Major Area Paper

Rise of the Machines: Causal Policy Analysis, Modern Econometrics, and Machine Learning

Jared Greathouse

Committee:
Dr. Jason Coupet (Chair)
Dr. Andrew Heiss
Dr. Jonathan Oxley
Dr. Lindsey Bullinger

November 6, 2024

Rise of the Machines: Causal Policy Analysis, Modern Econometrics, and Machine Learning*

Jared Greathouse[a]

[a]Georgia State University

November 6, 2024

Abstract

Probably because of the advances of computing power and statistical software, causal inference has become indispensable to public policy analysts in academia, industry, and government. The most common causal inference tools in the toolkit of policy analysts are the difference-in-differences and the synthetic control methodologies. Over the last five or six years, both methods have developed at a rapid pace, with each month bringing forth new applied and theoretical advances. These advances make it hard for policy analysts to stay abreast of the state-of-the-art estimators that are commonly taught (usually) in graduate school. This article reviews recent developments in the difference-in-differences and synthetic control methods (especially, but not limited to, those which employ machine learning techniques) and how they may be used by policy analysts. I first begin by discussing the standard, off the shelf iterations of DID and SCM as they are commonly employed in the policy analysis literature, with focus on their identification assumptions, estimation, and use cases. Then, I discuss some of the newly developed methods which are competitive or better than the off the shelf estimators in certain cases. I conclude by posing questions for future research on the basis of my review of the literature.

*Jared Greathouse: jgreathouse3@student.gsu.edu

1 Introduction

Given the advent of modern computing in the 20th century, researchers gained the ability to collect, store, access, and analyze data at a grand scale. Taking advantage of computing developments, public policy, economics, political science, and other fields, researchers became more interested in using advanced statistical methods to conduct data analyses to test hypotheses of interest, especially for causal inference (Spirtes, 2010; Shi et al., 2017). The so called "credibility revolution" has completely changed the thought process and practices many empirical researchers use to try and answer research questions (Ashworth et al., 2021). Given these shifting trends, statisticians have naturally attempted to create tools to meet big datasets so that they may be analyzed to draw inferences.

Given the fact that genuine randomized experiments are unethical or not possible for other reasons, policy analysts commonly resort to the tools of econometrics to estimate the casual impacts of policy. Rigorous methods for treatment effect estimation are indispensable weapons in the policy analyst's arsenal (Dague and Lahey, 2018; Imbens, 2024). This places quasi-experimental methods, and their advances from their original formulations, in high demand. The two methods which have ascended most in popularity among policy analysts are the difference-in-differences (DID) and synthetic control methods (SCM). Both methods are quasi-experimental designs created for situations where we wish to construct a counterfactual for a treated unit/group using a set of control units. Specifically, DID and SCM are both weighted regression estimators which exploit *pre-intervention correlations* between a treated group and their respective controls so as to calculate the average treatment effect on the treated.

Likely owing to implementation in popular statistical software like R and Stata, the last few years have witnessed substantial development in the theoretical and applied econometrics of DID and SCM. New designs for either method are developed almost by the month which are intended so supplement or replace originally established norms so that researchers may estimate causal effects in a more principled way. Recent papers simply *reviewing* their advances have amassed many thousands of citations (Abadie, 2021; Roth et al., 2023). Even these papers are quickly outdate as even newer insights are developed for things such as inference, staggered treatment adoption, statistical power, and other important factors.

However, computer science has also joined the development of both DID and SCM, in conjunction with broader trends that are well described by econometricians (Heckman and Pinto, 2022, 2024). In particular, many such methodological advances employ machine learning techniques to augment the off the shelf DID/SCM estimators. Machine learning methods represent a curious break from the traditional econometric model; their emphasis on prediction, error bounds, and convergence rates, deviate from the traditional econometric tradition which usually assumes some underlying data generating process. However, there is a kind of disconnect between these advances and public policy analysis. Sometimes, the tools used to implement them are written in softwares that are not commonly adopted in public policy, such as Python or MATLAB. Many of these methodological advances are frequently published in non-policy analysis journals [e.g., (Amjad et al., 2018; Li, 2024; Viviano and Bradic, 2023)]. This means policy analysts either likely do not know of these advances or cannot used these advances, even if the state-of-the-art methods could address methodological issues commonly faced in practice by public policy researchers.

The key contribution of this article is to overview some of the more recent tools that have been developed for SCM and DID that I feel should be part of the core of the graduate causal inference curriculum for policy data analysis courses. In particular, I list only some of these tools that have recently been developed. So naturally, such a list is subjective and I do not pretend for it to be comprehensive. This review seeks to highlight how some of the machine learning extensions to the baseline designs may, in some cases, assist policy analysis in a few key dimensions of their research design. Especially, I note how they may be employed for control group selection, use of regularization to combat overfitting, and the selection of covariates or robustness to not relying on them.

Layout I provide a brief review of the identifying assumptions and estimation strategies for DID and SCM. I take note of when policy analysts may face some of the issues regarding the dimensions I just discussed. Then, following Athey and Imbens (2019), I review the basic goals and terminology of machine-learning based methods to provide setup for the subsequent discussion. After, I detail some of the recent advances in both methods. A central point is that there are gaps in the literature which raise unanswered questions which should inspire future research. Beyond simply describing the new estimation methods, I emphasize why using these advances and answering questions surrounding them improves causal inference for policy analysis.

2 The Models

Difference-in-Differences The classic DID method exploits differences between the control group and treatment group in the pre and post treatment periods.[1] As mentioned above, it is a weighted regression

[1] To reduce clutter, I refer the reader to Section A for the precise notations I use.

estimator where the weights are uniform, $\mathcal{W}_{\text{DID}} := \left\{ \mathbf{w} \in \mathbb{R}^{N_0} \mid \mathbf{w} = \frac{1}{N_0} \right\}$. Here is DID's objective function

$$\underset{\mathbf{w} \in \mathcal{W}_{\text{DID}}, \widehat{\boldsymbol{\beta}}_0}{\operatorname{argmin}} \; \|\mathbf{y}_1 - \mathbf{Y}_0 \mathbf{w} - \boldsymbol{\beta}_0\|_2^2 \quad \forall t \in \mathcal{T}_1 \tag{2.1}$$

where $\boldsymbol{\beta}_0$ is the constant. The $\mathcal{W}_{\text{DID}}$ constraint reduces the solution to this problem to be the average difference between the treated and control group in the pre-intervention period. The way we compute the out of sample counterfactual, then, is simply the in-sample average difference between the treated and control group plus the empirical average of the control group. The key identifying assumption of DID is that absent the treatment, the expected difference between the treated and control groups would be constant, $\mathbf{y}_1^{\text{DID}} - \mathbf{Y}_0 \mathbf{w} = \boldsymbol{\beta}_0 + \epsilon$ (Roth et al., 2023). Here, $\mathbf{y}_1^{\text{DID}}$ is the model's predictions.

