Quantitative Economics 15 (2024), 999–1034 1759-7331/20240999

Design-robust two-way-fixed-effects regression for panel data

Dmitry Arkhangelsky
CEMFI and CEPR

Guido W. Imbens
Graduate School of Business and Department of Economics, Stanford University, SIEPR, and NBER

Lihua Lei
Graduate School of Business and Department of Statistics, Stanford University and SIEPR

Xiaoman Luo
Instacart

We propose a new estimator for average causal effects of a binary treatment with panel data in settings with general treatment patterns. Our approach augments the popular two-way-fixed-effects specification with unit-specific weights that arise from a model for the assignment mechanism. We show how to construct these weights in various settings, including the staggered adoption setting, where units opt into the treatment sequentially but permanently. The resulting estimator converges to an average (over units and time) treatment effect under the correct specification of the assignment model, even if the fixed- effect model is misspecified. We show that our estimator is more robust than the conventional two-way estimator: it remains consistent if either the assignment mechanism or the two-way regression model is correctly specified. In addition, the proposed estimator performs better than the two-way-fixed-effect estimator if the outcome model and assignment mechanism are locally misspecified. This strong robustness property underlines and quantifies the benefits of modeling the assignment process and motivates using our estimator in practice. We also discuss an extension of our estimator to handle dynamic treatment effects.

Keywords. Fixed effects, panel data, causal effects, treatment effects, double robustness, staggered adoption.

JEL classification. C14, C21, C22, C23, C90.

Dmitry Arkhangelsky: darkhangel@cemfi.es
Guido W. Imbens: imbens@stanford.edu
Lihua Lei: lihualei@stanford.edu
Xiaoman Luo: xmluo@ucdavis.edu

We are grateful for helpful comments and feedback from the two anonymous referees, as well as Susan Athey, Lanier Benkard, Peng Ding, Avi Feller, Dalia Ghanem, Jonathan Roth, Sophie Sun, Stefan Wager, Ye Wang, Ruonan Xu, and Yiqing Xu. Generous support from the Office of Naval Research through ONR Grants N00014-17-1-2131 and N00014-19-1-2468, and a gift from Amazon are gratefully acknowledged. Lihua Lei is grateful for the support of National Science Foundation Grant DMS-2338464.

 https://doi.org/10.3982/QE1962

1. Introduction

Difference-in-difference (DiD) methods (e.g., Ashenfelter and Card (1985), Angrist and Krueger (1999)) are commonly used in empirical economics to establish causal relationships (see Currie, Kleven, and Zwiers (2020) for some evidence regarding the usage in the empirical literature). In particular, researchers estimate regression functions of the form

$$Y_{it} = \mu + \alpha_i + \lambda_t + \beta^\top X_{it} + \tau W_{it} + \epsilon_{it} \tag{1.1}$$

using ordinary least squares (OLS), treating α_i and λ_t as fixed parameters—the fixed effects, leading to the two-way fixed-effect (TWFE) estimator. Here, Y_{it} is the outcome variable of interest, W_{it} is a binary treatment, X_{it} are observed exogenous characteristics, and τ is the main object of interest. Practitioners routinely justify regression (1.1) by appealing to "quasi-experimental" variation in treatment paths $\boldsymbol{W}_i = (W_{i1}, \ldots, W_{iT})$. Formal and informal arguments are invoked to make a case that this variation is not associated with unobserved unit and time-specific components ϵ_{it}. In other words, to motivate (1.1), researchers reason about the underlying model for $\boldsymbol{W}_i$. This model, however, does not explicitly enter the estimation process. Moreover, econometric assumptions that justify the OLS estimation apply conditionally on $\boldsymbol{W}_i$ and do not appeal to randomness in the treatment paths (e.g., Arellano (2003)).

In this paper, we develop new methods for estimating causal effects that explicitly incorporates design assumptions on the assignment process without abandoning the transparency and simplicity of the two-way model. We incorporate assumptions about the assignment mechanism by augmenting the specification (1.1) with unit-specific weights γ_i, leading to

$$\hat{\tau}(\gamma) = \underset{\tau,\mu,\alpha_i,\lambda_t,\beta}{\arg\min} \sum_{it} \big(Y_{it} - \mu - \alpha_i - \lambda_t - \beta^\top X_{it} - \tau W_{it}\big)^2 \gamma_i. \tag{1.2}$$

We compute the weights γ_i using the assignment model for $\boldsymbol{W}_i$.

We start our analysis by assuming that the assignment process for $\boldsymbol{W}_i$ is known. In Section 2, we show how to use this knowledge to construct oracle weights $\gamma^\star$ and conduct design-based inference. Under the correct specification of the assignment model, our inference procedure is valid regardless of the underlying model for potential outcomes, and in particular, we do not rely on the validity of the equation (1.1). Our results substantially generalize the properties established in Athey and Imbens (2022), allowing for an arbitrary assignment process (subject to overlap restrictions).

To construct $\gamma^\star$, we need to solve a nonlinear equation that depends on the support of $\boldsymbol{W}_i$. Practically, this means that the construction and the values of the weights vary across different types of assignment processes. In Supplemental Appendix C (Arkhangelsky, Imbens, Lei, and Luo (2024)), we provide solutions for several prominent examples, including staggered adoption, that is, a situation where units opt into treatment sequentially. Another input we need for $\gamma^\star$ is the probability distribution of $\boldsymbol{W}_i$ (generalized propensity score, Imbens (2000)).

After establishing design-based properties of the oracle estimator $\hat{\tau}(\gamma^\star)$ based on knowledge of the assignment process, we turn to the robustness—the behavior of the

estimator in settings where the postulated assignment model can be incorrect. At this point, we use the structure of the regression problem (1.2) to demonstrate that $\hat{\tau}(\gamma^\star)$ has a strong double-robustness property (Robins, Rotnitzky, and Zhao (1994), Kang and Schafer (2007), Bang and Robins (2005), Chernozhukov et al. (2018)): it has a small bias whenever either the assignment or the regression model is approximately correct. We view these results as the primary motivation for using our estimator in practice, where we cannot expect either the TWFE model or the assignment model to be fully correct.

In practice, the assignment model is rarely completely known—unless W_i-s are assigned in the controlled experiment (i.e., Attanasio, Meghir, and Santiago (2012), Broda and Parker (2014), Colonnelli and Prem (2022)) and has to be estimated. We use the insights from the known assignment setting as a building block in Section 3, where the assignment process is unknown but can be estimated consistently from the data. In Section 5, we use an empirical example to show how to estimate this distribution for the staggered adoption design using duration models. This approach is connected to Shaikh and Toulis (2019) that uses a duration model to test a sharp null hypothesis that specifies no treatment effects. Our general strategy of explicitly using the assignment model for estimation is directly connected to the recent literature on quasi-experimental designs (e.g., Borusyak and Hull (2023)). Our results on robustness are especially appealing in such contexts because in quasi-experimental settings researchers cannot rule out the misspecification of the assignment model.

Our focus on TWFE regression (1.2) is motivated by its increased popularity in economics (Currie, Kleven, and Zwiers (2020)). In applications, this model provides an effective and parsimonious approximation for the baseline outcomes, allowing researchers to capture unobserved confounders and to improve the efficiency of the resulting estimator by reducing noise. At the same time, recent research shows that regression estimators for average treatment effects based on TWFE models might have undesirable properties, particularly negative weights for unit-time specific treatment effects. These concerns are particularly salient in settings with heterogeneity in treatment effects and general assignment patterns (e.g., De Chaisemartin and d'Haultfoeuille (2020), Goodman-Bacon (2021), Sun and Abraham (2021), Callaway and Sant'Anna (2021), Borusyak, Jaravel, and Spiess (2024)). Our results show that the concerns raised in this literature regarding negative weights lose some of their force under random assignment, or more generally once we properly reweight the observations.

Our main analysis assumes that the treatment affects only contemporaneous outcomes, ruling out dynamic effects. We make this choice to crystallize the connection between the TWFE regression model (1.2) and the assignment process. We do not restrict heterogeneity in contemporaneous treatment effects that can vary over units and periods. To test for, or estimate, dynamic treatment effects, one has to compare units that receive treatment at different times. Such comparisons are justified only if we restrict individual heterogeneity in treatment effects or treat the assignment as random. Consequently, and this is of course a key insight from the causal inference literature in cross-section settings since Rosenbaum and Rubin (1983), it is imperative to model both the assignment mechanism and the outcome model. In Bojinov, Rambachan, and Shephard

(2021) the authors show how to use the assignment process to estimate dynamic treatment effects (see also Blackwell and Yamauchi (2021) for the related analysis in large-T setup). Our results suggest that a fruitful approach may be to construct robust estimators by combining Bojinov, Rambachan, and Shephard (2021) approach to estimation with conventional dynamic panel regression models using the weighting methods derived in the current paper for the static case. We discuss a particular realization of this in Section 4.

Our results are related to recent literature on doubly robust estimators with panel data. Conceptually, the closest paper to us is Arkhangelsky and Imbens (2022) that also emphasizes the role of the assignment process in the same setting and shows double robustness. In Arkhangelsky and Imbens (2022), the focus is on a class of estimators defined as a linear function of realized outcomes, with the coefficients in that linear representation chosen to lead to consistent estimators for average treatment effects under either assumptions on the outcome model or on the assignment mechanism. Here, we start with a different class of estimators, restricted to weighted versions of the TWFE estimator in (1.2). We also show how to estimate a flexible class of average treatment effects with user-specified weights over units and time. The robustness property in our paper is distinct from the double robustness analyzed recently in the difference-in-difference literature (e.g., Sant'Anna and Zhao (2020)): our estimator is robust to arbitrary violations of parallel trends assumptions, as long as the assignment model is correctly specified. At the same time, our estimator is not necessarily semiparametrically efficient in environments where, as in Sant'Anna and Zhao (2020), the conditional parallel trends assumption holds.

We also connect to recent work on the causal panel model with experimental data (e.g., Athey and Imbens (2022), Bojinov, Rambachan, and Shephard (2021), Roth and Sant'Anna (2023)). Similar to these papers, we establish properties of regression estimators under design assumptions. Importantly, we consider a general setting without restricting our attention to staggered adoption design. Our contribution to this literature is the characterization of the behavior of $\hat{\tau}(\gamma)$ for a large class of weighting functions and general designs. By establishing a connection between weighting functions and limiting estimands, we allow users to construct consistent estimators for a pre-specified weighted average treatment effect of interest.

Finally, the form of our estimator (1.2) connects it to the Synthetic Difference-in-Differences (SDID) estimator introduced in Arkhangelsky, Athey, Hirshberg, Imbens, and Wager (2021). The difference between these two procedures is in the way they construct the weights $\gamma^\star$. The SDID estimator uses pretreatment outcomes to build a synthetic control unit that follows the path of the average treated unit as closely as possible (up to an additive shift). This strategy is infeasible if W_{it} varies over time. However, precisely in situations with enough variation in $\boldsymbol{W}_i$, we can estimate the assignment process and use it to construct the weights $\gamma^\star$. As a result, the two estimators are complementary and can be used in applications with different assignment patterns.

Throughout the paper, we adopt the standard probability notation $O(\cdot)$, $o(\cdot)$, $O_{\mathbb{P}}(\cdot)$, $o_{\mathbb{P}}(\cdot)$. For any vector v, denote by $v^\top$ the transpose of v, $\|v\|_2$ the L_2 norm of v, and by $\mathrm{diag}(v)$ the diagonal matrix with the coordinates of v being the diagonal elements. For a

pair of vectors v_1, v_2, we write $\langle v_1, v_2 \rangle$ for their inner product $v_1^\top v_2$. Furthermore, let $[m]$ denote the set $\{1, \ldots, m\}$, I_m the $m \times m$ identity matrix, and $\mathbf{1}_m$ the m-dimensional vector with all entries 1. Finally, the support of a discrete distribution F is the set of elements with positive probabilities under F.

2. Reshaped IPW estimator with known assignment mechanisms

We consider a setting with n units and each unit is characterized by potential outcomes $\boldsymbol{Y}_i(1) = (Y_{i1}(1), \ldots, Y_{iT}(1))$, $\boldsymbol{Y}_i(0) = (Y_{i1}(0), \ldots, Y_{iT}(0))$ and a set of covariates $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iT})$.[1] By writing the potential outcomes in this form, we assume away any dynamic effects of past treatments on current outcomes, thus focusing on static models. Analysis of such models is useful both theoretically and practically. First, they constitute a building block for more general environments, which we consider in Section 4. Second, when the treatment is irreversible, as in staggered adoption designs, we are likely interested in its average (over time) effect on the outcome rather than the transitory dynamics. This makes the static model a reasonable approximation for a more complicated dynamic model. Finally, if we observe the data at a lower frequency than the one that is relevant for dynamics (e.g., days vs. months), then the static model is the only available option.

Given the realized treatment assignment W_{it}, the observed outcomes are defined in the usual way:

$$Y_{it} = Y_{it}(1)W_{it} + Y_{it}(0)(1 - W_{it}). \tag{2.1}$$

Throughout the paper, we treat covariates as fixed and consider $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ as a random vector (jointly) drawn from a distribution (conditional on $\{\boldsymbol{X}_i : i \in [n]\}$). We let $\mathbb{P}$ denote the joint distribution of the entire random vector $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ (conditional on $\{\boldsymbol{X}_i : i \in [n]\}$) and $\mathbb{E}$ denote the expectation over this distribution. We consider the asymptotic regime with n going to infinity and fixed $T \geq 2$.

This structure nests the conventional sampling-based framework, which is common in panel data analysis, going back to (Chamberlain (1984)), and which was used to establish statistical results in the recent DiD literature (e.g., Abadie (2005), Callaway and Sant'Anna (2021)). It also extends the standard fixed-effects framework, where the distribution for each unit is characterized by unit-specific parameters, but units themselves are usually assumed independent (e.g., Neyman and Scott (1948), Lancaster (2000)). Even in the absence of any covariates, we do not assume that unit-level observations $(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i)$ are independent or exchangeable, which brings two practical advantages. First, it allows us to accommodate correlated potential outcomes among units, which is natural in applications involving networks or multilevel structures. Second, it allows the assignments to be correlated across units, which is natural for many commonly used experimental designs. We elaborate on this point in the next section.

In this section, we study a special case where the assignment mechanism is known. This assumption is natural for experimental settings (Brown and Lilford (2006), Attanasio, Meghir, and Santiago (2012), Broda and Parker (2014), Hemming, Haines, Chilton,

[1] Time-invariant covariates can be handled by letting $X_{i1} = \cdots = X_{iT}$.

Girling, and Lilford (2015), Chandar, Gneezy, List, and Muir (2019), Chandar, Hortaçsu, List, Muir, and Wooldridge (2019), Colonnelli and Prem (2022)), but it has also been used to analyze the quasi-experimental settings (e.g., Borusyak and Hull (2023)). It allows us to derive inferential results under mild assumptions. We will consider the case of unknown designs in Section 3 at the cost of stronger (yet standard) assumptions.

