international debt to GDP ratio reduction to be at around 204 percentage points, which are higher than the 111-percentage-point reduction estimated by SCM, but the SDID method gives tighter standard errors. Regardless, all results here confirms that the 2011 policy had a statistically significant effect on reducing the international debt level.

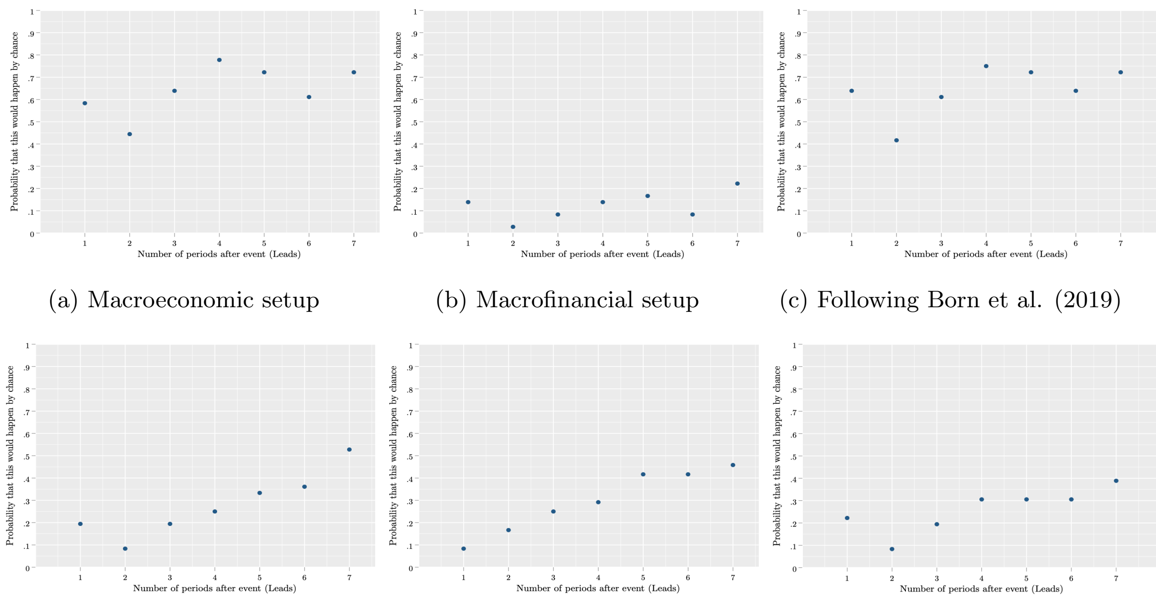
- Income inequality: Neither the cross-validation procedure nor the SDID estimation dramatically improve the lack of statistical significance of the previous SCM estimates, the *p*-values for which are particularly high. The income shares for each of Iceland's income bracket (bottom 50%, middle 40%, and top 10%) largely trended in line with its OECD peers after 2011. In sum, the estimations are not statistically significant across the board. This could be due to a variety of factors. One is that the measurement for income share and inequality may be subject to all kinds of heterogeneity across OECD countries, and this issue could be especially pronounced for a small economy with a tiny population of 350,000. Meanwhile, as explained in Section 6.1, income inequality is likely not a first-order outcome that fiscal and monetary policymakers consider when they design policy responses to crises, so it is hard to extrapolate the debt jubilee's *immediate* effect on income shares. Lastly, to draw any definitive conclusion about income inequality, I would likely need to change my empirical strategy and collect more micro-level data to make my model more fitting for identifying specific channels that influence inequality. It is also not a central mission of my thesis to examine the debt jubilee's impact on income inequality, so I would not worry too much about the lack of statistical significance here. I feel relatively comfortable leaving this as a question that needs further research by me and other scholars.

1. Unemployment

i. p-values for Synthetic Controls Estimation

Figure B21 shows the p-values under six different setups of predictors. The synthetic controls estimation in the main write-up (Figure 1) corresponds to Figure B21f as they all use the same set of labor market predictors for RMSPE optimization. We see that the p-values are not too high, mostly fluctuating between 0.1 to 0.4 in the years following the treatment. Meanwhile, the p-values are as low as less than 0.1 under some other predictor constructions such as using macrofinancial variables for estimation (Figure B21b). All this shows that though the estimation is not as statistically significant as we wish to be, the likelihood that the unemployment improvement happened by chance was still relatively low, and it seems reasonable to conclude that the 2011 policy had a substantive effect on reducing unemployment.

Figure B21: p-values for synthetic controls estimation of unemployment rate



(a) Macroeconomic setup (b) Macrofinancial setup (c) Following Born et al. (2019)

(d) Following Abadie et al. (2015) (e) Banking & housing setup (f) Labor market setup

Notes: For a detailed list of predictor variables that make up each of the above setup, see Section 6.1.

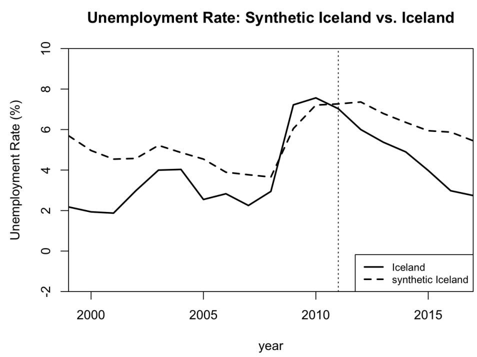
ii. Synthetic Controls with Cross-Validation

Figure B22 displays the SCM trajectory graphs *with* cross-validation, which can be compared with Figure 1, which shows the SCM estimation results *without* cross-validation. We see here that cross-validation seems to validate our previous results. We still observe that Iceland's unemployment rate seems to have dropped more quickly than synthetic Iceland after 2011.

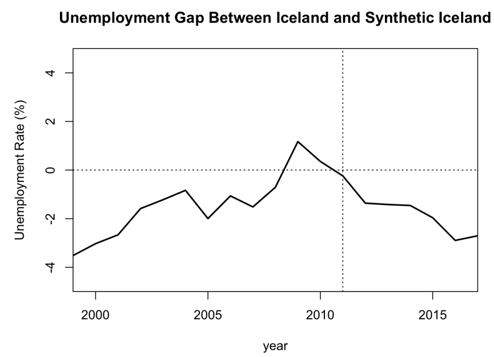
The in-time placebo estimation in Figure B22c shows that Iceland the 2013 placebo debt jubilee has no perceivable effect on the outcome, which suggests that the gap estimated in Figure B22b reflects the impact of the actual 2011 debt jubilee and not a potential lack of predictive power of the synthetic control. Meanwhile, the "leave-one-out" in-place placebo estimation in Figure B22d shows that Iceland significantly outperformed its placebo peers in bringing down the unemployment level after 2011. I would consider the effect of the debt jubilee on Iceland to be significant since the estimate effect for Iceland is unusually large relative to the distribution of placebo effects.

Table B2 and B3 showcase the $\boldsymbol{V}$ matrix results for predictors from the cross-validation procedure and the subsequent $\boldsymbol{W}$ matrix results for countries. There is nothing out-of-the-ordinary to report here, and the logic for interpretation is the same as for the previous SCM regression weights output tables, so I will not over-explain here.

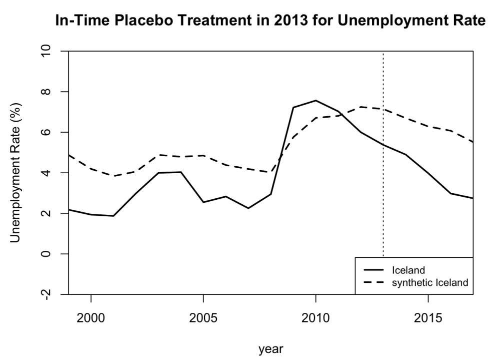
Figure B22: Cross-validation results for unemployment rate



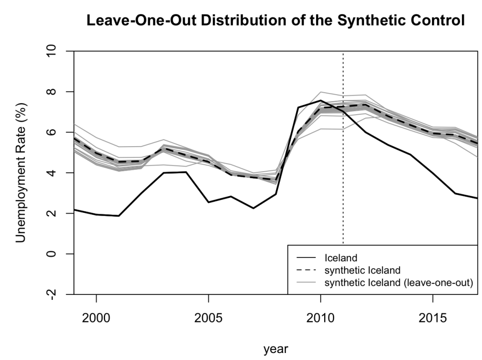
(a) SCM trajectory



(b) SCM effect



(c) In-time placebo test assigned to 2013



(d) In-place placebo “leave-one-out” estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland and synthetic Iceland. Figure(b) shows the effects gap. Figure (c) shows the in-time placebo result, where I rerun the model for the placebo case when the 2011 policy is reassigned to year 2013, two years after the policy actually occurred. Figure (d) shows the in-place “leave-one-out” placebo result, where I use the same out-of-sample validation technique to compute the synthetic control.

Table B2: Cross-validation synthetic and regression weights for Iceland: unemployment rate

	Country	Synthetic Control Weights	Regression Weights	Country	Synthetic Control Weights	Regression Weights
1	Australia	0.00	0.06	Korea, Rep.	0.00	-0.08
2	Austria	0.00	0.07	Latvia	0.00	0.11
3	Belgium	0.00	-0.10	Lithuania	0.00	-0.04
4	Canada	0.00	-0.04	Luxembourg	0.00	0.02
5	Chile	0.00	-0.09	Mexico	0.00	0.06
6	Colombia	0.00	0.18	Netherlands	0.00	0.18
7	Czech Republic	0.00	0.02	New Zealand	0.36	0.17
8	Denmark	0.58	0.19	Norway	0.00	0.03
9	Estonia	0.00	0.06	Poland	0.00	-0.04
10	Finland	0.00	-0.00	Portugal	0.00	0.25
11	France	0.00	0.01	Slovak Republic	0.00	-0.07
12	Germany	0.00	0.02	Slovenia	0.00	0.10
13	Greece	0.00	0.04	Spain	0.00	-0.06
14	Hungary	0.00	-0.10	Sweden	0.00	0.11
15	Ireland	0.00	-0.03	Switzerland	0.06	0.20
16	Israel	0.00	-0.09	Turkey	0.00	-0.15
17	Italy	0.00	-0.11	United Kingdom	0.00	0.10
18	Japan	0.00	0.00			

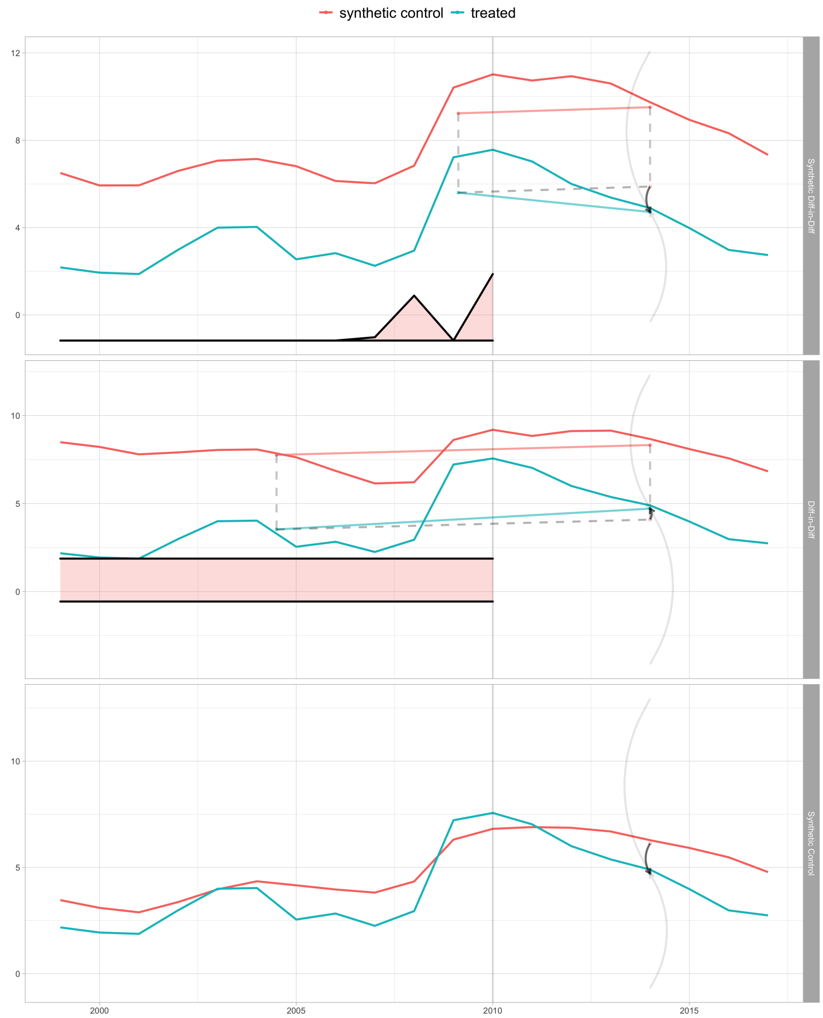
Table B3: Cross-validation predictor means before treatment: unemployment rate

	Treated	Synthetic	Custom V Weights
Unemployment rate	5.0	5.2	0.2
Labor participation rate	85.3	79.2	0.1
Trade	86.9	84.9	0.2
Industry share	21.5	21.9	0.3
Employment in agriculture	5.5	4.3	0.0
Labor productivity growth	11.2	11.4	0.2
Secondary education completion	31.0	37.1	0.0

iii. Synthetic Difference-in-Differences (SDID)

Figure B23 and Table B4 display the results from the SDID estimation for unemployment rate alongside estimations for DiD and SCM for comparison purposes. We may observe that SDID captures the parallel trend much better than DiD and SCM do, but the post-treatment effects are not statistically significant given the low point estimate and high standard errors. However, SDID yields a lower point estimate and tighter standard errors than SCM and still confirms the conclusion that the 2011 debt jubilee lowered unemployment. In sum, SDID seems to have captured the "diff-in-diff" relationship better than the other estimators, but it is just that the effect is not as statistically significant as we wish it to be.

