Comparing Labour Market Outcomes across OECD Countries using Synthetic Control Methods

Bachelor Thesis

In partial Fulfillment of the Requirements
for the Degree of

Bachelor of Science (B.Sc.) in Economics

School of Business and Economics
Humboldt-Universität zu Berlin

submitted by

Dominik Prugger
Student number: 567214

Supervisor:

Prof. Dr. Jochen Kluve
Institute of Empirical Labor Economics

Berlin, 1 September, 2017

Contents

1 Introduction

"Arguably the most important innovation in the evaluation literature in the last fifteen years is the synthetic control method...this method builds on DD estimation, but uses arguably more attractive comparisons to get causal effects"

Susan Athey and Guido Imbens [Athey and Imbens, 2017]

If Susan Athey and Guida Imbens are right, can synthetic control methods also be applied to evaluate the impact of structural labour market reforms? Which target variables provide the best fit? Evaluating the impact of reforms on aggregated outcomes such as the national unemployment rate or gross domestic product (GDP) is an important matter both to the public and to policy makers in gerneral. The public often has a legitimate interest to know whether their lives have improved or detoriated due to a specific reform that was instated by the government or other authorities. At the same time, policy makers need to know whether or not the consequences of their reforms lead to the desired outcomes, should be abandoned in the future and what their exact impact was on different areas.

In the last 20 years, there as been a substantial increase in the scientific literature about methods that try to evaluate reforms, both on the individual (micro) and the aggregate (macro) level. One of the main challenges for both cases is to construct a clearly defined counterfactual: what would have happened to the treated unit in the absence of the treatment? When evaluating small-scale reforms or policies, Randomized Control Trials (RCTs) have become increasingly popular to answer this important question. One of the many applications has been to evaluate the impact of aid programmes. [Banerjee and Duflo, 2011] In a properly conducted RCT, the counterfactual has no statistically significant deviation from the treated units and thus represents a suitable unit of comparison in absence of the intervention. However, when the target of evaluation is an outcome on a national and broader level, RCTs do not prove useful because it is usually not practicable to randomly assign treatment status, e.g. a national minimum wage, for different countries.

One of the main challenges when evaluating aggregate outcomes of, for example, a new policy that was implemented in a specific country is to find a suitable counterfactual. A good example In their paper concerning the mariel boatlift, David Card [1990] broadly introduced the concept of the "Difference in differences"(Diff-in-diff) estimator. The general idea is to measure the impact of immigration on wages and the Miami labour market in general, using a comparison between Miami, where the Mariel Boatlift happenend, and four other cities that were similar to Miami in certain variables and located in geographical proximity. The main assumption is that these cities were comparable and, in absence of a sudden inflow of immigrants, each a suitable counterfactual for Miami. Although highly applicable in this context, one shortcoming of the Diff-in-diff is that it uses just on control unit at a time for comparison, in this example only one city for a comparison of Miami. As every city or country differs in many aspects and thus can never be equal in all regards. A potential improvement is to use a weighted group of control units that approximates the treated unit better than each one could itself.

In this bachelor's thesis, I conduct three evaluations of labour market reforms in Europe that were implemented after the financial crisis of 2007-2009. The question I want to answer is whether I am able to estimate useful and robust results of each reform using synthetic control methods and what limitations arise in the process. The idea and methodology I mainly follow was first introduced in 2003 [Abadie and Gardeazabal, 2003] and further developed seven years later. [Abadie et al., 2010] The synthetic control method allows researchers to construct the counterfactual out of several aggregated units and calculate the weights of each possible control unit through a data-driven process. The reasoning being that it is often not possible to identify a single unit of comparison which best approximates the defined characteristics of the exposed unit, whereas combinations of different control units provide better fits. To provide an example: To evaluate the effect of a new set of laws aiming to reduce the sales of cigarettes in California in 1988, Abadie, Diamond and Hainmueller [Abadie et al., 2010] construct a counterfactual combining data from five U.S. states to estimate the level of cigarette sales that California would have had without the policy. Given the high quality and availability of online macro data, e.g. through databases from Eurostat/OECD/IMF etc., synthetic control methods seem promising for estimating aggregated outcomes. After its initial usage, synthetic control methods have been implemented by researchers in a variety of cases, which include the effects of market liberalization [Billmeier and Nannicini, 2013], the impact of the German reunification on GDP per capita [Abadie et al., 2015], the local impacts of nuclear power plants [Ando, 2015] and the effects of social connections in the stock market. [Acemoglu et al., 2016]

I implement the synthetic control method for three countries where large labour market reforms have taken place: I try to estimate the outcome of the introduction of a mandatory minimum wage in Germany starting in 2015. Aditionnally, I use the method to estimate effects of the structural reforms in Spain and Hungary, both in 2012, which caused far reaching changes to labour market institutions, unemployment and dismissal regulations, the degree of centralisation, as well as the collective bargaining system. [Tóth, 2012, Bentolila et al., 2012] To the best of my knowledge, this is the first attempt to estimate the impact of these reforms using synthetic control methods. The thesis is organized as follows: In Section 2 the main ideas are introduced, the mathematical model and assumptions of the synthetic control method, as well as a way to measure the significance of the results. Section 3 describes the data used to construct a suitable counterfactual for each of the three countries. Section 4 gives a brief summary of each of the different reforms and summarizes the resulting estimations for both, the unemployment rate and GDP per capita. As the results suggests that Spain has the best fit for the synthetic control methods I followingly perform inference tests to see if the results are robust. Section 5 discusses the insights gained and a possible interpretation of the results as causal effects. All calculations, plots and graphics were produced in “R” taking advantage of the package “Synth”. [Abadie et al., 2011]I include the most relevant graphs and tables inside the main body. Further informative graphics on the different calculations can be found in Appendix A.

2 Synthetic Control Method

2.1 Main Model

The evaluation of the three different labour market reforms in my thesis mainly rely on the synthetic control design that was introduced by Alberto Abadie et al. and evaluated in the first cases the impact of terrorism and the effect of an anti-tobacco legislation. [Abadie and Gardeazabal, 2003, Abadie et al., 2010] My aim is to estimate a counterfactual for three different labour market reforms in Germany, Hungary and Spain, which behaves as the treated country would in all aspects but the status of treatment. To explain briefly the methodology behind the mechanism, I firstly derive the treatment estimators and the specifications of the model[1]. In the following, I assume that only one unit is affected by the respective reform of interest.[2] Moreover, I use the words "unit" and "country" interchangeably, while "treatment" or "treated" means that this country is affected by the reform. I use the terms "control countries" and "donor countries/pool" for countries that were not affected by the reform. Let Y_{it}^N be the outcome variable for country i at the time t in the absence of a reform, for countries $i = \{1, .., J+1\}$ with $i = 1$ as the treated unit and years $t = \{1, ..., T\}$ with T_0 as the number of years before the reform was implemented; $1 \leq T_0 \leq T$. Let Y_{it}^I be the outcome that can be measured for country i at time t, if country i is exposed to the intervention in periods $T_0 + 1$ to T.

Let α_{it} be the impact of the reform for country i at time t and D_{it} an indicator, that is one if country i is exposed to the intervention at time t and zero otherwise. The outcome that can be observed for country i at time t then is

$$Y_{it} = Y_{it}^N + \alpha_{it} D_{it}$$

Only country $i = 1$ is exposed to the reform and only after T_0. Thus the aim is to estimate $(\alpha_{1T_0+1}, ..., \alpha_{1T})$. For $t > T_0$ it can be written as

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

Y_{it}^I is the observable outcome. To estimate α_{1t}, it is therefore essential to estimate the unobservable outcome Y_{1t}^N. Following the methodology from the main paper on synthetic control models [Abadie et al., 2010], Y_{1t}^N can be sufficiently described by the factor model:

$$Y_{1t}^N = \gamma_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it} \tag{1}$$

where γ_t is an unknown common factor, containing constant factor loadings across all countries observed, Z_i is a $(r \times 1)$ vector of observable covariates that are not affected by the intervention, θ_t is a $(1 \times r)$ vector of unknown parameters, λ_t a $(1 \times F)$ vector of unobserved common factors μ_i is a $(F \times 1)$ vector of unknown factor loadings, and ϵ_{it} a zero-mean error term of unobserved transitory shocks[3] at the desired aggregate level. Both r and F are elemenet of $\mathbb{N}$.

Next, consider a vector $W = (w_2, ..., w_{J+1})'$ that has the two important characteristics $w_j \geq 0$ and $w_2 + w_3 +, ..., w_{J+1} = \sum_{j=2}^{J+1} w_j = 1$ for all $j = 2, 3, ...J+1$. Multiplication with (1) results in:

$$\sum_{j=2}^{J+1} w_j Y_{it} = \gamma_t + \theta_t \sum_{j=2}^{J+1} w_j Z_i + \lambda_t \sum_{j=2}^{J+1} w_j \mu_i + \sum_{j=2}^{J+1} w_j \epsilon_{it}$$

To be more precise: The weights $w_j^* = \left(w_2^*, ..., w_{J+1}^*\right)$ need to be so, that

$$\sum_{j=2}^{J+1} w_j^* Y_{j1}^N = Y_{11}, ..., \sum_{j=2}^{J+1} w_j^* Y_{jT0} = Y_{1T0} \ and \ \sum_{j=2}^{J+1} w_j^* Z_j = Z_1 \tag{2}$$

[1]Note that the following derivations are a replication from [Abadie et al., 2010]. However, I find it useful to understand the possibilities and limitations of the synthetic control model and added additional explanations wherever I found them useful

[2]If this does not hold true, it would be possible to aggregate all regions that are affected.

[3]Transitory shocks in the sense, that the effect of the shock declines and eventually dissappears over time, e.g.

The first part means that for all pre-intervention periods, the weighted outcomes of the control units have to be equal to the pre-intervention outcomes of the treated unit. The second part states that the weighted vector of the observable covariates of the control countries has to be equal to the observed covariates of the treated unit.

Because this thesis is focused on the practical application of the model, rather than its theroetical verification, I skip the following proof and refer to the Appendix B of the main paper on the theory of synthetic control methods, [Abadie et al., 2010] where it is shown that under certain standard conditions and if the matrix $\sum_{t=1}^{T_0} \lambda_t' \lambda_t$ is non-singular, the difference between the unobserved outcome for the treated unit without treatment and the weighted control group is approximately zero. Another requirement for this is that the amount of pre-interventions is large relative to transitory shocks. So, approximately:

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{it} = 0$$

Therefore it is possible to use $\hat{\alpha}_{1t} = Y_{1t}^I - \sum_{j=2}^{J+1} w_j^* Y_{it}$ as an estimator α_{1T}.