2.1 *A design-based causal framework*

We assume that, for any $i \in \{1, \dots, n\}$ and $\boldsymbol{w} \in \{0, 1\}^T$,

$$\mathbb{P}\big(\boldsymbol{W}_i = \boldsymbol{w} | \boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0)\big) = \boldsymbol{\pi}_i(\boldsymbol{w}), \tag{2.2}$$

where $\boldsymbol{\pi}_i$ is a distribution known to the analyst. We call it the generalized propensity score (Imbens (2000), Athey and Imbens (2022), Bojinov, Rambachan, and Shephard (2021), Bojinov, Simchi-Levi, and Zhao (2023))—the marginal probability of the treatment path.

This structure allows for covariate-adaptive designs, where the probability of W_{it} depends on past covariates. However, we rule out sequentially-adaptive designs where the assignment can depend on past outcomes, even if the randomization protocol is known.[2] Furthermore, our framework places no restriction on the support of $\boldsymbol{W}_i$ and substantially generalizes the previous works that focus on simple random sampling for nonstaggered difference-in-differences (Rambachan and Roth (2020)) and staggered adoption (Athey and Imbens (2022), Roth and Sant'Anna (2023)).

If the treatment paths $\{\boldsymbol{W}_i, i \in [n]\}$ are independent across units, then the marginal distributions $\{\boldsymbol{\pi}_i(\boldsymbol{w}), i \in [n]\}$ characterize the joint distribution of $\{\boldsymbol{W}_i, i \in [n]\}$. However, as discussed above, we allow the assignments to be correlated across units. In practice, this correlation can range from being very mild, as in the case of completely randomized experiments with a fixed share of treated units (Neyman (1923/1990)), to being sizable, as in cases of cluster-level randomization such as cluster randomized design (Abadie, Athey, Imbens, and Wooldridge (2023)) and two-stage randomization. We impose technical restrictions on the dependence across units in Section 2.3.

2.2 *Causal estimands*

We define the unit and time-specific treatment effect as

$$\tau_{it} \triangleq \mathbb{E}\big[Y_{it}(1) - Y_{it}(0)\big]. \tag{2.3}$$

Note that τ_{it} can vary with both i and t since we assume neither identically distributed units nor time-homogeneous treatment effects. For time period t, we define the time-specific ATE as

$$\tau_t \triangleq \frac{1}{n} \sum_{i=1}^{n} \tau_{it}, \tag{2.4}$$

[2]Even if units are i.i.d. and $\mathbb{P}(W_{it}|Y_{i1}, \dots, Y_{i(t-1)})$ is known, $\mathbb{P}(W_{it}|Y_{i1}, \dots, Y_{iT})$ would depend on the unknown conditional distribution of $(Y_{it}, \dots, Y_{iT})$ given W_{it}.

and consider a broad class of weighted average of time-specific ATE:

$$\tau^*(\xi) \triangleq \sum_{t=1}^{T} \xi_t \tau_t \tag{2.5}$$

for some user-specified deterministic weights $\xi = (\xi_1, \ldots, \xi_T)^\top$ such that

$$\sum_{t=1}^{T} \xi_t = 1, \quad \xi_t \geq 0. \tag{2.6}$$

We refer to (2.5) as a doubly average treatment effect (DATE). For example, the weights $\xi_t = 1/T$ yield the usual ATE over units and time periods. In the difference-in-differences setting with two time periods, $\xi_t = \mathbf{1}_{t=2}$. In a particular application, one might also be interested in an effect with time discounting factor that puts more weight on initial periods, that is, $\xi_t \propto \beta^t$ for some $\beta < 1$.

REMARK 2.1. We can further generalize DATE by allowing for unequal unit weights:

$$\tau^*(\xi; \zeta) = \sum_{i=1}^{n} \sum_{t=1}^{T} (\zeta_i \xi_t) \tau_{it}, \tag{2.7}$$

where $\zeta = (\zeta_1, \ldots, \zeta_n)$, $\sum_{i=1}^{n} \zeta_i = 1$ and $\zeta_i \geq 0$. In particular, ζ_i can be i-specific, for example, a function of the ith covariates, but cannot depend on outcomes and treatment assignments. Using appropriate propensity-based weights ζ, one can build estimands that target a given subpopulation.

2.3 *Technical assumptions*

We allow $\boldsymbol{W}_i$ to be dependent across units to capture different assignment processes. Such dependence arises in applications, sometimes for technical reasons (e.g., in case of sampling without replacement as in Athey and Imbens (2022)), and sometimes by the nature of the assignment process (spatial experiments). To quantify this dependence as well as the dependence among the potential outcomes, we follow Rényi (1959) and define the maximal correlation:

$$\rho_{ij} \triangleq \sup_{f,g} \big\{ \operatorname{corr}\big(f\big(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i\big), g\big(\boldsymbol{Y}_j(1), \boldsymbol{Y}_j(0), \boldsymbol{W}_j\big)\big) \big\}, \tag{2.8}$$

where the supremum is taken over all real-valued measurable functions f, g.

In the standard design-based framework where potential outcomes are assumed fixed, it reduces to the ρ-mixing coefficient between $\boldsymbol{W}_i$ and $\boldsymbol{W}_j$. In the main text, we maintain a simplified restriction on $\{\rho_{ij}\}_{ij}$ leaving a more general one to Supplemental Appendix A. The assumption is stated as follows.

ASSUMPTION 2.1. *There exists $q \in (0, 1]$, such that as n approaches infinity, the following holds*:

$$\frac{1}{n^2}\sum_{i,j=1}^{n}\rho_{ij} = O\big(n^{-q}\big). \tag{2.9}$$

By definition, $\frac{1}{n} \le (1/n^2)\sum_{i,j=1}^{n}\rho_{ij} \le 1$ with lower bound being attained if the observations are independent, and the upper bound being attained if they are perfectly dependent. As a result, one can view q as measuring the strength of the correlation. When $(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i)$ are independent across units, (2.9) holds with $q = 1$. More generally, when $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ have a network dependency with $\rho_{ij} = 0$ if there is no edge between i and j, (2.9) is satisfied if the number of edges is $O(n^{2(1-q)})$. Note that it imposes no constraint on the maximum degree of the dependency graph. Even if the network is fully connected, it can still hold if the pairwise dependence is weak, for example, sampling without replacement; see Supplemental Appendix A.4. On the other hand, (2.9) excludes the case where all units are perfectly correlated or equicorrelated with a positive maximal correlation that is bounded away from 0.

We also impose minimal overlap restrictions on each $\boldsymbol{\pi}_i$.

ASSUMPTION 2.2. *There exists a universal constant $c > 0$ and a nonstochastic subset $\mathbb{S}^* \subset \{0, 1\}^T$ with at least two elements and at least one element not in $\{\mathbf{0}_T, \mathbf{1}_T\}$, such that*

$$\boldsymbol{\pi}_i(\boldsymbol{w}) > c, \quad \forall \boldsymbol{w} \in \mathbb{S}^*, i \in [n]. \tag{2.10}$$

Our final assumption restricts the second moment of outcomes.

ASSUMPTION 2.3. *There exists $M < \infty$ such that* $\max_{i,t,w}\mathbb{E}[Y_{it}^2(w)] < M$.

It is presented here only for simplicity. We relax it substantially in Supplemental Appendix A.

2.4 *Reshaped IPW estimator*

We consider a class of weighted TWFE regression estimators without covariates. We refer to them as reshaped inverse propensity weighted (RIPW) estimators, and formally define them as follows:

$$\hat{\tau}(\boldsymbol{\Pi}) \triangleq \underset{\tau,\mu,\sum_i \alpha_i = \sum_t \lambda_t = 0}{\arg\min} \sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it} - \mu - \alpha_i - \lambda_t - W_{it}\tau)^2 \frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)}{\boldsymbol{\pi}_i(\boldsymbol{W}_i)}, \tag{2.11}$$

where $\boldsymbol{\Pi}(\boldsymbol{w})$ is a density function on $\{0, 1\}^T$, that is,

$$\sum_{\boldsymbol{w}\in\{0,1\}^T}\boldsymbol{\Pi}(\boldsymbol{w}) = 1. \tag{2.12}$$

We refer to the distribution $\mathbf{\Pi}$ as a reshaped distribution, and the weight $\mathbf{\Pi}(\boldsymbol{W}_i)/\boldsymbol{\pi}_i(\boldsymbol{W}_i)$ as a RIP weight. To ensure that the RIPW estimator is well-defined, we require $\mathbf{\Pi}$ to be absolutely continuous with respect to each $\boldsymbol{\pi}_i$, that is,

$$\mathbf{\Pi}(\boldsymbol{w}) = 0 \quad \text{if } \boldsymbol{\pi}_i(\boldsymbol{w}) = 0 \text{ for some } i \in [n]. \tag{2.13}$$

The estimator (2.11) is feasible for any such $\mathbf{\Pi}$ because $\boldsymbol{\pi}_i$ is assumed to be known.

Adding covariates to the objective function (2.11) is relatively straightforward. However, it considerably complicates the notation without contributing substantially to the primary narrative. We will explicitly incorporate covariates in the objective function in Section 3. Note that the covariates still play a role in the RIPW estimator through $\boldsymbol{\pi}_i$ for covariate-adaptive designs.

The reshaped distribution $\mathbf{\Pi}$ can be interpreted as an experimental design. If $\boldsymbol{W}_i \sim \mathbf{\Pi}$, then $\boldsymbol{\pi}_i = \mathbf{\Pi}$ and (2.11) reduces to the standard unweighted TWFE regression. If this is not the case, then $\mathbf{\Pi}(\boldsymbol{W}_i)/\boldsymbol{\pi}_i(\boldsymbol{W}_i)$ acts like a likelihood ratio that changes the original design to one provided by $\mathbf{\Pi}$. For cross-sectional data, we would like to shift the distribution to uniform $\{0, 1\}$, making the weights equal to $1/2\boldsymbol{\pi}_i(\boldsymbol{W}_i)$ if the fixed effects are not included. This would yield the standard IPW estimator. However, as we alluded to in the Introduction, the situation is more complicated with panel data, and shifting toward the uniform design might not deliver consistent estimators for the DATE of interest. We explore this formally in the next section, where we characterize the set of $\mathbf{\Pi}$ that one can use. This interpretation of $\mathbf{\Pi}$ has one caveat: RIP weights only shift the marginal distribution of $\boldsymbol{W}_i$ to $\mathbf{\Pi}$, but they do not say anything about the joint distribution of $\{\boldsymbol{W}_i, i \in [n]\}$, which can remain complicated.

2.5 *DATE equation and consistency of RIPW estimators*

We now derive sufficient conditions under which the RIPW estimator is a consistent estimator for a given DATE of interest. The following theorem presents a precise condition for consistency of $\hat{\tau}(\mathbf{\Pi})$ for $\tau^*(\xi)$.

THEOREM 2.1. *Let $J = I_T - \mathbf{1}_T\mathbf{1}_T^\top/T$ and $\boldsymbol{\tau}_i = (\tau_{i1}, \ldots, \tau_{iT})^\top$; fix ξ that satisfies (2.6). Under Assumptions 2.1–2.3, for any reshaped distribution $\mathbf{\Pi}$ with support $\mathbb{S}^*$ defined in Assumption 2.2, as n tends to infinity,*

$$\hat{\tau}(\mathbf{\Pi}) - \tau^*(\xi) = O_{\mathbb{P}}\big(\mathrm{Bias}_\tau(\xi)\big) + o_{\mathbb{P}}(1),$$

where

$$\mathrm{Bias}_\tau(\xi) = \left\langle \mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\big(\mathrm{diag}(\boldsymbol{W}) - \xi\boldsymbol{W}^\top\big)J\big(\boldsymbol{W} - \mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)\big], \frac{1}{n}\sum_{i=1}^{n}\big(\boldsymbol{\tau}_i - \tau^*(\xi)\mathbf{1}_T\big)\right\rangle,$$

and $\langle v_1, v_2\rangle$ denotes their inner product $v_1^\top v_2$.

This result has two user-specified parameters: time weights ξ, and the reshaped distribution $\mathbf{\Pi}$. They are naturally connected; to guarantee consistency for $\tau^*(\xi)$, we can select $\mathbf{\Pi}$ such that the following holds:

$$\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\big(\operatorname{diag}(\boldsymbol{W})-\xi\boldsymbol{W}^\top\big)J\big(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)\big]=0. \tag{2.14}$$

Alternatively, for a given $\mathbf{\Pi}$, we can look for ξ such that (2.14) is satisfied. We call (2.14) the DATE equation hereafter. For a fixed ξ, it is a quadratic system with $\{\mathbf{\Pi}(\boldsymbol{w}):\boldsymbol{w}\in\{0,1\}^T\}$ being the variables. Together with the density constraint (2.12) and the support constraint in Theorem 2.1 that $\mathbf{\Pi}(\boldsymbol{w})=0$ for $\boldsymbol{w}\notin\mathbb{S}^*$, there are $T+1+2^T-|\mathbb{S}^*|$ equality constraints and $|\mathbb{S}^*|$ inequality constraints that impose the positivity of $\mathbf{\Pi}(\boldsymbol{w})$ for each $\boldsymbol{w}\in\mathbb{S}^*$. We will show in Supplemental Appendix C that the DATE equation has closed-form solutions in various examples and provide a generic solver based on nonlinear programming in Supplemental Appendix C.5.

Without further restrictions on $\boldsymbol{\tau}_i$, we can show that the DATE equation is also a necessary condition for consistency of $\hat{\tau}(\mathbf{\Pi})$ for $\tau^*(\xi)$. To see this, assume that

$$\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\big(\operatorname{diag}(\boldsymbol{W})-\xi\boldsymbol{W}^\top\big)J\big(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)\big]=z. \tag{2.15}$$

for some vector z that is not proportional to ξ. Because we can vary individual treatment effects without changing the average one, we can find a set $\{\boldsymbol{\tau}_i:i\in[n]\}$ that yields the same DATE but $\langle z,(1/n)\sum_{i=1}^n(\boldsymbol{\tau}_i-\tau^*(\xi)\mathbf{1}_T)\rangle\neq 0$, leading to inconsistency. For $z=b\xi$, we get that the inner product of the LHS of (2.15) and $\mathbf{1}_T$ is 0 because $\mathbf{1}_T^\top(\operatorname{diag}(\boldsymbol{W})-\xi\boldsymbol{W}^\top)=W^\top(1-\mathbf{1}_T^\top\xi)=0$, while that of the right-hand side and $\mathbf{1}_T$ is equal to b. This implies that z has to be equal to zero, thus proving the necessity of DATE equation.

Notably, when the DATE equation has a solution, our estimator is consistent without any restrictions on the potential outcomes, except Assumption 2.3. This is in sharp contrast to usual results about TWFE estimators, which typically require the trends to be parallel among units, at least conditionally on observed covariates (e.g., Callaway and Sant'Anna (2021), Sant'Anna and Zhao (2020)). Theorem 2.1 shows that if the assignment process is known and the DATE equation has a solution, we can correct the potentially misspecified TWFE regression model by simply reweighting the objective function. We want to stress that this result relies on the knowledge of the assignment process, whereas the analysis based on conditional parallel trends does not require such knowledge.