Figure B23: Estimation trajectories of SDID vs. DiD vs. SCM: unemployment rate



Notes: The three figures are the SDID, DiD, and SCM estimation trajectories in the order from top to bottom. All figures are produced in R using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Table B4: Regression results of SDID vs. DiD vs. SCM: unemployment rate

	SDID	DiD	SCM
Point estimate	-1.167	0.624	-1.416
Standard error	3.095	3.071	3.424
95% confidence interval	(-7.233, 4.899)	(-5.395, 6.644)	(-8.127, 5.296)

Notes: As detailed in Section 4.3 on the methodology of SDID, the point estimates for SDID, DiD, and SCM are the coefficient values for $\hat{\tau}$ in the estimation equations (12)-(14), and standard errors are calculated accordingly using the methods and statistical tools provided by Arkhangelsky et al. (2020).

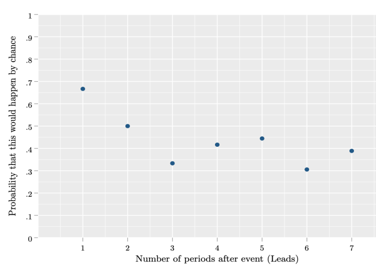
2. GDP per capita

i. p-values for Synthetic Controls Estimation

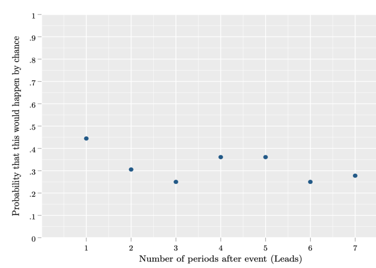
Figure B24 shows the p-values under six different setups of predictors. The synthetic controls estimation in the main write-up (Figure 2) corresponds to Figure B24d as they all use the same set of predictors for RMSPE optimization following Abadie et al. (2015). We see that the p-values are quite high and fluctuate between 0.4 to 0.8 in the years following the treatment. This shows that the estimation is not statistically significant.

Meanwhile, the p-values are at least above 0.4 across the different predictor constructions. Because we are ultimately interested in the Icelandic financial crisis, one might assume that including more financial sector variables like bank concentration, domestic savings, international debt to GDP ratio could improve our pre-trend match, which is the purpose of including in Figure B24b and B24e. Unfortunately, doing so has only slightly improved the p-values compared to the estimates using other predictors, and there is still quite a high probability of at least 0.4 that Iceland's GDP per capita post-treatment improvement happened by chance, much higher than what we would ideally like to see in a placebo test. It seems reasonable to conclude that our estimation for GDP per capita is simply not statistically significant across the board.

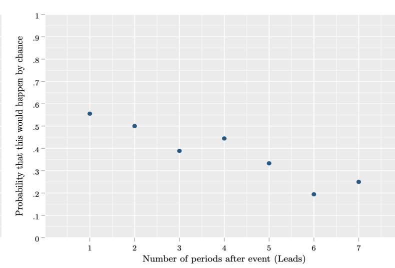
Figure B24: p-values for synthetic controls estimation of GDP per capita



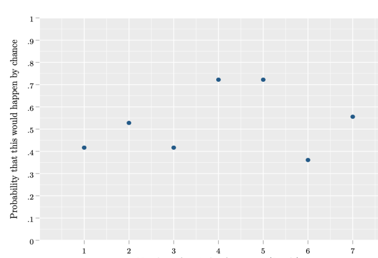
(a) Macroeconomic setup



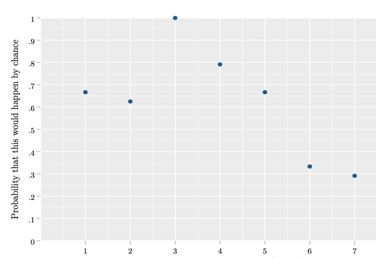
(b) Macrofinancial setup



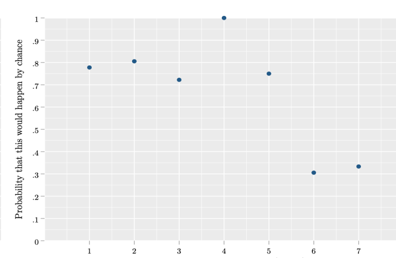
(c) Following Born et al. (2019)



(d) Following Abadie et al. (2015)



(e) Banking & housing setup



(f) Labor market setup

Notes: For a detailed list of predictor variables that make up each of the above setup, see Section 6.1.

ii. Synthetic Controls with Cross-Validation

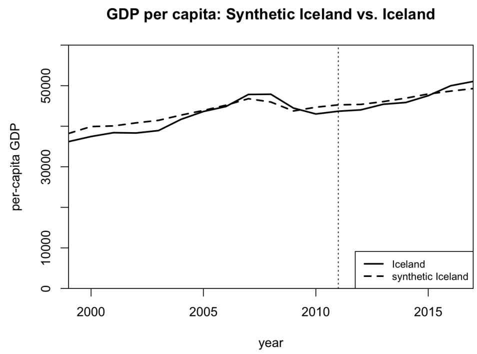
Figure B25 display the SCM trajectory graphs *with* cross-validation, which can be compared with Figure 2, which displays the SCM estimation results *without* cross-validation. We see here that cross-validation does not seem to improve our previous results. Fortunately, the results are not contradictory. We still observe that Iceland's per-capita GDP seems to have recovered more quickly than synthetic Iceland after 2011.

The in-time placebo estimation in Figure B25c shows that Iceland the 2013 placebo debt jubilee has no perceivable effect on the outcome, which suggests that the gap estimated in Figure B25b reflects the impact of the actual 2011 debt jubilee and not a potential lack of predictive power of the synthetic control. Nevertheless, the "leave-one-out" in-place placebo estimation in Figure B25d shows that Iceland's GDP per capita performance was largely in line with its placebo peers. From there it would be hard to consider the effect of the debt jubilee on Iceland to be significant since the estimated effect for Iceland is not larger relative to the distribution of placebo effects. The placebo trajectories (gray lines in the graph) seem to suggest that there was indeed effect in the sense that Iceland had a more rapidly increasing per-capita GDP before the 2008 crisis, followed by a sharp drop and a quick bounce-back to be growing faster than its peers again. In other words, Iceland may have suffered more than its synthetic peers immediately after 2008, it did recover and catch up more quickly after 2011. The pre-trend match is not good under the classical SCM framework, which inhibits us from isolating out the "diff-in-diff" effect, but we can observe somewhat of an effect that Iceland performed better relative to its peers after 2011.

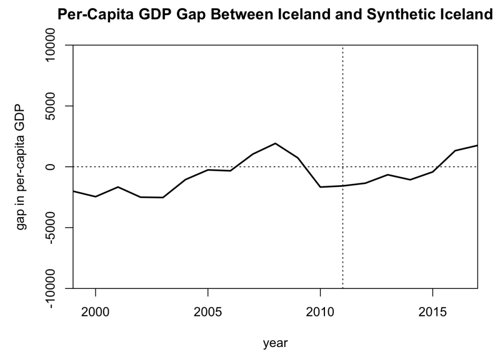
In sum, the in-time and in-place placebo estimates fail to add more certainty to our estimation, but they still show that Iceland's GDP per capita trended upwards after 2011 after all. This is likely because cross-validation still relies on fitting pre-trends and does not entirely relax the parallel pre-trend assumptions needed for synthetic controls estimations. The SDID estimation might thus be able to better capture the "diff-in-diff" effect, as we shall see later.

Table B5 and B6 showcase the $\boldsymbol{V}$ matrix results for predictors from the cross-validation procedure and the subsequent $\boldsymbol{W}$ matrix results for countries. The logic for interpretation is the same as for the previous SCM regression weights output tables, so I will not over-explain here. It is surprising to see from Table B6 that the cross-validation process tells us that per-capita GDP and imports matter the most as a predictor, while exports, industry share, and private credit barely do. One would expect that the $\boldsymbol{V}$ matrix weights be somewhat equally divided amongst at least more than two characteristic variables. This, however, should not be the reason why the standard errors for the estimation have not improved under cross-validation. There are likely other underlying structural reasons related to SCM that have caused that.

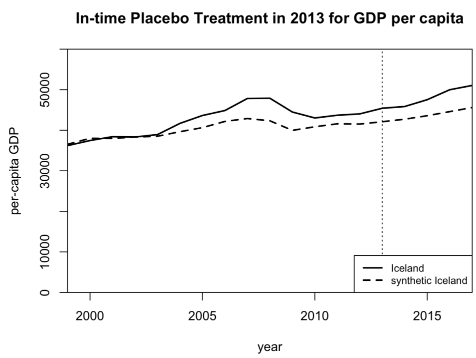
Figure B25: Cross-validation results for GDP per capita



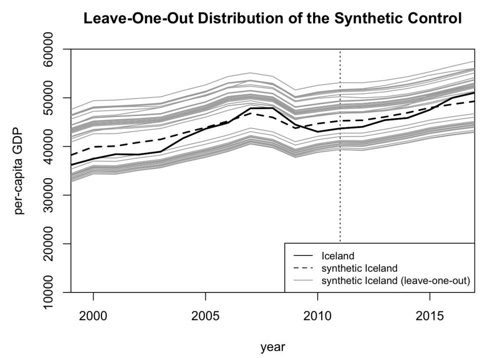
(a) SCM trajectory



(b) SCM effect



(c) In-time placebo test assigned to 2013



(d) In-place placebo "leave-one-out" estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland and synthetic Iceland. Figure(b) shows the effects gap. Figure (c) shows the in-time placebo result, where I rerun the model for the placebo case when the 2011 policy is reassigned to year 2013, two years after the policy actually occurred. Figure (d) shows the in-place "leave-one-out" placebo result, where I use the same out-of-sample validation technique to compute the synthetic control.

Table B5: Cross-validation synthetic and regression weights for Iceland: GDP per capita

	Country	Synthetic Control Weights	Regression Weights	Country	Synthetic Control Weights	Regression Weights
1	Australia	0.00	0.14	Korea, Rep.	0.00	0.17
2	Austria	0.00	-0.06	Latvia	0.12	0.35
3	Belgium	0.00	-0.16	Lithuania	0.00	0.12
4	Canada	0.00	0.05	Luxembourg	0.15	0.10
5	Chile	0.00	0.05	Mexico	0.00	-0.10
6	Colombia	0.00	-0.05	Netherlands	0.00	-0.07
7	Czech Republic	0.00	-0.18	New Zealand	0.00	0.12
8	Denmark	0.00	0.25	Norway	0.05	0.18
9	Estonia	0.00	0.18	Poland	0.00	-0.19
10	Finland	0.00	-0.09	Portugal	0.00	0.00
11	France	0.00	-0.20	Slovak Republic	0.00	-0.15
12	Germany	0.00	-0.14	Slovenia	0.00	-0.09
13	Greece	0.00	-0.03	Spain	0.00	0.16
14	Hungary	0.00	0.11	Sweden	0.00	-0.03
15	Ireland	0.00	0.07	Switzerland	0.00	0.05
16	Israel	0.00	-0.14	Turkey	0.27	0.41
17	Italy	0.00	-0.15	United Kingdom	0.00	0.22
18	Japan	0.00	-0.14	United States	0.42	0.24

