# Comparative Politics and the Synthetic Control Method Revisited

Joseph Francis

June 2, 2025
(v1.0)

Abstract

Abadie, Diamond, and Hainmueller (2015) have made errors of design and implementation in the application of the Synthetic Control Method (SCM) to comparative politics. When studying the effects of reunification on West Germany's GDP per capita, they have removed countries they considered to have a treatment effect, potentially predetermining the *p*-value calculated from their permutation tests. This decision seems to be based on a misunderstanding of the nature of causal inference. Testing its effects is, moreover, impossible because Abadie *et al.* have used the incorrect variable of interest. Rather than using real GDP per capita, they use current-price GDP per capita. When this data handling error is corrected, their application of the SCM to comparative politics no longer appears robust.

Keywords: causal inference, comparative politics, data handling, econometrics, synthetic control method (SCM)

JEL codes: C12, C22, C52, C80

Alberto Abadie, Alexis Diamond, and Jens Hainmueller (2015) have sought to revolutionize the study of comparative politics through the use of the Synthetic Control Method (SCM). They apply this econometric technique to determine the effects of reunification after 1990 on West Germany's GDP per capita. An algorithm builds a Synthetic West Germany out of a pool of suitable comparison countries—the so-called "donors"—to estimate the treatment effect. This technique, Abadie *et al.* (2015, 495–496, 500) state,

---

Joseph Francis is an economic historian from Wales. He is an honorary research fellow at the University of Birmingham and can be contacted at joefrancis505@gmail.com. Discussions with Daniel B. Klein aided the writing of this paper. Supplementary materials are available at https://github.com/joefrancis505/Francis_Comparative_Politics.

allows them to address "a widespread consensus [that] has emerged about the necessity of establishing bridges between quantitative and qualitative approaches to empirical research in political science." Crucially, the SCM "provides a systematic way to choose comparison units in comparative case studies." In doing so, it "opens the door to the possibility of precise quantitative inference." It allows, moreover, for "a quantitative comparison between the distribution of placebo effects and the synthetic control estimate [that] can be operationalized through the use of p-values." Abadie *et al.* thereby imply that they are making comparative politics more scientific.

Yet there are considerable issues in Abadie *et al.*'s research design, especially related to donor pool selection—a significant, but often overlooked, problem in the SCM (Francis 2025). When selecting their donor pool, Abadie *et al.* (2015, 501n14) "started with the 23 OECD member countries in 1990 (excluding West Germany)." Luxembourg and Iceland were excluded "because of their small size and because of the peculiarities of their economies." And Turkey was then dropped due to its low GDP per capita. Then, however, the choices become more questionable. Ireland is taken out because it "experienced a rapid Celtic Tiger expansion period in the 1990s." And, finally, Canada, Finland, and Sweden are excluded as a result of the "profound financial and fiscal crises at the beginning of the 1990s." Alarm bells then begin to ring because these seem to be precisely the countries that could, like Germany, have struggled in the 1990s.

By this point, the strategies of identification and inference have become somewhat confused. An identification strategy is a response to the problem of being unable to observe counterfactuals. In this case, there was no untreated West Germany to observe after reunification. As such, the SCM seeks to build a counterfactual through a weighted average of similar countries, based on algorithmic selection. The GDP per capita of this Synthetic West Germany can then be compared to the actual GDP per capita of the relevant regions of post-reunification Germany. If the respective GDP per capita figures diverge after 1990, it suggests reunification may have had an effect. To test that impression, however, an inference strategy is needed. For the SCM, this consists of permutation tests, where the algorithm constructs synthetic controls for all donors in the pool, treating them as placebo cases. If Synthetic West Germany's divergence is unusually extreme relative to the placebo cases, causality can be inferred. The problem, however, is that Abadie *et al.* appear to have preemptively removed from the donor pool the countries that they considered likely to also have a post-1990 treatment effect. They have decided they are inappropriate based on what happens

after the treatment begins, potentially predetermining the significance of the permutation tests. As a result, it is unclear whether any causality can be robustly inferred from their research design.[1]

A more qualitative detour into economic history illustrates why the assumptions underlying this decision are problematic. Abadie *et al.*'s (2015, 500) theoretical justification for excluding countries that might show a treatment effect is questionable. They note that "units that may have suffered large idiosyncratic shocks to the outcome of interest during the study period should also be excluded if such shocks would have not affected the treated unit in the absence of the treatment." Yet there is no way of knowing whether West Germany would have succumbed to similar "shocks" as the excluded countries in the absence of reunification. Furthermore, there is no way of separating out the effects of reunification from the other factors that might have affected West Germany and the excluded countries alike. Finland, after all, was in the same European Monetary System as Germany, while Sweden pegged its own currency to the Deutschmark. Even Canada was part of the same highly integrated core of North Atlantic countries. For this reason, it is impossible to conclude *a priori* that the causes of these countries' problems were not also potential causes of West Germany's problems, which Abadie *et al.* attribute solely to reunification. Their causal inference strategy therefore stumbles when it encounters the realities of globalization in the late twentieth century.