Of course, this setup may be generalized to incorporating multiple treated units where the treatment is introduced at different times (Wooldridge, 2023; de Chaisemartin and D'Haultfœuille, 2024; Callaway and Sant'Anna, 2021; Wing et al., 2024a,b). DID is a popular method among policy analysts fore precisely this reason (Pac and Berger, 2024; Connolly et al., 2024; Myers, 2024; Churchill et al., na; Slopen, 2024; Courtemanche et al., 2017; Pichler et al., 2021; Cawley and Frisvold, 2017; Gihleb et al., 2024; Willage et al., 2023). At the simplest level, DID is estimated via a two-way fixed effects linear model (Myers, 2024; Pac and Berger, 2024), where a set of unit and time dummies are included in the regression models. However, there is good reason to doubt the validity of these models, particularly in settings where the treatment is introduced at different time points (Goodman-Bacon, 2021) due to treatment effect heterogeneity and other concerns. As a result, it is common in the policy analysis literature to implement estimators that are robust heterogeneous treatment effects (Roth et al., 2023; de Chaisemartin and D'Haultfœuille, 2022). Gihleb et al. (2024) for example uses the estimators by Callaway and Sant'Anna (2021) and Borusyak et al. (2024) to study the impact of right-to-work laws on hours worked across different labor market sectors. Abouk et al. (2023) use the same to study the causal impact of recreational cannabis laws on pain management. Similar methods have also been used to study the impact of in-person schooling on COVID-19 cases during the pandemic (Goldhaber et al., 2022).

Synthetic Control Methods SCM was originally developed by Abadie and Gardeazabal (2003) to study the causal impact of terrorism in the Basque Country on Basque GDP. Later, SCM was formalized by Abadie et al. (2010) and Abadie et al. (2015) to study the causal impact of anti-tobacco legislation and German Reunification, respectively, on tobacco consumption and GDP per Capita. SCM is also a weighted regression estimator like DID. However, the unit weights are subject to $\mathcal{W}_{\text{SC}} := \left\{ \mathbf{w} \in \mathbb{R}_{\geq 0}^{N_0} \mid \|\mathbf{w}\|_1 = 1 \right\}$.

Here is SCM's objective function when covariates are not specified

$$\underset{\mathbf{w} \in \mathcal{W}_{\mathrm{SC}}}{\operatorname{argmin}} \; \|\mathbf{y}_1 - \mathbf{Y}_0 \mathbf{w}\|_2^2 \; \forall \, t \in \mathcal{T}_1 \tag{2.2}$$

where the unit non-negative and sum to 1. We refer to this as the *convex hull* restriction. This practically means that the counterfactual will not extrapolate beyond the support of the control group, in contrast to approaches like OLS or matching which suffer from extrapolation biases.

We can instantly notice a few similarities between DID and SCM: for one, some of the weights will be positive (all of them in the case of DID) and both sets of weights add up to 1. However, the key distinction with this method is that the weights are not uniform (i.e., do not need to be the arithmetic average of controls). In practical terms, DID imposes the restriction that the average of all controls are equally valid comparison units. SCM posits that some controls should not matter at all whereas some should matter to varying degrees. Plus, the objective function does not include an intercept [usually, Li and Shankar (2024)], meaning that we force the counterfactual to be strictly on the support of the control group. This means that the identification for SCM is that the counterfactual would simply *be* the weighted average of controls, instead of an intercept shifted average. We usually test for the validity of this assumption based on pre-treatment fit, $\mathbf{y}_1 \approx \mathbf{Y}_0 \mathbf{w}^* \, \forall \, t \in \mathcal{T}_1$. While SCM is a biased estimator, its bias is bounded the extent to which it has quality pre-intervention fit over a long pre-intervention time series (Abadie et al., 2010; Abadie, 2021; Abadie and Vives-i-Bastida, 2022; Ferman and Pinto, 2021).

SCM has become extraordinarily popular in recent years, originally marketed as a tool to conduct comparative case studies in settings where one or a few units were treated in the sample. SCM has been used to study non-profits (Alonso and Andrews, 2020), healthcare provision (Vogler, 2020), immigration (Harris and Jerch, ress), transport policy (Grossi et al., 2024; Wallimann et al., 2023; Li et al., 2022), COVID-19 policy (Friedson et al., 2021; Yang, 2021; Cohn et al., 2022; Cho, 2020), paid family leave (Noh, 2024), marine policy (Fox and Swearingen, 2021), right to work laws (Jordan et al., 2021; Eren and Ozbeklik, 2016), and other fields (Coupet, ress; Gilchrist et al., 2023; Bahar et al., 2024; Krijestorac et al., 2020; Maseland and Spruk, 2023). Much as with DID, numerous theoretical (Zhang et al., 2022a, 2024; Arkhangelsky and Hirshberg, 2023) and practical (Arkhangelsky et al., 2021; Zheng and Chen, 2024; Pang et al., 2022; Xu, 2017; Li and Shankar, 2024) advances have been made with the methodology.

2.1 Limitations of Existing Methods

Control Group Selection When conducting DID or SCM studies, control group selection is critical. At present, DID methods tend to focus on "clean" controls in the sense of not currently treated units to units that were treated before (Roth et al., 2023). However, the specific control group matters too. Suppose a policy analyst is concerned with the impact of some anti-crime initiative upon local robbery rates and wishes to use DID; if she employs a control group whose average difference of the robbery rate is not constant with respect to the treatment group in the pre-treatment period, this violates parallel trends assumption. Of course, analysts recognize this, and often make attempts to circumvent the issue of poor controls. For example. in Liu et al. (2022) the authors exclude already treated cities from their sample, but they construct their control group by picking cities "with similar meteorological and geographical conditions" to the 28 treated cities using a mean comparison. Elsewhere, researchers use alternate control groups as robustness checks (Pan et al., 2020). The selection of controls is even known to matter for spatial DID designs (Ferman, 2023). Some methods have been suggested to get around the problem of parallel trends violations (Callaway and Tsyawo, 2023; Ham and Miratrix, 2024; Rambachan and Roth, 2023; Basu and Small, 2020). However, there are no agreed upon ways to estimate DID models in the presence of parallel trends violations stemming from a poorly constructed control groups.

