Electoral Insecurity & Federal Spending Procurement: Causal Inference with Panel Matching and Synthetic Control Methods

Tom (Hyeon Seok) Yu*

Last Update: December 2021

Abstract

How do legislators respond, if at all, to changes in their electoral prospects? Most existing studies adopt a difference-in-differences design that exploits redistricting as an exogenous shock to estimate the causal effect of electoral insecurity on legislators' federal spending procurement for their districts. This project employs matching and synthetic control methods that produce more comparable counterfactuals to derive the causal estimate of interest. Nearly all matching and SC methods yield improved covariate balance. In addition, these methods return mostly null results while the conventional difference-in-differences method returns statistically significant results, which suggests the importance of ensuring comparability of treatment and control groups. Finally, a negative outcome analysis is conducted to compare the performance of different synthetic control methods.

*Ph.D. Candidate, Stanford Graduate School of Business. Email: tom.hs.yu@stanford.edu

Introduction

Competitive elections provide a means through which voters can hold elected officials accountable (Ashworth 2012), and legislators, as electorally motivated individuals (Mayhew 1974), take various actions to impress their voters with the aim of improving their electoral prospects. But what actions do they take, and how responsive are they to the wishes of their voters? Addressing whether and how legislators respond to changes in their electoral prospects – resulting from changes in voter preferences, for instance – is important, as it directly speaks to the state of electoral accountability. While the empirical literature has generally found patterns of electorally vulnerable legislators working harder for their constituents (e.g., securing more federal spending), a rigorous within-legislator analysis is difficult to find, and those that do rely exclusively on the difference-in-differences ("DiD") framework, which requires assumptions such as the parallel-trends assumption.

This project fills the empirical gap by employing methods that construct more comparable counterfactuals. Taking the decennial redistricting as an exogenous shock to legislators' electoral insecurity, this project implements matching and synthetic control ("SC") methods after separating legislators into treatment and control groups based on the degree to which their districts' average preference shifted (i.e., treatment intensity). Evaluating different matching methods based on covariate balance reveals that the matching refinement based on Mahalanobis distance generally outperforms others. More interestingly, the causal estimates from matching and SC methods are null results, while conventional DiD estimates suggest statistically significant effects. Finally, the negative outcome that assumes a pseudo-treatment period compares different SC methods and finds that the plain method outperforms a generalized method, which likely reflects the importance of meeting data requirements – having a sufficient number of pre-treatment periods, in particular – for such methods.

Related Work

Political scientists have long noted that legislators are electorally motivated individuals (Mayhew 1974). In order to win more votes, House members take various actions, one of which includes securing federal funding for their districts. Ashworth and Mesquita (2006) provides a formal account of how an increase in electoral insecurity could induce legislators to secure more federal funding. In their game-theoretic model, legislators face an allocation problem where they divide their resources between (1) providing constituency service, which only benefit their local districts and (2) providing global public goods, an example of which includes producing legislation that benefits the whole country. Aside from ideological congruence, one key factor that affects voters' decision on reelecting the incumbent is their belief on the latter's capacity, which is increasing in the amount of constituency service provided.

A number of empirical studies have already found evidence seemingly consistent with the stated theoretical result.[1] Existing works that employ panel data rely heavily on the DiD

[1] With a descriptive observation that "legislators who 'bring home the bacon' are rewarded for their efforts at the ballot box," Stein and Bickers (1994) find that electorally vulnerable House incumbents, specifically

framework that exploits redistricting as an one-time exogenous "treatment" (shock) to legislators' electoral insecurity to estimate the causal effect of interest. This project aims to extend the literature by implementing methods that make the parallel trends assumption more plausible.

Data

The panel dataset of this study consists of 3,688 representative-year pairs with 346 unique House incumbents from years 2003 to 2018. Since the research design takes advantage of 2010 redistricting as an exogenous shock that changes each legislators' electoral insecurity to analyze within-legislator variations, House incumbents need to have served in the 112th (2011-2012) and at least a part of the 113th Congress (2013-2014) sessions to be included in the sample. This means members who either were not present in this period or retired/resigned before 2012, the year 2010 redistricting became effective for the House elections, are excluded.

