Ellipsoidal Optimal Recovery: A Minimax Approach to Robust Counterfactual Estimation

Anonymous authors
Paper under double-blind review

Abstract

Consider the problem of quantifying the causal effects of an intervention to determine whether the intervention achieved desired outcomes. Researchers address this problem using statistical, machine learning, or signal processing techniques that have limitations of high bias or need of expert knowledge. We present a new minimax geometric approach called *ellipsoidal optimal recovery (EOpR)* for estimating the unobservable outcome of a treatment unit. It is an approximation-theoretic technique that recovers unknown observations given a learned signal/principal vector and a set of known observations. The significance of our approach is that it improves pre-treatment fit and mitigates bias of the post-treatment estimate relative to other methods in causal inference. Beyond recovery of the unit of interest, an advantage of EOpR is that it produces worst-case limits over the estimates produced. We assess our approach on synthetically-generated data, on standard datasets commonly used in the econometrics (synthetic control) literature, and in the context of the COVID-19 pandemic, showing better performance than baseline techniques.
Keywords: causal inference, optimal recovery, synthetic control

1 Introduction

A big part of policy evaluation is to estimate the effects of an implemented policy, so as to know whether it achieved its goals. Estimating effects further yields inferences of causal relationships between interventions and their outcomes. Quantifying the effect of a treatment has been of interest not only in policymaking but also across different domains in health sciences, social sciences, and engineering. Typically, the effect is measured by looking at the difference between outcomes before and after an intervention of a treated unit. However, for a given object (e.g., geographical region) at a given time, only one of the outcomes is observed and not both. Thus, we aim to recover and estimate the outcome that was not observed.

Several causal inference methods have been developed for observational studies to estimate the unobserved outcomes for a given intervention. For example *synthetic control* (SC) by Abadie (2021) constructs a weighted average of control units to act as a synthetic control unit to compare with the treated unit. Recently, there has been a growing literature that approaches causal inference from a matrix completion perspective. Proposals include approximating the control unit matrix using nuclear-norm minimization (Athey et al., 2021), using singular value decomposition (Amjad et al., 2018), or by finding nearest neighbors (Agarwal et al., 2021) for missing entries of a matrix to best match control units and the treated unit of interest.

A main limitation in previous work is that when there is insufficient data, especially data from only a short period of time, methods are unable to recover the true estimate (Amjad et al., 2019). Further, the insufficiency or the low quality of data tend to create a poor pre-treatment fit which is a main source of bias in estimates in SC (Abadie, 2021).

Our Contributions. This work has three main contributions: algorithmic, theoretical, and empirical contributions.

Algorithmic. Our first contribution is the development of an approximation-theoretic approach called *ellipsoidal optimal recovery* (EOpR) that minimizes the worst-case error to optimally recover the unobserved

outcome. This also addresses the problem of insufficiency of data, specifically with short pre-intervention period. Taking inspiration from signal processing, we adapt the *optimal recovery* algorithm (Muresan & Parks, 2004) to recover missing outcomes. Unlike statistical approaches (Abadie & Gardeazabal, 2003; Abadie et al., 2010) or low-rank matrix decomposition approaches (Amjad et al., 2018; 2019; Athey et al., 2021) which minimize the average error, our approach minimizes the maximum error over the known samples, which is intended to make EOpR a robust algorithm, among other properties. It has been previously proven that minimizing the maximum error produces robust and low-bias estimators (Kassam & Poor, 1985; Muresan & Parks, 2005). To do so, we assume signals are from a particular ellipsoidal signal class, which is a common assumption in economics (Armstrong & Kolesár, 2018).

Theoretical. Our second contribution is the derivation of desired properties given the geometrical assumptions underlying EOpR. Having the ellipsoidal assumption allows us to make use of various useful properties: (a) The ellipsoidal assumption creates a quadratic closed form solution. (b) The optimization problem is min-max optimization by the definition of the circle/ellipsoid as we see throughout the paper. (c) It ensures that the solution is bounded given the boundary conditions of the ellipsoid, such that we prove the consistency of the estimator and the derivation of worst-case estimates of outcomes given the geometrical properties of the algorithm.

All such properties are novel to the causal inference literature and have proven to be desirable. Synthetic control literature has been mainly based on regression-style solutions with a minimization objective only and never considers min-max objective with worst-case estimation.

Empirical. Our third contribution is the extensive evaluation of our algorithm and other baselines. We conduct two sets of experiments: (a) on synthetically generated data following Amjad et al. (2018); (b) on existing case studies from real-world datasets used by Abadie & Gardeazabal (2003); Abadie et al. (2010); Abadie (2021) and additional real-world COVID-19 experiment that we developed. Synthetic data helps validate the efficacy of our method, since it is impossible to simultaneously observe the treated unit and its counterfactual. Our algorithm outperforms baseline methods especially under varying lengths of pre-treatment periods.

Organization of the paper. Section 2 reviews some aspects of causal inference and related work. Section 3 formally states our problem setup. Section 4 describes our estimation strategy in the context of comparative studies data. Section 5 supports the efficacy of our approach via artificial and empirical experiments.

2 Related Work

To fix concepts and common terms, we briefly overview causal inference and discuss related work.

Causal analysis takes a step beyond standard statistical analysis, inferring beliefs under changing conditions to uncover causal relationships among variables (Pearl, 2009). Several frameworks have been proposed to tackle causality analysis, such as structural models (Pearl, 2010) and the potential outcome framework (Rubin, 1974), which we focus on here.

2.1 Potential Outcomes Framework

This framework assumes effects are tied to a treatment or an intervention. To reveal the causal effects of an intervention, Rubin (1974) proposed to measure the difference of two potential outcomes. The outcome for a unit without being exposed to an intervention and the outcome after an intervention is applied. So, the causal effect is the difference between the two outcomes. However, in real applications, we can never observe both outcomes for the same unit under the same conditions, as only one of the two will take place at a given time. Therefore, one of the potential outcomes will always be *missing*, and the core objective of the framework to estimate the missing outcome.

Let us introduce the main terms used in the potential outcomes literature, which are used throughout. A *unit* is the atomic object in the framework, which can be any object, whether it is a patient or a city. A

treatment[1] is the action applied on a unit to change its state, whether administering medicine or enforcing a lockdown order. Treatment is usually thought of as binary, so one group receives the treatment (the *treated* group) and the other does not (the *control* group). *Panel data* is another word for cross-sectional time-series data and comparative studies.

2.2 Synthetic Control Methods

Synthetic control (SC) is a method that proposes a particular way to measure the missing observable potential outcome to estimate causal effects (Abadie et al., 2010). Instead of using a single control unit or a simple average of a set of control units, SC creates a *synthetic* unit to act as a control group by selecting appropriate weights for selected control units. The choice of weights should result in a synthetic control unit that best resembles the pre-intervention values of the treated unit.

Therefore, SC is subject to the curse of dimensionality, in which the probability of exact weight matching vanishes as the number of time periods increases (Ferman & Pinto, 2021; Ben-Michael et al., 2021). The *demeaned SC* (DSC) was proposed to relax SC constraints on weights to allow for a good pre-treatment fit when the length of pre-intervention is very large (Ferman & Pinto, 2021). Similarly, the panel data approach improves performance with increasing pre-intervention periods, using unconstrained regressions relaxing the SC assumptions (Wan et al., 2018). On the other hand, for shorter pre-intervention periods, SC weights fail to reproduce the trajectory of the treated unit Abadie (2021) or overfit to idiosyncratic errors (Sun et al., 2024).

Further, SC tends to heavily depend on the expert selection of control units; therefore, estimates are biased by noisy control units if they exist. *Robust* SC (RSC) enhanced SC by denoising the set of control units using a latent variable model (Amjad et al., 2018). The observable outcomes of the control group are obtained by a low-rank approximation.

Recent work on SC found that adding multiple outcomes metrics, e.g., GDP and education level, would improve inference (Amjad et al., 2019). Instead of having a separate SC model for each outcome, Sun et al. (2024) proposed to have a common set of weights across multiple outcome series, either by balancing a vector of all outcomes or an average of them. Though this approach is interesting, we do not consider it in the present work.

Another variation of synthetic control is to deal with dynamic data collection setting, e.g., personalized recommendation system, by updating the principal components as new data arrives (Agarwal et al., 2023), or by solving the model in a reinforcement learning fashion (Dwivedi et al., 2024).

Here, we consider a linear panel data setting, with concentration on a single metric outcome, and static data. For this problem setting, none of the recently-proposed methods address the problem of emerging bias and overfitting in short pre-intervention periods.