There are some important assumptions that have to be fulfilled for $\hat{\alpha}_{1t}$ being a useful estimator: First, there should be no spill-over effects from the treated country to other countries that contribute to the counterfactual. Second, there have to be no or relatively small anticipation effects, so that the reform has no effect to the treated country prior to T_0. Both of these assumptions are in reality not always fulfilled and a discussion of their relevance for the reforms I evaluate can be found in section 5. Third, the treated unit has to be inside the convex hull of both the outcome variable and the covariats of the countries that constitute the donor pool. Or mathematically, if $(Y_{11}, ..., Y_{1T_0}, Z_1')$ is inside the convex hull of $\{(Y_{21}, ..., Y_{2T_0}, Z_2'), ..., (Y_{J+11}, ..., Y_{J+1T_0}, Z_{J+1}')\}$. To give an example: If the treated country has an unemployment rate of 4% in 2000 and 3% in 2001 in the pre-intervention period, the synthetic control method fails to reproduce this trend if the two control countries have a trend in the opposite direction (4% and 5% in 2000 and 5% and 6% in 2001 respectively) or if the control countries have consistently higher levels (6% and 5% in 2000 and 5% and 4% in 2001 respectively). The importance of this assumption for the three reforms can be calculated or graphically solved for each case and will be as well discussed in section 5.

Even if there is a synthetic control group that provides a good fit for the treated unit, the interpolation of other countries to construct a hypothetical untreated country may be biased if the linear model does not hold over the entire set of countries. Additionally, some interpolation problems can arise if the relationship between the outcome variable of the treated and the explanatory variables of the donor countries is highly nonlinear. For example, a combination consisting of a country with 95% completed secondary education and a country with 5% secondary education approximates the outcome of a state with 50% secondary education only if that outcome is to some degree linear in the percentage of education. One practical solution for these interpolation biases is to restrict the donor pool to countries with similar characteristics to the country exposed by the reform.

One main difference from the synthetic control method to the more usual Difference-in-difference estimations, which are used to a great extent in the empiricial studies about aggregated outcomes, is that the effects of confounding unobserved characteristics can vary with time. [Abadie et al., 2010]. This difference can be seen in equation (1), where λ_t can vary over time. When using a Synhtetic Control Method, taking time differences does not neutralise the unobserved μ_i. Ideally, a weighted control group would have the properties

$$\sum_{j=2}^{J+1} w_j^* Z_j = Z_1 \; and \; \sum_{j=2}^{J+1} w_j^* \mu_j = \mu_1 \tag{3}$$

Because μ_j is not observable, these properties can not be proven. Again, it is shown, [Abadie et al., 2010] that under the same standard conditions as mentioned before, the factor model (equation 1) implies that if the synthetic control method provides a good fit to Z_1 and a range of outcomes before the reform, $Y_{11}, ...Y_{1T_0}$, then it fits both Z_1 and μ_1. Thus equation (3) holds approximately.

2.2 Implementation

Next, I summarize briefly how the synthetic control method is implemented and calculated through a data-driven approach. W is a $(J \times 1)$ vector of weights, with the same properties as defined before. Each value stands for one explicit weight and importance of a donor country. For the purpose of estimating labour market reforms, I chose to exclude the possibility of the weights being either negativ or greater than one, while theoretically these can be used to gain a higher fit at the cost of extrapolation problems. Y_{1t} and Y_{jt} are the outcome variables of interest for time periods $(t = 1, ...T_0, ...T)$ and unaffected regions $(j = 2, ...J + 1)$. Then, Y_1 is a $(T_1 \times 1)$ vector of outcomes for the treated country after the intervention at T_0 and Y_0 a $(T_1 \times J)$ matrix of outcomes for the potential control countries after the reform, with $T_1 = T - T_0$ being the number of years after the intervention.

For the next part, a $(T_0 \times 1)$ vector $K = (k_1, ..., k_{T_0})'$ is defined which represents a linear combination of all the outcomes which happened before the reform: $\bar{Y}_i^K = \sum_{s=1}^{T_0} k_s Y_{is}$. For example, if $k_s = 1/T_0$ then $\bar{Y}_i^K$ is just the average of all pre-intervention outcomes. There are M $(M \in \mathbb{N})$ such combinations, given by the vectors $K_1, ..., K_M$. $X_1 = \left(Z_1', Y_1'^{K_1}, ..., Y_1'^{K_M}\right)'$ is then a $(k \times 1)$ vector which consists of the characteristics for the treated country or unit, with $k = r + M$. To remember, $Z'1$ is the $(1 \times r)$ vector of observed covariates that are not affected or changed by the reform. Doing the same for the untreated countries, X_0 is the $(k \times J)$ matrix that contains pre-intervention characteristics for those countries, where J is the number of untreated countries. The vector with the optimal weights W^* is then chosen to minimize $\|X_1 - X_0W\|$.

To find out how well the resulting weights fit the treated country, V is established as a $(k \times k)$ symmetric and positive semidefinite matrix to measure $\|X_1 - X_0W\|_V = \sqrt{(X_1 - X_0W)' V (X_1 - X_0W)}$. V thus reflects the relative importance of the different predictor variables in estimating the outcome variable. The algorithm assigns larger weights to those pre-treatment variables the higher their predictive power for the outcome variable. There are different approaches to obtain an optimal value of V. In this thesis, I focus on the most commonly used approach: V will be chosen to minimize the mean squared prediction error of the synthetic control estimator.

The derivations in this chapter is not my own work but a summary of the mathematical method that was introduced with the evaluation of the anti-tabacco legislation, [Abadie et al., 2010]. However, it is important to understand how the synthetic control works and I added explanations whenever I found them useful. In the following, I will use the word "good fit" and "bad fit" in the meaning that the synthetic control group is able to mimic the pre-intervention development of the treated country to a reasonable good extent. This means that the deviations between both is near zero in all pre-intervention years.

2.3 Inference

Like in all estimations of causal consequences, it is important to calculate the significance of the results. One main concern for synthetic control methods is that the control group is not adequate to reproduce the desired counterfactual. So called "Placebo Tests" are proposed [Abadie and Gardeazabal, 2003, Abadie et al., 2010]to gauge the significance of the estimations instead of the more usual standard erros used in regressions. To measure if the resulting treatment effects are disproportionally large or rather small relative to a random effect, the synthetic control method gets iteratively applied to every other country of the control group. The authors claim that "this inferential exercise is exact in the sense that, regardless of the number of available comparison regions, time periods, and whether the data are individual or aggregate, it is always possible to calculate the exact distribution of the estimated effect of the placebo interventions. " [Abadie et al., 2010] In other words, with these "Placebo Tests", one is able to assess the relative size of the estimated effect of a reform compared to how all other donor countries would behave if those had been assigned the same reform in the same year. The effect of the labour market reform is seen as significant if the estimated effect is large compared to those placebo effects.

Second, I compare the ratio of post/pre-treatment mean square prediction erros (MSPE) for each country. This ratio measures the magnitude of the gap in outcome between the actual and synthetic country. The ratio between post/pre is needed, because a large post-intervention MSPE only relates to a large treatment effect if the synthetic control can closely mimic the outcome in the pre-intervention period, thus producing a low pre-intervention MSPE.

Third, I reassign the treatment year to different pre-intervention periods to check whether the resulting effect is driven by a random point in time. If the effect is really driven by the evaluated reform, then one would expect no clear deviation between the synthetic control group and the treated country in the years before the reform was implemented. Forth, it is proposed [Abadie et al., 2015] to change the composition and selection of control countries and see if the effect is robust to these changes. One possibility is to vary the original composition of the control countries, for example by iteratively exclude one country out of pool of donor countries, another is to remove some or all of the original countries and select countries that may provide a better fit.

3 Selection of Data and comparable Countries

3.1 Data

To estimate the causal effect of the three labour reforms I create a cross-country panel data set from 1990 to 2016 that consists of every european country that is a member of the OECD, as well as Canada and the U.S.A. The outcome of interest is the harmonized unemployment rate and GDP per capita in US $, purchase power parity 2010. I assume that structual labour market reforms, as happened in the three countries I selected, would have a profound impact on unemployment as it is one problems most pressing in the treated countries. Second, I wanted to compare the different fits of the synhtetic control methods and therefore chose to include another target variable with GDP per capita measured in U.S. dollar and purchasing power parity (PPP) of 2010 as a proxy for productivity and economic growth. As explanatory variables I chose macro indicators that reflect the general development of the country, namely the inflation rate, share of exports of the GDP as a proxy for trade openness, logged GDP and the annual primary surplus as a percentage of GDP. Because I try to evaluate labour market reforms I additionaly use variables specific to labour market outcomes, namely the labour participation rate, the percentage of total governmental spending on social programs, an index of employment protection and the percentage of the population above 25 that attained at least secondary education.

Except the percentage of social spending, all variables are reported on an annual basis. The primary source of data is the statistic office of the OECD from where I obtained all variables used except the schooling level and the share of exports, which I obtained from the World Bank dataset. A detailed table of all variables used can be found in the Appendix B. To cope with issues of missing data, I use two-year averages of the percentage of social spending and use the level of schooling only for the years where most countries have data, namely 2007 to 2012, 2014 and 2015.

3.2 Selection of Cases

For every reform, I define a donor pool of countries, of which all shall be comparable to the treated country and hopefully resemble the same macroeconomic trends as the treated country. Second if there are any missing values for the outcome variable then these countries are dropped.

- Introduction of minimum wages in Germany 2015:

  The selection of countries for a potential counterfactual for Germany is straightforward. To estimate the effect of a sudden introduction of a mandatory minimum wage in the whole country, I remove all countries from the dataset that had a national minimum wage before or after 2015. In other words, the donor pool only consists of countries that have had no national minimum wage before or after 2015. These countries are Finnland, Norway, Sweden, Denmark, Austria, Switzerland and Italy.

- Introduction of a new Labor Code in Hungary 2012:

  To obtain a set of comparable countries that share most of the unobserved covariates like population, political and institutional frameworks or macroeconomic shocks, I limit the control group to countries of Central and Eastern Europe, namely Germany, Austria, Czech Republic, Slovenia, Slovak Republic, Poland, Greece and Estonia. None of these countries had an entire rehaul of their labour code in the time period of interest.