The key outcome and explanatory variables of interest are the federal spending (i.e., pork) procured by each legislator and a measure of electoral insecurity induced by redistricting, respectively. First, the data of the amount of federal spending channeled to a district comes from publicly available federal assistance data known as Federal Assistance Award Data System ("FAADS"), employed widely among political scientists.[2] Second, the measure of electoral insecurity relies on Tausanovitch and Warshaw (2013)'s multilevel regression and poststratification ("MRP") data that estimate the average of constituents' policy preferences at the congressional district-level before and after redistricting. For a member i:[3]

$$\text{Electoral Insecurity (MRP)}_i = \begin{cases} MRP_{i,2010} - MRP_{i,2000} \text{ if Democrat.} \\ \text{-}(MRP_{i,2010} - MRP_{i,2000}) \text{ if Republican.} \end{cases}$$

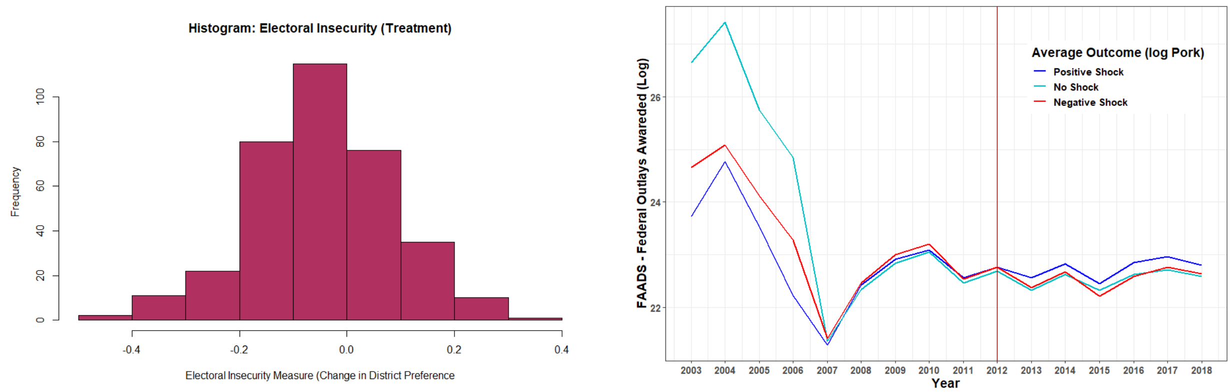
As a keen reader might have noted, the treatment is continuous; the change in electoral insecurity induced by redistricting as measured in MRP represents a degree. Figure 1a shows the distribution of the "shock" values, which seems close to a bell shape. To apply the methods proposed, a discretization is applied based on quantiles to divide data into three groups: 1) the "positive" treatment group (above 66 percentiles) consisting of those who saw the positive increase in their electoral insecurity measures, 2) the "negative" treatment group (below 33 percentiles), and 3) the control group (center 33-66 percentiles). Figure 1b shows the trends of the outcome variable (pork) by the three different groups. In the Appendix, table 4 reports means of each variable by treatment groups, and table 5 shows the result of difference-in-means test for these variables between each treatment and control groups. The

those who won their last elections by small margins, are more likely to seek an increase in projects to their districts. Bickers and Stein (1996), Herron and Theodos (2004), Lazarus (2009), and Dropp and Peskowitz (2012) also find similar results in different contexts.

[2] Some examples include Bickers and Stein (1996), Berry, Burden, and Howell (2010), Dynes and Huber (2015), and Berry and Fowler (2016).

[3] This continuous measure is constructed so that a higher value indicates a greater degree to which a district's average preference shifted against the given representative's favor ex post redistricting. For a Democratic incumbent, for instance, her MRP-based measure would be positive if the district's MRP-measure increased (i.e., moved rightward) from 2000 to 2010, while the opposite would be the case for a Republican counterpart.

covariates seem mostly balanced, but some statistically significant differences appear for the share of Democrats and the share of legislators holding some committee leadership positions.