3 Problem Formulation

Consider panel data that is a collection of time series with respect to an aggregated metric of interest (e.g., country GDP). The data includes N units observed over T periods of time. Let T_0 be the intervention time, which splits the time period into a pre-intervention period over $1 \leq T_0 < T$ and a post-intervention period with length $T_1 = T - T_0$. We fix $i = 1$ ($i \in \{1, \ldots, N\}$) for the treated unit at time t, hence, let $s_{1t} \in \mathbb{R}^T$ be the treated unit vector. The remaining units $i = 2, \ldots, N$ are the controls that are not affected by the intervention. Let $\boldsymbol{S} \in \mathbb{R}^{N-1 \times T}$ be the control units matrix.

Let the outcomes of the control and treated units follow a factor model, a common model in the econometrics literature (Abadie, 2021). Let x_{it} denote an aggregated metric for a unit i at time t. In the absence of covariates and unobserved outcomes, the factor model is the following:

$$x_{it} = s_{it} + \epsilon_{it} \ . \tag{1}$$

[1]The terms *treatment* and *intervention* are used interchangeably.

Following the literature in data imputation (Athey et al., 2021; Amjad et al., 2018) and optimal recovery for missing values (Muresan & Parks, 2004), we consider $s_{it} = \eta_i \psi_t$, where $\eta_i \in \mathbb{R}^N$ and $\psi_t \in \mathbb{R}^T$ are the latent features that capture the unit and time specifications respectively for observed outcomes, with independent random zero-mean noise, ϵ_{it}, and variance $Var(\epsilon_{1j}) \leq \sigma^2$. Equation 1 can be rewritten in matrix form as $\mathbf{X} = \mathbf{S} + \mathcal{E}$.

To distinguish pre- and post-intervention periods, let $\boldsymbol{A} = [\boldsymbol{A}^-, \boldsymbol{A}^+]$, where $\boldsymbol{A}^- = \{a_{it}\}_{2 \leq i \leq N, t \leq T_0}$, and $\boldsymbol{A}^+ = \{a_{it}\}_{2 \leq i \leq N, T_0 < t \leq T}$. Vectors are defined in the same manner, i.e. $a_i = [a_i^+, a_i^-]$. The inverse of $\boldsymbol{A}$ is $\boldsymbol{A}^{-1}$. The Moore-Penrose pseudo-inverse of $\boldsymbol{A}$ is $\boldsymbol{A}^\dagger$. The transpose of $\boldsymbol{A}$ is $\boldsymbol{A}^\top$. The $\boldsymbol{A}$-norm of a vector u, denoted $||u||_A$, is the value of $u^\top \boldsymbol{A} u$.

The same notational convention holds for $\boldsymbol{S}$ and $\boldsymbol{X}$. We keep $\boldsymbol{A}$ as a generic variable. We use $\boldsymbol{S}$ to reflect a theoretical setting to develop the basic geometric algorithm and its properties. We consider $\boldsymbol{X}$ to reflect a noisy setting, i.e. the latent factor model in equation 1, for actual deployment of simulations and real-world experiments.

Given the noisy observations $\boldsymbol{X}$, and the partially observed treated unit s_1^- (in pre-intervention), our algorithm approximates s_1^+, the missing part of the treated unit vector (in post-intervention).

4 Algorithm

This section describes our algorithm. First, we briefly introduce the *optimal recovery* algorithm from signal processing, which is a fundamental building block of our approach. Then, we describe our estimation strategy.

Optimal recovery was introduced to effectively approximate a function known to belong to a certain signal class with limited information about it (Micchelli & Rivlin, 1976). It has been applied to estimate missing or corrupted pixels in images (Muresan & Parks, 2004) and missing values in biological data (Dean & Varshney, 2021). Optimal recovery estimates the missing value using a learned principal vector and a set of known values. One way the principal vector is constructed is using an ellipsoidal set of vectors that pass through a hyperplane.

Definition 1 (Ellipsoids) *Given a matrix Q, an ellipsoid K is a bounded convex set which has the form*

$$K = \{u \in \mathbb{R}^n : u^\top Q u \leq h\} \tag{2}$$

where $Q = Q^\top$, and $Q \succ 0$. That is, Q is symmetric and positive definite (Boyd & Vandenberghe, 2004).

The value h is the ellipsoid radius. The matrix Q determines how far the ellipsoid extends in every direction from the center (the ellipsoid semi-axis lengths are determined by eigenvalues of Q).

4.1 Ellipsoidal Optimal Recovery (EOpR)

Here, we extend *optimal recovery* with ellipsoidal signal class to estimate causal effects in panel data. Under the potential outcomes framework, we treat the estimation of causal effects as a missing data problem (Athey et al., 2021). Therefore, we use optimal recovery from approximation theory to recover the missing outcomes for causal inference, a setting in which optimal recovery has not been considered before.

Geometrically, we assume control units vectors in $\boldsymbol{A}$ belong to an ellipsoidal class K, and the pre-intervention of control units $\boldsymbol{A}^-$ belong to a hyperplane $\mathcal{H}$. We consider the treated unit a_1 as a partially-observed vector, where $a_1^+ = \{a_{1t}\}_{t > T_0}$ (in the post-intervention) is unknown and requires approximation. The vector a_1 lies in C, the intersection of K and $\mathcal{H}$. The aim is to find an estimator that minimizes the worst-case error. This is equivalent to finding the Chebyshev center of C (Boyd & Vandenberghe, 2004). Figure 1 illustrates the geometry of optimal recovery.

The benefit of the Chebyshev center is that it provides a minimax optimal solution for the recovery problem (Boyd & Vandenberghe, 2004), as in the following theorem.

Theorem 1 (Minimax Optimality) *Let $C \in \mathbb{R}^n$ be an ellipse that represents an intersection of an ellipsoid and a hyperplane. Ellipse C is a bounded and convex set with nonempty interior. A Chebyshev center*

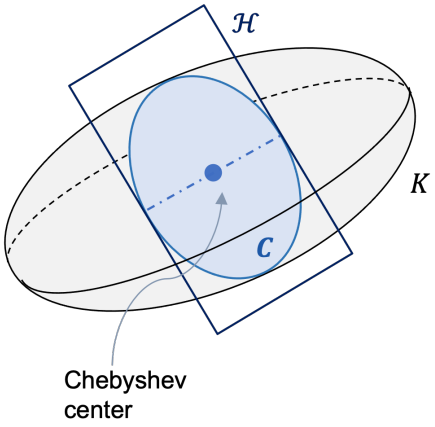


Figure 1: Geometric illustration for the optimal recovery algorithm with hyperplane $\mathcal{H}$ intersecting ellipsoid K, creating ellipse C, with a Chebyshev center

is a point inside C and is the minimal farthest distance from other points in C. For a given point $a_1 \in C$, the estimator $\hat{a}_1 \in C$

$$\min_{\hat{a}_1} \max_{a_1 \in C} ||\hat{a}_1 - a_1||. \tag{3}$$

is a Chebyshev center and it is the minimax estimator for a_1.

Proof. Follows directly from the definition of the Chebyshev center (Boyd & Vandenberghe, 2004; Micchelli & Rivlin, 1976).

4.2 Estimation

To find the Chebychev center, i.e. minimax estimator $\hat{a}_1$, we take two steps: (1) **Learning** which includes learning the ellipsoidal class K as in equation 2, and learning the hyperplane space for intersection, namely the *representers*, which we will elaborate on next. (2) **Extrapolation** given the learned spaces and the partially-known vector a_1^-.

4.2.1 Learning

1. Learning the ellipsoid We start by constructing a covariance matrix $\boldsymbol{\Sigma}$ from the data matrix $\boldsymbol{A}$, such that $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^\top + \lambda \boldsymbol{I}$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{T \times T}$, λ is a scalar, and $\boldsymbol{I}$ is the identity matrix.

To learn the ellipsoidal class K in equation 2, K needs to have the most stretch in the same direction of $\boldsymbol{\Sigma}$, hence we let the eigenvalues of $\boldsymbol{Q}$ in equation 2 be the reciprocal of the eigenvalues of $\boldsymbol{\Sigma}$,

$$\boldsymbol{Q} = \boldsymbol{\Sigma}^\dagger. \tag{4}$$

Choice of parameter. Based on the ellipsoid definition equation 1, the matrix $\boldsymbol{Q}$ must be positive definite. To ensure that eigenvalues of $\boldsymbol{Q}$ are strictly positive, we add a small perturbation $\lambda \in (0, 1]$ to the diagonal of $\boldsymbol{A}\boldsymbol{A}^\top$, such that it minimizes the ℓ_2-norm between outcomes in the pre-intervention period.

