Submitted to *Econometrica*

# SAME ROOT DIFFERENT LEAVES: TIME SERIES AND CROSS-SECTIONAL METHODS IN PANEL DATA

Dennis Shen
Simons Institute for the Theory of Computing, University of California, Berkeley

Peng Ding
Department of Statistics, University of California, Berkeley

Jasjeet Sekhon
Departments of Statistics & Data Science and Political Science, Yale University

Bin Yu
Departments of Statistics and EECS, University of California, Berkeley

One dominant approach to evaluate the causal effect of a treatment is through panel data analysis, whereby the behaviors of multiple units are observed over time. The information across time and units motivates two general approaches: (i) horizontal regression (i.e., unconfoundedness), which exploits time series patterns, and (ii) vertical regression (e.g., synthetic controls), which exploits cross-sectional patterns. Conventional wisdom often considers the two approaches to be different. We establish this position to be partly false for estimation but generally true for inference. In the absence of any assumptions, we show that both approaches yield algebraically equivalent point estimates for several standard estimators. However, the source of randomness assumed by each approach leads to a distinct estimand and quantification of uncertainty even for the same point estimate. This emphasizes that researchers should carefully consider where the randomness stems from in their data as it has direct implications for the accuracy of inference.

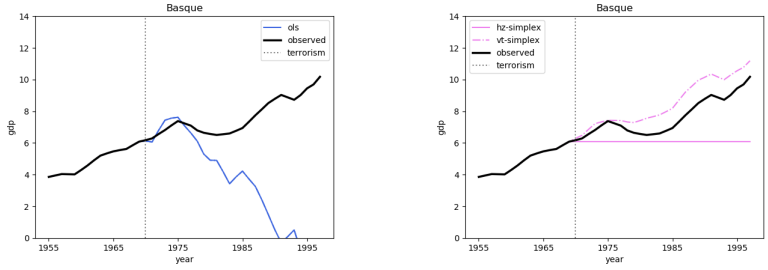
Keywords: horizontal regression, vertical regression, unconfoundedness, synthetic controls, causal inference, minimum norm estimators.

## 1. INTRODUCTION

In a seminal paper, Abadie and Gardeazabal (2003) set out to investigate the economic impact of terrorism in Basque Country. Prior to the outset of terrorist activity in the early 1970's, Basque Country was considered to be one of the wealthiest regions in Spain. After thirty years of turmoil, however, its economic activity dropped substantially relative to its neighboring regions. Although intuition affirms that Basque Country's economic downturn can be attributed, at least partially, to its political and civil unrest, it is difficult to quantitatively isolate the economic costs of conflict. In response to this challenge, Abadie and Gardeazabal (2003) introduced the synthetic

Dennis Shen: dshen24@berkeley.edu
Peng Ding: pengdingpku@berkeley.edu
Jasjeet Sekhon: jasjeet.sekhon@yale.edu
Bin Yu: binyu@berkeley.edu

We sincerely thank Alberto Abadie, Avi Feller, and Devavrat Shah for their thoughtful comments and insightful feedback. We gratefully acknowledge support from NSF grants 1945136, 1953191, 2022448, 2023505 on Collaborative Research: Foundations of Data Science Institute (FODSI), and ONR grant N00014-17-1-2176. The data and code to reproduce the results in this article are available at https://github.com/deshen24/panel-data-regressions.




(a) Symmetric regressions. (b) Asymmetric regressions.

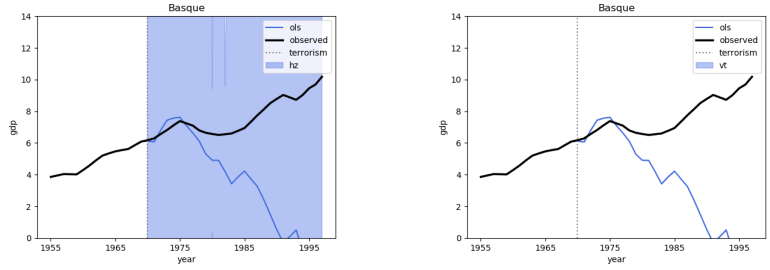
FIGURE 1.—1a: Estimates of OLS with minimum $\ell_2$-norm; see Section 3.1.1. 1b: Estimates of simplex regression. HZ and VT estimates correspond to colored solid and dashed-dotted lines, respectively. The outset of terrorism is the vertical line and Basque Country's observed GDP is in solid black. Notably, the OLS point estimates likely suffer from overfitting.

controls framework. At its core, synthetic controls constructs a synthetic Basque Country from a weighted composition of control regions that are largely unaffected by the instability to estimate Basque Country's economic evolution in the absence of terrorism. This novel concept has inspired an entire subliterature within econometrics that is "arguably the most important innovation in the policy evaluation literature in the last 15 years" (Athey and Imbens, 2017).

Researchers have historically tackled problems of this flavor using repeated observations of units across time, i.e., panel data, where a subset of units are exposed to a treatment during some time periods while the other units are unaffected. In the study above, the per capita gross domestic product (GDP) of 17 Spanish regions are measured from 1955–1998. Basque Country is the sole treated unit and the remaining regions are the control units; the pre- and post-treatment periods are defined as the time horizons before and after the first wave of terrorist activity, respectively.

Synthetic controls has become a cornerstone for panel studies in recent years and across numerous fields. Beforehand, the unconfoundedness approach (Rosenbaum and Rubin, 1983, Imbens and Wooldridge, 2009) served as a common workhorse. Whereas synthetic controls posits a relation between treated and control units that is stable across time, unconfoundedness posits a relation between treated and pretreatment periods that is stable across units. Accordingly, synthetic controls exploits cross-sectional correlation patterns while unconfoundedness exploits time series correlation patterns. Considering the panel data format, unconfoundedness and synthetic controls based methods are commonly referred to as horizontal (HZ) and vertical (VT) regressions, respectively. Given their conceptual and computational distinctions, the two approaches are considered to be different from one another (Athey et al., 2021).

