Comparative Politics and the Synthetic Control Method

Alberto Abadie – Harvard University and NBER
Alexis Diamond – International Finance Corporation
Jens Hainmueller – Massachusetts Institute of Technology

October 2011

Abstract

As a result of a recent and highly prominent methodological debate (King, Keohane, and Verba, 1994; Tarrow, 1995; Brady and Collier, 2004; George and Bennett, 2005) a widespread consensus has emerged about the necessity of establishing bridges between the quantitative and the qualitative approaches to empirical research in political science. In this article, we discuss the use of the synthetic control method (Abadie and Gardeazabal, 2003; Abadie, Diamond, and Hainmueller, 2010) as a way to bridge the quantitative/qualitative divide in comparative politics. The synthetic control method provides a systematic way to chose comparison units in comparative case studies. This systematization (advocated by King, Keohane, and Verba, 1994, among others) opens the door to precise quantitative inference in small-sample comparative studies, without precluding the application of qualitative approaches. That is, borrowing a colorful expression from Tarrow (1995), the synthetic control method allows researchers to put "qualitative flesh on quantitative bones". We illustrate the main ideas behind the synthetic control method with an application where we study the economic impact of the 1990 German reunification in West Germany.

Alberto Abadie, John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge MA 02138, USA. E-mail: alberto_abadie@harvard.edu. Alexis Diamond, International Finance Cooperation, 1737 Cambridge Street, Cambridge, MA 02138. E-mail: ADiamond@ifc.org. Jens Hainmueller, Department of Political Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: jhainm@mit.edu. In addition, all authors are affiliated with Harvard's Institute for Quantitative Social Science (IQSS). The title of this article pays homage to Lijphart (1971), one of the earliest and most influential studies on the methodology of the comparative method in political science.

Companion software developed by the authors (*Synth* package for MATLAB, R, and Stata) is available at `http://www.mit.edu/~jhainm/synthpage.html`.

I. Introduction

Starting with Alexis de Tocqueville's *Democracy in America* comparative case studies have become distinctly associated to empirical research in political science (Tarrow, 2010). Comparative researchers base their studies on the meticulous description and analysis of the characteristics of a small number of selected cases, as well as of their differences and similarities. By carefully studying a small number of cases, comparative researchers gather evidence at a level of granularity that is impossible to incorporate to quantitative studies, which tend to focus on larger samples but employ much coarser descriptions of the sample units.[1] However, large-sample quantitative studies are often favored in the social sciences because they provide precise numerical results, which can easily be compared across studies, and because they are better adapted to traditional methods of statistical inference.[2]

As a result of a recent and highly prominent methodological debate (King, Keohane, and Verba, 1994; Tarrow, 1995; Brady and Collier, 2004; George and Bennett, 2005), a widespread consensus has emerged about the necessity of establishing bridges between the quantitative and the qualitative approaches to empirical research in political science. In particular, there have been calls for the development and use of quantitative methods that complement and facilitate qualitative analysis in comparative studies (Gerring, 2007; Tarrow, 1995, 2010; Sekhon, 2004).[3] At the other end of the methodological spectrum, a recent strand of the quantitative literature is advocating for research designs that, like in Mill's Method of Difference, carefully select the comparison units in order to reduce biases in observational studies (Card and Krueger, 1994; Rosenbaum, 2005).

In this article we discuss how synthetic control methods (Abadie and Gardeazabal, 2003;

[1]See Lijphart (1971), Collier (1993), Mahoney and Rueschemeyer (2003), George and Bennett (2005), and Gerring (2004, 2007) for careful treatments of case study research in the social sciences.

[2]In this respect, Lord Kelvin's dictum,

> "... when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind"

(Thompson, 1889), is ubiquitously quoted by proponents of quantitative methods.

[3]The qualitative analysis technique of Ragin (1987) is an important earlier contribution motivated in part by the desire of bridging the gap between the quantitative and qualitative methods in the social sciences.

Abadie, Diamond, and Hainmueller, 2010) can be applied to complement and facilitate comparative case studies in political science. Following Mill's Method of Difference, we focus on a study design based on the comparison of outcomes between units representing the case of interest, defined by the occurrence of a specific event or intervention that is the object of the study, and otherwise similar but unaffected units.[4] In this design, comparison units are intended to reproduce the counterfactual of the case of interest in absence of the event or intervention under scrutiny.[5]

The selection of comparison units is a step of crucial importance in comparative case studies, because using inappropriate comparisons may lead to erroneous conclusions. If comparison units are not sufficiently similar to the units representing the case of interest, then any difference in outcomes between these two sets of units may be a mere reflection of the disparities in their characteristics (King, Keohane, and Verba, 1994; Geddes 2003; George and Bennett 2005). The synthetic control method provides a systematic way to choose comparison units in comparative case studies. Formalizing the way comparison units are chosen not only represents a way of systematizing comparative case studies (as advocated, among others, by King, Keohane, and Verba, 1994), it also has profound implications for inference. We demonstrate that the main barrier to quantitative inference in comparative studies comes not from the small-sample nature of the data, but from the absence of an explicit mechanism that determines how comparison units are selected. By carefully specifying how units are selected for the comparison group, the synthetic control method opens the door to the possibility of precise quantitative inference in comparative case studies, without precluding qualitative approaches to the same data set.

One distinctive feature of comparative political science is that the units of analysis are

[4]This is the "most similar" design in the terminology of Przeworski and Teune (1970) and the "comparable-cases strategy" of Lijphart (1971, 1975).

[5]It is important to recognize that comparative politics is "a river of many currents" (Hall, 2003) and researchers therefore may have different motivations for selecting cases beyond the goal of establishing a valid controlled comparison to remove selection bias (Collier and Mahoney, 1996; Bennett and Elman, 2006; Hall, 2003). For example, researchers may select particular cases in order to examine causal mechanisms through within-case methods such as process tracing (George and Bennett, 2005) or causal process observations (Collier, Mahoney, and Seawright, 2004). We do not intend to critique these approaches, as we see our proposals as complementary to existing methods.

usually aggregate entities, like countries or regions, for which suitable single comparisons often do not exist (Lijphart, 1971; Collier 1993; George and Bennett, 2005; Gerring, 2007). The synthetic control method is based on the observation that, when the units of analysis are a few aggregate entities, a combination of comparison units (which we term "synthetic control") often does a better job reproducing the characteristics of unit or units representing the case of interest than any single comparison unit alone. Motivated by this consideration, the comparison unit in the synthetic control method is selected as the weighted average of all potential comparison units that best resembles the characteristics of the case of interest.

Relative to regression analysis, the synthetic control method has important advantages. Using a weighted average of units as a comparison precludes the type of extrapolation exercises that regression results are often based on.[6] In section II.B we show that the regression estimator can be expressed also as a weighted average of the outcomes of comparison units, with weights that sum to one. However, regression weights are not restricted to lie in between zero and one, allowing extrapolation. Moreover, like in small sample comparative studies and in contrast to regression analysis techniques, the synthetic control method makes explicit the contribution of each comparison unit to the counterfactual of interest. This allows researchers to use quantitative *and* qualitative techniques to analyze the similarities and differences between the units representing the case of interest and the synthetic control.

In this section we have briefly described and motivated the synthetic control method. We finish it by taking stock of the main advantages of the synthetic control method. Relative to small sample studies, the synthetic control method helps in the selection of comparison cases and opens the door to a method of quantitative inference. Relative to large sample regression-based studies, the synthetic control method avoids extrapolation biases and allows a more focused description and analysis of the similarities and differences between the case of interest and the comparison unit. We carefully elaborate on these points later in the article.

The rest of the article is organized as follows. Section II describes the synthetic control

[6] See King and Zeng (2006) for a discussion of the dangers of extrapolation in regression analysis.

estimator, provides a formal comparison between this estimator and a conventional regression estimator, and discuss inferential techniques. Section III illustrates the main points of the article applying the synthetic control method to the study of the economic effects of the 1990 German reunification in West Germany. Section IV concludes. Data sources for the empirical example are provided in an appendix.

II. Synthetic Control Method for Comparative Case Studies

A. Constructing Synthetic Comparison Units

Suppose that there is a sample of $J+1$ units (e.g., countries) indexed by j, among whom unit $j=1$ is the case of interest and units $j=2$ to $j=J+1$ are potential comparisons.[7] Borrowing from the medical literature, we will say that $j=1$ is the "treated unit", that is, the unit exposed to the event or intervention of interest, while units $j=2$ to $j=J+1$ constitute the "donor pool", that is, a reservoir of potential comparison units. Studies of this type abound in political science (Gerring, 2007; Tarrow, 2010). Because comparison units are meant to approximate the counterfactual of the case of interest without the intervention, it is important to restrict the donor pool to units with outcomes that are thought to be driven by the same structural process as the unit representing the case of interest and that were not subject to structural shocks to the outcome variable during the sample period of the study. In the application explored later in this article we investigate the effects of the 1990 German reunification on the economic prosperity in West Germany. In that example, the case of interest is West Germany around 1990 and the set of potential comparisons is a sample of OECD countries around the same time.

We assume that the sample is a balanced panel, that is, a longitudinal data set where all units are observed at the same time periods, $t=1,\ldots,T$.[8] We also assume that the

[7]For expositional simplicity, we focus on the case where only one unit is exposed to the event or intervention of interest. This is done without a loss of generality. In cases where multiple units are affected by the event of interest, our method can be applied to each of the affected units separately or to the aggregate of all affected units.

[8]This is typically the case in political science applications, where sample units are large administrative entities like nation-states or regions, for which data are periodically collected by statistical agencies. We do not require, however, that the sample periods are equidistant in time.

sample includes a positive number of pre-intervention periods, T_0, as well as a positive number of post-intervention periods, T_1, with $T = T_0 + T_1$. The goal of the study is to measure the effect of the event or intervention of interest on some post-treatment outcome.

