A *p*-Hacker's Guide to the Synthetic Control Method

Joseph Francis

June 2, 2025
(v1.0)

Abstract

The Synthetic Control Method (SCM) can easily be *p*-hacked by exploiting the significant researcher degrees of freedom in the construction of the donor pool. As recently shown in the *Journal of Political Economy* by Magness and Makovi (2023), the *p*-hacking does not even need to be intentional. It is necessary—at a minimum—to revise Abadie's (2021) guide by introducing stronger guardrails around donor pool selection. And even that may ultimately be insufficient, given the incentives economists face today and the dysfunction in academic publishing.

Keywords: economics, *p*-hacking, synthetic control method (SCM)

JEL codes: C12, C22, C52

In a recent article in the *Journal of Political Economy*, Phillip W. Magness and Michael Makovi (2023) illustrate how easily the Synthetic Control Method (SCM) can be *p*-hacked. Probably without realizing it, Magness and Makovi have exploited the considerable researcher degrees of freedom in the construction of donor pools. According to Alberto Abadie (2021), donors should only be included in the donor pool if they are similar to the treated unit in both their observed (Z_j) and unobserved (μ_j) attributes. Abadie's (2021, 401) warning is clear: "Including in the donor pool units that are regarded by the analyst to be unsuitable controls (because of large discrepancies in the values of their observed attributes Z_j or because of suspected large differences in the values of the unobserved attributes μ_j relative to the treated unit) is a recipe for bias." The problem, however, is

---

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. Supplementary materials are available at https://github.com/joefrancis505/Francis_A_p-Hackers_Guide.

that Abadie's guide to the SCM provides no objective criteria for determining whether a donor is suitable or not, such as what "large discrepancies" in Z_j may mean in practice. It is left to subjective opinion; whatever is "regarded by the analyst" as suitable is fine, leading to considerable researcher degrees of freedom. Permissive guidelines can then combine with misunderstandings of causal inference to result in research designs that are unintentionally p-hacked from the beginning. Such studies can, moreover, pass peer review by a "top five" journal—supposedly the gold standard for truth claims in economics.[1]

Magness and Makovi's (2023) article provides an introduction to the problem. The treated unit is Karl Marx, who Magness and Makovi (2023, 1509) describe as "an occasionally acknowledged but relatively minor figure between his death and the events of 1917." Following Abadie's (2021) guide, it would therefore appear sensible to include similarly minor figures in a donor pool designed to test the effects of the Russian Revolution on Marx's intellectual influence, which Magness and Makovi (2023) proxy by his share of n-grams in the Google Ngram Viewer. Similarly minor figures would, it can be assumed, be close to Marx in terms of both Z_j and μ_j. Nonetheless, without any explanation why, Magness and Makovi's (2023, 1520–1521, Table 1) donor pool includes Abraham Lincoln, the world-famous American president; Adam Smith, the world's most famous economist; Aeschylus, the ancient Greek tragedian; Aesop, the ancient Greek fabulist and storyteller; Alessandro Manzoni, the widely known Italian poet and philosopher; Alexander Hamilton, an American Founding Father; and so on. These are the various names whom Magness and Makovi have subjectively decided resembled Marx in terms of Z_j and μ_j in the pretreatment period.

The stated criteria for choosing these names could easily be perceived as incoherent. Magness and Makovi (2023, 1519) note that they "brainstormed a list of relevant economic, sociopolitical, and socialist thinkers up to the time of Marx's death," added some more from "two primary-source readers in political philosophy," then "almost all authors from the first 39 volumes of the 50-volume *Harvard Classics*," an early twentieth-

[1] As Stefan and Schönbrodt (2023, 1) put it, p-hacking is typically defined as "a compound of strategies targeted at rendering non-significant hypothesis testing results significant." This paper's p-hacker's guide to the SCM is different from the one provided by Ferman et al. (2020), which focuses on how researchers can cherry-pick predictor variables to achieve a good pre-treatment fit between the control and the treated unit. Neither Abadie's (2021) original guide nor the updated but unpublished guidelines in Abadie and Vives-i-Bastida (2021) address either issue.