- Structural labour market reforms in Spain, 2011 and 2012:

  Applying the same logic to the structural labour market reforms in Spain, I restric the potential control countries first to the ones that which economy declined the most during the financial crisis and the ongoing years. The resulting donor pool consists of Italy, Portugal, Ireland, Greece, Estonia and the Slovak Republic. To confirm the results, a second estimation uses the countries that are geographically and from their population and institutional frameworks the most similiar. Then, all countries that underwent major structural labour market reforms are excluded. I identify countries without structural reforms with the help of the website of the European Trade Union Institute and the variance of the employment protection index as a proxy for structural reforms. The resulting donor pool consists all original countries except Germany, Portugal, Greece, Estonia, Slovak Republic, Slovenia, Hungary, Slovenia, Austria and the U.S. because of missing data.

4 Reforms and results

4.1 Introduction of a Minimum Wage in Germany, 2015

4.1.1 Background

Germany introduced a mandatory statuary minimum wage of 8.50 per working hour on 1 January 2015, following the coalition agreement of the Social Democratic Party (SPD) and the Christian Democratic Union (CDU/CSU). The introduction of a minimum wage was a central demand of the SPD during the election campaign und subsequently agreed upon. Prior to 1 January 2015, minimum wages were only specific to regions or industries and varied in the level of the wages. The primary goal of the introduction, as stated in the coalition agreement was "good working conditions for everybody - safe and well-paid" as stated by the Federal Ministry of Labour and Social Affairs. With a level of 8.50 per hour, the German minimum wage is one of the highest in Europe, considering the ratio between minimum wage and median wage in a variety of measurements.

There are several groups that are excempted from the minimum wage regulations: Juveniles younger than 18 years old, trainees, interns, if their internship is not longer than three months or mandatory, volunteers and longtime unemployed persons. The simultaneously instated comission that supervises the regulations, consists of three representatives of each employee and employer side as well as two scientific advisors. [Kluve, 2013] Following its decision, the hourly minimum wage was raised on 1 January 2017 to 8.84 per hour.

The impact of minimum wages on employment is one of the most throroughly researched topic in labour economics, although there is little consensus about the effect of minimum wages on employment. While some meta-analyses hint that there is either no discernible or only a small causal effect of minimum wages on employment, e.g. [Schmitt et al., 2013] or [Neumark et al., 2014] state on the other side of the spectrum that minimum wages pose a tradeoff between higher wages for the employed vs. job losses for others.

4.1.2 Effect of the minimum wage on unemployment

I run synthetic control group estimates to estimate the effect of the mandatory minimum wage on both the harmonised unemployed rate and GDP per capita. Figure 1 shows on the left hand side the trajectories of both the unemployment rate in Germany and in the resulting synthetic control from 1995 to 2016, while the right hand side shows the annual gaps between the two.

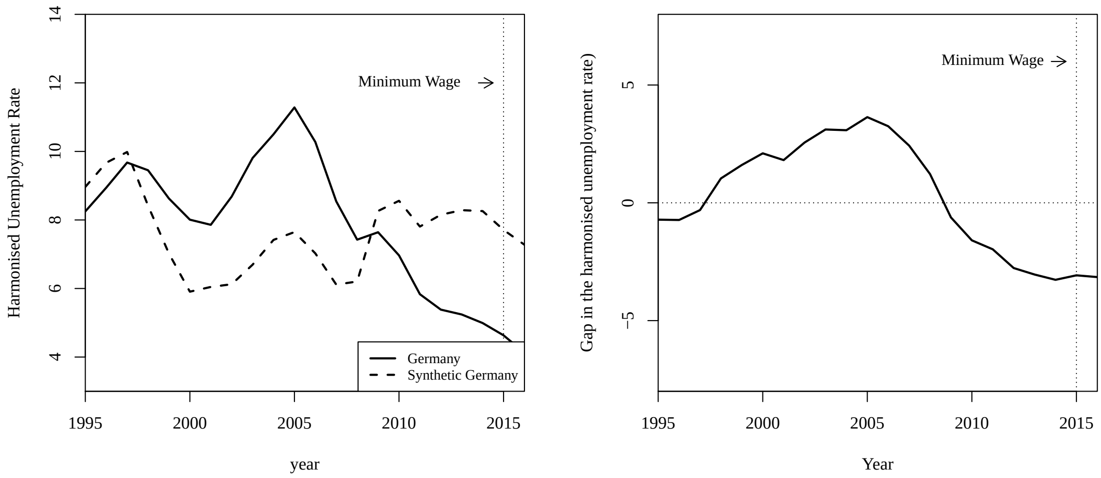


Figure 1: Pathplot and Gapsplot for Germany and its synthetic control: Effect of minimum wages on unemployment

The synthetic control consists of nearly only Sweden (93%) and Italy (7%), with the bi-annual schooling level and the primary surplus being the two most predictive variables. However, the synthetic control is not able to mimic the pre-reform unemployment rate of Germany. Neither the spike between 2003 - 2006, that was partly due to a change in measurement, nor the decline in the unemployment rate after 2010 is reproduced by the synthetic Germany. This implies that one or several of the assumptions discussed in section 2.1 do not hold. One potential reason is that the outcome of interest in the period 1995-2016 is not inside the convex hull of observation of the control group. Not a single country from the donor pool shares the same spike in unemployment between 2003 and 2007. Then, following the financial crisis after 2010, the unemployment rate decreases in Germany, which is ,in this magnitude, not shared by any other control country (See Figure 11 in Appendix A.1 for the plotted unemployment rates). The labour market of Germany is simply too unique to be evaluated via synthetic control methods. Figure 1 suggests that a potential reform took place around 2009, the time when the two graphs diverge. However, as both graphs diverge considerably before 2009, this can not be seen as a potential causal effect of anything. When looked again at the impact of the minimum wage in 2015 that I tried to estimate, the synthetic control method does not provide any useful estimations.

4.1.3 Effect of the minimum wage on GDP per capita

Next, I estimate the effect of the introduction of minimum wages on GDP per capita. The resulting synthetic control is a linear combination of Austria (58 percent), Italy (23 percent) and Sweden (19 percent) while the variables of logged annual GDP, labour participation rate and the percentage of local schooling had the highest predictive power.

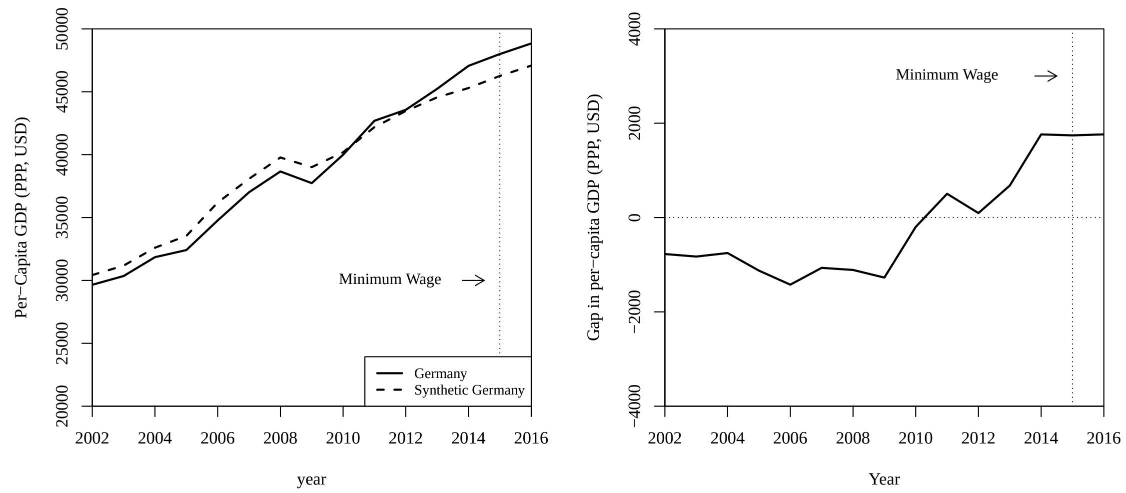


Figure 2: Pathplot and Gapsplot for Germany and its synthetic control: Effect of minimum wages on GDP per capita

Figure 2 shows that the synthetic control is able to mimic the pre-reform level of GDP per capita in Germany to some extend. The level of the synthetic control lies underneath the actual level before 2012, reaching a maximum deviation of -1423$ per capita in 2006. Between 2012 and 2015 it then rises above the actual level and rises until a deviation of 1762.4 $ per capita in 2014, which stays constant for the following two years. The graph on the right hand side shows again the annual deviations between Germany and its synthetic control. If all assumptions of the synthetic control method hold, there would be close to no deviation at all before the year of the introduction of the minimum wage, 2015. However, it seems as a potential reform happend in 2011/2012 instead, as the gaps before these years was close to zero and deviated afterwards. Once again, I think that the most plausible explanation for this would be the exceptionally well-perfoming German labour market. Another possiblity could be that the linear combination of the three countries is simply not able to reproduce the real trend in a way that there is a significant effect of the introduction of minimum wages. This can be seen by the steady positive gap of around 1750$ between 2014 and 2016.

As the fit for both cases, the unemployment rate and GDP per capita is rather poor, and the resulting effects of the introduction of minimum wages thus not plausible, I do not bother with explicit interference tests for these results.

4.2 The new Hungarian Labour Code, 2012

4.2.1 Background

The new labour code, labeld as the "2012 I. Act", came into effect on July 1 2012. The old labour code set a combination of strict minimum labour standards and flexible regulations on the other hand. The main goals of the new code was to increase flexibility for employers and to cut the power of unions in the collective wage bargaining system.

It states that wage bargaining should take place between sectoral unions and workplace-level unions. One of its main topics was to increase flexibility and allow for individual or collective agreements that can deviate from what is stipulated by the law. While the old code explicitly stated that these deviations could only happen to the benefit of the employee, 2012 I. Act allowed deviations for the benefit of the employer as well. [Tóth, 2012] Extending on this idea, the new labour code extensivly shifted risks from the employer to the employee, e.g.: an employee is not longer entitled to salary, in case of an unavoidable external force, such like a power cut due to extreme weather conditions. Additionally the code tightened termination and recruitment rules for the employee in several ways, e.g. increased the period of probation, allowed employers to dismiss workers during their sick leave, decreased the lump sum paid to dismissed employees or terminated the necessity of re-employ unlawfully dismissed workes. The new reform added flexibility to working times through expanding the amount of possible overtime hours and allowing collective agreements to overrule sectoral or regional agreements.