As stated above, the pre-intervention characteristics of the treated unit can often be much more accurately approximated by a combination of untreated units than by any untreated unit alone. We define a synthetic control as a weighted average of the units in the donor pool. That is, a synthetic control can be represented by a $(J \times 1)$ vector of weights $W = (w_2, \ldots, w_{J+1})'$, with $0 \leq w_j \leq 1$ for $j = 2, \ldots J$ and $w_2 + \cdots + w_{J+1} = 1$. Choosing a particular value for W is equivalent to choosing a synthetic control. Following Mill's Method of Difference, we propose selecting the value of W such that the characteristics of the treated unit are best resembled by the characteristics of the synthetic control. Let X_1 be a $(k \times 1)$ vector containing the values of the pre-intervention characteristics of the treated unit that we aim to match as closely as possible, and let X_0 be the $k \times J$ matrix collecting the values of the same variables for the units in the donor pool. The differences between the pre-intervention characteristics of the treated unit and a synthetic control is given by the vector $X_1 - X_0 W$. We select the synthetic control, W^*, that minimizes the size of this difference. This can be operationalized in the following manner. For $m = 1, \ldots, k$, let X_{1m} be the value of the m-th variable for the treated unit and let X_{0m} be a $1 \times J$ vector containing the values of the m-th variable for the units in the donor pool. Abadie and Gardeazabal (2003) and Abadie, Diamond and Hainmueller (2010) choose W^* as the value of W that minimizes:

$$\sum_{m=1}^{k} v_m (X_{1m} - X_{0m} W)^2, \tag{1}$$

where v_m is a weight that reflects the predictive power of the m-th variable on the outcome.[9]

[9]More formally, let $\|\cdot\|$ be a norm or seminorm in $\mathbb{R}^k$. One example is the Euclidean norm, defined as $\|u\| = \sqrt{u'u}$ for any $(k \times 1)$ vector u. For any positive semidefinite $(k \times k)$ matrix, V, $\|u\| = \sqrt{u'Vu}$ defines a seminorm. The synthetic control $W^* = (w_2^*, \ldots, w_{J+1}^*)'$ is selected to minimize $\|X_1 - X_0 W\|$, subject to $0 \leq w_j \leq 1$ for $j = 2, \ldots J$ and $w_2 + \cdots + w_{J+1} = 1$. Typically, V is selected to weight covariates in accordance to their predictive power on the outcome (see Abadie and Gardeazabal, 2003; Abadie, Diamond and Hainmueller, 2010). If V is diagonal with main diagonal equal to $(v_1, \ldots, v_k)$, then W^* is equal to the value of W that minimizes equation (1). Because W^* is invariant to scale changes in $(v_1, \ldots, v_k)$, these weights can always be normalized to sum to one.

Let Y_{jt} be the outcome of unit j at time t. In addition, let Y_1 be a $(T_1 \times 1)$ vector collecting the post-intervention values of the outcome for the treated unit. That is, $Y_1 = (Y_{1\,T_0+1}, \ldots, Y_{1\,T})'$. Similarly, let Y_0 be a $(T_1 \times J)$ matrix, where column j contains the post-intervention values of the outcome for unit $j+1$. The synthetic control estimator of the effect of the treatment is given by the comparison of post-intervention outcomes between the treated unit, which is exposed to the intervention, and the synthetic control, which is not exposed to the intervention, $Y_1 - Y_0 W^*$. That is, for a post-intervention period t (with $t \geq T_0$) the synthetic control estimator of the effect of the treatment is given by the comparison between the outcome for the treated unit and the outcome for the synthetic control at that period:

$$Y_{1\,t} - \sum_{j=2}^{J+1} w_j^* Y_{j\,t}.$$

The matching variables in X_0 and X_1 are meant to be predictors of post-intervention outcomes, which are themselves not affected by the intervention. Critics of Mill's Method of Differences rightfully point out that the applicability of the method may be limited by the presence of unmeasured factors affecting the outcome variables as well as heterogeneity in the effect of observed and unobserved factors. However, using a linear factor model, Abadie, Diamond, and Hainmueller (2010) argue that if the number of pre-intervention periods in the data is large, matching on pre-intervention outcomes (that is, on the pre-intervention counterparts of Y_0 and Y_1) helps controlling for the unobserved factors affecting the outcome of interest as well as for the heterogeneity of the effect of the observed and unobserved factors on the outcome of interest. The intuition of this result is immediate: only units that are alike in both observed and unobserved determinants of the outcome variable as well as in the effect of those determinants on the outcome variable should produce similar trajectories of the outcome variable over extended periods of time. Once it has been established that the unit representing the case of interest and the synthetic control unit have similar behavior over extended periods of time prior to the intervention, a discrepancy in the outcome variable following the intervention is interpreted as produced by the intervention itself.[10]

B. Comparison to Regression

Constructing a synthetic comparison as a linear combination of the untreated units with coefficients that sum to one may appear unusual. We show below, however, that a regression-based approach also uses a linear combination of the untreated units with coefficients that sum to one as a comparison, albeit in an implicit way. In contrast with the synthetic control method, the regression approach does not restrict the coefficients of the linear combination that define the comparison unit to be in between zero and one, therefore allowing extrapolation outside the support of the data.

The proof is as follows. A regression-based counterfactual of the outcome for the treated unit in the absence of the treatment is given by the $(T_1 \times 1)$ vector $\widehat{B}'X_1$, where $\widehat{B} = (X_0X_0')^{-1}X_0Y_0'$ is the $(k \times T_1)$ matrix of regression coefficients of Y_0 on X_0.[11] As a result, the regression-based estimate of the counterfactual of interest is equal to Y_0W^{reg}, where $W^{\text{reg}} = X_0'(X_0X_0')^{-1}X_1$. Let ι be a $(J \times 1)$ vector of ones. The sum of the regression weights is $\iota'W^{\text{reg}}$. Notice that $(X_0X_0')^{-1}X_0\iota$ is the $(k \times 1)$ vector of coefficients of the regression of ι on X_0. Assume that, as usual, the regression includes an intercept, so the first row of X_0 is a vector of ones.[12] Then $(X_0X_0')^{-1}X_0\iota$ is a $(k \times 1)$ vector with the first element equal to one and all the rest equal to zero. The reason is that $(X_0X_0')^{-1}X_0\iota$ is the vector of coefficients of the regression of ι on X_0. Because ι is a vector of ones and because the first row of X_0 is also a vector of ones, the only non-zero coefficient of this regression is the intercept, which takes value equal to one. This implies that $\iota'W^{\text{reg}} = \iota'X_0'(X_0X_0')^{-1}X_1 = 1$ (because the first element of X_1 is equal to one).

That is, the regression estimator is a weighting estimator with weights that sum to one.

[10]In this respect, the synthetic control method combines the synchronic and diachronic approaches outlined in Lijphart (1971). As pointed out by Gerring (2007), this approach is close in spirit to comparative historical analysis methods (Pierson and Skocpol, 2002; Mahoney and Rueschemeyer, 2003).

[11]That is, each column r of the matrix $\widehat{B}$ contains the regression coefficients of the outcome variable at period $t = T_1 + r - 1$ on X_0.

[12]It is easy to extend the proof to the more general case where the unit vector, ι, belongs to the subspace of $\mathbb{R}^{J+1}$ spanned by the rows of $[X_1\, X_0]$.

However, regression weights are unrestricted and may take on negative values or values greater than one. As a result, estimates of counterfactuals based on linear regression may extrapolate beyond the support of comparison units. Even if the characteristics of the case of interest cannot be approximated using a weighted average of the characteristics of the potential controls, the regression weights extrapolate to produce a perfect fit. In more technical terms, even if X_1 is far from the convex hull of the columns of X_0, regression weights extrapolate to produce $X_0 W^{\mathrm{reg}} = X_0 X_0'(X_0 X_0')^{-1} X_1 = X_1$.

Regression extrapolation can be detected if the weights W^{reg} are explicitly calculated, because it results in weights outside the $[0, 1]$ interval. We do not know, however, of any previous article that explicitly computes regression weights, as we are also unaware of previous results casting regressions as weighting estimators with weights that sum to one. Because regression weights are not calculated in practice, the extent of the extrapolation produced by regression techniques is typically hidden from the analyst. In the empirical section below we provide a comparison between the unit synthetic control weights and the regression weights for the German reunification example. For that example we show that the regression-based counterfactual relies on extrapolation. Extrapolation is, however, unnecessary in the context of the German reunification example. We show that there exist a synthetic control that closely fits the values of the characteristics of the units and that does not extrapolate outside of the support of the data.[13]

C. Inference with the Synthetic Control Method

The use of statistical inference in comparative case studies is complicated by the small sample nature of the data, the absence of randomization, and by the fact that probabilistic

[13] While using weights that sum to one and fall in the $[0, 1]$ interval prevents extrapolation biases, interpolation biases may be severe in some cases, especially if the donor pool contains units of characteristics that are very different from those of the unit representing the case of interest. Interpolation biases can be minimized by restricting the donor pool to units that are similar to the one representing the case of interest and/or complementing the $\|X_1 - X_0 W\|$ objective function for the weights with penalty terms that reflect the discrepancies in characteristics between the unit representing the case of interest and the units with positive weights in the synthetic control. This type of penalty terms can also be useful to select a synthetic control in cases when the minimization of $\|X_1 - X_0 W\|$ has multiple solution because X_1 falls in the convex hull of the columns of X_0.

sampling is not employed to select sample units. These limitations render traditional approaches to statistical inference unfeasible.[14] However, by systematizing the process of estimating the counterfactual of interest, the synthetic control method enables researchers to conduct a wide array of falsification exercises, which we term "placebo studies", that provide the building blocks for an alternative mode of qualitative and quantitative inference. This alternative model of inference is based on the premise that our confidence that a particular synthetic control estimate reflects the impact of the intervention under scrutiny would be severely undermined if we obtained estimated effects of similar or even greater magnitudes in cases where the intervention did not take place.