2. Learning the representers Given the pre-intervention matrix $\boldsymbol{\Sigma}^-$, we derive the representers $\boldsymbol{\Phi} \in \mathbb{R}^{T_0 \times T_0}$, by the Riesz representation theorem (Muresan & Parks, 2004), as

$$\boldsymbol{\Phi} = {\boldsymbol{\Sigma}^-}^\top \boldsymbol{Q} \boldsymbol{\Sigma}^-. \tag{5}$$

We can think of representers, geometrically, as the hyperspace $\mathcal{H}$ that intersects with K to create the space C.

4.2.2 Extrapolation

Now we calculate $\hat{a}_1$, the Chebyshev center of C. Given that the representer vectors lie in a subspace that is parallel to the center of the ellipse C, and a known vector a_1^-, $\hat{w}$ is a linear combination of the inverse of representers $\mathbf{\Phi}$, such that

$$\hat{w} = (\mathbf{\Phi})^{-1} a_1^-, \tag{6}$$

then, the Chebyshev center, the estimated outcome $\hat{a}_1$ is

$$\hat{a}_1 = \mathbf{\Sigma}^\top \hat{w}. \tag{7}$$

4.3 Worst-case estimates

In situations of deep uncertainty, cf. Lempert et al. (2006), it is important to acknowledge the most severe possible outcomes that could occur for a given policy. EOpR can also estimates the worst-case scenario as following.

Given that the estimator $\hat{a}_1$ is at the center of the intersection C, the vectors on the boundary of C are the worst-case estimates. To attain worst-case vectors (minimax control of the estimates), let y be the unit norm in $\mathcal{Z}$, a parallel subspace to representers $\mathbf{\Phi}$, determined by

$$y = \mathbf{\Phi}^{-1} \mathbf{\Sigma}^{-^\top} \boldsymbol{Q} \mathbf{\Sigma}. \tag{8}$$

The worst-case estimates $\bar{a}_1$ are:

$$\bar{a}_1 = \hat{a}_1 \pm (\varepsilon - \|\hat{a}_1\|_Q)^{\frac{1}{2}} y, \tag{9}$$

with a very small ε, and the $\boldsymbol{Q}$-norm $\|\hat{a}_1\|_Q = \hat{a}_1^\top Q \hat{a}_1$.

4.4 Properties of the estimator

Recall the latent factor model in equation 1, where $x_{it} = s_{it} + \epsilon_{it}$. To estimate the vector of interest s_1, we apply the learning step on $\boldsymbol{X}$ (in place of the generic $\boldsymbol{A}$), and extrapolate using the actual partially observed values in s_1^-, which is the pre-treatment vector. This leads to $\hat{s}_1$, which is an optimal Chebyshev solution. Hence, we rewrite equation 6-7 as following

$$w^* = \mathbf{\Sigma}^\top s_1^-, \tag{10}$$

and

$$\hat{s}_1 = \mathbf{\Sigma}^\top w^*. \tag{11}$$

Based on the optimal recovery approach, the extrapolation step produces an estimator that is the Chebyshev center of the ellipse C with desirable properties of unbiasedness and consistency.

Theorem 2 *Let r denote the rank of matrix $\mathbf{\Sigma}$, for $\lambda \geq 0$, and random noise ϵ with variance σ^2, the estimation error can be bounded as*

$$MSE(s_1, \hat{s_1}) \leq \frac{2\sigma^2 r + \sigma^2}{T} \tag{12}$$

.

Proof. Proof is in Appendix 8.1.

We say that the *MSE* error converges to zero if T grows without bound, however, in case $r = T$, then the bound is $[0, 2\sigma^2 + \frac{\sigma^2}{T}]$.

The Chebyshev center has also been proven to be a consistent and unbiased estimator geometrically (Halteman, 1986) for spherically-symmetric noise distributions ϵ, but it extends directly for non-spherical noise, full proof is in Appendix 8.2.

5 Empirical Analysis

We compare the accuracy of our ellipsoidal optimal recovery (EOpR) approach with other causal inference methods used in policy evaluation: SC (Abadie & Gardeazabal, 2003), RSC (Amjad et al., 2018), DSC (Ferman & Pinto, 2021), and SDID (Arkhangelsky et al., 2019). We first evaluate on simulated data to demonstrate properties of EOpR under certain settings. We then evaluate on two classical panel datasets commonly used in the SC literature (California Proposition 99 (Abadie et al., 2010) and Basque Country (Abadie & Gardeazabal, 2003)). We finally apply EOpR in the context of the COVID-19 pandemic to estimate the number of confirmed cases in New York State.

5.1 Evaluation Metrics

To measure the quality of estimation, we use two metrics. First, we measure the root-mean-square error (RMSE) of estimated vectors. The pre-intervention (training) error is for $1 \leq t \leq T_0$, and a post-intervention (testing) error is for $T_0 < t \leq T$

$$\text{RMSE}(u, \hat{u}) = \left(\frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} (u - \hat{u})^2 \right)^{1/2}, \tag{13}$$

where $\mathcal{T}$ is the size of the selected time period, and u and $\hat{u}$ are two dummy variables.

Second, Abadie et al. (2010) proposed a test statistic to evaluate the reliability of the estimates by running *placebo tests*. One placebo test considers one control unit as a placebo treated unit and apply the estimation algorithm. Since control units are assumed to not be affected by the examined intervention, one would expect that the estimated signal for the placebo unit does not diverge from its corresponding control unit. Further, the gaps between each placebo estimation and its corresponding control unit should be less divergent than the gap between the original treated unit and its estimation. Placebo tests are applied on the classical econometrics case studies.

5.2 Simulations

We conduct synthetic simulations to demonstrate the properties of EOpR estimates in both the pre- and post-intervention periods. We show that EOpR performs well and better than existing causal inference methods under various settings.

Experimental setup. Consider a data generating process, which is frequently considered for low-rank matrix decomposition solutions, similar to Amjad et al. (2019), as follows. First we create two sets of row and column features, B_r, B_c, where $B_r = \{b_k | b_k \sim \text{Unif}(0,1), 1 \leq k \leq 10\}$ and $B_c = \{b_k | b_k \sim \text{Unif}(0,1), 1 \leq k \leq 10\}$. For each unit $2 \leq i \leq N$, we assign a parameter θ_i drawn from B_r (with replacement), and for each time $1 \leq t \leq T$ we assign a parameter ρ_t drawn from B_c (with replacement). We use the following formula to generate a data point $\tilde{s}_{it} = f(\theta_i, \rho_t)$ to construct the control units

$$f(\theta_i, \rho_t) = \frac{10}{1 + \exp\left(-\theta_i - \rho_t - (\theta_i \rho_t)\right)} + \epsilon_{it}, \tag{14}$$

where $\epsilon_{it} \sim \mathcal{N}(0,1)$, an independent Gaussian noise.

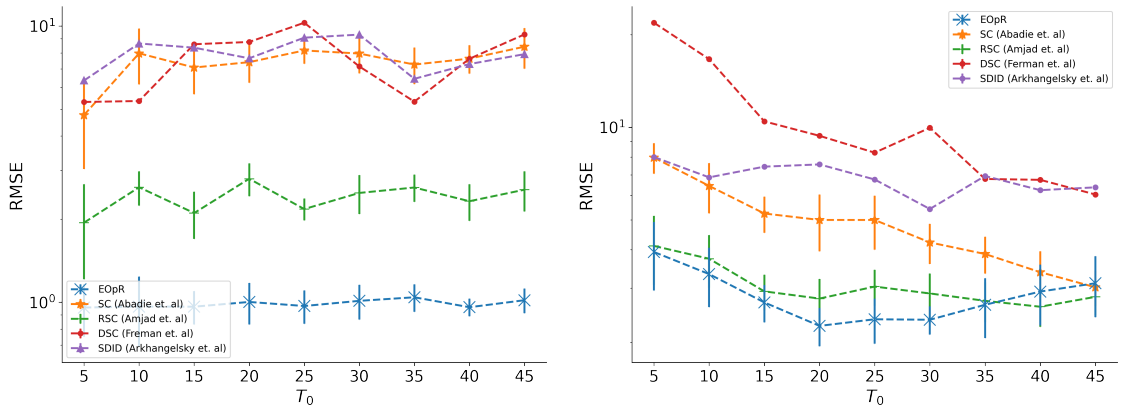
The data-generating process follow the latent factor model in equation 1, such that we are generating noisy $\boldsymbol{X}$ variables and applying the algorithm to them, but measuring performance for the true mean vector s_1.