It is, moreover, impossible to test the effects of their decision to exclude these countries because Abadie *et al.* (2015) have used the wrong outcome variable in their study. Their article has been cited thousands of times, but no one seems to have previously noticed that Abadie *et al.* (2015, 501) cannot have used PPP GDP per capita "in 2002 U.S. dollars," as they believe. Why can be seen in Figure 1, which is reproduced from their article. It shows West Germany's GDP per capita increasing from around $2,300 in 1960 to $20,500 in 1990, implying a compound annual growth rate of 7.6 percent per year, which is unlikely. What happened can be seen in Figure 2, which compares the annual rates of change in all Abadie *et al.*'s series for GDP per capita with the equivalents that are currently available from the

[1] Abadie *et al.* (2015, 501n14) report that "when included in the sample, these four countries obtain zero weights in the synthetic control for West Germany. Therefore, our main results are identical whether or not we exclude these countries." Yet that is irrelevant to the inference strategy, which is based on using all donors as placebos, irrespective of whether they are given weights in the control.

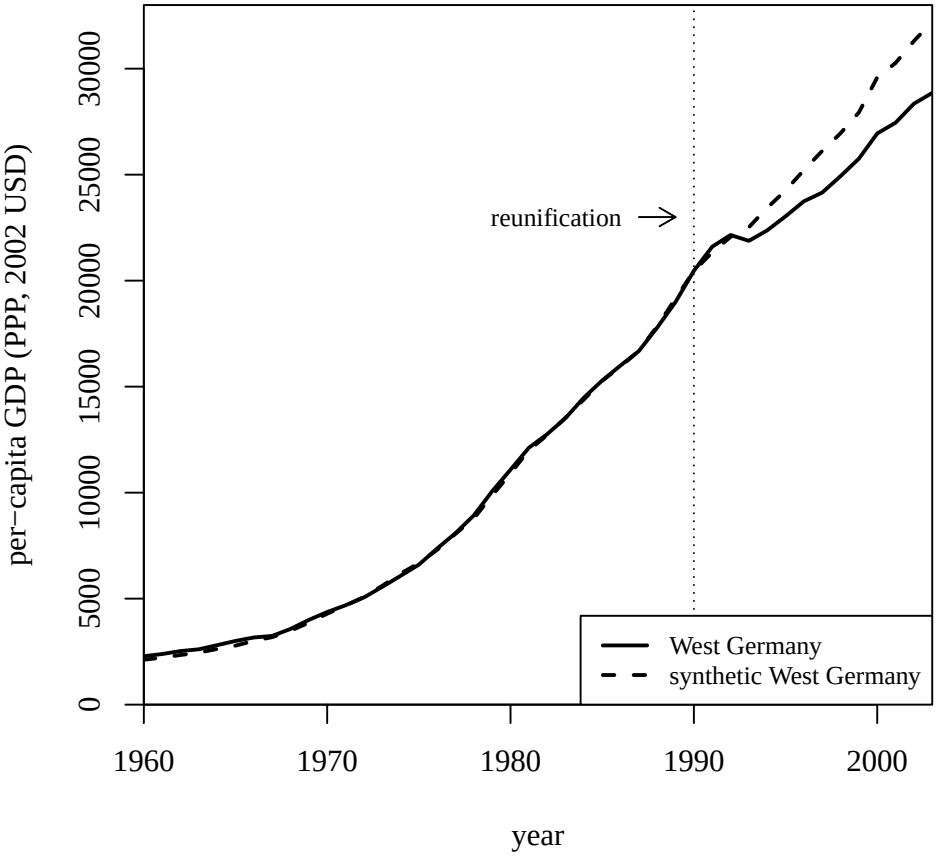



*Figure 1*
Abadie *et al.*'s West Germany

Note: Calculated from data underlying Abadie *et al.* (2015, 503, Figure 2).

OECD.[2] Panel (a) shows that there is a poor correlation when Abadie *et al.*'s series are compared to the OECD's constant-price series, while Panel (b) indicates that there is a close correlation with the current-price series. It would appear, then, that they have accidentally used PPP-adjusted GDP per capita in current prices rather than constant prices.

Aside from any concerns about research design, then, this data handling error undermines the validity of the SCM's results. Hence, Figures 3 to 7 compare the results from Abadie *et al.*'s original study with a replication using constant 2015-price GDP per capita from the World Bank's *World Development Indicators*. Otherwise, all other variables are identical to their study, except for West Germany's GDP per capita from 1991 onward,

[2] Abadie *et al.* (2015, 509) report having found GDP per capita for 1960 to 1970 from the OECD Health Database, but these data are no longer available on the OECD's website, while the archived version Health Statistics at https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7596 is not functional.