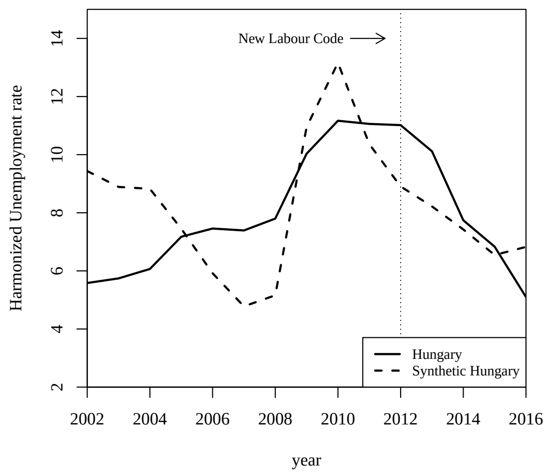
On the other hand, the Labour Code 2012 drastically cut the rights and entitlements of unions, especially of union activists. It vastly reduced the number of officials of each firm that were granted legal protection, cut several sources of funding for unions and stripped union activists of working time exemptions. In addition it ended the traditional legal right of unions to be informed in several ways by the employer and cut their influence on the electoral commitee.[4]

As the panel data contains annual or bi-annual variables, the first challenge is to set the starting year of the new labour act, as it was set in motion by the parliament in the middle of the year 2012. To account for anticipation effects, I select 2012 as the starting year for the synthetic control method. I choose 2002 as the starting year for the pre-reform time, to allow for a pre-reform period that is long enough to see whether there are considerable deviations.

4.2.2 Effect of the Labour Code 2012 on the unemployment rate

The resulting synthetic control of Hungary is a convex combination of four countries: Estonia (66%), Austria (20%), Slovenia (13 %) and the Slovak Republic (1 %). For all other control countries the resulting optimal weight is zero. (See figures in A.3 for the original tables). Figure 3 displays the trajectories for the unemployment rate of Hungary and the synthetic control group between 2002 and 2016. As shown, the convex combination of control countries is not able to closely reproduce the unemployment rate of Hungary in the whole pre-reform timeframe. In 2002, the trajectories start with a difference of around three percentage points. Afterwards the synthetic control shows clear deviations in both directions, the extremes being 2007 and 2010. After the introduction of the new labour code in 2012, the unemployment rate of the constructed counterfactual is first lower than Hungary's, then approximates Hungary's unemployment rate in the years 2014 and 2015 and consequently exceeds it. This would imply a negative effect of the refom on unemployment at first and a positive effect after the year 2015, but the poor fit in the pre-reform years prohibits a causal statement and undermines the validity of the results.

[4] For the fulltext version of the new labour code, see: http://www.ilo.org/dyn/travail/docs/2557/Labour%20Code.pdf



	Treated	Synthetic	Sample Mean
schooling	139	139.2	131.3
emp_protect	2	2.3	2.5
particip	54.5	63.4	58.6
bipacc	−4.5	−4.5	−3
exports	72.8	63.5	51.9
infl	5.2	3.6	2.9

Figure 3: GDP per capita: Synthetic control and Germany

Figure 3 shows the different trajectories and as well compares the predictor variables of Hungary with the respective variables for the synthetic control group. In the table on the right, it can be seen that the two variables that received the highest predictive power (see Figures in A.3) to predict Hungary's unemployment rate, "schooling" and "bipacc" are very close in comparison.[5] This is anticipated as the synthetic control method gives the most importance in similarity to the variables with the highest predictive power. So far as I understand the outcome, the reason for the bad fit lies partly in the fact that these two variables were given nearly all of the predictive power and that a linear combination of countries that had the same level of schooling and primary surplus produced a very different trajectory in the donor countries as it did in Hungary. This points to the potential problem that the two variables do not fulfill the assumption of a linear relationship towards GDP per capita. As in the case for Germany, it seems that the unemployment rate of a country is too specific and unique to the characteristics of the countries, as that it would be possible to be predicted and reconstructed with a synthetic control method.

4.3 Effect of the Labour Code 2012 on GDP per capita

As in the case for Germany, I proceed with an inspection of the effect on GDP per capita. As previously, the timeframe is set from 2002 - 2016 with the reform (2012 I. Act) set to begin in 2012. The resulting synthetic control groups consists of Poland (77%), Greece (20%) and Estonia (2%). Thus, this synthetic control deviates considerably in its compositio from the one in the previous section. Unfortunately, the synthetic control displays considerable deviations from the real Hungary. (See Figures A.4) For example, the averaged unemployment rate of the synthetic control group is 12.8% instead of 8.2% and the average logged GDP 3.2% instead of 1.6%. This stems again from the fact that the algorithm set most of the predictive power to two variables, the level of employment protection and labour participation and combines the countries accordingly. Thus, these two variables are closely matched (2.3 instead of 2 and 54.7% instead of 54.5%).

[5]Note that these values are averaged over the whole timespan.

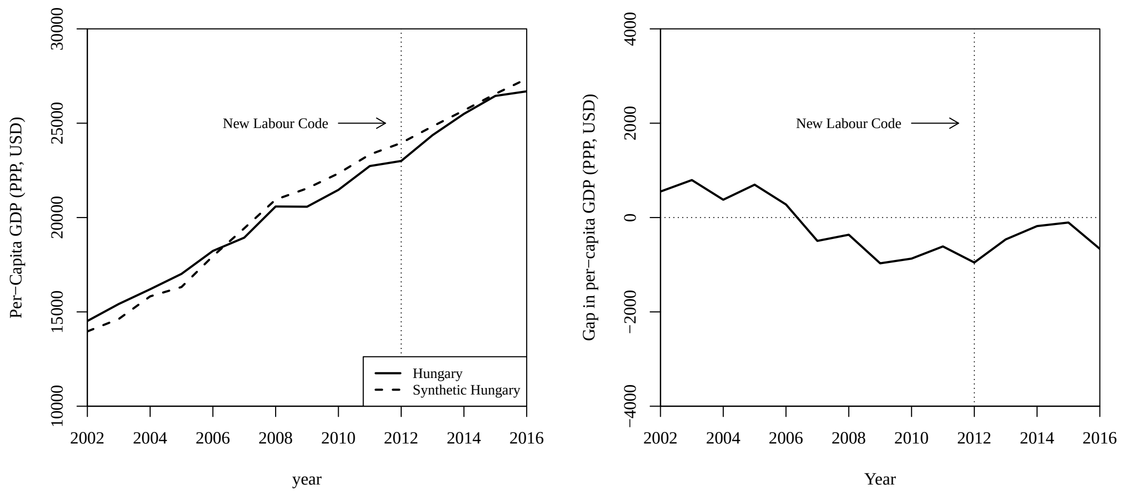


Figure 4: Pathplot and Gapsplot for Hungary and its synthetic control group

Figure 4 shows the trajectories of GDP per capita from 2002 until 2016 for both Hungary and the resulting synthetic control group. The synthetic Hungary can approximate the levels of GDP per capita to some extend, but not perfectly. From 2002-2006 it is below Hungary's original level, while it exceeds Hungary's trajectory from 2006 - 2012. Surprisingly, both lines show less deviation after the new labour code than before it.If all assumptions hold, this would be interpreted as that the new Labor Coude in Hungary had no effect on GDP per capita at all. As their is no perfect fit for the synhtetic control model and nearly all the predictive weight of the variables was set to labour participation (69%), it is questionable to what extend this result can be interpreted as a nonexisting causal cause, given that it changed primarily the rules of dismissals, flexibility of employers and union laws. This will be discussed further in Section 5. As the synthetic control model does not imply any effect of the new labour code on GDP per capita and a bad fit for the unemployment rate, there is no need to apply inference or placebo tests.

4.4 Structural Labour Reforms in Spain 2012

4.4.1 Background

Spain is considered as one of the European countries that experienced the worst consequences from the financial crisis 2008-2010. [Karanikolos et al., 2013] In the years following the crisis, the unemployment rate skyrocketed from 8.2 % (2007) to 24.8% (2012), the second highest between all OECD countries, only exceeded by Greece. In real terms, this meant that more than four million people became unemployed during and after the crisis. Simultaneously, the GDP per capita fell from 32577.61 $ (2007) to 31991.01 $ (2012). Interestingly, the cost per unit labour in the business sector remained or slightly increased during the same time, despite the huge increase in demand. According to the OECD, European Central Bank and especially the IMF the rigid and indynamic labour market played a crucial role in these devolepments and the slow recovery. As the pressure for reforms increased, the leftwing government 2010 passed a first labour reform that, among others, eased the implementation of temporary contracts and generalised the severance pay subsidy scheme.

Two years later, the Spanish Parliament, that had then a conservative majority, passed a second reform, the "Ley 3/2012 de medidas urgentes para la reforma del mercado laboral" (Ley 3/2012) in July 2012, that implemented far-reaching structural changes to the wage bargaining system, the structure of payments to unemployed, laws of dismissals and employment regulations. The overall aim was to give firms a higher degree of flexibility and allowing them to modify the amount of employees and the level of wages according the their economic performance. [Bentolila et al., 2012] Some key components shall be highlighted:

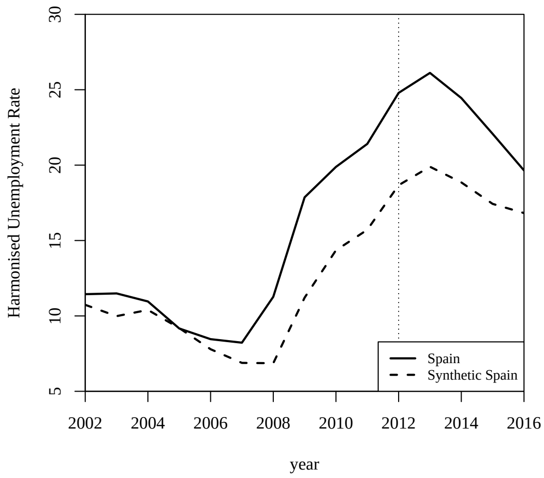
- Changes to the system of collective bargaining: Priority is given to firm-level agreements over national or sectoral agreements; collective agreements can be prolonged for a maximum of one year after their end date.
- Dismissal legislation/Collective redundancy: Dismissals become justified if the firm faces a persistent decline or in revenue or income for over three consecutive quarters and collective dismissals no longer require administrative approval beforehand.
- Monetary compensation: Reduction or removal of procedural or interim wages. Replacement rate and unemployment
- "Contrato de Apoyo a Emprendedores": A new full-time permanent contract for smaller employers that beside others allows a extended trial period of twelve months.

These changes reflect some of the adjustments, while the reform included many others on a smaller and more specific scale. As in the case of Hungary, I define the starting date of the reform as January 1, 2012, because the synthetic control method works with annual data and to account for anticipation effects. I have to exclude the annual primary surplus from the pool of predictive variables because Ireland has missing data for all respective time periods.