In the following experiments, we investigate EOpR resistance to bias in comparison to other algorithms under different numbers of units N and time periods T_0 and T. For each combination of N, T_0, and T we generate ten simulations and average the resulting RMSE scores for the estimated pre- and post-intervention signals of $\tilde{s}_{1t}$.

5.2.1 Length of pre-intervention period

When the time of intervention T_0 starts very early in a period of time, it creates a short pre-intervention period. It has been discussed earlier that if T_0 is too short it fails to reproduce the trajectory of the treated unit in synthetic control methods (Abadie, 2021). Here, we test the robustness of our method under very limited length of pre-intervention period T_0. We fix the size of units N and post-intervention length T_1. We vary the time of intervention T_0 to be as short as 10% or as large as close to T_1.

Figure 2 shows the effect of different pre-treatment lengths on the algorithm's ability to estimate, fixing $N = 50$ and $T_1 = 50$. When having either a small or large number of pre-treatment periods, EOpR recovers the original treated signal with the smallest error compared to other algorithms. Specifically at $T_0 \leq 30$, where the pre-treatment period is relatively short, EOpR extrapolates beyond the training periods with the lowest estimation bias, whereas other algorithms have higher bias in the post-intervention estimation.



(a) Pre-intervention Error

(b) Post-intervention Error

Figure 2: Growing size of pre-intervention period T_0 with fixed post-intervention $T_1 = 50$ and $N = 50$

5.2.2 Length of post-intervention period

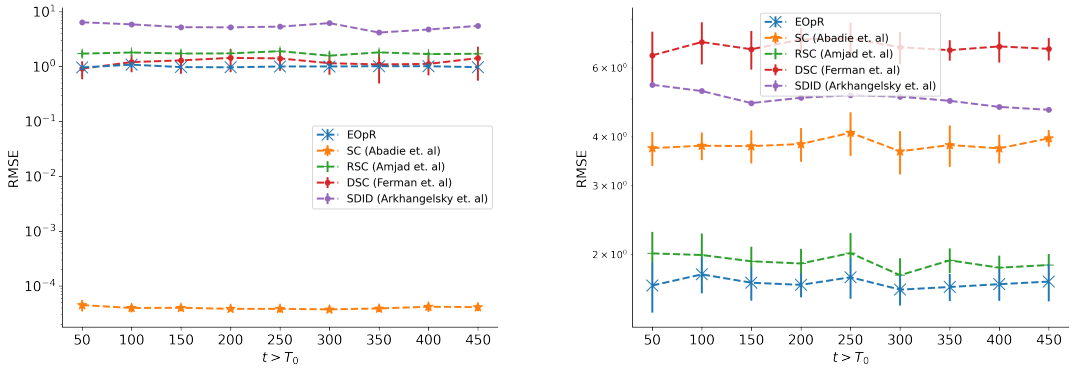
We further investigate the ability of EOpR to estimate the trajectory of the post-intervention for an extended period of time. The ability to estimate for an extended period of time indicates an algorithm is robust and consistent as $T \to \infty$.

Figure 3 shows the estimation errors with fixed $N = 100$ and $T_0 = 50$, and tested over multiple lengths of post-intervention. EOpR consistently achieves a low estimation error in both pre- and post-intervention estimation, especially at longer periods of time, e.g. $(t = 450)$.

The remaining simulated experiments in Appendix 9.1.

5.3 Real-World Experiments

We explore two econometric real-world case studies: Basque Country case study Abadie et al. (2010) and the California 99 Proposition case study Abadie & Gardeazabal (2003) (in Appendix 9.2). Both case studies showcase the ability of the original synthetic control estimator to produce reliable counterfactual estimation. We use the two case studies to demonstrate the robustness of our proposed algorithm in producing a reliable counterfactual. Given that the ground truth is missing in these case studies, econometricians consider placebo tests to show reliability of estimator, and we follow the same procedure for evaluation. We also consider a third case-study that we develop for COVID-19 data and lockdown interventions.



(a) Pre-intervention Error

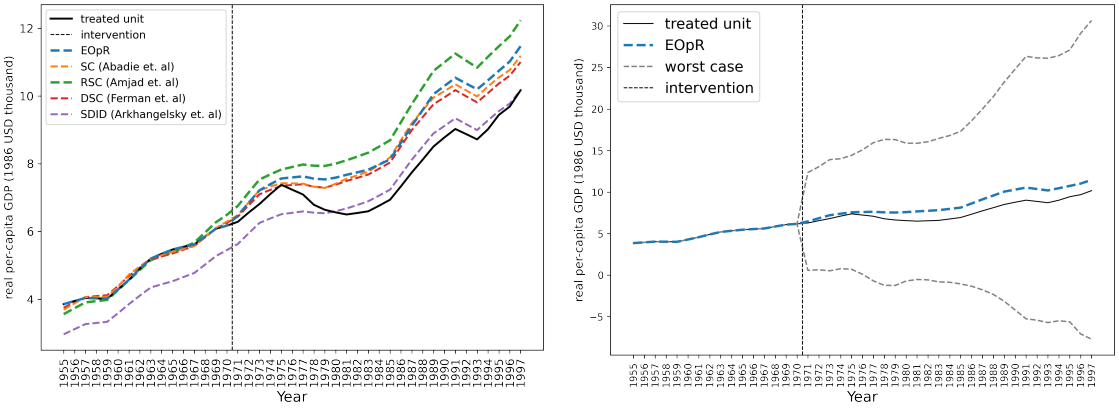
(b) Post-intervention Error

Figure 3: Growing post-intervention periods, with $N = 100$ and $T_0 = 50$

5.3.1 Basque Country

The objective of this case study is to investigate the effect of terrorist attacks on the Basque Country economy compared to other Spanish regions. Terrorist activities started by 1970. There was a significant negative impact on the economy of Basque Country measured by per-capita GDP. Abadie et al. (2010) showed that the economy would be better without terrorism.

Results. Figure 4 shows the actual trajectory of the Basque Country economy in black, with a degradation after 1970. In comparison to other methods, our EOpR method recovers Basque Country estimate, with more accurate fit on the pre-intervention values of the treated unit. Figure 4 shows the estimated worst-case potential outcomes from EOpR.



(a) Comparison of methods

(b) Solution with worst cases

Figure 4: Trends in per-capita GDP between Basque Country vs. synthetic Basque Country

Placebo Tests. We create placebo tests, similar to Abadie et al. (2010). Note that Abadie et al. (2010) excluded five regions that had poor fit in the pre-intervention, but we keep all regions. We plot the differences between our estimates and the observations of all regions as placebo and Basque Country (the actual treated unit). Figure 5 shows the differences for all regions compared to Basque Country (solid black line in the

figure). The divergence for Basque Country was the largest, thus, the derived estimates by the EOpR are reliable.

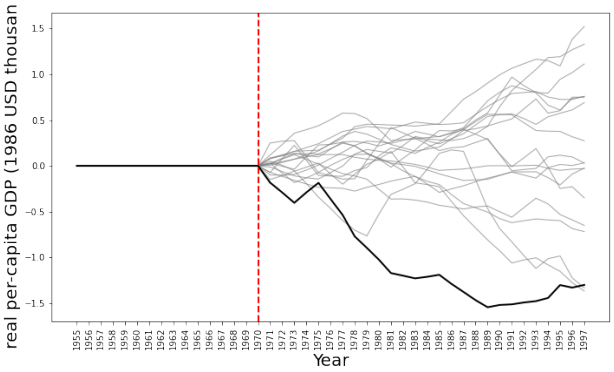


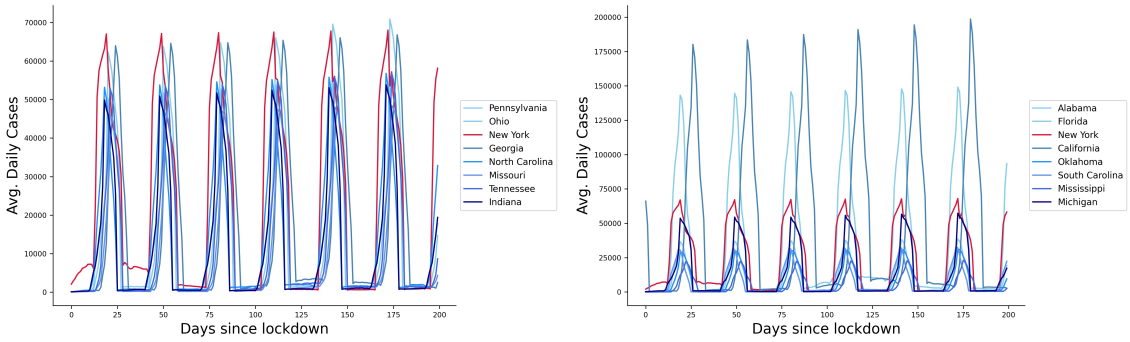
Figure 5: Placebo tests including all regions

5.3.2 COVID-19 in New York

New York was one of the earliest American states that turned to an epicenter of COVID-19 during 2020 (Thompson et al., 2020). We aim to estimate the New York COVID-19 cases trajectory with states as control units using EOpR and other methods. We select the states where lockdowns were imposed, yielding a total of 43 states.