which comes from the 2024 edition of *Volkswirtschaftliche Gesamtrechnungen der Länder*. As can be seen, once inflation is accounted for, the pre-treatment fit is far less impressive, while the treatment effect is less significant—the mechanism by which German reunification affected Australia to such a degree remains to be explained. Yet it is also the case that constant-price PPP GDP per capita is an intrinsically problematic metric for comparing the wealth of nations, given the major index-number problems its calculation entails (Deaton and Heston 2010; Bolt, Inklaar, et al. 2018). Figures 8 to 12 therefore use an alternative measure. In this version, Abadie *et al.*'s (2015) current-price PPP GDP per capita have been normalized so that the United States equals 100 in each year. Nonetheless, the results of the SCM are once again unimpressive. In both replications, the $p$-value rises from the 0.06 found by Abadie *et al.* to 0.12. Overall, it is unlikely that their article would have been published if constant-price PPP GDP per capita had been used.[3]

But there is a precautionary twist to the tale. Such are the researcher degrees of freedom in the SCM that Abadie *et al.* could probably have engineered their study to make its results convincing enough for publication, even if they had used constant-price data for GDP per capita. Most worryingly, they would not have necessarily understood what they were doing. In the published version, they decided, for instance, to leave out Canada, Finland, and Sweden without realizing how it would affect their causal inference strategy. And the SCM lacks guardrails to prevent these kinds of decisions from being taken. Hence, Francis (2025) documents how the subjectivity of donor pool selection is a major issue, while Bruno Ferman, Cristine Pinto, and Vitor Possebom (2020) describe how predictor variables can be cherry-picked to achieve a pre-treatment fit between the control and the treated unit. It may be, then, that the SCM is not the revolutionary force in comparative politics that Abadie *et al.* (2015) hope. It is not as scientific as it appears because the researcher degrees of freedom are simply too great. Sometimes the old methods are best.

---

[3] The full replication can be found in the supplementary materials at https://github.com/joefrancis505/Francis_Comparative_Politics. Further replications are done using two iterations of the Maddison Project constant-price PPP GDP per capita data (Bolt and van Zanden 2020; 2025). In the 2023 iteration $p$=0.06, but visual fits are poor. With the 2020 iteration, $p$=0.18.

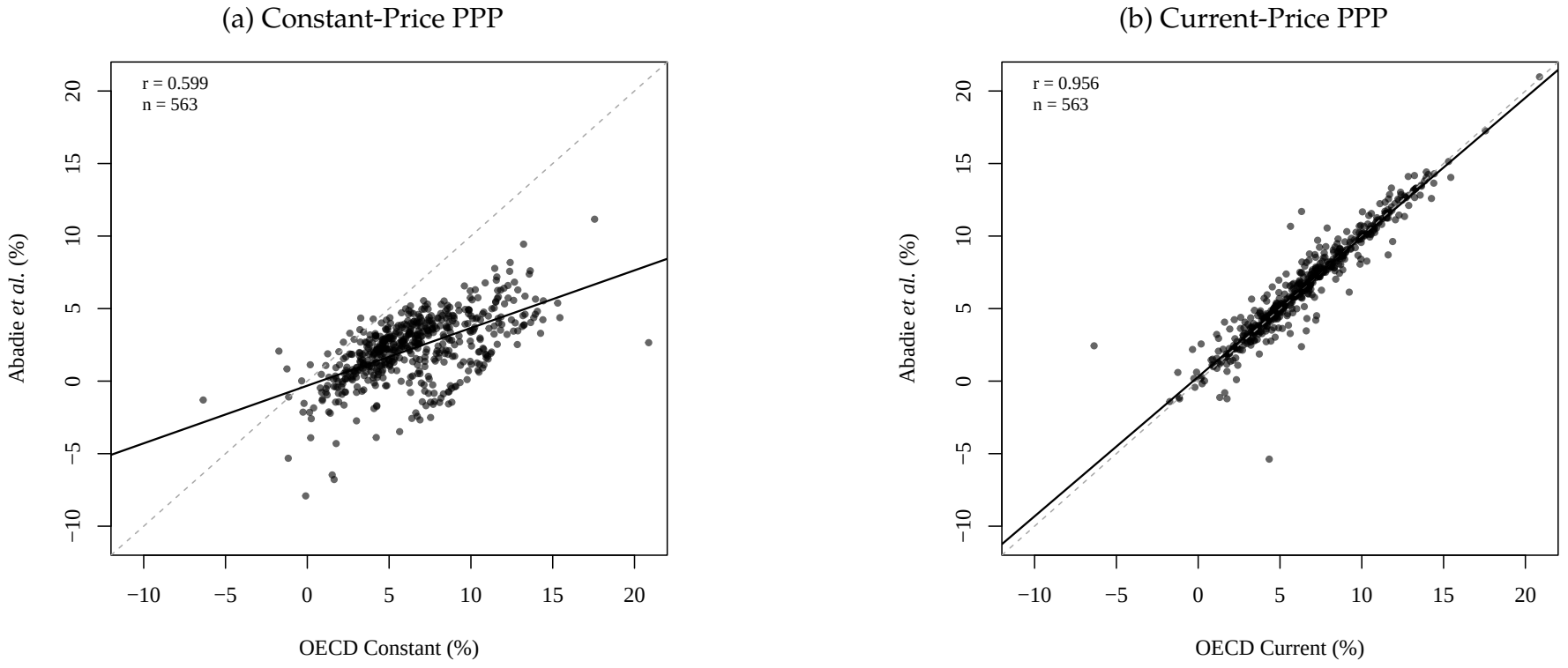



*Figure 2*

GDP per Capita Growth Rates

Note: Calculated from data underlying Abadie *et al.* (2015, available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/24714) and the OECD's *Data Explorer* (https://data-explorer.oecd.org/).

