

Triply Robust Panel Estimators

Susan Athey, Guido Imbens, Zhaonan Qu & Davide Viviano

September 12, 2025

This paper studies estimation of causal effects in a panel data setting. We introduce a new estimator, the Triply ROBust Panel (TROP) estimator, that combines (i) a flexible model for the potential outcomes based on a low-rank factor structure on top of a two-way-fixed effect specification, with (ii) unit weights intended to upweight units similar to the treated units and (iii) time weights intended to upweight time periods close to the treated time periods. We study the performance of the estimator in a set of simulations designed to closely match several commonly studied real data sets. We find that there is substantial variation in the performance of the estimators across the settings considered. The proposed estimator outperforms two-way-fixed-effect/difference-in-differences, synthetic control, matrix completion and synthetic-difference-in-differences estimators. We investigate what features of the data generating process lead to this performance, and assess the relative importance of the three components of the proposed estimator. We have two recommendations. Our preferred strategy is that researchers use simulations closely matched to the data they are interested in, along the lines discussed in this paper, to investigate which estimators work well in their particular setting. A simpler approach is to use more robust estimators such as synthetic difference-in-differences or the new triply robust panel estimator which we find to substantially outperform two-way fixed effect estimators in many empirically relevant settings.

Keywords: Causal Inference, Panel Data, Factor Models, Difference in Differences, Synthetic Controls

Susan Athey, Stanford University; Guido Imbens, Stanford University; Zhaonan Qu, Columbia University; Davide Viviano, Harvard University. We thank the Office of Naval Research for support under grant numbers N00014-17-1-2131 and N00014-19-1-2468 and Amazon for a gift. An early version of this paper was presented as the Journal of Applied Econometrics lecture at the ASSA meetings in San Francisco in January 2025. We are grateful for comments by Isaiah Andrews, Dmitry Arkhangelsky, Jesse Shapiro and Aico van Vuren.

1. Introduction

In recent years, there has been a resurgence of interest in methods for analyzing panel data. The focus in this literature has been on estimating causal effects, often of binary interventions. One important part of this literature has been the development of Synthetic Control (SC) methods [AG03, ADH10, ADH15a], with subsequent work proposing modifications of the basic SC approach. Another recent strand of the panel data literature has focused on modifications of the older Difference-in-Differences (DID) and Two-Way-Fixed-Effect (TWFE) methods [CS20, GB21, SA20, RSBP23, CSX25, BCC⁺25, SZ20, AS18, dCd19].

The range of commonly used panel data estimators is based on different assumptions that are difficult to validate or even compare in practice. DID methods invoke parallel trends, possibly controlling for time heterogeneity; Matrix Completion (MC) and interactive fixed effects estimators rely on a low-rank factor model; SC emphasizes alignment on pretreatment outcomes. Empirical researchers therefore face the difficult choice of which estimator to use when the validity of these assumptions is uncertain.

A second, related, challenge concerns how existing methods weight observations across units and over time. Classical SC assigns a single set of unit weights that must suit every post-treatment period and implicitly treats all pre-periods as equally informative, while DID imposes uniform weights for each time period and unit. Both thus ignore the fact that outcomes in different periods may have different predictive power for the target counterfactual.

A third concern is that SC methods, by relying on balancing weights, do not naturally extend to general assignment patterns.

Motivated by these observations, we introduce the Triply RObust Panel (TROP) estimator, a class of methods that is flexible in choosing counterfactual comparison groups. The core idea of the TROP estimator is to incorporate three components: time weights, unit weights, and a flexible outcome model. Each of these three components depends on a data-driven tuning parameter.

Specifically, TROP predicts outcomes for treated unit-time pairs using a parsimonious regression model (e.g., regularized low-rank factor structure) in combination with unit and time fixed effects, while in addition jointly learning two sets of data-driven weights. These weights can, in principle, take many forms and encompass existing estimators as special cases. In practice, we choose the first set of unit weights to minimize imbalance in pretreatment trajectories between treated and control units, and the sec-

ond set of time weights to discount less informative periods to allow for the possibility that more recent observations carry greater influence. However, other weights, such as time weights that include seasonal components, are also possible. Both weighting schemes depend on tuning parameters chosen via cross-validation and can accommodate complex assignments.

Together with the regression adjustment that may capture latent common factors, these elements give TROP its defining property of triple robustness: the estimator is asymptotically unbiased if either time weights, unit weights, or the regression adjustment remove underlying biases. Specifically, we show in Theorem 5.1 that the bias of the estimator depends on the *product* of unit-level imbalance between treated and control units, time-level imbalance between treatment and control periods, and bias arising from a (possibly) misspecified regression adjustment. This multiplicative bias bound is tighter than the bias bounds that govern leading alternatives such as TWFE/DID, SC, and SDID, providing TROP with stronger robustness properties.

To set the stage we compare the TROP estimator to the Synthetic Difference In Differences (SDID), Synthetic Control (SC), Difference In Differences (DID), Matrix Completion (MC) and Doudchenko-Imbens-Fernan-Pinto (DIFP) estimators in a range of realistic settings based on real data. The complete description of the simulation designs is deferred to Section 3. We compare the estimators in terms of their Root-Mean-Squared-Error (RMSE) to predict the *counterfactual*, *i.e.*, the potential control outcome in the treated periods for the treated units.¹ The counterfactual control outcomes, unobserved in a real-world study, are known to us by design and set so that treatment effects are all zero.

Table 1 reports results for twenty-one settings that are our primary focus in this study. In twenty out of the twenty-one cases the TROP estimator is the best in terms of RMSE, with in the one remaining case the standard SC estimator better. The table documents substantial heterogeneity in the performance of existing methods. Compared to the TROP estimator, DID can have an RMSE that is 900% higher, for the Basque GDP data with random assignment. The RMSE of SC can be more than 500% times bigger, for the same case. MC and SDID do somewhat better, never more than 300% worse or 93% worse than the TROP estimator, respectively. Note however, that, like the SC estimator, the SDID estimator does not extend naturally to general assignment mechanisms, whereas the TROP estimator, like the MC estimator, allows for general assignment patterns. In contrast, the TROP estimator is the best one across all designs but in one case where it

¹The RMSE here takes the square root of the average of the squared the unit/period specific errors.

only underperforms in RMSE by 24% relative to the best competitor.

The results in Table 1 reveal striking amounts of heterogeneity in performance. For some data sets the DID estimator has substantially lower RMSE than the SC estimator, and for others the opposite is true. The results clearly show that the choice between some of the estimators cannot be made once and for all. The fact that TROP performs well across environments is evidence that it effectively learns which features (time balancing, unit balancing, regression adjustments, or combinations of these) are most relevant to improve the accuracy of the estimator.

Given these results, it is natural to ask: What features of the data generating processes in Table 1 matter the most for determining performance? What components of the TROP estimator drive its better performance? In what we see as a template for conducting informative semi-synthetic simulations, in Section 4 we systematically explore the drivers of the relative performance of our method, focusing on both the features of the data generating process and the components of the estimator that lead to a better performance. We find that DID is particularly sensitive to the presence of interactive fixed effects, which lead to violations of parallel trend assumptions. Empirically, we find that interactive fixed effects are important for predicting outcomes in many of the chosen applications, spanning examples with both large and small N and T . We also find that in settings with few treated periods and few treated units other estimators become more competitive with the TROP estimator. Regarding the three components of the TROP estimator we find that the presence of the time weights, as well as combinations of the time weights and the regression adjustment can play a key role in significantly reducing RMSE. Unit weights reduce RMSE only in some applications, although they play a less prominent role.

From a theoretical perspective, the use of the three components of TROP estimator provides us with a strong triple-robustness property.

This paper connects to a broad literature on panel data models, including (i) Synthetic Control and matching-type unit weighting [AG03, ADH10], (ii) two-way fixed effect estimators [see e.g., RSBP23, DCd23]; (iii) doubly-robust combinations of the two [AAH⁺19, BMFR18, IV23]; and (iv) interactive fixed-effects and matrix-completion outcome modeling [e.g. Bai09, MW15, BN13, ABD⁺21a]. However, none of this literature has studied triple robust estimators.

Specifically, SC methods and related works [see also BMAF⁺23, SSP⁺19, BGK⁺15, HX18, CWZ18, CFT21, FP16, HS17, KMPT21] do not accommodate (i) complex assignment patterns and (ii) heterogeneous weights over time. Recognizing the complemen-

Data set	Outcome	Treatment	N	T	Normalized Out-of-sample RMSE					
					TROP	SDID	SC	DID	MC	DIFP
CPS	log-wage	min wage	50	40	1.00	1.14	1.44	1.91	1.26	1.22
CPS	urate	min wage	50	40	1.00	1.05	1.11	1.89	1.10	1.09
CPS	hours	min wage	50	40	1.00	1.19	1.22	1.25	1.11	1.20
CPS	log-wage	gun law	50	40	1.00	1.15	1.13	2.04	1.33	1.32
CPS	log-wage	abortion	50	40	1.00	1.14	1.43	1.99	1.25	1.24
CPS	log-wage	random	50	40	1.00	1.12	1.10	1.95	1.24	1.17
PennWT	log-GDP	democracy	111	48	1.00	1.44	1.59	7.85	1.76	1.54
PennWT	log-GDP	education	111	48	1.00	1.51	2.10	6.90	1.60	1.57
PennWT	log-GDP	random	111	48	1.00	1.32	1.51	4.31	1.41	1.50
Germany	GDP	random	17	44	1.00	1.46	2.82	3.58	1.56	2.46
Germany	GDP	simulated	17	44	1.00	1.79	5.82	7.58	1.93	4.72
Germany	GDP	treated unit	17	44	1.00	1.13	1.82	5.17	1.27	1.89
Basque	GDP	random	18	43	1.00	1.02	4.55	9.11	1.70	2.47
Basque	GDP	simulated	18	43	1.00	1.55	2.11	3.05	1.30	1.66
Basque	GDP	treated unit	18	43	1.00	1.35	2.79	1.25	3.02	2.94
Smoking	packs pc	random	39	31	1.00	1.22	1.48	2.16	1.14	1.45
Smoking	packs pc	simulated	39	31	1.00	1.28	1.93	2.88	1.44	1.41
Smoking	packs pc	treated unit	39	31	1.00	1.04	1.72	2.18	1.27	1.30
Boatlift	log-wage	random	44	19	1.00	1.34	1.62	1.35	1.04	1.62
Boatlift	log-wage	simulated	44	19	1.00	1.50	1.31	1.46	1.12	1.58
Boatlift	log-wage	treated unit	44	19	1.24	1.93	1.00	1.35	1.17	2.07

TABLE 1. RMSEs of six estimators for semi-synthetic simulations with data generated using CPS, PWT, Germany reunification, Basque, Smoking and Boatlift data as described in Section 3. TROP refers to our proposed estimator, SDID to the Synthetic Difference in Differences, SC to Synthetic Control, DID to Difference in Differences, MC to Matrix Completion, and DIFP to the SC estimator after recentering its mean (*i.e.*, with an intercept). RMSE calculations are out-of-sample (the square-root of the average of the squared errors at the unit/period level) and based on 1000 simulation replications. They are normalized by dividing by the smallest RMSE among all estimators. The simulation design is the same as that used in Table 2 in [AAH⁺19] and formally described in Section 3. Specifically, for CPS and PWT dataset, the treatment is simulated by estimating the probability of treatment as in Equation (8) below using data from the treatment variable in the data as indicated in the table. For the remaining designs random denotes a random treatment assignment across all units, simulated denotes a simulated treatment with probabilities corresponding to the estimated weights from a synthetic control regression and treated unit corresponds to use the true treated unit to evaluate the performance of the method (where the RMSE is evaluated for the treated unit over a pre-treatment period where therefore treatment effects must equal zero). The outcome variable is simulated from a rank 4 factor model as in Equation (6) estimated using the observed outcome from the data and as indicated in the table. Treatment effects are set to be equal to zero. N and T denotes the total number of units and periods. For each simulated dataset we consider ten treated periods and ten treated units.

tary roles of SC-style balancing and outcome modeling, SDID achieves double robustness: bias vanishes if either unit imbalance or time imbalance is negligible [AAH⁺19]. Augmented SC [BMFR18] and related augmented-inverse-propensity approaches follow a similar spirit. However, SC and SDID are (i) not applicable to contexts where treatments may switch on and off; (ii) do not allow for a more flexible regression adjustment other than the two-way fixed structure. [Sch23] also allows for some time weighting.

Models using interactive fixed effects, which under mild conditions can approximate any outcome model with sufficient factors (that is, a high rank model), can address

some of the limitations of two-way fixed-effects models. Such approaches can perform well in moderate sample sizes when a low-rank structure (small number of factors) is a very good approximation, but the approaches ignore assignment mechanisms and often require strong conditions about the signal to noise ratio [Bai09, MW17]. TROP combines these ideas with synthetic control ideas: it leverages a regression adjustment such as a low-rank model. As a result, factor misspecification affects at most one of three channels, leaving the estimator consistent provided either unit or time balancing succeeds.

We organize this paper as follows. In Section 2 we introduce the main setup and estimator focusing on a single treated unit for expositional convenience. Section 3 presents a large collection of numerical studies calibrated to relevant empirical applications. Section 4 investigates the drivers of the superiority of the proposed TROP estimator, first in terms of features of the data to which we apply the methods, and second in terms of the features of the proposed estimator. Section 5 formalizes the triple robustness properties of the estimator and discuss theoretical properties. Section 6 studies the estimator with multiple treated units, covariates and present algorithms for inference. Section 7 concludes.

2. The Triply RObust Panel (TROP) Estimator

In this section we introduce the set up and our proposed estimation procedure.

2.1. Set Up and Estimand

We observe an $N \times T$ matrix of outcomes \mathbf{Y} , and an $N \times T$ matrix of binary treatment assignments. For ease of exposition, consider a single treated unit/period pair so that $\sum_{i,t} W_{it} = 1$. The single treated unit-period pair is not essential for the TROP estimator which can deal with general assignment patterns, but allows for easier comparison with some of the previously proposed estimators, in particular SC and SDID that have not been defined for general assignment patterns.

