INSTITUTE for REPLICATION

Reproduction and Replication Analyses of Andersson (2019): A Replication Report from the Toronto Replication Games

Carl Bonander

Niklas Jakobsson

Naimi Johansson

I4R DISCUSSION PAPER SERIES

I4R DP No. 26

Reproduction and Replication Analyses of Andersson (2019): A Replication Report from the Toronto Replication Games

Carl Bonander[1], Niklas Jakobsson[2], Naimi Johansson[3]

[1]*University of Gothenburg/Sweden*
[2]*Karlstad University/Sweden*
[3]*Region Örebro and Örebro University/Sweden*

MARCH 2023

Reproduction and replication analyses of Andersson (2019): A replication report from the Toronto Replication Games

Carl Bonander (University of Gothenburg), Niklas Jakobsson (Karlstad University), and Naimi Johansson (Region Örebro and Örebro University)*

Abstract

This report presents a replication of Andersson (2019) performed at the Toronto Replication Games in 2023. Andersson (2019) estimates the effect of carbon taxes on CO_2 emissions in Sweden using the synthetic control method. His findings indicate a 10.9 percent reduction in emissions during the 1990-2005 period, which equates to -0.29 metric tons of CO_2 per capita in an average year. The results from an in-space placebo test show that Sweden had the highest post/pre-mean squared prediction error (MSPE) ratio, resulting in a placebo-based p-value of 1/15=0.067. We successfully reproduce these findings and conduct a series of pre-specified replication analyses to examine how robust the findings are to model specification choices. We run 14 alternative specifications with various combinations of pre-treatment outcome values, with and without covariates. The median point estimate from our replication analyses is -0.28 metric tons of CO_2 per capita (min: -0.34, max: -0.17). Placebo-based p-values are equal to 1/15=0.067 in seven specifications, 2/15=0.13 in six, and 4/15=0.27 in one.

1. Introduction

In 1991, Sweden implemented a carbon tax on transport fuel emissions, as one of the first countries in the world. Already in 1990, the value-added tax (VAT) was extended to include gasoline and diesel. Andersson (2019) estimates the causal effect of introducing the carbon tax on carbon dioxide (CO_2) emissions from transports, using the synthetic control method. The data is an annual panel for 25 OECD countries over the years 1960–2005 of CO_2 transport emissions per capita, originally retrieved from the World Bank. Covariates include GDP per capita, number of motor vehicles per 1000 people, gasoline consumption per capita, and share of urban population.

Andersson's results show a reduction in emissions of 10.9 percent, or –0.29 metric tons CO_2 per capita, in an average year 1990–2005. Placebo-tests are conducted "in-time", "in-space", "leave-on-out", and "full sample" (p. 15). The results from the "in-space" placebo test show that Sweden has by far the largest ratio of post-treatment mean squared prediction error (MSPE) and pre-treatment MSPE, resulting in a placebo-based p-value of 1/15 = 0.067 (rank 1 out of 15).

In the present paper, we investigate whether Andersson's results are reproducible and test the robustness replicability using the 14 alternative synthetic control specifications suggested by Ferman et al. (2020). The argument by Ferman et al. (2020) is that the lack of guidance on how to specify the variables to include when implementing the synthetic control method gives opportunities for specification-searching, and thus suggests 14 specifications that should be implemented to assess the sensitivity of the results. Andersson's replication package and data were downloaded from the American Economic Association.[1] We successfully reproduce Andersson's findings. Our results of robustness replicability (14 specifications) yield a median point estimate of –0.28 metric tons of CO_2 per capita (min: –0.34, max: –0.17), which is very close to the original estimate. Placebo-based p-values are equal to 1/15=0.067 in seven specifications, 2/15=0.13 in six, and 4/15=0.27 in one.

---

[1] https://www.aeaweb.org/journals/dataset?id=10.1257/pol.20170144

The rest of this report is structured as follows: Section 2 details the results from an assessment of the R code in the study's replication package, Section 3 presents the replication results, and Section 4 concludes.

2. Assessment of replication code

We identified only one minor coding error in the study's replication package while reproducing the study. The 'synth' command in the replication package contains an undefined option called 'method="All"', which we believe is intended to allow the algorithm to select the best-performing optimizer among a set of alternative algorithms implemented in the 'optimx' package for R. The correct option in the current version of 'synth' (1.1.6) is 'optimxmethod="All"'. Since the option method is undefined in synth, it is passed to the "…" part of the function, which is typically used in R programming to pass undefined options to subfunctions. No error message is shown, and synth's default optimizers (BFGS and Nelder-Mead) are run instead.

