

FINAL PROJECT

Replication & Response of Synthetic Control Approach for

TX Prison Building Program

Minerva University | Professor Alexis Diamond

CS130 | 2022 - 2023 Spring Semester

Honglu Wang, Heying Wang

Executive Summary

Cunningham, S. & Kang, S. (2019) applied a synthetic control method to investigate the causal effect of increased prison construction on the African-American male incarceration rate in Texas. We replicated their original synthetic control method and found matching results with minor variations. In addition, we performed two extensions: an in-time placebo test and a robustness test by leaving out high-weighted control units. The extension results demonstrate that no observed factors drive the estimates and the absence of false treatment effects, further establishing our confidence in the robustness of the causal effect stated in the original work.

Introduction

This paper intends to replicate Scott Cunningham's "Studying the Effect of Incarceration Shocks on Drug Markets" (Cunningham & Kang, 2019).

The paper focuses on the US incarceration rates' effect on drug usage and markets, specifically targeting Texas' penitentiary system as an experiment. Texas was the target of an overcrowded lawsuit, ultimately leading to a massive prison expansion project. Subsequently, Texas' incarceration rates skyrocketed largely due to declining paroles. Cunningham explored the statistically significant results of Texas' imprisonment capacity utilization with different pathways that might lead to this increase. The remaining research builds on previous discoveries and focuses on the effect of prison expansion on drug markets (price), purity, and consumption, where no significant changes were found pre- and post-treatment (prison expansion).

The paper used a synthesized control approach (Abadie et al., 2010; Abadie & Gardeazabal, 2003) toward the incarceration rates in Texas without the expansion project, utilized to determine a causal effect of the prison expansion project on the incarceration rates in Texas. The inference is further used to support proceeding questions of incarceration of drug usage, where the fundamental problem is linking the expansion project to the actual criminal justice behavior of Texas and whether or not it propelled further influence on the drug market.

The result from its synthetic control study shows a significant positive causal effect of prison construction on African-American males' incarceration rate. The study applied in-place placebo tests and obtained a supporting p-value, demonstrating the statistical significance of the causal effect estimate.

However, this paper intends to replicate the causal question proposed by Cunningham as “Did Prison Expansion Increase Incarceration?” (Section 4.1, Cunningham 2019), with several extensions made on a replication of the original synthetic control, given the limited data set and information apart from this specific question that was included in the paper. The data on which this replication paper would build is derived from data disclosed in “Causal Inference: The Mixtape” (Cunningham, 2021).

Cunningham et al. gathered data from seven sources, including the National Prison Statistics (NPS). The data contains 816 observations, 16 from each state in the US (50 states and Washington DC), and contains 16 variables listed down below. The data can be accessed through the link in the Appendix.

Variable Name	Type	Format
---------------	------	--------

statefip	float	%9.0g
year	float	%8.0g
bmprison	double	%10.0g
wmprison	double	%10.0g
bmpop	long	%12.0g
wmpop	long	%12.0g
alcohol	float	%9.0g
income	long	%12.0g
ur	float	%9.0g
poverty	float	%9.0g
black	float	%9.0g
perc1519	float	%9.0g
aidscapita	float	%9.0g
state	str20	%20s
bmprate	float	%9.0g
wmprate	float	%9.0g

Table 1. Variables and data types

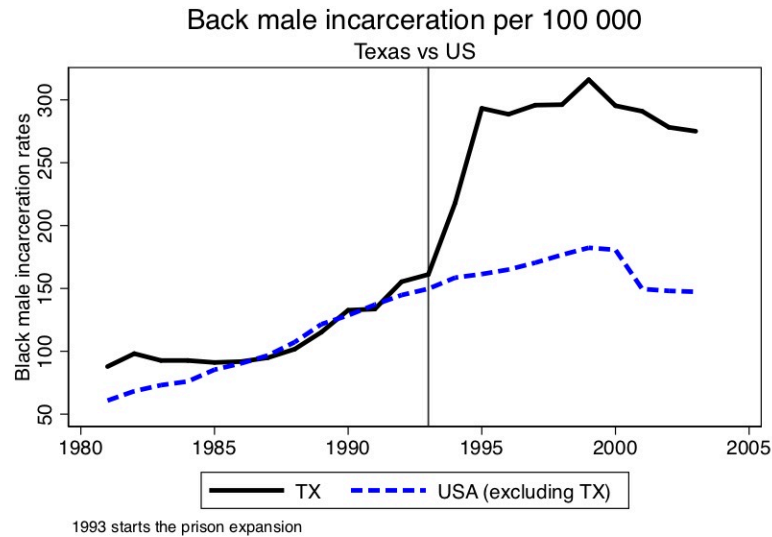


Figure 1. The original synthetic control graph by Cunningham et al. (source: https://mixtape.scunning.com/10-synthetic_control#fnref4)