Suppose, for example, that the synthetic control method estimates a sizeable effect for a certain intervention of interest. Our confidence about the validity of this result would all but disappear if the synthetic control method also estimated large effects when applied to dates when the intervention did not occur (Heckman and Hotz, 1989). We refer to these falsification exercises as "in-time placebos". These tests are feasible if there are available data for a sufficiently large number of time periods when no structural shocks to the outcome variable occurred. In the example of section III we consider the effect of the 1990 German reunification on per capita GDP in West Germany. The German reunification occurred in 1990, but we have data starting in 1960. As a result, we are able to test whether the synthetic control method produces large estimated effects when applied to dates earlier than the reunification, like 1970 or 1980. If we find estimated effects that are of similar or larger magnitude than the one estimated for the 1990 reunification, our confidence that the effect estimated for the 1990 reunification is attributable to reunification itself would greatly diminish (because in the 1960-1990 period Germany did not experience a structural shock to the economy of a magnitude that could potentially match that of the German reunification). In that case, the placebo studies would suggest that synthetic controls do not provide good predictors of the trajectory of the outcome in West Germany in periods when the reunification did not occur. Conversely, in section III we find a very large effect for the 1990 German reunification, but no effect at all when we artificially

[14] See Rubin (1990) for a description of the different modes of statistical inference for causal effects.

reassign the reunification period in our data to 1980 or 1970.

Another way to conduct placebo studies is to reassign the intervention not in time, but to units not directly exposed to the intervention. Here the premise is that our confidence that a sizeable synthetic control estimate reflects the effect of the intervention would disappear if similar or larger estimates arose when the intervention is artificially reassigned in the data set to units not directly exposed to the intervention.

A particular implementation of this idea consists of applying the synthetic control method to estimate placebo effects for every potential control unit in the donor pool. This creates a distribution of placebo effects against which we can then evaluate the effect estimated for the unit that represents the case of interest. Our confidence that a large synthetic control estimate reflects the effect of the intervention would be severely undermined if the magnitude of the estimated effect fell well inside the distribution of placebo effects. Like in traditional statistical inference, a quantitative comparison between the distribution of placebo effects and the synthetic control estimate can be operationalized through the use of p-values. In this context, a p-value can be constructed by estimating the effect of the intervention for each unit in the sample and then calculating the proportion of estimated effects that are greater or equal to the one estimated for the unit representing the case of interest. Notice that this inferential exercise reduces to classical randomization inference when the intervention is randomized (Rosenbaum, 2005). In absence of randomization, the p-value still has an interpretation as the probability of obtaining an estimate at least as large as the one obtained for the unit representing the case of interest when we reassign at random the intervention in our data set.

In the next section, we compare the reunification effect estimated for West Germany to the placebo effects estimated for all the other countries in the sample. The effect for West Germany clearly stands out as the largest negative effect when the synthetic control estimator is applied to every unit in the donor pool.

III. Application: The Economic Cost of the 1990 German Reunification

A. *The German Reunification and the West German Economy*

In this section, we apply the synthetic control method to estimate the impact of the 1990 German reunification, one of the most significant political events in post-war European history. After the crumbling of the Berlin Wall on November 9, 1989, the German Democratic Republic and the Federal Republic of Germany officially reunified on October 3, 1990. At that time, per capita GDP in West Germany was about three times higher than in East Germany (Lipschitz and McDonald, 1990). Given the large income disparity, the integration of both states after more than half a century of separation called for political and economic adjustments of unprecedented complexity and scale. The 1990 German reunification therefore provides an excellent case study to examine the economic consequences of political integration.

When policy makers pursue political integration such as monetary unions, mergers of sub-national units, or other related efforts to redraw political boundaries, they are often motivated by overarching political goals that can trump concerns about the possibly severe economic consequences of integration (Haas, 1958; Eichengreen and Frieden, 1994; Feldstein, 1997; Alesina and Spolaore, 2003). By estimating the economic costs of political integration, we gain a better understanding of how much political leaders are willing to sacrifice in terms of economic prosperity for their citizens in order to further broader national political goals. The tradeoff between political gains and economic sacrifice was particularly clear in the case of the German reunification where many observes at the time feared that West German taxpayers would suffer severely to "foot the bill" of reunification and avoid a "Mezzogiorno problem" of continuing fiscal transfers to the East (Dornbusch and Wolf, 1991; Akerlof et. al., 1991; Adams, Alexander and Gagonet, 1993; Hallett and Ma, 1993).

We construct a synthetic West Germany as a convex combination of other advanced industrialized countries chosen to resemble the values of economic growth predictors for West Germany prior to the reunification. The synthetic West Germany is meant to replicate the (counterfactual) per capita GDP trend that West Germany would have experienced in the absence of the 1990 reunification. We then estimate the effect of the reunification by comparing the actual (with reunification) and counterfactual (without reunification) trends in per capita GDP for West Germany.[15]

B. Data and Sample

We use annual country-level panel data for the period 1960-2003. The German reunification occurred in 1990, giving us a pre-intervention period of 30 years. Our sample period ends in 2003 because a roughly decade-long period after the reunification seems like a reasonable limit on the span of plausible prediction of the effect of reunification. Recall that the synthetic West Germany is constructed as a weighted average of potential control countries in the donor pool. Our donor pool includes a sample of 17 OECD member countries that are commonly used in the comparative political economy literature on advanced industrialized countries. The sample includes: Australia, Austria, Belgium, Canada, Denmark, France, Greece, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Switzerland, United Kingdom, the United States, and West Germany.[16]

We provide a list of all variables used in the analysis in the data appendix, along with data sources. The outcome variable, Y_{jt}, is the real per capita GDP. GDP is PPP-adjusted and measured in 2002 U.S. Dollars (USD, hereafter) in country j at time t. For the pre-reunification characteristics in X_{jt} we rely on a standard set of economic growth predictors: per capita GDP, inflation rate, industry share of value added, investment rate, education, and a measure of trade openness (see the appendix for details). For each variable we checked

[15]Additionally, one could also try to estimate the effect of reunification on East Germany. However, concerns about quality of the official East German statistics before the German reunification renders this a questionable endeavor. See Lipschitz and McDonald, 1990.

[16]To construct this sample we started with the 24 OECD-member countries in 1990. We first excluded Luxembourg and Iceland because of their small size and because of the peculiarities of their economies. We also excluded Turkey, which had in 1990 a level of per capita GDP well below the other countries in the sample. We finally excluded Finland, Sweden, Ireland because these countries were affected by profound structural shocks during the sample period. Ireland experienced a rapid *Celtic Tiger* expansion period in the 1990's. Finland and Sweden experienced profound financial crises at the beginning of the 1990's. However, the exclusion of these countries from the sample is rather innocuous because, when included in the sample, they obtain zero weights in the synthetic control for West Germany.

that the German data refers exclusively to the territory of the former West Germany.[17] We experimented with a wide set of additional growth predictors, but their inclusion did not change our results substantively.

Using the techniques described in Section II, we construct a synthetic West Germany with weights chosen so that the resulting synthetic West Germany best reproduces the values of the predictors of per capita GDP in West Germany in the pre-reunification period. The growth predictors are weighted according to their predictive power for the per capita GDP trajectory prior to reunification using a data-driven procedure. This ensures that the synthetic West Germany approximates West Germany most closely on the most important predictors.[18] We estimate the effect of the German reunification on per capita GDP in West Germany as the difference in per capita GDP levels between West Germany and its synthetic counterpart in the years following the reunification. Finally, we perform a series of placebo studies and robustness checks.

C. Constructing a Synthetic Version of West Germany

Table 1 shows the weights of each country in the synthetic version of West Germany. The synthetic West Germany is a weighted average of Austria, Switzerland, the United States, the Netherlands, and Japan with weights decreasing in this order. All other countries in the donor pool obtain zero weights. As a comparison, Table 1 also reports the weights that regression analysis employs implicitly when applied to the same data (these weights are backed out using the formulas in Section II.B). By construction, both sets of weights sum to one. The two sets of weights show some similarities. For example, Austria receives the highest weight in both approaches. Overall, however, the weights are very different. For example, regression weights Japan almost as much as Austria, while the weight obtained

[17]For that purpose, when necessary, our data set was supplemented with data from the German Federal Statistical Office (*Statistisches Bundesamt*).

[18]More formally, we choose V among all positive definite and diagonal matrices such that the resulting synthetic West Germany best approximates (in a minimum mean squared prediction error sense) the per capita GDP trajectory of the actual West Germany during the pre-reunification period, 1960-1989. This is the similar to the method used in Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010). In section III.E we propose an alternative way to choose V based on out-of-sample validation and demonstrate the robustness of our results to the choice rule for V.

by Austria in the synthetic control is more than four times larger than that of Japan. Moreover, regression assigns negative weights to six of the 17 control units in the donor pool, including some rather large negative weights for Italy (-0.17), Portugal (-0.14), New Zealand (-0.08), and Norway (-0.07).

Table 2 compares the pre-reunification characteristics of West Germany to those of the synthetic West Germany, and also to those of a population-weighted average of the 17 OECD countries in the donor pool. The synthetic West Germany approximates the pre-1990 values of the economic growth predictors for West Germany far more accurately than the average of our sample of other OECD countries. The synthetic West Germany is very similar to the actual West Germany in terms of pre-1990 per capita GDP, schooling, investment rate, and industry share. Compared to the average of OECD countries, the synthetic West Germany also matches West Germany much closer on the inflation rate. Because West Germany had the lowest inflation rate in the sample during the pre-reunification years, this variable cannot be perfectly fitted using a combination of the comparison countries. There is a noticeable discrepancy between West Germany and its synthetic counterpart in terms of the trade openness variable. Trade openness, however, has a low predictive power on per capita GDP, reflected in a low value for v_m (not reported, see section II for the definition of v_m). Overall, Table 2 suggests that the synthetic West Germany provides a much better comparison for West Germany than the average of our sample of other OECD countries. As seen in Figure 1, even before the German reunification, the OECD average experienced a trend in per capita GDP different than the trend for West Germany. We will show, however, that the synthetic control described in this section accurately reproduces the pre-1990 per capita GDP trend for West Germany.