Experimental setup For lockdown dates, we use data from the COVIDVis project [2] that tracks policy interventions at the state level. We consider the dates of *shelter-in-place* mandate. For COVID-19 cases, we consider state-level case load data from The New York Times (2021). Note that since reported cases depend on COVID-19 testing, our analysis is limited by the fact there was widespread shortage of available tests in different regions at different times.

Since each state imposed a lockdown at different times, we aligned states based on the days differences between the time a lockdown took place and the rest of the dates during the period of interest, following Bayat et al. (2020). Figure 6 shows the trend of New York and 7 other states and their moving averages (over 7 days).



(a) States with relatively similar trends

(b) States with dissimilar trends

Figure 6: Moving average of COVID-19 confirmed cases of New York (in red) and 7 other states

Based on the literature of SC (Abadie et al., 2010), one would need to select the control units based on their similar trends to the target unit to ensure a correct construction of the treated unit. In the following

[2] `https://covidvis.berkeley.edu/#lockdown_section`

experiments, we estimate New York COVID-19 cases using 43 states, for a period of 200 days, $T = 200$ to approximate the period between March and August in 2020. We vary the lengths of pre-treatment period by varying the time of intervention $T_0 \in \{5, 10, 15, 20, 25, 30\}$ days.

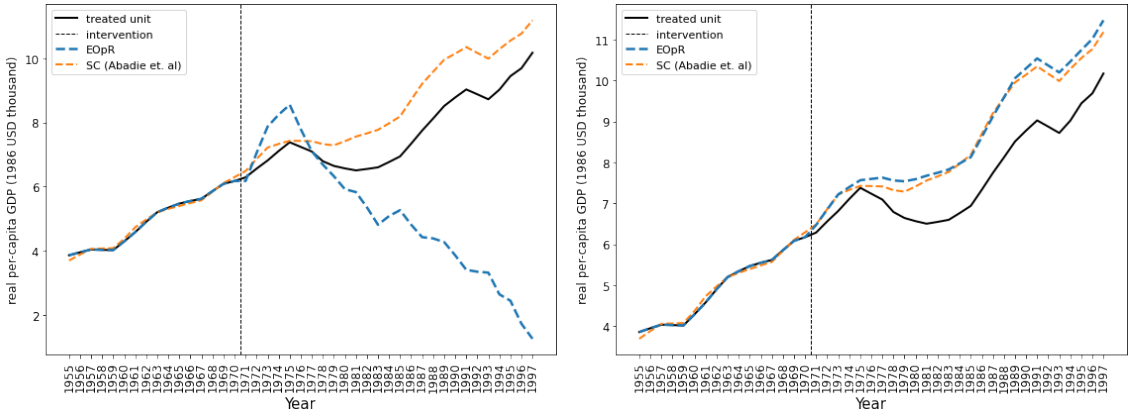
Table 1: Pre- and Post-intervention RMSE for COVID-19 case study

	Pre-intervention (Training) Error					
T_0	5	10	15	20	25	30
EOpR	**0.017**	**0.021**	0.631	**0.499**	0.706	0.831
SC (Abadie et al.)	0.043	0.074	**0.316**	0.501	**0.501**	**0.525**
RSC (Amjad et al.)	0.094	0.139	0.434	0.905	0.890	0.834
DSC (Ferman et al.)	0.061	0.129	0.377	0.745	0.659	0.685
SDID (Arkhangelsky et. al)	0.045	0.063	0.397	0.922	1.005	0.964
	Post-intervention (Testing) Error					
T_0	5	10	15	20	25	30
EOpR	**0.842**	**0.852**	0.676	**0.456**	0.633	0.803
SC (Abadie et al.)	0.921	0.959	0.896	0.512	**0.491**	**0.477**
RSC (Amjad et al.)	0.874	0.893	0.936	0.854	0.788	0.790
DSC (Ferman et al.)	0.958	0.879	**0.602**	0.730	0.611	0.645
SDID (Arkhangelsky et. al)	0.947	0.958	0.964	0.935	0.922	0.926

Results. Table 1 shows a comparison of estimation performance of our algorithm and the other four algorithms. At short pre-intervention periods, e.g. $T_0 = 5$ and $T_0 = 10$, EOpR has produced the lowest RMSE in pre- and post-treatment trends in comparison to other algorithms. This reflects its ability to recover the true estimates even under limited time in pre-intervention, or for an early intervention decision. Figure 13 (in appendix) shows the estimation trajectories for each choice of T_0.

5.4 Ablation Study

We empirically study the effect of the added perturbation λ to the covariate matrix $\boldsymbol{X}\boldsymbol{X}^\top$. This small perturbation is added to ensure $\boldsymbol{Q}$ is positive definite in equation 4, following the definition of ellipsoids in equation 1. When $\lambda = 0$, we see the resulting predictions suffer drastically, deviating from expected possible outcomes (Figure 7a). Using an appropriate λ as in Figure 7b, balances the model complexity which helps safeguard the algorithm from potentially underfitting the training data, and producing a biased estimate.

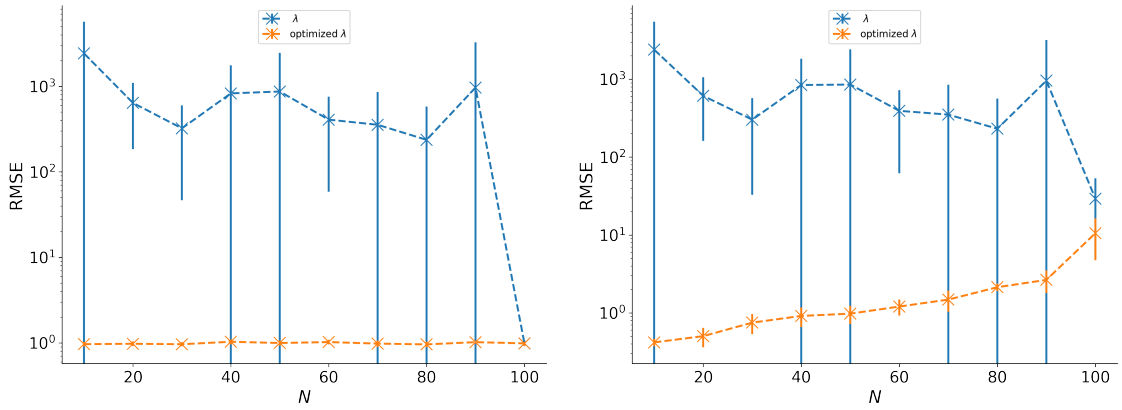


(a) $\lambda = 0$ (b) Optimized λ

Figure 7: Impact of adding λ to the covariance matrix $\boldsymbol{X}\boldsymbol{X}^\top$ for Basque case study

A similar observation is seen for simulated data, fixing $T_0 = 100$ and $T_1 = 10$ and varying N, and for each combination we generate 10 simulations and average the resulting error scores. Figure 8 shows that with

optimized λ the bias and variance are consistently low, whereas estimations with non-optimized λ, i.e. $\lambda = 0$ estimations suffer from high error.



(a) pre-intervention error

(b) post-intervention error

Figure 8: Impact of adding λ to the covariance matrix $\boldsymbol{X}\boldsymbol{X}^\top$ for synthetically-generated data, fixed $T_0 = 100$ and $T_1 = 10$ with varying N.