Yet, contrary to conventional wisdom, it turns out that HZ and VT regressions can yield identical point estimates. As Figure 1a shows, when the regression models are learned via ordinary least squares (OLS), then the two approaches produce the same economic evolution for Basque Country in the absence of terrorism. Figure 1b, by contrast, shows that when the regression models are enforced to lie within the simplex—as proposed by Abadie and Gardeazabal (2003) for VT regression—then the two approaches output contrasting economic trajectories. To make matters more intriguing, Figure 2 indicates that even when the two regressions arrive at the same point estimate, the corresponding confidence intervals can be markedly different under different sources of randomness. The juxtaposition of these figures beg two questions:




(a) OLS under a HZ model. (b) OLS under a VT model.

FIGURE 2.—Confidence intervals for OLS with minimum $\ell_2$-norm constructed from HZ-based (left) and VT-based (right) generative models; see Section 4.1.1. The VT-based confidence intervals are degenerate.

**Q1:** "When are HZ and VT point estimates identical?"
**Q2:** "When the point estimates are identical, how does the source of randomness impact inference?"

**Contribution.** This article provides a technical contribution to Q1 and a conceptual contribution to Q2. For Q1, we classify various widely studied regression formulations into (i) a symmetric class that yields algebraically identical point estimates and (ii) an asymmetric class that yields algebraically contrasting point estimates. These results hold without any assumptions on the data generating process or data configuration.

With this result in place, we proceed to tackle Q2 by staying within the symmetric class and studying properties of the estimator with randomness stemming from (i) time series patterns, (ii) cross-sectional patterns, and (iii) both patterns simultaneously. We conduct our analysis from a *model-based* perspective, which attributes randomness to the potential outcomes, and a *design-based* perspective, which attributes randomness to the mechanism assignment of treatment. Even under the most stylized assumptions within each framework, we demonstrate that each source of randomness leads to a distinct estimand and variance for the same point estimate.

While the specific estimands and variances may vary with the underlying assumptions, the general message remains invariant. In this spirit, we construct distinct confidence intervals for each source of randomness based on our model-based assumptions and asymptotic analysis. Though these intervals are unlikely to be practical for real-world settings, they are a useful device to conduct data-inspired simulations and empirical applications in illustrating our key concept. Indeed, our findings highlight that the confidence interval developed for one estimand often has incorrect coverage for another. Altogether, our results emphasize that the source of randomness that researchers assume has direct implications for the accuracy of inference that can be conducted. This further motivates the need for a principled framework to check researchers' assumptions, which is left as important future work.

**Organization.** Section 2 overviews the panel data framework. Sections 3–4 provide one set of answers for Q1–Q2. Section 5 illustrates concepts developed in this article. Section 6 summarizes our findings. We relegate mathematical proofs to Appendices A–B.

**Notation.** Let $\boldsymbol{I}$ be the identity matrix. Let $\mathbf{1}$ and $\mathbf{0}$ be the vectors of ones and zeros, respectively. The curled inequality denotes $\succeq$ the generalized inequality, i.e., componentwise inequality between vectors and matrix inequality between symmetric matrices.

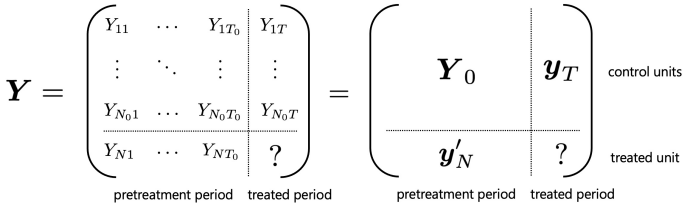



FIGURE 3.—Panel data format with rows and columns indexed by units and time, respectively.

Let $\circ$ denote the elementwise product. For vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, let $\langle a, b\rangle = a'b$ denote the inner product. Let $\mathrm{tr}(\boldsymbol{A})$ denote the trace of $\boldsymbol{A}$. We define $0/0 = 0$ when applicable.

## 2. THE PANEL DATA FRAMEWORK

We anchor on the Basque study to introduce the panel data framework and relevant notations. Panel data contains observations of $N$ units over $T$ time periods. The Basque study, for instance, consists of per capita GDP across $N = 17$ Spanish regions over $T = 43$ years. In each time period $t$, each unit $i$ is characterized by two potential outcomes, $Y_{it}(0)$ and $Y_{it}(1)$, which correspond to its outcome in the absence and presence of a binary treatment, respectively. The potential outcomes framework posits that each region possesses two possible levels of economic activity each year, one that is immune to terrorism and one that is affected by terrorism. In reality, however, we can only observe one economic state—this is the fundamental challenge of causal inference.

Let $A_i \in \{0, 1\}$ and $B_t \in \{0, 1\}$ be the indicator variables for whether the $i$th unit and $t$th period are treated. We write the observed outcome as

$$Y_{it} = A_i B_t \cdot Y_{it}(1) + (1 - A_i B_t) \cdot Y_{it}(0). \tag{1}$$

Often, we observe all $N$ units without treatment (control) for $T_0$ time periods, i.e., $A_i = B_t = 0$ for all $i \le N$ and $t \le T_0$. For the remaining $T_1 = T - T_0$ time periods, $N_1$ units receive treatment while the remaining $N_0 = N - N_1$ units remain under control, i.e., if we label the first $N_0$ units as the control group, then $A_i = B_t = 0$ for all $i \le N_0$ and $t > T_0$, and $A_i = B_t = 1$ for all $i > N_0$ and $t > T_0$. In the study of Abadie and Gardeazabal (2003), Basque Country is the single treated unit, thus $N_1 = 1$ and $N_0 = 16$. The first wave of terrorist activity partitions the time horizon into pre- and post-treatment periods of lengths $T_0 = 15$ and $T_1 = 28$ years, respectively.

For ease of exposition, this article considers a *single* treated unit and *single* treated period indexed by the $N$th unit and $T$th time period, respectively, i.e., $A_N = B_T = 1$ and $A_i = B_t = 0$ for all other $i \le N_0$ and $t \le T$. However, our results hold for any $(i, t)$ pair where $i > N_0$ is a treated unit and $t > T_0$ is a treated period. We organize our observed control data into an $N \times T$ matrix, $\boldsymbol{Y} = [Y_{it}]$, as shown in Figure 3. In our example, $\boldsymbol{y}_N = [Y_{Nt} : t \le T_0] \in \mathbb{R}^{T_0}$ represents Basque Country's economic evolution prior to the outset of terrorism; $\boldsymbol{Y}_0 = [Y_{it} : i \le N_0, t \le T_0] \in \mathbb{R}^{N_0 \times T_0}$ represents the control regions' economic evolution prior to the outset of terrorism; and $\boldsymbol{y}_T = [Y_{iT} : i \le N_0] \in \mathbb{R}^{N_0}$ represents the control regions' economic evolution after the outset of terrorism. Our object of interest is Basque Country's counterfactual GDP in the absence of terrorism, $Y_{NT}(0)$.