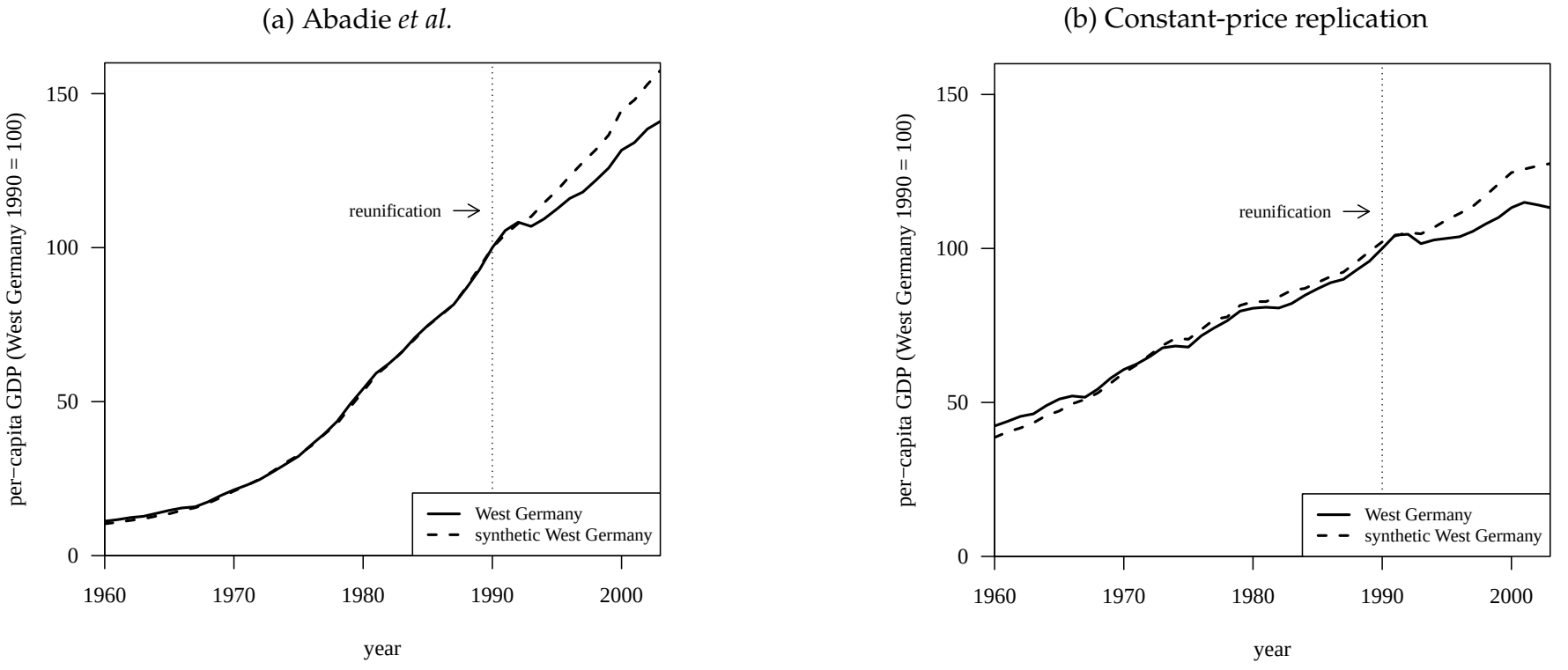



*Figure 3*

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

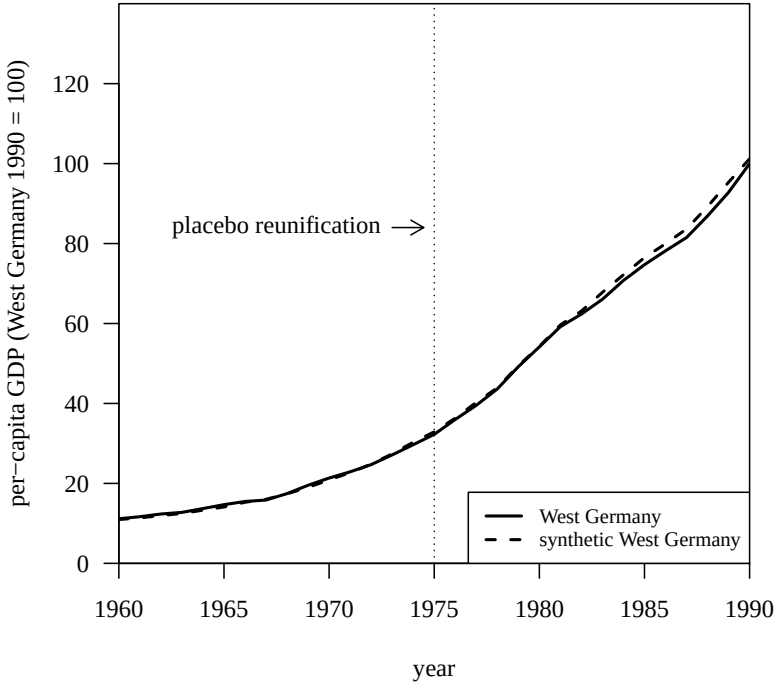
Notes: The GDP per capita series have been normalized so that West Germany in 1990 equals 100. Calculated from data underlying Abadie *et al.* (2015, available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/24714), Germany's *Volkswirtschaftliche Gesamtrechnungen der