We assume there are no spillovers and no dynamic effects. We postulate the existence of a pair of potential outcomes $(Y_{it}(0), Y_{it}(1))$ for each unit and time period with $\tau_{i,t} = Y_{it}(1) - Y_{it}(0)$ the unit/period pair specific treatment effect. We are interested in the

average effect for the treated,

$$(1) \quad \tau = \frac{\sum_{i=1}^N \sum_{t=1}^T W_{it} (Y_{it}(1) - Y_{it}(0))}{\sum_{i=1}^N \sum_{t=1}^T W_{it}}.$$

2.2. The TROP Estimator

The core idea behind our estimator is to introduce a set of time weights across different periods t , denoted by θ_t and different weights across units units j , denoted by ω_j , and to introduce a regression adjustment. We specify our working model as

$$Y_{it}(0) = \alpha_i + \beta_t + \mathbf{L}_{it} + \varepsilon_{it}, \quad \mathbb{E}[\varepsilon_{it} | \mathbf{L}] = 0$$

where we think of ε_{it} as idiosyncratic shocks and \mathbf{L} as a component we will try to predict (e.g., through a low-rank factor model). Although the unit and time fixed effects α_i and β_t can be incorporated into the low rank component \mathbf{L} (adding a rank two component), because in practice we introduce regularization to recover \mathbf{L} but not the unit and time fixed effects it will be convenient to include them separately unless otherwise specified. In principle, one could consider several estimators for \mathbf{L} , and our theory will consider a broad class of estimators.

In practice (and in our numerical studies), we propose estimating the treatment effect on unit (i, t) by minimizing

$$(2) \quad (\hat{\alpha}, \hat{\beta}, \hat{\mathbf{L}}) = \arg \min_{\alpha, \beta, \mathbf{L}} \sum_{j=1}^N \sum_{s=1}^T \theta_s^{i,t} \omega_j^{i,t} (1 - W_{js}) \left(Y_{js} - \alpha_j - \beta_s - L_{js} \right)^2 + \lambda_{nn} \|\mathbf{L}\|,$$

where $\omega_j^{i,t}, \theta_s^{i,t}$ allows for heterogeneous weights to different units and periods, and λ_{nn} imposes a penalization on the nuclear norm of the matrix \mathbf{L} , followed by estimating the treatment effect as

$$\hat{\tau}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_t - \hat{\mathbf{L}}_{it}.$$

The weights, can be implicitly functions of (i, t) and can take arbitrary form. We defer a comprehensive discussion on the choice of the weights to Section 2.3.

The formulation in Equation (2) encompasses DID, SC, MC, and SDID estimators all as special cases. That is, for $\lambda_{nn} = \infty$ and $\omega_j = \theta_s = 1$, we recover the DID/TWFE estimator. For $\omega_j = \theta_s = 1$, and $\lambda_{nn} < \infty$, we recover the MC estimator [ABD⁺21b]. For $\lambda_{nn} = \infty$ we recover the SC and SDID estimators for specific choices of the units and

time weights.

To gain further insight into the TROP estimator, note that, as we discuss further in the following section, we can interpret the weights ω as balancing weights, in the spirit of [Zub15, AIW18], the weights θ as time-varying weights, and the penalization of the nuclear norm on \mathbf{L} as imposing a low-rank outcome model structure. That is, the estimator combines flexible balancing with the flexible outcome model in the same spirit of doubly robust estimation [BR05, CCD⁺18].

2.3. Choice of the Weights

One of the challenge of our proposed estimation procedure is that, in principle, it may depend on several tuning parameters. In particular, it may depend on the choice of the weight as well as well as on the choice of the rank for \mathbf{L} . In practice, often due to data feasibility, we recommend to reduce the number of tuning parameters by imposing some particular functional form restrictions on the choice of the weights.

Here, we propose one specific choice of the weights, although our analysis is not specific to this choice of the weights, and, depending on the particular context, one may wish to generalize the time weights to accomodate seasonal patterns.

Given tuning parameters $\lambda = (\lambda_{\text{time}}, \lambda_{\text{unit}}, \lambda_{\text{nn}})$, the distance-based weights exhibit exponential decay for both the time and units weights:

$$(3) \quad \theta_s^{i,t}(\lambda) = \exp\left(-\lambda_{\text{time}} \text{dist}^{\text{time}}(s, t)\right), \quad \text{and} \quad \omega_j^{i,t}(\lambda) = \exp\left(-\lambda_{\text{unit}} \text{dist}_{-t}^{\text{unit}}(j, i)\right).$$

For the time distance we take a simple function that directly measures the number of periods:

$$\text{dist}^{\text{time}}(s, t) = |t - s|.$$

For the unit weights we use the square root of the average squared difference for the periods where both units are in the control group:

$$\text{dist}_{-t}^{\text{unit}}(j, i) = \left(\frac{\sum_{u=1}^T \mathbb{1}\{u \neq t\} (1 - W_{iu})(1 - W_{ju})(Y_{iu} - Y_{ju})^2}{\sum_{u=1}^T \mathbb{1}\{u \neq t\} (1 - W_{iu})(1 - W_{ju})} \right)^{1/2}.$$

where, with an abuse of notation, we write the weights as an explicit function of λ .

For this particular choice of weights, units that far apart (either over the time dimension or the unit-level dimension) with respect to the treated units should be weighted less in our final estimator.

Our primary task is designing an efficient method to choose the triplet $(\lambda_{time}, \lambda_{unit}, \lambda_{nn})$ based on leave-one-out cross validation. The core idea of the leave-one-out cross validation we propose here is that the predicted treatment effect on control units should be close to zero if were to consider control units as if they were treated. Define, given

$$I_{j,s}^{i,t} = 1\{(j, s) = (i, t)\},$$

(4)

$$\hat{\tau}_{it}^{\text{loocv}}(\lambda) = \arg \min_{\tau} \min_{\alpha, \beta, \mathbf{L}} \sum_{j, s: W_{j,s}=0} \omega_j^{i,t}(\lambda) \theta_s^{i,t}(\lambda) \left(Y_{js} - \alpha_j - \beta_s - L_{js} - \tau \cdot I_{j,s}^{i,t} \right)^2 + \lambda_{nn} \|\mathbf{L}\|$$

Here $I_{j,s}^{i,t}$ denotes the indicator that unit (j, s) corresponds to (i, t) and $\hat{\tau}_{it}^{\text{loocv}}(\lambda)$ corresponds to the estimated treatment effects on a given unit (i, t) , *as if* unit (i, t) was the treated unit, after removing the outcomes for the other treated observations. Intuitively, because we estimate the treatment effect $\hat{\tau}_{it}^{\text{loocv}}(\lambda)$ only over control units, such effect captures the difference between the true outcome and predicted counterfactual, after leaving out from such regression information at time t .

This intuition motivates the choice of $(\lambda_{time}, \lambda_{unit}, \lambda_{nn})$ that minimizes the criterion function

$$(5) \quad Q(\lambda) = \sum_{i=1}^N \sum_{t=1}^T (1 - W_{it}) \left(\hat{\tau}_{it}(\lambda) \right)^2$$

over a grid of values. This is equivalent to choosing the tuning parameters with the smallest out-of-sample squared error for predicting the potential outcome under control on the control units.²

3. Semi-synthetic Simulations

The goal of this section is to study the performance of our method in realistic settings. To this end, we first consider simulation studies identical to those already considered in [AAH⁺19]. These simulations are calibrated to datasets representative of those typically used for panel data studies.

²In practice, we start with the following procedure: letting $\lambda_{nn} = \infty$ and $\lambda_{unit} = 0$, we minimize Q over a grid of values of λ_{time} , and similarly for the other two penalty parameters. These optimized values provide *upper* bounds for a finer grid for joint search. Next, starting from these initial values, we select the penalty parameters by cycling through the three parameters. In other words, we successively update the three parameters holding the other two at the most recent optimal values.

Algorithm 1 TROP estimator for treatment effect with single treated unit

Require: Grid of values \mathcal{G} for $(\lambda_{time}, \lambda_{unit}, \lambda_{nn})$, treatments \mathbf{W} , outcomes \mathbf{Y}

- 1: **for each** $\lambda \in \mathcal{G}$ **do**
 - 2: For each (i, t) such that $W_{it} = 0$ estimate $\hat{\tau}_{it}(\lambda)$ as in Equation (4)
 - 3: Compute the function $Q(\lambda)$ in Equation (5)
 - 4: **end for**
 - 5: Find $\hat{\lambda} \in \arg \min_{\lambda \in \mathcal{G}} Q(\lambda)$
 - 6: Compute the estimator $\hat{\tau}(\hat{\lambda})$ as in Equation (2) with λ for the weights replaced by $\hat{\lambda}$ found in the previous step
-

3.1. Simulation design

As a first step, we summarize the simulation designs from [AAH⁺19] used to create placebo studies based on (1) PWT World Table (PWT) data for international log-GDP, and (2) aggregated Current Population Survey (CPS) data for U.S. states.³ In both settings, we generate potential outcomes via a latent factor model and assign treatment via a nonrandom mechanism that replicates empirically observed policy patterns. As noted in [AAH⁺19], the first simulation mimics the relevant settings for when DID is used in practice, and the second for when synthetic control is used in practice.

Both the CPS-based and PWT-based placebo studies yield balanced panel data with: (1) outcomes driven by a latent-factor structure plus autocorrelated errors, and (2) treatment indicators that depend non-randomly on either fixed effects or the interactive factor component. These designs allow for systematic confounding that would invalidate simpler difference-in-means comparisons, while still retaining enough structure (through the factor model) to examine estimators that correct for confounding.

Latent-Factor Outcome Model. Let Y_{it}^* be the observed outcome (e.g., average log-wage) in state i and year t . As in [AAH⁺19], we approximate Y_{it}^* via a low-rank factor model

$$L_{it} = \arg \min_{L: \text{rank}(L) \leq 4} \sum_{i,t} (Y_{it}^* - L_{it})^2,$$

thus obtaining a matrix $\mathbf{L} = (L_{it})$ that captures common trends and cross-sectional variation. We estimate an autoregressive model for the residuals $Y_{it}^* - L_{it}$ to capture residual serial correlation within each state. Denote the resulting covariance matrix by

³For CPS data, we use data for female workers in the March outgoing rotation groups of the Current Population Survey (CPS), spanning selected years in 1979–2019. The original dataset includes average (log) wages by state-year cells, where i indexes states and t indexes years.

Σ . The simulated outcomes Y_{it} in the placebo experiments then follow

$$(6) \quad Y_{it} = L_{it} + \varepsilon_{it},$$

with $(\varepsilon_{i,1}, \dots, \varepsilon_{i,T}) \sim \mathcal{N}(0, \Sigma)$ from a multivariate Normal distribution with the estimated covariance structure Σ . As in [AAH⁺19], for interpretation, we further decompose \mathbf{L} into an additive component and interactive component, where

$$(7) \quad F_{i,t} := \alpha_i + \beta_t = \frac{1}{T} \sum_{t=1}^T L_{it} + \frac{1}{N} \sum_{i=1}^N L_{i,t} - \frac{1}{NT} \sum_{it} L_{it}, \quad M_{i,t} := L_{i,t} - F_{i,t}.$$

We combine the factor structure with idiosyncratic noise exhibiting serial correlation (modeled as an AR(2) process) to reflect realistic dynamics. All outcomes are normalized to have mean zero and unit variance for comparability. We consider treatment effects equal to zero.

Treatment Assignment. We create a (placebo) treatment indicator D_i for each state i by employing a logistic regression on the latent factors and fixed effects. Specifically,

$$(8) \quad D_i \mid (\alpha_i, M_i, \varepsilon_{i,\cdot}) \sim \text{Bernoulli}(\pi_i), \quad \text{where} \quad \pi_i = \frac{\exp(\phi_\alpha \alpha_i + \phi_M M_i)}{1 + \exp(\phi_\alpha \alpha_i + \phi_M M_i)},$$

and α_i and M_i respectively capture additive and interactive components of the fitted factor model. We estimate (ϕ_α, ϕ_M) by regressing actual state-level policy indicators on α_i and M_i . Once D_i is drawn, we treat states with $D_i = 1$ as “exposed” to a placebo intervention for all $t > T_0$.

For the CPS data, the model for the treatment is estimated using existing policies such as minimum wage policy (default), gun law policies and abortion policies [AAH⁺19]. For the PWT Data, we estimate the model using democracy status and education metrics also similar to [AAH⁺19]. We examine both a realistic assignment settings, as well as randomized assignments.

In our baseline setup with CPS data, we simulate panel data with $N = 50$ units and $T = 40$ periods (including $T_{\text{pr}} = 30$ pre-treatment periods and $T_{\text{post}} = 10$ post-treatment periods), with at most $N_{\text{tr}} = 10$ number of treated units. For the PWT data, we have $N = 111$, $T = 48$, and keep $T_{\text{post}} = 10$ and $N_{\text{tr}} = 10$. In addition, we vary $T_{\text{post}}, N_{\text{tr}}$ as we describe below.

Four additional datasets: Germany, Basque, Smoking and Boatlift datasets. In addition to the CPS and PWT Dataset, we also conduct simulations using the Germany reunification dataset studied by [ADH15b], the Basque country dataset studied in [ADH10], the smoking dataset in [ADH10] and the boatlift dataset [PY15].

A key challenge in these applications is that there is a single treated unit, making the simulation design harder. Therefore, we consider three DGPs for the treatment assignment matrix. First, we consider settings where the treatment is randomly drawn from the pool of control units (“uniform random”). Second, we consider the one where treatment probability equals the weight estimated through a penalized Synthetic Control regression. This emulates settings where treatment may be confounded by some unobservable components. Third, we report results in terms of predicting the outcome of the treated unit over pre-treatment periods (therefore fixing the choice of the treated unit as the one in the given application). We simulate the outcomes following the same design as for the CPS and PWT dataset, calibrated here to the corresponding Germany, Basque, Smoking, and Boatlift datasets.

3.2. Main simulation results

Table 1 collect results across all specification considered (both across different outcomes and treatments), twenty-one in total and focusing on ten treated units and ten treatment periods. The TROP estimator outperforms all competitors in all but only one specification, for which it underperforms by 24%. In contrast, DID can have RMSE up 900% higher than the best competitor, SC 580%, MC 300%, and SDID 90%. (In addition, Appendix Table A2 shows that both TROP and SDID have substantially smaller (no) bias compared to the other estimators focusing for brevity on CPS and PWT data. A standard TWFE estimator shows noticeable bias when the treatment is not randomly assigned.)