To further parse the DATE equation, we discuss two alternative interpretations. First, fix ξ and let $\mathbf{\Pi}$ be the solution of the DATE equation. Then consider a class of complete randomized experiments where all propensity scores $\boldsymbol{\pi}_i$ are identical and are equal to $\mathbf{\Pi}$. Then, by definition, the RIPW estimator with reshaped distribution $\mathbf{\Pi}$ reduces to the standard (unweighted) TWFE estimator. Theorem 2.1 guarantees that this estimator converges to $\tau^*(\xi)$. Since the DATE equation is a necessary condition, all experimental designs that do not satisfy this restriction cannot lead to a consistent estimator for $\tau^*(\xi)$. As a result, DATE equation characterizes all complete randomized experiments under which the unweighted two-way estimator converges to a given estimand. This

can be interpreted as a general converse of the results established in Athey and Imbens (2022).

As an alternative interpretation, consider a fixed $\mathbf{\Pi}$ instead. For any such $\mathbf{\Pi}$, the equation (2.14) can be rewritten as

$$\big(\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\boldsymbol{W}^\top J\big(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)\big]\big)\xi=\mathbb{E}\big[\operatorname{diag}(\boldsymbol{W})J\big(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)\big]. \tag{2.16}$$

It is easy to see that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\boldsymbol{W}^\top J\big(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)\big]&=\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\big(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)^\top J\big(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)\big]\\&=\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\big\|\tilde{\boldsymbol{W}}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\tilde{\boldsymbol{W}}]\big\|_2^2\big],\end{aligned} \tag{2.17}$$

where $\tilde{\boldsymbol{W}}=J\boldsymbol{W}$. Since the support of $\mathbf{\Pi}$ involves a point $\boldsymbol{w}\notin\{\mathbf{0}_T,\mathbf{1}_T\}$, for which $\boldsymbol{w}'\neq 0$ the quantity in (2.17) is strictly positive. Therefore, (2.16) implies that

$$\xi=\frac{\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\operatorname{diag}(\boldsymbol{W})J\big(\boldsymbol{W}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\boldsymbol{W}]\big)\big]}{\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}\big[\big\|\tilde{\boldsymbol{W}}-\mathbb{E}_{\boldsymbol{W}\sim\mathbf{\Pi}}[\tilde{\boldsymbol{W}}]\big\|_2^2\big]}. \tag{2.18}$$

By Theorem 2.1, in a randomized experiment with $\boldsymbol{\pi}_i\triangleq\mathbf{\Pi}$ (Athey and Imbens (2022), Roth and Sant'Anna (2023)), the effective estimand of the unweighted TWFE regression is the DATE with weight vector ξ.

Remark 2.2. To illustrate this result, we consider the experiment conducted by Uber in 2017 to test the effect of in-app tipping on labor supply (Chandar et al. (2019,?)). They introduced the in-app tipping feature in a staggered fashion across 209 operational cities in the United States and Canada to avoid bugs in the product. Three cities were randomized to launch this feature on June 20, 2017, followed by 103 cities on July 6, 2017, and the remaining 103 cities on July 17, 2017. We can treat it as a two-period experiment, with June 20–July 5 being the first period and July 6–July 16 being the second period. The possible assignments include $\{1,1\}$, $\{0,1\}$, $\{0,0\}$ and $\boldsymbol{\pi}_i(\{1,1\})=3/209$, $\boldsymbol{\pi}_i(\{0,1\})=\boldsymbol{\pi}_i(\{0,0\})=103/209$. By (C.1) in Supplemental Appendix C, (2.18) implies that $\xi_1=3/106$, $\xi_2=103/106$ for the unweighted TWFE regression, which they applied to estimate the treatment effect. Thus, their analysis is essentially focused on the second period.

The following result shows that the induced weights are guaranteed to be nonnegative for arbitrary design.

Proposition 2.1. *Let ξ be defined in (2.18). Then for any $\mathbf{\Pi}$ on $\{0,1\}^T$, $\xi_t\geq 0$ for all t.*

This result generalizes the conventional cross-sectional logic that says that in randomized experiments, regression estimators are consistent for average effects (e.g., Lin (2013)). However, in the case of the TWFE regression, the situation is more nuanced. While the resulting estimand always corresponds to a weighted average effect with nonnegative weights, it still depends on the experimental design. As a result, if two analysts

were to split a given population into two random subpopulations and conduct two experiments with different designs on each part, the resulting estimands would have been different.

There are two reasons for this unusual behavior. First, in the cross-sectional case, $\boldsymbol{W}_i$ has two points of support, while in the panel case the support of $\boldsymbol{W}_i$ ranges from 2 to 2^T points (as long as Assumption 2.2 is satisfied). For example, if none of the units are treated in the first period, it is impossible to identify any DATE that puts positive weight on the first period. Second, fixed effects lead to a familiar incidental parameter problem (Neyman and Scott (1948)), albeit in a mild form. To see this, consider $\boldsymbol{\pi}_i = \boldsymbol{\Pi} \propto 1$, in which case the RIPW estimator corresponds to the conventional TWFE regression. The effective estimand for this regression is equal to the solution of (2.18) and is different from the effective estimand for the regression without the unit fixed effects.

REMARK 2.3. To estimate the generalized DATE defined in (2.7), we only need to mildly adjust the RIPW estimator:

$$\hat{\tau}(\boldsymbol{\Pi};\zeta) \triangleq \underset{\tau,\mu,\sum_i \alpha_i=\sum_t \lambda_t=0}{\arg\min} \sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it}-\mu-\alpha_i-\lambda_t-W_{it}\tau)^2 \frac{\zeta_i \boldsymbol{\Pi}(\boldsymbol{W}_i)}{\boldsymbol{\pi}_i(\boldsymbol{W}_i)}. \tag{2.19}$$

In Supplemental Appendix A.7, we prove that the adjusted RIPW estimator $\hat{\tau}(\boldsymbol{\Pi};\zeta)$ consistently estimates $\tau^*(\xi;\zeta)$ under the same set of assumptions as in Theorem 2.1, provided that $n\|\zeta\|_\infty = O(1)$, namely that all entries of ζ are on the same scale.

2.6 *Inference on RIPW estimators*

To enable statistical inference of DATE, we first present an asymptotic expansion showing the asymptotic linearity of RIPW estimators.

THEOREM 2.2. *Let $\boldsymbol{Y}_i$ be the vector $(Y_{i1},\dots,Y_{iT})$. Further, let $\Theta_i = \boldsymbol{\Pi}(\boldsymbol{W}_i)/\boldsymbol{\pi}_i(\boldsymbol{W}_i)$, and*

$$\Gamma_\theta \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i, \qquad \Gamma_{ww} \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i \boldsymbol{W}_i^\top J \boldsymbol{W}_i, \qquad \Gamma_{wy} \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i \boldsymbol{W}_i^\top J \boldsymbol{Y}_i,$$

and

$$\boldsymbol{\Gamma}_w \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i J \boldsymbol{W}_i, \qquad \boldsymbol{\Gamma}_y \triangleq \frac{1}{n}\sum_{i=1}^{n}\Theta_i J \boldsymbol{Y}_i.$$

Under the same settings as Theorem 2.1,

$$\mathcal{D}\cdot\sqrt{n}\big(\hat{\tau}(\boldsymbol{\Pi})-\tau^*(\xi)\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]\big) + O_{\mathbb{P}}\big(n^{1/2-2q}\big),$$

where $\mathcal{D} = \Gamma_{ww}\Gamma_\theta - \boldsymbol{\Gamma}_w^\top \boldsymbol{\Gamma}_w$, and

$$\begin{aligned}\mathcal{V}_i = \Theta_i\Big\{&\big(\mathbb{E}[\Gamma_{wy}] - \tau^*(\xi)\mathbb{E}[\Gamma_{ww}]\big) - \big(\mathbb{E}[\boldsymbol{\Gamma}_y] - \tau^*(\xi)\mathbb{E}[\boldsymbol{\Gamma}_w]\big)^\top J\boldsymbol{W}_i \\ &+ \mathbb{E}[\Gamma_\theta]\boldsymbol{W}_i^\top J\big(\boldsymbol{Y}_i - \tau^*(\xi)\boldsymbol{W}_i\big) - \mathbb{E}[\boldsymbol{\Gamma}_w]^\top J\big(\boldsymbol{Y}_i - \tau^*(\xi)\boldsymbol{W}_i\big)\Big\}\end{aligned}$$

Note that the asymptotic linear expansion holds under a fairly general dependency structure in the treatment assignments. Below, we derive a valid confidence interval for $\tau^*(\xi)$ when $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ are independent. The general case is discussed in Supplemental Appendix A.4. If $\{\mathcal{V}_i : i \in [n]\}$ are well-behaved Theorem 2.2 implies that

$$\frac{\mathcal{D} \cdot \sqrt{n}\big(\hat{\tau}(\boldsymbol{\Pi}) - \tau^*(\xi)\big)}{\sigma_n^*} \approx N(0, 1), \quad \text{where } \sigma_n^{*2} = (1/n)\sum_{i=1}^n \mathrm{Var}(\mathcal{V}_i),$$

where $\mathcal{D}$ is known by design.[3] If $\{\mathcal{V}_i : i \in [n]\}$ were known, a natural estimator for σ_n^{*2} would be the empirical variance:

$$\hat{\sigma}_n^{*2} = \frac{1}{n-1}\sum_{i=1}^n (\mathcal{V}_i - \bar{\mathcal{V}})^2, \quad \text{where } \bar{\mathcal{V}} = \frac{1}{n}\sum_{i=1}^n \mathcal{V}_i.$$

We should not expect the difference between $\hat{\sigma}_n^*$ and σ_n^* to converge to zero since $\mathbb{E}[\mathcal{V}_i]$ in general varies over i. Nonetheless, $\hat{\sigma}_n^*$ is an asymptotically conservative estimate of σ_n^* since

$$\mathbb{E}\big[\hat{\sigma}_n^{*2}\big] \approx \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\left(\mathcal{V}_i - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathcal{V}_i]\right)^2\right] \approx \sigma_n^{*2} + \underbrace{\frac{1}{n-1}\sum_{i=1}^n \left(\mathbb{E}[\mathcal{V}_i] - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\mathcal{V}_i]\right)^2}_{\text{empirical variance of } \mathbb{E}[\mathcal{V}_i]}, \quad (2.20)$$

where the second term measures the heterogeneity of $\mathbb{E}[\mathcal{V}_i]$ and is always nonnegative, implying that $\hat{\sigma}_n^{*2}$ is a conservative estimator for σ_n^{*2}. This is unsurprising because even in the cross-section case, the asymptotic design-based variance is only partially identifiable due to the unknown correlation structure between two potential outcomes; see, for example, Neyman's variance formula (Neyman (1923/1990), Rubin (1974)).

In general, $\mathcal{V}_i$ is unknown due to $\tau^*(\xi)$ and the expectation terms. Nonetheless, we can estimate $\mathcal{V}_i$ by replacing each expectation with the corresponding plug-in estimate, that is,

$$\begin{aligned}\hat{\mathcal{V}}_i = \Theta_i\big\{&(\Gamma_{wy} - \hat{\tau}\Gamma_{ww}) - (\boldsymbol{\Gamma}_y - \hat{\tau}\boldsymbol{\Gamma}_w)^\top J\boldsymbol{W}_i \\ &+ \Gamma_\theta \boldsymbol{W}_i^\top J(\boldsymbol{Y}_i - \hat{\tau}\boldsymbol{W}_i) - \boldsymbol{\Gamma}_w^\top J(\boldsymbol{Y}_i - \hat{\tau}\boldsymbol{W}_i)\big\},\end{aligned} \quad (2.21)$$

and use them to compute the variance:

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (\hat{\mathcal{V}}_i - \bar{\hat{\mathcal{V}}})^2, \quad \text{where } \bar{\hat{\mathcal{V}}} = \frac{1}{n}\sum_{i=1}^n \hat{\mathcal{V}}_i. \quad (2.22)$$

This yields a Wald-type confidence interval for $\tau^*(\xi)$ as

$$\hat{C}_{1-\alpha} = \big[\hat{\tau}(\boldsymbol{\Pi}) - z_{1-\alpha/2}\hat{\sigma}/(\sqrt{n}\mathcal{D}),\ \hat{\tau}(\boldsymbol{\Pi}) + z_{1-\alpha/2}\hat{\sigma}/(\sqrt{n}\mathcal{D})\big], \quad (2.23)$$

[3] By well-behaved $\mathcal{V}_i$, we mean that they are sufficiently regular for the appropriate version of the central limit theorem to hold. In the simplest case, when data is i.i.d., this reduces to standard moment restrictions.

where z_η is the ηth quantile of the standard normal distribution. Properties of this confidence interval are established in the next theorem.

THEOREM 2.3. *Assume that* $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ *are independent with*

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(\mathcal{V}_i) \geq v_0, \quad \textit{for some constant } v_0 > 0. \tag{2.24}$$

Then under Assumptions 2.2 and 2.3, for any $\alpha \in (0, 1)$,

$$\liminf_{n\to\infty}\mathbb{P}\big(\tau^*(\xi) \in \hat{C}_{1-\alpha}\big) \geq 1 - \alpha.$$

In Supplemental Appendix A.4, we discuss a generic result for general dependent assignments (Theorem A.6), which covers completely randomized experiments, blocked and matched pair experiments, two-stage randomized experiments, and so on. We present a detailed result (Theorem A.7) for completely randomized experiments where potential outcomes are fixed and $\boldsymbol{W}_i$'s are sampled without replacement from a user-specified subset of $\{0, 1\}^T$.[4] This substantially generalizes the setting of Athey and Imbens (2022) and Roth and Sant'Anna (2023), where the assignments are sampled without replacement from the set of $T+1$ staggered assignments. At the same time, compared to Roth and Sant'Anna (2023), we cannot provide efficiency guarantees for our estimator.

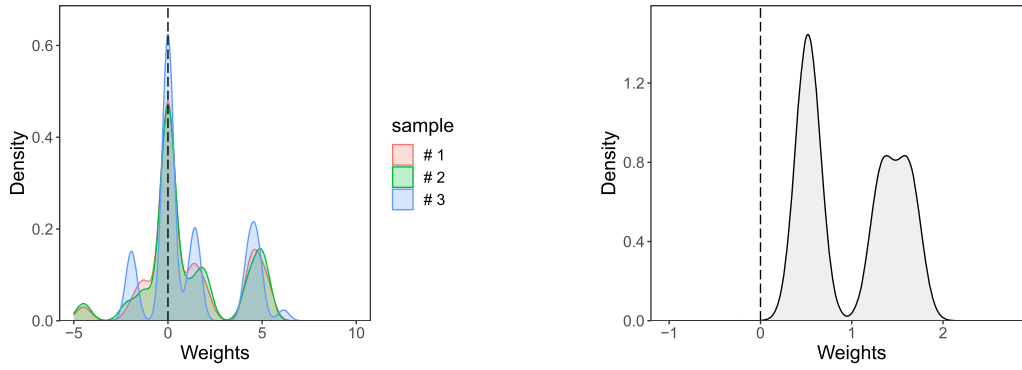
2.7 *Discussion*

Theorem 2.1 and Proposition 2.1 might appear counterintuitive given well-understood problems of TWFE estimators (e.g., De Chaisemartin and d'Haultfoeuille (2020), Goodman-Bacon (2021), Sun and Abraham (2021)). To put our result in context, we emphasize two important features of the setup. First, we restrict attention to static models, and second, we use the randomness that is coming from $\boldsymbol{W}_i$. Both of these restrictions play a key role in Theorem 2.1. The absence of dynamic effects implies that we can meaningfully average units with different histories of past treatments. A version of this assumption is inescapable if we want the method to work for general designs where controlling for past history is practically infeasible. As we explain below, the randomness of assignments helps to resolve the issue that TWFE estimators put negative weights on some individual treatment effects.