Länder* (https://www.statistikportal.de/de/vgrdl/ergebnisse-laenderebene/bruttoinlandsprodukt-bruttowertschoepfung#alle-ergebnisse, Table 8.3), and the World Bank's *World Development Indicators*.


*Figure 4*

Per Capita GDP Gap between West Germany and Synthetic West Germany

Note: See Figure 3 for details.

(a) Abadie *et al.*




(b) Constant-price replication

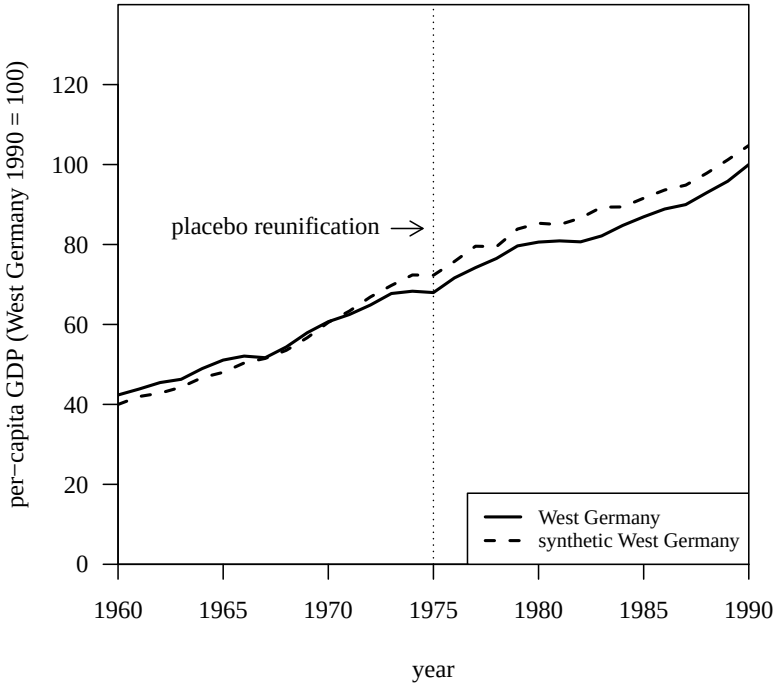



*Figure 5*

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

Note: See Figure 3 for details.