century anthology of famous authors and world-historical figures, and finally some more from unspecified "German-language encyclopedic anthologies," arriving at a list of 227 donors. Yet beyond this description of what they did, there is minimal explanation of why they did it. And the criteria that are offered are questionable. Most importantly, there are the numerous famous names taken from the *Harvard Classics*, which Magness and Makovi (2023, 1519n20) consider "a rough approximation of the common intellectual canon at the turn of the century," while noting that "Marx was not included among its volumes." Hence, by definition, these donors do not seem similar to Marx in terms of Z_j, given that they were canonical and he was not. Furthermore, Magness and Makovi's stated criteria for choosing full names or surnames are contradicted by their actual list of donors. "For every author," Magness and Makovi (Magness and Makovi, 1525) write, "we selected the most easily identifiable iteration that was least likely to be conflated with that of another famous individual." Yet they include "Hume" and not "David Hume," "Newton" instead of "Isaac Newton," "Kafka" instead of "Franz Kafka," with numerous other instances, as well.[2] For some, then, Magness and Makovi's description of the donor pool selection might seem nonsensical. But there is no way to demonstrate this objectively because there are no guardrails on donor pool selection in Abadie's (2021) guide. It offers no objective criteria to ensure that both Z_j and μ_j align between the treated unit and the donors. Rather, it is simply a question of opinion and personal choice.[3]

---

[2] As Francis (2025b, 97–99) notes, the inclusion of the surname "Kafka" is particularly odd because Magness and Makovi's (2023, 1519) stated criteria for the German-language writers is to include those "whose lives preceded or overlapped Marx's," yet Franz Kafka was born 111 days after Marx died.

[3] Magness and Makovi (2023, Online Appendices (A29)) seem to half-recognize that their donor pool is problematic when they run an SCM for the surname "Marx." The result gives a large weight in the control to the surname "Nietzsche," which leads them to question its validity because "there is no clear reason why Nietzsche should be such an important counterfactual for Marx." The next step would be to ask why the surname "Nietzsche" is included in the donor pool at all if they regard it as an unsuitable control, as explained by Francis (2025b, 93). But this step is not taken, and Magness and Makovi (2025, 121–122n13) entirely miss the point in their response. They appear to believe that the synth package's ability to "achieve adequate indicator balance" in Stata as a sign that a donor pool has been well-designed. Nonetheless, Abadie (2021) warns that "interpolation biases may still be important if the synthetic control matches the characteristics of the affected unit by averaging away large discrepancies between the characteristics of the affected unit and the characteristics of the units in the synthetic control," such as by combining a whale like Abraham Lincoln with a minnow like Ferdinand Lassalle, as Francis (2024, 370–375) discusses

Given this lack of guardrails, p-hacking becomes so easy that researchers may not even be aware that they are doing it. To understand why, it is necessary to look at how causal inference works in an SCM. As Abadie (2021, 403–405) explains, causation can be inferred from permutation tests. Placebo effects are estimated for each unit in the donor pool by iteratively reassigning the treatment to them and constructing a synthetic control for each. The estimated effect for the actual treated unit is then compared to the distribution of these placebo effects. If the effect for the treated unit is larger in magnitude than most placebo effects, it is considered significant. Hence, a p-value is calculated as:

$$p = \frac{1}{J+1} \sum_{j=1}^{J+1} I_+ \left(r_j - r_1 \right) \tag{1}$$

where p is the proportion of units in the entire sample—the treated unit plus all donors (J)—for which the calculated ratio of post-intervention Root Mean Squared Prediction Error (RMSPE) to pre-intervention RMSPE (r_j) is greater than or equal to the same ratio observed for the actual treated unit (r_1). Why the donor pool construction matters, then, is because all donors are included in the denominator of the p-value calculation. And the lack of guardrails around donor selection means that anyone can p-hack an SCM by including many donors who are unlikely to have a treatment effect.