4.4.2 Effect of the Reforms on Unemployment

The resulting synthetic control group which could be used as a counterfactual for spanish unemployment is constructed with only two out of the potential six donor countries, namely Greece (60%) and Estonia (40%). As these two countries are located in South-East Europe and East-Europe respectively and geographically far away from Spain, it is not the best start to mimic real Spain.

The highest predictive power of the variables to forecast the unemployment rate is given to the labour participation rate (36%), employment protection (27%), share of exports (10%) and the share of social spending (7%). Accordingly, Figure 5 shows on the right hand side that the average of these variables is very similiar except the share of exports, which is 26% in Spain and 40% in the synthetic control group. In fact, all variables besides exports show little deviation which indicates a good fit at first glance.



	Treated	Synthetic	Sample Mean
infl	2.7	3.5	3
loggdp	1.3	1.4	1.7
emp_protect	2.3	2.5	2.6
exports	26	40.4	51.6
particip	58.6	57.9	58.3
special.schooling.2007.2015	156.6	165.9	166.8
special.socsp.1995.2013	20.5	20	19.3

Figure 5: Unemployment rate: Synthetic control and Spain

Contradictingly, the left hand side of Figure 5 shows that the synthetic control group is able to mimic the unemployment rate of Spain accordingly from 2002 to 2006 but deviates severly after. Following 2007, the graph of Spain rises over the synthetic control group and displays differences between the two graphs from 2007 until 2012 of roughly 4.4 to 6.7 percent. After 2012 the difference reaches a maximum of 6.2 percent in 2013 and then falls until 2.8 percent in 2016. Thus the graph implies that a reform other than Ley 3/2012 happened around 2007, which led to the following divergence. The most likely condentant for this is the financial crisis. The probable reason for the synthetic method to fail, is that the spanish unemployment rate had one of the highest spikes and consistent high levels of all Europe. In the years 2007 - 2012 it lies outside the convex hull of observations, while in the years 2012-2016 it does lie inside the convex hull of observations, but is only topped by Greece and thus hard to replicate by a linear combination of donor countries. Thus, the predictor variables are not able to explain the high spike during the financial crisis, nor is the combination of Greece and Estonia.

4.4.3 Effect of the Reforms on GDP per capita

The last contender for a suitable synthetic control group is the Spanish GDP per capita. All other cases were not appropriate, either for a lack of fit or a small/not existing difference. In the case of Germany and Hungary the evaluation of the impact of a reform yielded better results for the GDP per capita trajectory than for the unemployment rate, and thus I was interested if that development would hold for Spain too.

The resulting synthetic control group is the first case, where all potential donor countries are assigned a weight greather than zero. The country with the heighest weight is again Greece (39%), followed by Ireland (24%), Slovak Republic (14%), Italy (15%) and Estonia and Portugal with 3% each. In comparison to the unemployment rate, this combination seems at first glance more likely to serve as a counterfactual for Spain because of the inclusion of Italy and Portugal, two countries of southern Europe. Interestingly, the predictor variables rely entirely on logged GDP (52 %) and the employment protection index (47 %). To some degree, this seems logical, as the reform was created in the spirit of decreasing the impact of the crisis and easing strict employment regulations.

	Treated	Synthetic	Sample Mean
Inflation	2.7	3	3
Logged GDP	1.3	1.3	1.7
Employment Protection	2.3	2.3	2.6
Share of Exports	26	49.1	51.6
Labour Participation	58.6	56.1	58.3
Schooling	20.2	20	18.8
Social Spending	156.6	163.9	166.8

Figure 6: Comparison: Spain and its synthetic control group (GDP per capita)

Figure 6 shows the averaged properties of Spain and its synthetic control group. As the most importance is given to the logged GDP and the employment protection index, these two variables are completly equal. Similarly, the other variables show just little deviation, with the exception of the share of exports, as before. (Spain: 26%, synthetic control group: 49.1%). As the sample mean is 51.6% the most possible explanation again is that the share of exports of Spain lies outside the convex hull or is smaller than many of its donor countries. Looking at the export share over time (Figure 17, Appendix A), the second point is valid. In all years Spain has one of the three or two lowest shares of exports.

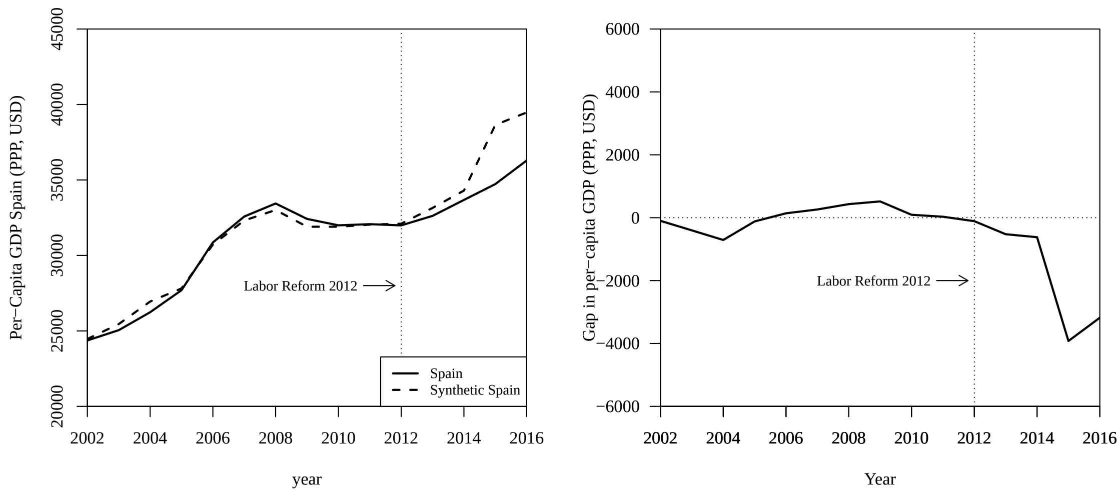


Figure 7: Pathplot and Gapsplot for Spain and its synthetic control group

The good fit that the predictive variables and the weights imply at first glance is greatly supported by the trajectories of GDP per Capita. The highest deviation between the two trajectories is 708 $ per capita in 2004. Compared to all other cases, this clearly provides a good fit. Other studies see deviations in this scale as a good fit as well. [Abadie et al., 2015] As can be seen in Figure 7, the graphs of Spain and its synthetic counterpart overlap nearly perfectly until the start of Ley 3/2012. After the reform, the trajectory of the synthetic control group becomes greater than Spain's and the gap steadily grows until 2015 where Spain's GDP per capita is 3912 $ lower than its synthetic counterpart. In 2016 it falls to 3181 $ per capita. The question, if the negative impact of the structural reforms is temporarily and becomes positive in the long run or if it persists for the years to come can only be answered at a more advanced time. This estimation contradicts to some extend a hypothetical estimation done by OECD [Bentolila et al., 2012] in 2012 that estimated that in the long term the 2012 reform would imply a faster economic growth of 15 percentage points than without the reform. On the other hand, it is possible that the lowered employment protection leads to more dismissals in the short run, as more firms can get rid of employees in the aftermath of the financial crisis, leading to a temporal decrease in per capita GDP. As the economy recovers, following the years after 2016 the reform then could lead to a higher economic growth as the labour productivity increases and the firms are confident to employ more, due to the increased flexibility and reduced labour costs. Note that the estimation by the OECD was done directly after the reform and can only be regarded as rough lower-bound estimate.

4.4.4 Inference Tests

As the synthetic control method for the evaluation of the structural reforms in Spain 2012 and their impact on GDP per capita yield the best fit out of all cases examined, I conduct inference tests to gauge the significance and credibility, following the idea of Placebo Test from Abadie et al. [Abadie and Gardeazabal, 2003, Abadie et al., 2010]

First, I conduct a so called "In-time Placebo Studie", reassigning the year of the reform to ten years before 2012. I run the same synthetic control method as above, only changing the starting year to 2002.[6]As there were no reforms to the extent of the Ley 3/2012 I expect no deviation in the trajectory of GDP per capita after 2002.

[6]Due to missing data, I had to exclude Estonia out of the donor pool. As the assigned weight for Estonia for the actual reform in 2012 is only 3%, this seems not to hurt the significance of the result

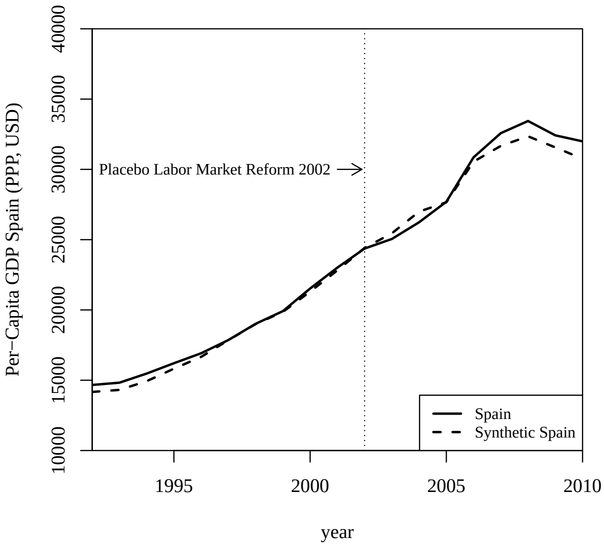


Figure 8: Inference: Reassigning the year of treatment to 2002

Figure 8 shows, that there is nearly no deviation between Spain and the synthetic control group before the placebo reform in 2002. More interestingly and important to the significance of the negative estimations derived above there is no very little deviation after 2002 too. In contrast, the actual reform, happening in 2012, the placebo structural reform has no perceivable effect on GDP per capita. This suggests, that the gap estimated in Figure 7 is not due to a random time effect but to the actual reform in 2012.[7]

Another way to calculate the significance of the estimated negative effects of the reform on GDP per capita in Spain is to compare the ratio of post/pre-treatment mean square prediction erros (MSPE) for each country, as described in 2.3. Figure 21 in Appendix A.6 shows that Spain has the highest ratio of 5.7, which means that its post-treatment MSPE is about 5.7 times the pre-treatment MSPE. The closest contender in this category is Greece, with a ratio of around 2.4, less than half the value of Spain.

Next, I run placebo tests, reassigning the structural labour market reforms of Spain 2012 to each of the countries in the donor pool respectively, and compare the results of these estimations to the negative impacts derived in 4.4.3. If the estimated effect of the structural reforms on Spain is unusually large relative to the distribution of the placebo effects, it is an indication of significance. In other words, this answers the question if the effect would be the same when I assign the treatment status at random. I iteratively apply the same synthetic control method with the same weights for the predictor variables on one of the six countries in the donor pool, while leaving Spain in the donor pool. I then display a distribution of the simulated gaps for these countries.