We have not verified if this is a change in syntax from later or earlier versions of the 'synth' package. But with the (potential) coding error ('method="All"'), we are able to reproduce the findings from the paper exactly. With 'optimxmethod="All"', 'synth' picks an alternative optimizer (Rvmmin) in the main specification, which results in a small difference in the synthetic control weights (Table 1). The estimates are, however, almost identical (-0.30 instead of -0.29 metric tons of CO_2 per capita in an average year). We therefore perform the rest of the replication without correcting the coding error (thus, in effect, running 'synth' with its default optimizers in R).

3. Replication

We now turn our attention to our replication. First, we conduct a computational reproduction of the main findings, using the code and data available in the replication package for the article. Second, for robustness replicability, we test the robustness of the results in Andersson (2019) to the 14 synthetic control specifications suggested by Ferman et al. (2020). The specifications they suggest are using (1) all pre-treatment outcome values, (2) the first three-fourths of the pre-treatment outcome values, (3) the first half of the pre-

treatment outcome values, (4) odd pre-treatment outcomes, (5) even pre-treatment outcomes, (6) pre-treatment outcome mean, and (7) the first, the middle, and the last outcome value. These specifications are estimated (a) without covariates, and (b) with covariates. We implement the replication by adapting the code provided in the paper's replication package. Andersson notes (p. 13) that he has conducted some sensitivity analyses with different specifications without any meaningful changes in results, but the resulting estimates are not detailed in the paper or appendix.

Our decision to conduct these robustness checks was taken after reading Andersson's paper but prior to observing the codes. We communicated our analysis plan with the organizers of the Toronto Replication Games by email on February 3, 2023.

3.1 Computational reproduction

We successfully reproduce Andersson's findings using his code (Table 2, first specification).

3.2 Robustness replication

The results from the 14 synthetic control specifications suggested by Ferman et al. (2020) yield a median point estimate of –0.28 metric tons of CO_2 per capita, with point estimates ranging from –0.34 and –0.17. Placebo-based p-values are equal to 1/15=0.067 in seven specifications, 2/15=0.13 in six, and 4/15=0.27 in one (Table 2).

4. Concluding remarks

We successfully reproduce Andersson's main findings. In our replicability analyses, we find that the median emission reduction from our specifications is almost identical to the one found in the original paper, and all estimates are in the same direction. Additionally, we find that Sweden has the largest ratio of all 15 countries in 50 percent of our specifications, and the largest or second largest in all but one.

References

Andersson, Julius J. 2019. "Carbon Taxes and CO_2 Emissions: Sweden as a Case Study." *American Economic Journal: Economic Policy*, 11(4): 1-30.

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

Tables

Table 1. Weights for the synthetic Sweden from Andersson
(2019) and weights with optimxmethod="All"

Country	Andersson (2019)	optimxmethod="All"
Australia	0.001	0
Belgium	0.195	0.165
Canada	0	0
Denmark	0.384	0.387
France	0	0
Greece	0.090	0.114
Iceland	0.001	0
Japan	0	0
New Zealand	0.177	0.215
Poland	0.001	0
Portugal	0	0
Spain	0	0
Switzerland	0.061	0.031
United States	0.088	0.087

Table 2. Reproduction of Andersson (2019) and robustness to Ferman et al. (2020) specifications

Specification	Emissions reduction 2005	Average emissions reduction 1990-2005	Rank, in-space placebo test
Andersson (2019)	-0.35	-0.29	1
1a	-0.35	-0.27	2
2a	-0.36	-0.28	2
3a	-0.19	-0.17	4
4a	-0.36	-0.28	2
5a	-0.33	-0.27	2
6a	-0.29	-0.25	2
7a	-0.23	-0.19	1
1b	-0.35	-0.28	1
2b	-0.35	-0.28	1
3b	-0.36	-0.27	1
4b	-0.35	-0.28	1
5b	-0.34	-0.28	1
6b	-0.38	-0.34	2
7b	-0.33	-0.26	1
Median	-0.35	-0.28	
Min	-0.38	-0.34	
Max	-0.19	-0.17	

Note: Andersson is the reproduced main estimate from Andersson (2019, p. 14). 1a-7b are the specifications suggested by Ferman et al. (2020). Specifications ending with a do not include covariates, while those ending with b include them. The rank after in-space placebo tests shows that the effect was largest in Sweden in 7 of 15 new specifications, second largest in 6 specifications, and forth largest in one specification.

8