For this reason, Magness and Makovi (2023) were able to p-hack their SCM from the beginning—probably without realizing what they were doing. They sought to explain why a "minor figure" like Marx became influential. Only the Russian Revolution made his ideas important, they claim. Yet to test this hypothesis, Magness and Makovi (2023, 1520–1521, Table 1) construct a donor pool that mainly consists of massively famous figures, without saying why they resemble Karl Marx in terms of either Z_j or μ_j. And if any other researcher suggests that these donors are overwhelmingly unsuitable, it does not matter because, as Abadie (2021, 401)

---

for Magness and Makovi's (2023) headline Synthetic Marx. The response by Magness and Makovi (2025, 121) is to point towards their "socialists-only" Synthetic Marx in their article's Online Appendices, and then rhetorically ask "if our original paper's headline, main result had been our "socialist-only" test, and if all our other tests had been relegated to the robustness sections or the Appendix, would Francis have ever written any response for *Econ Journal Watch* at all?" Obviously, from Francis' perspective, it would have been better to try to build a more sensible donor pool and use that as the headline result. There are, however, similar problems with the donor pool in the "socialists-only" Synthetic Marx, as will be discussed below.

puts it, whatever is "regarded by the analyst" as suitable is fine. Given the lack of objective criteria for donor pool selection, all Magness and Makovi's (2023) choices are permissible. They are free to add donors who some would say were extremely dissimilar to their own description of Marx and—crucially—were extremely unlikely to have been affected by the Russian Revolution. In doing so, Magness and Makovi mechanically lower their *p*-values by packing the denominator with placebos that have a low r_j in the permutation tests. Inadvertently, then, they have revealed how open the SCM methodology is to *p*-hacking.

Specification searching is also easy. As Francis (2025b, 96–97) shows, *p*-values can be manipulated to produce the desired results. His replication of Magness and Makovi's (2020) original working paper returned p=0.01, using their initial pool of 97 donors.[4] Francis then found that when 45 donors were added from "German-language anthologies," as in Magness and Makovi's (2025) final article, the permutation test returned p=0.064. It was only by adding another 49 donors from the *Harvard Classics* anthology of famous authors and historical figures that p fell back to 0.046, which is close to Magness and Makovi's (2023, 1530, Table 4) final headline figure of p=0.047. It is easy to speculate, then, that peer reviewers may have insisted that more German donors be included to make the SCM more realistic. But when that happened, the headline result returned p>0.05. Only adding more famous authors and world historical figures from the *Harvard Classics* to dilute the donor pool produced p<0.05, thereby allowing the headline result to meet the conventional level of statistical significance. That such a chain of events may have taken place is, of course,pure speculation, but its very plausibility is a symptom of there being no objective criteria for why donors were included or not. What made Magness and Makovi (2023) add more donors from the *Harvard Classics* is unstated. The entire research design is a black box, but it is permissible because, according to Abadie (2021), which donors should be included in the SCM is based on the subjective opinion of the researcher. Essentially, any choice is permissible.[5]

---

[4] The replication reported here is slightly different from that of Francis (2025b, 96–97) because he incorrectly left out Karl Marx from the donor pool when calculating the placebos. Marx has now been added back in, which reinforces the pattern described.

[5] In their final article, Magness and Makovi (2023) also introduced unconventional *p*-values that were not present in their working paper, which only relied on the permutation tests shown in Equation 1. As Francis (2025b, 96n4) notes, their use of these unconventional *p*-values is problematic. Magness and Makovi's (2025, 113, fn4) response is that they are aware of the issues in the "asymptotically exact p-value (AEP)," but use it anyway, while also reporting the Simes *p*-value. Both the AEP and

Supposed robustness tests are then undermined by the initial issues in the research design. Magness and Makovi report, for example, a "socialists-only" Synthetic Marx that appears to address some of the issues related to their choice of donors. By limiting the pool to only the 19 donors whom they code as socialists, it seems to make the donor pool more like Marx in terms of Z_j. It returns, moreover, p=0.[6] Yet a replication of this SCM soon reveals questionable choices. The initial run using 19 donors, for instance, depends heavily on Oscar Wilde, whose biography as a playwright and high-society dilettante raises questions over whether he can really be considered a socialist comparable to Marx. With Wilde excluded, the SCM then turns to the surname "Sombart," which Magness and Makovi decided to include instead of the full name Werner Sombart, without, of course, explaining why. Sombart is also a questionable choice because he died in 1941, almost six decades after Marx. As such, some analysts would see his inclusion as evidence of what can be called "life-cycle bias," given that there is a common tendency for fame to increase until death and diminish thereafter (Francis 2024, 372–374; 2025b, 94–95). And this is a more general problem in the sample of 19 donors whom Magness and Makovi have subjectively decided were socialists. Nine of them lived after Marx died, and the SCM tends to rely heavily on those particular donors. More fundamentally, still, the results of the SCM could easily be changed by adding other donors or taking some away, coding or uncoding them as socialist, using surnames or full names. All these decisions are subjective, which makes the meaningfulness of Magness and Makovi's robustness test a simple matter of opinion.[7]

Simes *p*-values are nevertheless unconventional in the context of an SCM. Indeed, their use seems like a clear instance of multiple testing—another kind of *p*-hacking. Again, however, it must be stressed that Magness and Makovi probably do not know what they are doing, making the *p*-hacking unintentional.