[7]I ran the same in-time placebo tests for the years 1999, 2004 and 2007. All three showed the same similarities as described for the year 2002.

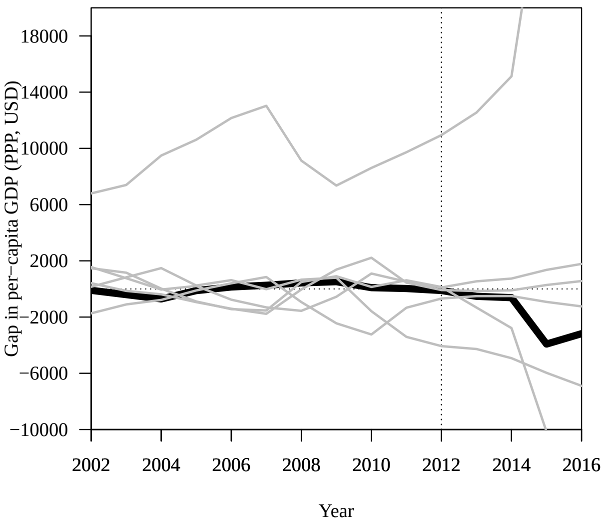


Figure 9: Placebo Studies: Black line: Spain; grey lines: Control countries

Figure 9 shows the estimated gaps for Spain (thick black line) and the placebo-gaps for the donor countries (grey lines). While the effect for Spain is larger than for the three donor countries, it is dominated in value by three donor countries, namely Portugal,Italy and Greece. This may implie a limit to the significance of the results. On the other hand, I think it points to a problem that comes with the selection of the donor pool countries. The three countries, whose placebo gap is larger than Spain, have each undergone major structural reforms themselves. The synthetic control method therefore does not imply a valid counterfactual that simulates how Spain would habe behaved if it did not implement a structural labour market reform in 2012. I therefore run a second synthetic control method for the structural reforms in Spain, this time using the general sample and excluding countries that have undergone large structural labour market reforms.

4.4.5 Change in the selection of control countries

Out of my sample of OECD countries, I identify Belgium, Canada, Denmark, Finland, France, Luxembourg, Netherlands, Norway, Poland, Sweden, Switzerland, Turkey, United Kingdom and the United States as countries which have not undergone large structural labour market reforms. I then have to exclude the United States because of missing data for the level of schooling. Next, I re-run the synthetic control method for the spanish GDP per capita with all the aforementioned ramifications concering variables and timeframes.

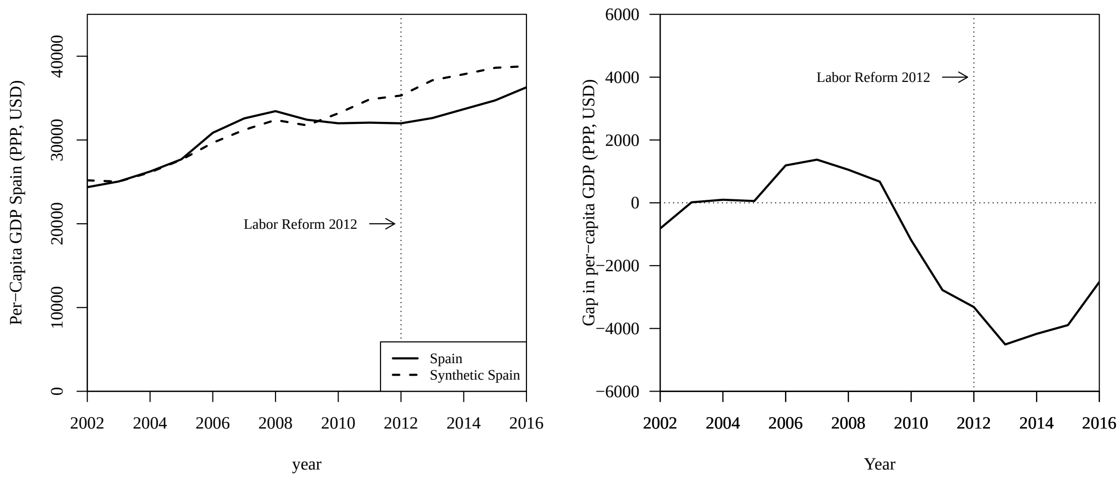


Figure 10: Pathplot and Gapsplot for Spain and its synthetic control group

The resulting synthetic control group consists of only five countries. The main part consists of France (73%) and Turkey (20%) and small percentages of Canada (4 %) and Norway (3 %). The trajectories itself do not posess the same good fit as with the donor pool of countries consisting only of the "crisis countries", as the lines for GDP per capita diverge considerably after 2010. However, this was anticipated as this is the aftermath of the financial crisis in which Spain lies on the extremes of the convex hull of observations.[8] Nevertheless, as an inference measure, this control group, which consists of not one of the donor countries in 4.4.3 points to the same direction of the estimation. Following the Implementation of the reform, the GDP per capita in Spain is lower compared to its synthetic control counterpart. Then, in 2015 and 2016 the gap becomes smaller, which could lead to a longterm higher level of GDP per capita in the future. However, this is speculation from my part and not supported by the data.

5 Discussion

The first result of my evaluation of the three structural labour market reforms in Germany, Hungary and Spain is that the assessment of GDP per capita consistingly leads to better fits than the evaluation of the unemployment rate. The unemployment rate seems to be too country-specific as that a linear combination of other countries is able to mimic its trajectory and different levels over time. A good example for this is the spike in unemployment in Germany between 2002 and 2007, partly grounded in a change of measurement and the following decline during and after the financial crisis that was partly due to new regulations implemented through the Hartz-Reforms and other unique features of the German labour market. [Burda and Hunt, 2011] This limitation can be observed for the two other countries as well and remains even when changing the selection of variables or control countries. The unemployment rate appears to be difficult to reconstruct with synthetic control methods

[8] The trade-off between good fit and endogeneity problems stems from the fact that the countries that share the same negative impact of the financial crisis as Spain nearly all implemented structural labour market reforms, partly under pressure from the OECD/ECB/IMF and to cope with the high unemployment rates during that time.

Using synthetic control methods to evaluate GDP per capita levels leads to considerably better fits, especially in the case of Hungary and Spain. The evaluation of GDP per capita in Hungary exhibits no change at all following the new Labour Code 2012, while in the case of Spain, the result points to a negative short-term impact with a maximum of -3921 $ PPP three years after the reform. To what extent can this be interpreted as a causal effect? In the following, I will focus on the case of Spain, as it provided the best fit for the synthetic control method. The estimations for Hungary are highly interesting as they suggest that there is no substantial effect of the new labour code on GDP. However, most inference tests that are suitable for synthetic control methods require an effect greater than zero. The inference tests that I applied to the results for Spain point in two different directions: The "In-time Placebo" supports the hypothesis that the effect is not driven by temporal effects and the higher MSPE ratio points for Spain than for any other country indicates that the synthetic control methods provides a good fit for Spain. Other impact studies that use the synthetic control method show similar or higher ratios for the pre/post ratio of the treated country and their next-best control unit, although the ratio itself is sometimes a lot larger.[9] The placebo studies used to check if the effect of the Ley 3/2012 is driven by a random assignment to other control countries do not necessarily yield a good fit. This is probably due to a trade-off between a good pre-intervention fit and a control group of appropriate countries. Most of the countries that are similar to the economy of Spain and exhibit the same economic performance during the financial crisis implemented their own structural labour market reforms. Using a modified selection of donor countries provides a lower fit of the synthetic control method through a higher deviation between Spain and its synthetic control, but shares the negative impact on GDP per capita for Spain with a maximum of around -4500 $ PPP per capita. This can be interpreted as an indication of a negative short-time impact of the reform, although the estimation of its magnitude should be treated with caution and has to be embedded in other and future empirical research. This interpretation is in line with past findings that the short-term effect of structural labour market reforms may be negative if the treated country is in a current depression [Bouis et al., 2012] which was the case for Spain in the years directly before and after 2012. [Isabel et al., 2017]

Further, one may question if the assumptions to synthetic control methods discussed in 2.1 hold in the case of Spain. As the reforms were implemented simulteanously in the whole country, there should be no spillover effects inside Spain. However, it is possible that the reforms also affected neighbouring countries. Firms may have hired more foreign workers and possible unregistered employees because of the more flexible employment regulations. This would overestimate the effect for the national GDP per capita, as the OECD defines "capita" as "registered head of population". While this could be possible, the percentage of foreign workers in Spain dropped after 2012 from 54.5% to 53.3%. [OECD, 2017a] It rose again in 2014, but this occured simultaneously with a global increase in migration, which does not support the hypothesis that the Ley 3/2012 caused a higher immigration towards Spain.

Other possible factors that might flaw the size of the result is anticipation effects and labour market reforms that were implemented before and after the year of the reform evaluated. Anticipation effects may materialize if firms restrain their hirings and dismissals before the date of implementation because they await the better conditions after the labour market reform. In the case of the Ley 3/2012, I think that most anticipation effects are accounted for by choosing 2012 as the starting year for the synthetic control method. One would expect the highest anticipation

[9] See for example [Abadie et al., 2010], where the post/pre ratio for California is about 1.5 times higher than of the next-best state. The ratio itself is around 130

effects in the period between the governmental approval of the reform and the day that the law was passed by the Spanish Parliament. In the case of Spain, this timeframe lies between February and July 2012. As my synthetic control method uses variables that are reported on an annual basis, I define the starting date as the first of January 2012 and therefore account for the entire timeframe. Anticipation effects that occured before the government passed the Ley 3/2012 are possible but probably small enough in size to have no substantial effect on the resulting change in GDP per capita, as some studies about the short term effect of labour market policies suggest. [Agell, 1999]

The other factor that poses a possible bias is the variety of different labour market reforms in the case of Spain. One year before the implementation of Ley 3/2012, another structural labour market reform was passed, the regulatory changes of which were not as far-reaching as the one in 2012 but could attenuate its effect. How can one be certain that the effect of the Ley 3/2012 was not partially driven by the previous reform? To account for the previous labour market reform, I ran another model in which the treatment date was reassigned to 2011. The resulting synthetic control shows no clear deviation following 2011 up until 2014. (See Figure 23 in Appendix A.6)

However, the fact that many additional labour market reforms were implemented in 2012 and the years to come points to a larger problem when implementing synthetic control methods to labour market reforms: From my understanding, synthetic control methods should best be implemented for the evaluation of a policy or historical event that is unique, is not reversed inside the observed timespan and mostly persists in its original form over the next years. A good example for such an event is the effect of the German unification on the GDP per capita of Germany. [Abadie et al., 2015] This can be regarded as a sudden, far-reaching policy that was not changed substantially at a later point in time. However, policies regarding the labour market, its participants and instituions are a topic of an ever ongoing debate and changes. The Ley 3/2012, for example, was followed by several modifications in 2013 and 2014 (for more detail see [OECD, 2013]). Additionally, the government passed several other policies in 2012 that tried to combat the negative consequences of the financial crisis and the poor performance of the labour market and which likely also impacted the GDP per capita.[10]Therefore, one can argue that labour market reforms are not suited for an evaluation with synthetic control methods due to a lack of uniqueness and immutability. On the other hand, the synthetic control method supposedly captures all labour market and crisis reforms that were passed in 2012 and that were not applied in this form in the control countries. At this stage, I am not able to quantify to what extend this problem undermines my estimated negative effect of the Spanish reform on GPD per capita.