We then zoom in into seven specifications, one with CPS data using log-wage and unemployment rate as outcomes, and minimum wage as treatment; PWT data using log-GDP as outcome and democracy as treatment, and the Smoking, Boatlift, Basque and Germany datasets with randomly assigned treatments (keeping in mind that Table 1 and the Appendix show robust results for a wide range of other non-random assignment mechanisms). Specifically, Table 2 zoom-in on six specific designs, with Table 2 providing details about the outcome-generating process and reporting RMSE in absolute terms (instead of relative terms), together with cross-validated values of parameters $(\lambda_{unit}, \lambda_{time}, \lambda_{nn})$ for the TROP estimator. It collects results for both one and ten treated

units and treatment periods. The TROP estimator also with one treated unit and period outperforms all competitors in half of the cases, and it slightly under-performs (by no more than 25%) in the remaining cases. Appendix Table A1 reports results for one treated periods, and ten treated units and vice-versa, illustrating similar results to those observed with ten treatment periods and ten treated units.

In addition, across all designs, we find substantial variability in the optimal choice of the tuning parameters. Unit, time, and regression adjustments all may play a role for the estimator depending on the choice of the outcome and dataset, in some but not all cases. These results are suggestive of the importance of constructing an estimator that learns the optimal components (time, unit weights, and factor model) from the data directly. In the next section we explore the main drivers of the performance.

Dataset	Treatment	$T_{\text{post}} = 10, N_{\text{tr}} = 10$												
		N	T	$\frac{\ F\ _F}{\sqrt{NT}}$	$\frac{\ M\ _F}{\sqrt{NT}}$	$\sqrt{\frac{\text{Tr}(\Sigma)}{T}}$	AR(2)	$(\lambda_{\text{unit}}, \lambda_{\text{time}}, \lambda_{nn})$	RMSE					
									TROP	SDID	SC	DID	MC	DIFP
CPS logwage	min wage	50	40	0.992	0.10	0.098	(0.010, -0.057)	(0, 0.1, 0.9)	0.025	0.029	0.037	0.049	0.032	0.032
CPS urate	min wage	50	40	0.770	0.40	0.604	(0.039, 0.011)	(1.6, 0.35, 0.011)	0.203	0.214	0.226	0.384	0.224	0.221
PWT	democracy	111	48	0.972	0.229	0.069	(0.913, -0.221)	(0.3, 0.4, 0.006)	0.023	0.036	0.040	0.198	0.045	0.039
Germany	random	17	44	0.994	0.108	0.032	(0.840, -0.200)	(1.2, 0.2, 0.011)	0.025	0.037	0.065	0.073	0.032	0.038
Basque	random	18	43	0.986	0.167	0.037	(1.228, -0.577)	(0, 0.35, 0.006)	0.041	0.070	0.109	0.145	0.069	0.082
Smoking	random	39	31	0.935	0.337	0.166	(0.506, -0.058)	(0.25, 0.4, 0.011)	0.085	0.108	0.132	0.192	0.102	0.129
Boatlift	random	44	19	0.913	0.344	0.276	(0.031, -0.087)	(0.2, 0.2, 0.151)	0.115	0.155	0.187	0.156	0.121	0.187

Dataset	Treatment	$T_{\text{post}} = 1, N_{\text{tr}} = 1$												
		N	T	$\frac{\ F\ _F}{\sqrt{NT}}$	$\frac{\ M\ _F}{\sqrt{NT}}$	$\sqrt{\frac{\text{Tr}(\Sigma)}{T}}$	AR(2)	$(\lambda_{\text{unit}}, \lambda_{\text{time}}, \lambda_{nn})$	RMSE					
									TROP	SDID	SC	DID	MC	DIFP
CPS logwage	min wage	50	40	0.992	0.10	0.098	(0.010, -0.057)	(0, 0.2, 0.105)	0.111	0.121	0.127	0.156	0.123	0.121
CPS urate	min wage	50	40	0.770	0.40	0.604	(0.039, 0.011)	(2, 0.03, 0.325)	0.637	0.713	0.688	0.629	0.701	0.666
PWT	democracy	111	48	0.972	0.229	0.069	(0.913, -0.221)	(1.8, 4.4, 0.01)	0.052	0.056	0.116	0.294	0.059	0.083
Germany	random	17	44	0.994	0.108	0.032	(0.840, -0.200)	(0.1, 3.8, 0.01)	0.037	0.044	0.105	0.136	0.056	0.076
Basque	random	18	43	0.986	0.167	0.037	(1.228, -0.577)	(0.3, 0.2, 0.006)	0.039	0.033	0.180	0.333	0.055	0.072
Smoking	random	39	31	0.935	0.337	0.166	(0.506, -0.058)	(0.5, 0.5, 0.0163)	0.200	0.190	0.279	0.464	0.208	0.218
Boatlift	random	44	19	0.913	0.344	0.276	(0.031, -0.087)	(0.1, 0.4, 0.2014)	0.324	0.362	0.394	0.494	0.351	0.363

TABLE 2. RMSEs of six estimators for semi-synthetic simulations with data generated using CPS, PWT, Germany reunification, Basque, Smoking and Boatlift data as described in Section 3. TROP refers to our proposed estimator, SDID to the Synthetic Difference in Differences, SC to Synthetic Control, DID to Difference in Diferences, MC to Matrix Completion, and DIFP to the SC estimator after recentering its mean (*i.e.*, with an intercept). RMSE calculations are out-of-sample (the square-root of the average of the squared errors at the unit/period level) and based on 1000 simulation replications. The simulation design is the same as that used in Table 2 in [AAH⁺19] and formally described in Section 3. Specifically, the outcome variable is simulated from a rank-4 factor model as in Equation (6) estimated using the observed outcome in the data (log-wage and unemployment rate for CPS data, log-GDP for PWT data, GDP for Germany and Basque dataset, cigarettes sold for the smoking dataset and log-wage for the Boatlift dataset). For CPS and PWT dataset, the treatment is simulated by estimating the probability of treatment as in Equation (8) using data from the treatment variable (minimum wage and democracy). For Germany, Basque, Smoking and Boatlift dataset, treatment is randomly assigned with uniform probabilities across units. Columns 5 - 8 defines features of the generating process for the outcome estimated as described in Section 3.1. $(\lambda_{\text{unit}}, \lambda_{\text{time}}, \lambda_{nn})$ denote the estimated tuning parameters for TROP, estimated via cross-validation. N and T denotes the total number of units and periods. The first panel corresponds to ten post-treatment periods and ten treated units and the second panel corresponds to one post-treatment period and one treated unit.

4. Drivers of the Superior Performance of the TROP Estimator

Why do the six estimators we compare perform so differently across the simulation designs, and in particular, why does the TROP estimator perform so well relative to the other estimators? Is it features of the data that drive this, and might it be different for other data sets? Is it features of the estimator that particularly matter, and could one build those features into other estimators? In this section we investigate these questions by systematically exploring changes in the data generating processes and in the TROP estimator. The goal is to give the empirical researcher assistance in choosing the estimator in their own research (which may differ substantially from that covered by the six designs explored in the previous section).

4.1. Critical Features of the Data Generating Process

First, we inspect the key components in the data-generating process that drive the performance of the estimators. To do so, we first “shut down” relevant components in the data-generating process, with the results presented in Table 3 for the CPS and PWT datasets. Specifically, starting from the baseline specification (for both CPS and PWT data), we first remove correlation in the residuals and observe similar behavior as in our main simulation results.

We then remove interactive effects (the M component in Equation (7)). In this case we see that DID (and other methods such as MC) perform competitively with TROP. For DID this behavior is expected as DID or related methods are designed to tackle additive (non-interactive) unobserved confounders. For MC this may be tied to the regularization we use for the factor component. This result highlights the importance of interactive fixed effects in driving the superior performance of our estimator over existing ones.

Given the sensitivity of DID to the presence of interactive components, we next investigate how important interactive fixed effects are in empirical relevant settings. To study this question, we document how much residual variation is explained by an interactive fixed effect model over a TWFE model. We report results in Table 4. Across all our datasets, adding even a single additional interactive fixed effect can significantly decrease the root-mean-squared error (RMSE) between 10 and 60%.

We conduct a formal test to determine whether to include an interactive fixed effect as follows: for each application, we divide the panel into two halves based on the total number of observations, T . We then test whether unit-fixed effects (for each unit) estimated in the first half of observations are the same as in the second half. Rejection

Design	$\frac{\ F\ _F}{\sqrt{NT}}$	$\frac{\ M\ _F}{\sqrt{NT}}$	$\sqrt{\frac{\text{tr}(\Sigma)}{T}}$	AR(2)	$(\lambda_{unit}, \lambda_{time}, \lambda_{nn})$	RMSE						
						TROP	SDID	SC	DID	MC	DIFP	
Panel A: CPS Data (Y: log-wage; W: min wage)												
Baseline	0.992	0.10	0.098	(0.010, -0.057)	(0, 0.1, 0.9)	0.025	0.029	0.037	0.050	0.032	0.032	
<i>Simpler outcome models</i>												
No Auto Corr	0.992	0.10	0.098	(0, 0)	(0.7, 0.25, 0.6)	0.027	0.029	0.037	0.049	0.033	0.032	
No M	0.992	0	0.098	(0.010, -0.057)	(0.1, 0.025, 0.121)	0.015	0.018	0.018	0.013	0.015	0.017	
No F	0	0.10	0.098	(0.010, -0.057)	(1.4, 0.25, 0.301)	0.026	0.029	0.022	0.049	0.032	0.032	
Only Noise	0	0	0.098	(0.010, -0.057)	(1.8, 0.005, 0.4)	0.015	0.018	0.014	0.013	0.015	0.017	
Panel B: PWT Data (Y: log-GDP; W: democracy)												
Baseline	0.972	0.229	0.069	(0.913, -0.221)	(0.3, 0.4, 0.006)	0.023	0.036	0.040	0.198	0.046	0.039	
<i>Simpler outcome models</i>												
No Auto Corr	0.972	0.229	0.069	(0, 0)	(0.2, 0.3, 0.016)	0.027	0.030	0.033	0.197	0.046	0.033	
No M	0.972	0	0.069	(0.913, -0.221)	(0, 0.12, 0.051)	0.015	0.016	0.019	0.016	0.014	0.018	
No F	0	0.229	0.069	(0.913, -0.221)	(1.6, 0.2, 0.041)	0.028	0.036	0.028	0.198	0.046	0.039	
Only Noise	0	0	0.069	(0.913, -0.221)	(2, 0.1, 0.142)	0.015	0.016	0.016	0.016	0.014	0.018	

TABLE 3. *Simple outcome models with CPS log-wage (Panel A) and PWT log-GDP (Panel B) data: RMSEs of six estimators for semi-synthetic simulations with data generated using CPS, PWT, Germany reunification, Basque, Smoking and Boatlift data as described in Section 3. TROP refers to our proposed estimator, SDID to the Synthetic Difference in Differences, SC to Synthetic Control, DID to Difference in Differences, MC to Matrix Completion, and DIFP to the SC estimator after recentering its mean (*i.e.*, with an intercept). RMSE calculations are out-of-sample (the square-root of the average of the squared errors at the unit/period level) and based on 1000 simulation replications. The simulation design is the same as that used in Table 2 in [AAH⁺19] and formally described in Section 3. Specifically, the treatment is simulated by estimating the probability of treatment as in Equation (8) using data from the treatment variable. The outcome and treatment variable used in simulations is log-wage for CPS data and log-GDP for PWT data, and treatments are minimum wage for CPS and democracy for PWT data. Columns 2 - 4 defines features of the generating process for the outcome estimated as described in Section 3.1. $(\lambda_{unit}, \lambda_{time}, \lambda_{nn})$ denote the estimated tuning parameters for TROP, estimated via cross-validation. Here we consider ten treated periods and ten treated units.*

of this hypothesis is supportive of interactive fixed effects. We reject this hypothesis for more than 58% of units in each of these applications. This provides further evidence of the importance of interactive fixed effects across all these applications.

We complement these tests by reporting the bias in simulations corresponding to failing to account for interactive fixed effects via a DID estimator. Figure 1 shows the bias of a DID estimator across 1000 simulation replications as the treatment is generated by the logistic assignment with a single-factor model. The figure shows that selection into treatment can strongly depend on an unobserved interactive component, in which case the DID may suffer a large bias.

Overall, these results are suggestive that, although methods such as DID can perform well in the absence of interactive effects, interactive effects are important in applications and can introduce large misspecification bias for models that ignore interactive fixed effect components.

Finally, as we remove the fixed effect component (F in Equation (7)), Table 3 shows that the Synthetic Control method performs well and similar (or slightly better) to the TROP estimator. This is suggestive that the advantage of the TROP estimator is to jointly capture interactive and fixed effect components through weighting and regression adjustments.

dataset	(N, T)	RMSE			rejection % at 5% level
		TWFE	plus factor	% decrease	
CPS logwage	(50, 40)	0.13	0.10	23%	76%
CPS unemp-rate	(50, 40)	0.70	0.62	11%	58%
PennWT	(111, 48)	0.25	0.11	56%	91%
Germany	(17, 44)	0.14	0.06	57%	88%
Basque	(18, 43)	0.19	0.08	58%	100%
Smoking	(39, 31)	0.37	0.20	46%	79%
Boatlift	(44, 19)	0.48	0.39	19%	66%

TABLE 4. Out of sample RMSE of TWFE and TWFE plus a leading factor estimated from the data. The column “% decrease” indicates the percentage decrease in RMSE by introducing a single-factor for predicting the outcome across the different applications. The last column indicates the percentage of rejections of hypothesis that unit fixed effects do not vary in the first vs. second halves of time periods. These results are suggestive of the relevance of a factor model to better approximate the outcomes in the applications considered in our main simulations.