### 2.1. *Time Series Versus Cross-Sectional Based Regressions*

The information across time and units motivates two natural ways to estimate the missing $(N,T)$th entry. These perspectives are explored in two large and mostly separate bodies of work (Athey et al., 2021).

#### 2.1.1. *Horizontal Regression and Unconfoundedness*

The unconfoundedness literature operates on the concept that "history is a guide to the future". As such, unconfoundedness methods express outcomes in the treated period as a weighted composition of outcomes in the pretreatment periods. This is carried out by regressing the control units' treated period outcomes $\boldsymbol{y}_T$ on its lagged outcomes $\boldsymbol{Y}_0$ and applying the learned regression coefficients to the treated unit's lagged outcomes $\boldsymbol{y}_N$ to predict the missing $(N,T)$th outcome. Following Athey et al. (2021), we refer to such methods as horizontal (HZ) regression.

#### 2.1.2. *Vertical Regression and Synthetic Controls*

The synthetic controls literature is built on the concept that "similar units behave similarly". Therefore, synthetic controls methods express the treated unit's outcomes as a weighted composition of control units' outcomes. This is carried out by regressing the treated unit's lagged outcomes $\boldsymbol{y}_N$ on the control units' lagged outcomes $\boldsymbol{Y}_0'$ and applying the learned regression coefficients to the control units' treated period outcomes $\boldsymbol{y}_T$ to predict the missing $(N,T)$th outcome. Following Athey et al. (2021), we refer to such methods as vertical (VT) regression.

#### 2.1.3. *Conventional Wisdom*

The dimensions of the data often guide the choice of estimator. In fact, the unregularized forms of HZ and VT regressions are cautioned against when $T > N$ and $N > T$, respectively, due to overfitting (Abadie et al., 2015, Doudchenko and Imbens, 2016, Li and Bell, 2017, Athey et al., 2021). With regularization, however, Athey et al. (2021) argues the two approaches can be applied to the same setting. This allows HZ and VT regressions to be systematically compared through methods such as cross-validation.

In parallel, the growth rates of the two literatures have also exhibited asymmetry. While the development of the unconfoundedness literature has seemingly plateaued, the synthetic controls literature continues to rapidly expand. Across many domains, synthetic controls based methods are arguably the de facto approach for panel studies.

## 3. POINT ESTIMATION

**Q1:** "When are HZ and VT point estimates identical?"

We tackle Q1 by studying algebraic properties of the HZ and VT point estimates. Below, we denote the singular value decomposition of $\boldsymbol{Y}_0$ as $\boldsymbol{Y}_0 = \sum_{\ell=1}^{R} s_\ell \boldsymbol{u}_\ell \boldsymbol{v}_\ell' = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}'$, where $\boldsymbol{u}_\ell \in \mathbb{R}^{N_0}$ and $\boldsymbol{v}_\ell \in \mathbb{R}^{T_0}$ are the left and right singular vectors, respectively, $s_\ell \in \mathbb{R}$ are the ordered singular values, and $R = \operatorname{rank}(\boldsymbol{Y}_0) \leq \min\{N_0, T_0\}$. Let $\boldsymbol{U} \in \mathbb{R}^{N_0 \times R}$ and $\boldsymbol{V} \in \mathbb{R}^{T_0 \times R}$ be the matrices formed by the left and right singular vectors, respectively, and $\boldsymbol{S} \in \mathbb{R}^{R \times R}$ be the diagonal matrix of singular values. The pseudoinverse is $\boldsymbol{Y}_0^\dagger = \sum_{\ell=1}^{R} (1/s_\ell) \boldsymbol{v}_\ell \boldsymbol{u}_\ell' = \boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{U}'$.

### 3.1. *Classifying Notable Least Squares Formulations*

We present several of the most widely studied regression formulations in the HZ and VT literatures. This list is far from exhaustive given the vastness of these literatures.

#### 3.1.1. *Description of Least Squares Formulations*

**Penalized least squares.** A large class of formulations are expressed as follows:

(a) HZ regression: for $\lambda_1, \lambda_2 \geq 0$,

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{y}_T - \boldsymbol{Y}_0\boldsymbol{\alpha}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\|_1 + \lambda_2\|\boldsymbol{\alpha}\|_2^2, \quad \text{with} \quad \widehat{Y}_{NT}^{\text{hz}}(0) = \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}} \rangle. \quad (2)$$

(b) VT regression: for $\lambda_1, \lambda_2 \geq 0$,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{y}_N - \boldsymbol{Y}_0'\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_2^2, \quad \text{with} \quad \widehat{Y}_{NT}^{\text{vt}}(0) = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}} \rangle. \quad (3)$$

We overview common choices for $(\lambda_1, \lambda_2)$ and describe the corresponding strategy.

*I: Ordinary least squares (OLS).* Arguably, the mother of all regressions is OLS, where $\lambda_1 = \lambda_2 = 0$. OLS is an unconstrained problem with possibly infinitely many solutions. In such settings, one particular solution is the vector with minimum $\ell_2$-norm; this minimizer is unique and described through the pseudoinverse. OLS has been analyzed in numerous panel study works, including Hsiao et al. (2012), Li and Bell (2017), and Li (2020).

*II: Principal component regression (PCR).* To formalize PCR, let $\boldsymbol{Y}_0^{(k)} = \sum_{\ell=1}^{k} s_\ell \boldsymbol{u}_\ell \boldsymbol{v}_\ell'$ denote the rank $k < R$ approximation of $\boldsymbol{Y}_0$ that retains the top $k$ principal components. HZ and VT PCR corresponds to replacing $\boldsymbol{Y}_0$ with $\boldsymbol{Y}_0^{(k)}$ within (2) and (3), respectively, with $\lambda_1 = \lambda_2 = 0$. In words, PCR first finds a $k$ dimensional representation of the covariate matrix via principal component analysis; then, PCR performs OLS with the compressed $k$ dimensional covariates. Within the synthetic controls literature, Amjad et al. (2018, 2019) and Agarwal et al. (2021) utilize PCR.