One of the central points of this article is that the synthetic control method provides the qualitative researcher with a quantitative tool to select or validate comparison units. In our analysis, Austria, Switzerland, the United States, the Netherlands, and Japan emerge, in this order, as potential comparisons to West Germany. Regression analysis fails to provide such a list. In a regression analysis, typically all units contribute to the regression fit, and the contribution of units with large positive regression weights may be compensated or eliminated by the contributions of units with negative weights.

D. The Effect of the 1990 Reunification

Figure 2 displays the per capita GDP trajectory of West Germany and its synthetic counterpart for the 1960-2003 period. The synthetic West Germany almost exactly reproduces the per capita GDP for West Germany during the entire pre-reunification period. This close fit for the pre-reunification per capita GDP and the close fit that we obtain for the GDP predictors in Table 2 demonstrates that there exists a combination of other industrialized countries that reproduces the economic attributes of West Germany before the reunification. That is, it is possible to closely reproduce economic characteristics of West Germany before the 1990 reunification without extrapolating outside of the support of the data for the donor pool.

Our estimate of the effect of the German reunification on per capita GDP in West Germany is given by the difference between the actual West Germany and its synthetic version, visualized in Figure 3. We estimate that the German reunification did not have much of an effect on West German per capita GDP in the first two years immediately following reunification. In this initial period per capita GDP in the synthetic West Germany is even slightly lower than in the actual West Germany, which is broadly in line with arguments about an initial demand boom (see, for example, Meinhardt et al., 1995). From 1992 onwards, however, the two lines diverge substantially. While per capita GDP growth decelerates in West Germany, for the synthetic West Germany per capita GDP keeps ascending at a pace similar to that of the pre-unification period. The difference between the two series continues to grow until the end of the sample period. Thus, our results suggest a pronounced negative effect of the reunification on West German income. We find that over the entire 1990-2003 period, per capita GDP was reduced by about 1400 USD per year on average, which amounts to approximately 7 percent of the 1990 baseline level. In 2003, per capita GDP in the synthetic West Germany is estimated to be about 11 percent higher than in the actual West Germany.

It is possible that these estimates are conservative. If the German reunification had negative spillover effects on the per capita GDP of the countries included in the synthetic control, then the synthetic control would provide an underestimate of the counterfactual per capita GDP trajectory for West Germany in the absence of the reunification and, therefore an underestimate of the effect of the reunification on per capita GDP in West Germany.

E. Placebo Studies

To evaluate the credibility of our results, we conduct a battery of placebo studies where the event of interest, that is the German reunification, is reassigned in the data set to years different than 1990 and countries different than West Germany. We first compare the reunification effect estimated above for West Germany to placebo effects obtained after reassigning in our data the German reunification to a period before the reunification actually took place. Large placebo estimates would undermine our confidence that the results in Figure 2 are indeed indicative of the economic cost of the reunification and not merely driven by lack of predictive power.

We first show results for the case when reunification is reassigned to the year 1980, ten years earlier than it actually occurred. In this placebo study the pre-treatment period is 1960-1979, and to compute the synthetic control we lag the predictors variables accordingly. The upper panel in Figure 4 displays the results of this "in-time placebo" study. The synthetic West Germany almost exactly reproduces the evolution of per capita GDP in the actual West Germany for the 1960-1979 period. Most importantly, the per capita GDP trajectories of West Germany and its synthetic counterpart do not diverge considerably after 1980. That is, in contrast to the actual 1990 German reunification, our 1980 placebo reunification has no perceivable effect. This suggests that the gap estimated in Figure 2 reflects the impact of the German reunification and not a potential lack of predictive power of the synthetic control.

The lower panel in Figure 4 displays the results of a second similar in-time placebo study where we reassign in our data the German reunification to the year 1970, twenty years earlier than it actually occurred. We run the same model as before (with appropriately lagged covariates), but the pre-treatment period is constrained to be 1960-1969. Again, in contrast to the actual 1990 German reunification, the 1970 placebo reunification shows no perceivable effect.

An alternative way to conduct placebo studies is to artificially reassign in the data the event of interest, that is the German reunification, to one of the comparison units. In this way we obtain a synthetic control estimate for a country that did not experience the event of interest. Applying this idea to each country in the donor pool allows us to judge whether the reunification effect estimated for West Germany is unusually large, compared to placebo effects obtained for countries that did not experience the event of interest.

This type of placebo study is very stringent in the sense that it will be affected by the cross-country heterogeneity in per capita GDP shocks. While the OECD countries in the donor pool did not experience events similar to the German reunification during 1990-2003, they may have experienced other types of country-specific macroeconomic shocks during this period. If those country-specific shocks had large effects on per capita GDP, the estimated placebo effects for the OECD donor pool countries will be large and perhaps comparable to the estimated effect of the German reunification. In order for this placebo exercise to be informative, the estimated effect of the 1990 reunification needs to be of larger magnitude than other, naturally occurring, placebo effects.

Figure 5 presents the results for this "across units" placebo study. The top panel shows the estimated placebo effects for all countries in the OECD donor pool, that is, the differences between per capita GDP paths in the OECD countries and in their synthetic counterparts. The synthetic control estimate for West Germany is clearly on the negative side of the placebo distribution. Some countries, like Switzerland, New Zealand, and Greece, have negative estimated placebo effects that are as large or larger than the effect estimated for West Germany. However, these countries show also non-zero "effects" before 1990, because their per capita GDP path is not accurately approximated by their synthetic controls. For example, for most of the pre-1990 period Switzerland was the richest country in the sample. Therefore, it is not possible to track the Swiss per capita GDP trajectory with a convex combination of the other countries in the sample. A similar issue applies to other countries such as the US, which is the second richest country for the pre-reunification period but then rapidly supersedes Switzerland in the 90's.[19] Portugal, the country with the lowest per capita GDP during the pre-1990 period, is also poorly approximated by its synthetic control.

If the synthetic control cannot reproduce the pre-1990 per capita GDP path for a particular country, then the placebo effect estimated for that country after 1990 does not provide a good benchmark to evaluate the significance of the post-1990 gap estimated for West Germany (a country whose per capita GDP is accurately reproduced by the the synthetic control for the entire pre-1990 period). In the lower panel of Figure 5, we report placebo effects only for those countries whose pre-1990 per capita GDP is accurately reproduced by their synthetic controls. In particular, we discard all countries for the which the root mean square prediction error (RMSPE) for the 1960-1989 period is more than three times larger than the RMSPE for West Germany for the same period.[20] Among the 11 remaining countries, West Germany clearly stands out as the country with the lowest gap line in the post-1990 period indicating that the negative gap is very unusual compared to the results we obtain for other countries.

Figure 6 shows an alternative way to compare the effect estimated for West Germany and the distribution of the placebo effects. For each country we divide the post-reunification RMSPE by its per-reunification RMSPE. This metric obviates the need to discard countries based on the pre-1990 fit since countries that do not fit well prior to the reunification are down-weighted in the ratio. Again, West Germany clearly stands out as the country with

[19]In fact the weights for the synthetic version of Switzerland are equal to one for the US and zero for all other countries. On the flip-side, Switzerland receives a large weight in the synthetic USA. This explains why after 1990 we estimate placebo effects of opposite signs for Switzerland and the US.

[20]The RMSPE measures lack of fit between the path of the outcome variable for any particular country and its synthetic counterpart. The pre-1990 RMSPE error for West Germany is defined as:

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

The RMSPE can be analogously defined for other countries or time periods.

the highest ratio. For West Germany the post-reunification gap is about 17 times larger than the pre-reunification gap. If one were to pick a country at random from the sample, the chances of obtaining a ratio as high as this one would be $1/18 \simeq 0.06$.

F. Robustness Tests

In this section we run a robustness check to test the sensitivity of our main results to the changes in the country weights, W^*. Recall from Table 1 that the synthetic West Germany is estimated as a weighted average of Austria, Switzerland, the United States, the Netherlands, and Japan, with weights decreasing in this order. Here we iteratively re-estimate the baseline model to construct a synthetic West Germany omitting in each iteration one of the countries that received a positive weight in Table 1. The motivation is to check if the estimates in section III.D are sensitive to the exclusion of any particular country from our sample. That is, with this sensitivity check we evaluate to which extent our results are driven by any particular country. The upper panel in Figure 7 displays the results. This panel reproduces Figure 2 (solid and dashed black line) incorporating the leave-one-out estimates (dashed grey lines). The lower panel in Figure 7 reproduces Figure 3 (solid black line) incorporating the leave-one-out estimates (solid grey lines). These two plots show that the results of the analysis in section III.D are fairly robust to the exclusion of any particular country from our sample of comparison countries.

G. Using Out-of-Sample Validation to Choose V

We next check the robustness of our results to changes in the way we choose V. In particular, in this section we choose V using an out-of-sample validation technique. We first divide the pre-treatment years into a training period from 1970-79 and a validation period from 1980-89. Next, using predictors measured in the training period, we select the weights v_m such that the resulting synthetic control minimizes the RMSPE over the validation period. Finally, we use the set of v_m weights selected in the previous step and predictor data measured in 1980-89 to estimate a synthetic control for West Germany. Tables 3 and 4 and Figure 8 show the results. The resulting synthetic control weights in Table 3 are very similar to those in section III.C (Table 1). Table 4 shows that the resulting synthetic West Germany closely reproduces the economic attributes of West Germany before the reunification. In fact, inflation and trade openness (averaged now over 1980-89) are more closely reproduced than in Table 2, where those averages are measured over the entire pre-reunification period (1960-89). The upper panel in Figure 8 shows the per capita GDP path of West Germany and the synthetic West Germany with v_m weights chosen by out-of-sample validation, and the lower panel shows the GDP per capita gap. Overall the results are very similar to those in Figures 2 and 3.