Replication¹

The synthetic control method is used to analyze the effect of an intervention in a comparative case study with just one treated unit. A synthetic control unit is formed by combining and assigning weights to other units that are similar to the treated unit in the pre-treatment period (in both the outcomes and the covariates) that did not receive the treatment to see the difference in the post-treatment outcome between the treated unit and the controlled unit. In this case, the treated unit is Texas, while other states where the program was not introduced are in the synthetic control donor pool. The following replication reached similar findings as the original paper.

¹ **#studyreplication:** replication is done to produce multiple graphs in the original study (original synthetic control path plot, gap plots, in-space placebo and MSPE frequency distribution). Thoroughly explained the goal and method of the original study; proper extensions were built on the replication to further explore the robustness of the method.

First, we reproduced the previous synthetic control path plot (Figure 1). In this case, the outcome variable is the number of African Americans in prison (bmprison), shown on the y-axis. The predictors (covariates) included in the analysis are poverty and income, the same as in the original paper. The treatment (the start of building more prisons in Texas) happened in 1993. The weights assigned in the replication are shown in the table below.

State name	Weights
California	0.34
Florida	0.30
Louisiana	0.36

Table 2: Assigned weights for synthetic Texas

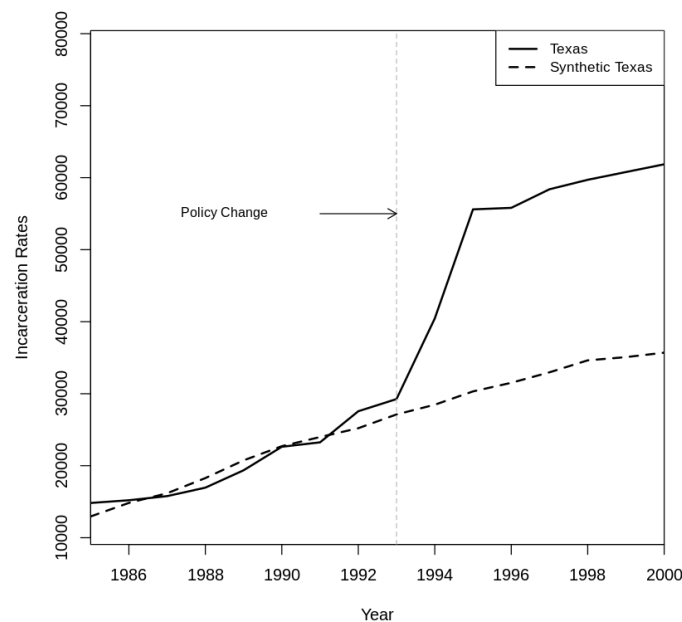


Figure 2: Synthetic control path plot

The graph above (Figure 2) is the replicated synthetic control graph, similar to Figure 1. Before 1993, the black man incarceration rates in Texas and the synthetic control Texas had very similar trajectories; however, things changed quickly after 1993, as the incarceration rate in Texas increased enormously, with its numbers almost doubled compared to the synthetic unit. This confirmed the findings produced by Cunningham et al.

Apart from the path plot, we also reproduced the gap plot, which shows the difference in black man incarceration rates between the treated unit (Texas) and the synthetic control unit. As expected, the difference remained close to 0 before the treatment but quickly spiked to more than 20000 two years after the program's onset.