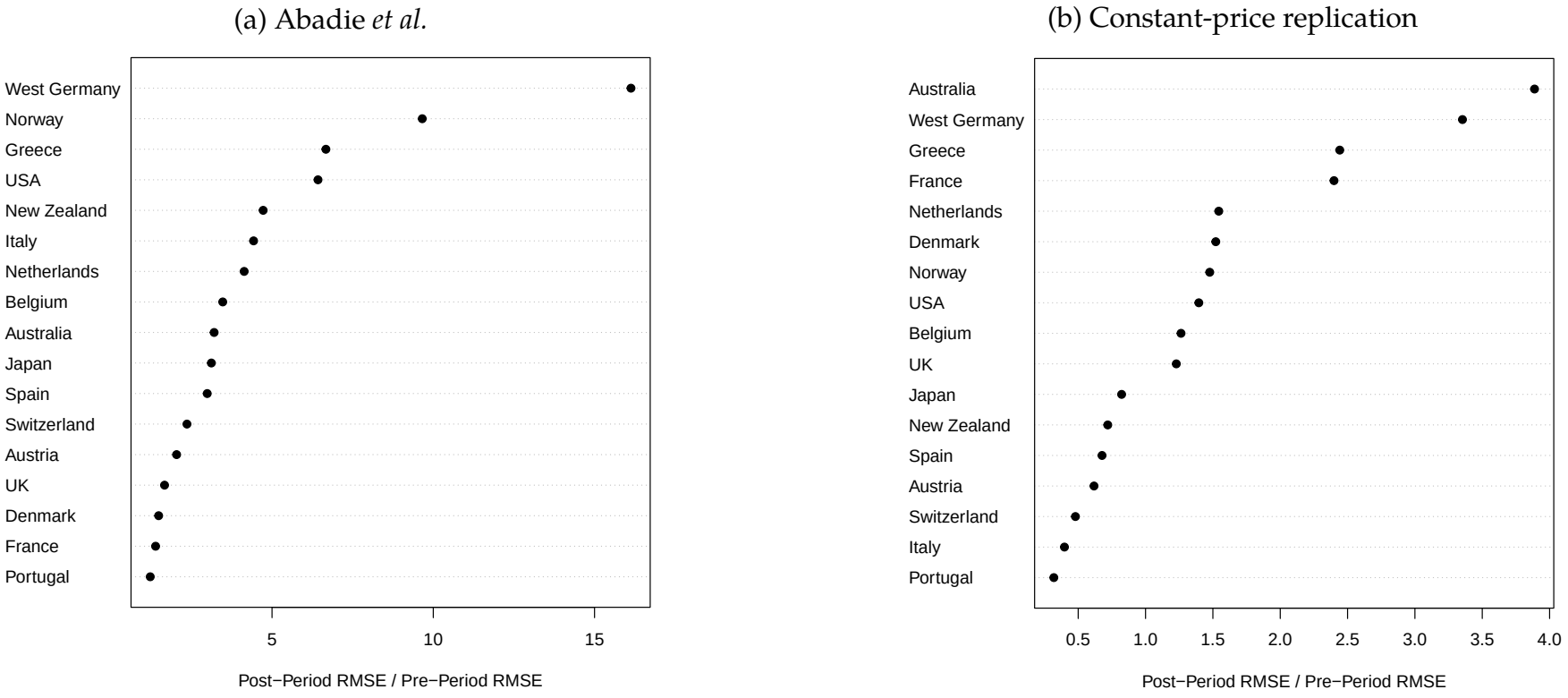
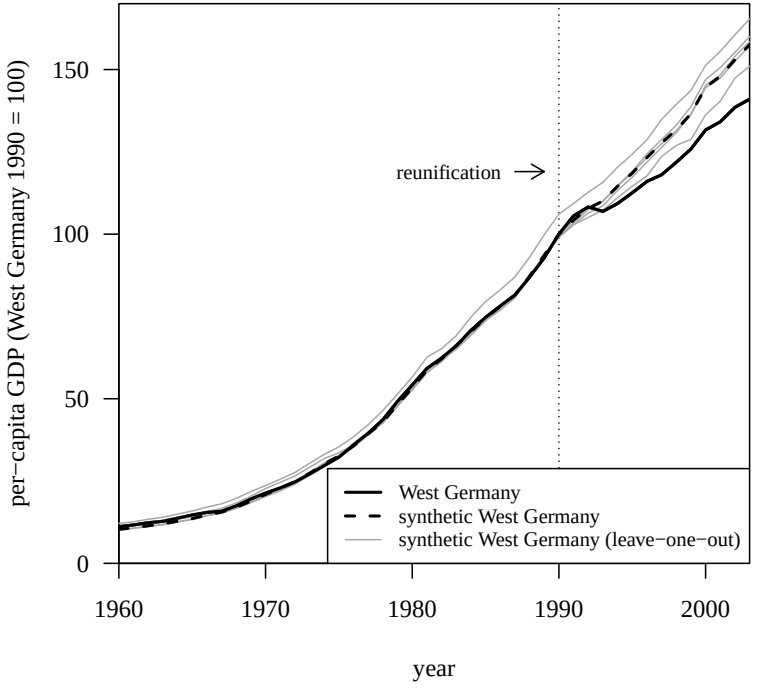



*Figure 6*

Ratio of Postreunification RMSPE to Prereunification RMSPE: West Germany and Control Countries

Note: See Figure 3 for details.

(a) Abadie *et al.*




(b) Constant-price replication

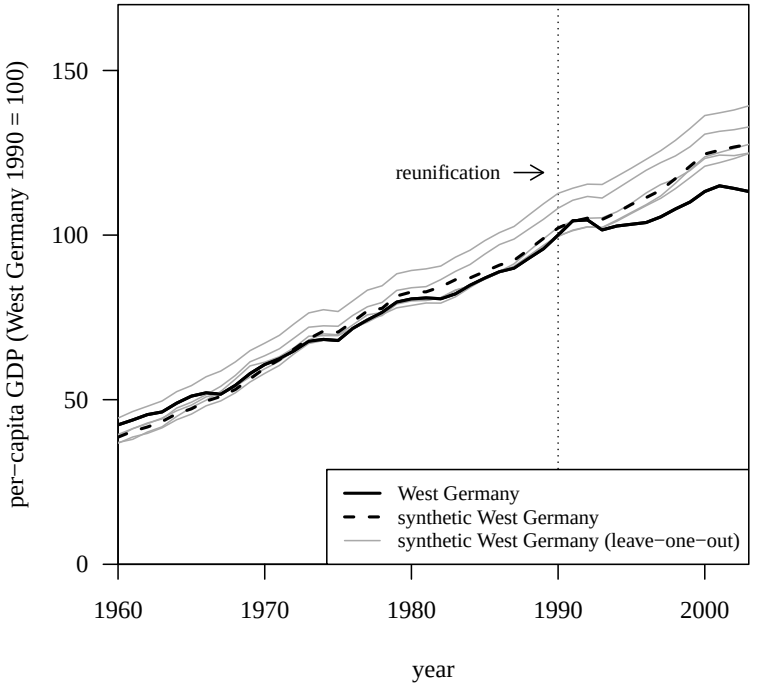



*Figure 7*

Leave-One-Out Distribution of the Synthetic Control for West Germany

Note: See Figure 3 for details.