4.2. Varying T_{pre} , N_{co} , T_{tr} , and N_{tr}

In this section, we explore the role that the sample size of the different subsamples plays in the performance of each estimator. For the sake of brevity, we focus on the

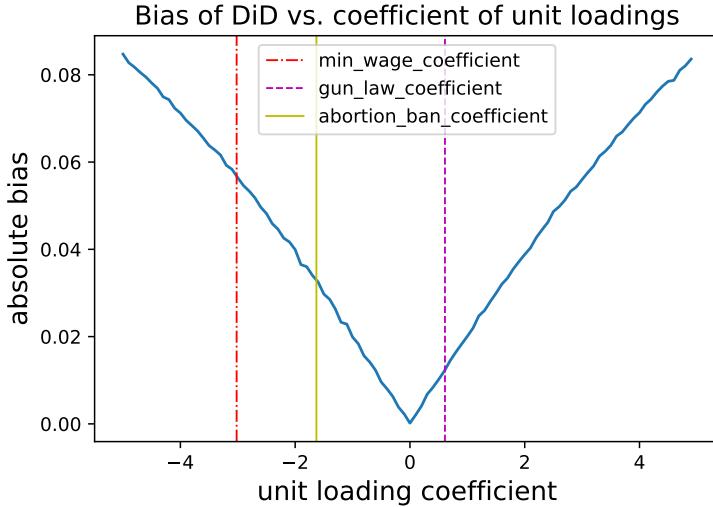


FIGURE 1. Bias of DID estimator on CPS data obtained through 10000 replications where the treatment assignment mechanism follows a logistic model as a function of unobserved factor loadings as described in Equation (8).

PWT data with the treatment and outcome generated as described in Section 3.1 using for the outcome log-GDP and for the treatment democracy.

In the baseline simulation, we have $N_{\text{tr}} = 10$ treated units, $N_{\text{co}} = 101$ control units, $T_{\text{post}} = 10$ post-treatment periods, and $T_{\text{pre}} = 38$ pre-treatment periods. In this baseline case, the RMSE of the six estimators is 2.5% (TROP), 3.6% (SDID), 4% (SC), 19.8% (DID), 4.6% (MC), and 3.9% (DIFP) (in percentages of the variance of the outcome).

We vary N_{co} , N_{tr} , T_{post} , and $T_{\text{pre}} = 38$, one at a time, to see how the relative performance of the estimators changes. In each case, we plot the RMSE of the six estimators (TROP, SDID, SC, DID, MC, and DIFP) as a function of the single feature we are varying. In all cases, the plot includes the baseline case with $N_{\text{tr}} = 10$, $N_{\text{co}} = 101$, $T_{\text{post}} = 10$, and $T_{\text{pre}} = 38$. This case is always on the far right of the four figures.

For the first plot in Figure 2 (left-panel), we vary the number of control units between 10 to 101. For a fixed value of N_{co} , we randomly select N_{co} control units from the set of 101 that are available in the full sample. On the right panel in Figure 2, we replicate the same exercise where we increase the number of pretreatment time periods from 5 to 38. In Figure 3, we vary the number of treated units from 1 to 10 (the left panel) and the number of treated periods from 1 to 10 (the right panel).

For a small number of control units, we find that SDID, DIFP, and TROP perform similarly, while MC, SC, and DID perform poorly. As the number of control units increases, TROP dominates all other competitors considered. The DID estimator benefits slightly from more control units (not surprisingly as two of the four averages in that

estimator will now be estimated more precisely), but not much because the increase in the number of control units does not improve the bias. The estimators that assign different weights to different units benefit more than the DID estimator because the increase in the number of control units will remove some of the bias (conditional on the unit selected for treatment).

Increasing the number of pretreatment periods has multiple effects. One effect is increasing the sample size and therefore improving precision. However, with the treated periods always at the end of the panel, the bias for estimators that do not allow for variation in the time weights is likely to increase. We see that clearly with the DID estimator. The other estimators do not deteriorate as much, but all do to some extent, suggesting that further research into optimal time weights beyond our simple exponential decline might be productive.

To further explore the behavior of the estimators, in Figure 3 we vary the number of treated units from one to ten and similarly the number of treatment periods. Increasing the number of treated units slightly improves the DID estimator, but leads to a much bigger improvement for the estimators that select the control units more deliberately, and leads to a larger relative improvement in the TROP estimator compared to the other estimators. Increasing the number of treatment periods, with these periods always coming after the pre-treatment periods, leads to a deterioration in the performance of all the estimators, but the relative performance of the TROP estimator improves.

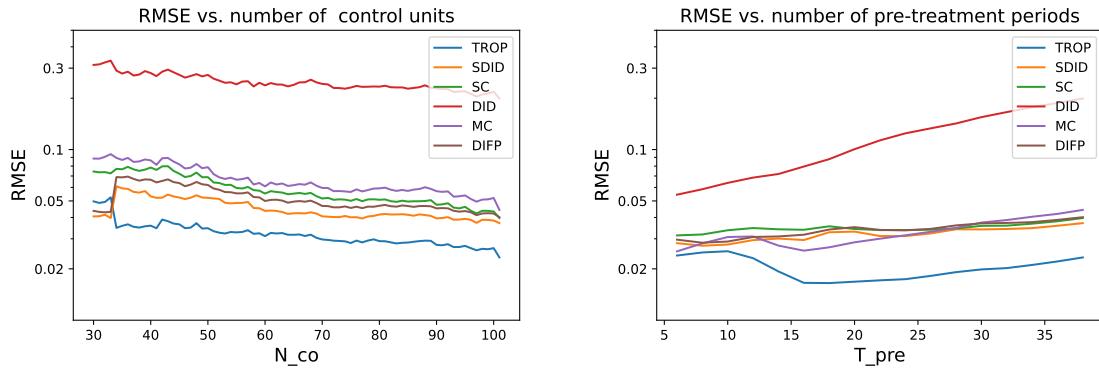


FIGURE 2. RMSEs of estimators on the PWT dataset with the outcome and treatment generated as described in Section 3.1, using as the outcome the log-GDP as the outcome and as the treatment democracy. In the figure, we vary the number of control units (left) or the number of pre-treatment periods (right). In both cases, the number of control periods and units is ten, with the treatment periods corresponding to the last ten periods in the simulated panel. When we vary the number of control units, we select randomly N units from the full PWT panel while keeping the total number of period $T = 48$. When we select the number of pre-treatment periods, we construct shorter panels by selecting the last T periods of the full PWT panel while keeping $N = 111$.

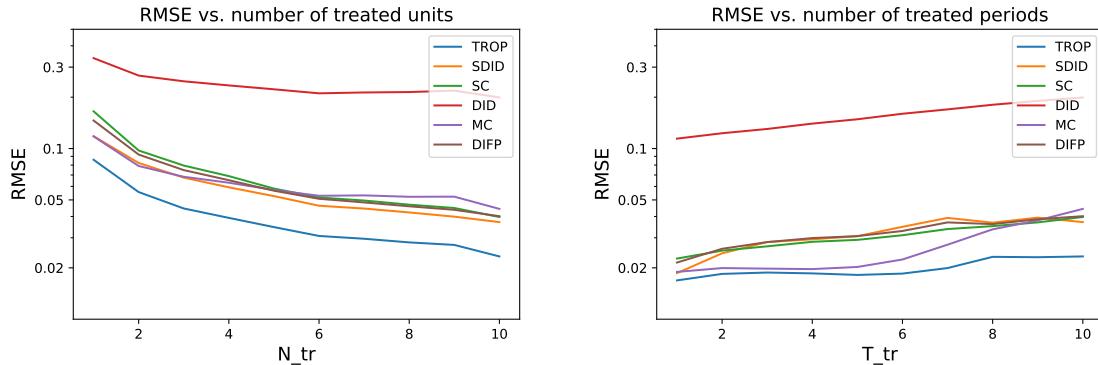


FIGURE 3. RMSEs of estimators on the PWT dataset with the outcome and treatment generated as described in Section 3.1, using as the outcome the log-GDP as the outcome and as the treatment democracy. Here we vary the number of treated units (left) or the number of treatment periods units (right). For the panel on the left, we use $N_{co} = 101$ and $T_{co} = 38$, while varying $N_{tr} = 1, \dots, 10$ with $T_{tr} = 10$. For the panel on the right, we use $T_{co} = 38$ and $N_{co} = 111$, while varying $T_{tr} = 1, \dots, 10$ with $N_{tr} = 10$.

4.3. Critical Features of the TROP Estimator

To better understand what drives the strong performance of TROP, we consider several variants where we shut down components of the TROP estimator. This includes shutting down the factor model (selecting $\lambda_{nn} = \infty$), or the time weights (selecting $\lambda_{time} = 0$) or the unit weights (selecting $\lambda_{unit} = 0$) or combinations of those.

Table 5 illustrates the results across the six main simulation designs. As in our main specifications, we focus on CPS data using treatment generated from minimum wage and unemployment rate, PWT dataset using democracy for treatment, and for the remaining four datasets, we consider treatment assigned randomly across the units. We consider ten treated units and periods (Appendix Table A4 reports results for one treated unit and period with similar results to the ones described here). We report the RMSE relative to the original TROP estimator across the different dataset (Appendix Table A3 reports the RMSE in absolute term and bias of the estimator).

For ten treated units and post-treatment periods, we find that removing the regression adjustment ($\lambda_{nn} = \infty$) or the time weights ($\lambda_{time} = 0$) increases significantly the RMSE, up to 85% for $\lambda_{nn} = \infty$ and 91% for $\lambda_{time} = 0$. Interestingly, removing unit weights ($\lambda_{unit} = 0$) increases RMSE only in some but not all cases, with a maximum increase in RMSE 10%. Removing both λ_{unit} and λ_{nn} or both λ_{unit} and λ_{time} can significantly increase RMSE to 90%. Setting $\lambda_{time} = 0$ and $\lambda_{nn} = \infty$ can have the largest effect on RMSE among all combinations of two-out-of-three parameters, with an increase in RMSE up-to five times greater.

At the same time, we observe substantial heterogeneity across the different applications, where the different weights or regression adjustments can play different roles

depending on the particular application. For one treated unit and period (Panel B in Table 5) we similarly find large heterogeneity in the role played by each of the three components. (Also, for a few specifications with one treated unit and period, TROP with no constraints slightly underperforms by one to three percentage points to introducing some constraints due to approximation error.)

In summary, Table 5 highlights the key advantage of the TROP of learning from the data how to weight observations across units and time, combined with a flexible predictor model: failing to include time weights and regression adjustment together, for example, may more than double RMSE in some applications, while in others a similar drop in performance occurs when failing to use time and unit weights.

Panel A: $T_{\text{post}}, N_{\text{tr}} = 10$	CPS logwage	CPS urate	PWT	Germany	Basque	Smoking	Boatlift
TROP	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$(\lambda_{nn} = \infty)$	1.00	1.21	1.65	1.48	1.85	1.46	1.55
$(\lambda_{unit} = 0)$	1.00	1.10	1.04	1.00	1.00	1.04	1.00
$(\lambda_{time} = 0)$	1.28	1.11	1.91	1.28	1.51	1.20	1.05
$(\lambda_{unit} = \lambda_{time} = 0)$	1.28	1.10	1.91	1.28	1.51	1.20	1.05
$(\lambda_{unit} = 0, \lambda_{nn} = \infty)$	1.00	1.22	1.70	1.48	1.85	1.46	1.55
$(\lambda_{time} = 0, \lambda_{nn} = \infty)$	1.92	1.78	5.35	2.84	3.54	2.26	1.36
DID	1.96	1.89	8.61	2.92	3.54	2.26	1.36
$(\lambda_{unit} = \lambda_{time} = 0, \lambda_{nn} = \infty)$							

Panel B: $T_{\text{post}}, N_{\text{tr}} = 1$	CPS logwage	CPS urate	PWT	Germany	Basque	Smoking	Boatlift
TROP	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$(\lambda_{nn} = \infty)$	1.00	0.99	1.00	1.00	1.00	1.02	1.13
$(\lambda_{unit} = 0)$	1.00	1.03	1.03	1.00	1.09	1.08	1.12
$(\lambda_{time} = 0)$	1.01	0.99	1.15	1.51	4.03	1.15	1.01
$(\lambda_{unit} = \lambda_{time} = 0)$	1.04	1.09	1.15	1.51	1.31	1.17	1.00
$(\lambda_{unit} = 0, \lambda_{nn} = \infty)$	1.01	0.99	1.03	1.00	1.09	1.08	1.14
$(\lambda_{time} = 0, \lambda_{nn} = \infty)$	1.23	0.99	9.85	3.08	6.51	3.38	1.42
DID	1.41	0.99	11.06	3.68	4.37	3.61	1.38
$(\lambda_{unit} = \lambda_{time} = 0, \lambda_{nn} = \infty)$							

TABLE 5. Simulation designs as in Table 1, where we report the TROP estimator (first row) and the TROP estimator constraining one or combinations of λ_{time} , λ_{unit} to be zero (corresponding to no time or unit weights) and λ_{nn} to be infinity (corresponding to no regression adjustment). Each column reports the RMSE relative to TROP within each dataset column (lower is better; TROP = 1.00). The number of treated units is ten and the number of treated periods is ten for Panel A and one for each in Panel B. The treatment simulated as in Equation (8) is calibrated for CPS log-wage and CPS unemployment rate to minimum wage, for PWT to the democracy indicator, and for Germany, Basque, Smoking and Boatlift is randomly assigned.

Table 6 further illustrates key features of the TROP estimator with ten treated periods and units, showing how the cumulative weight between units and time periods changes between the different datasets. Across all applications more than 50% of the weight is assigned to the closest five periods, while unit weights are less concentrated in applications. This suggests the relevance of time weighting in particular to improve the performance in realistic applications.

Dataset	treatment	N	T	last 5 periods	Cumulative Weight	
					closest 5 units	closest half units
CPS logwage	min wage	50	40	0.56	0.13	0.59
CPS unemp-rate	min wage	50	40	0.66	0.23	0.67
PWT log-GDP	democracy	111	48	0.70	0.06	0.59
Germany	random	17	44	0.50	0.37	0.55
Basque	random	18	43	0.66	0.28	0.50
Smoking	random	39	31	0.70	0.15	0.56
Boatlift	random	44	19	0.51	0.13	0.54

TABLE 6. Cumulative weights across time or units of the TROP estimator based on the cross-validated tuning parameters across applications. The number of treated units and the number of treated periods are both 10.