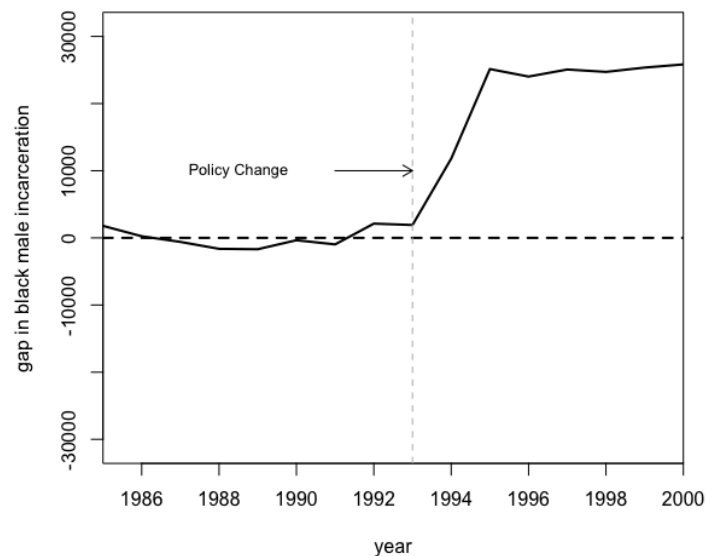


Figure 3. Gap plot of synthetic control

After constructing the synthetic control graph, Cunningham et al. performed a robustness test. The reason for such tests is to ensure that such a big difference does not happen by chance and, therefore, could result from the treatment. The test performed in the original paper is an

in-space placebo test, meaning that we perform the same steps as we did to the original treated unit (Texas), but now to every other unit as we treat them as the treated unit. Then we would have many gap plots showing the difference between each treated unit and their respected synthetic control unit, which eventually forms Figure 4. Then, Post and Pre treatment MSPEs (Median Squared Prediction Error) are calculated for each country, illustrating the ratio of discrepancies between the pre and post-treatment periods. The higher the Post/Pre ratio, the larger the discrepancies. The p-value comes from comparing how extreme the observed effect (MSPE ratio) on Texas is relative to the in-space placebo effects. As seen in Figure 5, Texas has the third-highest ratio among all 51 states. Therefore, the p-value could be calculated as $3/51 = 0.05$, which is extremely small, and indicates that it is hard for units to have such significant discrepancies when adopting this model. This also paves the way for further robustness testing in the following section.