In De Chaisemartin and d'Haultfoeuille (2020), Goodman-Bacon (2021), Sun and Abraham (2021), the authors show that treated units are averaged with potentially negative weights, but these results are conditional on the assignments $\boldsymbol{W} = (\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n)$ being fixed. Let $\xi_{it}(\gamma; \boldsymbol{W})$ be these weights for the general weighted least squares estimator $\hat{\tau}(\gamma)$ defined in (1.2) such that

$$\mathbb{E}\big[\hat{\tau}(\gamma)|\boldsymbol{W}\big] = \sum_{i=1}^{n}\sum_{t=1}^{T}\xi_{it}(\gamma; \boldsymbol{W})\tau_{it},$$

[4]Specifically, given any support $\mathbb{S}^*$ and a prespecified vector $\{n_{\boldsymbol{w}} : \boldsymbol{w} \in \mathbb{S}^*\}$ with $\sum_{\boldsymbol{w}\in\mathbb{S}^*} n_{\boldsymbol{w}} = n$, the experimenter sample assignments $(\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n)$ with probability $\prod_{\boldsymbol{w}\in\mathbb{S}^*} n_{\boldsymbol{w}}!/n!$.



(a) Histograms of $(nT) \cdot \xi_{it}(\gamma; \boldsymbol{W})$'s (b) Histogram of $(nT) \cdot \xi_{it}(\gamma)$'s

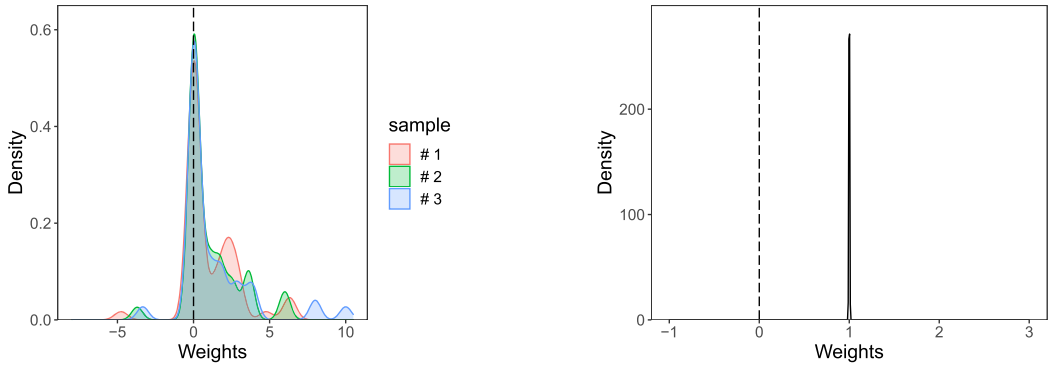
FIGURE 1. Effect weights for the unweighted TWFE estimator.

where we now explicitly allow the weights to depend on $\boldsymbol{W}$. When the assignments are treated as random, the large sample limit of $\hat{\tau}(\gamma)$ is

$$\mathbb{E}\big[\hat{\tau}(\gamma)\big] = \sum_{i=1}^{n} \sum_{t=1}^{T} \xi_{it}(\gamma) \tau_{it},$$

where $\xi_{it}(\gamma) = \mathbb{E}_{\boldsymbol{W}}[\xi_{it}(\gamma; \boldsymbol{W})]$. While $\{(i, t) : \xi_{it}(\gamma; \boldsymbol{W}) < 0\}$ is nonempty almost surely for every realization of $\boldsymbol{W}$, it is still possible that all $\xi_{it}(\gamma)$ are positive due to the averaging over $\boldsymbol{W}$. For illustration, we consider a simulation study with $n = 100$, $T = 4$, and other details specified in Section 5.1. We consider the conditional and unconditional weights induced by the unweighted and RIP-weighted TWFE estimator in Figure 1 and Figure 2, respectively. We plot the histograms of $\{(nT) \cdot \xi_{it}(\gamma; \boldsymbol{W}) : i \in [n], t \in [T]\}$ for three realizations of $\boldsymbol{W}$ and the histogram of $\{(nT) \cdot \xi_{it}(\gamma) : i \in [n], t \in [T]\}$, approximately by averaging over a million realizations of $\boldsymbol{W}$, where the multiplicative factor nT is chosen to normalize the weights into a more interpretable scale. Clearly, despite the large fraction of negative weights in each realization, their averages do not have any negatives. Therefore, the criticism on TWFE estimators does not apply in this case. Indeed, it never applies to the RIPW estimator by Proposition 2.1. In this study, all weights are designed to be $1/nT > 0$ when $\boldsymbol{\Pi}$ is a solution of the DATE equation with $\xi = \mathbf{1}_T/T$, as shown in Figure 2(b), regardless of the data generating process.

The discussion above demonstrates that while for each cell (i, t), a particular realization of weights can be negative, this fact is not systematic. If we use the RIPW estimator designed for the equally weighted DATE, then all cells will receive the same weight on average. An alternative description of the same phenomenon is that once correctly weighted, the realized treatment paths $\boldsymbol{W}_i$ are independent of potential outcomes. This independence implies that there cannot be systematic differences in treatment effects among units with distinct assignment paths, and thus negative weights do not create complications for the interpretation of the estimates. As we illustrate in Section 4, this interpretation remains valid even when certain dynamic effects are present.



(a) Histograms of $(nT)\cdot\xi_{it}(\gamma;\boldsymbol{W})$'s (b) Histogram of $(nT)\cdot\xi_{it}(\gamma)$'s

FIGURE 2. Effect weights for our RIPW estimator.

REMARK 2.4. One might ask if Figure 1(b) presents a general feature of unweighted TWFE estimators with random assignments. For completely randomized experiments where $\boldsymbol{\pi}_i \equiv \boldsymbol{\Pi}$, the standard TWFE estimator is equivalent to the RIPW estimator with reshaped distribution $\boldsymbol{\Pi}$. By Proposition 2.1, all weights are guaranteed to be nonnegative. When $\boldsymbol{\pi}_i$ varies across units, the weights are not guaranteed to be nonnegative. Consider the extreme case where $\boldsymbol{\pi}_i$ assigns $1-\epsilon$ mass on one assignment pass and ϵ mass on all others. As $\epsilon \to 0$, this approaches the case of fixed-treatment assignments, for which the unconditional weights are almost the same as the conditional weights, which always include negative ones.

3. RESHAPED IPW ESTIMATOR WITH UNKNOWN ASSIGNMENT MECHANISMS

In this section, we move to nonexperimental settings where the assignment mechanism is not controlled by the researcher and is unknown. We assume that researchers constructed unit-level estimates $\{\hat{\boldsymbol{\pi}}_i, i \in [n]\}$. In addition, we assume that the researchers have access to a set of estimates $\{(\hat{\boldsymbol{\mu}}_i(0), \hat{\boldsymbol{\mu}}_i(1)) : i \in [n]\}$ of $\{(\mathbb{E}[\boldsymbol{Y}_i(0)], \mathbb{E}[\boldsymbol{Y}_i(1)]) : i \in [n]\}$. Further, let $\hat{m}_{it}$ be the double-centered version of $\hat{\mu}_{it}(0)$ and $\hat{\nu}_{it}$ be a shifted version of $\hat{\mu}_{it}(1) - \hat{\mu}_{it}(0)$:

$$\hat{m}_{it} \triangleq \hat{\mu}_{it}(0) - \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_{it}(0) - \frac{1}{T}\sum_{t=1}^{T}\hat{\mu}_{it}(0) + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\mu}_{it}(0), \tag{3.1}$$

$$\hat{\nu}_{it} \triangleq \big(\hat{\mu}_{it}(1) - \hat{\mu}_{it}(0)\big) - \sum_{t=1}^{T}\frac{\xi_t}{n}\sum_{i=1}^{n}\big(\hat{\mu}_{it}(1) - \hat{\mu}_{it}(0)\big). \tag{3.2}$$

For notational convenience, we write $\hat{\boldsymbol{m}}_i$ for the vector $(\hat{m}_{i1}, \ldots, \hat{m}_{iT})$ and $\hat{\boldsymbol{\nu}}_i$ for the vector $(\hat{\nu}_{i1}, \ldots, \hat{\nu}_{iT})$. Given a set of estimates $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$, we define the RIPW esti-

mator as

$$\hat{\tau}(\mathbf{\Pi}) \triangleq \underset{\tau,\mu,\sum_i \alpha_i=\sum_t \lambda_t=0}{\arg\min} \sum_{i=1}^{n}\sum_{t=1}^{T}\big((Y_{it}-\hat{m}_{it}-\hat{\nu}_{it}W_{it})-\mu-\alpha_i-\lambda_t-W_{it}\tau\big)^2 \frac{\mathbf{\Pi}(\boldsymbol{W}_i)}{\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i)}. \quad (3.3)$$

The above estimator generalizes (2.11) by allowing for regression adjustment. Throughout the rest of the paper, we will abuse the notation by denoting it as $\hat{\tau}(\mathbf{\Pi})$. This two-stage formulation replaces the regression with covariates by regression on the modified outcome $(Y_{it}-\hat{m}_{it}-\hat{\nu}_{it}W_{it})$ without covariates, yielding a simplified structure, which allows us to use previously established results. In the rest of this section, we discuss formal properties they need to satisfy to guarantee consistency and asymptotic normality of $\hat{\tau}(\mathbf{\Pi})$.

In the previous section, we assumed that the researcher controlled the assignment process, which led to the restriction (2.2). In observational studies, the assignment process is unknown, and we must substitute this restriction with a different assumption. Throughout this section, we impose a high-level restriction on the relationship between unit-specific potential outcomes and assignment paths.

ASSUMPTION 3.1. (UNIT-SPECIFIC MEAN IGNORABILITY)

$$\mathbb{E}\big[\big(\boldsymbol{Y}_i(1),\boldsymbol{Y}_i(0)\big)|\boldsymbol{W}_i\big]=\mathbb{E}\big[\big(\boldsymbol{Y}_i(1),\boldsymbol{Y}_i(0)\big)\big], \quad i=1,\dots,n. \quad (3.4)$$

Recall that we do not assume that $(\boldsymbol{Y}_i(1),\boldsymbol{Y}_i(0),\boldsymbol{W}_i)$ are identically distributed across units. As a result, Assumption 3.1 imposes n separate restrictions, one for each unit. It follows the tradition of the part of the panel data literature that treats unit-specific unobservables as fixed parameters (e.g., Lancaster (2000), Hahn and Newey (2004)), rather than random variables as in (Chamberlain (1984)). It is trivially satisfied in an extreme case where $(\boldsymbol{Y}_i(1),\boldsymbol{Y}_i(0))$ has a degenerate distribution for each $i\in[n]$, which corresponds to the finite population analysis (e.g., Abadie, Athey, Imbens, and Wooldridge (2020)). In applications where $(\boldsymbol{Y}_i(1),\boldsymbol{Y}_i(0))$ is random, this assumption imposes a strict exogeneity restriction. It describes the average behavior of the outcomes conditional on the whole treatment path and does not allow the current treatment to depend on past outcomes. To illustrate this connection, consider the classical linear TWFE model where

$$Y_{it}=\mu+\alpha_i+\lambda_t+X_{it}^{\top}\beta+\tau W_{it}+\epsilon_{it}, \quad \text{where } \sum_{i=1}^{n}\alpha_i=\sum_{t=1}^{T}\lambda_t=0. \quad (3.5)$$

Assumption 3.1 is equivalent to $\mathbb{E}[\epsilon_{it}|\boldsymbol{W}_i]=0$ for $t=1,\dots,T$, which is a strict exogeneity restriction. In contrast, if $\{\epsilon_{it}: i\in[n], t\in[T]\}$ only satisfies contemporaneous restrictions $\mathbb{E}[\epsilon_{it}|W_{it}]=0$, Assumption 3.1 does not necessarily hold. We want to note that in the DiD literature, it is common to impose restrictions only on $\boldsymbol{Y}_i(0)$, while Assumption 3.1 restricts both potential outcomes. This is necessary given our focus on the ATE, defined in Section 2.2.

Assumption 3.1 is also related to the recent cross-sectional literature on quasi-experimental designs (e.g., Borusyak and Hull (2023)). A typical restriction in that literature is that while the distribution of the treatment of interest varies over units in a

complicated way, it still can be estimated and then used to construct counterfactuals. For this approach to be valid, one needs to impose a version of Assumption 3.1. In the panel data literature, this type of quasi-experimental variation was also exploited. For example, Wojtaszek and Kofoed (2022) studied the effect of military bonuses on charitable giving and found that the timing of receiving the bonus is (nearly) as-if random. Depending on the choice of outcome variable, the bonus can be viewed as a staggered or one-off treatment with a uniform generalized propensity score.

To construct estimators $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$, we use the observed covariates $\{\boldsymbol{X}_i, i \in [n]\}$. Our assumptions implicitly restrict the set of feasible covariates. In particular to respect Assumption 3.1, we do not allow any parts of the observed outcomes $\boldsymbol{Y}_i$ to be used as covariates. The situation is more delicate for $\boldsymbol{W}_i$, and we allow functions of $\boldsymbol{W}_i$ to be part of $\boldsymbol{X}_i$ as long as Assumption 2.2 holds. We elaborate on this in the next two sections.

3.1 *Assignment model estimation*

In strictly exogenous panel models, the distribution of $\boldsymbol{W}_i$ is commonly left unspecified and the analysis is based on the outcome model alone. In particular, the distribution of $\boldsymbol{W}_i$ can be degenerate for each $i \in [n]$, which is another extreme case where Assumption 3.1 trivially holds. However, researchers often informally appeal to random or quasi-random variation in $\boldsymbol{W}_i$ as a source of identification, even though they continue using outcome-based methods, such as the TWFE regression. We interpret these informal statements as statistical restrictions on $\{\boldsymbol{\pi}_i, i \in [n]\}$ that go beyond Assumption 3.1.

Precisely, because the arguments used in the applied work are often informal, we cannot offer and analyze a general methodology of how to use them to construct $\{\hat{\boldsymbol{\pi}}_i, i \in [n]\}$. Instead, we discuss several strategies that are potentially relevant for a large class of applications. Our goal is to demonstrate how to utilize the information used to construct the outcome-based estimators, and thus is readily available. In practice, researchers can have other sources of information that we do not incorporate in our analysis. After this discussion, we continue our formal analysis under high-level assumptions on $\{\hat{\boldsymbol{\pi}}_i, i \in [n]\}$.