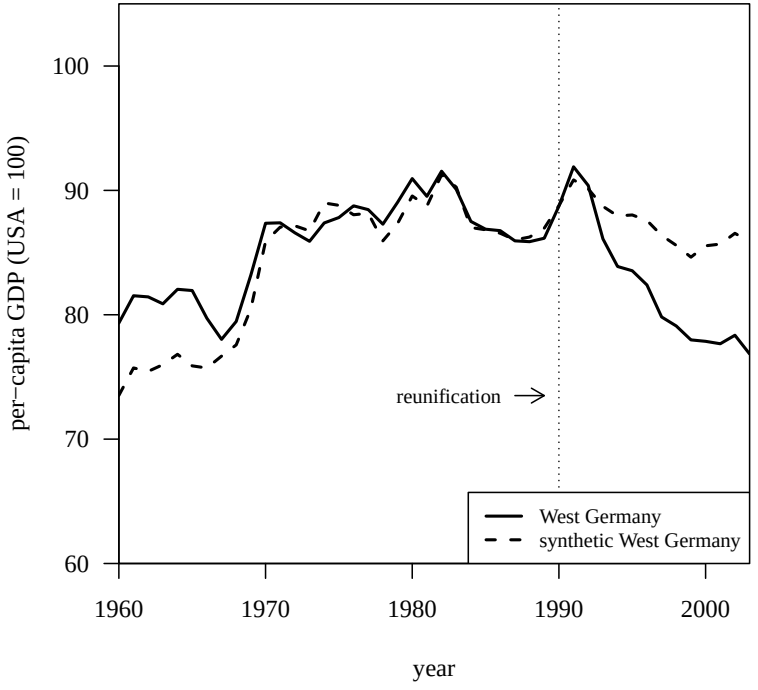



*Figure 8*

Trends in per Capita GDP: West Germany versus Synthetic West Germany (USA=100)

Notes: The GDP per capita series have been normalized so that USA in 1990 equals 100. Calculated from data underlying Abadie *et al.* (2015).


*Figure 9*

Per Capita GDP Gap between West Germany and Synthetic West Germany (USA=100)

Note: See left figure for details.

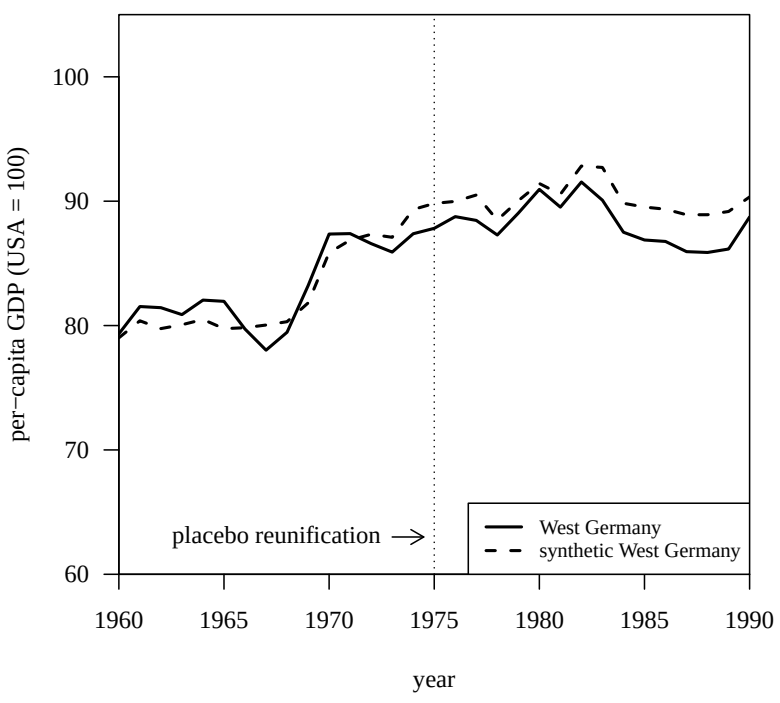



*Figure 10*
Placebo Reunification 1975–Trends in per Capita GDP: West Germany versus Synthetic West Germany (USA=100)

Notes: See above figure for details.