(a) Electoral Insecurity (Treatment) Distribution (b) Outcome Trend by Treatment Groups

Figure 1: Preliminary Analysis: Treatment Dist. and Outcome Trends

Methods

A credible claim for causal effect requires comparable treatment and control groups. In the current setting, seemingly apparent discrepancies in some attributes (e.g., the share of democrats) shown in table 5 suggests a relatively low comparability. Matching and synthetic control methods attempt to mitigate this issue by constructing counterfactuals similar to given treated units. The plausibility of assumptions required by each method is assessed at the end of this section.

Panel Matching

Matching a treated unit to the most similar control unit (or multiple units in the case of 1-to-many matching) based on pre-defined set of covariates offers one way of addressing the comparability issue. However, given the time dimension in the panel data, the proper way of constructing a matched set is not entirely obvious. Imai, Kim, and Wang (2021) proposes a method specifically for this problem.

Before delving into the method, let us define some notations. Suppose there are N units of legislator/district i, and define $t = 1, 2, \cdots T$ as time indicator, whose unit is year in this setting. Let T_0 be the set of all pre-treatment years, and T_1 that of the post-treatment.[4]. Finally, define Y_{it} as the outcome, the amount of federal spending awarded to a district i at time t, D_{it} as the treatment indicator, and X_{it} as K time-varying covariates.

This particular method utilizes "treatment history" as the central criterion for matching: for a given treated unit, its matched set only consists of control units that share the identical treatment history.[5] Figure 2 shows the distribution of treatment history for both positive

[4]2012 is the first year of treatment, so $T_0 = 2003 : 2011$, and $T_1 = 2012 : 2018$

[5]However, this does not mean an exact matching where all control units must have exactly the same history.

and negative treatment groups. Then, based on these matched control units, the average treatment effect on the treated ("ATT"), τ can be estimated as:

$$\tau_{it}(F,L) = E[Y_{i,t+F}(D_{it}=1, D_{i,t-1}=0, \{D_i, t-l\}_{l=2}^{L}) - Y_{i,t+F}(D_{it}=0, D_{i,t-1}=0, \{D_i, t-l\}_{l=2}^{L})]$$

where F and L refer to the lead and lag periods, respectively. For example, $\tau(2,5)$ is the ATT of two periods after treatment depending only on the treatment history up to five time periods back.[6] The latter quantity is estimated based on the matched set, M:

$$M_{it} = \{i' : i' \neq i, D_{i't} = 0, D_{i't'} = D_{it'} \forall t' = t-1, \cdots t-L\}$$

for the treated unit i with $D_{it} = 1$ and $D_{i,t-1} = 0$. This current matched set only relies on the treatment history, but it can be further refined by employing the following methods that compute some distance $S_{it}(i')$ for each $i' \in M_{it}$:

- **Mahalanobis distance** matching:

$$S_{it}(i') = \frac{1}{L}\sum_{l=1}^{L}\sqrt{(V_{i,t-l} - V_{i',t-l})^{\intercal}\Sigma_{i,t-l}^{-1}(V_{i,t-l} - V_{i',t-l})}$$

  where $V_{it'}$ is the vector of time-varying covariates and $\Sigma it'$ is the sample covariance matrix of $V_{it'}$.

- **Propensity score** matching: This method employs logistic regression as a treatment assignment model to compute propensity score.

- **Covariate balancing propensity score** ("**CBPS**"): Imai and Ratkovic (2014) provides another method where the propensity score "is estimated such that it maximizes the resulting covariate balance as well as the prediction of treatment assignment."

Then, based on the distance computed, the refined matched set only consists of J selected control units with the lowest distance measure. All of these matching methods can be evaluated by comparing covariate balances achieved.

(a) Positive and Control Treatment History

(b) Negative and Control Treatment History

Figure 2: Treatment History by Treatment Groups

That is, if a given control unit had more pre-treatment data while sharing the same treatment history for the treated unit's pre-treatment period, then it will be included in the matched set.