*III: Ridge regression.* Ridge considers $\lambda_1 = 0$ and $\lambda_2 > 0$. When $\boldsymbol{Y}_0$ is rank deficient, the gram matrix, i.e., $\boldsymbol{Y}_0'\boldsymbol{Y}_0$ for HZ regression and $\boldsymbol{Y}_0\boldsymbol{Y}_0'$ for VT regression, is ill-conditioned. This can discourage the usage of OLS. Ridge provides a remedy by adding a *ridge* on the diagonal of the gram matrix, which increases all eigenvalues by $\lambda_2$, thus removing the singularity problem. Ben-Michael et al. (2021) explores properties of a doubly robust estimator that utilizes HZ ridge regression.

*IV: Lasso regression.* Lasso considers $\lambda_1 > 0$ and $\lambda_2 = 0$. Lasso is a popular tool for estimating sparse linear coefficients in high-dimensional regimes. Because the criterion not strictly convex, there are possibly infinitely many solutions. Thus, for our analysis of lasso only, we make the mild assumption that the entries of $\boldsymbol{Y}_0$ are drawn from a continuous distribution. This guarantees the lasso solution to be unique (Tibshirani, 2013). Several notable works in the synthetic controls literature, e.g., Li and Bell (2017), Carvalho et al. (2018), and Chernozhukov et al. (2021), analyze the lasso.

*V: Elastic net regression.* Elastic net considers $\lambda_1, \lambda_2 > 0$. At a high level, elastic net selects variables similar to the lasso, but deals with correlated variables more gracefully as with ridge. When $\lambda_2 > 0$, the criterion is strictly convex so the solution is unique. Doudchenko and Imbens (2016) propose an elastic net synthetic controls variant.

**Constrained least squares.**
*VI: Simplex regression.* The next formulation constrains the regression weights to lie within the simplex, i.e., the weights are nonnegative and sum to one:
(a) HZ regression: for $\lambda \geq 0$,

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}:\boldsymbol{\alpha}'\mathbf{1}=1,\boldsymbol{\alpha}\geq\mathbf{0}}{\operatorname{argmin}} \|\boldsymbol{y}_T - \boldsymbol{Y}_0\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2, \quad \text{with} \quad \widehat{Y}_{NT}^{\text{hz}}(0) = \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}}\rangle. \tag{4}$$

(b) VT regression: for $\lambda \geq 0$,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}:\boldsymbol{\beta}'\mathbf{1}=1,\boldsymbol{\beta}\geq\mathbf{0}}{\operatorname{argmin}} \|\boldsymbol{y}_N - \boldsymbol{Y}_0'\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2, \quad \text{with} \quad \widehat{Y}_{NT}^{\text{vt}}(0) = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}\rangle. \tag{5}$$

We consider a vanishing $\ell_2$ penalty since $\lambda = 0$ (standard formulation) can induce multiple minima (Abadie and L'Hour, 2021). When $\lambda > 0$, the criterion becomes strictly convex and the solution is unique. Simplex regression is the original formulation set forth in the pioneering works of Abadie and Gardeazabal (2003), Abadie et al. (2010, 2015), and its properties continue to be actively studied today. Attractive aspects of simplex regression include interpretability, sparsity, and transparency (Abadie, 2021).

### 3.1.2. *Classification Results*

To answer Q1, we classify the regression formulations into a (i) *symmetric* class and an (ii) *asymmetric* class. We use the shorthand $\mathsf{HZ} = \mathsf{VT}$ if the two approaches produce algebraically identical point estimates and $\mathsf{HZ} \neq \mathsf{VT}$ otherwise.

**I: Symmetric class.** We first state the symmetric formulations.

THEOREM 1: $\mathsf{HZ} = \mathsf{VT}$ *for (i) OLS with* $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ *as the minimum* $\ell_2$*-norm solutions:*

$$\widehat{Y}_{NT}^{\text{hz}}(0) = \widehat{Y}_{NT}^{\text{vt}}(0) = \langle \boldsymbol{y}_N, \boldsymbol{Y}_0^{\dagger}\boldsymbol{y}_T\rangle = \sum_{\ell=1}^{R}(1/s_\ell)\langle \boldsymbol{y}_N, \boldsymbol{v}_\ell\rangle\langle \boldsymbol{u}_\ell, \boldsymbol{y}_T\rangle;$$

*(ii) PCR with the same choice of* $k < R$*:*

$$\widehat{Y}_{NT}^{\text{hz}}(0) = \widehat{Y}_{NT}^{\text{vt}}(0) = \langle \boldsymbol{y}_N, (\boldsymbol{Y}_0^{(k)})^{\dagger}\boldsymbol{y}_T\rangle = \sum_{\ell=1}^{k}(1/s_\ell)\langle \boldsymbol{y}_N, \boldsymbol{v}_\ell\rangle\langle \boldsymbol{u}_\ell, \boldsymbol{y}_T\rangle;$$

*(iii) ridge regression with the same choice of* $\lambda_2 > 0$*:*

$$\widehat{Y}_{NT}^{\text{hz}}(0) = \widehat{Y}_{NT}^{\text{vt}}(0) = \langle \boldsymbol{y}_N, (\boldsymbol{Y}_0'\boldsymbol{Y}_0 + \lambda_2\boldsymbol{I})^{-1}\boldsymbol{Y}_0'\boldsymbol{y}_T\rangle = \sum_{\ell=1}^{R}\frac{s_\ell}{s_\ell^2 + \lambda_2}\langle \boldsymbol{y}_N, \boldsymbol{v}_\ell\rangle\langle \boldsymbol{u}_\ell, \boldsymbol{y}_T\rangle.$$

Theorem 1 might seem familiar at first glance. As observed in Abadie et al. (2015) and Ben-Michael et al. (2021, Lemma 2), the point estimates associated with HZ OLS and HZ ridge can be written as linear combinations of the elements in $\boldsymbol{y}_T$, which take the same *linear forms* as the corresponding VT point estimates. However, their results stop short of establishing algebraic equivalence as in Theorem 1. In this view, Theorem 1 is perhaps surprising. It demonstrates that the HZ and VT perspectives—while appearing to be different—are, in fact, two equivalent ways of approaching the same problem

when the regression model belongs to the symmetric class. Notably, the identity holds without any assumptions on the data generating process.