IV. Conclusion

There is a widespread consensus among political methodologists about the necessity to integrate and exploit complementarities between qualitative and quantitative tools for empirical research in political science. However, some of the efforts in this direction have been denounced by qualitative methodologists as attempts to impose quantitative templates on qualitative research that disregard or do not make use of the many genuine advantages of qualitative research (Brady and Collier, 2004; George and Bennett, 2005). The synthetic control method discussed in this article 'falls in between' the qualitative and quantitative methodologies and provides a potentially useful tool for researchers of both traditions. On the one hand, the synthetic control method provides a systematic way to select comparison units in quantitative comparative case studies. In this way, like in Card and Krueger (1994) and Rosenbaum (2005), the synthetic control method brings to quantitative studies the careful selection of cases that is done in qualitative analysis. In addition, by explicitly specifying the set of units that are used for comparison, the method does not preclude but facilitates detailed qualitative analysis and comparison between the case of interest and the set of comparison units selected by the method. That is, the synthetic control method can be used to guide the selection of comparison units in qualitative studies, allowing what Tarrow (1995) calls "qualitative inference with quantitative bones".

References

Abadie, A., A. Diamond, and J. Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*. 105(490): 493-505.

Abadie, A. and J. Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review*. 93(1): 112-132.

Adams, G., L. Alexander, and J. Gagon. 1993. "German unification and the European monetary system: A quantitative analysis." *Journal of Policy Modeling*. 15(4): 353-392.

Alesina, A., and E. Spolaore, 2003. *The Size of Nations*. Cambridge, MA: MIT Press.

Akerlof, G., A. K. Rose, J. L. Yellen, and H. Hessenius. 1991. "East Germany in from the Cold: The Economic Aftermath of Currency Union." *Brookings Papers on Economic Activity*. 1: 1-105.

Barro, R.J. and J. Lee. 1994. "Data Set for a Panel of 138 Countries." Available at: `http://www.nber.org/pub/barro.lee/`.

Barro, R.J. and J. Lee. 2000. "International Data on Educational Attainment: Updates and Implications." *CID Working Paper No. 42, April 2000 – Human Capital Updated Files.*

Bennett, A. and C. Elman. 2006. "Qualitative Research: Recent Developments in Case Study Methods." *Annual Review of Political Science*. 9: 455-76.

Brady, H.E. and D. Collier, Eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield.

Card, D. and A.B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review*. 84(4): 772-93.

Collier D. 1993. "The comparative method." In A. Finifter. *Political Science: The State of the Discipline II*. Washington, DC: American Political Science Association: 105-19.

Collier D and J. Mahoney. 1996. "Insights and pitfalls: selection bias in qualitative research." *World Politics*. 49(1): 56-91.

Collier D, J. Mahoney, and J. Seawright. 2004. Claiming too much: warnings about selection bias. In Brady, H.E. and D. Collier. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Eds. Lanham, MD: Rowman & Littlefield: 85-102.

Dornbusch, R. and H. Wolf. 1992. "Economic Transition in Eastern Germany." *Brookings Papers on Economic Activity*. 1: 235-261.

Eichengreen, B. and J.A. Frieden. 2000. *The Political Economy of European Monetary Unification*. Westview Press.

Feldstein, M. 1997. "The Political Economy of the European Economic and Monetary Union: Political Sources of an Economic Liability." *Journal of Economic Perspectives.* 11(4): 23-42.

Geddes B. 2003. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics.* Ann Arbor: Univ. Mich. Press.

George, A.L. and A. Bennett. 2005. *Case Studies and Theory Development in the Social Sciences.* Cambridge, MA: MIT Press.

Gerring, J. 2004. "What Is a Case Study and What Is It Good for?" *American Political Science Review* 98(2): 341-354.

Gerring, J. 2007. *Case Study Research: Principles and Practices.* Cambridge, MA: Cambridge University Press.

Haas, E.B. 1958. *The uniting of Europe: political, social, and economic forces.* Stanford University Press.

Hall, Peter. 2003. "Aligning Ontology and Methodology in Comparative Politics." In Mahoney, J. and D. Rueschemeyer. *Comparative Historical Analysis in the Social Sciences.* Cambridge, UK: Cambridge University Press: 373-429.

Hallett, A. J. and Y. Ma. 1993. "East Germany, West Germany, and Their Mezzogiorno Problem: A Parable for European Economic Integration." *Economic Journal.* 103(417): 416-428.

Heckman, J.J. and V.J. 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.

King, G., R.O. Keohane, and S. Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton, NJ: Princeton University Press.

King, G., and L. Zeng. 2006 "The Dangers of Extreme Counterfactuals." *Political Analysis.* 14(2): 131-159.

Lijphart, A. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review.* 65(3): 682-693.

Lijphart, A. 1975. "The Comparative-Cases Strategy in Comparative Research." *Comparative Political Studies.* 8(2): 158-177.

Lipschitz, L., and D. McDonald. 1990. "German Unification: Economic Issues." *International Monetary Fund Occasional Paper.* No 75.

Mahoney, J. and D. Rueschemeyer. 2003. *Comparative Historical Analysis in the Social Sciences.* Cambridge, UK: Cambridge University Press.

Meinhardt, V., B. Seidel, F. Stile, and D. Teichmann. 1995. "Transferleistungen in die neuen Bundesländer und deren wirtschaftliche Konsequenzen." *Deutsches Institut für Wirtschaftfroschung, Sonderheft 154.*

Mueller, G. 2000. *Nutzen und Kosten der Wiedervereinigung.* Baden-Baden: Nomos.

Pierson, P. and T. Skocpol. 2002. "Historical Institutionalism in Contemporary Political Science." In Katznelson, I. and H. Millner. *Political Science: The State of the Discipline.* New York: W.W. Norton.

Przeworski, A., M. Alvarez, J. Cheibub, and F. Limongi. 2000. *Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990.* Cambridge, UK: Cambridge University Press.

Przeworski, A. and H. Teune. 1970. *The Logic of Comparative Social Inquiry.* New York, NY: Wiley.

Ragin, C.C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies.* Berkeley, CA: University of California Press.

Rosenbaum, P.R. 2005. "Heterogeneity and Causality: Unit Heterogeneity and Design Sensitivity in Observational Studies." *The American Statistician.* 59(2): 147-152.

Rubin, D.B. 1990. "Formal Modes of Statistical Inference for Causal Effects." *Journal of Statistical Planning and Inference.* 25: 279-292.

Siebert, H. 1992. *Das Wagnis der Einheit: Eine Wirtschaftspolitische Therapie.* Stuttgart: Deutsche Verlags-Anstalt.

Sekhon, J.S. 2004. "Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals." *Perspectives on Politics.* 2(2): 281-293.

Tarrow, S. 1995. "Bridging the Quantitative-Qualitative Divide in Political Science." *American Political Science Review.* 89(2): 471-474.

Tarrow, S. 2010. "The Strategy of Paired Comparison: Toward a Theory of Practice." *Comparative Political Studies.* 43(2): 230-259.

Thompson, W. 1889. *Popular Lectures and Addresses Vol. I.* London, UK: MacMillan & Co.

Data Appendix

The data sources employed for the application are:

- GDP per capita (PPP 2002 USD). Source: OECD National Accounts (retrieved via the OECD Health Database). Data for West Germany was obtained from Statistisches Bundesamt 2005 (Arbeitskreis "Volkswirtschaftliche Gesamtrechnungen der Länder") and converted using PPP monetary conversion factors (retrieved from the OECD Health Database).
- Investment Rate: Ratio of real domestic investment (private plus public) to real GDP. The data is reported in five year averages. Source: Barro and Lee (1994).
- Schooling: Percentage of secondary school attained in the total population aged 25 and older. The data is reported in five year increments. Source: Barro and Lee (2000).
- Industry: industry share of value added. Source: World Bank WDI Database 2005 and Statistisches Bundesamt 2005.
- Inflation: annual percentage change in consumer prices (base year 1995). Source: World Development Indicators Database 2005 and Statistisches Bundesamt 2005.
- Trade Openness: Export plus Imports as percentage of GDP. Source: World Bank: World Development Indicators CD-ROM 2000.