[6] The authors suggest that F relies on the analyst's institutional knowledge, while L is an empirical quantity that embodies bias-variance tradeoff. In this setting, we select $F = 5$ given the long time it generally takes for legislators to secure federal spending for their districts, and $L = 5$ based on pre-treatment period.

Synthetic Controls

In addition to the matching method, the synthetic control ("SC") method proposed by Abadie, Diamond, and Hainmueller (2010) is another way of constructing comparable counterfactuals. Similar to matching, it relies on control units (the donor pool), but instead of taking control units as they are, SCM constructs a synthetic counterfactual by taking a convex combination of selected control units. More specifically, the key is to compute weights that achieve better balance in covariates while minimizing the mean squared prediction error ("MSPE") in the pre-treatment years.[7]

One major shortcoming of the plain SCM is its inefficiency in the case of multiple treated units: one needs to construct synthetic control for each treated unit, requiring N_{tr}, the number of treated units, number of computations that often involve expensive constrained quadratic optimization. The generalized SC ("GSC") method proposed by Xu (2017) offers a much more efficient way of computing counterfactuals. Specifically, it constructs counterfactuals using control unit data "based on a linear interactive fixed effects model that incorporates unit-specific intercepts interacted with time-varying coefficients," the model of which can be estimated in a single run. More importantly, it also offers a clearer way of achieving uncertainty estimates, not provided by plain SC.[8]

Assumptions: Parallel Trends & Exogeneity

Both matching and SC methods require a number of assumptions for causal identification, and this section discusses two major ones. As aforementioned, the main requirement of DiD is the parallel trends assumption (Imbens and Lemieux 2008): for it to be the case that changes in electoral insecurity resulting from redistricting are responsible for changes in federal spending procurement, the latter should exhibit parallel trends in the pre-treatment period. Both matching and SCM, to an extent, mitigate the issue by constructing counterfactuals that follow similar trends in the pre-treatment period.

Another major assumption required for both methods is the strict exogeneity assumption. Formally,

$$\epsilon_{it} \perp\!\!\!\perp D_{js}, X_{js} \ \forall i, j, t, s$$

which states that the error term of any unit at any time period must be independent of treatment assignment and observed covariates.[9] In particular, there should not be unobserved factors affecting both covariates and the treatment assignment (redistricting) systematically, thereby making the assignment non-random. For this setting, there could be unobservables such as an incumbent's "well-connectedness" in his/her state legislature, which is in charge of re-drawing of districts, affecting the treatment intensity.[10] Given their unobervability, they are difficult to account for, and these potential violations are acknowledged. The pro-

[7] For a detailed formal treatment of the method, see Abadie (2021).

[8] The plain SCM relies on placebo tests, where the treatment is randomly assigned to a control unit, to obtain uncertainty estimates, but these tests can be arbitrary.

[9] For the GSCM that employs a linear fixed effects model, the assumption extends to unobserved cross-sectional and temporal heterogeneities (see Xu 2017, p. 62).

[10] Indeed, if the treatment assignment is not random, one can no longer attribute the changes in pork procurement to the changes in electoral insecurity.

posed methods are pursued under the ground that they render the first assumption more defensible.

Experiments & Results

The main experiment comprises three parts: 1) the balance of covariates achieved by each method, 2) ATT estimates relative to the DiD baseline, and 3) negative outcome analysis employing year "2009-2011" as the pseudo-treatment period. In comparing each method, one important caveat to note is that samples/matched sets considered by each class of methods differed to an extent, meaning slight discrepancies in the sample size.

Covariate Balance

The covariate balance is computed by taking the average of the difference between the values of the specified covariates for the treated unit(s) and the weighted average of the control units across all matched sets.[11] Table 1 reports the results, which are standardized to reflect standard deviations for each covariate, and the almost all methods return improved balance compared to raw data. Among refinement methods, performance is largely mixed; the refinement using Mahalanobis generally outperforms other methods on average. Most matching methods return better balance, especially for the "democrat" attribute, as shown in figures 4 and 5 in the Appendix.