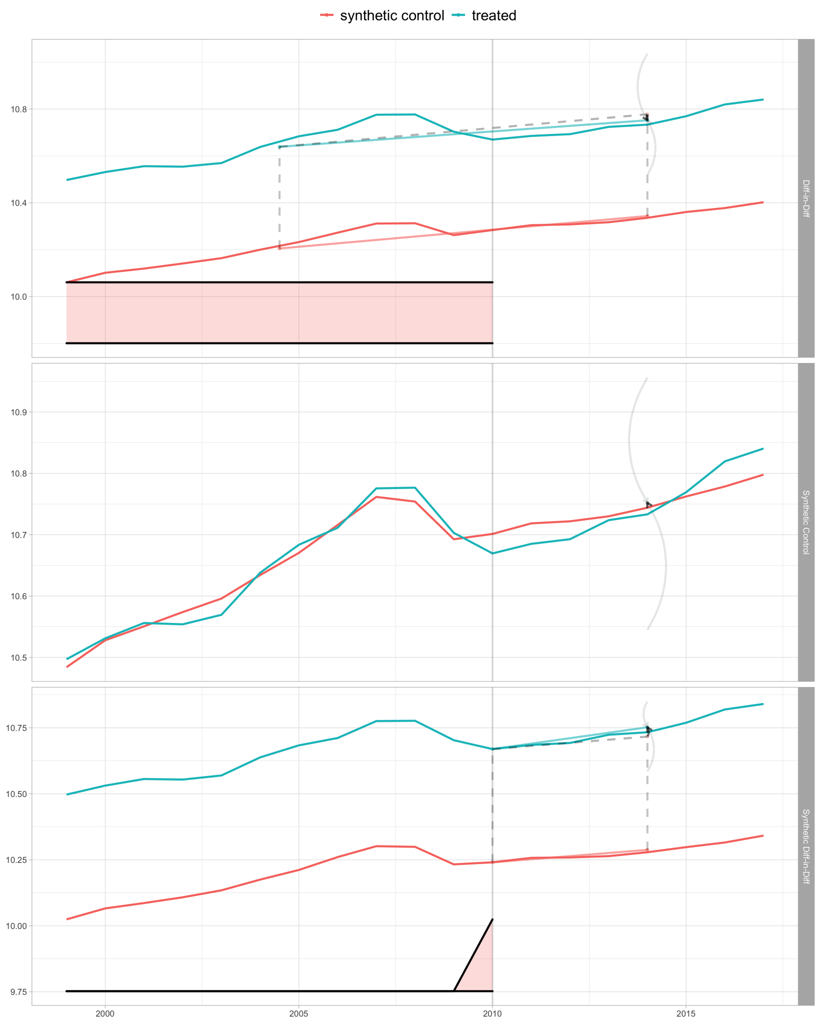
Table B6: Cross-validation predictor means before treatment: GDP per capita

	Treated	Synthetic	Custom V Weights
GDP per-capita	45300.3	45278.0	0.6
Imports	43.6	43.5	0.3
Exports	43.3	45.1	0.0
Industry share	21.5	21.5	0.0
Private credit	203.6	120.1	0.0
Inflation rate	7.6	4.4	0.0

iii. Synthetic Difference-in-Differences (SDID)

Figure B26 and Table B7 display the results from the SDID estimation for GDP per capita alongside estimations for DiD and SCM for comparison purposes. We may observe that SDID has successfully captured the parallel trend much better than SCM, but the post-treatment effects are not statistically significant given the low point estimate and high standard errors. In sum, SDID has successfully captured the "diff-in-diff" relationship better than the other estimators, but it is just that the effect of GDP per capita is not significant enough in magnitude.

Figure B26: Estimation trajectories of SDID vs. DiD vs. SCM: GDP per capita



Notes: The three figures are the SDID, DiD, and SCM estimation trajectories in the order from top to bottom. All figures are produced in R using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Table B7: Regression results of SDID vs. DiD vs. SCM: GDP per capita

	SDID	DiD	SCM
Point estimate	0.035	-0.025.0	0.002
Standard error	0.063	0.133	0.106
95% confidence interval	(-0.087, 0.158)	(-0.286, 0.235)	(-0.206, 0.209)

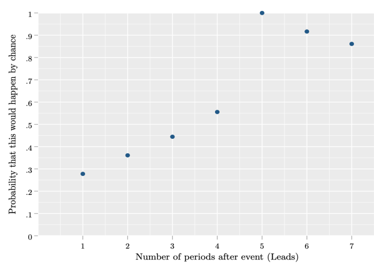
Notes: As detailed in Section 4.3 on the methodology of SDID, the point estimates for SDID, DiD, and SCM are the coefficient values for $\hat{\tau}$ in the estimation equations , and standard errors are calculated accordingly using the methods and statistical tools provided by Arkhangelsky et al. (2020).

3. NPL

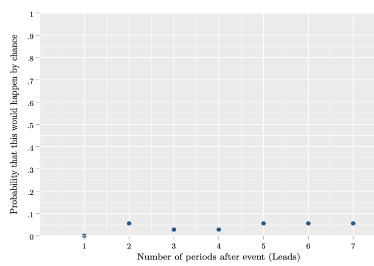
i. p-values for Synthetic Controls Estimation

Figure B27 shows the p-values under six different setups of predictors. The synthetic controls estimation in the main write-up (Figure 3) corresponds to Figure B21b as they all use the same set of macrofinancial predictors for RMSPE optimization. We see that the p-values are below 0.1 and near zero in the years following the treatment. The p-values are higher under some other predictor constructions, but the macrofinancial setup is sufficient in showing that the estimation is statistically significant. The likelihood that the NPL ratio reduction happened by chance was very low, and it seems reasonable to conclude that the 2011 policy had a substantive effect.

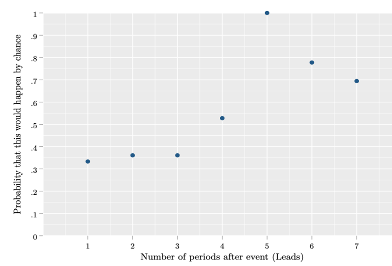
Figure B27: p-values for synthetic controls estimation of NPL ratio



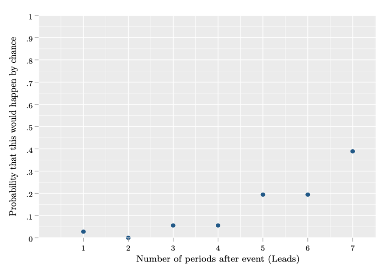
(a) Macroeconomic setup



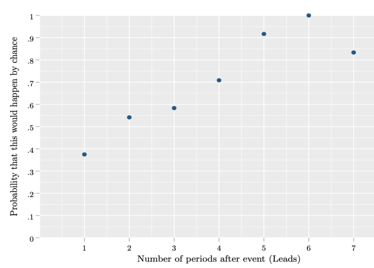
(b) Macrofinancial setup



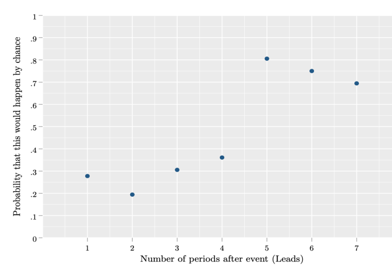
(c) Following Born et al. (2019)



(d) Following Abadie et al. (2015)



(e) Banking & housing setup



(f) Labor market setup

Notes: For a detailed list of predictor variables that make up each of the above setup, see Section 6.1.

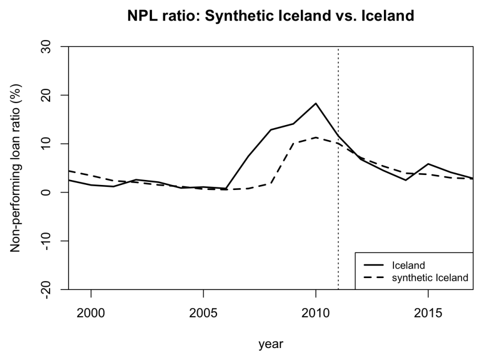
ii. Synthetic Controls with Cross-Validation

Figure B28 displays the SCM trajectory graphs *with* cross-validation, which can be compared with Figure 3, which shows the SCM estimation results *without* cross-validation. We see here that Iceland's NPL ratio dropped as quickly as synthetic Iceland after 2011, which are estimates are less large in magnitude than the previous results.

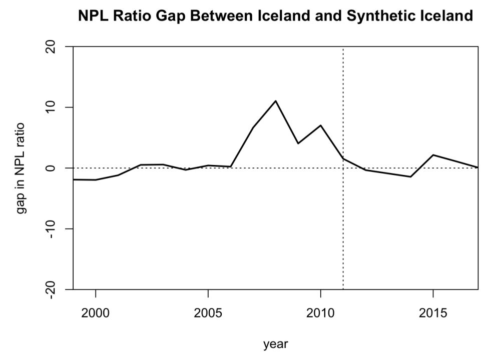
The in-time placebo estimation in Figure B28c shows that Iceland the 2013 placebo debt jubilee has no perceivable effect on the outcome, which suggests that the gap estimated in Figure B28b reflects the impact of the actual 2011 debt jubilee and not a potential lack of predictive power of the synthetic control. Meanwhile, the "leave-one-out" in-place placebo estimation in Figure B28d shows that Iceland significantly outperformed its placebo peers in bringing down the NPL ratio level after 2011. I would consider the effect of the debt jubilee on Iceland to be significant since the estimate effect for Iceland is unusually large relative to the distribution of placebo effects.

Table B8 and B9 showcase the $\boldsymbol{V}$ matrix results for predictors from the cross-validation procedure and the subsequent $\boldsymbol{W}$ matrix results for countries. There is nothing out-of-the-ordinary to report here, and the logic for interpretation is the same as for the previous SCM regression weights output tables, so I will not over-explain here.

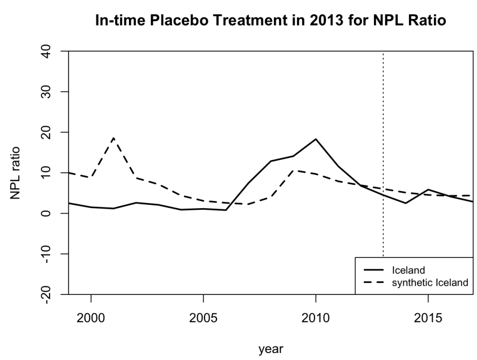
Figure B28: Cross-validation results for NPL ratio



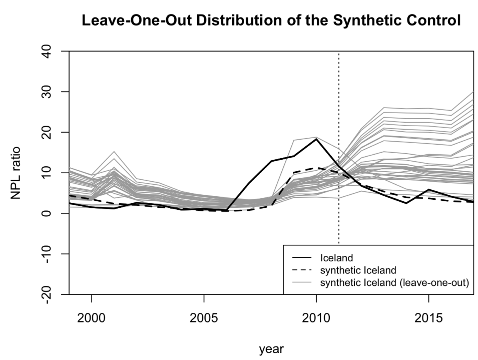
(a) SCM trajectory



(b) SCM effect



(c) In-time placebo test assigned to 2013



(d) In-place placebo "leave-one-out" estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland and synthetic Iceland. Figure(b) shows the effects gap. Figure (c) shows the in-time placebo result, where I rerun the model for the placebo case when the 2011 policy is reassigned to year 2013, two years after the policy actually occurred. Figure (d) shows the in-place "leave-one-out" placebo result, where I use the same out-of-sample validation technique to compute the synthetic control.

Table B8: Cross-validation predictor means before treatment: NPL ratio

	Country	Synthetic Control Weights	Regression Weights	Country	Synthetic Control Weights	Regression Weights
1	Australia	0.00	0.20	Korea, Rep.	0.00	0.03
2	Austria	0.00	0.24	Latvia	0.61	0.19
3	Belgium	0.00	-0.20	Lithuania	0.00	0.33
4	Canada	0.00	-0.07	Luxembourg	0.00	-0.15
5	Chile	0.00	0.04	Mexico	0.00	0.04
6	Colombia	0.00	0.18	Netherlands	0.00	0.46
7	Czech Republic	0.00	-0.33	New Zealand	0.00	-0.32
8	Denmark	0.18	-0.08	Norway	0.00	0.26
9	Estonia	0.00	-0.07	Poland	0.00	-0.30
10	Finland	0.00	-0.16	Portugal	0.00	0.43
11	France	0.00	-0.26	Slovak Republic	0.00	-0.33
12	Germany	0.00	0.00	Slovenia	0.00	-0.36
13	Greece	0.00	0.19	Spain	0.00	0.42
14	Hungary	0.00	-0.00	Sweden	0.00	0.05
15	Ireland	0.00	0.97	Switzerland	0.00	-0.08
16	Israel	0.00	-0.51	Turkey	0.00	0.51
17	Italy	0.00	-0.08	United Kingdom	0.21	0.56
18	Japan	0.00	-0.62	United States	0.00	-0.20

Table B9: Cross-Validation Economic Growth Predictor Means before 2011 Policy Treatment: NPL Ratio