A consequence of SCM's convex hull restriction is that while the predictions may not extrapolate beyond the support of the control group, interpolation biases can still be quite severe when the control group is too dissimilar from the target unit. Furthermore, high-dimensions also poses a non-uniqueness problem, where many controls may admit a convex representation of the treated unit. The recommendation consistently given to address this is like DID in the sense that it mostly *ad hoc*, where researchers may limiting the units in the control group to ones that are similar to the treated unit (Abadie et al., 2015; Abadie and Vives-i-Bastida, 2022).

However, how would one define similar? Even if we could collect rich covariate data on all of our units of interest, how can we select control units that are similar both in terms of the pre-intervention outcome trend and auxillary covariates? Especially in extremely high dimensions, not uncommon for policy analysis (Yu et al., 2023). As Amjad et al. (2018) notes, we would need an expert in the field to establish some kind of selection criteria for our control units; and even this poses a problem since experts tend to disagree on matters like this. Which covariates should we deem to be most important and why?

Overfitting and Generalizing SCM was developed for low-dimensional settings (Robbins et al., 2017), with an asymptotic bias bound that is valid in the $T_0 >> N_0$ framework. However, in very short pre-intervention periods, SCM can face interpolation biases and overfit to the pre-intervention period. A consequence of overfitting is that the SCM will not give an unbiased estimate of the counterfactual due to high estimation variance, resulting in poor counterfactual predictions. Small pre-intervention periods are quite common in the public policy sphere. For example, Coupet (ress) has 10 pretreatment periods, Noh (2024) has 4 pretreatment periods, and Friedson et al. (2021) has 6. While this does not necessarily mean that SCM is onverfitting in these specific instances, it is certainly a common problem that policy analysts face. Overfitting is also a risk factor in experimental settings where SCM may be used (Abadie and Zhao, 2024).

The same issue of non-uniqueness mentioned above persists in the high dimensional setting too (Abadie and L'Hour, 2021). If we have a very large donor pool and are matching on covariates, there may be many solutions to the optimization problem which poses problems for the interpretability of the SCM's predictions. The reason this is harmful is because in the very first place, synthetic control methods were meant to provide a sparse, interpretable solution to the problem of selecting a control group so as to enable quality post-intervention predictions. If we have a setting where multiple units admit a convex combination of the treated unit's pre-intervention outcomes, then this harms the analysts ability to conduct in-space placebo robustness checks and other quality checks commonly demanded of researchers who employ SCM.

Covariates Use of covariates for both methods is still a kind of thorny issue. Use of covariates in DID models is actually not so simple as is commonly employed in the two way fixed effects setup. For one, the notion of "bad controls" is still an issue in the DID framework (Cinelli et al., 2024), albeit for different reasons than in the literature of sequential ignorability. Parallel trends for DID can hold sometimes unconditionally, but be violated with their inclusion. Regression to the mean is another documented situation where including covariates may lead to violatoins of parallel trends (Daw and Hatfield, 2018). As a result, estimators going beyond two-way fixed effect models have been proposed to tackle this issue (Callaway and Sant'Anna, 2021; Caetano and Callaway, 2024; Caetano et al., 2024). This literature is exciting, and there is not agreement on how to properly incorporate covariates in DID designs as of yet.

The issue is fairly worse for SCM; while SCMs typically include outcome lags, they also typically include additional covariates that predict the pre-intervention time series for the treated unit's outcomes

(Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). The optimization takes the form of

$$\underset{\mathbf{w}\in\mathcal{W}_{\mathrm{SC}}}{\operatorname{argmin}} \|\mathbf{y}_1 - \mathbf{Y}_0\mathbf{w}\|_{\mathbf{V}} = \sqrt{(\mathbf{y}_1 - \mathbf{Y}_0\mathbf{w})^\top \mathbf{V}(\mathbf{y}_1 - \mathbf{Y}_0\mathbf{w})} \tag{2.3}$$

where $\mathbf{V}$ is a diagonal matrix of weights that represent the importance of covariates in predicting the pre-intervention values of the treated unit. Much as with the unit weights, the covariate weights also must sum to 1 and be nonnegative. This practically means that covariates (be they averages of the outcome or other theoretically relevant predictors) directly affect the donor units that are assigned weight. In fact in many cases, including covariates is necessary to converge to a solution (Amjad et al., 2018; Ferman et al., 2020).

However, recent literature shows that this is a much more challenging problem than was first imagined. Upon further investigation, there are complex mathematical issues regarding SCM's optimization (Malo et al., 2023; Albalate et al., 2021; Becker and Klößner, 2018). Oftentimes, the bilevel optimization solved by standard solvers will not converge to the optimal solution, instead offering a corner solution. Differing methods have been proposed to tackle this issue. Malo et al. (2023) for example proposes a solution based on an iterative algorithm, where a non-Archimedean scalar is added to the optimization problem.

Technical issues aside, the above discussion poses an issue regarding SCM's broader applicability: what if the policy analyst does not have access to a rich set of covariates with which to match upon? For example, covariates may be suppressed for privacy reasons in the business context. Germà Bel and Mazaira-Font (2022) study the case of a hotel moratorium on hotel prices in Barcelona. In this case, the treatment is exogenous, but the only data they have access to, provided by Booking.com in this instance, are the price in each city, the date, and whether the city is a Mediterranean city. This was done to preserve the privacy of the cities in question.

Elsewhere, privacy is less of a concern. In Kim (2024), who studies the economic impact of redistributing public institutions on urban congestion in South Korean cities, the treatment assignment was based on three criteria, however the specific contextual variables were kept secret for purposes of fairness and objectivity. In addition to the control group selection problem mentioned above, both of these are examples where potentially useful covariates *cannot* be employed to answer a public policy question. This of course does not mean use of predictors is impossible or never advisable; however given current computational challenges and the inherent subjectivity in covariate selection, policy analysts would benefit from employing SCMs which are less reliant upon covariates in order to generate counterfactuals.

3 Machine Learning as a Tool

Below, I provide brief discussion of key terminology employed in machine-learning methods, in the manner of Athey and Imbens (2019), so the discussion of advances for DID and SCM is straightforward.