Table 1: Covariate Balance by Methods

	No Matching	Unweighted	Mahalanobis	Prop. Score	CBPS	SC
Positive Treatment						
Chamber Tenure	1.41	0.16	**0.01**	0.15	0.06	-0.10
Committee Leadership	-0.82	-0.08	-0.42	-0.18	-0.18	**-0.01**
Democrat	-0.02	0.68	**-0.01**	-0.06	**-0.01**	0.03
State Population (log)	-0.58	0.12	**0.03**	0.12	0.04	**0.03**
Negative Treatment						
Chamber Tenure	-1.48	**-0.05**	-0.06	-0.11	-0.11	-0.60
Committee Leadership	0.60	-0.23	-0.42	-0.49	-0.50	**-0.01**
Democrat	0.24	-0.28	**-0.02**	-0.06	-0.03	**-0.02**
State Population (log)	0.65	0.02	0.02	0.02	**-0.00**	0.03

Panel ATT Estimates

Table 2 reports the desired causal estimate of interest for each method for both positive and negative treatment groups. It is notable that while the conventional DiD return estimates with economically significant magnitudes in the predicted directions, except GSC for the positive treatment, neither matching nor SC shows statistically significant estimates.[12] These estimates reflect an average across post-treatment periods, so certain years might show estimates significantly different from zero. However, both matching and SC methods

[11] For the panel matching methods, the computation relies on Imai, Kim, and Wang (2021)'s `PanelMatch` package in `R`, and the same approach is applied to achieve the corresponding estimates for SC.

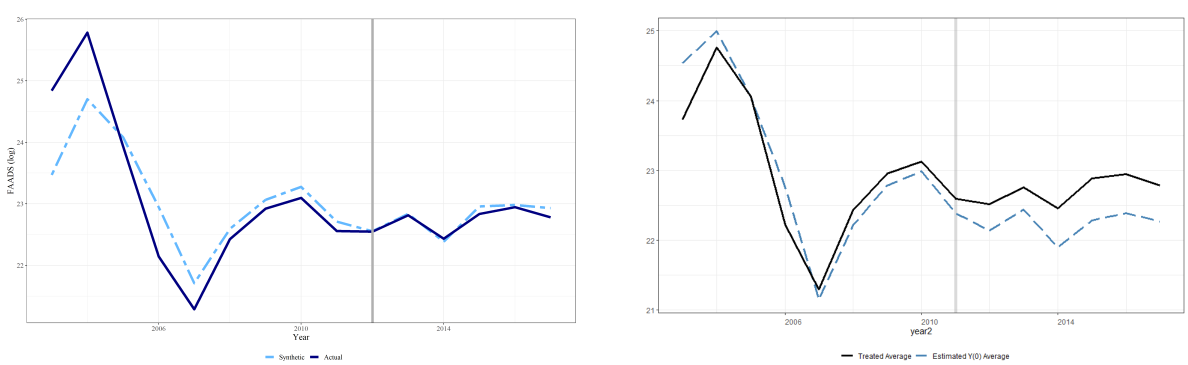
[12] The uncertainty estimates for the plain SC are omitted due to the time constraint; obtaining bootstrap standard errors is computationally expensive, and the validity of the resulting estimates is also unclear.

do not provide support for the claim that a greater electoral insecurity leads to more federal spending procurement.