5. Formal analysis

In this section we provide an intuitive description of properties of our estimator. To facilitate our comparison with existing procedures, consider estimating the counterfactual in a setting where $W_{i,t} = 1\{i > N_0\}1\{t > T_0\}$ for fixed (N_0, T_0) . Let $T_1 = T - T_0 = 1, N_1 = N - N_0 = 1$. Therefore, we have a single treated unit and treatment period. We denote the weights as

$$\omega \in \mathbb{R}^{N_0}, \theta \in \mathbb{R}^{T_0}$$

denoting unit and time weights, where we suppressed their dependence with the index of the treated unit and period given our focus on a single treated unit (the analysis extends to multiple treated units and treatment periods, omitted for expositional convenience).

We will discuss theoretical properties for *given* weights, interpreting these as probability limits of estimated weights similar to [AAH⁺19] and [IV23]). As [AAH⁺19] show, assuming deterministic weights is almost without loss whenever $N_0 \gg N_1, T_0 \gg T_1$, that is, whenever N_0 and T_0 are sufficiently larger than the error from the post-treatment period and treated units. This is because the estimation error of the weights depends on N_0 and T_0 whereas the estimation error of the counterfactual also depends on N_1 and T_1 typically much smaller. Therefore, we will focus on settings where the weights converge to *some* probability limit, with their corresponding estimation error being negligible relative to the error of the post-treatment idiosyncratic shocks, as we further discuss in Section 5.3. The weights sum up to one, so that $1^\top \omega = 1^\top \theta = 1$.

Assumption 1. Suppose that the following holds: for all (i, t)

(i) $W_{i,t} = 1\{i > N_0\}1\{t > T_0\}$ for some constants (N_0, T_0) .

(ii) for some $\Gamma_i, \Lambda_t \in \mathbb{R}^K$ (for arbitrary K)

$$Y_{i,t}(0) = \mathbf{L}_{i,t} + \varepsilon_{i,t}, \quad \mathbf{L}_{i,t} = \Gamma_i^\top \Lambda_t, \quad \mathbb{E}[\varepsilon_{i,t} | \mathbf{L}] = 0.$$

Define $\Gamma = [\Gamma_1, \dots, \Gamma_N]^\top$ and $\Lambda = [\Lambda_1, \dots, \Lambda_T]^\top$, so that $\mathbf{L} = \Gamma \Lambda^\top$.

Assumption 1(ii) imposes a factor model structure on the outcome of interest with K being arbitrary (possibly growing with the sample size). Note that for arbitrary factor models, Γ and Λ are defined up-to a rotation. This will not affect our analysis given that our goal is to predict $\mathbf{L}_{i,t}$, and not separately each factor. Finally, because here we consider general estimators, without loss (and with a slight abuse of notation) we directly incorporate unit and time fixed effects as part of Γ_i, Λ_t (assuming that the first entry is constant for Λ_t and the second entry is constant for Γ_i).

For our analysis, we introduce some notation discussed in Table 7.

TABLE 7. Notation.

Description	Mathematical Formulation
Average Control Loadings	$\bar{\Gamma}_0 = \frac{1}{N_0} \sum_{j \leq N_0} \Gamma_j$
Average Control Factors	$\bar{\Lambda}_0 = \frac{1}{T_0} \sum_{s \leq T_0} \Lambda_s$
Weighted Average Control Loadings	$\bar{\Gamma}_0(\omega) = \sum_{j \leq N_0} \omega_j \Gamma_j$
Weighted Average Treatment Loadings	$\bar{\Lambda}_0(\theta) = \sum_{s \leq T_0} \theta_s \Lambda_s$
Unit Discrepancy	$\Delta^u(\omega, \Gamma) = \bar{\Gamma}_0(\omega) - \Gamma_N$
Time Discrepancy	$\Delta^t(\theta, \Lambda) = \bar{\Lambda}_0(\theta) - \Lambda_T$

5.1. TROP as a balancing estimator and triple robustness

Define

$$(9) \quad \hat{\tau}^{\text{TROP}}(0, \theta, \omega),$$

the estimated average counterfactual from the TROP estimator as a function of given weights θ, ω and potential outcome under control on the treated unit, absent idiosyncratic noise. We can write

(10)

$$\hat{\tau}^{\text{TROP}}(0, \theta, \omega) = \hat{\mathbf{L}}_{NT} + \left(\sum_{t \leq T_0} \theta_t (Y_{Nt} - \hat{\mathbf{L}}_{Nt}) + \sum_{i \leq N_0} \omega_i (Y_{iT} - \hat{\mathbf{L}}_{iT}) - \sum_{t \leq T_0} \sum_{i \leq N_0} \theta_t \omega_i (Y_{it} - \hat{\mathbf{L}}_{it}) \right).$$

That is, we can interpret the estimated counterfactual through the lens of regression adjustments: the weights θ, ω serve as balancing weights over time and units respectively. The estimator $\hat{\mathbf{L}}$ serves as a regression adjustment. Different from previous sections, here $\hat{\mathbf{L}}$ can be constructed using several estimators of interest (without necessarily using a nuclear norm penalization).

The estimated ATT reported by our proposed procedure reads as

$$\hat{\tau} = Y_{NT} - \hat{\tau}^{\text{TROP}}(0, \theta, \omega).$$

It is natural to ask whether, by including an additional regression adjustment, TROP is asymptotically unbiased as $N_0, T_0 \rightarrow \infty$ for τ even in the absence of perfect balance between units or time. The following theorem shows that the proposed estimator inherits a triple robustness property.

To establish this result we impose the following restriction on the class of estimators for $\mathbf{L}_{i,t}$.

Assumption 2 (Regression adjustment estimators). There is a matrix B of dimension $K \times K$ such that

$$\mathbb{E} [\hat{\mathbf{L}} | \mathbf{L}] = \Gamma(\mathbb{I} + B)\Lambda^\top.$$

where, Γ, Λ are defined in Assumption 1 and \mathbb{I} is the identity matrix of dimension $K \times K$.

Assumption 2 introduces a broad class of regression adjustments. The matrix B summarizes the bias of $\hat{\mathbf{L}}$ relative to \mathbf{L} : whenever B is the zero-matrix, $\hat{\mathbf{L}}$ is an unbiased estimator, whereas B different from the zero matrix captures shrinkage and rank truncation (e.g., if $\text{rank}(\mathbf{L}) = K' > K$, a rank- K procedure typically yields a non-zero bias B). Assumption 4 is satisfied, under isotropic additive noise, by several methods that include PCA/truncated SVD, nuclear-norm penalized least squares, and general singular-value–shrinkage estimators whose predictions rotate equivariantly with the singular vectors of \mathbf{L} [CCS10, CSLT13, GD17, NW11, Nad14].⁴ In this formulation, B characterizes

⁴Formally, let $Y = L + \varepsilon$ with $\text{rank}(L) = K'$ and SVD $L = U\Sigma V^\top$. Noise isotropy implies that for all

the bias of the regression adjustment.⁵

THEOREM 5.1 (Triple robustness). *Let Assumption 1 hold. Then for fixed (not data dependent) weights θ, ω (and conditional on T_0, N_0),*

$$\left| \mathbb{E}[\hat{\tau} - \tau | \mathbf{L}] \right| \leq \underbrace{\|\Delta^u(\omega, \Gamma)\|_2}_{(i): \text{Units imbalance}} \times \underbrace{\|\Delta^t(\theta, \Lambda)\|_2}_{(ii): \text{Time imbalance}} \times \underbrace{\|B\|_*}_{(iii): \text{misspecification of the regression adjustment}} ,$$

where $\|\cdot\|_2$ denotes the l_2 -norm and $\|\cdot\|_*$ denotes the spectral norm and \mathbb{I}_K denotes the identity matrix and B as defined in Assumption 2.

PROOF. See Appendix 8. □

Theorem 5.1 shows that we can decompose the bias of the estimator into *the product* of three main components: (i) unit imbalance, (ii) time imbalance and (iii) bias of the estimator \mathbf{L} , expressed through nuclear norm operator in the final expression. The first two components coincide with the time and unit imbalance components studied, for instance, in [IV23]. The last component captures the misspecification error from estimating the factors and loadings when, e.g., imposing a low-rank regularization. As discussed above, here we consider non-data dependent weights, interpreting θ, ω as probability limits of the estimated weights [see e.g. Hir21].⁶

The following corollary formalizes the triple robustness property.

COROLLARY 1. *Let Assumption 1 hold. Suppose that any one of the following conditions hold:*

- (a) $\sum_{i \leq N_0} \Gamma_i \omega_i = \Gamma_N$ (*balance over unit loadings*); or
- (b) $\sum_{s \leq T_0} \theta_s \Lambda_s = \Lambda_T$ (*balance over factor loadings*); or
- (c) $B = \mathbf{0}_K$ where $\mathbf{0}_K$ is a matrix of dimension $K \times K$ with all zero entries (*correct regression adjustment specification*).

Then $\mathbb{E}[\hat{\tau} - \tau | \mathbf{L}] = 0$.

orthogonal $Q_1, Q_2, Q_1 \varepsilon Q_2^\top \mid L =_d \varepsilon \mid L$; and (ii) estimator equivariance implies that for all orthogonal Q_1, Q_2 , the estimator $\hat{\mathbf{L}}$ can be written as a function $F(\cdot)$ of the data such that $F(Q_1 Y Q_2^\top) = Q_1 F(Y) Q_2^\top$.

⁵Also, note that it is possible to relax Assumption 2 and assume that the expectation equals $\Gamma(\mathbb{I} + B)\Lambda^\top$ up to a small vanishing term, in which case our robustness properties should be interpreted asymptotically as the additional term converges to zero. We omit this for expositional convenience only.

⁶Note that because τ is also random as it depends on potential outcomes, the expectation is both over $\hat{\tau}$ and τ .

Finally, note that it is possible to consider general estimators instead of imposing restrictions as in Assumption 2. This is formalized in Theorem A1 where we characterize the estimation error for general estimators of $\hat{\mathbf{L}}$ (at the expense of more elaborate notation). With general estimators, the estimation error depends on generic properties of the estimator. We provide an example in Section 6 and Appendix 8.4, where in addition to the low-rank factor models, we also consider additive covariates multiplying linear coefficients (without imposing restrictions on the particular choice of the linear estimator). We leave to future research a comprehensive analysis of arbitrary non-linear estimators.

5.2. Comparison with existing estimators

Theorem 5.1 allows us to directly compare our proposed estimator to both Synthetic Control and Synthetic DID. To gain further insight, we illustrate the source of the bias for the DID, Synthetic Control and Synthetic DID estimator under Assumption 1 alone. These estimators (for given weights) take the form

$$\hat{\tau}^{\text{SC}} = Y_{NT} - \sum_{i \leq N_0} \omega_i Y_{iT}, \quad \hat{\tau}^{\text{sDID}} = Y_{NT} - \left(\sum_{t \leq T_0} \theta_t Y_{Nt} + \sum_{i \leq N_0} \omega_i Y_{iT} - \sum_{t \leq T_0} \sum_{i \leq N_0} \omega_i \theta_t Y_{it} \right).$$

We interpret the Synthetic Control as a balancing estimator that imposes balance over units, and the Synthetic DID as a balancing estimator across units and time.

As noted in [IV23], with θ, ω not data dependent for simplicity, we can write

$$(11) \quad \begin{aligned} B^{\text{DID}} &:= \mathbb{E}[\hat{\tau}^{\text{DID}} - \tau | \mathbf{L}] &= (\Gamma_N - \bar{\Gamma}^0)^\top (\Lambda_T - \bar{\Lambda}^0), \\ B^{\text{SC}} &:= \mathbb{E}[\hat{\tau}^{\text{SC}} - \tau | \mathbf{L}] &= (\bar{\Gamma}_0(\omega) - \Gamma_N)^\top \Lambda_T, \\ B^{\text{sDID}} &:= \mathbb{E}[\hat{\tau}^{\text{sDID}} - \tau | \mathbf{L}] &= (\bar{\Gamma}_0(\omega) - \Gamma_N)^\top (\bar{\Lambda}_0(\theta) - \Lambda_T). \end{aligned}$$

That is, the bias of the DID estimator depends on the imbalance between the average loadings on the treated and control group, times the imbalance between the factor in the treatment and control period. The bias of the Synthetic control estimator depends on imbalances between the loadings of the treated and control, *reweighted* by unit-specific weights (which typically are estimated to minimize such imbalance). The bias of Synthetic DID is doubly robust to imbalance either across time periods or across treated and control units. In particular, both the Synthetic Control and Synthetic DID leverage possibly non-uniform weights to decrease the imbalance between the treated

and control units. The Difference-in-Differences estimator instead suffer a bias due to using uniform weights over the control units and control periods.

Different from all such estimator, our proposed procedure inherits robustness not only from the choice of the weights, but also through the additional regression adjustment. The combination of these three components guarantee a triple robustness properties to the TROP not encountered in previous estimators.