6 Conclusion

Classical synthetic control has been noted as effective for causal inference in comparative studies. Here, we propose an approximation-theoretic approach for synthetic control—ellipsoidal optimal recovery (EOpR)—that estimates the causal effect given a policy intervention. Furthermore, given the properties of EOpR, we derive worst-case estimates, which are themselves very useful for policy evaluation. A limitation in EOpR is that given the ellipsoid assumption (a common assumption in economics (Armstrong & Kolesár, 2018)), it is important to correctly specify the shape of the ellipsoid, which we do from data. Real-world data may exhibit heteroskedasticity or other non-quadratic characteristics, in which data cannot be modeled adequately using an ellipsoid (but optimal recovery can be extended to other geometric priors (Dean & Varshney, 2021)). Our approach of EOpR has less estimation error for pre- and post-intervention periods, especially with short pre-intervention periods. This is demonstrated through comparisons on simulated data, classical case studies in econometrics, and a new health-relevant setting.

References

Alberto Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425, June 2021. doi: 10.1257/jel.20191450.

Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1):113–132, March 2003. doi: 10.1257/000282803321455188.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 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, 2010. doi: 10.1198/jasa.2009.ap08746.

Anish Agarwal, Munther Dahleh, Devavrat Shah, and Dennis Shen. Causal matrix completion. arXiv:2109.15154, 2021.

Anish Agarwal, Keegan Harris, Justin Whitehouse, and Steven Z. Wu. Adaptive principal component regression with applications to panel data. In *Advances in Neural Information Processing Systems*, volume 36, pp. 77104–77118. Curran Associates, Inc., 2023. URL `https://proceedings.neurips.cc/paper_files/paper/2023/file/f37265d7493377170a3b4ba91823119a-Paper-Conference.pdf`.

Muhammad Amjad, Devavrat Shah, and Dennis Shen. Robust synthetic control. *Journal of Machine Learning Research*, 19(22):1–50, September 2018.

Muhammad Amjad, Vishal Misra, Devavrat Shah, and Dennis Shen. MRSC: Multi-dimensional robust synthetic control. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(2): 1–27, June 2019. doi: 10.1145/3341617.3326152.

Dmitry Arkhangelsky, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. Synthetic difference in differences. Working paper, National Bureau of Economic Research, 2019. URL `http://www.nber.org/papers/w25532`.

Timothy B. Armstrong and Michal Kolesár. Optimal inference in a class of regression models. *Econometrica*, 86(2):655–683, March 2018.

Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536): 1716–1730, 2021. doi: 10.1080/01621459.2021.1891924.

Niloofar Bayat, Cody Morrin, Yuheng Wang, and Vishal Misra. Synthetic control, synthetic interventions, and COVID-19 spread: Exploring the impact of lockdown measures and herd immunity. arXiv:2009.09987, 2020.

Eli Ben-Michael, Avi Feller, and Jesse Rothstein. The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536):1789–1803, 2021.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004.

Rebecca Chen Dean and Lav R. Varshney. Optimal recovery of missing values for non-negative matrix factorization. *IEEE Open Journal of Signal Processing*, 2:207–216, March 2021. doi: 10.1109/OJSP.2021.3069373.

Raaz Dwivedi, Katherine Tian, Sabina Tomkins, Predrag Klasnja, Susan Murphy, and Devavrat Shah. Counterfactual inference for sequential experiments, 2024. URL `https://arxiv.org/abs/2202.06891`.

Bruno Ferman and Cristine Pinto. Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221, November 2021. doi: https://doi.org/10.3982/QE1596.

Edward J Halteman. The Chebyshev center: A multidimensional estimate of location. *Journal of Statistical Planning and Inference*, 13:389–394, 1986.

Saleem A Kassam and H Vincent Poor. Robust techniques for signal processing: A survey. *Proceedings of the IEEE*, 73(3):433–481, 1985.

Robert J. Lempert, David G. Groves, Steven W. Popper, and Steven C. Bankes. A general, analytic method for generating robust strategies and narrative scenarios. *Management Science*, 52(4):514–528, April 2006. doi: 10.1287/mnsc.1050.0472.

Charles A. Micchelli and Theodore J. Rivlin. A survey of optimal recovery. In *Optimal Estimation in Approximation Theory*, pp. 1–54. Plenum Press, New York, NY, USA, 1976.

D Darian Muresan and Thomas W Parks. Demosaicing using optimal recovery. *IEEE Transactions on Image Processing*, 14(2):267–278, 2005.

Darian D. Muresan and Thomas W. Parks. Adaptively quadratic (AQua) image interpolation. *IEEE Transactions on Image Processing*, 13(5):690–698, May 2004. doi: 10.1109/TIP.2004.826097.

Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009. doi: 10.1214/09-SS057.

Judea Pearl. Causal inference. In *Proceedings of Workshop on Causality: Objectives and Assessment*, pp. 39–58, 2010.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. doi: 10.1037/h0037350.

Liyang Sun, Eli Ben-Michael, and Avi Feller. Using multiple outcomes to improve the synthetic control method, 2024. URL `https://arxiv.org/abs/2311.16260`.

The New York Times. Coronavirus (Covid-19) data in the United States, 2021. URL `https://github.com/nytimes/covid-19-data`. dataset.

Corinne N Thompson, Jennifer Baumgartner, Carolina Pichardo, Brian Toro, Lan Li, Robert Arciuolo, Pui Ying Chan, Judy Chen, Gretchen Culp, Alexander Davidson, et al. COVID-19 outbreak—New York City, February 29–June 1, 2020. *Morbidity and Mortality Weekly Report*, 69(46):1725, 2020.

Shui-Ki Wan, Yimeng Xie, and Cheng Hsiao. Panel data approach vs synthetic control method. *Economics Letters*, 164:121–123, 2018. ISSN 0165-1765. doi: https://doi.org/10.1016/j.econlet.2018.01.019. URL `https://www.sciencedirect.com/science/article/pii/S0165176518300272`.

7 Appendix

8 Consistency and Unbiasedness

8.1 Analytical Proof

To show consistency, here we bound the ℓ_2 error of the estimation. We will drop the dependency on λ for the sake of simplicity in proof. Recall that the noise term ϵ_1 is a zero-mean independent random variable that satisfies $\mathbb{E}(\epsilon_{ij}) = 0$ for all i and j by assumption, with variance $Var(\epsilon_{ij}) = \sigma^2$.

Lemma 1 *Suppose $x_1 = s_1 + \epsilon_1$ with $\mathbb{E}(\epsilon_{1j}) = 0$ and $Var(\epsilon_{1j}) \leq \sigma^2$ for all $j \in \{1, \ldots, T\}$. Let w^* be the min-max optimizer for $\hat{s}_1$. Then for any $\lambda \geq 0$,*

$$\mathbb{E}\|s_1 - \hat{s}_1\| \leq 2\sigma^2 r + \sigma^2. \tag{15}$$

Proof. Recall from equation 1 that the treatment row is $x_1 = s_1 + \epsilon_1$. By definition of the Chebyshev center and its properties of unique estimation, consider $\hat{s}_1 = \mathbf{\Sigma}^\top w^*$ from equation 11. By the generic definition in equation 7, $\hat{w} = \Sigma^{\dagger^\top} x_1$, ($x_1$ in place of the generic a_1), and so, $\hat{w}$ is sub-optimal for $\hat{s}_1$.

$$\begin{aligned}
\|s_1 - \hat{s}_1\|^2 &= \|(x_1 - \epsilon_1) - \mathbf{\Sigma}^\top w^*\|^2 \\
&\leq \|(x_1 - \mathbf{\Sigma}^\top \hat{w} + (-\epsilon_1)\|^2 \\
&\leq \|x_1 - \mathbf{\Sigma}^\top \hat{w}\|^2 + \|\epsilon_1\|^2 + 2\langle -\epsilon_1, x_1 - \mathbf{\Sigma}^\top \hat{w}\rangle \\
&\leq \|(\mathbf{\Sigma}^\top \hat{w} + \epsilon_1) - \mathbf{\Sigma}^\top \hat{w}\|^2 + \|\epsilon_1\|^2 \\
&\quad + 2\langle -\epsilon_1, x_1 - \mathbf{\Sigma}^\top \hat{w}\rangle \\
&\leq 2\|\epsilon_1\|^2 + 2\langle -\epsilon_1, x_1 - \mathbf{\Sigma}^\top \hat{w}\rangle.
\end{aligned} \tag{16}$$

Taking expectations, we arrive at

$$\mathbb{E}\|s_1 - \hat{s}_1\|^2 = 2\mathbb{E}\|\epsilon_1\|^2 + 2\mathbb{E}\langle -\epsilon_1, x_1 - \mathbf{\Sigma}^\top \hat{w}\rangle. \tag{17}$$