3.1 Sample Splitting

One of the key differences between traditional econometrics and machine learning is their respective emphases on parameter estimation and prediction. Econometricians generally care about estimating β, in our case the ATT for some treatment, in an unbiased manner. Machine learning engineers, on the other hand, care about precisely predicting $\widehat{y}$. This leads to the notion of splitting one's dataset into a *training*/in sample dataset versus a *test*/out of sample dataset. In the training dataset, the policy analyst considers a particular model of interest. Then then estimate the model many times according to some criteria, a process called *cross validation*, and use the best of these training models to predict the values of a test dataset, or the dataset that the purification training models were unexposed to.

The same notions extend very easily to panel data, where we denote the training dataset as $\mathcal{D}_{\text{train}} \coloneqq \{\mathbf{y}_{1,\mathcal{T}_1}, \mathbf{Y}_{0,\mathcal{T}_1}\}$ and test dataset as $\mathcal{D}_{\text{test}} \coloneqq \{\mathbf{y}^{?}_{1,\mathcal{T}_2}, \mathbf{Y}_{0,\mathcal{T}_2}\}$. In this setting, we consider estimating the causal effect of interest as a prediction problem, where we denote $\mathbf{y}^{?}_{1,\mathcal{T}_2}$ as the out of sample predictions made by either DID or SCM. In a sense, we are using machine learning's focus on prediction "in service" of treatment effect estimation (Mullainathan and Spiess, 2017). Since the models are estimated only by using the data from $\mathcal{D}_{\text{train}}$, the idea is that out post-intervention predictions, the ones we care most about, will not be affected by any post-intervention confounders. Note that by construction, DID and SCM already function like this. After all, their respective optimizations are done over the pre-intervention period only, and the estimated weights the analyst derives are used to predict the post-intervention/out of sample counterfactual. In fact, the idea of splitting the sample and employing cross validation was first argued for, in the context of SCM, by Abadie et al. (2015). In their context, cross-validation was used to choose the covariate/importance weights such that they could better predict the out of sample values of West Germany.

Part of the goal of sample splitting is to avoid overfitting too much to idiosyncratic trends in the pre-intervention period. One way machine learning models tend to avoid overfitting is by employing what is called regularization for feature selection. This is the subject of the next section.

3.2 Feature Selection

By now, I've discussed the identification for both DID and SCM, which respectively are parallel trends and a convex combination of controls being able to fit as closely as possible to the pre-treatment time series. I've also discussed the problems associated with poor control groups in both settings. This means that policy analysis may be augmented by machine learning's feature selection capabilities. In our case, instead of a generic variable selection problem, we are more precisely concerned with a *control unit* selection problem. To formalize this, we work within the framework of

Assumption 1. *We suppose that our counterfactual may be expressed via the linear model*

$$y_{1t}^0 = \mathbf{Y}_0(?)\mathbf{w} + \beta_0 + \epsilon_t \tag{3.1}$$

where $\mathbf{w} \in [0, 1]$ *is some generic weight vector and* $\beta_0 \in \mathbb{R}$ *is some constant that may take any real value.*

The question, hence the question mark for the control matrix, is "*what control units should we include in our regression model?*". In the case of DID we are interested in selecting controls such that we satisfy, as best as possible, the parallel trends assumption. For SCM, we wish to select controls such that our treatment unit may best be expressed as a convex combination of control units, limiting the control group as recommended by Abadie (2021). In this section, I will go over two ways of doing this, broadly referred to as penalization and dimensionality reduction. The former implies some statistical learning penalty that is within the estimation of the DID/SCM model, whereas the latter refers to conducting preprocessing prior to estimation.

Statistical Learning Penalty To select controls, we can employ statistical learning penalties

$$\begin{aligned}
&\text{Lasso:} && \underset{\mathbf{w} \in \mathbb{R}^{N_0}}{\operatorname{argmin}} \; \|\mathbf{y}_1 - \mathbf{Y}_0\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1 \\
&\text{Ridge:} && \underset{\mathbf{w} \in \mathbb{R}^{N_0}}{\operatorname{argmin}} \; \|\mathbf{y}_1 - \mathbf{Y}_0\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_2^2 \\
&\text{Elastic Net:} && \underset{\mathbf{w} \in \mathbb{R}^{N_0}}{\operatorname{argmin}} \; \|\mathbf{y}_1 - \mathbf{Y}_0\mathbf{w}\|_2^2 + \lambda_1\|\mathbf{w}\|_1 + \frac{(1-\lambda_1)}{2}\|\mathbf{w}\|_2^2
\end{aligned}$$

The key thing to note here, is the properties of the ℓ_1 and ℓ_2 norms.[2] The former is the Least Absolute Shrinkage and Selection Operator, or the LASSO. The LASSO encourages a sparse solution by penalizing

[2] The technical details of these may be found elsewhere (Kinn, 2018); for a review directed toward public policy scholars specifically, see Lechner (2023) and Chen et al. (2021).

higher order coefficients, shrinking the least predictive ones to zero and keeping the ones with the highest predictive power. In other words, $\mathbf{Y}_0(?)$ is replaced by the control group selected by the LASSO. The second is the "Ridge" penalty, which does not shrink any controls to 0, but instead reduces the least predictive ones towards zero. The final one is the elastic net, which is a combination of both penalties. Here, some elements of $\mathbf{Y}_0(?)$ will be zero and others will not, seeking to trade off sparsity and overfitting. In all three cases, the degree to which sparsity is emphasized is mitigated by the lambda hyperparameter, or λ. Note that we can still place constraints upon the weights if we wish. These methods have seen some use among public policy scholars in the context of synthetic controls (Duncan et al., 2025; Fantoni, na). DID, in contrast, has seen less emphasis on statistical learning penalties, however some scholarship on it exists (Chang, 2020; Zhang et al., 2022b).

Dimensionality Reduction Another approach which machine learning offers for policy analysts is the dimensional reduction approach. The above methods impose hard constraints on the weighting scheme; here, controls are removed entirely prior to the estimation being conducted. One such way is principal component analysis, or PCA. For a simple introduction to PCA, interested readers may consult Saccenti (2024). PCA essentially estimates a set of scores/singular values that explain a given percentage of the variance of a set of variables, returning the matrix $\mathbf{Y}_0^k$. In our case, this would be the pre-intervention control group matrix that have been reduced to its principal components (also called a low rank structure). We can have as many scores as there are control units, and then use these in regression to assign weights to the control group. Typically rules of thumb, such as the "elbow method" are used to select the number of scores to use.