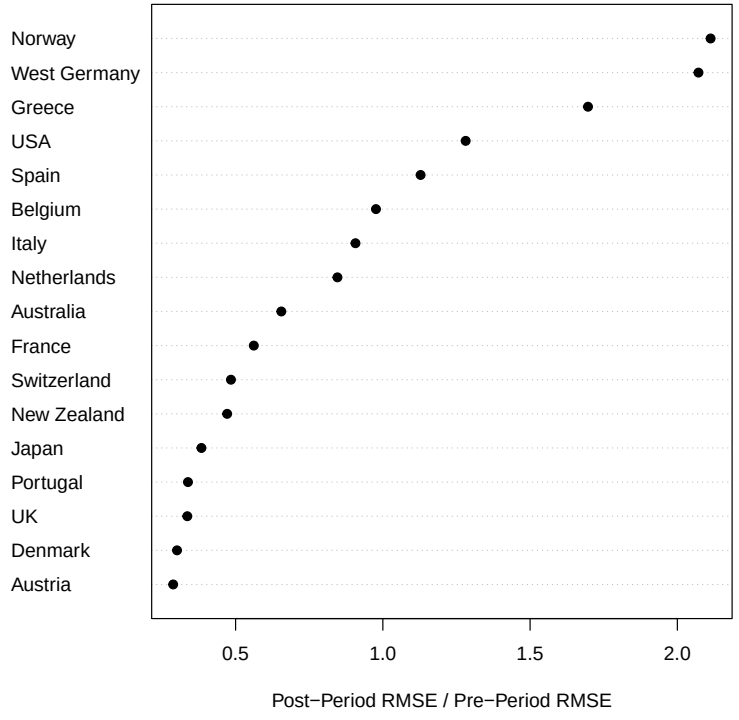



*Figure 11*
Ratio of Postreunification RMSPE to Prereunification RMSPE: West Germany and Control Countries (USA=100)

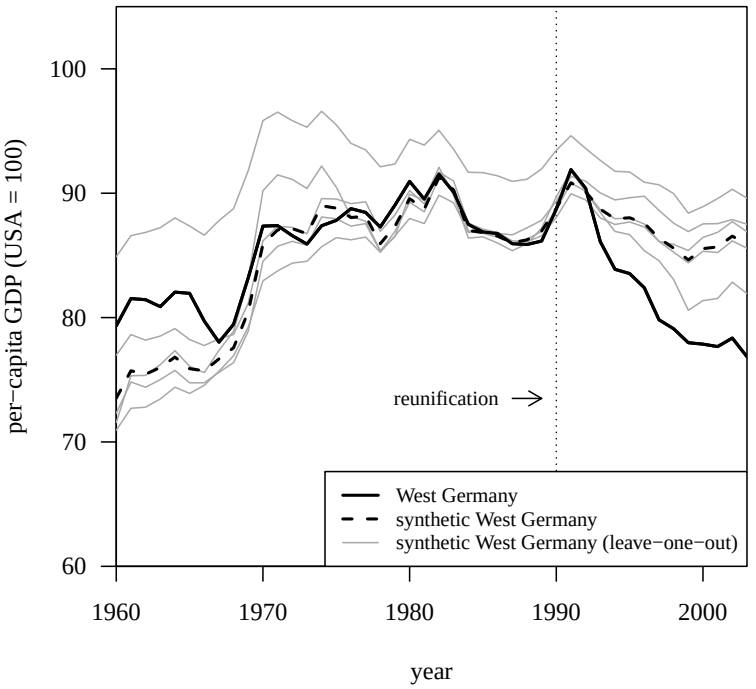



*Figure 12*
Leave-One-Out Distribution of the Synthetic Control for West Germany (USA=100)

Note: See first figure for details.

## References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2015) "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science*. 59(2), 495–510.

Bolt, Jutta, Robert Inklaar, Herman de Jong, and Jan Luiten van Zanden (2018) "Rebasing 'Maddison': New Income Comparisons and the Shape of Long-Run Economic Development." Unpublished paper, GGDC Research Memorandum 174, Jan. https://www.rug.nl/ggdc/html_publications/memorandum/gd174.pdf.

Bolt, Jutta and Jan Luiten Van Zanden (2020) "Maddison style Estimates of the Evolution of the World Economy: A New 2020 Update." Unpublished paper, Oct. https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020.

— (2025) "Maddison style Estimates of the Evolution of the World Economy: A New 2023 Update." *Journal of Economic Surveys*. 39(2), 631–671.

Deaton, Angus and Alan Heston (2010) "Understanding PPPs and PPP-based National Accounts." *American Economic Journal: Macroeconomics*. 2(4), 1–35.

Ferman, Bruno, Cristine Pinto, and Vitor Possebom (2020) "Cherry Picking with Synthetic Controls." *Journal of Policy Analysis and Management*. 39(2), 510–532.

Francis, Joseph (2025) "A *p*-Hackers Guide to the Synthetic Control Method." Unpublished paper, v1.0, June 2. https://github.com/joefrancis505/Francis_A_p-Hackers_Guide.