We use $\{(\boldsymbol{W}_i, \boldsymbol{X}_i) : i \in [n]\}$ to estimate $\boldsymbol{\pi}_i$. At first glance, it might appear to be challenging to estimate the distribution of the whole vector. Nevertheless, treatment paths often have restricted support with a size much smaller than 2^T, such as staggered adoption and/or special structures that reduce the complexity of the distribution, such as the Markov structure. We present a few examples below for illustration.

In the staggered adoption designs, $\boldsymbol{W}_i$ is equivalent to an adoption time $A_i \in \{1, \ldots, T, \infty\}$, where $A_i = \infty$ for never-treated units and $A_i = t$ for units initially treated at time t. Then A_i can be viewed as an event or during outcome, and one can apply any survival or duration model, such as the Cox proportional hazard model and accelerated failure time model, to estimate its distribution which yields $\boldsymbol{\pi}_i$ by taking the difference between the consecutive points; see Section 5.2 for an empirical illustration that uses this strategy and additional discussion. For transient treatments that occur at most once during the study period, $\boldsymbol{W}_i$ can be expressed by the adoption time $A_i \in \{1, \ldots, T, \infty\}$ as above. The propensity score $\boldsymbol{\pi}_i$ can then be estimated via a discrete choice model.

For general designs where the treatment can be alternated on and off, $\boldsymbol{\pi}_i$ can be reparametrized as a sequence of conditional distributions $\mathbb{P}(W_{it}|W_{i(t-1)}, \ldots, W_{i1}, \boldsymbol{X}_i)$ and estimated by a Markov model. In particular, Arkhangelsky and Imbens (2022) show that if $\boldsymbol{X}_i$ incorporates appropriate sufficient statistics, then conditioning on $\boldsymbol{X}_i$ eliminates the ex ante present unobserved heterogeneity from the distribution of $\boldsymbol{W}_i$. Aguirregabiria, Gu, and Luo (2021) show that these assumptions are satisfied by a large class of models that are widely used in economic applications, including structural models with forward-looking agents as well as myopic (backward-looking) dynamic logit models. They also provide explicit characterizations for sufficient statistics in such models. These results can be directly applied in our setting.

Given an estimate $\hat{\boldsymbol{\pi}}_i$, we say that it estimates the assignment model well if $\hat{\boldsymbol{\pi}}_i$ is close to $\boldsymbol{\pi}_i$ in L^2 distance. Specifically, for each unit i we define the accuracy of $\hat{\boldsymbol{\pi}}_i$ as

$$\delta_{\pi i} \triangleq \sqrt{\mathbb{E}\big[\big(\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) - \boldsymbol{\pi}_i(\boldsymbol{W}_i)\big)^2\big]}. \tag{3.6}$$

Here, the expectation is taken over both $\boldsymbol{W}_i$ and $\hat{\boldsymbol{\pi}}_i$ (conditional on $\{\boldsymbol{X}_i : i \in [n]\}$). In the setting of Section 2, $\delta_{\pi i} = 0$ because $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$.

3.2 *Outcome model estimation*

In this section, we discuss the construction of the terms $\{(\hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$, which we use to build the estimator (3.3). We start with unit specific quantities $(\hat{\boldsymbol{\mu}}_i(0), \hat{\boldsymbol{\mu}}_i(1))$, which we view as estimators for $(\mathbb{E}[\boldsymbol{Y}_i(0)], \mathbb{E}[\boldsymbol{Y}_i(1)])$. There are many ways of constructing such estimators, and our results require only high-level restrictions on these objects. For example, one can consider a generalization of the linear TWFE model (3.5):

$$\mathbb{E}\big[Y_{it}(w)\big] = \mu + \alpha_i + \lambda_t + X_{it}^\top \beta + \big(\tau + X_{it}^\top \phi\big)w, \quad \text{where } \sum_{i=1}^{n} \alpha_i = \sum_{t=1}^{T} \lambda_t = 0. \tag{3.7}$$

Then $(\hat{\mu}_{it}(0), \hat{\mu}_{it}(1))$ can be chosen as

$$\hat{\mu}_{it}(w) = \hat{\mu} + \hat{\alpha}_i + \hat{\lambda}_t + X_{it}^\top \hat{\beta} + \big(\hat{\tau} + X_{it}^\top \hat{\phi}\big)w, \tag{3.8}$$

where the parameters are estimated by regressing Y_{it} on X_{it}, W_{it}, the covariate-treatment interaction $X_{it}W_{it}$, and a set of fixed effects. When we estimate $\hat{\mu}_{it}(w)$ for a new unit whose unit fixed effect is not estimated, we can simply set $\hat{\mu}_{it}(w) = \hat{\mu} + \hat{\lambda}_t + X_{it}^\top \hat{\beta} + (\hat{\tau} + X_{it}^\top \hat{\phi})w$.

In the cross-sectional case, an estimate $(\hat{\boldsymbol{\mu}}_i(0), \hat{\boldsymbol{\mu}}_i(1))$ is considered an accurate estimate of $(\mathbb{E}[\boldsymbol{Y}_i(0)], \mathbb{E}[\boldsymbol{Y}_i(1)])$ if $\{\|\hat{\boldsymbol{\mu}}_i(0) - \mathbb{E}[\boldsymbol{Y}_i(0)]\|_2 + \|\hat{\boldsymbol{\mu}}_i(1) - \mathbb{E}[\boldsymbol{Y}_i(1)]\|_2 : i \in [n]\}$ is small on average (e.g., Robins, Rotnitzky, and Zhao (1994), Kang and Schafer (2007)). Constructing such estimators for panel models with fixed effects and a finite number of periods is impossible. Thus, the standard approach of measuring accuracy does not apply in our setting, and we need to consider alternative measures.

We start by defining the estimands (m_{it}, ν_{it}) that $(\hat{m}_{it}, \hat{\nu}_{it})$ attempt to estimate:

$$m_{it} = \mathbb{E}\big[Y_{it}(0)\big] - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big[Y_{it}(0)\big] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[Y_{it}(0)\big] + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\big[Y_{it}(0)\big] \tag{3.9}$$

$$\nu_{it} \triangleq \tau_{it} - \tau^*(\xi). \tag{3.10}$$

To measure the degree of mispecification of the outcome model, we introduce the following quantity:

$$\delta_{yi} \triangleq \sqrt{\mathbb{E}\big[\|\hat{\boldsymbol{m}}_i - \boldsymbol{m}_i\|_2^2\big] + \mathbb{E}\big[\|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i\|_2^2\big]}. \tag{3.11}$$

The first term captures the estimation accuracy of $\boldsymbol{m}_i$, and the second term captures the estimation accuracy of $\boldsymbol{\tau}_i$. By definition, δ_{yi} is invariant if we replace $\mathbb{E}[Y_{it}(0)]$ by $\mathbb{E}[Y_{it}(0)] + \mu' + \alpha_i' + \lambda_t'$ and τ_{it} by $\tau_{it} + \tau'$ for any μ', τ', $\{\alpha_i' : i \in [n]\}$, and $\{\lambda_t' : t \in [T]\}$. Thus, requiring δ_{yi} to be small is strictly less stringent than requiring the standard measure of outcome model accuracy for cross-sectional data to be small.

In the simplest TWFE model (3.5) without covariates, $\delta_{yi} = 0$ if we choose $\hat{\mu}_{it}(0) = \hat{\mu}_{it}(1) = 0$. For the more general TWFE model (3.7), regardless whether unit i is used for fitting the TWFE regression,

$$\delta_{yi} = \sqrt{\mathbb{E}\left[\sum_{t=1}^{T}\big\{(X_{it} - \bar{X}_{i\cdot} - \bar{X}_{\cdot t} + \bar{X}_{\cdot\cdot})^\top(\hat{\beta} - \beta)\big\}^2 + \left\{\left(X_{it} - \sum_{t=1}^{T}\xi_t \bar{X}_{\cdot t}\right)^\top(\hat{\phi} - \phi)\right\}^2\right]},$$

Standard assumptions (e.g., Arellano (2003), Wooldridge (2010)) guarantee that $(\hat{\beta}, \hat{\phi})$ are consistent for (β, ϕ) even with a finite number of periods. We can further generalize the model by replacing $X_{it}^\top\beta$ and $X_{it}^\top\phi$ with nonlinear functions $g(X_{it})$ and $\tau(X_{it})$ and estimate them by nonparametric TWFE regressions (Boneva, Linton, and Vogt (2015)).

The requirement that $\delta_{yi} \approx 0$, at least on average, puts restrictions on the treatment effects. These requirements, however, can be redundant, depending on the structure of $\boldsymbol{X}_i$. For example, Wooldridge (2021) shows that the problems with heterogeneous treatment effects can be solved, under conditional parallel trends and linearity, by including a sufficiently rich set of controls, which includes functions of $\boldsymbol{W}_i$. In the staggered adoption case, one needs to include interactions with all the adoption dates. Unfortunately, including such interactions into $\boldsymbol{X}_i$ violates the overlap Assumption 2.2.

3.3 *Consistency of RIPW estimators*

In this and the next subsection, we consider a simplified case where the estimates $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$ are independent of the data (e.g., obtained from external data). While this is not always possible in practice, the theory of consistency and asymptotic normality can be stated without much mathematical complication. Moreover, these results are building blocks for the theory of cross-fitting estimator described at length in Supplemental Appendix B. To ease implementation, we provide a self-contained description of the (derandomized) cross-fitting RIPW estimator in Algorithm 1 at the end of the next subsection.

ASSUMPTION 3.2. *There exists $c > 0$ such that, for the same $\mathbb{S}^*$ defined in Assumption 2.2,*

$$\hat{\boldsymbol{\pi}}_i(\boldsymbol{w}) \geq c, \quad \forall \boldsymbol{w} \in \mathbb{S}^*, i \in [n], \text{ almost surely.}$$

ASSUMPTION 3.3. *There exists $M < \infty$ such that $\max_{i,t,w} \mathbb{E}[\hat{m}_{it}^2 + \hat{\nu}_{it}^2] \leq M$.*

Theorem 2.1 implies that the RIPW estimator with $\mathbf{\Pi}$ being a solution of the DATE equation, if any, is a consistent estimator of DATE without any outcome model when $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$ is known. On the other hand, when the outcome model is correctly specified, $Y_{it} - \hat{m}_{it} - \hat{\nu}_{it} W_{it} \approx Y_{it} - m_{it} - (\tau_{it} - \tau^*(\xi)) W_{it}$ is a linear model with two-way fixed effects and a single predictor W_{it} and $\hat{\tau}$ is approximately a weighted least squares estimator, which is consistent under mild conditions on the weights (e.g., Wooldridge (2010)). This shows a weak double robustness property that $\hat{\tau}(\mathbf{\Pi})$ is consistent if either the outcome model or the assignment model is exactly correct.

For cross-sectional data, the augmented IPW estimator enjoys a strong double robustness property, which states that the asymptotic bias is the product of estimation errors of the outcome and assignment models (e.g., Robins, Rotnitzky, and Zhao (1994), Kang and Schafer (2007), Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, and Newey (2017), Chernozhukov et al. (2018)). Clearly, this implies the weak double robustness. It further implies the estimator has higher asymptotic precision than estimators based on merely the outcome or assignment modeling when both models are estimated well. The next result provides a sufficient condition for strong double robustness of $\hat{\tau}(\mathbf{\Pi})$ when the estimated treatment and outcome models are independent of the data.

THEOREM 3.1. *Assume that $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$ are independent of the data. Under Assumptions 2.1–2.3 and 3.1–3.3, conditional on the estimates,*

$$\hat{\tau}(\mathbf{\Pi}) = \tau^*(\xi) + O_{\mathbb{P}}(\bar{\delta}_\pi \bar{\delta}_y), \quad \text{where } \bar{\delta}_\pi = \sqrt{\frac{1}{n}\sum_{i=1}^n \delta_{\pi i}^2}, \ \bar{\delta}_y = \sqrt{\frac{1}{n}\sum_{i=1}^n \delta_{yi}^2}.$$

In particular, $\hat{\tau}(\mathbf{\Pi})$ is a consistent estimator of $\tau^*(\xi)$ if $\bar{\delta}_\pi \bar{\delta}_y = o(1)$.

Assumptions 2.3 and 3.3 guarantee that $\bar{\delta}_y$ is bounded. Thus, the RIPW estimator is consistent whenever $\boldsymbol{\pi}_i$ is consistently estimated without any requirement on the rate of convergence. On the other hand, under the TWFE model (3.7) or nonparametric TWFE models discussed in the last subsection, $\bar{\delta}_y = o(1)$ and the estimator is consistent even if the assignment model is globally misspecified.

3.4 *Inference with independent model estimates*

Similar to Theorem 2.2, we can derive an asymptotic linear expansion for $\mathcal{D} \cdot \sqrt{n}(\hat{\tau}(\mathbf{\Pi}) - \tau^*(\xi))$.

THEOREM 3.2. *Assume that* $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$ *are independent of the data. Let* $\Gamma_\theta, \Gamma_{ww}, \boldsymbol{\Gamma}_w$, *and* $\mathcal{D}$ *be defined as in Theorem 2.2 with* $\hat{\boldsymbol{\pi}}_i$ *in the definition of* Θ_i. *Redefine* $\Gamma_{wy}, \boldsymbol{\Gamma}_y$, *and* $\mathcal{V}_i$ *by replacing* $(\boldsymbol{Y}_i(0), \boldsymbol{Y}_i(1))$ *with* $(\tilde{\boldsymbol{Y}}_i(0), \tilde{\boldsymbol{Y}}_i(1)) = (\boldsymbol{Y}_i(0) - \hat{\boldsymbol{m}}_i, \boldsymbol{Y}_i(1) - \hat{\boldsymbol{m}}_i - \hat{\boldsymbol{\nu}}_i)$. *Under Assumptions 2.1–2.3 and 3.1–3.3,*

$$\mathcal{D} \cdot \sqrt{n}\big(\hat{\tau}(\mathbf{\Pi}) - \tau^*(\xi)\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\mathcal{V}_i - \mathbb{E}[\mathcal{V}_i]\big) + O_{\mathbb{P}}\big(n^{1/2-2q} + \bar{\delta}_\pi \bar{\delta}_y\big). \tag{3.12}$$

In particular, the last term is $o_{\mathbb{P}}(1/\sqrt{n})$ *if* $q > 1/2$ *and* $\bar{\delta}_\pi \bar{\delta}_y = o(1/\sqrt{n})$.

Similar to Section 2, we can estimate $\hat{\mathcal{V}}_i$ and the asymptotic variance via (2.22) and construct the Wald-type confidence interval as (2.23) when units are independent. This is a special case of Theorem A.6 in Supplemental Appendix A.4 for general dependent designs.