The practical meaning of this is that these PC scores would penalize the donors differently, focusing on their "signal" (i.e., the ones which explain most of the pre-intervention variation) as opposed to noise. One of the methods which we will discuss in more detail below, developed by Amjad et al. (2018) and Agarwal et al. (2021), uses this exact method. As far as I am aware, this approach has been developed mostly for SCM, with more limited application for DID methods (Chan and Kwok, 2022). A closely related technique, called "Robust" PCA, was also employed by Bayani (2021) for the SCM, and this method is also discussed below. Of course, PCA based methods are but a few of the recent advances which policy analysts may use. Other control group selection mechanisms are also possible. Another one worthy of mention (the first method will discuss in terms of new estimators) is forward selection. This was developed for the DID method by Li (2024) and implemented for Stata by Greathouse et al. (2024).

4 Recent Advances in DID and SCM

4.1 Forward Difference-in-Differences

Description Forward Difference-in-Differences (FDID) is a novel DID estimator which was proposed by Li (2024). FDID uses an iterative forward selection method to choose the control group. It starts by estimating one DID model per each control unit (if we have 50 controls, we have 50 one control unit models). We then see which control unit maximizes the pre-intervention R^2 statistic. This becomes the first selected control group. Then, we estimate $N_0 - 1$ two control unit DID models. In these, we include the first selected unit and the remaining unselected controls as predictors. The model which maximizes the pre-intervention R-squared statistic, in conjunction with the first selected control, becomes the second selected control group. We continue with this process until we have N_0 "selected" control groups. The final model FDID uses is whichever of the $1, 2 \ldots N - 1$ "selected" models maximizes the pre-intervention R^2 statistic. That control group is the one employed by FDID, and naturally is the one used to estimate the ATT.

Utility The first obvious benefit to public policy researchers is FDID's control group selection algorithm. As I have discussed above, selection of a control group is critical to causal analysis, in this case for satisfying the parallel trends assumption. For example, in the paper by Greathouse et al. (2024) which develops FDID for Stata, the authors employ the classic example of Proposition 99 used by Abadie et al. (2010). The FDID selection method selects the same controls as the SCM does, except for Utah. In addition, FDID gets a lower pre-intervention Root Mean Squared Error than the SCM for that specific example.

The additional problem that FDID can solve, in certain cases, is its robustness to lack of covariate information. Abadie et al. (2010) develop SCM precisely because the normal DID method's unconditional parallel trends assumption does not hold. In the Prop 99 example, Abadie et al. (2010) use four covariates to aid the SCM in its choice of control group selection. In contrast, FDID uses none in the selection of controls nor its estimation of the ATT. The selection and estimation are completely data driven based on the outcome of interest. Thus, FDID enables policy analysts to conduct program evaluation in cases without covariates while enabling the selection of a control group. FDID also requires no tuning parameters, unlike now complicated procedures. Of course, the validity of FDID depends on if the difference between the treated unit and a selected subset of controls would be constant absent the treatment.

4.2 Principal Component Regression

Description Principal Component Regression (PCR) is form of SCM proposed by Agarwal et al. (2021). It is a variant of the *Robust Synthetic Control* estimator by Amjad et al. (2018). To my knowledge, PCR is among the first form of SCM to incorporate machine-learning methodologies in its estimation process. PCR relies on extracting the *low rank structure* of the pre-intervention donor matrix using PCA (Udell and Townsend, 2019; Lin, 2016). We denote this matrix as $\mathbf{L}$, estimated via singular value decomposition

$$\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top} \tag{4.1}$$

where $\mathbf{U} \in \mathbb{R}^{N \times N}$, $\mathbf{\Sigma} \in \mathbb{R}^{N \times p}$, and $\mathbf{V} \in \mathbb{R}^{p \times p}$. To select the top number of singular values to mitigate the tradeoff between bias and variance, Agarwal et al. (2021) employ Universal Singular Thresholding as discussed in Chatterjee (2015). As with other SCMs, PCR applies a weighting scheme to its control units which in this case may be any real number, $\mathcal{W}_{\text{PCR}} \coloneqq \{\mathbf{w} \in \mathbb{R}^{N_0}\}$. PCR weighs donors via

$$\underset{\mathbf{w} \in \mathcal{W}_{\text{PCR}}}{\operatorname{argmin}} \; \|\mathbf{y}_1 - \mathbf{L}\mathbf{w}\|_2^2 \tag{4.2}$$

where the unit weights are learnt via OLS. Like for the other methods, the estimation of the weights is carried our for the pre-intervention period only. Then, we use those weighted to estimate the out of sample counterfactual predictions.

Utility Agarwal et al. (2021) note a few benefits of PCR which are quite useful for policy analysts, stemming from PCR's regularization procedure. For one, PCR is robust to noise and missing data, two major limitations of the original SCM. The low-rank estimation, aided by the thresholding approach, mitigates the dangers of overfitting, thus increasing the validity of our out of sample predictions. PCR also implicitly does control group selection. While the authors advocate for OLS (where all control units receive weight), we also may employ the convex hull constraints should we desire a sparse solution. Another benefit, as Agarwal et al. (2021) and Amjad et al. (2018) note, is that the method does not require covariates to obtain similar out of sample predictions as from Abadie and Gardeazabal (2003) and Abadie et al. (2010). The de-biasing procedure enables policy analysts to mitigate concerns about noise while using a robust algorithm to estimate the out of sample counterfactual.

4.3 Robust PCA SCM

Description Robust PCA Synthetic Control (RPCA-SCM) was developed by Bayani (2021) as a direct extension of Amjad et al. (2018) and the matrix completion method by Athey et al. (2021). This estimator is broken into two steps. The first step is control group selection. First, we calculate the functional PCA scores of the outcome matrix for all units [see Bayani (2021) for the technical details of this] using the pre-intervention time series of outcome data. We then use k-means clustering over this low-dimensional representation, where each unit (including the treatment unit) falls into a cluster based on its low-rank structure. We then keep the cluster which contains the treated unit. Next, we employ the Robust PCA method to separate the selected control group into its low rank structure versus it sparse/noise component. To do this, we solve the following optimization problem

$$\begin{aligned} \underset{\mathbf{L},\,\mathbf{S}}{\text{minimize}} \quad & \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \\ \text{subject to} \quad & \mathbf{Y} = \mathbf{L} + \mathbf{S}. \end{aligned} \tag{4.3}$$

where $\mathbf{L}$ and $\mathbf{S}$ respectively are the low-rank structure and noise components of the donor matrix. As before, the low-rank structure simply represents a simplified projection of the temporal and cross sectional patterns in our dataset. Using this, we can estimate our weights, $\mathcal{W}_{\text{RPCA}} \coloneqq \left\{ \mathbf{w} \in \mathbb{R}_{\geq 0}^{N_0} \right\}$, via least-squares

$$\underset{\mathbf{w} \in \mathcal{W}_{\text{RPCA}}}{\text{argmin}} \quad \|\mathbf{y}_1 - \mathbf{L}\mathbf{w}\|_2^2 \tag{4.4}$$