Table 2: Panel ATT Estimate by Methods and Treatment Groups

	ATT Estimate	Standard Error	95% Confidence Interval
Positive Treatment			
DiD (No Matching)	0.361	0.202	(-0.035, 0.756)
Unweighted	0.141	0.134	(-0.117, 0.407)
Maha. Dist	0.165	0.158	(-0.143, 0.477)
Prop. Score	0.099	0.184	(-0.261, 0.462)
CBPS	0.109	0.201	(-0.295, 0.479)
SC	-0.049	-	-
GSC	0.564	0.293	(-0.011, 1.14)
Negative Treatment			
DiD (No Matching)	-0.588	0.192	(-0.964, -0.211)
Unweighted	0.006	0.133	(-0.250, 0.268)
Maha. Dist.	0.045	0.135	(-0.222, 0.307)
Prop. Score	0.072	0.167	(-0.262, 0.400)
CBPS	0.081	0.171	(-0.253, 0.415)
SC	-0.087	-	-
GSC	-0.006	0.159	(-0.318, 0.307)

Figure 3 provides visual results for both pre- and post-treatment periods: the solid line represents the actual outcomes observed, and the dashed line those of the synthetic counterfactuals constructed by each method.[13] Comparison of the two methods suggests that GSC seems to have a slightly better pre-treatment fit, and it does show a positive gap in the post-treatment period. For the negative treatment group, however, both SC and GSC show no gap in figure 6 in the Appendix, and the latter in particular returns a very close synthetic control that essentially mirrors the actual outcomes in both pre- and post-treatment periods.



(a) Synthetic Control

(b) Generalized Synthetic Control

Figure 3: Synthetic Control Methods (Positive Treatment)

[13]The discrepancy in the actual outcomes results from the aforementioned difference in samples considered by each method. The difference in the treatment period indicator (gray bar) is simply the result of difference in the package.

Sensitivity Analysis (Negative Outcomes)
SC methods are now further evaluated by implementing a negative outcome analysis, where treatment is assumed to take place at a different time. Specifically, this analysis assumes 2009-2011 as the pseudo-treatment period. As the analyst observes the actual non-treated outcomes for this period, a method that yields lower MSPE should be deemed superior. Table 3 shows that SC far outperforms GSC, which is surprising given the relatively good fit exhibited by the latter in figures 3 and 6. The relatively poor performance may be attributable to the small size of pre-treatment years; although SC also needs a sufficient amount of pre-treatment data, GSC seems particularly sensitive to the availability.[14]

Table 3: Negative Outcome Analysis: MSPE(2009-2011)

	2009	2010	2011	Overall
Positive Treatment				
SC	0.001	0.005	0.015	0.007
GSC	0.411	0.454	0.438	0.434
Negative Treatment				
SC	0.005	0.005	0.051	0.020
GSC	0.309	0.365	0.331	0.335

Conclusion

Re-analyzing the panel data using matching and SC methods returns ATT estimates that are distant from the conventional DiD estimates while achieving better balance across all covariates. Considering that these methods render the parallel trends assumption more defensible, their corresponding results appear more convincing. That said, having a relatively short pre-treatment period remains an important caveat that should pause one from taking these results at face value. Nevertheless, a large change in results from applying these methods suggests the importance of making the control and the treatment groups comparable.

There are several limitations of the current project. First, as noted by Fong, Hazlett, and Imai (2018), an arbitrary discretization of the treatment drastically changes the interpretation of the causal estimate while losing the possibility of uncovering nuanced and substantive results. While the current methods employed are limited to binary treatments, an extension that allows analyzing continuous treatments seems both feasible and profitable. Second, both matching and SC methods inevitably result in the removal of unmatched units in the data. One should confirm that the out-of-sample units are comparable to those that remain in the data. Finally, given that the design that exploits redistricting is bound to have a relatively short pre-treatment period, a different design that overcomes such a limitation could be more suitable for future analysis addressing a similar question.

[14] Xu (2017) notes that with the small number of pretreatment period, "incidental parameters can lead to biased estimates," meaning it can yield biased counterfactuls.

References

Abadie, Alberto (2021). “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects”. *Journal of Economic Literature* 59.2, pp. 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, pp. 493–505.

Ashworth, Scott (2012). “Electoral Accountability: Recent Theoretical and Empirical Work”. *Annual Review of Political Science* 15.1, pp. 183–201.

Ashworth, Scott and Ethan Bueno de Mesquita (2006). “Delivering the Goods: Legislative Particularism in Different Electoral and Institutional Settings”. *The Journal of Politics* 68.1, pp. 168–179.