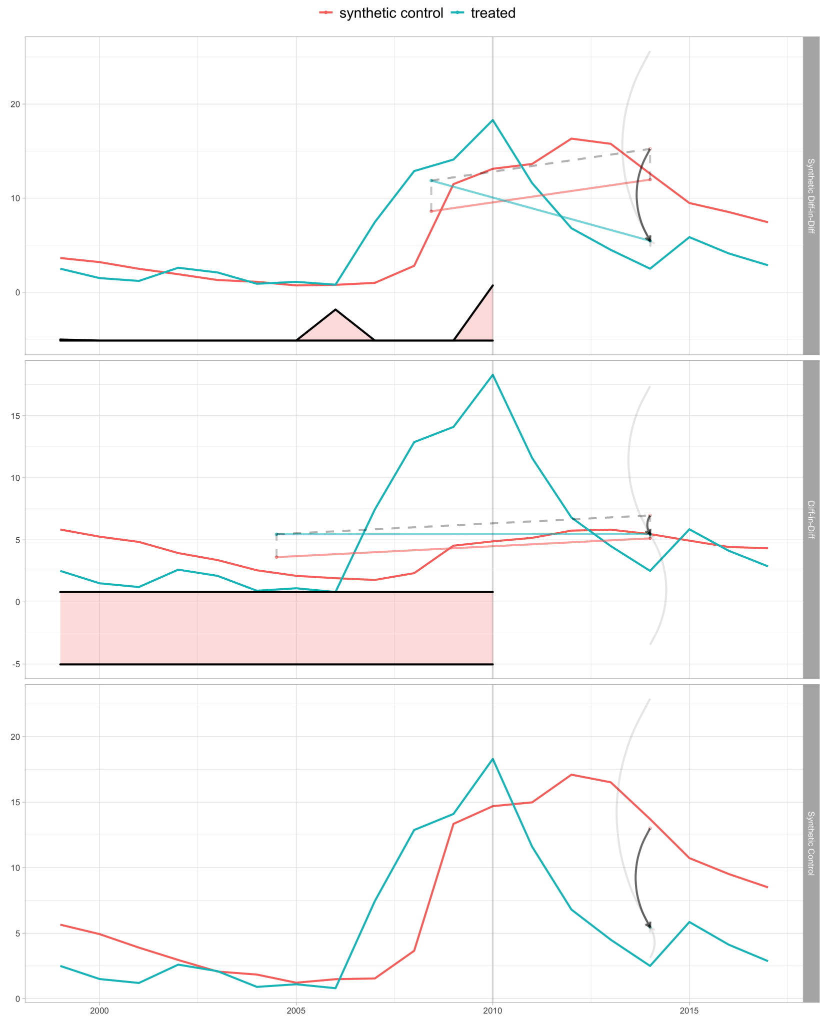
	Treated	Synthetic	Custom V Weights
NPL ratio	10.9	5.8	0.1
Imports	43.6	48.4	0.3
Exports	43.3	42.5	0.3
Industry share	21.5	20.8	0.1
Private credit	203.6	124.9	0.1
Inflation rate	7.6	4.9	0.1
Bank concentration	99.7	61.0	0.0
Domestic savings	22.5	20.6	0.1
International debt	310.5	31.2	0.0

iii. Synthetic Difference-in-Differences (SDID)

Figure B29 and Table B10 display the results from the SDID estimation for NPL ratio alongside estimations for DiD and SCM for comparison purposes. We may observe that SDID captures the parallel trend much better than DiD and SCM do. More specifically, while the DiD estimates are not statistically significant, both SDID and SCM's estimates are, which testify to the power of these two statistical methods.

Meanwhile, SDID seems to have captured the "diff-in-diff" relationship better than the other estimators. SDID yields a more pronounced point estimate on the NPL ratio reduction of around 10 percentage points, while SCM estimates the reduction to be around 7.5 percentage points. In other words, not only does SDID confirm the previous conclusion that the 2011 debt jubilee lowered NPL ratio, it actually claims that the effect could potentially be even greater than SCM tells us.

Figure B29: Estimation trajectories of SDID vs. DiD vs. SCM: NPL ratio



Notes: The three figures are the SDID, DiD, and SCM estimation trajectories in the order from top to bottom. All figures are produced in R using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Table B10: Regression results of SDID vs. DiD vs. SCM: NPL ratio

	SDID	DiD	SCM
Point estimate	-9.769	-1.508	-7.545
Standard error	4.117	4.529	3.947
95% confidence interval	(-17.838, -1.700)	(-10.385, 7.370)	(-15.282, 0.192)

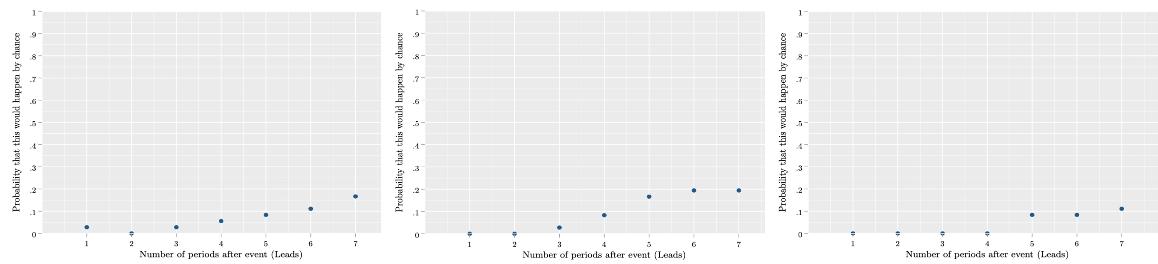
Notes: As detailed in Section 4.3 on the methodology of SDID, the point estimates for SDID, DiD, and SCM are the coefficient values for $\hat{\tau}$ in the estimation equations (12)-(14), and standard errors are calculated accordingly using the methods and statistical tools provided by Arkhangelsky et al. (2020).

4. Private credit

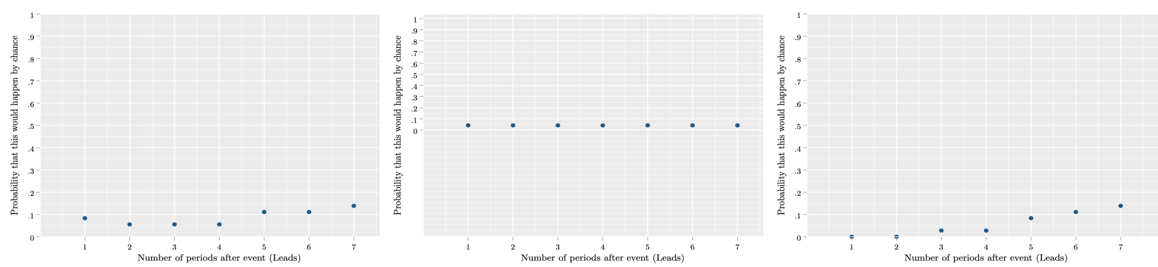
i. p-values for Synthetic Controls Estimation

Figure B30 shows the p-values under six different setups of predictors. The synthetic controls estimation in the main write-up (Figure 4) corresponds to Figure B30b as they all use the same set of macrofinancial predictors for RMSPE optimization. We see that the p-values are below 0.1 and near zero in the years following the treatment. The p-values are low across other predictor constructions as well and make it sufficient to show that the estimation is statistically significant. The likelihood that the private credit boom normalization happened by chance was very low, and it seems reasonable to conclude that the 2011 policy had a substantive effect in bringing down the private sector credit boom.

Figure B30: p-values for synthetic controls estimation of private credit



(a) Macroeconomic setup (b) Macrofinancial setup (c) Following Born et al. (2019)



(d) Following Abadie et al. (2015) (e) Banking & housing setup (f) Labor market setup

Notes: For a detailed list of predictor variables that make up each of the above setup, see Section 6.1.

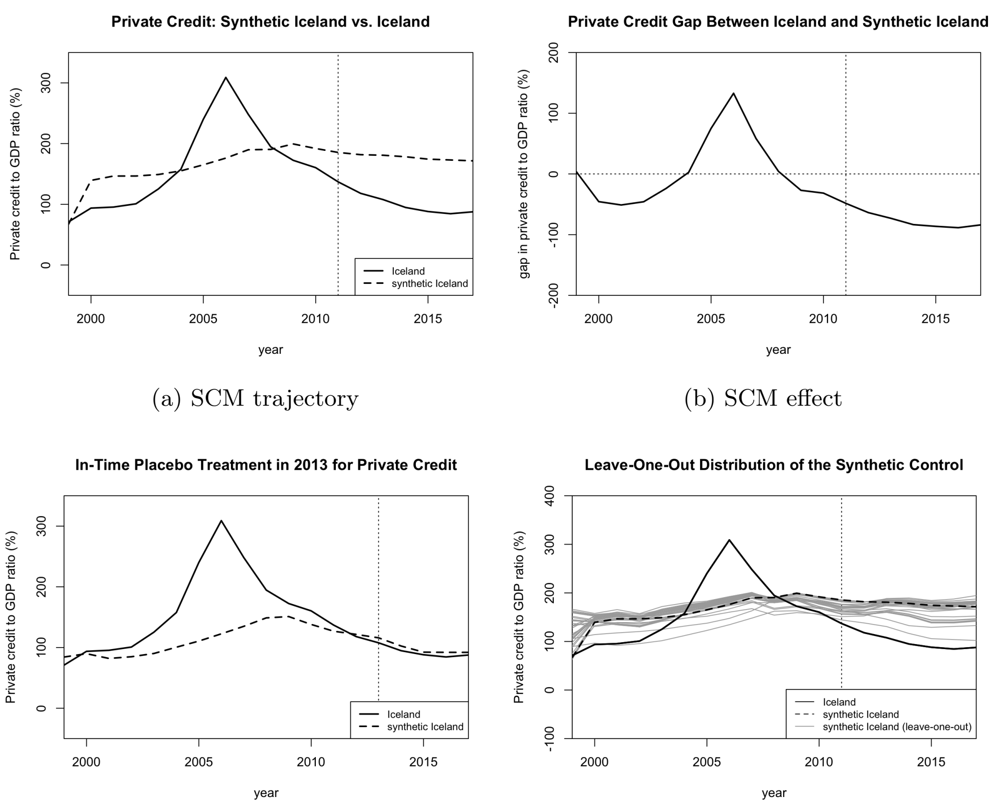
ii. Synthetic Controls with Cross-Validation

Figure B31 displays the SCM trajectory graphs *with* cross-validation, which can be compared with Figure 4, which shows the SCM estimation results *without* cross-validation. We see here that Iceland's private credit to GDP ratio dropped more quickly as synthetic Iceland after 2011, but the pre-trend fit is arguably worse than the fit without cross-validation, so the cross-validation procedure did not improve the pre-trend match.

The in-time placebo estimation in Figure B31c shows that Iceland the 2013 placebo debt jubilee has no perceivable effect on the outcome, which suggests that the gap estimated in Figure B31b reflects the impact of the actual 2011 debt jubilee and not a potential lack of predictive power of the synthetic control. Meanwhile, the "leave-one-out" in-place placebo estimation in Figure B31d shows that Iceland significantly outperformed its placebo peers in bringing down the private credit to GDP ratio level after 2011. I would consider the effect of the debt jubilee on Iceland to be significant since the estimate effect for Iceland is unusually large relative to the distribution of placebo effects.

Table B11 and B12 showcase the $\boldsymbol{V}$ matrix results for predictors from the cross-validation procedure and the subsequent $\boldsymbol{W}$ matrix results for countries. What is potentially worrisome here is that all the $\boldsymbol{W}$ weights shown in Table B11 have concentrated on Denmark and the United States, likely because it is very difficult to find countries with similarly high levels of private credit boom like Iceland. The $\boldsymbol{V}$ matrix results in Table B12 also validate that private credit matters the most amongst all variables, occupying a weight of 0.6. In other words, because very few other countries could match Iceland's private credit level, in order to construct a decent synthetic Iceland pre-trend fit for private credit, the statistical program had to prioritize this variable above all others. This is why the predictor means do not match very well. As a result, the cross-validation procedure ended up choosing *only* Denmark and the United States as the countries to make up the synthetic Iceland. This does not point to the flaw of cross-validation but just testifies to the difficulty of approximating for a variable that is at an usually high level.

Figure B31: Cross-validation results for private credit



(a) SCM trajectory

(b) SCM effect

(c) In-time placebo test assigned to 2013

(d) In-place placebo "leave-one-out" estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland and synthetic Iceland. Figure(b) shows the effects gap. Figure (c) shows the in-time placebo result, where I rerun the model for the placebo case when the 2011 policy is reassigned to year 2013, two years after the policy actually occurred. Figure (d) shows the in-place "leave-one-out" placebo result, where I use the same out-of-sample validation technique to compute the synthetic control.