Utility RPCA-SC presents numerous benefits for policy analysts. For one, it has an explicit control group selection step that is independent of the method we use to estimate the unit weights. The clustering method groups together the treated and control units by their pre-intervention similarities. As far as I am aware, this is a novelty in the literature. Additionally, RPCA-SC is robust to noise and missing entries in the outcome matrix as a result of the RPCA method, more so than normal PCA because PCA is easily corrupted by outliers. Like the other methods we have discussed, RPCA-SC is also flexible in the sense that it does not demand the policy analyst collect covariate data in order to obtain similar out of sample predictions to those of Abadie et al. (2015). A unique element of RPCA-SC is that it is more robust to changes in the base model. Bayani (2021) shows that the estimated treatment effect of RPCA-SC is less sensitive to in-time/in-space placebos and dropping the originally selected donors than the classic SCM or the PCR method.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

Abadie, A., Diamond, A., and 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.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132.

Abadie, A. and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536):1817–1834.

Abadie, A. and Vives-i-Bastida, J. (2022). Synthetic controls in action.

Abadie, A. and Zhao, J. (2024). Synthetic controls for experimental design.

Abouk, R., Ghimire, K. M., Maclean, J. C., and Powell, D. (2023). Pain management and work capacity: Evidence from workers' compensation and marijuana legalization. *Journal of Policy Analysis and Management*, 42(3):737–770.

Agarwal, A., Shah, D., Shen, D., and Song, D. (2021). On robustness of principal component regression. *Journal of the American Statistical Association*, 116(536):1731–1745.

Albalate, D., Bel, G., and Mazaira-Font, F. A. (2021). Decoupling synthetic control methods to ensure stability, accuracy and meaningfulness. *SERIEs*, 12(4):549–584.

Alonso, J. M. and Andrews, R. (2020). Government-Created Nonprofit Organizations and Public Service Turnaround: Evidence from a Synthetic Control Approach. *Journal of Public Administration Research and Theory*, 31(2):346–362.

Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *The Journal of Machine Learning Research*, 19(1):802–852.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

Arkhangelsky, D. and Hirshberg, D. (2023). Large-sample properties of the synthetic control method under selection on unobservables.

Ashworth, S., Berry, C. R., and de Mesquita, E. B. (2021). *Theory and Credibility: Integrating Theoretical and Empirical Social Science.* Princeton University Press, Princeton.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536):1716–1730.

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(Volume 11, 2019):685–725.

Bahar, D., Hauptmann, A., Özgüzel, C., and Rapoport, H. (2024). Migration and Knowledge Diffusion: The Effect of Returning Refugees on Export Performance in the Former Yugoslavia. *The Review of Economics and Statistics*, 106(2):287–304.

Basu, P. and Small, D. S. (2020). Constructing a more closely matched control group in a difference-in-differences analysis: its effect on history interacting with group bias. *Observational Studies*, 6(1):103–130.

Bayani, M. (2021). Robust pca synthetic control. working paper.

Becker, M. and Klößner, S. (2018). Fast and reliable computation of generalized synthetic controls. *Econometrics and Statistics*, 5:1–19.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, page rdae007.

Caetano, C. and Callaway, B. (2024). Difference-in-differences when parallel trends holds conditional on covariates.

Caetano, C., Callaway, B., Payne, S., and Rodrigues, H. S. (2024). Difference in differences with time-varying covariates.

Callaway, B. and Sant'Anna, P. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

Callaway, B. and Tsyawo, E. S. (2023). Treatment effects in staggered adoption designs with non-parallel trends.

Cawley, J. and Frisvold, D. E. (2017). The pass-through of taxes on sugar-sweetened beverages to retail prices: The case of berkeley, california. *Journal of Policy Analysis and Management*, 36(2):303–326.

Chan, M. K. and Kwok, S. S. (2022). The pcdid approach: Difference-in-differences when trends are potentially unparallel and stochastic. *Journal of Business & Economic Statistics*, 40(3):1216–1233.

Chang, N.-C. (2020). Double/debiased machine learning for difference-in-differences models. *The Econometrics Journal*, 23(2):177–191.

Chatterjee, S. (2015). Matrix estimation by Universal Singular Value Thresholding. *The Annals of Statistics*, 43(1):177 – 214.

Chen, J. C., Rubin, E. A., and Cornwall, G. J. (2021). *Data science for public policy*. Springer.

Cho, S.-W. S. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. *The Econometrics Journal*, 23(3):323–344.

Churchill, B. F., Henkhaus, L. E., and Lawler, E. C. (na). Effect of vaccine recommendations on consumer and firm behavior. *Journal of Policy Analysis and Management*, in press(in press):na.

Cinelli, C., Forney, A., and Pearl, J. (2024). A crash course in good and bad controls. *Sociological Methods & Research*, 53(3):1071–1104.

Cohn, E., Chimowitz, M., Long, T., Varma, J. K., and Chokshi, D. A. (2022). The effect of a proof-of-vaccination requirement, incentive payments, and employer-based mandates on covid-19 vaccination rates in new york city: a synthetic-control analysis. *The Lancet Public Health*, 7(9):e754–e762.

Connolly, L., Hampton, M., and Lenhart, O. (2024). Labor mobility and the affordable care act: Heterogeneous impacts of the preexisting conditions provision. *Journal of Policy Analysis and Management*, 43(1):157–191.

Coupet, J. (inpress). To make or buy commercialization: A synthetic control approach. *Journal of Policy Analysis and Management*, na(na):1–25.

Courtemanche, C., Marton, J., Ukert, B., Yelowitz, A., and Zapata, D. (2017). Early impacts of the affordable care act on health insurance coverage in medicaid expansion and non-expansion states. *Journal of Policy Analysis and Management*, 36(1):178–210.

Dague, L. and Lahey, J. N. (2018). Causal Inference Methods: Lessons from Applied Microeconomics. *Journal of Public Administration Research and Theory*, 29(3):511–529.

