A Review of "Comparative Politics and the Synthetic Control Method" by Abadie et al. (2015)

Reviewer 2

January 14, 2026

v1

isitcredible.com

Disclaimer

This report was generated by Reviewer 2, an automated system that uses large language models to assess academic texts. It has been read and approved by a human editor on behalf of The Catalogue of Errors Ltd. The report's goal is to facilitate the discovery of knowledge by identifying errors in the existing literature. Comments can be made here. Any errors will be corrected in future revisions.

> I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.
>
> Plato, *The Apology of Socrates*, 21d

> To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.
>
> Karl Popper, 'Knowledge and the Shaping of Reality'

Overview

Citation: Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, Vol. 59, No. 2, pp. 495–510.

URL: https://onlinelibrary.wiley.com/doi/full/10.1111/ajps.12116

Abstract Summary: This article discusses the synthetic control method as a systematic way to choose comparison units in comparative case studies, bridging the quantitative/qualitative divide and enabling precise quantitative inference in small samples. The authors illustrate the method by estimating the pronounced negative economic impact of the 1990 German reunification on West Germany.

Key Methodology: Synthetic Control Method (SCM) applied to panel data, using cross-validation for weight selection and placebo studies (in-time and in-space) for inference and robustness checks.

Research Question: How can the synthetic control method be used to bridge the quantitative/qualitative divide in comparative politics, and what was the economic impact of the 1990 German reunification on West Germany?

Summary

Is It Credible?

The contribution made by Abadie et al. is methodologically significant and empirically grounded, offering a credible advancement in the toolkit of comparative politics. The article successfully argues that the Synthetic Control Method (SCM) provides a more transparent and defensible alternative to standard regression for comparative case studies. By restricting weights to the zero-to-one interval, the authors convincingly demonstrate how SCM avoids the extrapolation biases that plague regression analyses, where negative weights can produce unintelligible counterfactuals. The formalization of selecting comparison units—moving from ad hoc qualitative selection to a data-driven "synthetic" combination—is a genuine improvement in rigor. Yet, while the method itself is robust, the specific empirical application regarding the costs of German reunification reveals the limitations of the "precise quantitative inference" the authors claim to deliver.

The primary credibility gap lies in the precision of the estimated economic impact of reunification. The authors report a reduction in West German per capita GDP of approximately US$1,600 per year. While the *direction* of this effect appears robust, the *magnitude* is highly sensitive to the composition of the donor pool. As the authors' own robustness checks reveal, excluding the United States from the donor pool causes the estimated effect to plummet to approximately US$630—a reduction of roughly 60 percent. This instability suggests that the specific quantification of the cost depends heavily on the inclusion of a single, highly influential comparator. While the authors argue that the remaining effect is still "substantive," the difference between a US$1,600 loss and a US$630 loss is economically vast. This sensitivity undermines the claim that the method yields precise point estimates in small-sample contexts, indicating instead that it provides a credible sign (negative) but a fragile magnitude.

This fragility is compounded by the issue of spillover effects, which the article acknowledges but does not empirically resolve. The synthetic control relies heavily on Austria, the Netherlands, and Switzerland—countries with deep economic integration with Germany. The au-

thors argue that if reunification harmed these neighbors (negative spillovers), the estimated effect on Germany is conservative. They do not, however, rigorously rule out the possibility of positive spillovers, such as a demand boom for neighbors' exports driven by German reconstruction. If the control units benefited from the treatment, the synthetic counterfactual would be artificially inflated, exaggerating the negative gap observed for West Germany. Given the reliance on trade partners in the donor pool, the assumption of no interference between units is a strong one that remains theoretically defended rather than empirically tested.

Ultimately, the article succeeds in its primary goal: establishing a systematic bridge between qualitative and quantitative approaches. The SCM forces researchers to be explicit about the contribution of each comparison unit, allowing for the kind of "qualitative flesh on quantitative bones" the authors promise. The limitations found in the German application—sensitivity to unit selection and potential spillovers—are inherent to the difficulty of small-sample inference in an interconnected world. They do not invalidate the method, but they do suggest that the specific numerical estimates produced by SCM should be interpreted as indicative ranges rather than precise calculations. The article is a credible and valuable methodological contribution, even if the certainty of its specific empirical finding is overstated.