Theorem 1 also holds for any data configuration. Thus, it clarifies that HZ and VT OLS are *not* invalid when $T > N$ and $N > T$, respectively. In fact, the OLS estimate can even be written as $\langle \widehat{\boldsymbol{\alpha}}, \boldsymbol{Y}_0' \widehat{\boldsymbol{\beta}} \rangle$, which incorporates both regression models. It is likely that the prior misconception arose from the fact that infinitely many solutions can exist depending on the dimensions of the data and chosen approach. Among these solutions, however, is the unique minimum $\ell_2$-norm model, which is the solution when the objective is optimized via gradient descent. This phenomena is known as "implicit regularization", where the optimization algorithm is biased towards a particular solution even though the bias is not explicit in the objective function (Neyshabur et al., 2015, Gunasekar et al., 2017).

Through its connection to the $\ell_2$-penalty, the minimum $\ell_2$-norm also offers a high-level intuition for the root of symmetry. More specifically, observe that the ridge model converges to the OLS model with minimum $\ell_2$-norm as $\lambda_2 \to 0$. Since the PCR model is precisely the OLS minimum $\ell_2$-norm model that is restricted to the space spanned by the top $k$ principal components, we conjecture that the geometry of the $\ell_2$-ball is the likely source of symmetry for HZ and VT point estimation.

**II: Asymmetric class.** Next, we state the class of asymmetric formulations.

THEOREM 2: HZ $\neq$ VT *for (i) lasso, (ii) elastic net, and (iii) simplex regression.*

A common thread of the objective functions in the asymmetric class is a penalty or constraint that promotes sparse models. Such regularizers are noticeably absent in the symmetric formulations. This suggests that geometries of the $\ell_1$-ball and simplex that encourage sparsity are likely sources of asymmetry for HZ and VT point estimation.

### 3.2. *Doubly Robust Regression*

These recent years have seen a surge of interest in doubly robust (DR) estimators. Within panel data, we discuss two prominent works that are rising in popularity.

#### 3.2.1. *Synthetic Difference-in-Differences*

An important approach that continues to dominate empirical work in panel data is the difference-in-differences (DID) estimator (Ashenfelter, 1978). In essence, DID posits an additive outcome model with unit- and time-specific fixed effects, known colloquially as the "parallel trends" assumption. Arkhangelsky et al. (2021) anchors on the DID principle and brings in concepts from the unconfoundedness and synthetic controls literatures to derive a DR estimator called synthetic difference-in-differences (SDID). In our setting, the SDID prediction for the missing $(N, T)$th potential outcome is

$$\begin{aligned} \widehat{Y}_{NT}^{\text{sdid}}(0) &= \sum_{i \le N_0} \widehat{\beta}_i Y_{iT} + \sum_{t \le T_0} \widehat{\alpha}_t Y_{Nt} - \sum_{i \le N_0} \sum_{t \le T_0} \widehat{\beta}_i \widehat{\alpha}_t Y_{it} \\ &= \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}} \rangle + \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}} \rangle - \langle \widehat{\boldsymbol{\beta}}, \boldsymbol{Y}_0 \widehat{\boldsymbol{\alpha}} \rangle, \end{aligned} \tag{6}$$

where $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ represent general HZ and VT models, respectively. Observe that $\widehat{\boldsymbol{\alpha}} = (1/T_0)\mathbf{1}$ and $\widehat{\boldsymbol{\beta}} = (1/N_0)\mathbf{1}$ recovers DID. Moving beyond simple DID to performing a weighted two-way bias removal, Arkhangelsky et al. (2021) propose to learn $\widehat{\boldsymbol{\alpha}}$ via simplex regression and $\widehat{\boldsymbol{\beta}}$ via simplex regression with an $\ell_2$-penalty.

3.2.2. *Augmented Synthetic Controls*

Ben-Michael et al. (2021) introduce the augmented synthetic control (ASC) estimator, which uses an outcome model to correct the bias induced by the classical synthetic controls estimator.[1] Concretely, the ASC estimator predicts the missing $(N,T)$th potential outcome as

$$\widehat{Y}^{\text{asc}}_{NT}(0) = \widehat{M}_{NT}(0) + \sum_{i \le N_0} \widehat{\beta}_i (Y_{iT} - \widehat{M}_{iT}(0)), \tag{7}$$

where $\widehat{M}_{iT}(0)$ is the estimator for the $(i,T)$th entry. Ben-Michael et al. (2021) instantiate $\widehat{M}_{iT}(0) = \sum_{t \le T_0} \widehat{\alpha}_t Y_{it}$. Plugging this HZ outcome model into (7) then gives

$$\widehat{Y}^{\text{asc}}_{NT}(0) = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}} \rangle + \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}} \rangle - \langle \widehat{\boldsymbol{\beta}}, \boldsymbol{Y}_0 \widehat{\boldsymbol{\alpha}} \rangle. \tag{8}$$

We consider this particular variant of ASC since it takes the same form as SDID, as seen in (6). In contrast to Arkhangelsky et al. (2021), Ben-Michael et al. (2021) learns $\widehat{\boldsymbol{\alpha}}$ via ridge regression and $\widehat{\boldsymbol{\beta}}$ via simplex regression.

3.2.3. *Connecting DR Regression to HZ and VT Regressions*

We refer to SDID and ASC, as defined in (6) and (8), respectively, as DR regression. To complement the existing results on DR regression for panel data, we leverage Theorem 1 to study properties of DR regression when $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ come from the symmetric class.

COROLLARY 1: DR = HZ = VT *for (i)* $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ *as the OLS minimum* $\ell_2$*-norm solutions and (ii)* $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ *as the PCR solutions with the same choice of* $k < R$.

Corollary 1 offers two interpretations. On the one hand, DR regression implicitly exploits only *one* pattern in the data if $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$ are implicitly regularized or learned via PCR. On the other hand, HZ and VT regressions implicitly exploit *both* patterns in the data for the same considerations on $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$. We keep with the latter perspective as it is similar in spirit to the "OLS is doubly robust" argument in Robins et al. (2007) but for the panel data setting. As Section 4 discusses, the second interpretation is further justified when randomness stems from both patterns of the data as well.