Table B11: Cross-validation synthetic and regression weights for Iceland: private credit

	Country	Synthetic Control Weights	Regression Weights	Country	Synthetic Control Weights	Regression Weights
1	Australia	0.00	0.20	Korea, Rep.	0.00	0.03
2	Austria	0.00	0.24	Latvia	0.00	0.19
3	Belgium	0.00	-0.20	Lithuania	0.00	0.33
4	Canada	0.00	-0.07	Luxembourg	0.00	-0.15
5	Chile	0.00	0.04	Mexico	0.00	0.04
6	Colombia	0.00	0.18	Netherlands	0.00	0.46
7	Czech Republic	0.00	-0.33	New Zealand	0.00	-0.32
8	Denmark	0.76	-0.08	Norway	0.00	0.26
9	Estonia	0.00	-0.07	Poland	0.00	-0.30
10	Finland	0.00	-0.16	Portugal	0.00	0.43
11	France	0.00	-0.26	Slovak Republic	0.00	-0.33
12	Germany	0.00	0.00	Slovenia	0.00	-0.36
13	Greece	0.00	0.19	Spain	0.00	0.42
14	Hungary	0.00	-0.00	Sweden	0.00	0.05
15	Ireland	0.00	0.97	Switzerland	0.00	-0.08
16	Israel	0.00	-0.51	Turkey	0.00	0.51
17	Italy	0.00	-0.08	United Kingdom	0.00	0.56
18	Japan	0.00	-0.62	United States	0.24	-0.20

Table B12: Cross-validation predictor means before treatment: private credit

	Treated	Synthetic	Custom V Weights
NPL ratio	10.9	2.4	0.0
Imports	43.6	39.2	0.0
Exports	43.3	41.8	0.3
Industry share	21.5	21.0	0.0
Private credit	203.6	188.9	0.6
Inflation rate	7.6	2.3	0.0
Bank concentration	99.7	69.8	0.0
Domestic savings	22.5	23.9	0.0
International debt	310.5	37.6	0.0

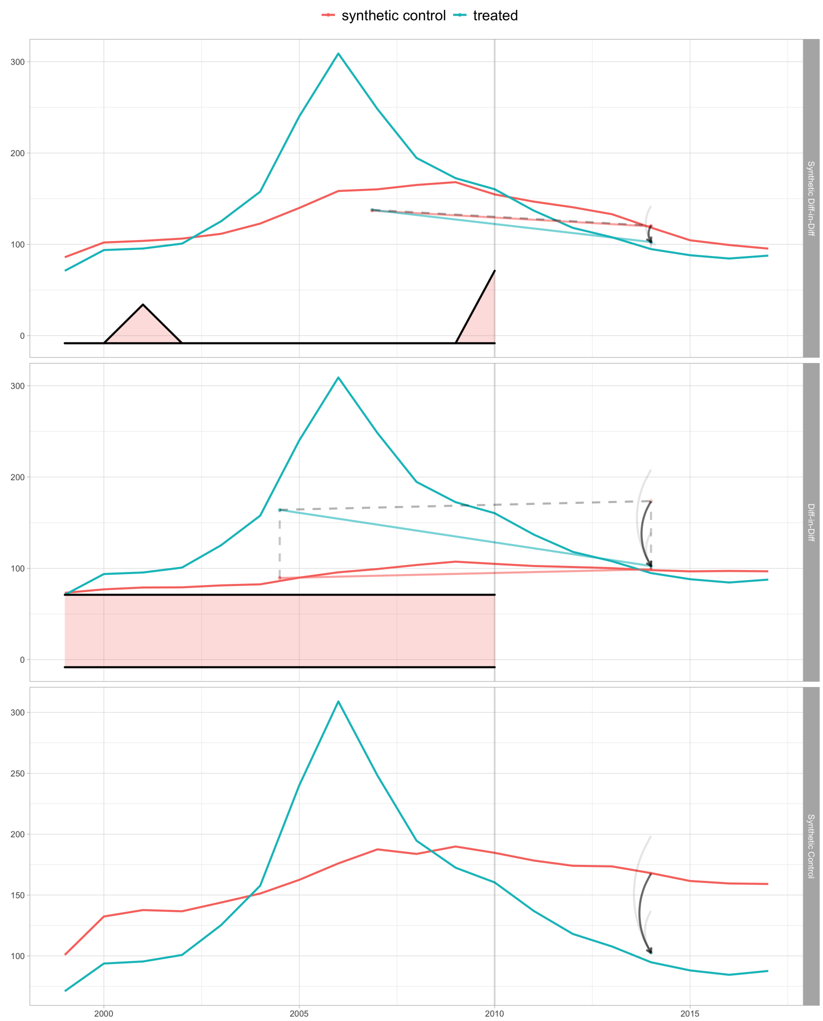
iii. Synthetic Difference-in-Differences (SDID)

Figure B32 and Table B13 display the results from the SDID estimation for private debt to GDP ratio alongside estimations for DiD and SCM for comparison purposes. We may observe that SDID still struggles to capture the parallel trend like DiD and SCM given the unusually high level of private credit in Iceland.

While all three estimators give statistically significant results, SDID gives a point estimate that is much lower in magnitude while having a lower standard error than the other two methods. More specifically, while DiD estimates a private credit reduction of around 71 percentage points and SCM of around 65 percentage points, SDID puts the reduction estimate at merely 18 percentage points. This could potentially be a more accurate estimate since SDID likely captures the "diff-in-diff" relationship better than the other estimators.

Nevertheless, the focus of this paper is not to produce a single most accurate point estimate, which would in fact be too difficult given the assumptions and modeling approach I adopt. The point is to gauge whether there was indeed a sadistically significant effect to Iceland's 2011 policy, which in this case all three methodologies confirm that the policy indeed lowered the private credit level substantively, which is in my view sufficient for conclusion.

Figure B32: Estimation trajectories of SDID vs. DiD vs. SCM: private credit



Notes: The three figures are the SDID, DiD, and SCM estimation trajectories in the order from top to bottom. All figures are produced in R using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Table B13: Regression results of SDID vs. DiD vs. SCM: private credit

	SDID	DiD	SCM
Point estimate	-17.991	-71.088	-65.175
Standard error	11.051	18.954	15.993
95% confidence interval	(-39.651, 3.669)	(-108.237, -33.939)	(-96.521, -33.830)

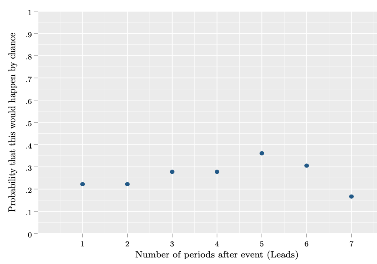
Notes: As detailed in Section 4.3 on the methodology of SDID, the point estimates for SDID, DiD, and SCM are the coefficient values for $\hat{\tau}$ in the estimation equations (12)-(14), and standard errors are calculated accordingly using the methods and statistical tools provided by Arkhangelsky et al. (2020).

5. International debt to GDP ratio

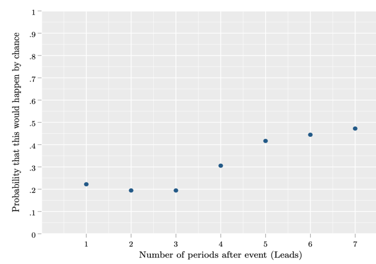
i. p-values for Synthetic Controls Estimation

Figure B33 shows the p-values under six different setups of predictors. The synthetic controls estimation in the main write-up (Figure 5) corresponds to Figure B33d as they all use the same set of macroeconomic and financial predictors for RMSPE optimization following Abadie et al. (2015). We see that the p-values are around 0.2 in the years following the treatment. The p-values are relatively low across other predictor constructions, especially using banking sector and housing variables (Figure B33e), which makes it sufficient to show that the estimation is relatively statistically significant at the 90% level. The likelihood that the private credit boom normalization happened by chance was low, and it seems reasonable to conclude that the 2011 policy had a substantive effect in bringing down the private international debt boom.

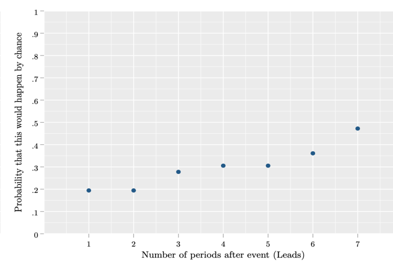
Figure B33: p-values for synthetic controls estimation of international debt



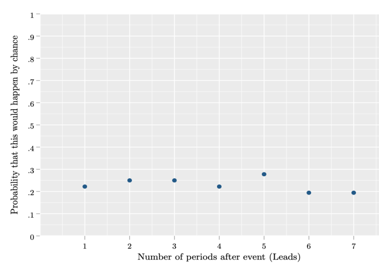
(a) Macroeconomic setup



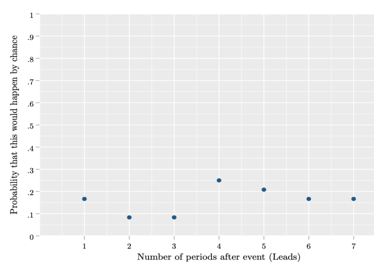
(b) Macrofinancial setup



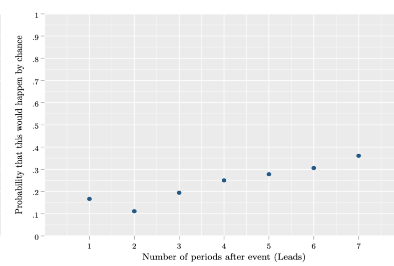
(c) Following Born et al. (2019)



(d) Following Abadie et al. (2015)



(e) Banking & housing setup



(f) Labor market setup

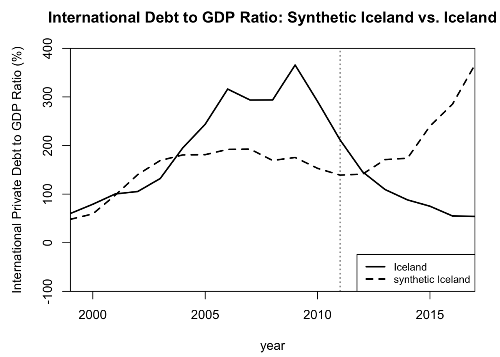
Notes: For a detailed list of predictor variables that make up each of the above setup, see Section 6.1.

ii. Synthetic Controls with Cross-Validation

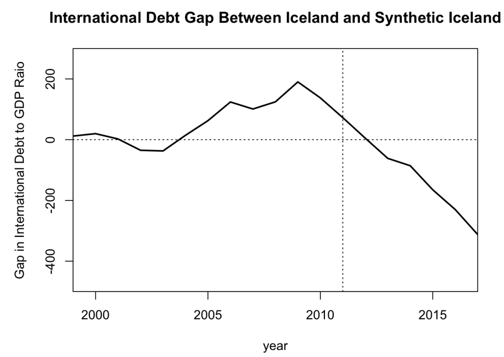
Figure B34 displays the SCM trajectory graphs *with* cross-validation, which can be compared with Figure 5, which shows the SCM estimation results *without* cross-validation. We see here that Iceland's private international debt to GDP ratio dropped more quickly as synthetic Iceland after 2011, but the pre-trend fit is arguably worse than the fit without cross-validation, so the cross-validation procedure did not improve the pre-trend match.

The in-time placebo estimation in Figure B34c shows that Iceland the 2013 placebo debt jubilee has no perceivable effect on the outcome, which suggests that the gap estimated in Figure B34b reflects the impact of the actual 2011 debt jubilee and not a potential lack of predictive power of the synthetic control. Meanwhile, the "leave-one-out" in-place placebo estimation in Figure B34d shows that Iceland significantly outperformed its placebo peers in bringing down the international debt ratio level after 2011. I would consider the effect of the debt jubilee on Iceland to be significant since the estimate effect for Iceland is unusually large relative to the distribution of placebo effects.

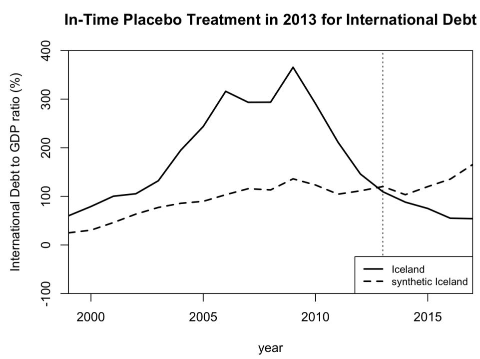
Figure B34: Cross-validation results for international debt



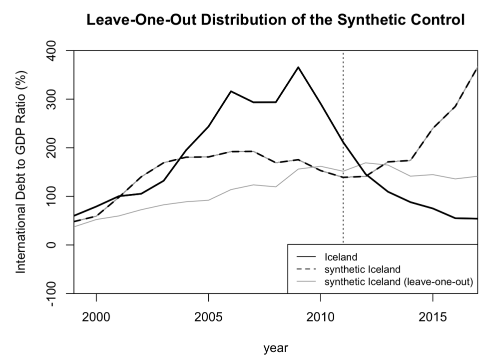
(a) SCM trajectory



(b) SCM effect



(c) In-time placebo test assigned to 2013



(d) In-place placebo "leave-one-out" estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland and synthetic Iceland. Figure(b) shows the effects gap. Figure (c) shows the in-time placebo result, where I rerun the model for the placebo case when the 2011 policy is reassigned to year 2013, two years after the policy actually occurred. Figure (d) shows the in-place "leave-one-out" placebo result, where I use the same out-of-sample validation technique to compute the synthetic control.

Table B14 and B15 showcase the $\boldsymbol{V}$ matrix results for predictors from the cross-validation procedure and the subsequent $\boldsymbol{W}$ matrix results for countries. What is potentially worrisome here is that all the $\boldsymbol{V}$ weights shown in Table B15 have concentrated on international debt with a ratio of 1.0, which means that this singular variable was so unique for Iceland such that it was the most important predictor to consider. In other words, because very few other countries could match Iceland's international debt level, in order to construct a decent synthetic Iceland pre-trend fit for international debt level, the statistical program had to prioritize this variable and this variable only above all others. This is why the predict means do not match very well. As a result, the cross-validation procedure ended up choosing *only* Luxembourg as the country to make up the synthetic Iceland. This does not point to the flaw of cross-validation but just testifies to the difficulty of approximating for a variable that is at an usually high level.