Figures

Figure 1: Trends in Per-Capita GDP: West Germany vs. Rest of OECD Sample

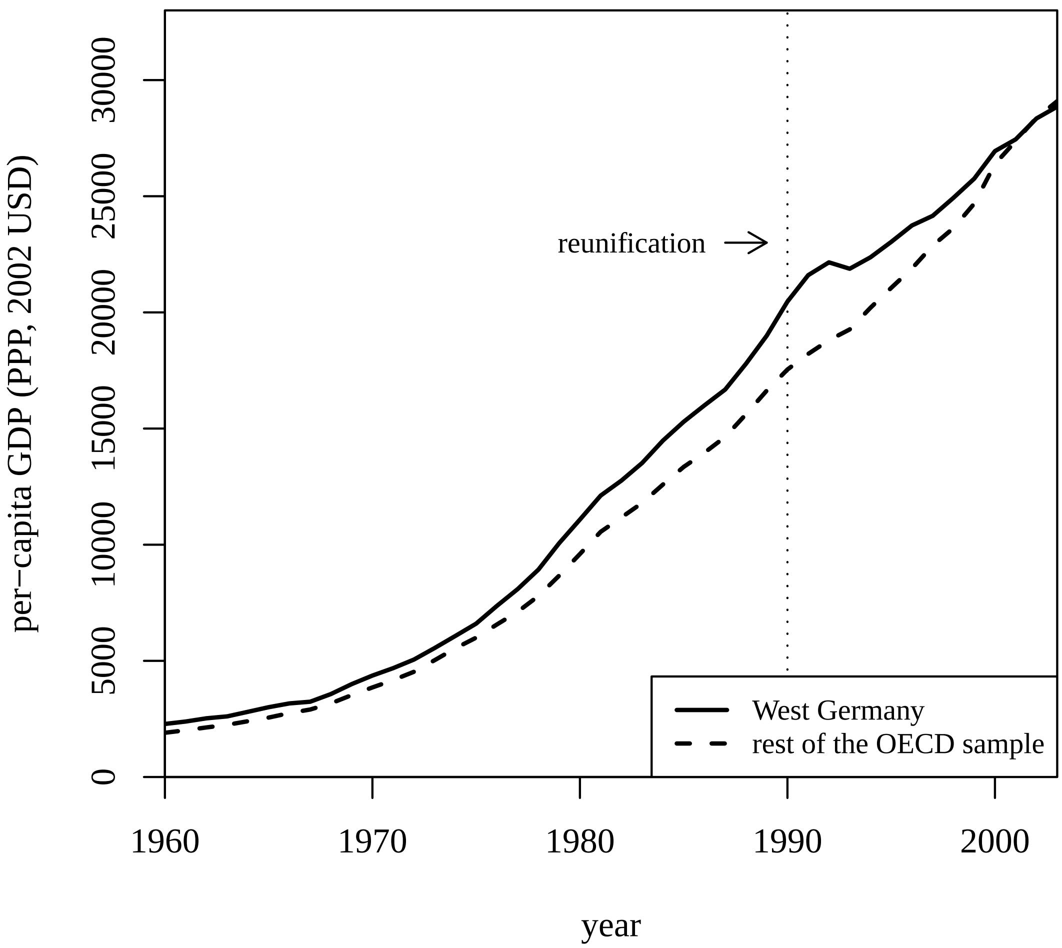


Figure 2: Trends in Per-Capita GDP: West Germany vs. Synthetic West Germany

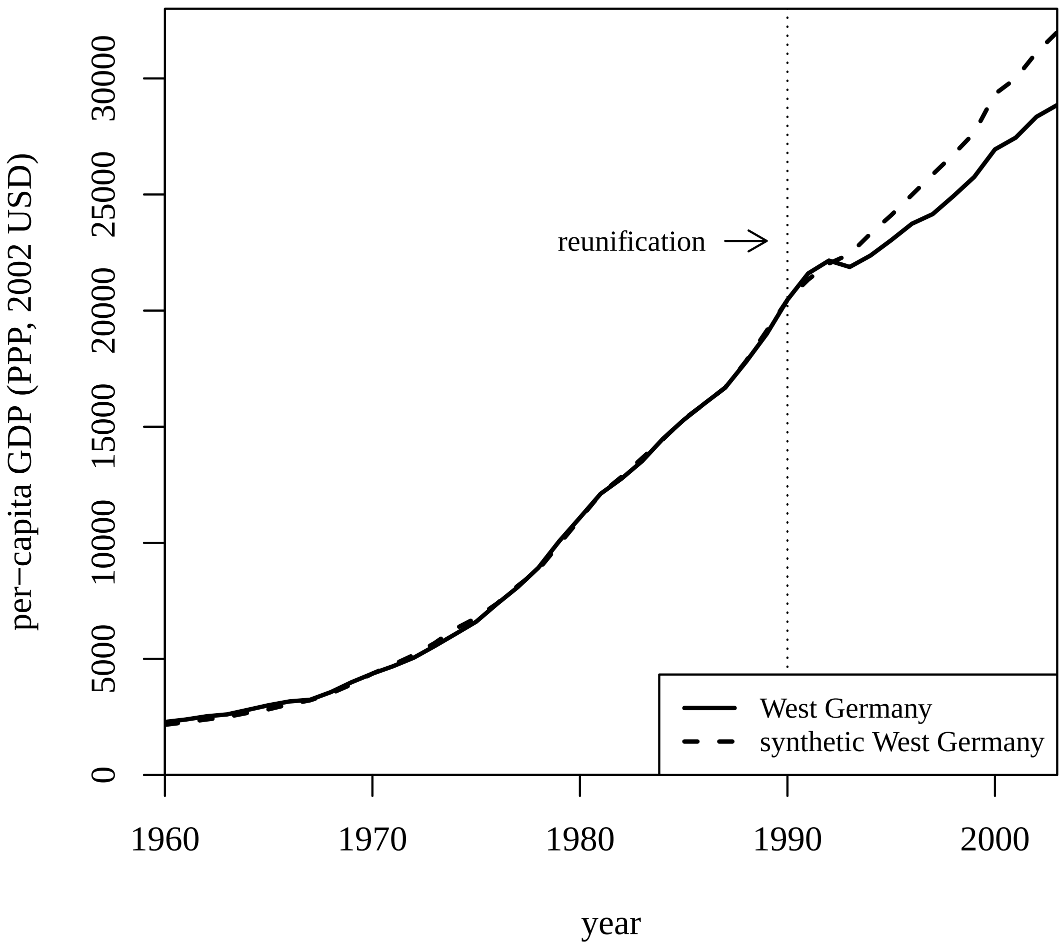


Figure 3: Per-Capita GDP Gap Between West Germany and Synthetic West Germany

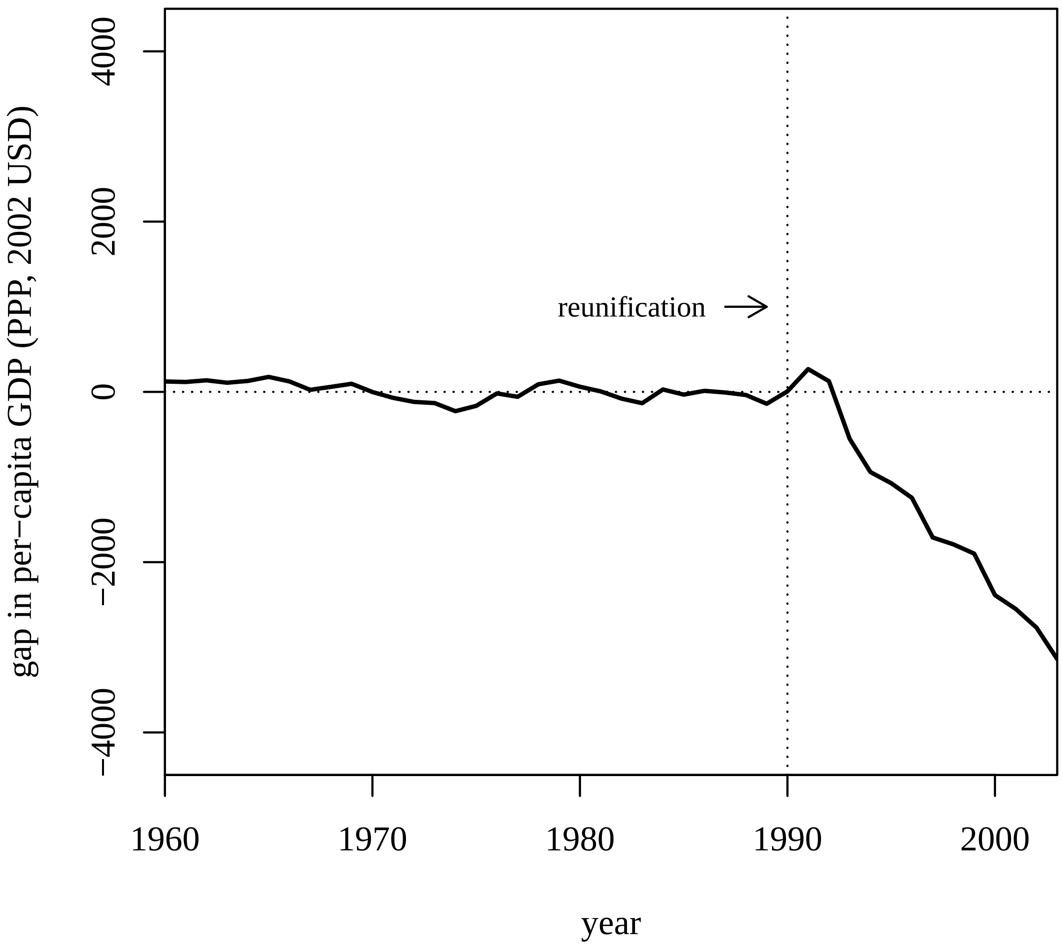


Figure 4: Placebo Reunifications 1980 and 1970 - Trends in Per-Capita GDP: West Germany vs. Synthetic West Germany