Daw, J. R. and Hatfield, L. A. (2018). Matching and regression to the mean in difference-in-differences analysis. *Health Services Research*, 53(6):4138–4156.

de Chaisemartin, C. and D'Haultfœuille, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*, 26(3):C1–C30.

de Chaisemartin, C. and D'Haultfœuille, X. (2024). Difference-in-Differences Estimators of Intertemporal Treatment Effects. *The Review of Economics and Statistics*, pages 1–45.

Duncan, D. R., Ross, J. M., and Mikesell, J. L. (2025). Can the federal budget process promote fiscal sustainability? evidence from the budget enforcement act of 1990. *Public Finance Review*, 53(1):29–61.

Eren, O. and Ozbeklik, S. (2016). What do right-to-work laws do? evidence from a synthetic control method analysis. *Journal of Policy Analysis and Management*, 35(1):173–194.

Fantoni, M. C. (na). At their fingertips: What is the impact of online reporting of domestic violence? *Public Administration Review*, n/a(n/a).

Ferman, B. (2023). Inference in difference-in-differences: How much should we trust in independent clusters? *Journal of Applied Econometrics*, 38(3):358–369.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221.

Ferman, B., Pinto, C., and Possebom, V. (2020). Cherry picking with synthetic controls. *Journal of Policy Analysis and Management*, 39(2):510–532.

Fox, H. K. and Swearingen, T. C. (2021). Using a difference-in-differences and synthetic control approach to investigate the socioeconomic impacts of oregon's marine reserves. *Ocean & Coastal Management*, 215:105965.

Friedson, A. I., McNichols, D., Sabia, J. J., and Dave, D. (2021). Shelter-in-place orders and public health: Evidence from california during the covid-19 pandemic. *Journal of Policy Analysis and Management*, 40(1):258–283.

Germà Bel, S. J. and Mazaira-Font, F. A. (2022). The effects of moratoriums on hotel building: an anti-tourism measure, or rather protection for local incumbents? *Applied Economics Letters*, 29(14):1319–1324.

Gihleb, R., Giuntella, O., and Tan, J. Q. (2024). The impact of right-to-work laws on long hours and work schedules. *Journal of Policy Analysis and Management*, 43(3):696–713.

Gilchrist, D., Emery, T., Garoupa, N., and Spruk, R. (2023). Synthetic control method: A tool for comparative case studies in economic history. *Journal of Economic Surveys*, 37(2):409–445.

Goldhaber, D., Imberman, S. A., Strunk, K. O., Hopkins, B. G., Brown, N., Harbatkin, E., and Kilbride, T. (2022). To what extent does in-person schooling contribute to the spread of covid-19? evidence from michigan and washington. *Journal of Policy Analysis and Management*, 41(1):318–349.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Greathouse, J., Coupet, J., and Sevigny, E. (2024). Greed is good: Estimating forward difference-in-differences in stata. [working paper].

Grossi, G., Mariani, M., Mattei, A., Lattarulo, P., and Öner, O. (2024). Direct and spillover effects of a new tramway line on the commercial vitality of peripheral streets: a synthetic-control approach. *Journal of the Royal Statistical Society Series A*, page qnae032.

Ham, D. W. and Miratrix, L. (2024). Benefits and costs of matching prior to a difference in differences analysis when parallel trends does not hold. *The Annals of Applied Statistics*, 18(3):2096 – 2122.

Harris, J. M. and Jerch, R. (press). The long-run effects of temporary protection from deportation. *Journal of Policy Analysis and Management*, na(na).

Heckman, J. and Pinto, R. (2024). Econometric causality: The central role of thought experiments. *Journal of Econometrics*, 243(1):105719.

Heckman, J. J. and Pinto, R. (2022). The econometric model for causal policy analysis. *Annual Review of Economics*, 14(Volume 14, 2022):893–923.

Imbens, G. W. (2024). Causal inference in the social sciences. *Annual Review of Statistics and Its Application*, 11(Volume 11):123–152.

Jordan, J., Mathur, A., Munasib, A., and Roy, D. (2021). Did right-to-work laws impact income inequality? evidence from u.s. states using the synthetic control method. *The B.E. Journal of Economic Analysis & Policy*, 21(1):45–81.

Kim, Y. (2024). Assessing the economic impact of innovative cities. *Policy Studies*, 0(0):1–25.

Kinn, D. (2018). Synthetic control methods and big data. [working paper].

Krijestorac, H., Garg, R., and Mahajan, V. (2020). Cross-platform spillover effects in consumption of viral content: A quasi-experimental analysis using synthetic controls. *Information Systems Research*, 31(2):449–472.

Lechner, M. (2023). Causal machine learning and its use for public policy. *Swiss Journal of Economics and Statistics*, 159(1):8.

Li, K. T. (2024). Frontiers: A simple forward difference-in-differences method. *Marketing Science*, 43(2):239–468.

Li, K. T. and Shankar, V. (2024). A two-step synthetic control approach for estimating causal effects of marketing events. *Management Science*, 70(6):3734–3747.

Li, X., Wu, Y.-J., and Khani, A. (2022). Investigating a small-sized bike-sharing system's impact on transit usage: a synthetic control analysis in tucson, arizona. *Public Transport*, 14(2):441–458.

Lin, Z. (2016). A review on low-rank models in data analysis. *Big Data & Information Analytics*, 1(2):139–161.

Liu, H., Wang, C., Zhang, M., and Wang, S. (2022). Evaluating the effects of air pollution control policies in china using a difference-in-differences approach. *Science of The Total Environment*, 845:157333.

Malo, P., Eskelinen, J., Zhou, X., and Kuosmanen, T. (2023). Computing synthetic controls using bilevel optimization. *Computational Economics*.

Maseland, R. and Spruk, R. (2023). The benefits of us statehood: an analysis of the growth effects of joining the usa. *Cliometrica*, 17(1):49–89.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

Myers, C. (2024). Forecasts for a post-roe america: The effects of increased travel distance on abortions and births. *Journal of Policy Analysis and Management*, 43(1):39–62.

Noh, K. (2024). Paid family leave policy and leave-taking: a synthetic control approach. *Applied Economics Letters*, 0(0):1–5.

Pac, J. and Berger, L. M. (2024). Quasi-experimental evidence on the employment effects of the 2021 fully refundable monthly child tax credit. *Journal of Policy Analysis and Management*, 43(1):192–213.

Pan, X., Li, M., Wang, M., Zong, T., and Song, M. (2020). The effects of a smart logistics policy on carbon emissions in china: A difference-in-differences analysis. *Transportation Research Part E: Logistics and Transportation Review*, 137:101939.