5.3. Inference: discussion

In the framework with a single treated unit, with idiosyncratic shock ε_{NT} , the estimation error of the weights is typically negligible (i.e., $N_0 \gg N_1, T_0 \gg T_1$), and as a result inference is driven by the variance of Y_{NT} . In the presence of multiple post-treatment periods, as we define Y_{NT} as the average outcome of the treated unit over such periods, we may assume that (keeping T_0 and N_0 as deterministic)

$$\varepsilon_{NT}(1) := Y_{NT}(1) - \mathbb{E}[Y_{NT}(1)|\mathbf{L}] \sim \mathcal{N}(0, \sigma_{NT}^2)$$

where the normal approximation holds from a central limit theorem argument assuming that the last period denotes the average over multiple post treatment periods. In particular, define

$$(12) \quad \bar{\varepsilon}^0(\theta, \omega) = \sum_{i \leq N_0, t > T_0} \omega_i \varepsilon_{it} + \sum_{t \leq T_0, i > N_0} \theta_t \varepsilon_{it} - \sum_{i \leq N_0, t \leq T_0} \omega_i \theta_t \varepsilon_{it}.$$

It follows that

$$\hat{\tau} - \mathbb{E}[\tau|\mathbf{L}] = \bar{\Delta}(\omega, \theta) + \varepsilon_{NT}(1) - \bar{\varepsilon}^0(\theta, \omega),$$

where $\bar{\Delta}(\omega, \theta) = \mathbb{E}[\hat{\tau} - \tau|\mathbf{L}]$. Typically, we may expect $\bar{\Delta}(\omega, \theta) \approx 0$ by the triple robustness property of our estimator. In addition, $\varepsilon^0(\omega, \theta) = O_p(\min\{\|\theta\|_2, \|\omega\|_2\})$, so that whenever $\|\theta\|_2, \|\omega\|_2 = o(\sigma_{NT})$, the variance contribution of $\varepsilon^0(\omega, \theta)$ is asymptotically negligible relative to the one from ε_{NT} . This occurs if the weights are sufficiently dispersed over the pre-treatment periods and units. Under such conditions

$$\hat{\tau} - \mathbb{E}[\tau|\mathbf{L}] = \varepsilon_{NT}(1) + o_p(\sigma_{NT}^2),$$

and therefore the variance of $\hat{\tau}$ only depends (asymptotically) on the variance of $\varepsilon_{NT}(1)$. Here, $\mathbb{E}[\tau|\mathbf{L}]$ defines the expected treatment effect on the treated unit, conditional on the latent factor \mathbf{L} and on T_0, N_0 .

Existing literature has proposed multiple procedures to estimate such variance components. We provide a formal algorithm to conduct inference with multiple treated units in Algorithm 3 below.

6. Extensions

6.1. Estimation with arbitrary treated units and periods

Next, we generalize this idea to multiple treated units and periods. As a first step, that will allow us to generalize the estimator to multiple treated units and periods, define

$$I_{j,s}^{i,t} = 1\{(i, t) = (j, s)\}$$

the indicator of whether observation (i, t) is the observation (j, s) . We will consider unit specific and time specific weights $\omega(\lambda), \theta(\lambda)$ parametrized by an arbitrary vector of parameters λ . Such parametrization will simplify the choice of the weights as we discuss further below.

For each unit (i, t) we construct an estimator

$$(13) \quad \hat{\tau}_{it}(\lambda) =$$

$$\arg \min_{\tau_{i,t}} \min_{\alpha, \beta, \mathbf{L}} \sum_{j=1}^N \sum_{s=1}^T \left[(1 - W_{j,s}) + W_{j,s} I_{j,s}^{i,t} \right] \omega_j^{i,t}(\lambda) \theta_s^{i,t}(\lambda) \left(Y_{js} - \alpha_j - \beta_s - L_{js} - \tau_{i,t} \cdot I_{j,s}^{i,t} \right)^2 + \lambda_{nn} \|\mathbf{L}\|$$

This corresponds to the estimated treatment effects on a given unit (i, t) , as if unit (i, t) was the only treated units, after removing the outcomes for the other treated observations over the treatment periods. We can write the final average treatment effect on the treated as

$$\hat{\tau} = \frac{1}{\sum_{i=1}^N \sum_{t=1}^T W_{i,t}} \sum_{i=1}^N \sum_{t=1}^T W_{i,t} \hat{\tau}_{i,t}(\lambda).$$

Such estimator generalizes the estimator we propose in the presence of a single treated unit and period. We summarize the estimation procedure in Algorithm 2.

To conduct inference, we propose a simple non-parametric bootstrap estimator, that is formally discussed in Algorithm 3. The bootstrap does not impose assumptions on time independence of idiosyncratic shocks within each unit, but its validity requires a growing number of treated units.

Algorithm 2 Triple robust panel estimator for treatment effect with multiple treated units

Require: Grid of values \mathcal{G} for $(\lambda_{time}, \lambda_{unit}, \lambda_{nn})$, treatments \mathbf{W} , outcomes \mathbf{Y}

1: **for each** $\Lambda \in \mathcal{G}$ **do**

- 2: For each (i, t) such that $W_{it} = 0$ estimate $\hat{\tau}_{it}(\Lambda)$ as in Equation (13)
- 3: Compute the function $Q(\Lambda)$ in Equation (5)

$$Q(\Lambda) = \sum_{i=1}^N \sum_{t=1}^T (1 - W_{it}) (\hat{\tau}_{it}(\Lambda))^2$$

4: **end for**

5: Find $\hat{\lambda} \in \arg \min_{\lambda \in \mathcal{G}} Q(\lambda)$

6: Compute the estimator

$$\hat{\tau} = \frac{1}{\sum_{i=1}^N \sum_{t=1}^T W_{i,t}} \sum_{i=1}^N \sum_{t=1}^T W_{i,t} \hat{\tau}_{i,t}(\hat{\lambda}).$$

where $\hat{\tau}(\lambda)$ are defined as in Equation (13).

Algorithm 3 Bootstrap Variance Estimation with multiple treated units

1: **Data:** Y, W, B

2: **Result:** Variance estimator \hat{V}_τ^{cb}

3: **for** $b \leftarrow 1$ to B **do**

- 4: Construct a bootstrap dataset $(Y^{(b)}, W^{(b)})$ by sampling N_0 rows of (Y, W) with replacement for the first N_0 units and sampling N_1 rows with replacement of (Y, W) from the last N_1 units.

5: Compute the TROP estimator $\hat{\tau}^{(b)}$ from $(Y^{(b)}, W^{(b)})$.

6: **end for**

7: Define

$$\hat{V}_\tau = \frac{1}{B} \sum_{b=1}^B (\hat{\tau}^{(b)} - \frac{1}{B} \sum_{b=1}^B \hat{\tau}^{(b)})^2.$$

Remark 6.1 (Alternative approaches to estimate the effect on multiple treated units). Note that here we estimate each individual's treatment effect $\hat{\tau}_{i,t}$ separately for each unit, hence allowing for heterogeneous effects across units and time periods. It is possible however to impose some additional modeling assumption to gain in terms of precision. For example, under homogeneous treatment effects, one could aggregate estimators for different units using different weighing mechanisms. We omit further details and leave a study of efficient estimation of treatment effects to future research. \square

6.2. The TROP Estimator with Covariates

In the presence of covariates, researchers may consider multiple extensions of the TROP estimator. One simple extension we propose is to include covariates additively into the model. That is, we can parametrize

$$L_{j,s} = X_{j,s}\beta + R_{j,s}$$

where $X_{j,s}$ defines observable covariates and $R_{j,s}$ denotes a low-rank component. The main advantage of controlling for covariates in this case is to relax the low rank assumption of $L_{j,s}$, assuming instead that $L_{j,s}$ is given by the sum of observed covariates and of a possibly low-rank component $R_{j,s}$. In this case the objective function for a treated unit/period (i^*, t^*) reads as

$$(14) \quad \hat{\tau}_{i^*,t^*} = \arg \min_{\tau} \min_{\alpha, \beta, \tilde{\beta}, \mathbf{L}} \sum_{j=1}^N \sum_{s=1}^T \theta_s^{i^*, t^*} \omega_j^{i^*, t^*} \left(Y_{js} - \alpha_j - \beta_s - X_{j,s}\beta - R_{j,s} - \tau W_{j,s} \right)^2 + \lambda_{nn} \|\mathbf{R}\|,$$

We can then follow verbatim our procedure also controlling for covariates in this case. Appendix Theorem 8.1 extends our triple-robustness results in the presence of covariates without assuming a particular estimator for β .

7. Conclusion

This paper develops the Triply ROBust Panel (TROP) estimator—a unified framework that combines data-driven unit weights, time-decay weights, and a regression adjustment to estimate counterfactuals in panel-data settings with potentially complex assignment patterns. The estimator inherits a triple robustness property, as its bias depends on the product of unit imbalance, time imbalance, and misspecification of the regression adjustment. We propose a leave-one-out cross-validation strategy that simultaneously tunes the unit and time weights, as well as the strength of the nuclear-norm penalty for estimating a low-rank factor model. Extensive simulations calibrated to canonical CPS and PWT-World-Table, as well as other applications, illustrate the benefits of the methods.

These results highlight two broader lessons for empirical work with panels. First, the recent past typically matters most, and estimators should adaptively adjust to learning

the most relevant periods without considering observations over time as exchangeable. Second, standard DID methods may approximate poorly relevant economic applications, and the combination of balancing and interactive fixed effects can provide more accurate rate counterfactuals.

Several open questions remain. First, alternative distance metrics (e.g., covariate-based Mahalanobis distances) could further refine unit weighting. Second, integrating TROP with augmented inverse-propensity weighting or proxy-variable strategies promises additional gains when assignment probabilities are partially observed. Finally, we leave the study of such estimators in contexts with dynamic treatment effects to future research.

References

- [AAH⁺19] Dmitry Arkhangelsky, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. Synthetic difference in differences. Technical report, National Bureau of Economic Research, 2019.
- [ABD⁺21a] 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.
- [ABD⁺21b] 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*, 2021.
- [ADH10] 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.
- [ADH15a] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, pages 495–510, 2015.
- [ADH15b] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, 2015.
- [AG03] Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(-):113–132, 2003.
- [AIW18] Susan Athey, Guido W Imbens, and Stefan Wager. Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):597–623, 2018.
- [AS18] Sarah Abraham and Liyang Sun. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. 2018.
- [Bai09] Jushan Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009.

- [BCC⁺25] Andrew Baker, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro HC Sant’Anna. Difference-in-differences designs: A practitioner’s guide. *arXiv preprint arXiv:2503.13323*, 2025.
- [BGK⁺15] Kay H Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, Steven L Scott, et al. Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1):247–274, 2015.
- [BMAF⁺23] Eli Ben-Michael, David Arbour, Avi Feller, Alexander Franks, and Steven Raphael. Estimating the effects of a california gun control program with multitask gaussian processes. *The Annals of Applied Statistics*, 17(2):985–1016, 2023.
- [BMFR18] Eli Ben-Michael, Avi Feller, and Jesse Rothstein. The augmented synthetic control method. *arXiv preprint arXiv:1811.04170*, 2018.
- [BN13] Jushan Bai and Serena Ng. Principal components estimation and identification of static factors. *Journal of econometrics*, 176(1):18–29, 2013.
- [BR05] Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.
- [CCD⁺18] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.
- [CCS10] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.
- [CFT21] Matias D Cattaneo, Yingjie Feng, and Rocio Titiunik. Prediction intervals for synthetic control methods. *Journal of the American Statistical Association*, 116(536):1865–1880, 2021.
- [CS20] Brantly Callaway and Pedro HC Sant’Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 2020.
- [CSLT13] Emmanuel J. Candès, Carlos A. Sing-Long, and Joshua D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE Transactions on Signal Processing*, 61(19):4643–4657, 2013.
- [CSX25] Xiaohong Chen, Pedro HC Sant’Anna, and Haitian Xie. Efficient difference-in-differences and event study estimators. *arXiv preprint arXiv:2506.17729*, 2025.
- [CWZ18] Victor Chernozhukov, Kaspar Wuthrich, and Yinchu Zhu. A t -test for synthetic controls. *arXiv preprint arXiv:1812.10820*, 2018.
- [dCd19] Clément de Chaisemartin and Xavier d’Haultfœuille. Two-way fixed effects estimators with heterogeneous treatment effects. Technical report, National Bureau of Economic Research, 2019.
- [DCd23] Clément De Chaisemartin and Xavier d’Haultfoeuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The econometrics journal*, 26(3):C1–C30, 2023.
- [FP16] Bruno Ferman and Cristine Pinto. Revisiting the synthetic control estimator. 2016.
- [GB21] Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of econometrics*, 225(2):254–277, 2021.

- [GD17] Matan Gavish and David L. Donoho. Optimal shrinkage of singular values. *IEEE Transactions on Information Theory*, 63(4):2137–2152, 2017.
- [Hir21] David A Hirshberg. Least squares with error in variables. *arXiv preprint arXiv:2104.08931*, 2021.
- [HS17] Jinyong Hahn and Ruoyao Shi. Synthetic control and inference. *Econometrics*, 5(4):52, 2017.
- [HX18] Chad Hazlett and Yiqing Xu. Trajectory balancing: A general reweighting approach to causal inference with time-series cross-sectional data. 2018.
- [IV23] Guido W Imbens and Davide Viviano. Identification and inference for synthetic controls with confounding. *arXiv preprint arXiv:2312.00955*, 2023.
- [KMPT21] Maxwell Kellogg, Magne Mogstad, Guillaume A Pouliot, and Alexander Torgovitsky. Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *Journal of the American statistical association*, 116(536):1804–1816, 2021.
- [MW15] Hyungsik Roger Moon and Martin Weidner. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579, 2015.
- [MW17] Hyungsik Roger Moon and Martin Weidner. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195, 2017.
- [Nad14] Raj Rao Nadakuditi. Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage. *IEEE Transactions on Information Theory*, 60(5):3002–3018, 2014.
- [NW11] Sahand Negahban and Martin J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *The Annals of Statistics*, 39(2):1069–1097, 2011.
- [PY15] Giovanni Peri and Vasil Yasenov. The labor market effects of a refugee wave: Applying the synthetic control method to the mariel boatlift. Technical report, National Bureau of Economic Research, 2015.
- [RSBP23] Jonathan Roth, Pedro HC Sant’Anna, Alyssa Bilinski, and John Poe. What’s trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244, 2023.
- [SA20] Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 2020.
- [Sch23] Timo Schenk. Time-weighted difference-in-differences: Accounting for common factors in short t panels. Technical report, Tinbergen Institute Discussion Paper, 2023.
- [SSP⁺19] Pantelis Samartsidis, Shaun R Seaman, Anne M Presanis, Matthew Hickman, and Daniela De Angelis. Assessing the causal effect of binary interventions from observational panel data with few treated units. *Statistical Science*, 34(3):486–503, 2019.
- [SZ20] Pedro HC Sant’Anna and Jun Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020.
- [Zub15] Jose R. Zubizarreta. Stable weights that balance covariates for estimation with

incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922, 2015.