On a broader note, the question remains wether synthetic control estimates can be seen as a true causal effect. It can be argued that a linear combination of, for example, France, Turkey, Canada and Norway is not comparable to Spain, due to differences in population, geographical specifications and political or institutional frameworks. However, it can be considered complementary to the common Diff-in-Diff estimations where in most cases only one neighbouring state or country is used. A potential approach to overcome interpolation biases that arise while combinating countries would be to implement the synthetic control method on a smaller geographic scale. For example, if one wants to evaluate regional policies, such as changes in the federal education system in Germany, using other German states as potential control units provide supposedly better fits than evaluations on a national level.

[10]For example the "Second Employment and Collective Bargaining Agreement 2012-2014" or the "Fifth Agreement on Autonomous Solutions to Labour Conflicts" in 2012.

6 Conclusion

To see wether synthetic control methods can be applied to structural labour market reforms, I conducted six evaluations of three different reforms in Germany, Hungary and Spain. The resulting synthetic control shows little or no deviation in the pre-intervention period in only two cases: the "New Labor Code 2012" in Hungary and the "Ley 3/2012" in Spain and their respective impact on GDP per capita. All other reforms did either not fulfill underlying assumptions of the synthetic control method or yielded a bad pre-intervention fit as the countries and their respective synthetic controls deviated severly. In the case of Hungary, estimated effects suggest that the reform did not have any effect on GDP per capita, while the Spanish case suggests that the introduction of the Ley 3/2012 led to a drop in GDP per capita during the subsequent four year with a maximum of -3921$ PPP in 2015. In 2016 the deviation started to shrink. However, given the date of my evaluation, there is not enough evidence to say that the reform will have a positive impact in the longer term. These findings can be of great importance to Hungarian and Spanish policy makers, as the reforms were intended to ease the severe economic problems at that time but apparently do not turn out to be helpful in the short run. To check whether or not effect of the Ley 3/2012 is robust, I conduct several inference tests that varied in their quality of robustness. It seems that there are probably better cases for the application of synthetic control methods than labour market policies as the institutions and specifications of labour markets vary greatly between countries. Additionally, labour market policies are constantly monitored and revised and new policies are implemented on a continous basis, thus implying a lack in immutability that synthetic control methods require. As evident by the example of Spain , it is sometimes hard to select an appropriate pool of control countries as the countries that share the economic performance of Spain all implemented similar reforms for exactly this reason.

Synthetic control methods provide new possibilities for future empirical research. Promising topics include, in my opinion, the evaluation of more specific labour market outcomes, such as schooling or job satisfaction, the application on a smaller geographic scale such as changes in federal education policies and a meta-analylis which cases provide the best fits for synthetic control methods and where to apply them best.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. *Journal of the American statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2011). Synth: An r package for synthetic control methods in comparative case studies.

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

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

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the united states. *Journal of Financial Economics*, 121(2):368–391.

Agell, J. (1999). On the benefits from rigid labour markets: norms, market failures, and social insurance. *The Economic Journal*, 109(453):143–164.

Ando, M. (2015). Dreams of urbanization: Quantitative case studies on the local impacts of nuclear power facilities using the synthetic control method. *Journal of Urban Economics*, 85:68–85.

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

Banerjee, A. V. and Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.

Bentolila, S., Dolado, J. J., and Jimeno, J. F. (2012). Reforming an insider-outsider labor market: the spanish experience. *IZA Journal of European Labor Studies*, 1(1):4.

Billmeier, A. and Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, 95(3):983–1001.

Bouis, R., Causa, O., Demmou, L., Duval, R., and Zdzienicka, A. (2012). The short-term effects of structural reforms.

Burda, M. C. and Hunt, J. (2011). What explains the german labor market miracle in the great recession?

Isabel, R.-P., Miguel, R.-B., Antonio, R.-G., and Oscar, M.-G. (2017). Economic crisis and suicides in spain. socio-demographic and regional variability. *The European Journal of Health Economics*, 18(3):313–320.

Karanikolos, M., Mladovsky, P., Cylus, J., Thomson, S., Basu, S., Stuckler, D., Mackenbach, J. P., and McKee, M. (2013). Financial crisis, austerity, and health in europe. *The Lancet*, 381(9874):1323–1331.

Kluve, J. (2013). So hoch wie möglich, so niedrig wie nötig: Was ist der optimale mindestlohn? Technical report, RWI Positionen.

Neumark, D., Salas, J. I., and Wascher, W. (2014). Revisiting the minimum wageemployment debate: Throwing out the baby with the bathwater? *ILR Review*, 67(3_suppl):608–648.

OECD (2013). The 2012 labour market reform in spain: a preliminary assessment.

OECD (2017a). International migration outlook 2017.

OECD (2017b). Oecd employment outlook 2017.

Schmitt, J. et al. (2013). Why does the minimum wage have no discernible effect on employment? *Center for Economic and Policy Research*, 22:1–28.

Tóth, A. (2012). The new hungarian labour code–background, conflicts, compromises. Technical report, Working Paper, Friedrich Ebert Foundation Budapest, June.

A Additional graphs for the different synthetic control groups

A.1 Germany, Unemployment

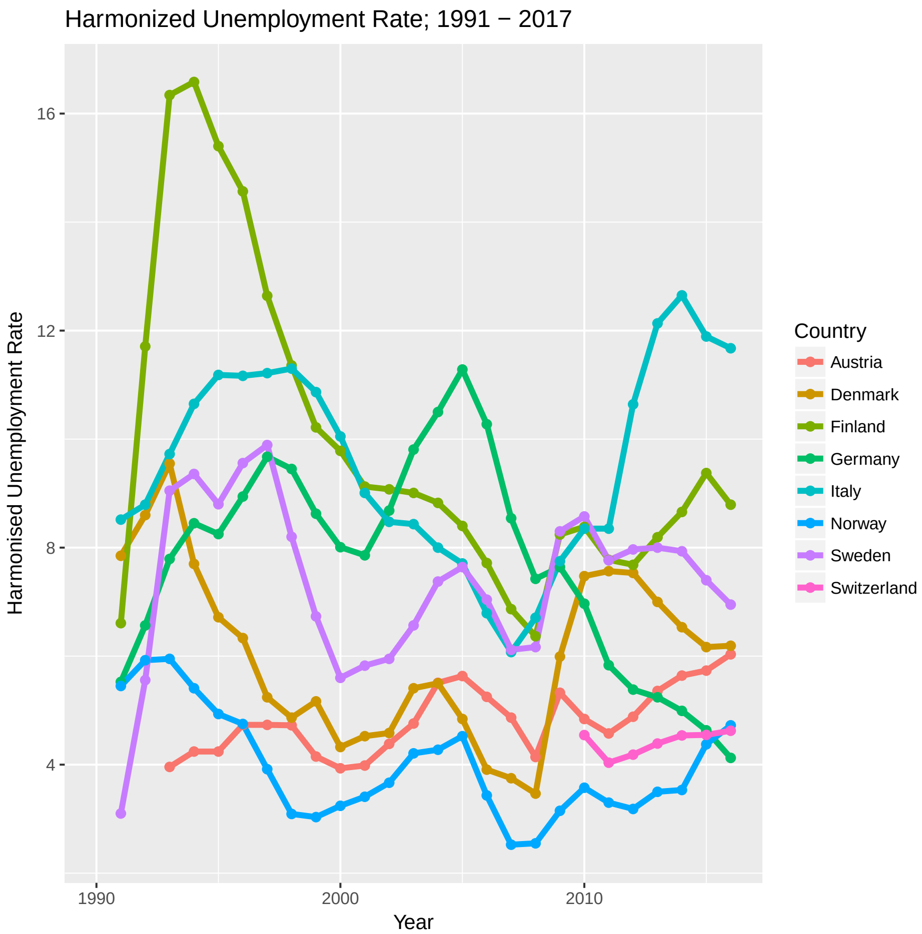


Figure 11: Different unemployment rates over time of all control countries and Germany

Weights	Country	Unit Number
0.00	Austria	1
0.00	Denmark	2
0.00	Finland	3
0.07	Italy	5
0.00	Norway	6
0.93	Sweden	7

Figure 12: Assigned weights for the control countries

A.2 Germany, GDP per Capita

Weights	Country	Unit Number
0.58	Austria	1
0.00	Denmark	2
0.00	Finland	3
0.23	Italy	5
0.00	Norway	6
0.19	Sweden	7

Figure 13: Assigned weights for the control countries

	Treated	Synthetic	Sample Mean
infl	1.5	1.7	1.7
loggdp	1.4	1	1
emp_protect	2.7	2.5	2.4
exports	43.1	45	43.1
particip	59.6	59.7	63.9
special.schooling.2007.2015	171.4	165.4	162.1
special.socsp.1995.2013	25.4	22.8	21.7

Figure 14: Comparison between synthetic control group and real Germany

A.3 Hungary, Unemployment Rate

Weights	Country	Unit Number
0.20	Austria	1
0.00	Czech Republic	2
0.66	Estonia	3
0.00	Germany	4
0.00	Greece	5
0.00	Poland	7
0.01	Slovak Republic	8
0.13	Slovenia	9

Figure 15: Assigned weights for the control countries

	Variable Weights
schooling	0.48
emp_protect	0.01
particip	0.00
bipacc	0.49
exports	0.01
infl	0.00