With the noise having variance σ^2, then

$$\mathbb{E}\|\epsilon_1\|^2 = \sigma^2. \tag{18}$$

We must address an inner product on the right side. First, we derive some useful facts. Recall the trace operator has the mapping property $tr(AB) = tr(BA)$, and the projection matrix P to be $P_1 = AA^\dagger$ and $P_2 = A^\dagger A$. Hence,

$$\begin{aligned}
\mathbb{E}[(\epsilon_1)^\top \mathbf{\Sigma}^\top \mathbf{\Sigma}^{\dagger^\top} \epsilon_1] &= \mathbb{E}[tr((\epsilon_1)^\top \mathbf{\Sigma}^\top \mathbf{\Sigma}^\dagger \epsilon_1)] \\
&= \mathbb{E}[tr(\mathbf{\Sigma}^\top \mathbf{\Sigma}^\dagger)\epsilon_1(\epsilon_1)^\top] \\
&= tr\big(\mathbb{E}[\mathbf{\Sigma}^\top \mathbf{\Sigma}^{\dagger^\top} \epsilon_1(\epsilon_1)^\top]\big) \\
&= tr\big(\mathbb{E}[\Sigma^\top \Sigma^{\dagger^\top}]\mathbb{E}[\epsilon_1(\epsilon_1)^\top]\big) \\
&= tr\big(\mathbb{E}[\Sigma^\top \Sigma^{\dagger^\top}]\sigma^2 I\big) \\
&= \sigma^2 \mathbb{E}[tr(\Sigma^\top \Sigma^{\dagger^\top})] \\
&= \sigma^2 \mathbb{E}[\text{rank}(\Sigma)] \\
&\leq \sigma^2 r,
\end{aligned} \tag{19}$$

which follows since the trace of a projection matrix equals the rank of the matrix, i.e. $tr(\Sigma^\top \Sigma^{\dagger^\top}) = \text{rank}(\Sigma^\top)$. Hence, the rank of Σ is at most r.

Returning to the inner product, by the generic definition in equation 7, $\hat{w} = \Sigma^{\dagger^\top} x_1$, ($x_1$ in place of the generic a_1). Recall from equation 1, $x_1 = s_1 + \epsilon_1$, then

$$\begin{aligned}
\mathbb{E}[\langle -\epsilon_1, x_1 - \Sigma\hat{w}\rangle] &= \mathbb{E}[(\epsilon_1)^\top \Sigma \hat{w}] - \mathbb{E}[(\epsilon_1)^\top x_1] \\
&= \mathbb{E}[(\epsilon_1)^\top \Sigma^\top \Sigma^{\dagger^\top} x_1] - \mathbb{E}[(\epsilon_1)]s_1 - \mathbb{E}[(\epsilon_1)^\top \epsilon_1] \\
&= \mathbb{E}[(\epsilon_1)^\top \Sigma^\top \Sigma^{\dagger^\top}]s_1 + \mathbb{E}[(\epsilon_1)^\top \Sigma^\top \Sigma^{\dagger^\top} \epsilon_1] - \mathbb{E}[(\epsilon_1)^\top \epsilon_1] \\
&= \mathbb{E}[(\epsilon)^\top][\Sigma^\top \Sigma^{\dagger^\top}]s_1 + \mathbb{E}[(\epsilon_1)^\top \Sigma^\top \Sigma^{\dagger^\top} \epsilon_1] - \mathbb{E}[(\epsilon_1)^\top \epsilon_1] \\
&= \mathbb{E}[(\epsilon_1)^\top \Sigma^\top \Sigma^{\dagger^\top} \epsilon_1] - \mathbb{E}\|\epsilon_1\|^2 \\
&\leq \sigma^2 r - \mathbb{E}\|\epsilon_1\|^2,
\end{aligned} \tag{20}$$

Finally, we replace the above terms into inequality equation 17 to arrive at

$$\begin{aligned}
\mathbb{E}\|s_1 - \hat{s}_1\|^2 &\leq 2\mathbb{E}\|\epsilon_1\|^2 + 2\mathbb{E}\langle -\epsilon, x_1 - \Sigma^\top \hat{w}\rangle \\
&\leq \sigma^2 + 2\sigma^2 r,
\end{aligned} \tag{21}$$

which completes the proof.

Now, let us consider replacing the above value in the mean squared error (*MSE*), which can be bounded by

$$\begin{aligned}
MSE(s_1, \hat{s}_1) &= \frac{1}{T}\|s_1 - \hat{s}_1\|^2 \\
&\leq \frac{2\sigma^2 r + \sigma^2}{T}.
\end{aligned} \tag{22}$$

8.2 Geometric proof

The following theorem (Halteman, 1986) demonstrates that the Chebyshev center is an unbiased and consistent estimator for the center of a any spherically-symmetric distribution over a k-sphere. The result can further be extended from a sphere to an ellipse.

Theorem 3 (Unbiasedness and consistency) *Suppose the points $W = \{w_i, \ldots, w_n\}$ are sampled from an independent and identically distributed (i.i.d.) spherical k-dimensional distribution to create a sphere $S(\omega, \pi)$, with center ω and radius π. Let $S(\bar{\omega}, \bar{\pi})$, be the smallest sphere containing the samples of W, (i.e. $\bar{\pi} < \pi$). The center $\bar{\omega}$, a Chebyshev center, has a spherical distribution and hence $\bar{\omega}$ is unbiased for ω, and consistent, i.e. $\bar{\omega} \to \omega$ with probability goes to unity.*

***Proof.* 1** *Given that distribution of $S(\omega, \pi)$ is i.i.d., it is invariant under any rotation about ω. The center $\bar{\omega}$, for any rotated sample in W, will be rotated by an equal amount of ω. Therefore, the distribution of $\bar{\omega}$ is rotationally invariant about the same ω, and thus the Chebyshev center $\bar{\omega}$ is unbiased.*

To prove consistency, let the sphere $S(\omega, \pi)$ have density function $f(w) = \frac{1}{n}$, and let H be the set of points at the boundary of $S(\bar{\omega}, \bar{r})$. Then, the maximum distance d between $w \in H$ and the sphere $S(\omega, \pi)$ converges to 0 for large n, and so $\bar{\omega} \to \omega$ with probability 1.

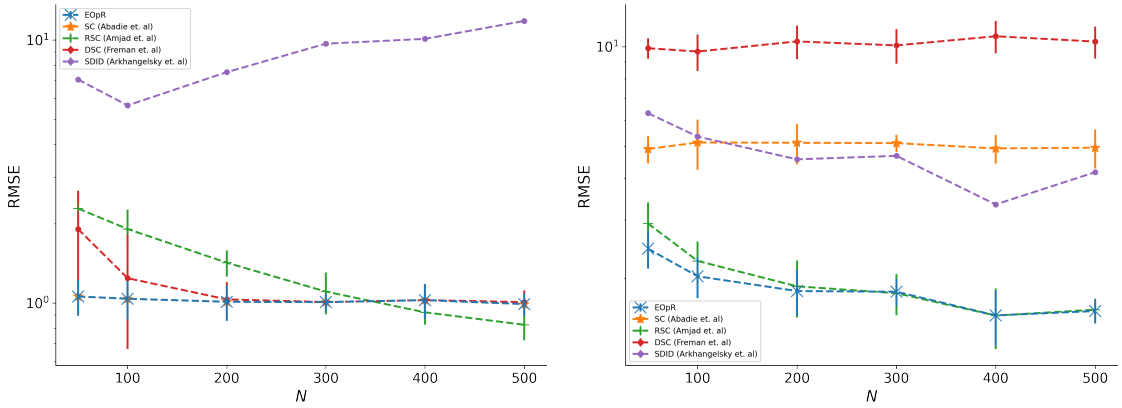
9 Empirical Results (continued)

9.1 Simulated Experiments

9.1.1 Number of control units

With a large number of units, the risk of overfitting increases, which produces a high potential of increased estimation bias (Abadie et al., 2010). Here, we model the increase of the number of units with fixed $T_0 = 25$ and total $T = 125$. Large number of control units, N, has shown to be challenging for SC as it exacerbates the bias of the estimator (Abadie, 2021).

Figure 9 shows the effect of a growing number of units N on algorithm performance. Given that EOpR also achieves low error at the post-intervention estimation at greater sizes of N, EOpR has the ability to reconstruct the true signal trend even with large control units and potentially noisy settings.