[6] This is possible because Magness and Makovi use the *synth runner* package, which follows Cavallo *et al.* (2013) in removing the treated unit from the denominator, rather than Abadie's (2021) more conservative approach. All the replicated *p*-values reported in this paper have been calculated in the same way as in Magness and Makovi. They would be higher if Marx were included in the denominator.

[7] The "socialists-only" Synthetic Marx is analyzed in the R script accompanying this paper. Once the surname Sombart is subtracted, the SCM depends on August Bebel, the German Social Democrat politician, who died in 1913, shortly before the Russian Revolution, again suggesting lifecycle bias. When Bebel is excluded, it turns to Henry George, the famous American popular economist who died in 1897. In this case, the life-cycle bias is somewhat attenuated, but other questions can be asked, such as whether George is a suitable control for a "minor figure" like Marx. Similarly, if

Whether intentionally or not, then, Magness and Makovi (2023) have exploited weaknesses in Abadie's (2021) guide to the SCM. The subjective approach to donor selection makes possible the construction of a large pool of donors who are unlikely to have a treatment effect, thereby predetermining the SCM's results. And having seen this approach legitimized by the *Journal of Political Economy*, any unscrupulous *p*-hacker can now use it to produce a SCM that provides the results that they desire, including for topics far more consequential than the Russian Revolution's effect on Karl Marx's n-gram share, with an article published in a "top five" economics journal cited as precedent. Consequently, Abadie's guide needs to be revised. Going forward, the Synthetic Marx affair suggests, for instance, that the SCM is more appropriate for cases where the donors are contemporaneous with the unit and the pool is naturally bounded. Such is, for example, the case for Spanish regions or American states, as in the classic SCM studies (Abadie and Gardeazabal 2003; Abadie, Diamond, et al. 2010). Indeed, those studies imply that much stricter criteria ought to be applied to donor pool construction than is allowed by Abadie's (2021, 401) more *laissez faire* approach. Constructing the donor pool based on whatever is "regarded by the analyst" to be suitable is not enough. A better guide with stronger guardrails is therefore required to prevent the SCM being *p*-hacked in the future.

At a minimum, the new guide to the SCM must not be naive. The root cause of the weaknesses in Abadie's (2021, 422) guide seems to be the assumption that researchers are competent and honest and will, as a result, heed his warning "that mechanical applications of synthetic controls that do not take into account the context of the investigation or the nature of the data are risky enterprises." Magness and Makovi (2023) have nevertheless shown that the rewards are sufficient that researchers will still take such risks. For young researchers, in particular, the temptation to do so is great, given the "tyranny of the top five"—that is, the need to publish in the five most prestigious economics journals, including the *Journal of Political Economy*, to achieve tenure (Heckman and Moktan 2020). Worryingly, making strong causal claims using econometric techniques, such as the SCM, is increasingly becoming a prerequisite for publication in those journals (Garg and Fetzer 2025). As a result, the risk-reward calculus suggests that *p*-hacking will continue, as long as there are not

George is excluded, the SCM relies on "Kropotkin," the Russian anarchist who died in 1921, and John Ruskin, the famous English polymath and philanthropist, who died in 1900.

stronger guardrails against it—not only in the SCM but in all econometric techniques.[8]