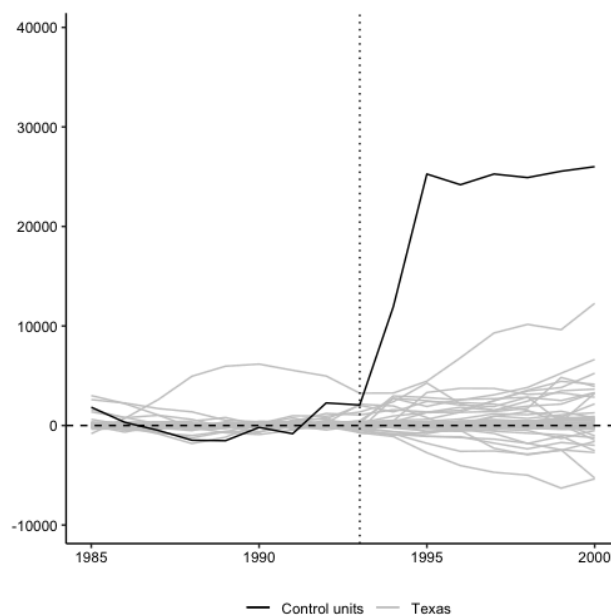


Figure 4: In-place placebo test graph of the Texas vs. All control units

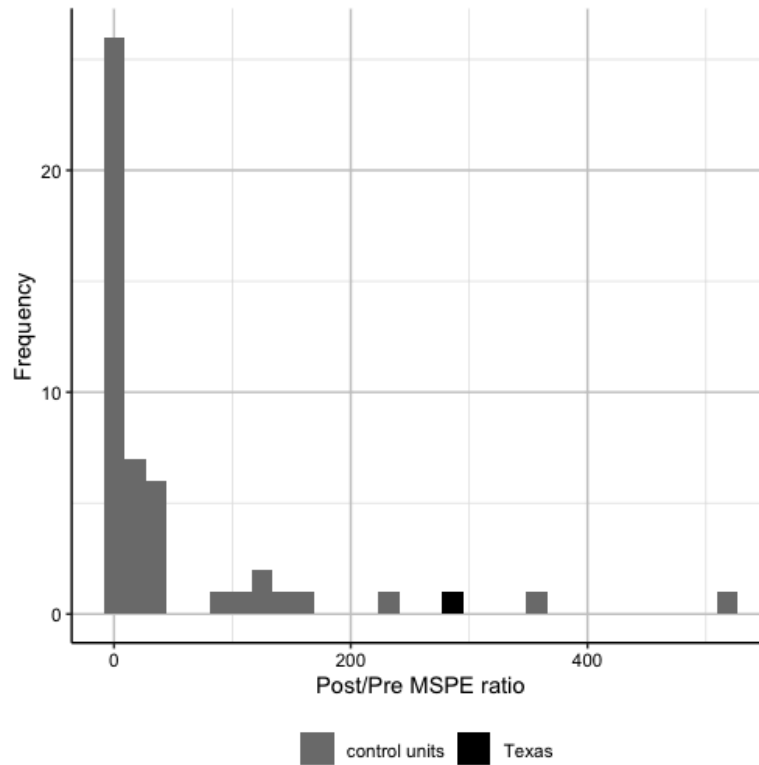


Figure 5: Frequency distribution plot of the Post/Pre MSPE Ratio

Extension

We performed several extensions on the study, building further on the replicated results. In the original paper, Cunningham mainly focused on an in-space placebo test with p-values as a robustness test. We intend to explore other factors further to determine the robustness of the model by applying: robustness tests by leaving out large contributing factors and one in-time placebo test. The original paper has done the extension of an in-space placebo test and found out that

In-time Placebo Test: The paper conducted an in-time placebo, feasible if there are available data for a sufficiently large number of periods when no structural, used to determine the robustness and validity of the model through implementing a placebo treatment and determine whether it acquires a higher estimated effect than the actual treatment (Abadie et al., 2014). The paper placed a placebo treatment in 1990, before 1993, and applied the synthesized control to the corresponding placebo treatment time. More specifically, we changed the “*time.predictors.prior*” and “*time.optimize.ssr*” factors from *1985:1993* to *1985:1990*, where we applied the false treatment in 1990. We observe, in Figure 6, that there is a large effect for the 1993 Texas prison expansion, but only a minor gap between the lines when we artificially reassign the expansion to a time before 1993 (expansion period to 1990). Thus, we can conclude that there is no false treatment effect, and we could be more confident in the robustness of our results.