THEOREM 3.3. *Assume that* $\{(\boldsymbol{Y}_i(1), \boldsymbol{Y}_i(0), \boldsymbol{W}_i) : i \in [n]\}$ *are independent. Under the same settings as in Theorem 3.2,*

$$\liminf_{n\to\infty} \mathbb{P}\big(\tau^*(\xi) \in \hat{C}_{1-\alpha}\big) \geq 1 - \alpha,$$

if, further, (2.24) *holds.*

Under Assumption 3.1, Theorem 3.2 and Theorem 3.3 strictly generalize Theorem 2.2 and Theorem 2.3—when $\boldsymbol{\pi}_i$ is known, $\bar{\delta}_\pi = 0$, and hence $\bar{\delta}_\pi \bar{\delta}_y = 0 = o(1/\sqrt{n})$ regardless of the accuracy of the outcome model estimates. When $\boldsymbol{\pi}_i$ is unknown, $\bar{\delta}_\pi$ and $\bar{\delta}_y$ are typically no less than $O(1/\sqrt{n})$ without external data. As a result, both models should be consistently estimated to achieve $\bar{\delta}_\pi \bar{\delta}_y = o(1/\sqrt{n})$ though the estimates can have a slower convergence rate than $O(1/\sqrt{n})$. For example, it would be satisfied if $\bar{\delta}_\pi, \bar{\delta}_y = o(n^{-1/4})$. We emphasize that this rate requirement is standard for inference with cross-sectional data (Chernozhukov et al. (2017, 2018)). Under this rate condition, by virtue of the asymptotic linear expansion in Theorem 3.2, the researcher can safely ignore the variability of the model estimates and use them in the variance calculation as if they are the truth.

Even when condition $\bar{\delta}_\pi \bar{\delta}_y = o(1/\sqrt{n})$ is violated, the asymptotically valid inference is still possible at the cost of more involved variance estimation. Doubly robust inference in this regime is generally hard (e.g., Benkeser, Carone, Vanm Der Laan, and Gilbert (2017)). We consider the setting where parametric models are used to fit the generalized propensity score and regression adjustment. This setting has been studied in the literature for cross-sectional data (e.g., Cao, Tsiatis, and Davidian (2009)). Our formal results are deferred in Supplemental Appendix A.5 due to the mathematical complication. Roughly speaking, if the estimators $\{(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i) : i \in [n]\}$ come from a smooth parametric model, then one can use their asymptotic expansion (around their limits, which do not necessarily correspond to the true parameters) to compute the asymptotic variance. Similar to the case discussed Section 2 and this section, we can obtain an asymptotically conservative variance estimator without knowing which model is misspecified apriori.

Algorithm 1 RIPW estimator with derandomized cross-fitting.

Input: data $\{(\boldsymbol{X}_i, \boldsymbol{W}_i, \boldsymbol{Y}_i) : i \in [n]\}$, number of folds K, number of data splits B, reshaped distribution $\boldsymbol{\Pi}$

procedure

for $b = 1, \ldots, B$ **do**

Randomly split $[n]$ into K folds $\mathcal{I}_1, \ldots, \mathcal{I}_K$ with $|\mathcal{I}_j| \in \{\lfloor n/K \rfloor, \lceil n/K \rceil\}$

for $k = 1, \ldots, K$ **do**

Fit the assignment model $\hat{\boldsymbol{\pi}}(\boldsymbol{w}; \boldsymbol{x})$ using data in $\bigcup_{j \neq k} \mathcal{I}_j$

Fit the outcome model $(\hat{\boldsymbol{m}}_t(\boldsymbol{x}), \hat{\boldsymbol{\nu}}_t(\boldsymbol{x}))$ using data in $\bigcup_{j \neq k} \mathcal{I}_j$

for $i \in \mathcal{I}_k$ **do**

$\hat{\boldsymbol{\pi}}_i(\boldsymbol{w}) \leftarrow \hat{\boldsymbol{\pi}}(\boldsymbol{w}; \boldsymbol{X}_i)$

$(\hat{\boldsymbol{m}}_{it}, \hat{\boldsymbol{\nu}}_{it}) \leftarrow (\hat{\boldsymbol{m}}_t(\boldsymbol{X}_i), \hat{\boldsymbol{\nu}}_t(\boldsymbol{X}_i))$ for each $t \in [T]$

Compute $\hat{\tau}^{(b)}(\boldsymbol{\Pi})$ via (3.3)

Compute $\mathcal{D}^{(b)}$ defined in Theorem 3.2

Compute $\{\hat{\mathcal{V}}_i^{(b)} : i \in [n]\}$ based on (2.21)

$\hat{\tau}(\boldsymbol{\Pi}) \leftarrow \sum_{b=1}^B \mathcal{D}^{(b)} \hat{\tau}^{(b)}(\boldsymbol{\Pi}) / \sum_{b=1}^B \mathcal{D}^{(b)}$

$\bar{\hat{\mathcal{V}}}_i \leftarrow \sum_{b=1}^B \hat{\mathcal{V}}_i^{(b)} / \sum_{b=1}^B \mathcal{D}^{(b)}$ for each $i \in [n]$

$\hat{\sigma}^2 \leftarrow$ sample variance of $\{\bar{\hat{\mathcal{V}}}_i : i \in [n]\}$

$\hat{C}_{1-\alpha} \leftarrow [\hat{\tau}(\boldsymbol{\Pi}) - z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}, \hat{\tau}(\boldsymbol{\Pi}) + z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}]$,

Output: the derandomized cross-fitting estimator $\hat{\tau}(\boldsymbol{\Pi})$ and confidence interval $\hat{C}_{1-\alpha}$

In practice, it is uncommon to obtain estimates of $(\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i)$ that are independent of the data, except in the design-based inference where $\hat{\boldsymbol{\pi}}_i = \boldsymbol{\pi}_i$ and $\hat{\boldsymbol{m}}_i = \hat{\boldsymbol{\nu}}_i = \mathbf{0}_T$, or when external data is available. Usually, these parameters need to be estimated from the data. The resulting dependence invalidates the assumptions of Theorem 3.2 and 3.3. However, as we show in Supplemental Appendix B similar results hold if we use a particular version of cross-fitting. Note that this implies that $\hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{m}}_i, \hat{\boldsymbol{\nu}}_i$ cannot contain unit-specific fixed effects. Moreover, we propose a simple approach to mitigate the randomness introduced by sample splitting. We describe the estimator in Algorithm 1. More details can be found in Supplemental Appendix B. We implemented this method in an `R` package `ripw` that is available at https://github.com/lihualei71/ripw.

4. Design-robust event study specifications

A key limitation of our analysis in previous sections is the focus on static models. This is important both theoretically and practically. Theoretically, some policies of interest are transient in nature, for example, a large infrastructure investment, but policymakers expect them to have a lasting impact, which requires a dynamic model. Practically, a large part of applied work in economics uses regression models that explicitly incorporate lags of treatment variables.

We consider a relatively simple class of linear potential outcome modes to address these concerns. For every i and t, we specify the potential outcomes as a function of the current treatment w and its p lags:

$$\mathbb{E}\big[Y_{it}(w_0, w_{-1}, \ldots, w_{-p})\big] = \mu_{it} + \sum_{l=0}^{p} \tau_{i,-l} w_{-l}. \tag{4.1}$$

As in Section 3, the expectation is conditional on covariates, and we do not require the units to be independent or identically distributed. This model does not restrict the baseline outcomes but puts structure on the dynamic effects of the treatment. First, the effect of the treatment is present only for p periods after it is implemented. Second, the effect is linear, that is, the causal effect of being treated one period ago, w_{-1}, does not depend on whether the unit was treated two periods ago w_{-2}. Finally, the effects are homogenous over time, meaning that $\tau_{i,l}$ do not depend on calendar time t. These restrictions are important: the first eliminates the possibility of long-term effects, while the other two eliminate state dependence. Still, we think this model is flexible enough to be useful for a large class of empirical applications.

Interestingly, if the treatment timing is fixed and common across units, and p is large enough, then (4.1) is a parametrization of all realizable potential outcomes, and thus does not impose any testable restrictions. To see this, let $q+1$ denote the adoption time and set $p = T - q + 1$. Then each unit i has $T + (T - q)$ potential outcomes $\{Y_{it}(\mathbf{0}_T) : t \in [T]\}$ and $\{Y_{it}(\mathbf{0}_q, \mathbf{1}_{T-q}) : t \in \{q+1, \ldots, T\}\}$. It is easy to see that (4.1) holds with $\mu_{it} = \mathbb{E}[Y_{it}(\mathbf{0}_T)]$, $\tau_{i,0} = \mathbb{E}[Y_{i(q+1)}(\mathbf{0}_q, \mathbf{1}_{T-q})] - \mathbb{E}[Y_{i(q+1)}(\mathbf{0}_T)]$, and

$$\begin{aligned}\tau_{i,-\ell} = {} & \mathbb{E}\big[Y_{i(q+\ell+1)}(\mathbf{0}_q, \mathbf{1}_{T-q})\big] - \mathbb{E}\big[Y_{i(q+\ell+1)}(\mathbf{0}_T)\big] \\ & - \big(\mathbb{E}\big[Y_{i(q+\ell)}(\mathbf{0}_q, \mathbf{1}_{T-q})\big] - \mathbb{E}\big[Y_{i(q+\ell)}(\mathbf{0}_T)\big]\big), \quad \ell \in [p].\end{aligned}$$

Similar logic extends to staggered adoption designs as long as we treat the assignment as fixed. However, it breaks if we assume that the adoption time is randomly assigned. In this case, we can test the static model from Section 2 and the dynamic model (4.1) by comparing outcomes across units that were previously treated at different periods. This emphasizes the importance of the assignment model for the analysis of dynamic effects.

In this case, it is natural to consider the RIPW estimator coupled with an event-study regression model, that is,

$$\begin{aligned}&(\hat{\tau}_0, \hat{\tau}_{-1}, \ldots, \hat{\tau}_{-p}) \\ &\quad = \underset{\tau_0, \tau_{-1}, \ldots, \tau_{-p}, \mu, \sum_i \alpha_i = \sum_t \lambda_t = 0}{\arg\min} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \mu - \alpha_i - \lambda_t - W_{it}\tau_0 - \sum_{j=1}^{p} W_{i(t-j)}\tau_{-j} \right)^2 \\ &\qquad \times \frac{\boldsymbol{\Pi}(\boldsymbol{W}_i)}{\boldsymbol{\pi}_i(\boldsymbol{W}_i)}\end{aligned} \tag{4.2}$$

where W_{it} is defined as 0 whenever $t \leq 0$. Our next result describes the probability limit of $(\hat{\tau}_0, \hat{\tau}_{-1}, \ldots, \hat{\tau}_{-p})$. The proof is presented in Supplemental Appendix A.8.

THEOREM 4.1. *Assume that* $Y_{it}(w_0, w_{-1}, \ldots, w_{-p})$ *satisfies* (4.1) *and the generalized propensity score* $\boldsymbol{\pi}_i(\boldsymbol{w}) \triangleq \mathbb{P}(\boldsymbol{W}_i = \boldsymbol{w} | \{Y_{it}(\tilde{\boldsymbol{w}}) : t \in [T], \tilde{\boldsymbol{w}} \in \{0,1\}^T\})$ *is known. Further assume that* $\mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[(\boldsymbol{W}_{ex} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{ex}])J(\boldsymbol{W}_{ex} - \mathbb{E}_{\boldsymbol{W}\sim\boldsymbol{\Pi}}[\boldsymbol{W}_{ex}])]$ *is positive definite, where*

$$\boldsymbol{W}_{ex} = (\boldsymbol{W}, \boldsymbol{W}_{-1}, \ldots, \boldsymbol{W}_{-p}) \in \{0,1\}^{T\times(p+1)}, \quad \boldsymbol{W}_{-k} = (0, \ldots, 0, W_1, \ldots, W_{T-k})^\top,$$

and $\boldsymbol{W} = (W_1, \ldots, W_T)$ *denote a generic random vector drawn from the distribution* $\boldsymbol{\Pi}$. *Then, under Assumptions* 2.1–2.3 (*with* $Y_{it}(w)$ *replaced by* $Y_{it}(w_0, w_{-1}, \ldots, w_{-p})$),

$$\hat{\tau}_{-k} = \frac{1}{n}\sum_{i=1}^{n} \tau_{i,-k} + o_{\mathbb{P}}(1), \quad k = 0, 1, \ldots, p.$$

This result justifies using the RIPW estimator in a large class of applications. If $\boldsymbol{\pi}_i$-s are unknown, then one can estimate them using one of the strategies discussed in the previous section. Similarly, one can introduce covariates in this model in the same way as before. Also, applied researchers often consider leads in addition to lags in their regressions, especially in the context of staggered adoption designs. To incorporate this practice into our framework, one simply needs to shift the treatment path $\boldsymbol{W}_i$ appropriately. The resulting estimators for the leads can then be used to test for the validity of the underlying model.

We do not establish analogs of Theorems 3.1–3.3 for this estimator, but we expect them to hold under appropriate technical conditions. In particular, under (4.1), if the TWFE model holds for the baseline potential outcomes such that $\mu_{it} = \mu + \alpha_i + \lambda_t$ and $\tau_{i,-\ell} = \tau_{-\ell}$, then (4.2) is consistent for $(\tau_0, \tau_{-1}, \ldots, \tau_{-p})$ since it is a weighted least squares estimator for a correctly specified linear model. Compared to our analysis in previous sections, the reshaping distribution $\boldsymbol{\Pi}$ does not play a major role in these results. The reason for this behavior is that the model for treatment effects is time-homogeneous. If we relax this assumption and allow for time-varying dynamic effects $\tau_{i,-l,t}$, then the distribution $\boldsymbol{\Pi}$ becomes important again. The corresponding DATE equation for this problem is more complicated than the one presented in Section 2, and its analysis is beyond the scope of this paper.

5. NUMERICAL STUDIES

In this section, we investigate the properties of our estimator in simulations and show how to apply it to real data sets. The R programs to replicate all results in this section is available at https://github.com/xiaomanluo/ripwPaper.

5.1 *Synthetic data*

To highlight the central role of the reshaping function in eliminating the bias, we focus on inference with known assignment mechanisms. Put another way, in such settings, the bias of the unweighted or IPW estimators is purely driven by the wrong reshaping function rather than other sources of variability. We consider the DATE with $\xi = \mathbf{1}_T/T$ for simplicity. We also design a simulation study with unknown assignment mechanisms

and present the results in Supplemental Appendix D, which involves all 2-by-2 settings with correct/incorrect assignment/outcome model and a detailed comparison between the RIPW estimator and several other competing estimators.