Berry, Christopher R., Barry C. Burden, and William G. Howell (2010). “The President and the Distribution of Federal Spending”. *American Political Science Review* 104.4, pp. 783–799.

Berry, Christopher R. and Anthony Fowler (2016). “Cardinals or Clerics? Congressional Committees and the Distribution of Pork”. *American Journal of Political Science* 60.3, pp. 692–708.

Bickers, Kenneth N. and Robert M. Stein (1996). “The Electoral Dynamics of the Federal Pork Barrel”. *American Journal of Political Science* 40.4, pp. 1300–1326.

Dropp, Kyle and Zachary Peskowitz (2012). “Electoral Security and the Provision of Constituency Service”. *The Journal of Politics* 74.1, pp. 220–234.

Dynes, Adam M. and Gregory A. Huber (2015). “Partisanship and the Allocation of Federal Spending: Do Same-Party Legislators or Voters Benefit from Shared Party Affiliation with the President and House Majority?” *American Political Science Review* 109.1, pp. 172–186.

Fong, Christian, Chad Hazlett, and Kosuke Imai (2018). “Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements”. *The Annals of Applied Statistics* 12.1.

Herron, Michael C. and Brett A. Theodos (2004). “Government Redistribution in the Shadow of Legislative Elections: A Study of the Illinois Member Initiative Grants Program”. *Legislative Studies Quarterly* 29.2, pp. 287–311.

Imai, Kosuke, In Song Kim, and Erik H Wang (2021). “Matching Methods for Causal Inference with Time-Series Cross-Sectional Data”. *American Journal of Political Science* Forthcoming.

Imai, Kosuke and Marc Ratkovic (2014). “Covariate balancing propensity score”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.1, pp. 243–263.

Imbens, Guido W. and Thomas Lemieux (2008). “Regression discontinuity designs: A guide to practice”. *Journal of Econometrics* 142.2, pp. 615–635.

Lazarus, Jeffrey (2009). “Party, Electoral Vulnerability, and Earmarks in the U.S. House of Representatives”. *The Journal of Politics* 71.3, pp. 1050–1061.

Mayhew, David R. (1974). *Congress: The Electoral Connection*. First Edition. New Haven: Yale University Press.

Stein, Robert M. and Kenneth N. Bickers (1994). “Congressional Elections and the Pork Barrel”. *The Journal of Politics* 56.2, pp. 377–399.

Tausanovitch, Chris and Christopher Warshaw (2013). “Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities”. *The Journal of Politics* 75.2, pp. 330–342.

Xu, Yiqing (2017). “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models”. *Political Analysis* 25.1, pp. 57–76.

Appendix

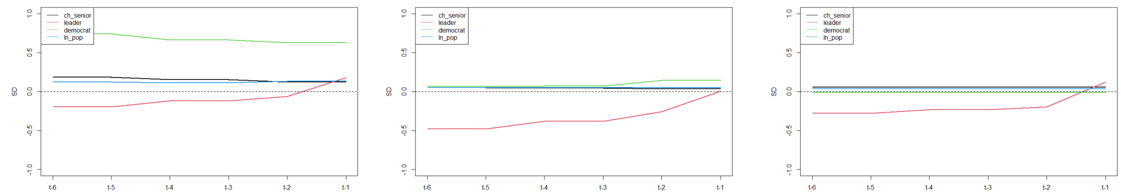
Additional Tables and Figures

Table 4: Pre-Treatment Means by Treatment Group

	Overall	Positive Treat	Negative Treat	Control
FAADS Awarded (log)	23.08	22.72	23.51	23.05
Electoral Insecurity	-0.04	0.10	-0.18	-0.04
Chamber Tenure	5.57	6.00	5.12	5.57
Committee Leadership (Binary)	0.12	0.12	0.09	0.15
Democrat (Binary)	0.45	0.68	0.28	0.39
Region (Midwest)	0.21	0.23	0.19	0.21
Region (Northeast)	0.19	0.24	0.12	0.20
Region (South)	0.38	0.34	0.46	0.34
Region (West)	0.22	0.18	0.23	0.25
State Population (log)	16.12	16.18	16.08	16.11