Appendix 8. Proofs

8.1. General notation

Recall the general setup: $Y_{it}(0) = L_{it} + \varepsilon_{it}$ with $\mathbb{E}[\varepsilon_{it} | L] = 0$ for all (i, t) . There is a single treated unit/period (N, T) ; all other cells are pre-treatment. Let $\omega \in \mathbb{R}^{N_0}$ and $\theta \in \mathbb{R}^{T_0}$ be fixed weights that each sum to one: $\sum_{i \leq N_0} \omega_i = \sum_{t \leq T_0} \theta_t = 1$, where $T_0 = T - 1$, $N_0 = N - 1$. (Note that our results extend beyond one treated unit/period, at the expense of more tedious notation).

We introduce the following notation:

$$(U_\omega)_i := \begin{cases} \omega_i, & i \leq N_0 \\ 0, & i > N_0, \end{cases} \quad (V_\theta)_t := \begin{cases} \theta_t, & t \leq T_0 \\ 0, & t > T_0, \end{cases}$$

and the canonical basis vectors $e_N \in \mathbb{R}^N$, $e_T \in \mathbb{R}^T$ so that e_N has the N^{th} entry equal to one and all other entries equal to zero and e_T has the T^{th} entry equal to one and all other entries equal to zero. Define

$$M(\theta, \omega) := (e_N - U_\omega)(V_\theta - e_T)^\top \in \mathbb{R}^{N \times T}.$$

We use the Frobenius inner product $\langle A, B \rangle = \sum_{i,t} A_{it}B_{it}$.

Let $\widehat{\mathbf{L}} = \{\widehat{\mathbf{L}}_{it}\}$ be any (possibly data-dependent) adjustment. Define the TROP counterfactual for (N, T) by

(A1)

$$\widehat{\tau}_{\text{TROP}}(0; \theta, \omega) := \widehat{\mathbf{L}}_{NT} + \sum_{t \leq T_0} \theta_t (Y_{Nt} - \widehat{\mathbf{L}}_{Nt}) + \sum_{i \leq N_0} \omega_i (Y_{iT} - \widehat{\mathbf{L}}_{iT}) - \sum_{i \leq N_0} \sum_{t \leq T_0} \omega_i \theta_t (Y_{it} - \widehat{\mathbf{L}}_{it}).$$

8.2. Main theorem

The following lemma provides a useful representation of the TROP estimator.

LEMMA A1. *With $M(\theta, \omega)$ as above,*

$$(A2) \quad \widehat{\tau}_{\text{TROP}}(0; \theta, \omega) = Y_{NT} + \left\langle Y - \widehat{\mathbf{L}}, M(\theta, \omega) \right\rangle.$$

Proof of Lemma A1. Expand the right-hand side of (A2). Since $M_{Nt} = \theta_t$ for $t \leq T_0$,

$M_{iT} = \omega_i$ for $i \leq N_0$, $M_{it} = -\omega_i \theta_t$ for $i \leq N_0$, $t \leq T_0$, and $M_{NT} = -1$, we obtain

$$\langle Y - \widehat{\mathbf{L}}, M \rangle = \sum_{t \leq T_0} \theta_t (Y_{Nt} - \widehat{\mathbf{L}}_{Nt}) + \sum_{i \leq N_0} \omega_i (Y_{iT} - \widehat{\mathbf{L}}_{iT}) - \sum_{i \leq N_0} \sum_{t \leq T_0} \omega_i \theta_t (Y_{it} - \widehat{\mathbf{L}}_{it}) - (Y_{NT} - \widehat{\mathbf{L}}_{NT}).$$

Adding Y_{NT} cancels the last term and leaves exactly (A1). \square

THEOREM A1. *Under Assumption 1 and for fixed (θ, ω) ,*

$$(A3) \quad \widehat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) = \langle \mathbf{L} - \widehat{\mathbf{L}}, M(\theta, \omega) \rangle + \bar{\varepsilon}_0(\theta, \omega) - \varepsilon_{NT},$$

with $\bar{\varepsilon}_0(\theta, \omega)$ defined in Equation (12). In particular, taking $\mathbb{E}[\cdot | \mathbf{L}]$,

$$(A4) \quad \mathbb{E}[\widehat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) | \mathbf{L}] = \langle \mathbf{L} - \mathbb{E}[\widehat{\mathbf{L}} | \mathbf{L}], M(\theta, \omega) \rangle.$$

Proof of Theorem A1. For short, we will refer to $M(\theta, \omega)$ as M . From Lemma A1 and Assumption 1, by letting Y denote the matrix with $(it)^{th}$ entry Y_{it} ,

$$(A5) \quad \widehat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) = \langle Y - \widehat{\mathbf{L}}, M \rangle + (Y_{NT} - \mathbf{L}_{NT}) - \varepsilon_{NT},$$

where we decompose $Y_{NT}(0) = \mathbf{L}_{NT} + \varepsilon_{NT}$. In addition, we can write

$$(A6) \quad \langle Y - \widehat{\mathbf{L}}, M \rangle = \langle \mathbf{L} - \widehat{\mathbf{L}}, M \rangle + \langle \varepsilon, M \rangle - (Y_{NT} - \mathbf{L}_{NT} - \varepsilon_{NT})$$

because $M_{NT} = -1$, the contribution of (N, T) inside $\langle \mathbf{L} + \varepsilon, M \rangle$ equals $-(\mathbf{L}_{NT} + \varepsilon_{NT})$ and the contribution of (N, T) inside $\langle Y, M \rangle$ equals $-Y_{NT}$.

Combining Equations (A5) and (A6), it follows that

$$\widehat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) = \langle \mathbf{L} - \widehat{\mathbf{L}}, M \rangle + \langle \varepsilon, M \rangle$$

which equals the desired result in Equation (A3) by definition of M .

Equation (A4) follows immediately after taking expectations, since $\mathbb{E}[\varepsilon_{it} | \mathbf{L}] = 0$ for all (i, t) and θ, ω are fixed (not data dependent). \square

8.3. Proof of Theorem 5.1

COROLLARY A1. *Let Assumptions 1 and 2 hold. Then*

$$(A7) \quad \begin{aligned} \mathbb{E}[\hat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) \mid \mathbf{L}] &= \Delta_u(\omega, \Gamma)^\top B \Delta_t(\theta, \Lambda) \\ &\leq \|\Delta_u(\omega, \Gamma)\|_2 \|\Delta_t(\theta, \Lambda)\|_2 \|B\|_\star. \end{aligned}$$

PROOF. Theorem A1 gives the identity

$$(A8) \quad \mathbb{E}[\hat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) \mid \mathbf{L}] = \left\langle \mathbf{L} - \mathbb{E}[\hat{\mathbf{L}} \mid \mathbf{L}], M(\theta, \omega) \right\rangle,$$

where $M(\theta, \omega) = (e_N - U_\omega)(V_\theta - e_T)^\top$.

Under Assumptions 1 and 2, $\mathbf{L} - \mathbb{E}[\hat{\mathbf{L}} \mid \mathbf{L}] = -\Gamma B \Lambda^\top$. Let $a := e_N - U_\omega$ and $b := V_\theta - e_T$. By Lemma A2,

$$(A9) \quad \left\langle \Gamma B \Lambda^\top, ab^\top \right\rangle = a^\top \Gamma B \Lambda^\top b = (\Gamma^\top a)^\top B (\Lambda^\top b).$$

Finally, note that by definition of $a = e_N - U_\omega$,

$$\Gamma^\top a = \Gamma^\top e_N - \Gamma^\top U_\omega = \Gamma_N - \sum_{i \leq N_0} \omega_i \Gamma_i = -(\bar{\Gamma}_0(\omega) - \Gamma_N) = -\Delta_u(\omega, \Gamma).$$

Similarly, with $b = V_\theta - e_T$,

$$\Lambda^\top b = \Lambda^\top V_\theta - \Lambda^\top e_T = \sum_{t \leq T_0} \theta_t \Lambda_t - \Lambda_T = \bar{\Lambda}_0(\theta) - \Lambda_T = \Delta_t(\theta, \Lambda).$$

Substituting these two identities into (A9) yields

$$-\left\langle \Gamma B \Lambda^\top, ab^\top \right\rangle = \Delta_u(\omega, \Gamma)^\top B \Delta_t(\theta, \Lambda),$$

proving Equation (A7).

Finally, the last inequality in Equation (A7) follows directly from Lemma A3. \square

Theorem 5.1 follows directly from Corollary A1 since $\mathbb{E}[\hat{\tau} - \tau \mid \mathbf{L}] = \mathbb{E}[\hat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) \mid \mathbf{L}]$.

8.4. Triple robustness with covariates

In the following theorem we generalize the triple robustness properties in the presence of covariates, allowing general (possibly penalized) regression methods in the presence of covariates.

Assumption 3 (Factor model with covariates). Suppose that for all (i, t) , $Y_{i,t}(0) = \mathbf{R}_{i,t} + X_{i,t}\beta + \varepsilon_{i,t}$, with $\mathbb{E}[\varepsilon_{i,t} | \mathbf{R}, \mathbf{X}] = 0$ and $\mathbf{R}_{i,t} = \Gamma\Lambda^\top$.

Assumption 3 generalizes Assumption 1 introducing additional observed covariates $X_{i,t}$.

Define the covariate contrast vector

$$(A10) \quad g(\theta, \omega) := \sum_{i,t} M(\theta, \omega)_{it} \mathbf{X}_{it} = (e_N - U_\omega)^\top \mathbf{X} (V_\theta - e_T) \in \mathbb{R}^p,$$

and the estimator's bias

$$\delta_\beta := \mathbb{E}[\hat{\beta} | \mathbf{X}, \mathbf{R}] - \beta \in \mathbb{R}^p.$$

Assumption 4 (Estimators with covariates). There exists a matrix $B \in \mathbb{R}^{K \times K}$ such that

$$\mathbb{E}[\hat{\mathbf{L}} - \mathbf{L} | \mathbf{X}, \mathbf{R}] = \mathbf{X}\delta_\beta + \Gamma B \Lambda^\top + E_\perp,$$

for an arbitrary matrix $E_\perp \in \mathbb{R}^{T \times N}$, with $\|E\|_{\max} = o(\eta)$ for some small $\eta \geq 0$.

Assumption 4 mimics Assumption 2 introducing additional covariates. Note that we impose no restrictions on the estimation of $\hat{\beta}$. Here, estimation of $\hat{\beta}$ interacts with properties of the estimator $\hat{\mathbf{R}}$. Therefore, in addition to the matrix B and bias component δ_β that capture misspecification of the low-rank regression adjustment and of the estimation of the coefficients, we also allow one additional term that may violate the invariance condition in Assumption 2. We think of E_\perp as a matrix with element-wise components of small order capturing any additional approximation error.

THEOREM 8.1 (Bias decomposition with covariates). *Let Assumptions 3 and 4 hold. Suppose that $\omega, \theta \geq 0$. Then*

$$(A11) \quad \mathbb{E}[\hat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) | \mathbf{X}, \mathbf{R}] = \Delta_u(\omega, \gamma)^\top B \Delta_t(\theta, \lambda) - g(\theta, \omega)^\top \delta_\beta + o(\eta).$$

Consequently,

$$\begin{aligned} \left| \mathbb{E}[\widehat{\tau}_{\text{TROP}}(0; \theta, \omega) - Y_{NT}(0) \mid \mathbf{X}, \mathbf{R}] \right| &\leq \|\Delta_u(\omega, \gamma)\|_2 \|B\|_* \|\Delta_t(\theta, \lambda)\|_2 \\ (A12) \quad &\quad + \|g(\theta, \omega)\|_2 \|\delta_\beta\|_2 + o(\eta). \end{aligned}$$

PROOF. We apply Theorem A1 with $\mathbf{L} := \mathbf{X}\beta + \mathbf{R}$ and adjustment $\widehat{\mathbf{L}} = \widehat{\mathbf{X}\beta} + \widehat{\mathbf{R}}$:

$$\mathbb{E}[\widehat{\tau}_{\text{TROP}} - Y_{NT}(0) \mid \mathbf{X}, \mathbf{R}] = \langle \mathbf{L} - \mathbb{E}[\widehat{\mathbf{L}} \mid \mathbf{X}, \mathbf{R}], M(\theta, \omega) \rangle.$$

Taking inner products with $M(\theta, \omega)$ and using yields

$$\mathbb{E}[\widehat{\tau}_{\text{TROP}} - Y_{NT}(0) \mid \mathbf{X}, \mathbf{R}] = -\langle X\delta_\beta, M(\theta, \omega) \rangle - \langle \Gamma B \Lambda^\top, (e_N - U_\omega)(V_\theta - e_T)^\top \rangle - \langle E_\perp, M(\theta, \omega) \rangle.$$

For the covariate term, by Lemma A1 $\langle X\delta_\beta, ab^\top \rangle = a^\top(X\delta_\beta)b$ with $a = e_N - U_\omega$, $b = V_\theta - e_T$, we get

$$\langle X\delta_\beta, M(\theta, \omega) \rangle = (e_N - U_\omega)^\top X(V_\theta - e_T) \delta_\beta = g(\theta, \omega)^\top \delta_\beta.$$

For the residual term, use the same identity with $A = I_K - B$ to obtain

$$-\langle \Gamma B \Lambda^\top, ab^\top \rangle = -(\Gamma^\top a)^\top B (\Lambda^\top b) = \Delta_u(\omega, \gamma)^\top B \Delta_t(\theta, \lambda),$$

because $\Gamma^\top a = \Gamma_N - \bar{\Gamma}_0(\omega) = -\Delta_u$ and $\Lambda^\top b = \bar{\Lambda}_0(\theta) - \Lambda_T = \Delta_t$. Finally, it is easy to show that by Holder's inequality

$$|\langle E_\perp, M \rangle| \leq |e_N^\top E V_\theta| + |e_N^\top E e_T| + |U_\omega^\top E V_\theta| + |U_\omega^\top E e_T| \leq \|E\|_{\max}(1 + \|\omega\|_1)(1 + \|\theta\|_1).$$

Combining the displays gives (A11), since $\|\omega\|_1 = \|\theta\|_1 = 1$ by assumption in the theorem.