### 3.3. *Intercepts*

Intercepts can be included in the HZ regression model by modifying the $\ell_2$-errors in (2) and (4) as $\|\boldsymbol{y}_T - \boldsymbol{Y}_0 \boldsymbol{\alpha} - \alpha_0 \mathbf{1}\|_2^2$; similarly, they can be included in the VT regression model by modifying $\ell_2$-errors in (3) and (5) as $\|\boldsymbol{y}_N - \boldsymbol{Y}_0' \boldsymbol{\beta} - \beta_0 \mathbf{1}\|_2^2$. We discuss the role of intercepts for point estimation below.

COROLLARY 2: HZ $\neq$ VT *for (i) OLS, (ii) PCR, and (iii) ridge with intercepts.*

We develop an intuition for Proposition 2 by interpreting intercepts in panel studies. A nonzero time intercept, $\alpha_0$, allows for a permanent constant difference between the

[1] See Abadie and L'Hour (2021) for a bias correction of synthetic controls through matching.

treated and pretreatment periods; a nonzero unit intercept, $\beta_0$, allows for a permanent constant difference between the treated and control units. These systematic structures then create an asymmetry between the two regressions. Below, we propose a methodology based on centering the data that allows for intercepts yet retains symmetry.

3.3.1. *Including Intercepts and Retaining Symmetry through Data Centering*

Let $\boldsymbol{Y}_0$ be twice centered, i.e., the rows and columns of $\boldsymbol{Y}_0$ are mean zero. This can be satisfied by applying $\boldsymbol{I} - (1/N_0)\mathbf{1}\mathbf{1}'$ and $\boldsymbol{I} - (1/T_0)\mathbf{1}\mathbf{1}'$ to the left and right, respectively, of $\boldsymbol{Y}_0$. Consider the following modified formulations.

(a) HZ regression: for $\lambda \geq 0$,

$$(\widehat{\alpha}_0, \widehat{\alpha}_1, \widehat{\boldsymbol{\alpha}}) = \underset{(\alpha_0, \alpha_1, \boldsymbol{\alpha})}{\operatorname{argmin}} \|\boldsymbol{y}_T - \boldsymbol{Y}_0\boldsymbol{\alpha} - \alpha_0\mathbf{1}\|_2^2 + \|\boldsymbol{y}_N - \alpha_1\mathbf{1}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2 \tag{9}$$

$$\widehat{Y}_{NT}^{\text{hz}}(0) = \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}} \rangle + \widehat{\alpha}_0 + \widehat{\alpha}_1. \tag{10}$$

(b) VT regression: for $\lambda \geq 0$,

$$(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\boldsymbol{\beta}}) = \underset{(\beta_0, \beta_1, \boldsymbol{\beta})}{\operatorname{argmin}} \|\boldsymbol{y}_N - \boldsymbol{Y}_0'\boldsymbol{\beta} - \beta_0\mathbf{1}\|_2^2 + \|\boldsymbol{y}_T - \beta_1\mathbf{1}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2 \tag{11}$$

$$\widehat{Y}_{NT}^{\text{vt}}(0) = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}} \rangle + \widehat{\beta}_0 + \widehat{\beta}_1. \tag{12}$$

Similar to before, OLS corresponds to $\lambda = 0$, PCR corresponds to OLS with $\boldsymbol{Y}_0^{(k)}$ for $k < R$ in place of $\boldsymbol{Y}_0$, and ridge regression corresponds to any $\lambda > 0$.

COROLLARY 3: HZ = VT *for the symmetric estimators in Theorem 1 under the formulations set in* (9) *and* (11) *with* $\boldsymbol{Y}_0$ *being twice centered.*

We inspect (10) and (12) to understand the implications of Corollary 3. First, we recall Theorem 1, which establishes that the HZ and VT estimates share the same "base" estimate, i.e., $\langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}} \rangle = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}} \rangle$. Next, we note that $\widehat{\alpha}_0 = \widehat{\beta}_1 = (1/N_0)\mathbf{1}'\boldsymbol{y}_T$ and $\widehat{\beta}_0 = \widehat{\alpha}_1 = (1/T_0)\mathbf{1}'\boldsymbol{y}_N$, which correspond to the time and unit fixed effects, respectively. Intuitively, the modified point estimates in (10) and (12) include both fixed effect models to compensate for $\boldsymbol{Y}_0$ being twice centered. Putting everything together, the modified HZ and VT point estimates are algebraically identical.

## 4. INFERENCE

**Q2:** "When the HZ and VT point estimates are identical, how does the source of randomness impact inference?"

To answer Q2, we study the inferential properties of the counterfactual prediction. Formal discussions for classical inference require an explicit postulation on the source of randomness. This article takes both a *model-based* approach, which makes assumptions about the distribution of the potential outcomes, and a *design-based* approach, which makes assumptions about the assignment mechanism of treatment. Within each setting, we consider a natural notion of randomness stemming from (i) time series patterns, (ii) cross-sectional patterns, and (iii) both patterns simultaneously.

Before we formalize these notions, we emphasize that the goal of this section is *not* to accurately model reality and propose novel confidence intervals for practice. The goal is to provide a simple example that illustrates the role of randomness in conducting inference; namely, that each source of randomness leads to a unique uncertainty quantification even for the same point estimate. Therefore, stylized though our assumptions may be, they offer an informative example to communicate this message.

**Setting.** To isolate the role of randomness, we focus on OLS and its minimum $\ell_2$-norm solutions: $\widehat{\boldsymbol{\alpha}} = \boldsymbol{Y}_0^{\dagger}\boldsymbol{y}_T$ and $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Y}_0')^{\dagger}\boldsymbol{y}_N$. Theorem 1 and Corollary 1 establish that the HZ, VT, and DR approaches all yield algebraically equivalent point estimates for this setting. As such, we denote the point estimate as $\widehat{Y}_{NT}(0)$ without any superscripts.

### 4.1. *Model-Based Inference*

Recall (1). The model-based perspective views the potential outcomes, $\{Y_{it}(0), Y_{it}(1)\}$, as stochastic and treatment assignments, $\{A_i, B_t\}$, as fixed. Within this framework, we consider a classical regression model. Though this postulation is certainly not always plausible, it is useful in studying how the assumed source of randomness affects the accuracy of inference that can be conducted.

#### 4.1.1. *Three Generative Models*

We now study properties of $\widehat{Y}_{NT}(0)$ under three different sources of randomness.