Table B14: Cross-validation synthetic and regression weights for Iceland: international debt

	Country	Synthetic Control Weights	Regression Weights	Country	Synthetic Control Weights	Regression Weights
1	Australia	0.00	0.37	Korea, Rep.	0.00	0.05
2	Austria	0.00	0.13	Latvia	0.00	0.16
3	Belgium	0.00	-0.06	Lithuania	0.00	0.22
4	Canada	0.00	0.13	Luxembourg	1.00	-0.06
5	Chile	0.00	-0.20	Mexico	0.00	0.19
6	Colombia	0.00	0.17	Netherlands	0.00	0.57
7	Czech Republic	0.00	-0.28	New Zealand	0.00	-0.43
8	Denmark	0.00	-0.44	Norway	0.00	0.08
9	Estonia	0.00	-0.21	Poland	0.00	-0.19
10	Finland	0.00	-0.18	Portugal	0.00	0.43
11	France	0.00	-0.15	Slovak Republic	0.00	-0.06
12	Germany	0.00	0.08	Slovenia	0.00	-0.27
13	Greece	0.00	0.14	Spain	0.00	0.49
14	Hungary	0.00	-0.22	Sweden	0.00	-0.07
15	Ireland	0.00	0.89	Switzerland	0.00	-0.05
16	Israel	0.00	-0.59	Turkey	0.00	0.32
17	Italy	0.00	-0.14	United Kingdom	0.00	0.60
18	Japan	0.00	-0.41			

Table B15: Cross-validation predictor means before treatment: international debt

	Treated	Synthetic	Custom V Weights
International private debt	260.2	172.5	1.0
GDP per-capita	44006.5	104055.0	0.0
Imports	41.9	138.8	0.0
Exports	39.5	168.5	0.0
Industry share	21.4	13.6	0.0
Inflation rate	6.1	2.4	0.0
Capital formation	23.6	19.3	0.0
Secondary education completion	31.0	28.7	0.0

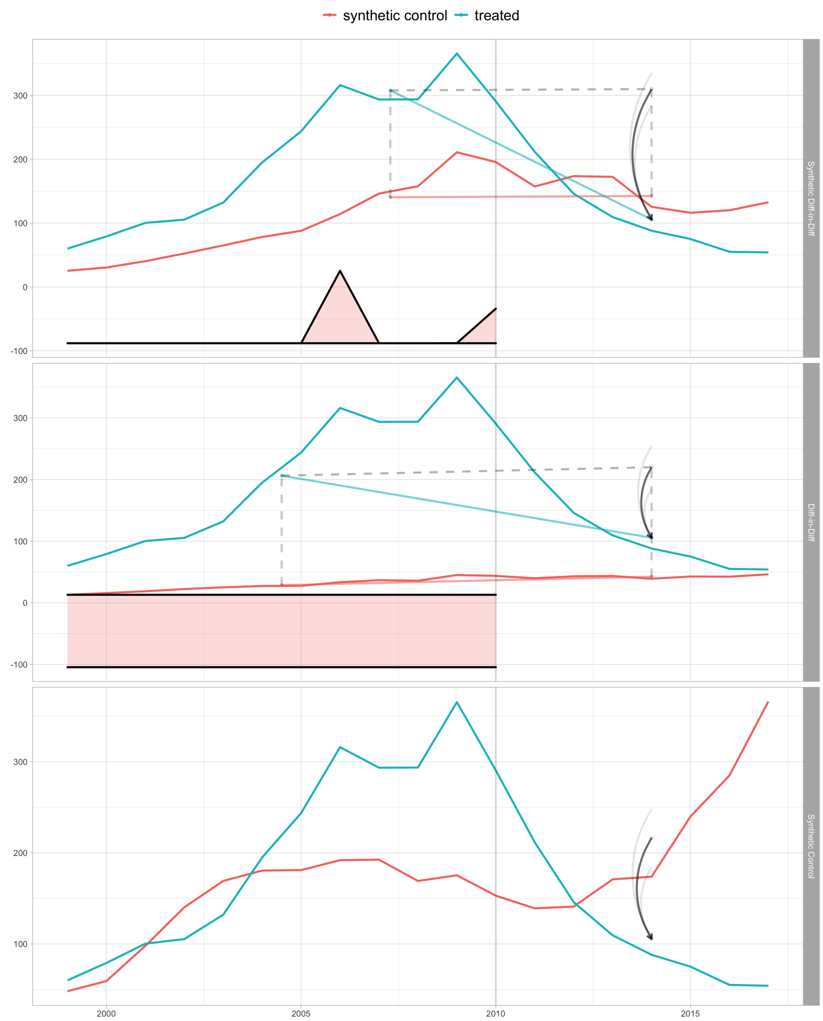
iii. Synthetic Difference-in-Differences (SDID)

Figure B35 and Table B16 display the results from the SDID estimation for international debt to GDP ratio alongside estimations for DiD and SCM for comparison purposes. We may observe that SDID captures the parallel trend much better than DiD and SCM do.

While all three estimators give statistically significant results, SDID gives a point estimate that is much larger in magnitude while having a lower standard error than the other two methods. More specifically, while DiD estimates a private credit reduction of around 114 percentage points and SCM of around 111 percentage points, SDID puts the reduction estimate at 204 percentage points. This could potentially be a more accurate estimate since SDID likely captures the "diff-in-diff" relationship better than the other estimators. In other words, not only does SDID confirm the previous conclusion that the 2011 debt jubilee lowered the international debt to GDP ratio, it actually claims that the effect could potentially be even greater than SCM tells us.

Nevertheless, the focus of this paper is not to produce a single most accurate point estimate, which would in fact be too difficult given the assumptions and modeling approach I adopt. The point is to gauge whether there was indeed a sadistically significant effect to Iceland's 2011 policy, which in this case all three methodologies confirm that the policy indeed lowered the international debt to GDP level substantively, which is in my view sufficient for conclusion.

Figure B35: Estimation trajectories of SDID vs. DiD vs. SCM: international debt



Notes: The three figures are the SDID, DiD, and SCM estimation trajectories in the order from top to bottom. All figures are produced in R using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Table B16: Regression results of SDID vs. DiD vs. SCM: international debt

	SDID	DiD	SCM
Point estimate	-204.276	-114.343	-110.995
Standard error	12.040	19.266	14.173
95% confidence interval	(-227.875, -180.677)	(-152.105, -76.581)	(-138.774, -83.217)

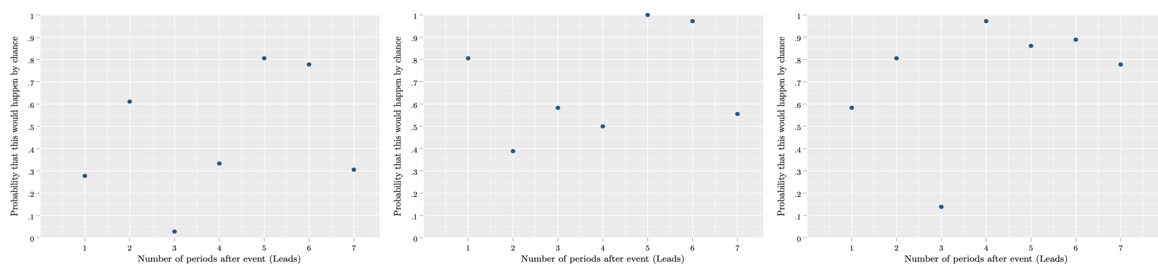
Notes: As detailed in Section 4.3 on the methodology of SDID, the point estimates for SDID, DiD, and SCM are the coefficient values for $\hat{\tau}$ in the estimation equations (12)-(14), and standard errors are calculated accordingly using the methods and statistical tools provided by Arkhangelsky et al. (2020).

6. Income inequality

i. p-values for Synthetic Controls Estimation

Figure B36 shows the p-values under six different setups of predictors. The synthetic controls estimation in the main write-up (Figure 8) corresponds to Figure B24d as they all use the same set of predictors for RMSPE optimization. We see that the p-values are quite high and fluctuate between 0.6 to 0.8 in the years following the treatment. This shows that the estimation is not statistically significant. In other words, the post-2011 changes in Iceland's income shares had a 60-80% probability of happening due to factors not limited to the 2011 debt jubilee policy, which is much higher than what we would ideally like to see in a placebo test. It seems reasonable to conclude that our estimation for income share is simply not statistically significant across the board.

Figure B36: p-values for synthetic controls estimation of income shares



(a) Bottom 50% income share (b) Middle 40% income share (c) Top 10% income share

Notes: For all the above estimations, I use the "macroeconomic setup." For the detailed list of predictor variables that make up of that setup, see Section 6.1.

ii. Synthetic Controls with Cross-Validation

For purpose of concision, here I only include my cross-validation results for the bottom 50% income share. The assumption is that the other two income brackets (middle 40% and top 10%) likely will not vary dramatically in statistical significant as the one for bottom 50%, and the additional p-values and SDID results should be sufficient in providing enough robustness checks for the outcome of income share.

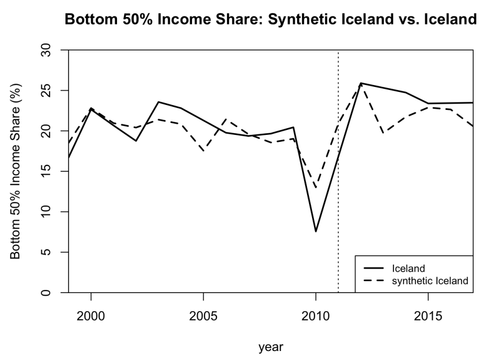
Figure B37 display the SCM trajectory graphs *with* cross-validation, which can be compared with Figure 8, which displays the SCM estimation results *without* cross-validation. We see here that cross-validation seems to have improved our previous results. We observe that Iceland's bottom 50% income share performed slightly better compared to synthetic Iceland after the treatment took place in 2011, but the pre-trend fit is still not perfect and the magnitude of the effect is less than 5%.

The in-time placebo estimation in Figure B37c shows that Iceland the 2013 placebo debt jubilee also had similar positive effect on the outcome, which suggests that the gap estimated in Figure B37b does not reflect a statistically significant impact of the actual 2011 debt jubilee. Meanwhile, the "leave-one-out" in-place placebo estimation in Figure B37d shows that Iceland's bottom 50% income share performance was largely in line with its placebo peers and only slightly better right after the treatment. From there it would be hard to consider the effect of the debt jubilee on Iceland to be significant since the estimated effect for Iceland is not larger relative to the distribution of placebo effects.

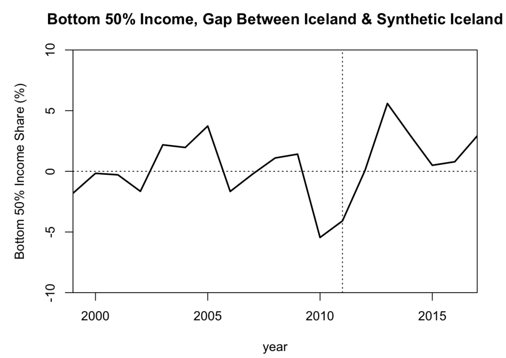
In sum, the in-time and in-place placebo estimates show that Iceland's bottom 50% income share largely trended in line with its OECD peers after 2011, and there is a potential lack of predictive power of the synthetic control. This could be because cross-validation still relies on fitting pre-trends and does not entirely relax the parallel pre-trend assumptions needed for synthetic controls estimations. The SDID estimation might thus be able to better capture the "diff-in-diff" effect, as we shall see later.

Table B17 and B18 showcase the $\boldsymbol{V}$ matrix results for predictors from the cross-validation procedure and the subsequent $\boldsymbol{W}$ matrix results for countries. There is nothing out-of-the-ordinary to report here, and the logic for interpretation is the same as for the previous SCM regression weights output tables, so I will not over-explain here.

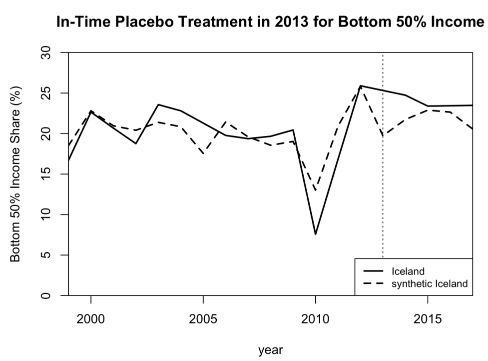
Figure B37: Cross-validation results for bottom 50% income share



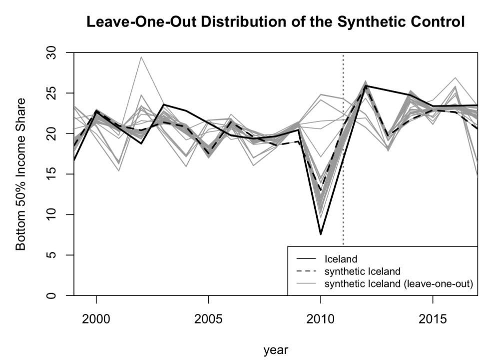
(a) SCM trajectory



(b) SCM effect



(c) In-time placebo test assigned to 2013



(d) In-place placebo "leave-one-out" estimates

Notes: Figure (a) shows the synthetic controls estimation trajectory for Iceland and synthetic Iceland. Figure(b) shows the effects gap. Figure (c) shows the in-time placebo result, where I rerun the model for the placebo case when the 2011 policy is reassigned to year 2013, two years after the policy actually occurred. Figure (d) shows the in-place "leave-one-out" placebo result, where I use the same out-of-sample validation technique to compute the synthetic control.