For the bound (A12), we apply Lemma A3:

$$|\Delta_u^\top B \Delta_t| = |\langle B, \Delta_u \Delta_t^\top \rangle| \leq \|B\|_* \|\Delta_u \Delta_t^\top\|_{\text{op}} = \|B\|_* \|\Delta_u\|_2 \|\Delta_t\|_2,$$

and Cauchy-Schwarz to the covariate term $|g^\top \delta_\beta| \leq \|g\|_2 \|\delta_\beta\|_2$. \square

8.5. Auxiliary lemmas

LEMMA A2. *For any $X \in \mathbb{R}^{N \times T}$ and vectors $a \in \mathbb{R}^N$, $b \in \mathbb{R}^T$. Then*

$$\langle X, ab^\top \rangle = a^\top X b.$$

PROOF. By definition of the Frobenius inner product,

$$\langle X, ab^\top \rangle = \text{tr}(X^\top ab^\top) = \text{tr}(b^\top X^\top a) = a^\top X b.$$

□

LEMMA A3. *For any $A \in \mathbb{R}^{K \times K}$ and vectors $x, y \in \mathbb{R}^K$,*

$$|x^\top A y| = |\langle A, xy^\top \rangle| \leq \|A\|_* \|xy^\top\|_{\text{op}} \leq \|A\|_* \|x\|_2 \|y\|_2.$$

PROOF. The equality $x^\top A y = \langle A, xy^\top \rangle$ follows from Lemma A2. Hölder's inequality for the dual pair $(\|\cdot\|_*, \|\cdot\|_{\text{op}})$, where $\|\cdot\|_{\text{op}}$ denotes the operator norm yields $|\langle A, Z \rangle| \leq \|A\|_* \|Z\|_{\text{op}}$ for all A, Z of matching size. Take $Z = xy^\top$ to get $|\langle A, xy^\top \rangle| \leq \|A\|_* \|xy^\top\|_{\text{op}}$. Finally, by definition,

$$\|xy^\top\|_{\text{op}} = \sup_{\|\nu\|_2=1} \|xy^\top \nu\|_2 = \sup_{\|\nu\|_2=1} \|x(y^\top \nu)\|_2 \leq \|x\|_2 \cdot \sup_{\|\nu\|_2=1} |y^\top \nu|.$$

By Cauchy–Schwarz, $\sup_{\|\nu\|_2=1} |y^\top \nu| = \|y\|_2$, with equality at $\nu = y/\|y\|_2$ (or any unit vector aligned with y). Therefore $\|xy^\top\|_{\text{op}} \leq \|x\|_2 \|y\|_2$. □

Appendix 9. Additional Simulation Results

Panel A: $T_{\text{post}} = 1$ (treated periods), $N_{\text{tr}} = 10$ (treated units)														
Dataset	Treatment	N	T	$\frac{\ F\ _F}{\sqrt{NT}}$	$\frac{\ M\ _F}{\sqrt{NT}}$	$\sqrt{\frac{\text{Tr}(\Sigma)}{T}}$	AR(2)	$(\lambda_{\text{unit}}, \lambda_{\text{time}}, \lambda_{nn})$	RMSE					
									TROP	SDID	SC	DID	MC	DIFP
CPS logwage	min wage	50	40	0.992	0.10	0.098	(0.010, -0.057)	(0, 0.3, 0.0178)	0.049	0.053	0.059	0.067	0.052	0.054
CPS urate	min wage	50	40	0.770	0.40	0.604	(0.039, 0.011)	(0, 0.2, 0.0031)	0.290	0.301	0.302	0.267	0.288	0.300
PWT	democracy	111	48	0.972	0.229	0.069	(0.913, -0.221)	(0.3, 3.4, 0.01)	0.017	0.020	0.024	0.167	0.020	0.023
Germany	random	17	44	0.994	0.108	0.032	(0.840, -0.200)	(0, 3.2, 0.01)	0.018	0.021	0.050	0.068	0.023	0.033
Basque	random	18	43	0.986	0.167	0.037	(1.228, -0.577)	(0.3, 0.2, 0.005)	0.020	0.023	0.047	0.129	0.027	0.041
Smoking	random	39	31	0.935	0.337	0.166	(0.506, -0.058)	(0.8, 0.3, 0.0316)	0.066	0.068	0.077	0.169	0.070	0.076
Boatlift	random	44	19	0.913	0.344	0.276	(0.031, -0.087)	(0.3, 0.4, 0.2014)	0.114	0.126	0.144	0.174	0.123	0.145

Panel B: $T_{\text{post}} = 10$ (treated periods), $N_{\text{tr}} = 1$ (treated unit)

Panel B: $T_{\text{post}} = 10$ (treated periods), $N_{\text{tr}} = 1$ (treated unit)														
Dataset	Treatment	N	T	$\frac{\ F\ _F}{\sqrt{NT}}$	$\frac{\ M\ _F}{\sqrt{NT}}$	$\sqrt{\frac{\text{Tr}(\Sigma)}{T}}$	AR(2)	$(\lambda_{\text{unit}}, \lambda_{\text{time}}, \lambda_{nn})$	RMSE					
									TROP	SDID	SC	DID	MC	DIFP
CPS logwage	min wage	50	40	0.992	0.10	0.098	(0.010, -0.057)	(0.6, 0.3, 0.009)	0.061	0.069	0.074	0.126	0.078	0.085
CPS urate	min wage	50	40	0.770	0.40	0.604	(0.039, 0.011)	(0.7, 0, 0.005)	0.370	0.391	0.400	0.517	0.372	0.382
PWT	democracy	111	48	0.972	0.229	0.069	(0.913, -0.221)	(0.3, 0.4, 0.005)	0.070	0.106	0.134	0.318	0.095	0.116
Germany	random	17	44	0.994	0.108	0.032	(0.840, -0.200)	(0.6, 0.2, 0.015)	0.054	0.080	0.103	0.148	0.073	0.095
Basque	random	18	43	0.986	0.167	0.037	(1.228, -0.577)	(0.6, 0.3, 0.01)	0.088	0.139	0.235	0.329	0.139	0.139
Smoking	random	39	31	0.935	0.337	0.166	(0.506, -0.058)	(0.4, 0.3, 0.0265)	0.230	0.283	0.370	0.507	0.279	0.315
Boatlift	random	44	19	0.913	0.344	0.276	(0.031, -0.087)	(0.1, 0.2, 0.255)	0.329	0.440	0.480	0.438	0.346	0.461

TABLE A1. Simulation design as in Table 2, varying the number of post-treatment periods and treated units. **Panel A** uses one treated period and ten treated units ($T_{\text{post}}=1$, $N_{\text{tr}}=10$). **Panel B** uses ten treated periods and one treated unit ($T_{\text{post}}=10$, $N_{\text{tr}}=1$).

Panel A: PWT data (Y: log-GDP)											
Treatment	$\frac{\ F\ _F}{\sqrt{NT}}$	$\frac{\ M\ _F}{\sqrt{NT}}$	$\sqrt{\frac{\text{Tr}(\Sigma)}{T}}$	AR(2)	$(\lambda_{\text{unit}}, \lambda_{\text{time}}, \lambda_{nn})$	bias					
						TROP	SDID	SC	DID	MC	DIFP
Democracy	0.972	0.229	0.069	(0.913, -0.221)	(0.3, 0.4, 0.006)	0.010	-0.011	0.001	0.176	0.035	-0.005
No AR	0.972	0.229	0.069	(0, 0)	(0.2, 0.3, 0.016)	0.011	0.010	0.001	0.181	0.041	-0.008
Education	0.972	0.229	0.069	(0.913, -0.221)	(0.75, 0.275, 0.026)	0.009	-0.003	0.028	0.168	0.040	-0.007
Random	0.972	0.229	0.069	(0.913, -0.221)	(0.4, 0.45, 0.003)	0.001	-0.002	0.0004	0.002	-0.004	0.0006

Panel B: CPS data (Y: log-wage)											
Treatment	$\frac{\ F\ _F}{\sqrt{NT}}$	$\frac{\ M\ _F}{\sqrt{NT}}$	$\sqrt{\frac{\text{Tr}(\Sigma)}{T}}$	AR(2)	$(\lambda_{\text{unit}}, \lambda_{\text{time}}, \lambda_{nn})$	bias					
						TROP	SDID	SC	DID	MC	DIFP
Minimum wage	0.992	0.10	0.098	(0.010, -0.057)	(0, 0.1, 0.9)	0.008	0.006	0.019	0.022	0.012	0.006
Gun Law	0.992	0.10	0.098	(0.010, -0.057)	(0, 0.35, 0.041)	-0.0002	0.008	-0.004	0.012	0.015	0.009
Abortion	0.992	0.10	0.098	(0.010, -0.057)	(0, 0.2, 0.281)	0.004	0.004	0.016	0.003	0.003	0.001
Random	0.992	0.10	0.098	(0.010, -0.057)	(0, 0.2, 0.21)	-0.0003	0.001	-0.001	-0.0007	0.001	-0.0006

Panel C: CPS data (Y: U-rate)											
Treatment	$\frac{\ F\ _F}{\sqrt{NT}}$	$\frac{\ M\ _F}{\sqrt{NT}}$	$\sqrt{\frac{\text{Tr}(\Sigma)}{T}}$	AR(2)	$(\lambda_{\text{unit}}, \lambda_{\text{time}}, \lambda_{nn})$	bias					
						TROP	SDID	SC	DID	MC	DIFP
Minimum wage	0.770	0.399	0.604	(0.039, 0.011)	(1.6, 0.35, 0.011)	0.113	0.156	0.165	0.356	0.187	0.159

TABLE A2. Simulation design as in Table 2 (top-panel) with ten treatment periods and ten treated units, collecting both the bias of each estimator as well as simulation for a wider range of treatments. For PWT data the outcome variable is log-GDP and for CPS data the outcome variable is log-wage in Panel B and unemployment rate in Panel C.

	CPS logwage		CPS urate		PWT		Germany		Basque		Smoking		Boatlift	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
TROP	0.025	0.014	0.203	0.104	0.023	0.011	0.025	0.0003	0.041	-0.002	0.085	0.002	0.115	-0.006
$(\lambda_{nn} = \infty)$	0.025	0.011	0.246	0.197	0.038	0.013	0.037	0.0005	0.076	-0.002	0.124	0.002	0.178	-0.004
$(\lambda_{unit} = 0)$	0.025	0.013	0.224	0.182	0.024	0.013	0.025	0.0003	0.041	-0.002	0.088	0.002	0.115	-0.006
$(\lambda_{time} = 0)$	0.032	0.012	0.226	0.184	0.044	0.034	0.032	0.0003	0.062	-0.003	0.102	0.004	0.121	-0.007
$(\lambda_{unit} = \lambda_{time} = 0)$	0.032	0.012	0.223	0.181	0.044	0.034	0.032	0.0002	0.062	-0.003	0.102	0.004	0.121	-0.007
$(\lambda_{unit} = 0, \lambda_{nn} = \infty)$	0.025	0.008	0.248	0.196	0.039	0.016	0.037	0.0006	0.076	-0.002	0.124	0.002	0.178	-0.004
$(\lambda_{time} = 0, \lambda_{nn} = \infty)$	0.048	0.030	0.362	0.331	0.123	0.137	0.071	-0.001	0.145	-0.006	0.192	0.009	0.156	-0.07
DID $(\lambda_{unit} = \lambda_{time} = 0, \lambda_{nn} = \infty)$	0.049	0.022	0.384	0.354	0.198	0.175	0.073	-0.001	0.145	-0.006	0.192	0.009	0.156	-0.007

TABLE A3. Simulation designs as in Table 1, where we report the TROP estimator (first row) and the TROP estimator constraining one or combinations of λ_{time} , λ_{unit} to be zero (corresponding to no time or unit weights) and λ_{nn} to be infinity (corresponding to no regression adjustment). Each column reports the RMSE and bias across different designs in absolute terms. The number of treated units is 10 and the number of treated periods is 10. The treatment simulated as in Equation (8) is for CPS log-wage and CPS unemployment rate is minimum wage, for PWT is democracy, and for Germany, Basque, Smoking and Boatlift is randomly assigned.

	CPS logwage		CPS urate		PWT		Germany		Basque		Smoking		Boatlift	
	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias
TROP	0.111	0.020	0.637	-0.022	0.033	0.001	0.037	0.0003	0.035	0.016	0.178	0.003	0.350	-0.040
$(\lambda_{nn} = \infty)$	0.111	0.020	0.632	-0.027	0.033	0.001	0.037	0.0003	0.035	0.016	0.182	0.007	0.397	-0.037
$(\lambda_{unit} = 0)$	0.111	0.020	0.655	-0.099	0.034	0.002	0.037	0.0004	0.038	-0.019	0.192	-0.010	0.391	-0.048
$(\lambda_{time} = 0)$	0.112	0.010	0.632	-0.024	0.038	-0.0001	0.056	0.010	0.141	0.121	0.205	0.029	0.352	-0.122
$(\lambda_{unit} = \lambda_{time} = 0)$	0.115	0.008	0.696	0.023	0.038	0.0003	0.056	0.009	0.046	0.001	0.208	0.032	0.351	-0.128
$(\lambda_{unit} = 0, \lambda_{nn} = \infty)$	0.112	0.024	0.630	-0.055	0.034	0.002	0.037	0.0005	0.038	-0.019	0.193	-0.007	0.400	-0.046
$(\lambda_{time} = 0, \lambda_{nn} = \infty)$	0.137	0.017	0.632	-0.027	0.325	0.014	0.114	0.008	0.228	0.215	0.601	0.212	0.497	-0.338
DID $(\lambda_{unit} = \lambda_{time} = 0, \lambda_{nn} = \infty)$	0.156	0.033	0.630	-0.055	0.365	0.022	0.136	0.013	0.153	0.044	0.642	0.202	0.484	-0.320

TABLE A4. Simulation designs as in Table 1, where we report the TROP estimator (first row) and the TROP estimator constraining one or combinations of λ_{time} , λ_{unit} to be zero (corresponding to no time or unit weights) and λ_{nn} to be infinity (corresponding to no regression adjustment). Each column reports the RMSE and bias across different designs in absolute terms. The number of treated units is 1 and the number of treated periods is 1. The treatment is simulated as in Equation (8) is for CPS log-wage and CPS unemployment rate is minimum wage, for PWT is democracy, and for Germany, Basque, Smoking and Boatlift is randomly assigned.