**I: HZ model.** The HZ model considers time series patterns as the source of randomness.

ASSUMPTION 1: *We have*

$$Y_{iT} = \sum_{t \le T_0} \alpha_t^* Y_{it} + \varepsilon_{iT}, \quad i = 1, \ldots, N_0. \tag{13}$$

*Here, $\boldsymbol{\alpha}^*$ is a vector of unknown coefficients and $\{\varepsilon_{iT}\}_{i=1}^{N_0}$ are zero mean idiosyncratic errors that are independent over $i = 1, \ldots, N_0$, conditional on $(\boldsymbol{y}_N, \boldsymbol{Y}_0)$.*

Assumption 1 posits the errors are zero mean and conditionally independent across space. The former property implies that the regressors, i.e., lagged outcomes, are uncorrelated with the errors; this is also known as strict exogeneity. Naturally, Assumption 1 motivates the HZ approach, whereby the statistical uncertainty of $\widehat{Y}_{NT}(0)$ is governed by the construction of $\widehat{\boldsymbol{\alpha}}$ from $(\boldsymbol{y}_T, \boldsymbol{Y}_0)$, i.e., the in-sample uncertainty.

**II: VT model.** The VT model considers cross-sectional patterns as the source of randomness.

ASSUMPTION 2: *We have*

$$Y_{Nt} = \sum_{i \le N_0} \beta_i^* Y_{it} + \varepsilon_{Nt}, \quad t = 1, \ldots, T_0. \tag{14}$$

*Here, $\boldsymbol{\beta}^*$ is a vector of unknown coefficients and $\{\varepsilon_{Nt}\}_{t=1}^{T_0}$ are zero mean idiosyncratic errors that are independent over $t = 1, \ldots, T_0$, conditional on $(\boldsymbol{y}_T, \boldsymbol{Y}_0)$.*

Assumption 2 is analogous to Assumption 1. Hence, the statistical uncertainty of $\widehat{Y}_{NT}(0)$ under the VT model is governed by the construction of $\widehat{\boldsymbol{\beta}}$ from $(\boldsymbol{y}_N, \boldsymbol{Y}_0)$.

**III: DR model.** We introduce a new model, the DR model, that considers aspects of the previous HZ and VT models. At a high level, the DR model considers time series and cross-sectional patterns as two distinct sources of randomness.

ASSUMPTION 3: *We have* (13) *and* (14). *Here,* $\{\varepsilon_{iT}\}_{i=1}^{N_0}$ *and* $\{\varepsilon_{Nt}\}_{t=1}^{T_0}$ *have zero mean and are independent over* $i = 1, \ldots, N_0$ *and* $t = 1, \ldots, T_0$, *conditional on* $\boldsymbol{Y}_0$.

Assumption 3 posits that $\boldsymbol{Y}_0$ contains all measured confounders and the errors are independent across both time and units. As a result, the statistical uncertainty of $\widehat{Y}_{NT}(0)$ under the DR model is governed by the constructions of both $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$.

4.1.2. *Model-Based Asymptotic Results on Inference*

Equipped with our three models, we offer a model-based response to Q2. We denote the error covariance matrices as $\boldsymbol{\Sigma}_T^{\text{hz}} = \text{Cov}(\boldsymbol{\varepsilon}_T|\boldsymbol{y}_N, \boldsymbol{Y}_0)$, $\boldsymbol{\Sigma}_N^{\text{vt}} = \text{Cov}(\boldsymbol{\varepsilon}_N|\boldsymbol{y}_T, \boldsymbol{Y}_0)$, $\boldsymbol{\Sigma}_T^{\text{dr}} = \text{Cov}(\boldsymbol{\varepsilon}_T|\boldsymbol{Y}_0)$, and $\boldsymbol{\Sigma}_N^{\text{dr}} = \text{Cov}(\boldsymbol{\varepsilon}_N|\boldsymbol{Y}_0)$, where $\boldsymbol{\varepsilon}_T = [\varepsilon_{iT} : i \leq N_0]$ and $\boldsymbol{\varepsilon}_N = [\varepsilon_{Nt} : t \leq T_0]$. Recalling the SVD of $\boldsymbol{Y}_0$ from Section 3, we denote $\boldsymbol{H}^u = \boldsymbol{U}\boldsymbol{U}'$ and $\boldsymbol{H}^v = \boldsymbol{V}\boldsymbol{V}'$ as the projections onto the columnspace and rowspace of $\boldsymbol{Y}_0$, respectively.

THEOREM 3—Informal (precise statement in Appendix C): *(i)* [HZ model] *Under Assumption 1 and suitable moment conditions, we have as* $N_0 \to \infty$

$$(v_0^{\text{hz}})^{-1/2} \cdot (\widehat{Y}_{NT}(0) - \mu_0^{\text{hz}}) \xrightarrow{d} \mathcal{N}(0, 1),$$

*where* $\mu_0^{\text{hz}} = \langle \boldsymbol{y}_N, \boldsymbol{H}^v \boldsymbol{\alpha}^* \rangle$ *and* $v_0^{\text{hz}} = \widehat{\boldsymbol{\beta}}' \boldsymbol{\Sigma}_T^{\text{hz}} \widehat{\boldsymbol{\beta}}$. *(ii)* [VT model] *Under Assumption 2 and suitable moment conditions, we have as* $T_0 \to \infty$

$$(v_0^{\text{vt}})^{-1/2} \cdot (\widehat{Y}_{NT}(0) - \mu_0^{\text{vt}}) \xrightarrow{d} \mathcal{N}(0, 1),$$

*where* $\mu_0^{\text{vt}} = \langle \boldsymbol{y}_T, \boldsymbol{H}^u \boldsymbol{\beta}^* \rangle$ *and* $v_0^{\text{vt}} = \widehat{\boldsymbol{\alpha}}' \boldsymbol{\Sigma}_N^{\text{vt}} \widehat{\boldsymbol{\alpha}}$. *(iii)* [DR model] *Under Assumption 3 and suitable moment conditions, we have as* $N_0, T_0 \to \infty$

$$(v_0^{\text{dr}})^{-1/2} \cdot (\widehat{Y}_{NT}(0) - \mu_0^{\text{dr}}) \xrightarrow{d} \mathcal{N}(0, 1),$$