We consider a short panel with $T = 4$ and sample size $n = 1000$. We generate a single time-invariant covariate $X_{it} = X_i$ with $P(X_i = 1) = 0.7$ and $P(X_i = 2) = 0.3$ and a single time-invariant unobserved confounder $U_{it} = U_i$ with $U_i \sim \text{Unif}(\{1, \ldots, 10\})$. Within each experiment, the covariates and unobserved confounders are only generated once and then fixed to ensure a fixed design. For treatment assignments, we consider a staggered adoption design, that is, $\boldsymbol{W}_i \in \mathcal{W}^{\text{sta}}$. We assume that $\boldsymbol{W}_i$ is less likely to be treated when $X_i = 1$. In particular,

$$\begin{aligned}&\big(\boldsymbol{\pi}_i(\boldsymbol{w}_{(0)}), \boldsymbol{\pi}_i(\boldsymbol{w}_{(1)}), \boldsymbol{\pi}_i(\boldsymbol{w}_{(2)}), \boldsymbol{\pi}_i(\boldsymbol{w}_{(3)}), \boldsymbol{\pi}_i(\boldsymbol{w}_{(4)})\big)\\ &= \begin{cases}(0.8, 0.05, 0.05, 0.05, 0.05) & (X_i = 1),\\ (0.1, 0.1, 0.2, 0.3, 0.3) & (X_i = 2).\end{cases}\end{aligned}$$

The potential outcome $Y_{it}(0)$ and the treatment effect τ_{it} are generated as follows:

$$Y_{it}(0) = \mu + \alpha_i + \lambda_t + m_{it} + \epsilon_{it}, \qquad m_{it} = \sigma_m X_i \beta_t, \qquad \tau_{it} = \sigma_\tau a_i b_t,$$

where $\mu = 0$, $\beta_t = t - 1$, $\alpha_i = 0.5U_i$, $\lambda_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, $b_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$, and $\epsilon_{it} \overset{i.i.d.}{\sim} N(0, 1)$. For a_i, we consider two settings: we either set $a_i = 1$, thus making τ_{it} unit-invariant; or $a_i \overset{i.i.d.}{\sim} \text{Unif}([0, 1])$, in which case τ_{it} varies over units and periods. As with the covariates X_i, the time fixed effects λ_t and factors a_i, b_t are generated once for each setting and then fixed over runs. In contrast, ϵ_{it} will be resampled in every run as the stochastic errors. Note that both m_{it} and τ_{it} are generated from rank-one factor models.

The parameters σ_m and σ_τ measure two types of deviations from the TWFE model: σ_m measures the violation of parallel trend because we will not adjust for X_i in the design-based inference, and σ_τ measures the violation of constant treatment effects. We consider two settings: we either set $\sigma_m = 1$, $\sigma_\tau = 0$—a model without parallel trends, but constant treatment effects; alternatively, we set $\sigma_m = 0$, $\sigma_\tau = 1$—a TWFE model with heterogeneous effects, but parallel trends. In the first setting, $\tau_{it} = 0$ regardless of the model for a_i, thus we have 3 different scenarios in total.

We consider three estimators: the unweighted TWFE estimator, the IPW estimator, and the RIPW estimator with $\boldsymbol{\Pi}$ given by (C.5). For each of the three experiments, we resample W_{it}'s and ϵ_{it}'s, while keeping other quantities fixed, for 1000 times and collect the estimates and the confidence intervals. Figure 3 presents the boxplots of the bias $\hat{\tau}(\boldsymbol{\Pi}) - \tau^*(\xi)$. In all settings, the unweighted estimator is clearly biased, demonstrating that both the parallel trend and treatment effect homogeneity are indispensible for classical TWFE regression. In contrast, the IPW estimator is biased when the treatment effects are heterogeneous, but unbiased otherwise even if the parallel trend assumption is violated. This is by no means a coincidence; in this case, $\boldsymbol{\tau}_i = \tau^*(\xi)\mathbf{1}_T$ for all i and, by Theorem 2.1, the asymptotic bias $\Delta_\tau(\xi) = 0$ for RIPW estimators with any reshaped function including the IPW estimator. Finally, as implied by our theory, the RIPW estimator is unbiased in all settings. Moreover, the coverage of confidence intervals for the RIPW

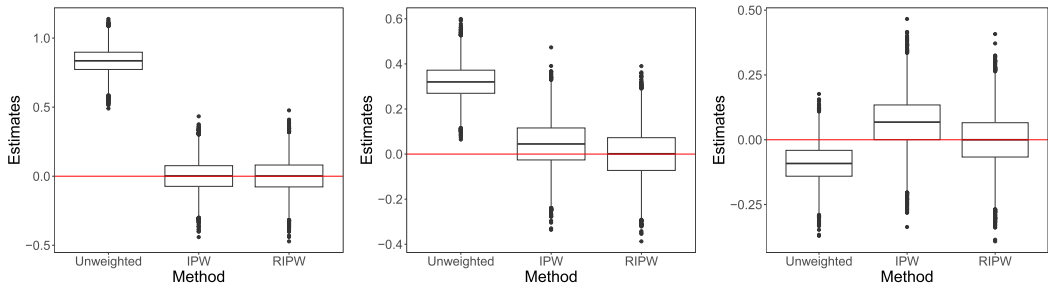


FIGURE 3. Boxplots of bias across 10,000 replicates for the unweighted, IPW, and RIPW estimators under (left) violation of parallel trend ($\sigma_m = 1$, $\sigma_\tau = 0$), (middle) heterogeneous treatment effect with limited heterogeneity ($\sigma_m = 0$, $\sigma_\tau = 1$, $a_i = 1$), and (right) heterogeneous treatment effect with full heterogeneity ($\sigma_m = 0$, $\sigma_\tau = 1$, $a_i \sim \text{Unif}([0, 1])$).

estimator is 94.6%, 95.2%, and 94.6% in these three settings, respectively, confirming the inferential validity stated in Theorem 2.3.

5.2 *Analysis of OpenTable data in the early COVID-19 pandemic*

On February 29, 2020, Washington declared a state of emergency in response to the COVID-19 pandemic. A state of emergency is a situation in which a government is empowered to perform actions or impose policies that it would normally not be permitted to undertake. It alerts citizens to change their behaviors and urges government agencies to implement emergency plans. As the pandemic has swept across the country, more states declared a state of emergency in response to the COVID-19 outbreak.

The state of emergency restricts various human activities. It would be valuable for governments and policymakers to get a sense of the short-term effect of this urgent action. Since mid-February 2020, OpenTable has been releasing daily data of year-over-year seated diners for a sample of restaurants on the OpenTable network through online reservations, phone reservations, and walk-ins. This provides an opportunity to study how the state of emergency affects the restaurant industry in a short time. The data covers 36 states in the United States, which we will focus our analysis on. Policy evaluation in the pandemic is extremely challenging due to the complex confounding and endogeneity issues (e.g., Chetty, Friedman, Hendren, and Stepner (2020), Chinazzi et al. (2020), Goodman-Bacon and Marcus (2020), Holtz et al. (2020), Kraemer et al. (2020), Abouk and Heydari (2021)). Fortunately, compared to the policies later in the pandemic, the state of emergency suffered from less confounding since it was the first policy that affected the vast majority of the public in the US. On the other hand, the restaurant industry is responding to the policy swiftly because the restaurants are forced to limit and change operations, thereby eliminating some confounders that cannot take effect in a few days.

Despite being more approachable, the problem remains challenging due to the effect heterogeneity and the difficulty of building a reliable model for the dine-in rates in a short time window. In contrast, the declaration time of the state of emergency is

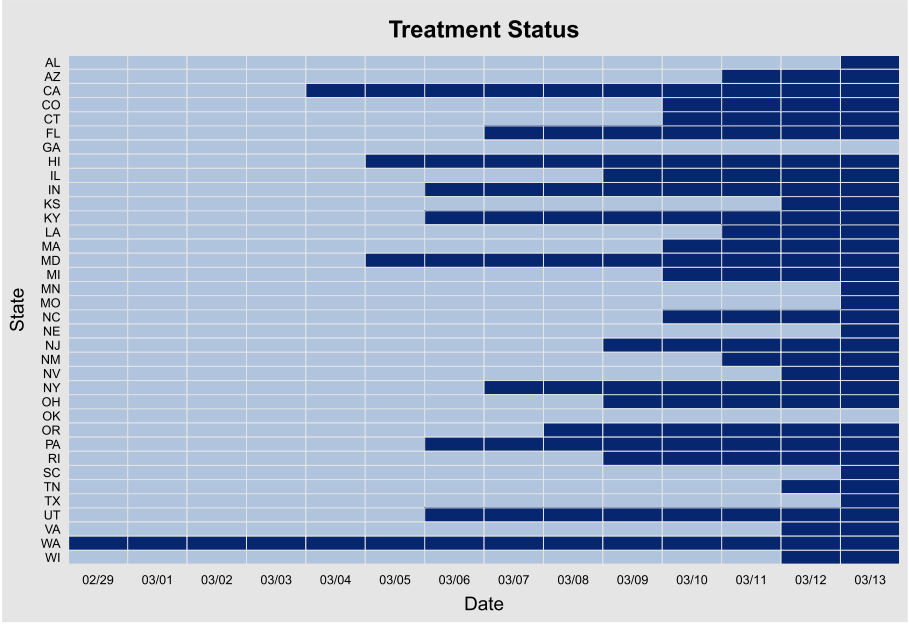


FIGURE 4. Treatment paths of each state. The darker color marks the treated days.

arguably less complex to model because it is mainly driven by the progress of the pandemic and the authority's attitude towards the pandemic.

We demonstrate our RIPW estimator on this data, which can be accessed through our `R` package `ripw`. The summary statistics and detailed descriptions of data sources (OpenTable, Perper, Cranley, and Al-Arshani, Dong, Du, and Gardner (2020), MIT Election Data and Science Lab (2018), Zemel, Eldridge, Bracco, King, and Siemer) can be found in Supplemental Appendix E. The outcome variable is the daily state-level year-over-year percentage change in seated diners provided by OpenTable. The treatment variable is the indicator of whether the state of emergency has been declared. We also include the state-level accumulated confirmed cases to measure the progress of the pandemic, the vote share of Democrats based on the 2016 presidential election data to measure the political attitude toward COVID-19, and the number of hospital beds as a proxy for the amount of regular medical resources. For demonstration purposes, we restrict the analysis to February 29–March 13, the first 14 days since the first declaration by Washington. As of March 13, 34 out of 36 states have declared a state of emergency; thus, the declaration times are right-censored. The treatment paths are plotted in Figure 4.

For the treatment model, we fit a Cox proportional hazard model on the declaration date to derive an estimate of the generalized propensity scores. Specifically, letting T_i be declaration time of state i, a Cox proportional hazard model with time-varying covariates X_{it} assumes that

$$h_i(t|X_{it}) = h_0(t)\exp\{X_{it}^\top \beta\},$$

where $h_i(t|\cdot)$ denotes the hazard function for state i, and $h_0(t)$ denotes a nonparametric baseline hazard function. The estimates $\hat{h}_0$ and $\hat{\beta}$ yield an estimate $\hat{F}_i(t)$ of the survival

TABLE 1. Parameter estimates and standard errors of parameter estimates (in parentheses) for the Cox regression with and without region fixed effects.

	w/o Region FE	w/ Region FE
log(confirmed cases)	0.225	0.166
	(0.257)	(0.245)
vote share	0.071	0.050
	(0.029)	(0.036)
log(beds)	−0.162	0.193
	(0.282)	(0.342)
region (South)		−0.884
		(0.810)
region (North Central)		−0.389
		(0.731)
region (West)		0.396
		(0.613)

function $\mathbb{P}(T_i \geq t)$ for state i, differencing which yields an estimate of the generalized propensity score[5]

$$\hat{\boldsymbol{\pi}}_i(\boldsymbol{W}_i) = \begin{cases} \hat{F}_i(T_i) - \hat{F}_i(T_i + 1) \\ \quad \text{(State } i \text{ declared state of emergency no later than 03/13)}, \\ 1 - \hat{F}_i(03/13) \\ \quad \text{(otherwise)}. \end{cases}$$

Here, we include as the time-varying covariates the logarithms of the accumulated confirmed cases and as the time-invariant covariates the logarithms of the number of hospital beds and the vote share. Note that fixed effects cannot be added into the Cox model because each state has only one outcome. To address unobserved heterogeneity, we include region fixed effects (Northeast, North Central, South, and West). While we will cross-fit the Cox model for the RIPW estimator, we fit the model on the entire data to illustrate the effect of covariates on the adoption time. Table 1 summarizes the exponentiated parameter estimates along with their standard errors with and without region fixed effects. It also reports the p-value of the joint significance test for the null hypothesis that all coefficients are zero. While most of the coefficients are not significant individually, they are jointly significant, suggesting that the generalized propensity score is nonconstant.

The proportional hazard assumption imposed by the Cox model is often controversial. Here, we apply the standard statistical tests based on Schoenfeld residuals (Schoenfeld (1980)) as a specification test for the Cox model. Figure 5 presents the p-values yielded by Schoenfeld's test. Clearly, none of them show evidence against the proportional hazard assumption. The p-value of Schoenfeld's test is 0.311, suggesting no evidence against the specification.

[5]For discrete event times, an alternative is the discrete Cox model introduced in Section 6 of Cox (1972). Here, we stick with the standard Cox model for simplicity.

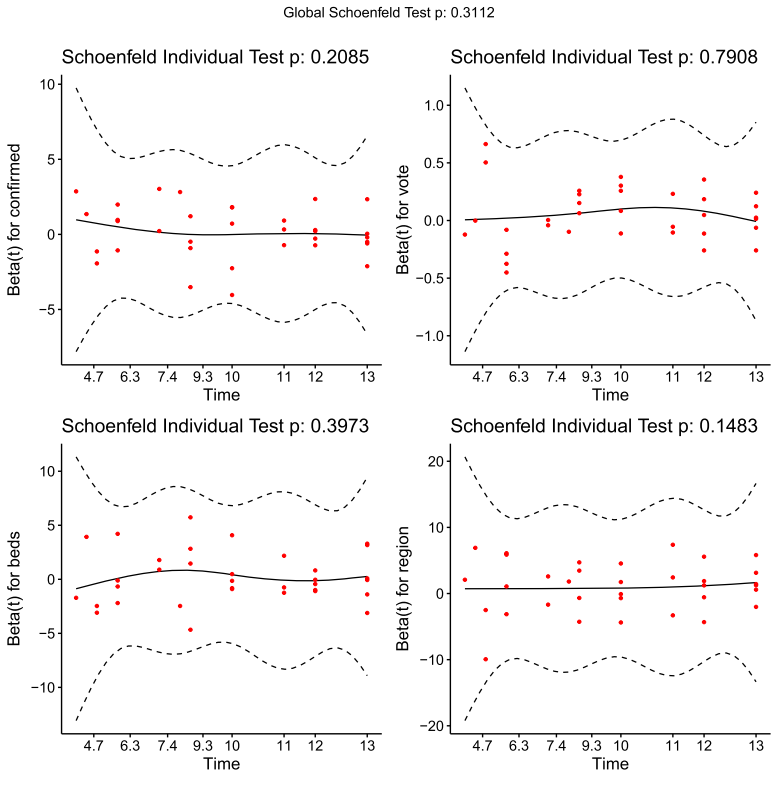


FIGURE 5. Diagnostics for the Cox proportional hazard model on adoption times.