The Bottom Line

This article presents a highly credible methodological innovation that improves upon traditional regression by preventing extrapolation and systematizing case selection. The Synthetic Control Method is a rigorous tool for constructing counterfactuals in comparative politics. However, the specific empirical claim—that German reunification cost West Germany US$1,600 per capita annually—is less definitive than presented. The estimate is highly sensitive to the inclusion of the United States in the comparison group and potentially biased by unmeasured economic spillovers to Germany's neighbors. Readers should accept the method as a new standard for the field but view the specific dollar-value cost of reunification as a likely upper bound rather than a precise fact.

Specific Issues

Sensitivity of the estimated effect magnitude: The precision of the main empirical finding is heavily dependent on the specific composition of the donor pool. The authors report a baseline average reduction of US$1,600 per capita GDP per year (p. 504). However, the "leave-one-out" robustness check reveals that excluding the United States—which carries a significant 22 percent weight—reduces the estimated effect to approximately US$630 per year (p. 506). This represents a decline of roughly 60 percent in the magnitude of the central finding. While the authors defend this by noting the effect remains "fairly large in substantive terms" (p. 506), such high sensitivity to a single unit suggests that the specific point estimate of US$1,600 is fragile and heavily contingent on the specific mix of the donor pool.

Risk of unaddressed spillover effects: The validity of the causal estimate is threatened by the potential for spillover effects among the highly integrated European economies in the donor pool. The authors acknowledge that reunification could impact countries like Austria, the Netherlands, and Switzerland (p. 504). While they argue that negative spillovers would bias the results conservatively, they do not empirically rule out positive spillovers—such as increased export demand from a reunifying Germany—which would inflate the counterfactual and exaggerate the estimated negative effect on West Germany. Given that Austria (weight 0.42) and the Netherlands (weight 0.09) are major trading partners, the assumption of independence is precarious.

Overstatement of inferential precision: The article claims that the method "opens the door to precise quantitative inference" (p. 495). However, the combination of the sensitivity issue noted above and the nature of the statistical tests suggests this claim is overstated. The inference relies on a single p-value of 0.059 derived from a small, non-random sample (p. 505), and the authors themselves admit that "traditional approaches to statistical inference" are difficult in this context (p. 499). The method provides a rigorous framework for falsification, but characterizing the resulting estimates as "precise" contradicts the demonstrated volatility of the magnitude.

Data and methodological limitations: There are several minor issues related to data con-

straints and procedural simplifications. The statistical inference relies on a p-value constructed from the rank of the RMSPE ratio, which is descriptive and limited by the small sample size of 17 countries (p. 505). Furthermore, there is a temporal inconsistency in the predictor data; while most variables are averaged over 1981–1990, the investment rate and schooling data rely on averages from 1980–1985 due to data availability (pp. 502, 509). Additionally, the authors admit that the synthetic control fails to match the pre-reunification inflation rate because West Germany's rate was lower than any country in the donor pool (p. 503). In the sparse control robustness check, the authors simplified the procedure by fixing the predictor weights rather than re-optimizing them, potentially underestimating the fit of the sparse controls (p. 506, fn. 20). Finally, the data source "Statistisches Bundesamt 2005" is cited in the Appendix but is missing from the References section (p. 509).

Future Research

Robustness through donor pool averaging: To address the sensitivity of the magnitude to single units like the United States, future research should develop a method for averaging across multiple valid synthetic controls. Rather than relying on a single "best" synthetic control vector, researchers could estimate a distribution of effects derived from iteratively dropping units or using different combinations of donors that meet a minimum fit threshold. This would produce a confidence interval for the magnitude of the effect, providing a more honest assessment of the uncertainty inherent in the donor pool composition.

Modeling spillover effects in trade networks: To resolve the ambiguity regarding spillovers, future research should explicitly model the economic linkages between the treated unit and the donor pool. By incorporating a spatial or network-based weight matrix (e.g., based on trade volume or geographic distance), researchers could adjust the synthetic control estimator to account for the likely propagation of shocks. This would allow for the empirical testing of whether spillovers are positive or negative, rather than relying on theoretical assumptions to sign the bias.