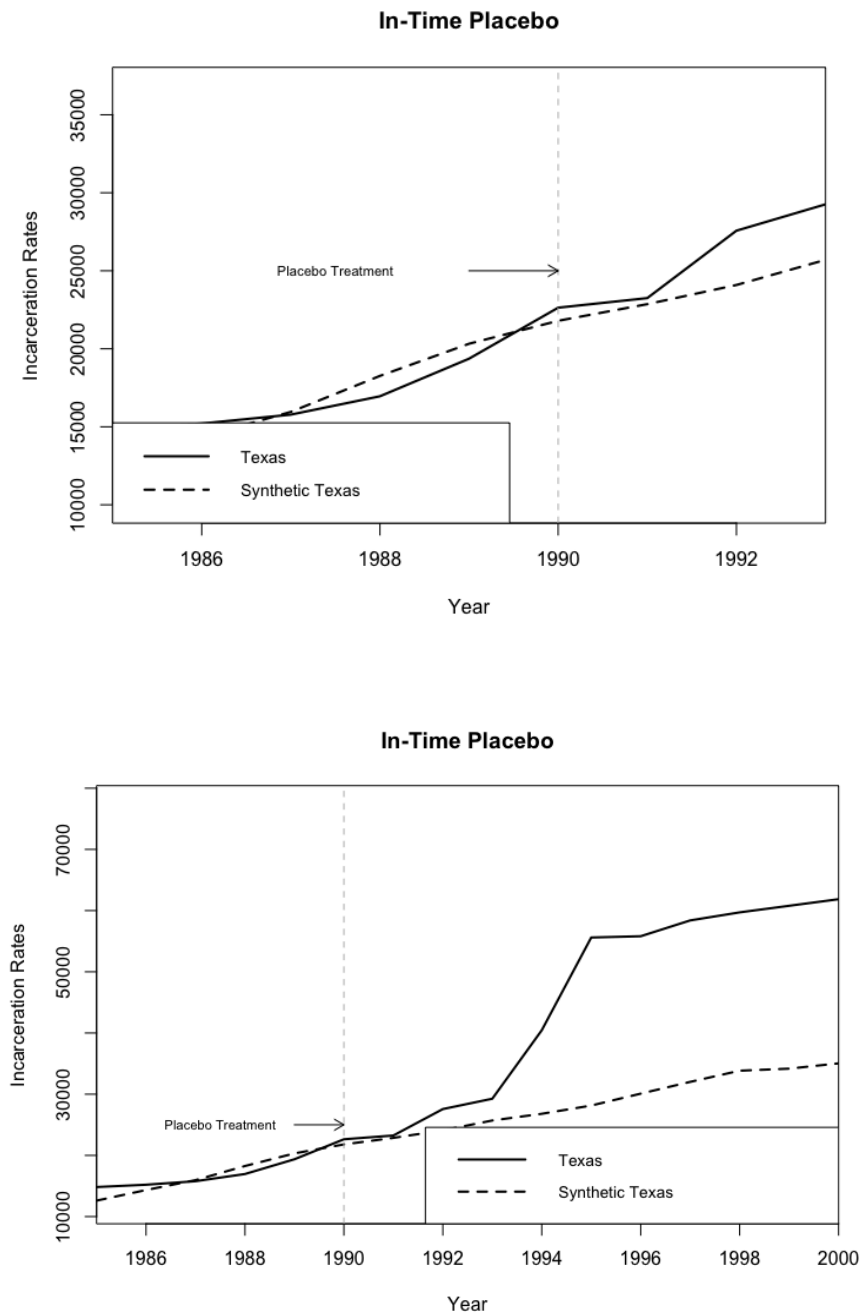


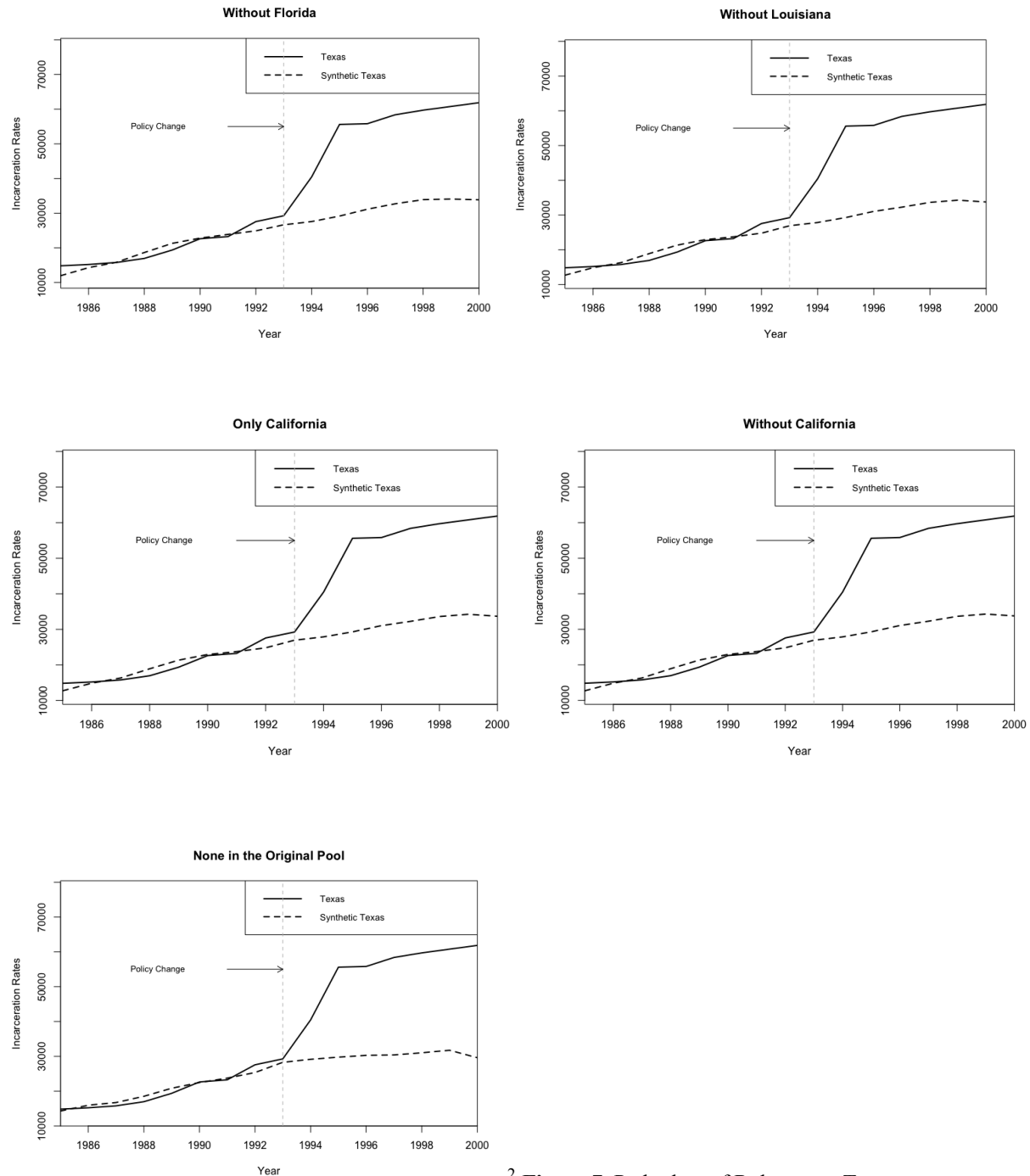
Figure 6. Path plot of an in-time placebo, assigned to 1990 before the actual expansion project in 1993. *Up*: Plot until 1993, the year of the expansion project; *Down*: Plot until 2000.