For this reason, the Synthetic Marx affair should be considered the result of a broader dysfunction in academic economics. There is evidence that *p*-hacking is still rife, despite the increased awareness of the issue (Brodeur et al. 2020). The chances of successfully publishing *p*-hacked research is, moreover, considerable because a broken peer review system makes publication like a lottery (Mastroianni 2022). Editors, moreover, have considerable influence, which means that a problematic paper's chances of being published will be greatly increased if they can be won over. Aggravating the situation, the editors of the "top five" then have strong incentives not to retract articles after publication. Hence, in the words of a recent article in *The Economist*, compared to problems of incompetence and misconduct in other academic disciplines, "[e]conomics has some protection owing to its record. The five leading journals have seen just four withdrawals in their combined 570-year history" (Anon. 2025). There are thus major incentives both for researchers to engage in *p*-hacking and for journals to ignore it after it has been identified. It is unlikely, for example, that the *Journal of Political Economy* will ever retract Magness and Makovi's (2023) article, despite its obvious flaws. To do so would make academic economics seem dysfunctional. But it is, which is why—at a minimum—stronger guardrails are needed.

References

Abadie, Alberto (2021) "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature*. 59(2), 391–425.

Abadie, Alberto, Alexis Diamond, and Jens 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.

[8] Worryingly, it may be that the author of the existing guide may not fully appreciate how research design affects causal inference. Abadie, Diamond and Hainmueller (2015, 502n14) look at the effects of reunification on West Germany's GDP per capita after reunification in 1990, but they exclude Canada, Finland, and Sweden from the donor pool because they experienced profound financial and fiscal crises at the beginning of the 1990s." In this way, they seem to arbitrarily remove donors who are likely to show large placebo effects, mechanically lowering their *p*-value. To justify their choice, they say that it makes little difference to their identification strategy, without appreciating how it affects causal inference. There is, in any case, a larger problem with that study: Abadie et al. accidentally used current price data for GDP rather than constant prices, invalidating their results, as discussed in Francis (2025a).

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

Abadie, Alberto and Javier Gardeazabal (2003) "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review*. 93(1), 113–132.

Abadie, Alberto and Jaume Vives-i-Bastida (2021) "Synthetic Controls in Action." Unpublished paper, Sept. 17.

Anon. (2025) "Sine Laude: What the Failure of a Superstar Student Reveals about Economics." *The Economist*, May 22. https://www.economist.com/finance-and-economics/2025/05/22/what-the-failure-of-a-superstar-student-reveals-about-economics.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes (2020) "Methods Matter: *p*-Hacking and Publication Bias in Causal Analysis in Economics." *American Economic Review*. 110(11), 3634–3660.

Cavallo, Eduardo, Sebastian Galiani, Han Noy, and Juan Pantano (2013) "Catastropic Natural Disasters and Economic Growth." *Review of Economics and Statistics*. 95(5), 1549–1561.

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 (2024) "Should We Quantify Karl Marx?" *Econ Journal Watch*. 21(2), 357–384.

— (2025a) "Comparative Politics and the Synthetic Control Method Revisited." Unpublished paper, v1.0, June 2. https://github.com/joefrancis505/Francis_Comparative_politics.

— (2025b) "From Synthetic Marx to Synthetic Kafka: A Rejoinder to Magness and Makovi." *Econ Journal Watch*. 22(1), 91–107.

Garg, Prashant and Thiemo Fetzer (2025) "Causal Claims in Economics." Unpublished paper, Jan. 12. http://arxiv.org/abs/2501.06873.

Heckman, James J. and Sidharth Moktan (2020) "Publishing and Promotion in Economics: The Tyranny of the Top Five." *Journal of Economic Literature*. 58(2), 419–470.

Magness, Phillip W. and Michael Makovi (2020) "The Mainstreaming of Marx: Measuring the Effect of the Russian Revolution on Karl Marx's Influence." Unpublished paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3578840.

— (2023) "The Mainstreaming of Marx: Measuring the Effect of the Russian Revolution on Karl Marx's Influence." *Journal of Political Economy*. 131(6), 1507–1545.

— (2025) "Synthetic Karl Marx and His Clumsy Critic." *Econ Journal Watch*. 22(1), 108–141.

Mastroianni, Adam (2022) "The Rise and Fall of Peer Review: Why the Greatest Scientific Experiment in History Failed, and Why That's a Great Thing." Blogpost, Dec. 13. https://www.experimental-history.com/p/the-rise-and-fall-of-peer-review.

Stefan, Angelika M. and Felix D. Schönbrodt (2023) "Big Little Lies: A Compendium and Simulation of *p*-Hacking Strategies." *Royal Society Open Science*. 10(220346).