Table 1: Predictive weights of predictor variables for the SCM

A.4 Hungary, GDP per Capita

Weights	Country	Unit Number
0.00	Austria	1
0.00	Czech Republic	2
0.02	Estonia	3
0.00	Germany	4
0.20	Greece	5
0.77	Poland	7
0.00	Slovak Republic	8
0.00	Slovenia	9

Figure 16: Assigned weights for the control countries

	Treated	**Synthetic**	**Sample Mean**
schooling	139	123.9	131.3
emp_protect	2	2.3	2.5
particip	54.5	54.7	58.6
unemp	8.2	12.8	9.5
bipacc	−4.5	−6	−3
loggdp	1.6	3.2	2.5
infl	5.2	2.9	2.9

Figure 17: Fit of predicted variables of Hungary and its synthetic control group

	Variable Weights
schooling	0.00
emp_protect	0.21
particip	0.69
unemp	0.00
bipacc	0.00
loggdp	0.07
infl	0.02

Table 2: Predictive weights of variables for the SCM

A.5 Spain, Unemployment

Weights	Country	Unit Number
0.4	Estonia	1
0.6	Greece	2
0.0	Ireland	3
0.0	Italy	4
0.0	Portugal	5
0.0	Slovak Republic	6

Figure 18: Assigned weights for the control countries

	Variable Weights
infl	0.00
loggdp	0.20
emp_protect	0.27
exports	0.10
particip	0.36
special.schooling.2007.2015	0.00
special.socsp.1995.2013	0.07

Table 3: Predictive weights of variables for the SCM

A.6 Spain, GDP per capita

Weights	Country	Unit Number
0.03	Estonia	1
0.39	Greece	2
0.24	Ireland	3
0.15	Italy	4
0.03	Portugal	5
0.17	Slovak Republic	6

Figure 19: Assigned weights for the control countries

	Variable Weights
infl	0.00
loggdp	0.52
emp_protect	0.47
exports	0.00
particip	0.00
special.socsp.1995.2011	0.00
special.schooling.2007.2015	0.00

Table 4: Predictive weights of variables for the SCM

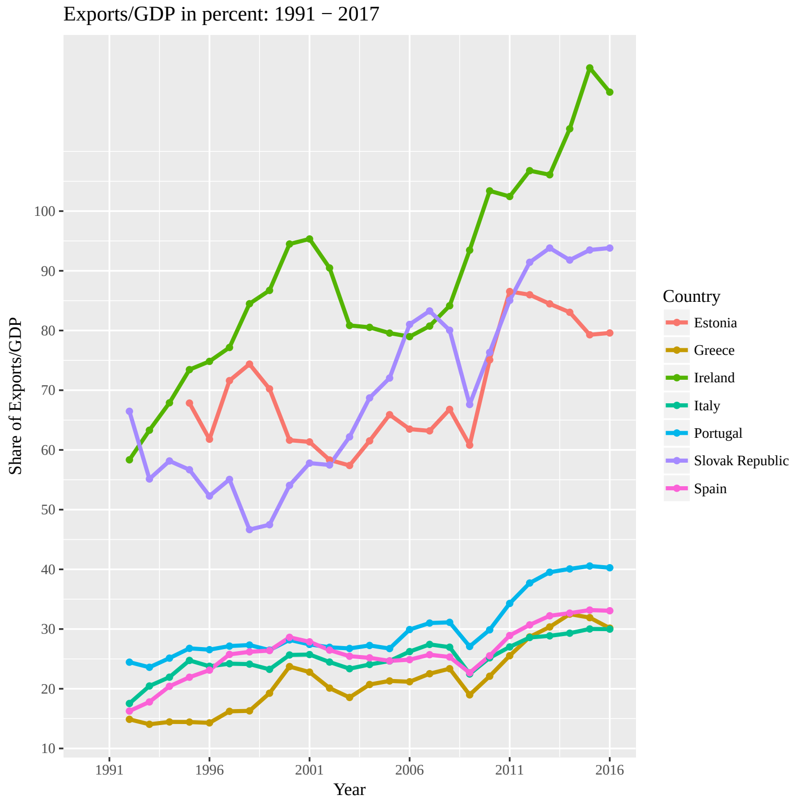


Figure 20: Assigned weights for the control countries

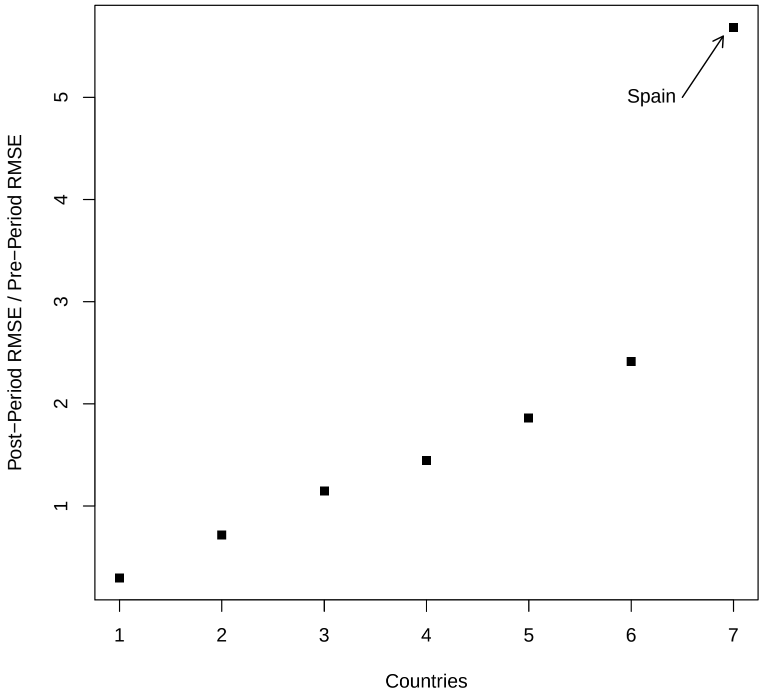


Figure 21: Ratio of pre/post-treatment MSPE for all countries, 1 = Slovak Republic, 2 = Portugal, 3 = Estonia, 4 = Ireland, 5 = Italy, 6 = Greece, 7 = Spain

Weights	Country	Unit Number
0.00	Belgium	1
0.04	Canada	2
0.00	Denmark	3
0.00	Finland	4
0.73	France	5
0.00	Luxembourg	6
0.00	Netherlands	7
0.03	Norway	8
0.00	Poland	9
0.00	Sweden	11
0.00	Switzerland	12
0.20	Turkey	13
0.00	United Kingdom	14

Figure 22: Assigned weights for the control countries in the modified version

	7
2002	-99.25
2003	-402.80
2004	-708.78
2005	-118.84
2006	139.94
2007	261.50
2008	431.91
2009	517.79
2010	92.77
2011	31.02
2012	-110.09
2013	-524.94
2014	-620.30
2015	-3920.57
2016	-3180.53

Table 5: Annual difference between Spain and its synthetic control

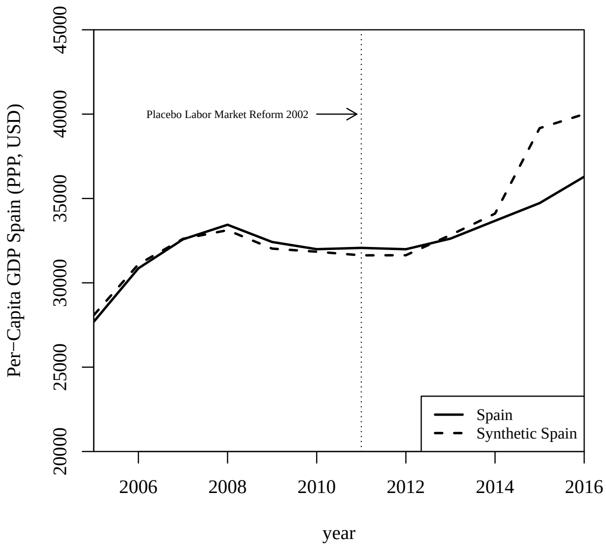


Figure 23: Reassigning the treatment year to 2011

B Data sources

In this Appendix, I describe the data used in the evaluation and provide its sources.

- Harmonized nemployment rate. Source: [OECD, 2017b], retrieved via OECD.stat (14.06.2017)
- GDP per capita (PPP, 2010 USD). Source: `http://stats.oecd.org/index.aspx?DataSetCode=PDB_LV#` (14.06.2017), retrieved via OECD.stat. GDP levels are converted to a common currency using Purchasing Power Parities (PPPs)
- Annual logged GDP. Source: `https://stats.oecd.org/Index.aspx?DataSetCode=PDB_GR` (27.06.2017), retrieved via OECD.stat
- Inflation Rate. Source: `https://stats.oecd.org/index.aspx?queryid=221` (14.06.2017)
- Primary surplus as a percentage of GDP (in %). Source: `https://stats.oecd.org/index.aspx?queryid=221#` (14.06.2017), retrieved from OECD.stat
- Trade opennes. Source: `http://databank.worldbank.org/data/reports.aspx?source=2&series=NE.EXP.GNFS.ZS&country=#` (21.06.2017), retrieved from the World Bank Dataset
- Labour participation rate. Source: [OECD, 2017b], retrieved via OECD.sat (21.06.2017). The labour force participation rates is calculated as the labour force divided by the total working-age population. The working age population refers to people aged 15 to 64.
- Percentage of government spending on social welfare. Source: `https://stats.oecd.org/Index.aspx?DataSetCode=IDD#`, retrieved via OECD.stat (14.06.2017) Social expenditure comprises cash benefits, direct in-kind provision of goods and services, and tax breaks with social purposes. The indicator is measured as a percentage of GDP
- Employment Protection. Source: `https://stats.oecd.org/Index.aspx?DataSetCode=EPL_OV`, retrieved via OECD.stat (13.07.2017). The OECD indicators of employment protection are synthetic indicators of the strictness of regulation on dismissals and the use of temporary contracts.
- Percentage of higher schooling. Source: `https://data.worldbank.org/indicator/SE.SEC.PRIV.ZS` (13.07.2017), retrieved via World Bank Dataset, Educational attainment, at least completed lower secondary, population 25+, total (%), (cumulative)

C Declaration of Authorship

I, Dominik Prugger, hereby declare that I have not previously submitted the present work for other examinations. I wrote this work independently. All sources, including sources from the Internet, that I have reproduced in either an unaltered or modified form (particularly sources for texts, mathematical derivations, statistical programs and interpretations), have been acknowledged by me as such. I understand that violations of these principles will result in proceedings regarding deception or attempted deception

Student's signature:

Name (in capitals): Dominik Prugger

Date of submission: 1 September, 2017