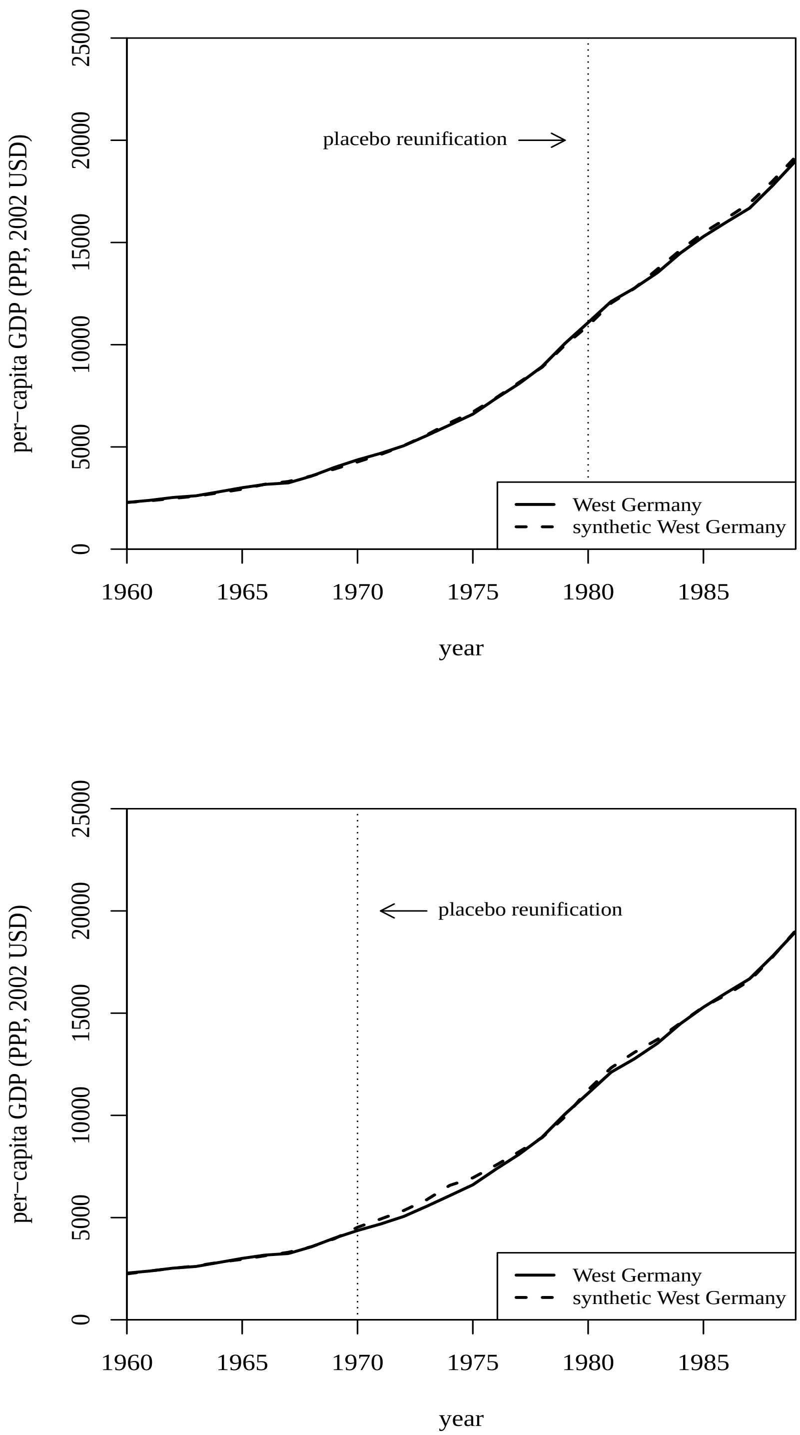


Figure 5: Per-Capita GDP gaps in West Germany and placebo gaps from applying the model to other countries (upper panel has all countries, lower panel discards countries with pre-reunification MSPE three times higher than Germany's).

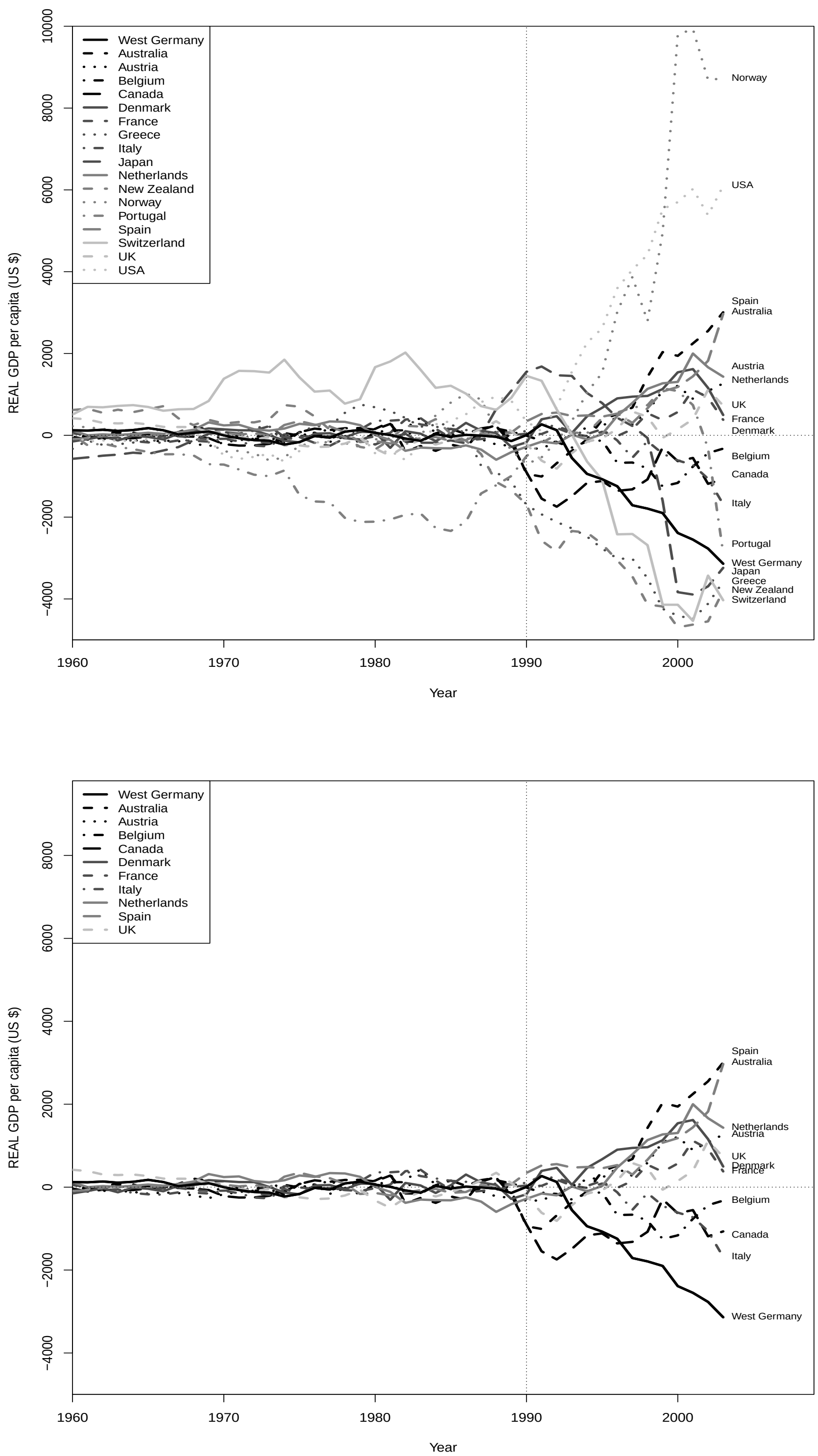


Figure 6: Ratio of post-reunification MSPE to pre-reunification MSPE: West Germany and all control countries.