(a) Pre-intervention Error

(b) Post-intervention Error

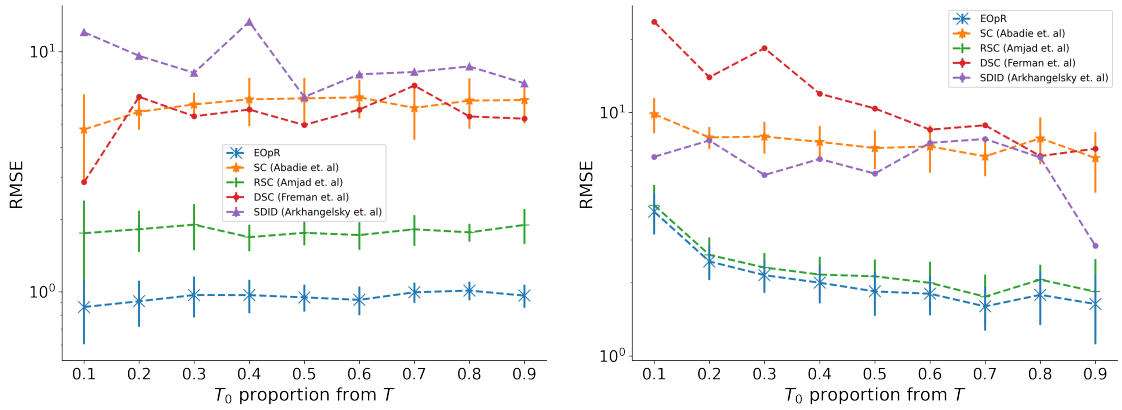
Figure 9: Growing size of N, with $T_0 = 25$, and post-intervention $T = 100$

9.1.2 Relative length of pre-intervention period

Consider when the number of pre-treatment units vary proportionally to T, fix the size of units N and the total size of time T. We vary the time of intervention T_0 between 10% to 90% of the entire time T.

Figure 10 shows the effect of different pre-treatment lengths on the algorithm's ability to estimate, with $N = 50$ and $T = 200$. When having either a small or large number of pre-treatment periods relative to T, EOpR recovers the original treated signal with the smallest error compared to other algorithms.

Additionally, at lengths of 40% and onwards, the number of pre-treament vectors are greater than the size of N, ($N \ll T_0$)—a common setting adapted by Abadie et al. (2010)—EOpR still maintains a lower bias in the post-intervention with consistent low training error.



(a) Pre-intervention Error

(b) Post-intervention Error

Figure 10: Percentage of T_0 length of the total T, with $N = 50$, and total $T = 200$

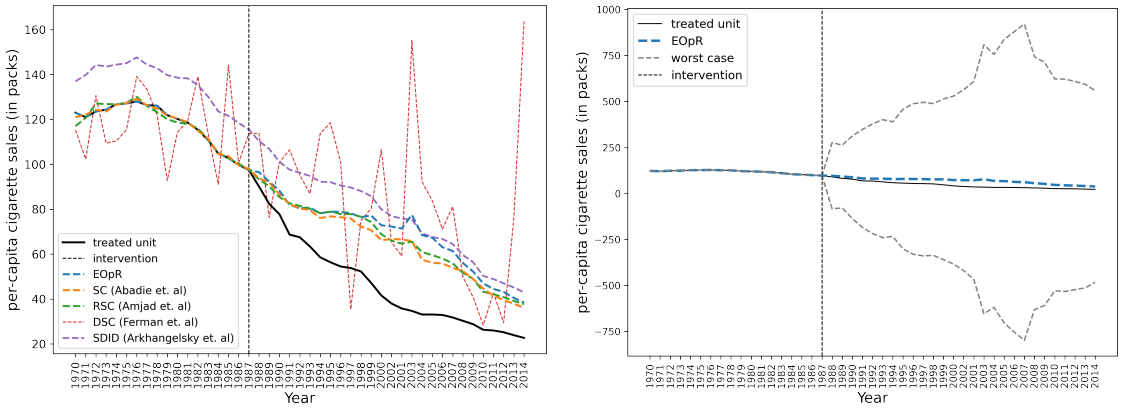
9.2 Real-world Experiment

9.2.1 California Proposition 99

The objective of this case study is to investigate the anti-tobacco legislation, Proposition 99, on the per-capita cigarette consumption in California in comparison to other states in the United States. The legislation took place in 1970. Without such legislation, the consumption of cigarettes in California would not have decreased (Abadie et al., 2010).

Results. Figure 11 shows the actual trajectory of California cigarette consumption in black. Our method recovers the estimated signal better than other methods with an adequate fit on the pre-intervention outcome, and also it derives the worst-case estimates.

Placebo Tests. We apply the same placebo test to the California case study. Abadie et al. (2010) excluded 12 regions, but we keep all of them. Figure 12 shows the divergence between estimations and observations of all regions with solid black line for California. This shows a similar observation as in Abadie et al. (2010).



(a) Comparison of methods

(b) Solution with worst cases

Figure 11: Trends in per-capita cigarette sales between California vs. synthetic California

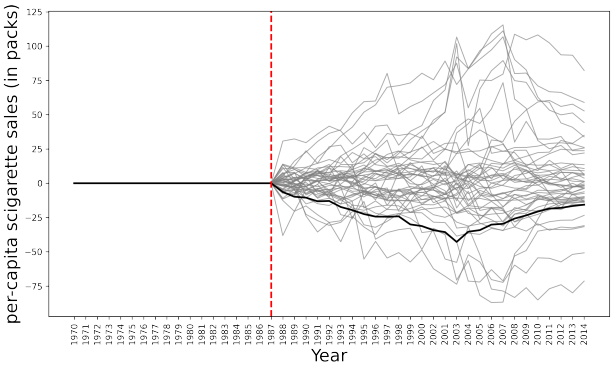
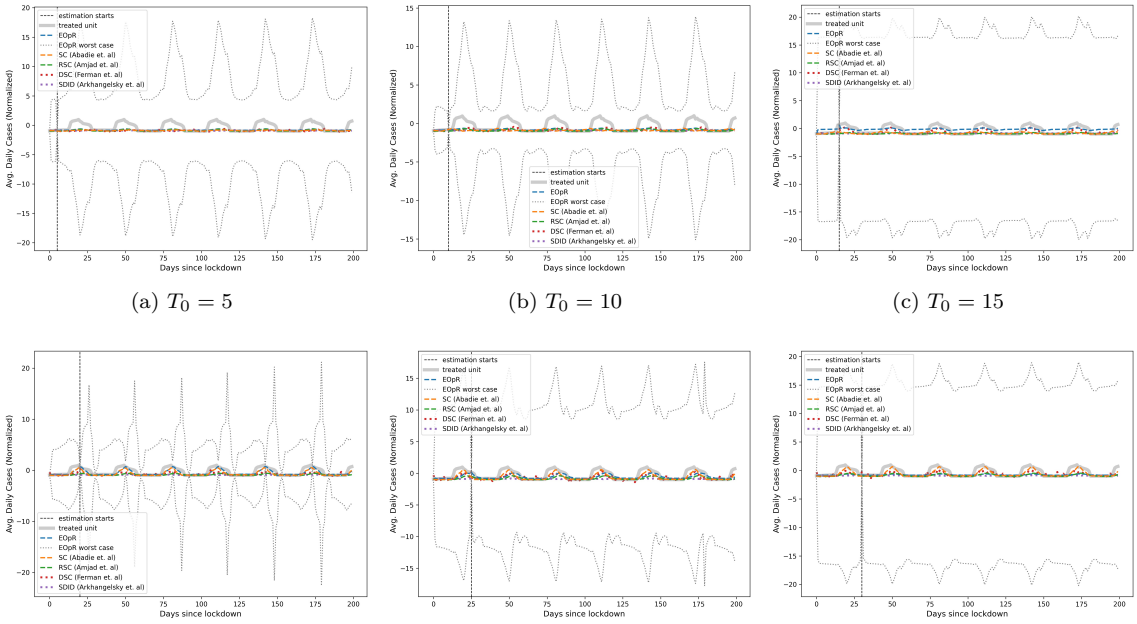


Figure 12: Placebo tests including all regions

9.3 COVID-19 Experiment (continued)

Post-treatment length = 60 Figure 13 below shows the estimated trend of daily average cases of New York in comparison to the actual New York trend (treated unit in the figure). The estimation starts at $T_0 \in \{5, 10, 15, 20, 25, 30\}$. The worst-case estimates are also plotted.



(a) $T_0 = 5$ (b) $T_0 = 10$ (c) $T_0 = 15$

(d) $T_0 = 20$ (e) $T_0 = 25$ (f) $T_0 = 30$

Figure 13: Comparisons of methods for estimating New York trend