Robustness Test (Leave-one-out & Largest Contributor): Furthermore, to test the validity of the synthetic model, we conducted the largest contributor test to inspect the possibility

that there is an overall strong contributor control unit that affects the model. We observed that there is there main contributing units, California, Louisiana, and Florida, among other weighted units which approximately rounds to zero. We conducted the synthesized approach iteratively to the main donor pool, reducing the least weighted within the three in each round. In addition, we removed one of the three largest contributors in the donor pool one by one, following a leave-one-out method. We approached this method by altering the specifications in *dataprep()* function on *controls.identifier* during each iteration of *dataprep()*, *synth()*, and *path.plot()* where we arrived at several plots for different scenarios. We observe, in Figure 7, that making changes and restricting the donor pool does not result in significant changes in the model's fit. The only observable change was completely dropping the original large contributors in the donor pool. Thus because of this consistency between the robustness tests in Figure 2, we can conclude that no controlling state is biasing the estimates, and the causal effect is robust.

Conclusion/Discussion

We replicated the results from the original Cunningham & Kang (2019) paper including the synthetic control plot and robustness tests comprised of an in-place p-value test (permutation) and an MSPE plot. The results are similar to the original paper with minor differences, affirming partially the validity of the study. With the extension of in-time placebo tests and robustness tests, the robustness is specifically confirmed. Going forward, additional research and exploration can be done in the paper's following questions, including studying the impact of incarceration on the drug market.



² **Figure 7.** Path plots of Robustness Test.

² **#dataviz:** The paper includes twelve plots including one cited plot, for the replications and extensions of the original study. The paper specified the axes, titles, and legends, and used the vertical line to indicate the treatment period. The plots are further connected with interpretations of each and their significance over the robustness of the causal effect. Some plots are intended to replicate the results included in the study, where minor discrepancies are observed but not significant enough to hinder the claims.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (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.
<https://doi.org/10.1198/jasa.2009.ap08746>
- Abadie, A., Diamond, A., & Hainmueller, J. (2011). Synth: AnRPackage for Synthetic Control Methods in Comparative Case Studies. *Journal of Statistical Software*, 42(13).
<https://doi.org/10.18637/jss.v042.i13>
- Abadie, A., Diamond, A., & Hainmueller, J. (2014). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2), 495–510.
<https://doi.org/10.1111/ajps.12116>
- Abadie, A., & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review*, 93(1), 113–132.
<https://doi.org/10.1257/000282803321455188>
- Cunningham, S. (2019). *Causal Inference The Mixtape - 10 Synthetic Control*. Mixtape.scunning.com. https://mixtape.scunning.com/10-synthetic_control
- Cunningham, S., & Kang, S. (2019). *Studying the Effect of Incarceration Shocks to Drug Markets*. https://www.scunning.com/files/mass_incarceration_and_drug_abuse.pdf
- Unpublished Manuscript

Appendix A: Project Code & Data

Link to our code: <https://gist.github.com/HongluWang/33172e686483c0edd240d47d3c4917a6>

Link to the code from the original paper:

https://github.com/scunning1975/mixtape/blob/master/R/synth_1.R

Link to data:

<https://github.com/scunning1975/mixtape/tree/bfc9653f833f8052bd53b5f3b75dc544e31f3f30/Texas>

Appendix B: Division of Work

Heying: Replication Code and Paragraphs, Review, Discussion & Conclusion

Honglu: Extension Code; Executive Summary, Introduction, and Extension Paragraphs

AI Disclaimer: No AI tools were used in the production of this assignment.