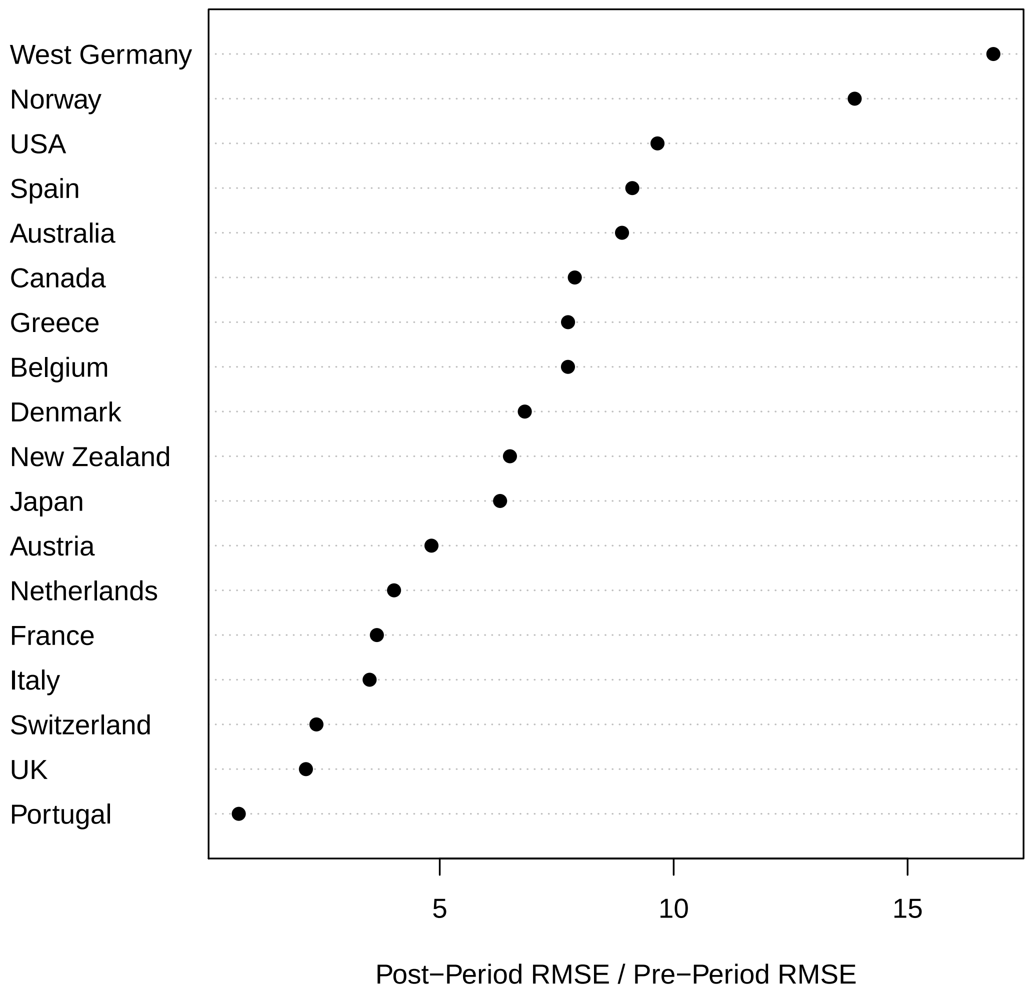


Figure 7: Leave one out: Trends and Gaps in Per-Capita GDP

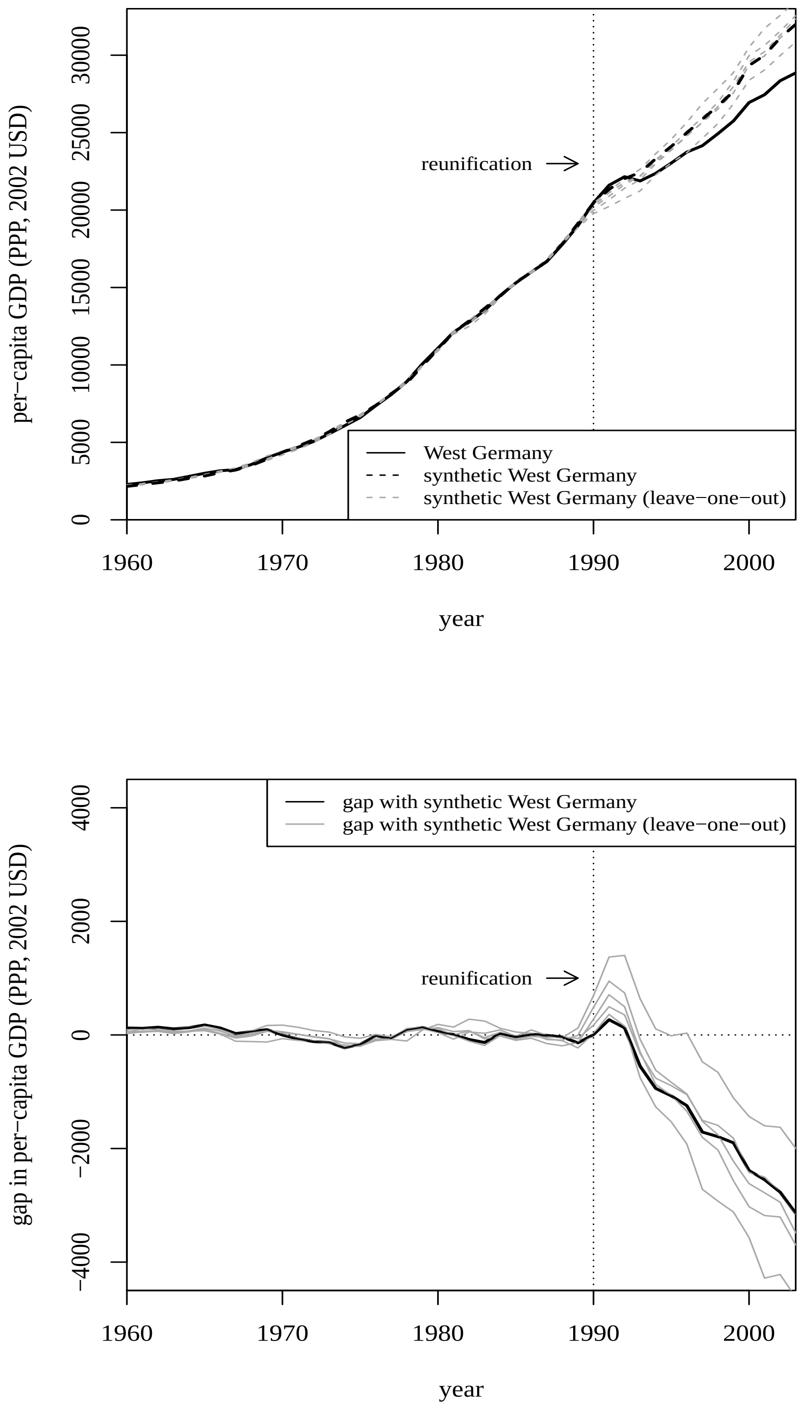
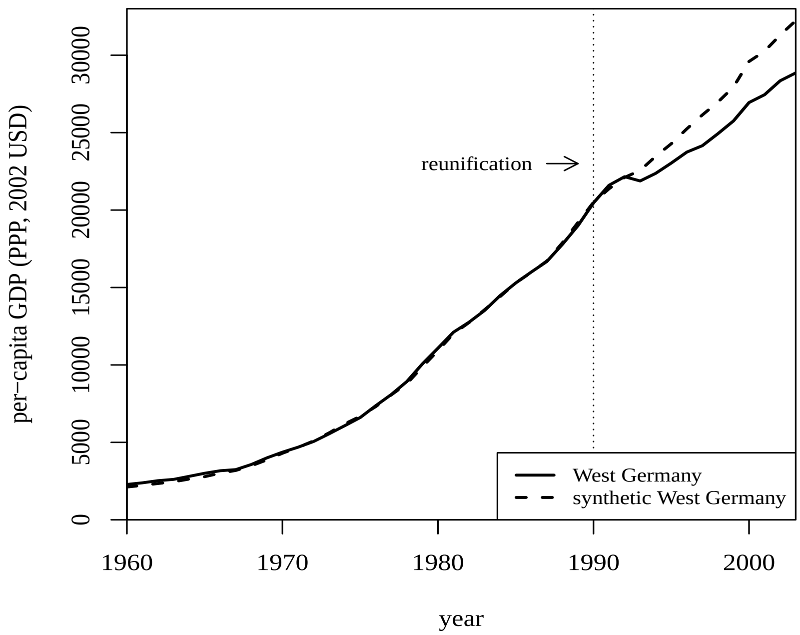
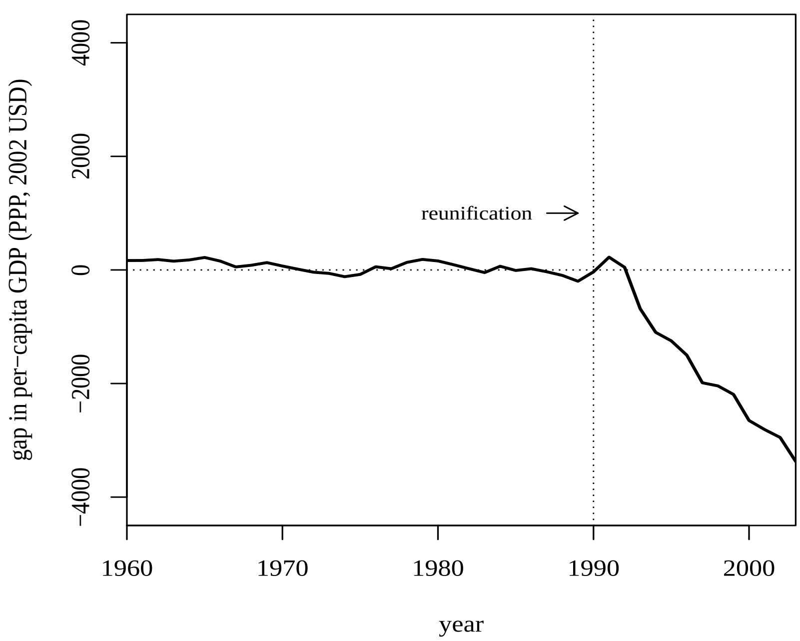


Figure 8: Out of Sample Validation: Trends and Gaps in Per-Capita GDP





Tables

Table 1: Synthetic and Regression Weights for West Germany

Country	Synthetic Control Weight	Regression Weight	Country	Synthetic Control Weight	Regression Weight
Australia	0	0.1	Netherlands	0.11	0.18
Austria	0.47	0.33	New Zealand	0	-0.08
Belgium	0	0.1	Norway	0	-0.07
Canada	0	0.09	Portugal	0	-0.14
Denmark	0	0.04	Spain	0	0
France	0	0.16	Switzerland	0.17	-0.06
Greece	0	0.02	UK	0	-0.04
Italy	0	-0.17	USA	0.14	0.21
Japan	0.11	0.32			

Note: The synthetic weight is the country weight assigned by the synthetic control method. The regression weight is the weight assigned by linear regression. See Section II for details.

Table 2: Economic Growth Predictor Means before the German Reunification

	West Germany	Synthetic West Germany	OECD Comparison Countries
GDP per-capita	8169.8	8163.1	8049.3
Trade openness	45.8	54.4	32.6
Inflation rate	3.4	4.7	7.3
Industry share	34.7	34.7	34.3
Schooling	55.5	55.6	43.8
Investment rate	27.0	27.1	25.9

Note: GDP per capita, inflation rate, and trade openness are averaged for the 1960–1989 period. Industry share is averaged for the 1980–1989 period. Investment rate and schooling are averaged for the 1980–1985 period. The last column reports a population weighted average for the 17 OECD countries in the donor pool.

Table 3: Out of Sample Validation: Country Weights in the Synthetic West Germany

Country	Synthetic Control Weight	Country	Synthetic Control Weight
Australia	0	Netherlands	0.08
Austria	0.42	New Zealand	0
Belgium	0	Norway	0
Canada	0	Portugal	0
Denmark	0	Spain	0
France	0	Switzerland	0.12
Greece	0	UK	0
Italy	0	USA	0.22
Japan	0.16		

Note: The synthetic weight is the country weight assigned by the synthetic control method. See Section II for details.

Table 4: Out of Sample Validation: Economic Growth Predictor Means before the German Reunification

	West Germany	Synthetic West Germany	OECD Comparison Countries
GDP per-capita	14870.7	14873.0	14145.7
Trade openness	56.3	56.5	34.5
Inflation rate	2.8	3.9	8.2
Industry share	34.7	34.8	34.3
Schooling	55.5	55.3	43.8
Investment rate	27.0	27.2	25.9

Note: GDP per capita, inflation rate, and trade openness are averaged for the 1980–1989 period. Industry share is averaged for the 1980–1989 period. Investment rate and schooling are averaged for the 1980–1985 period. The last column reports a population weighted average for the 17 OECD countries in the donor pool.