Table B17: Cross-validation synthetic and regression weights for Iceland: bottom 50% income share

	Country	Synthetic Control Weights	Regression Weights	Country	Synthetic Control Weights	Regression Weights
1	Australia	0.00	0.12	Korea, Rep.	0.00	0.15
2	Austria	0.00	0.01	Latvia	0.00	0.26
3	Belgium	0.00	-0.05	Lithuania	0.00	-0.05
4	Canada	0.26	0.22	Luxembourg	0.00	-0.01
5	Chile	0.00	-0.02	Mexico	0.00	-0.20
6	Colombia	0.00	-0.02	Netherlands	0.00	-0.10
7	Czech Republic	0.00	-0.15	New Zealand	0.00	0.03
8	Denmark	0.23	0.30	Norway	0.00	0.05
9	Estonia	0.20	0.17	Poland	0.00	-0.25
10	Finland	0.00	-0.06	Portugal	0.00	0.10
11	France	0.00	-0.09	Slovak Republic	0.00	-0.03
12	Germany	0.00	-0.06	Slovenia	0.00	-0.07
13	Greece	0.00	0.07	Spain	0.00	0.18
14	Hungary	0.00	0.12	Sweden	0.00	-0.07
15	Ireland	0.00	0.14	Switzerland	0.00	0.12
16	Israel	0.00	-0.15	Turkey	0.30	0.33
17	Italy	0.00	-0.15	United Kingdom	0.00	0.20
18	Japan	0.00	-0.04			

Table B18: Cross-validation predictor means before treatment: bottom 50% income share

	Treated	Synthetic	Custom V Weights
Bottom 50% income share	19.0	19.1	0.3
GDP per-capita	44006.5	31821.2	0.0
Imports	41.9	40.6	0.3
Exports	39.5	40.6	0.4
Industry share	21.4	25.3	0.0
Private credit	193.9	102.8	0.0
Inflation rate	6.1	4.8	0.0

iii. Synthetic Difference-in-Differences (SDID)

Figures B38 to B40 and Tables B19 to B21 display the results from the SDID estimation for the bottom 50%, middle 40%, and top 10% income shares alongside estimations for DiD and SCM for comparison purposes. We may observe that SDID has successfully captured the parallel trend much better and yielded lower standard errors than the other two methods, but the post-treatment effects are still not statistically significant enough given the low point estimate and high standard errors. In sum, SDID has successfully captured the "diff-in-diff" relationship better than the other estimators, but it is just that the 2011 debt jubilee policy effects on income shares may simply not be significant enough or that there are too many other confounding factors that could influence the income shares such that it is hard to conclude the isolated impact of the debt jubilee policy.

Figure B38: Estimation trajectories of SDID vs. DiD vs. SCM: bottom 50% income share

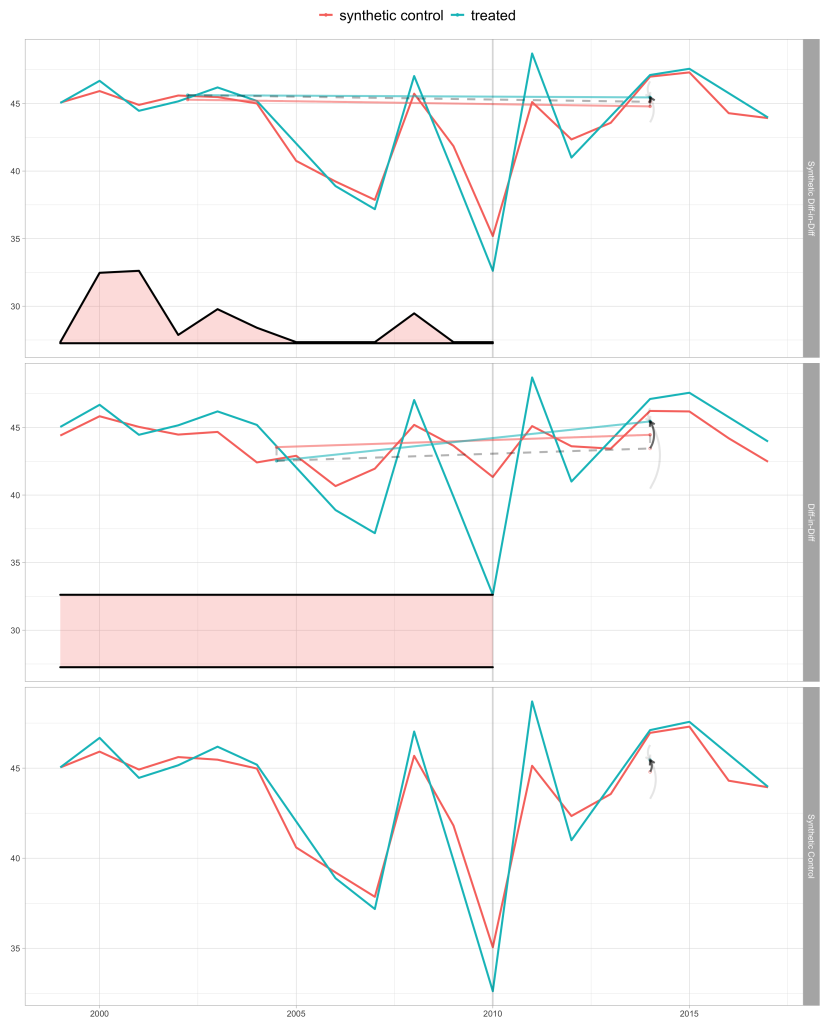
Notes: The three figures are the SDID, DiD, and SCM estimation trajectories in the order from top to bottom. All figures are produced in R using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Table B19: Regression results of SDID vs. DiD vs. SCM: bottom 50% income share

	SDID	DiD	SCM
Point estimate	0.363	1.607	0.265
Standard error	0.761	1.665	0.713
95% confidence interval	(-1.129, 1.854)	(-1.656, 4.870)	(-1.132, 1.663)

Notes: As detailed in Section 4.3 on the methodology of SDID, the point estimates for SDID, DiD, and SCM are the coefficient values for $\hat{\tau}$ in the estimation equations (12)-(14), and standard errors are calculated accordingly using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Figure B39: Estimation trajectories of SDID vs. DiD vs. SCM: middle 40% income share



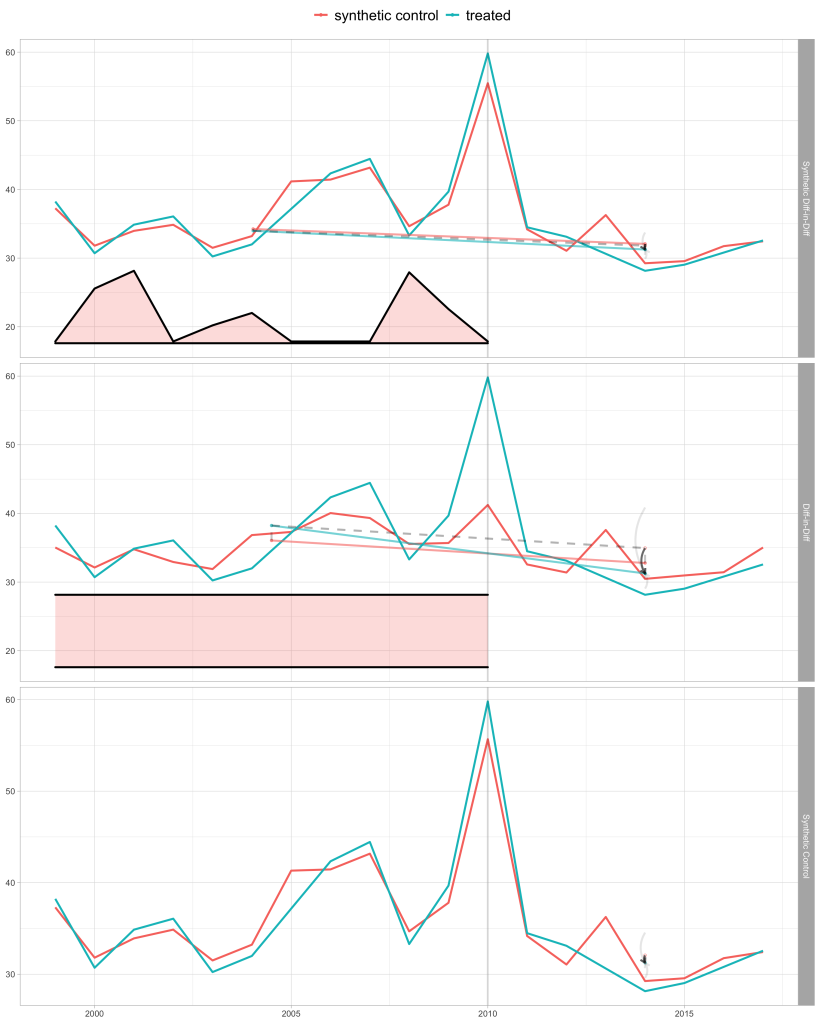
Notes: The three figures are the SDID, DiD, and SCM estimation trajectories in the order from top to bottom. All figures are produced in R using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Table B20: Regression results of SDID vs. DiD vs. SCM: middle 40% income share

	SDID	DiD	SCM
Point estimate	0.328	2.005	0.661
Standard error	0.793	1.492	0.728
95% confidence interval	(-1.227, 1.882)	(-0.918, 4.929)	(-0.767, 2.089)

Notes: As detailed in Section 4.3 on the methodology of SDID, the point estimates for SDID, DiD, and SCM are the coefficient values for $\hat{\tau}$ in the estimation equations (12)-(14), and standard errors are calculated accordingly using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Figure B40: Estimation trajectories of SDID vs. DiD vs. SCM: top 10% income share



Notes: The three figures are the SDID, DiD, and SCM estimation trajectories in the order from top to bottom. All figures are produced in R using the methods and statistical tools provided by Arkhangelsky et al. (2020).

Table B21: Regression results of SDID vs. DiD vs. SCM: top 10% income share

	SDID	DiD	SCM
Point estimate	-0.557	-3.698	-0.816
Standard error	1.237	3.074	1.210
95% confidence interval	(-2.981, 1.867)	(-9.724, 2.327)	(-3.187, 1.556)

Notes: As detailed in Section 4.3 on the methodology of SDID, the point estimates for SDID, DiD, and SCM are the coefficient values for $\hat{\tau}$ in the estimation equations (12)-(14), and standard errors are calculated accordingly using the methods and statistical tools provided by Arkhangelsky et al. (2020).

C. Discussion and Extensions

In this section, I bring up various debates related to my thesis, from the econometrics behind synthetic controls to Modern Monetary Theory. They do not intimately affect my core results and conclusions but could be helpful to those interested in learning more.

1. Challenges of Applying Synthetic Controls to Iceland

Since the beginning of my research process, I have aimed to apply frameworks from notable synthetic controls papers to Iceland without success. I have collected the same range of annual, cross-country data set that then allowed me to construct comparisons for characteristics like economic growth and labor market outcomes. However, my various estimation results seem to suggest that it has been particularly difficult to construct accurate pre-trends, and my estimates are often not as robust most published studies. The synthetic controls frameworks have in general struggled to produce a high-quality synthetic Iceland for various outcomes, and I now explore some of the reasons behind this.[15]

One reason could be that synthetic controls are not very suitable for outlier units or events. Most of the notable synthetic controls papers in cross-country contexts focus on regions that are "middle-of-the-pack" and with stable trends for data. For example, Abadie et al. (2010) estimates for California, Abadie et al. (2015) for West Germany, and Born et al. (2019) for the U.K. The time periods and policy events that these papers estimate are also not periods with dramatic fluctuations. Iceland, in contrast, is a fairly idiosyncratic country with its tiny population and remote geographical location amongst other characteristics, and the 2008 financial crisis also happened to be a time when Iceland's various financial metrics stood at levels that are multiple standard deviations above or below the OECD median. Therefore, these structural reasons inherently hinder SCM from constructing high-quality counterfactuals.

I have tried my best to choose predictor variables that have strong predictive powers – such as household consumption, housing price growth rate, banking sector concentration, labor participation rate and so on – and let my data inform my research. But these variables do not seem to have much predictive power. Not only is the causality of the 2011 policies hard to prove, many plausible constructions of SCM trajectories do not look very convincing, and overall the paper leaves us with some very uncertain results. Iceland is a tiny economy that does not seem to yield great fits based on most of the existing models and literature. There are very different dynamics between Iceland and the rest of OECD, so it was not a given that a combination of OECD countries would reproduce similar patterns like what was seen in Iceland and act as predictors for future trends.