Pang, X., Liu, L., and Xu, Y. (2022). A bayesian alternative to synthetic control for comparative case studies. *Political Analysis*, 30(2):269–288.

Pichler, S., Wen, K., and Ziebarth, N. R. (2021). Positive health externalities of mandating paid sick leave. *Journal of Policy Analysis and Management*, 40(3):715–743.

Rambachan, A. and Roth, J. (2023). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*, 90(5):2555–2591.

Robbins, M. W., Saunders, J., and Kilmer, B. (2017). A framework for synthetic control methods with high-dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention. *Journal of the American Statistical Association*, 112(517):109–126.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244.

Saccenti, E. (2024). A gentle introduction to principal component analysis using tea-pots, dinosaurs, and pizza. *Teaching Statistics*, 46(1):38–52.

Shi, J., Ai, X., and Cao, Z. (2017). Can big data improve public policy analysis? In *Proceedings of the 18th Annual International Conference on Digital Government Research*, dg.o '17, page 552–561, New York, NY, USA. Association for Computing Machinery.

Slopen, M. (2024). The impact of paid sick leave mandates on women's employment and economic security. *Journal of Policy Analysis and Management*, 43(4):1129–1151.

Spirtes, P. (2010). Introduction to causal inference. *Journal of Machine Learning Research*, 11(5).

Udell, M. and Townsend, A. (2019). Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160.

Viviano, D. and Bradic, J. (2023). Synthetic learner: Model-free inference on treatments over time. *Journal of Econometrics*, 234(2):691–713.

Vogler, J. (2020). Access to healthcare and criminal behavior: Evidence from the aca medicaid expansions. *Journal of Policy Analysis and Management*, 39(4):1166–1213.

Wallimann, H., Blättler, K., and von Arx, W. (2023). Do price reductions attract customers in urban public transport? a synthetic control approach. *Transportation Research Part A: Policy and Practice*, 173:103700.

Willage, B., Carlos, M., and Callison, K. (2023). Non-monetary obstacles to medical care: Evidence from postpartum contraceptives. *Journal of Policy Analysis and Management*, 42(4):1045–1064.

Wing, C., Freedman, S. M., and Hollingsworth, A. (2024a). Stacked difference-in-differences. Working Paper 32054, National Bureau of Economic Research.

Wing, C., Yozwiak, M., Hollingsworth, A., Freedman, S., and Simon, K. (2024b). Designing difference-in-difference studies with staggered treatment adoption: Key concepts and practical guidelines. *Annual Review of Public Health*, 45(Volume 45, 2024):485–505.

Wooldridge, J. M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, 26(3):C31–C66.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.

Yang, X. (2021). Does city lockdown prevent the spread of covid-19? new evidence from the synthetic control method. *Global Health Research and Policy*, 6:1–14.

Yu, L. B., Tran, T. M., and Lee, W.-S. (2023). Bridging the gap: Assessing the effects of railway infrastructure investments in northwest china. *China Economic Review*, 82:102076.

Zhang, L., Zhang, X., and Zhang, X. (2024). Asymptotic properties of the distributional synthetic controls.

Zhang, X., Wang, W., and Zhang, X. (2022a). Asymptotic properties of the synthetic control method.

Zhang, Y., Li, H., and Ren, G. (2022b). Quantifying the social impacts of the london night tube with a double/debiased machine learning based difference-in-differences approach. *Transportation Research Part A: Policy and Practice*, 163:288–303.

Zheng, X. and Chen, S. X. (2024). Dynamic synthetic control method for evaluating treatment effects in auto-regressive processes. *Journal of the Royal Statistical Society Series B*, 86(1):155–176.

Appendix to
"Rise of the Machines: Causal Policy Analysis, Modern Econometrics, and Machine Learning"

Jared Greathouse[a]

[a]Georgia State University

Section A includes the formal notations I use in the paper.

A Notation

Let ":=" be a definition. A scalar is an italicized lowercase letter, y. A vector is a bold lowercase letter $\mathbf{y}$. Let a matrix be a bold uppercase letter $\mathbf{Y}_0$. Let the cardinality of a discrete set be $|\cdot|$, say $S = |S|$. For any vector, say $\mathbf{w}$, let $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the standard ℓ_1 and ℓ_2 norms, respectively, where $\|\mathbf{w}\|_1 := \sum_i |w_i|$ and $\|\mathbf{w}\|_2 := \left(\sum_i w_i^2\right)^{0.5}$. Indexed by j, we observe $\mathcal{N} := \{1, 2, \ldots, N\}$ units where the set $\mathcal{N}$ has cardinality $N = |\mathcal{N}|$. $j = 1$ is the treated unit with the controls being $\mathcal{N}_0 := \mathcal{N} \setminus \{1\}$ whose cardinality is $N_0 = |\mathcal{N}_0|$. 1 Time periods are indexed by t. Let $\mathcal{T}_1 := \{1, 2, \ldots, T_0\}$ represent the pre-intervention periods, where T_0 is the final pre-intervention period, and $\mathcal{T}_2 := \{T_0 + 1, \ldots, T\}$ represents the post-intervention periods. Both of these sets have cardinalities $T_1 = |\mathcal{T}_1|$ and $T_2 = |\mathcal{T}_2|$. Let $\mathcal{T} := \mathcal{T}_1 \cup \mathcal{T}_2$ represent the full time series, with cardinality $T = |\mathcal{T}|$. Let $\mathbf{y}_j := [y_{jt}, \ldots, y_{jT}]^\top \in \mathbb{R}^T$ denote the generic vector of outcomes for unit $j \in \mathcal{N}$, where y_{jt} represents the outcome for unit j at time $t \in \mathcal{T}$. Furthermore, denote $\mathbf{y}_1 := (y_{1t})_{t \in \mathcal{T}}$ as the $T \times 1$ column vector of outcomes for the treated unit and $\mathbf{Y}_0 := (\mathbf{y}_j)_{j \in \mathcal{N}_0}$ as the $T \times N_0$ matrix of control unit outcomes. I define $\mathbf{w} := [w_2, \ldots, w_N] \in \mathbb{R}^{N_0}$ as some vector of unit weights, and $\mathcal{W}$ as a generic set which defines the support of the weights. Let the treatment effect be $\widehat{\Delta}_t := y_{1t} - \widehat{y}^0_{1t}$, where $\widehat{y}^0_{1t}$ is the estimated counterfactual. The causal estimand of interest is the ATT, defined as $\widehat{ATT} = T_2^{-1} \sum_{t \in \mathcal{T}_2} \widehat{\Delta}_t$.