For the outcome model, we fit an interacted TWFE regression in the form of (3.7) with the same set of covariates. Since unit fixed effects are included, no time-invariant covariate can be added to the main effects due to perfect collinearity. Thus, we add log confirmed cases, treatment, and the interactions between treatment and all variables, including region fixed effects, into the TWFE regression. Table 2 summarizes the results. The first row gives the treatment effect estimates by the TWFE regressions, though these estimates are irrelevant in the regression adjustment for our RIPW estimator, which only depends on the other rows. Table 1 reports the p-value of the joint significance test for the null hypothesis that all coefficients other than the two-way fixed effects are zero. Again, the null hypothesis that $m_{it} = 0$ for all (i, t) is rejected in both settings.

Finally, we compute the RIPW estimator for equally-weighted DATE with the reshaped distribution (C.5) in Supplemental Appendix C for staggered adoption and 10-fold cross-fitting that is described in Algorithm 1 and discussed at length in Supplemental Appendix B. Since this problem has a small sample size, the estimate exhibits large variation across different data splits. We thus apply the derandomization procedure discussed in Supplemental Appendix B.2 with 10,000 splits (i.e., $B = 10,000$ in Algorithm 1). Our derandomized cross-fitted RIPW estimate is reported in Table 3, together with the estimates obtained using the TWFE regressions reported in Table 2. It is significant at the 10% level and the magnitude is larger than that given by the unweighted

TABLE 2. Parameter estimates and standard errors (in parentheses) for the unweighted TWFE regression with and without region fixed effects.

	w/o Region FE × Treat	w/ Region FE × Treat
treat	−0.641	−0.619
	(1.640)	(1.640)
log(confirmed cases)	−3.022	−2.896
	(1.230)	(1.232)
log(confirmed cases) × treat	0.580	0.662
	(2.466)	(2.491)
vote share × treat	−0.251	−0.217
	(0.115)	(0.134)
log(beds) × treat	−0.925	−0.125
	(1.288)	(1.440)
region (South) × treat		4.813
		(3.477)
region (North Central) × treat		2.681
		(3.617)
region (West) × treat		7.316
		(3.219)

TWFE regressions shown in Table 2. Recall that the joint F-test p-value for the assignment model presents strong evidence of selection, and hence the difference between the RIPW estimator and the unweighted TWFE regression are likely due to the bias of the latter.

6. CONCLUSION

We demonstrate both theoretically and empirically that the unit-specific reweighting of the OLS objective function improves the robustness of the resulting treatment effects estimator in applications with panel data. The proposed weights are constructed using the assignment process (either known or estimated), and thus appropriate in situations with substantial cross-sectional variation in the treatment paths. Practically, our results allow applied researchers to exploit domain knowledge about outcomes and assignments, thus resulting in a more balanced approach to identification and estimation.

TABLE 3. Treatment effect estimates, standard errors (in parentheses), and confidence intervals.

	TWFE (w/o Region FE)	TWFE (w/ Region FE)	RIPW
Estimate	−0.641	−0.619	−3.403*
	(1.640)	(1.640)	(2.047)
90% CI	[−3.34, 2.06]	[−3.32, 2.08]	[−6.77, −0.04]
95% CI	[−3.86, 2.57]	[−3.83, 2.60]	[−7.42, 0.61]

References

Abadie, Alberto (2005), "Semiparametric difference-in-differences estimators." *The Review of Economic Studies*, 72 (1), 1–19. [1003]

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (2020), "Sampling-based versus design-based uncertainty in regression analysis." *Econometrica*, 88 (1), 265–296. [1015]

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (2023), "When should you adjust standard errors for clustering?" *The Quarterly Journal of Economics*, 138 (1), 1–35. [1004]

Abouk, Rahi and Babak Heydari (2021), "The immediate effect of covid-19 policies on social-distancing behavior in the United States." *Public health reports*, 136 (2), 245–252. [1025]

Aguirregabiria, Victor, Jiaying Gu, and Yao Luo (2021), "Sufficient statistics for unobserved heterogeneity in structural dynamic logit models." *Journal of Econometrics*, 223 (2), 280–311. [1017]

Angrist, Joshua D. and Alan B. Krueger (1999), "Empirical strategies in labor economics." In *Handbook of Labor Economics*, Vol. 3, 1277–1366, Elsevier. [1000]

Arellano, Manuel (2003), *Panel Data Econometrics*. Oxford university press. [1000, 1018]

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager (2021), "Synthetic difference-in-differences." *American Economic Review*, 111 (12), 4088–4118. [1002]

Arkhangelsky, Dmitry and Guido W. Imbens (2022), "Doubly robust identification for causal panel data models." *The Econometrics Journal*, 25 (3), 649–674. [1002, 1017]

Arkhangelsky, Dmitry, Guido W. Imbens, Lihua Lei, and Xiaoman Luo (2024), "Supplement to 'Design-robust two-way-fixed-effects regression for panel data'." *Quantitative Economics Supplemental Material*, 15, https://doi.org/10.3982/QE1962. [1000]

Ashenfelter, Orley and David Card (1985), "Using the longitudinal structure of earnings to estimate the effect of training programs." *The Review of Economics and Statistics*, 648–660. [1000]

Athey, Susan and Guido W. Imbens (2022), "Design-based analysis in difference-in-differences settings with staggered adoption." *Journal of Econometrics*, 226 (1), 62–79. [1000, 1002, 1004, 1005, 1009, 1012]

Attanasio, Orazio P., Costas Meghir, and Ana Santiago (2012), "Education choices in Mexico: Using a structural model and a randomized experiment to evaluate progresa." *The Review of Economic Studies*, 79 (1), 37–66. [1001, 1003]

Bang, Heejung and James M. Robins (2005), "Doubly robust estimation in missing data and causal inference models." *Biometrics*, 61 (4), 962–973. [1001]

Benkeser, David, Marco Carone, Mark J. Van Der Laan, and Peter B. Gilbert (2017), "Doubly robust nonparametric inference on the average treatment effect." *Biometrika*, 104 (4), 863–880. [1020]

Blackwell, Matthew and Soichiro Yamauchi (2021), "Adjusting for unmeasured confounding in marginal structural models with propensity-score fixed effects." arXiv preprint. arXiv:2105.03478. [1002]

Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard (2021), "Panel experiments and dynamic causal effects: A finite population perspective." *Quantitative Economics*, 12 (4), 1171–1196. [1001, 1002, 1004]

Bojinov, Iavor, David Simchi-Levi, and Jinglong Zhao (2023), "Design and analysis of switchback experiments." *Management Science*, 69 (7), 3759–3777. [1004]

Boneva, Lena, Oliver Linton, and Michael Vogt (2015), "A semiparametric model for heterogeneous panel data with fixed effects." *Journal of Econometrics*, 188 (2), 327–345. [1018]

Borusyak, Kirill and Peter Hull (2023a), "Nonrandom exposure to exogenous shocks." *Econometrica*, 91 (6), 2155–2185. [1001, 1004, 1015]

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024), "Revisiting event study designs: Robust and efficient estimation." *The Review of Economic Studies*, rdae007. [1001]

Broda, Christian and Jonathan A. Parker (2014), "The economic stimulus payments of 2008 and the aggregate demand for consumption." *Journal of Monetary Economics*, 68, S20–S36. [1001, 1003]

Brown, Celia A. and Richard J. Lilford (2006), "The stepped wedge trial design: A systematic review." *BMC medical research methodology*, 6 (1), 1–9. [1003]

Callaway, Brantly and Pedro HC Sant'Anna (2021), "Difference-in-differences with multiple time periods." *Journal of econometrics*, 225 (2), 200–230. [1001, 1003, 1008]

Cao, Weihua, Anastasios A. Tsiatis, and Marie Davidian (2009), "Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data." *Biometrika*, 96 (3), 723–734. [1020]

Chamberlain, Gary (1984), "Panel data." *Handbook of econometrics*, 2, 1247–1318. [1003, 1015]

Chandar, Bharat, Uri Gneezy, John A. List, and Ian Muir (2019), "The drivers of social preferences: Evidence from a nationwide tipping field experiment." Technical report, National Bureau of Economic Research. [1004, 1009]

Chandar, Bharat K., Ali Hortaçsu, John A. List, Ian Muir, and Jeffrey M. Wooldridge (2019), "Design and analysis of cluster-randomized field experiments in panel data settings." Technical report, National Bureau of Economic Research. [1004, 1009]

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey (2017), "Double/debiased/Neyman machine learning of treatment effects." *American Economic Review*, 107 (5), 261–265. [1019, 1020]

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018), "Double/debiased machine learning for treatment and structural parameters." *The Econometrics Journal*, 21 (1), C1–C68. [1001, 1019, 1020]

Chetty, Raj, John N. Friedman, Nathaniel Hendren, and Michael Stepner (2020), "The opportunity insights team. How did COVID-19 and stabilization policies affect spending and employment? A new real-time economic tracker based on private sector data." National Bureau of Economic Research Cambridge, MA. [1025]

Chinazzi, Matteo, Jessica T. Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun et al. (2020), "The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak." *Science*, 368 (6489), 395–400. [1025]

Colonnelli, Emanuele and Mounu Prem (2022), "Corruption and firms." *The Review of Economic Studies*, 89 (2), 695–732. [1001, 1004]

Cox, David R. (1972), "Regression models and life-tables." *Journal of the Royal Statistical Society: Series B (Methodological)*, 34 (2), 187–202. [1027]

Currie, Janet, Henrik Kleven, and Esmée Zwiers (2020), "Technology and big data are changing economics: Mining text to track methods." In *AEA Papers and Proceedings*, Vol. 110, 42–48. [1000, 1001]

De Chaisemartin, Clement and Xavier d'Haultfoeuille (2020), "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110 (9), 2964–2996. [1001, 1012]

Dong, Ensheng, Hongru Du, and Lauren Gardner (2020), "An interactive web-based dashboard to track covid-19 in real time." *The Lancet infectious diseases*, 20, 533–534. [1026]

Goodman-Bacon, Andrew (2021), "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225 (2), 254–277. [1001, 1012]

Goodman-Bacon, Andrew and Jan Marcus (2020), "Using difference-in-differences to identify causal effects of covid-19 policies." In *In Survey Research Methods*, Vol. 14, 153–158, European Survey Research Association, Southampton. [1025]

Hahn, Jinyong and Whitney Newey (2004), "Jackknife and analytical bias reduction for nonlinear panel models." *Econometrica*, 72 (4), 1295–1319. [1015]

Hemming, Karla, Terry P. Haines, Peter J. Chilton, Alan J. Girling, and Richard J. Lilford (2015), "The stepped wedge cluster randomised trial: Rationale, design, analysis, and reporting." *Bmj*, 350. [1004]

Holtz, David, Michael Zhao, Seth G. Benzell, Cathy Y. Cao, Mohammad Amin Rahimian, Jeremy Yang, Jennifer Allen, Avinash Collis, Alex Moehring, and Tara Sowrirajan (2020), "Interdependence and the cost of uncoordinated responses to covid-19." *Proceedings of the National Academy of Sciences*, 117 (33), 19837–19843. [1025]

Imbens, Guido (2000), "The role of the propensity score in estimating dose–response functions." *Biometrika*, 87 (0), 706–710. [1000, 1004]

Kang, Joseph and Joseph Schafer (2007), "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data." *Statistical science*, 22 (4), 523–539. [1001, 1017, 1019]

Kraemer, Moritz UG, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Louis Du Plessis, Nuno R. Faria, Ruoran Li, and William P. Hanage (2020), "The effect of human mobility and control measures on the covid-19 epidemic in China." *Science*, 368 (6490), 493–497. [1025]

Lancaster, Tony (2000), "The incidental parameter problem since 1948." *Journal of econometrics*, 95 (2), 391–413. [1003, 1015]

Lin, Winston (2013), "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *The Annals of Applied Statistics*, 295–318. [1009]

MIT Election Data and Science Lab, (2018), "County presidential election returns 2000–2020." 10.7910/DVN/VOQCHQ. [1026]

Neyman, Jerzy (1923/1990), "On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Translated by Dabrowska, D. M. and Speed, T. P." *Statistical Science*, 5, 465–472. [1004, 1011]

Neyman, Jerzy and Elizabeth L. Scott (1948), "Consistent estimates based on partially consistent observations." *Econometrica: Journal of the Econometric Society*, 1–32. [1003, 1010]

OpenTable, "The restaurant industry, by the numbers." https://www.opentable.com/state-of-industry. [1026]

Perper, Rosie, Ellen Cranley, and Sarah Al-Arshani, "Almost all us states have declared states of emergency to fight coronavirus—here's what it means for them." Available at https://www.businessinsider.com/california-washington-state-of-emergency-coronavirus-what-it-means-2020-3. [1026]

Rambachan, Ashesh and Jonathan Roth (2020), "Design-based uncertainty for quasi-experiments." arXiv preprint. arXiv:2008.00602. [1004]

Rényi, Alfréd (1959), "On measures of dependence." *Acta Mathematica Academiae Scientiarum Hungarica*, 10 (3–4), 441–451. [1005]

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao (1994), "Estimation of regression coefficients when some regressors are not always observed." *Journal of the American statistical Association*, 89 (427), 846–866. [1001, 1017, 1019]

Rosenbaum, Paul R. and Donald B. Rubin (1983), "The central role of the propensity score in observational studies for causal effects." *Biometrika*, 70 (1), 41–55. [1001]

Roth, Jonathan and Pedro HC Sant'Anna (2023), "Efficient estimation for staggered rollout designs." *Journal of Political Economy Microeconomics*, 1 (4), 669–709. [1002, 1004, 1009, 1012]

Rubin, Donald B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology*, 66 (5), 688. [1011]

Sant'Anna, Pedro HC and Jun Zhao (2020), "Doubly robust difference-in-differences estimators." *Journal of Econometrics*. [1002, 1008]

Schoenfeld, David (1980), "Chi-squared goodness-of-fit tests for the proportional hazards regression model." *Biometrika*, 67 (1), 145–153. [1027]

Shaikh, Azeem and Panagiotis Toulis (2019), "Randomization tests in observational studies with staggered adoption of treatment." University of Chicago, Becker Friedman Institute for Economics Working Paper, (2019-144). [1001]

Sun, Liyang and Sarah Abraham (2021), "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225 (2), 175–199. [1001, 1012]

Wojtaszek, Carl and Michael Kofoed (2022), "Sensitivity of charitable giving to realized income changes: Evidence from military bonuses and the combined federal campaign." [1016]

Wooldridge, Jeffrey M. (2010), *Econometric Analysis of Cross Section and Panel Data*. MIT press. [1018, 1019]

Wooldridge, Jeffrey M. (2021), "Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators." Available at SSRN 3906345. [1018]

Zemel, David, Kate Eldridge, Robert Bracco, Sam King, and Adam Siemer "COVID19 comparison." Available at https://github.com/rbracco/covidcompare. [1026]

Co-editor Stéphane Bonhomme handled this manuscript.

Manuscript received 3 August, 2021; final version accepted 19 July, 2024; available online 22 July, 2024.

The replication package for this paper is available at https://doi.org/10.5281/zenodo.12637091. The Journal checked the data and codes included in the package for their ability to reproduce the results in the paper and approved online appendices.