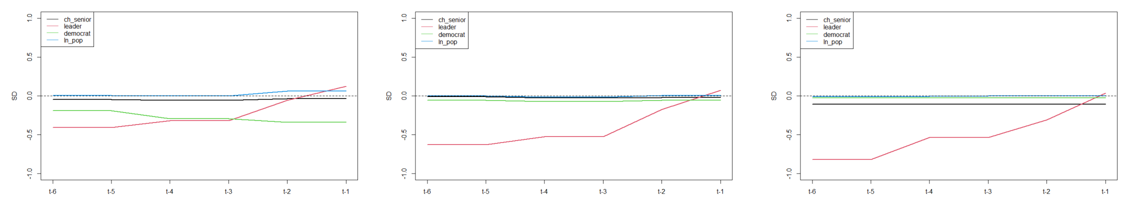
Table 5: Pre-Treatment Means Difference-in-Means

	Positive Treatment		Negative Treatment	
	t-stats	p-value	t-stats	p-value
FAADS Awarded (log)	-1.72	0.08	2.30	0.02
Electoral Insecurity	20.46	0.00	-18.89	0.00
Chamber Tenure	1.77	0.08	-1.80	0.07
Committee Leadership (Binary)	-1.56	0.12	-3.31	0.00
Democrat (Binary)	4.59	0.00	-1.84	0.07
Region (MidWest)	0.42	0.68	-0.35	0.72
Region (Northeast)	0.73	0.47	-1.80	0.07
Region (South)	0.00	1.00	1.94	0.05
Region (West)	-1.17	0.24	-0.34	0.73
State Population (log)	1.43	0.15	-0.81	0.42



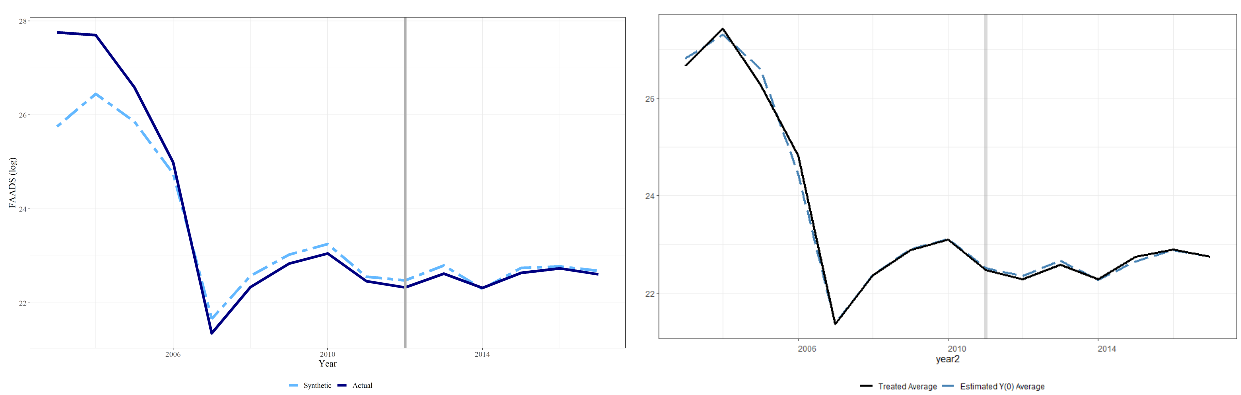
(a) Unweighted (b) Mahalanobis Distance (c) CBPS

Figure 4: Pre-Treatment Covariate Balance by Selected Methods (Positive Treatment)



(a) Unweighted (b) Mahalanobis Distance (c) CBPS

Figure 5: Pre-Treatment Covariate Balance by Selected Methods (Negative Treatment)



(a) Synthetic Control (b) Generalized Synthetic Control

Figure 6: Synthetic Control Methods (Negative Treatment)