The common approach to SCM, and even with cross-validation, essentially involves a lot of ad-hoc decisions by the researcher to try out estimations variable by variable and so on, which I have done with my six groups of predictors specified in Section 6.1. This procedure is underpinned by the assumption that there is some underlying one-dimensional process through which all the other OECD countries and Iceland are related to each other. And, by using some weighted average of characteristic variables, one can supposedly deconstruct this underlying process. But this might not be the case for a country like Iceland, and my estimations might still be far off from arriving at a "true" model that best approximates the results. After

[15] I am grateful for my econometrics professors Chris Sims and Mikkel Plagborg-Moller, who offered many of the thoughts below and truly broadened my thinking on the econometric frameworks for this paper.

all, my current procedure simply uses an estimator that is plausible, and I am essentially extrapolating a distribution theory for it without actually using the "true" model.

There are approaches to SCM that are more principled, where one can explicitly model the parameters and data to arrive at better estimates. For example, I can use a dynamic factor model, which captures the same ideas as SCM. However, both models assume that the pre-trends match well. So, if this assumption does not hold, which it does not seem to with SCM, then neither would the dynamic factor model work very well. Therefore, surely I could be getting imperfect matches because of the modeling approach, but it is also highly likely that Iceland is just too unique of a country to apply SCM or dynamic factor model on in general.

Another idea is to use local projections, which has become more popular in the macro-financial literature since Jordà (2005) and is just another way to estimate VAR IRFs (vector autoregression impulse response functions). Funke et al. (2020) study the question of how economies perform under populist leaders using SCM. But in additional to the classic SCM estimation, they also run a dynamic event study using the local projections model of Jordà (2005). The local projections allow the authors to trace the dynamic path of GDP per capita after a populist comes to power. They plot the cumulative path of a dependent variable in response to the event of interest (i.e. the start of a populist leadership episode) and compare this projected path to that of a placebo treatment (i.e. the entering into office of a non-populist government).

Unfortunately, I cannot entirely adapt this local projections model into my Icelandic context given that I only have one "shock" in my study while local projections and VAR require the time series to undergo multiple shocks. In my case, even though I have a fairly long data series, I do not have multiple examples of policy interventions that I am trying to estimate. In VAR and local projections, one is supposed to have many policies vary up and down in many different periods, so that the estimates could demonstrate their effects over time. With only one treatment unit (Iceland) and event (debt jubilee), I would only be able to replicate Funke et al. (2020) after constructing a much more complete and comprehensive data set reflecting the status and magnitude of debt forgiveness programs across OECD countries, which could be a subject matter for future studies.

2. Why do the W weights have to sum up to 1 and be non-negative?

Recall from Section 5.1.3 that a synthetic control is defined as a weighted average of the units in the donor pool. Formally, a synthetic control can be represented by a $J \times 1$ vector of weights, $\boldsymbol{W} = (w_2, ...w_{J+1})'$. Given a set of weights, $\boldsymbol{W}$, the synthetic control estimators of Y_{1t}^N and α_{1t} are:

$$\widehat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt} \tag{15}$$

$$\widehat{\alpha}_{1t} = Y_{1t} - \widehat{Y}_{1t}^N. \tag{16}$$

The weights are restricted to be non-negative and to sum to one, so synthetic controls are weighted averages of the units in the donor pool.

Why do the weights have to sum to one and be non-negative? This is a point of debate for many scholars. In the context of Iceland, which is a very small economy relative to other OECD countries, forcing the weights to sum to one is essentially artificially imposing that the smaller countries have a much higher likelihood to get selected for the synthetic control group, since even adding a tiny fraction of a variable value from larger countries could dramatically skew the synthetic control match. Meanwhile, Abadie (2020) believes that relaxing these constraints would allow for extrapolation, but some may say that the model is currently already allowing for extrapolation.

Prof. Chris Sims has told me that one of his many objections to the synthetic control methods is that there is no reason in principle why the weights have to sum up to one or be positive. In fact, say Iceland and Norway have competing fishing industries, then when Norway gets a lot of fish, Iceland should be getting less fish, and that should be reflected by putting a negative synthetic weight on Norway. In other words, negative weights do not make it worse for interpretation and should in fact be the more preferred approach. Because the coefficient scaling could be arbitrary, Prof. Sims believes that the requirement of non-negative weights is a bigger issue than the requirement of weights summing up to one. However, it is doubtful whether changing the technical constraints for weights would dramatically improve the estimation for Iceland. This warrants further discussion and research.

3. Other Frameworks Under Synthetic Controls Methods

A number of scholars have explored the shortcomings of SCM as detailed by Abadie himself in Abadie (2019). One of the main critiques of the method is that though it limits the extrapolation bias that can occur when units with different pre-treatment characteristics are combined using a traditional adjustment like in a linear regression, SCM could be susceptible to *interpolation* biases (Kellogg et al. 2020). However, because most other commonly used estimators, such as nearest neighbor matching, suffer from the opposite drawback of potentially extrapolating too much when suitable untreated units are unavailable, Kellogg et al. (2020) believed that one could combine the synthetic control (SC) and matching estimators to defend against the weaknesses and preserve the strengths in both kinds. They name it the matching and synthetic control (or MASC) as a model averaging estimator that combines the standard SC and matching estimators. The MASC estimator' cross-validated weight can be solved for in closed-form, making it only marginally more difficult to implement than the usual SC estimator.

Abadie and L'Hour (2019) propose the penalized SC estimator. The penalized SC and MASC estimators are different, but related in that both assign weights to untreated units while taking into consideration their distance from the treated unit in terms of pre-treatment characteristics. Kellogg et al. (2020) showed that the penalized SC estimator is the solution to a constrained version of the problem implicitly solved by the MASC. Thus, the MASC should represent a more flexible model than the penalized SC.

One way to improve the SC estimation and imperfect pre-trend fit could be through demeaning the data. Ferman and Pinto (2021) propose two alternatives to the original SC method. One of them is through a slight modification in the SC method where data are demeaned using the pre-intervention period, and then the SC estimator is constructed with the demeaned data. The authors show that, if selection into treatment is only correlated with individual fixed effects (which is essentially the identification assumption of the DiD model), then this demeaned SC estimator will be asymptotically unbiased. On the other hand, when the number of control units is fixed, the estimated SC weights will generally not converge to the weights that reconstruct the factor loadings of the treated unit, which means that the SC estimator will be asymptotically biased if treatment assignment is correlated with the unobserved heterogeneity. In my research, the robustness of my estimation unfortunately did not improve dramatically after demeaning the data, but I believe that this is an additional helpful framework that one can try out.

It is unclear whether machine learning could be helpful in improving my estimation. Athey et al. (2018) discuss matrix completion methods for causal panel data models and would fit into the context of SCM. Mühlbach and Nielsen (2021) propose a new "tree-based" synthetic controls method, in which they recast the synthetic controls for evaluating policies as a counterfactual prediction problem and replace its linear regression with a nonparametric model inspired by machine learning. Meanwhile, if one observes a large number of covariates in the data set, maybe it would be helpful to do some LASSO regularization that could potentially yield different estimation results. These additional machine learning frameworks are helpful in guiding through my research on synthetic controls, but I did not end up applying any of them to mine because machine learning likely would not have helped when I only have one treatment unit that also happens to be quite idiosyncratic like Iceland.

Lastly, one may be interested in synthetic control frameworks that do not require the parallel trends assumption and are more suitable for estimating causality under specific real-world, micro-level contexts. Li and Shankar (2020) study the rapid growth of omnichannel retailing, but the causal effect cannot be estimated through a randomized control field experiment. However, a crucial identifying assumption for the SC method is the parallel trends assumption, which may not hold in real data like in the omnichannel context. They hence propose a new two-step synthetic control (TSSC) method that comprises a new test for the parallel trends assumption in the first step, and the application of an appropriate synthetic control method in the second step. This approach unifies the synthetic control and the modified synthetic control (MSC) methods, and it may be helpful for researchers interested in studying these scenarios.

4. Was the Icelandic Debt Jubilee an MMT Application?

A further motivation for this research project is that the current discussions about debt forgiveness now take place under the greater backdrop of the increasing popularity of Modern Monetary Theory (MMT) in American political discourse, whose central argument is that large deficits can be money financed without major problems (Kelton 2019). But the scope of MMT proposals go beyond just deficits and also have implications on how traditional debt crises shall be handled and how economic growth shall be financed. Dalio (2019) writes that MMT is essentially a form of fiscal-monetary policy coordination aimed at directly providing printed money to spenders, and as long as the debts are denominated in domestic currencies, such spending could usually smooth out downturns (with many caveats).

According to Dalio (2019), some historical cases where quasi-MMT policies were implemented include: big debt write-down followed by big money creation (the 1930s Great Depression in the U.S., Iceland in 2008); helicopter money (US Veterans' Bonus during the Great Depression); central bank directly monetizes government programs and promises to cover debt (Germany in the 1930s under the Nazi regime, UK during WWI)... Many of these fiscal-monetary coordination programs are quite successful in their own ways, and MMT advocates have often pointed to them as proof of concept for widely applying MMT today.

For instance, the German central bank issued "Mefo bills" – treasury bills camouflaged as commercial bills. The central bank facilitated the issuance of these bills to keep the rearmament operation under wraps, and it provided a discountability guarantee, convertible into Reichsmark upon demand. Mefo bills had no actual existence or operations and was solely a balance sheet entity. The German government made sure that the bills were never exchanged for Reichsmark, so these bills enabled the German Reich to run a greater deficit than it would normally have been able to. This is a direct fiscal-monetary coordination where the central bank printed money to cover debt payments and monetized government programs. MMT economist Bill Mitchell wrote how the Mefo bills were a successful implementation of MMT.[16] Bossone and Labini (2016) wrote how the Mefo bills lend support to Italy's "fiscal money" proposal.[17] There are many similar articles arguing for the power and efficacy of Mefo bills.[18]

These writings are popular when monetary policy now seems to be out of tools, but very likely they're all just drawing simple correlations and not looking deeply enough into the caveats. The Mefo bill case was directly refuted by a prominent German economist historian, Albrecht Ritschl, who wrote to me that:

> "From a legal vantage point, the Mefo bills were treasury bills camouflaged as commercial bills. The central bank's role was twofold. On the one hand, it facilitated the issuance of these bills to keep the rearmament operation under wraps. On the other, it provided a discountability guarantee. The second aspect could be the interesting one but I don't think it really is. Work creation bills were issued openly by government agencies at the same time, and with similar conditions. The difference between those and Mefo is that the former did show up in the published central government debt accounts while the latter did not. Also, the Mefo bills were not money, as they didn't circulate or weren't discounted as collateral for monetary instruments."

Given how the 1930s Germany example seems to have been disproven as an application of MMT, it is likely that many other historical cases are also far from what MMT economists have claimed. Therefore, this thesis is partly motivated by my personal confusion and suspicion of MMT economists' tendency to take credit for these historical cases of fiscal-monetary coordination.

In the context of Iceland, Patrizia Baudino kindly offered her insights on whether the Icelandic debt jubilee has much to do with the portrayal by the media:

> "I see the discussion in the article by Krugman about the Icelandic experience. One word of caution though – the crisis response there relied on placing all foreign-held positions in the old banks, imposing losses on these creditors and taking a long time to wind down the old banks. For the Icelandic people, the flip side of the coin was to accept the restrictions of capital controls for a decade. It may also be tempting to interpret the debt relief to Icelandic borrowers as somewhat coordinated with the steps on the monetary policy front, but the reality is more complex, as these measures were taken in stages, and not necessarily in a coordinated, well-designed way. They could hold together, but I doubt Icelandic authorities would see them as a package from the start."

Based on my research and conversations with other scholars, I do not consider the Icelandic debt jubilee to be an application of MMT. As the thesis already explored, much of the foundation behind the efficacy of the jubilee is not ignoring the deficit constraint or sacrificing central bank independence, but rather sound

[16] Mitchell, Bill (2014). "Yet another solution for the Eurozone." September 25, 2014. http://bilbo.economicoutlook.net/blog/?p=29092.

[17] Bossone, B. and S. Labini (2016) "Macroeconomics in Germany: The forgotten lesson of Hjalmar Schacht." *VOX EU CEPR*, July, 1 2016. https://voxeu.org/article/macroeconomics-germany-forgotten-lesson-hjalmar-schacht.

[18] Pilkington, Philip (2013). "Hjalmar Schacht, Mefo Bills and the Restoration of the German Economy 1933-1939." *Fixing the Economists*, December 11, 2013. https://fixingtheeconomists.wordpress.com/2013/12/11/hjalmar-schacht-mefo-bills-and-the-restoration-of-the-german-economy-1933-1939/.

economic theories on risk sharing and debt forgiveness. Due to the time and word constraint of this project, I do not spend time examining the general claims by MMT theorists. I also do not consider this thesis as a direct response to MMT. My hope is that the case of Iceland can offer insights on debt forgiveness programs on the macro level and encourage policymakers to take bold actions in response to financial crises, which are just good economics, not Modern Monetary Theory.

Honor Pledge

This thesis represents my own work in accordance with University regulations.

Tiger Gao