*where* $\mu_0^{\text{dr}} = \langle \boldsymbol{\alpha}^*, \boldsymbol{Y}'_0 \boldsymbol{\beta}^* \rangle$ *and*

$$v_0^{\text{dr}} = (\boldsymbol{H}^u \boldsymbol{\beta}^*)' \boldsymbol{\Sigma}_T^{\text{dr}} (\boldsymbol{H}^u \boldsymbol{\beta}^*) + (\boldsymbol{H}^v \boldsymbol{\alpha}^*)' \boldsymbol{\Sigma}_N^{\text{dr}} (\boldsymbol{H}^v \boldsymbol{\alpha}^*) + \text{tr}(\boldsymbol{Y}_0^\dagger \boldsymbol{\Sigma}_T^{\text{dr}} (\boldsymbol{Y}'_0)^\dagger \boldsymbol{\Sigma}_N^{\text{dr}}).$$

Theorem 3 shows that the estimand and variance are controlled by time series patterns under the HZ model, cross-sectional patterns under the VT model, and both patterns under the DR model. It is critical to underline once more that the emphasis of Theorem 3 is *not* on the specific estimands and variances associated with each model, which are, of course, subject to our specific assumptions. Rather, the emphasis is this: *each model measures uncertainty with respect to a distinct estimand.* This message is invariant to the particular assumptions imposed by the researcher. That is, Assumptions 1–3 can be tweaked in numerous ways to yield different estimands and variances than those stated above. Nevertheless, we expect these quantities to change from one source of randomness to another. This clarifies that the assumed source of randomness has substantive implications for conducting inference.

#### 4.1.3. *Model-Based Confidence Intervals*

In Section 5, we look to breathe life into our message above by studying the trade-offs of conducting inference under different sources of randomness through data-inspired simulations and empirical applications. To this end, we construct separate HZ, VT, and DR confidence intervals based on the results in Theorem 3 under homoskedastic and heteroskedastic errors. These intervals are either unbiased or conservative under our assumptions. For ease of exposition, we present their formulations in Appendix C.2.

### 4.2. *Design-Based Inference*

The design-based perspective views the potential outcomes, $\{Y_{it}(0), Y_{it}(1)\}$, as fixed and treatment assignments, $\{A_i, B_t\}$, as stochastic. Within this framework, we consider the assignment mechanisms introduced in Bottmer et al. (2021). As Bottmer et al. (2021) notes, these assumptions are not always plausible, but they underlie the placebo tests that are commonly used in synthetic controls analyses.

#### 4.2.1. *Three Designs*

Let $\boldsymbol{A} \in \{0,1\}^N$ with $\mathbf{1}'\boldsymbol{A} = 1$ and $\boldsymbol{B} \in \{0,1\}^T$ with $\mathbf{1}'\boldsymbol{B} = 1$ be the indicator vectors for the treated unit and treated time period, respectively.

**I: HZ design.** The HZ design considers the treated period to be randomly selected.

ASSUMPTION 4: *We have*

$$\mathbb{P}(\boldsymbol{B} = \boldsymbol{b}) = \begin{cases} 1/T, & \text{if } b_t \in \{0,1\} \ \forall t, \ \mathbf{1}'\boldsymbol{b} = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

**II: VT design.** The VT design considers the treated unit to be randomly selected.

ASSUMPTION 5: *We have*

$$\mathbb{P}(\boldsymbol{A} = \boldsymbol{a}) = \begin{cases} 1/N, & \text{if } a_i \in \{0,1\} \ \forall i, \ \mathbf{1}'\boldsymbol{a} = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

**III: DR design.** The DR design considers both the treated period and treated unit to be randomly selected.

ASSUMPTION 6: *We have* (15) *and* (16)*, where* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are independent.*

#### 4.2.2. *Design-Based Estimator*

To conduct design-based analysis, we consider all possible treatment assignments, not only the realized assignment. Let $Y^*_{it}(0)$ be the OLS fit of $Y_{it}(0)$ using outcomes up to and including time $t$, but not thereafter, i.e., $Y^*_{it}(0) = \langle \boldsymbol{x}, \boldsymbol{W}^\dagger \boldsymbol{z} \rangle$, where $\boldsymbol{W} = [Y_{j\tau} : j \neq i, \tau < t]$, $\boldsymbol{x} = [Y_{i\tau} : \tau < t]$, and $\boldsymbol{z} = [Y_{jt} : j \neq i]$. The design-based estimator is

$$\widehat{Y}(0) = \sum_{i \leq N} \sum_{t \leq T} A_i B_t \cdot Y^*_{it}(0).$$

Again, the stochasticity of $\widehat{Y}(0)$ stems from the assignments since $Y_{it}^*(0)$ is fixed. Hence, the model-based and design-based estimators share the same point estimate for the realized assignment, but differ in their formulations and attributions of randomness.

#### 4.2.3. *Connecting Model-Based and Design-Based Perspectives*

In Table I, we summarize the estimands associated with the model-based and design-based estimators for the three sources of randomness in consideration. We examine the realized $(N, T)$th assignment for concreteness.

TABLE I

MODEL-BASED AND DESIGN-BASED ESTIMANDS UNDER DIFFERENT SOURCES OF RANDOMNESS. WE USE THE SHORTHAND $\tilde{\boldsymbol{\alpha}}^* = \boldsymbol{H}^v \boldsymbol{\alpha}^*$ AND $\tilde{\boldsymbol{\beta}}^* = \boldsymbol{H}^u \boldsymbol{\beta}^*$ AND CONSIDER THE REALIZED $(N, T)$TH ASSIGNMENT.

| Source of Randomness | Model-Based Estimand | Design-Based Estimand |
|---|---|---|
| Time | $\mathbb{E}[\widehat{Y}_{NT}(0)\|\boldsymbol{y}_N, \boldsymbol{Y}_0] = \sum_{t \le T_0} \tilde{\alpha}_t^* Y_{Nt}$ | $\mathbb{E}[\widehat{Y}(0)\|\boldsymbol{A}] = \frac{1}{T}\sum_{t \le T} Y_{Nt}^*(0)$ |
| Unit | $\mathbb{E}[\widehat{Y}_{NT}(0)\|\boldsymbol{y}_T, \boldsymbol{Y}_0] = \sum_{i \le N_0} \tilde{\beta}_i^* Y_{iT}$ | $\mathbb{E}[\widehat{Y}(0)\|\boldsymbol{B}] = \frac{1}{N}\sum_{i \le N} Y_{iT}^*(0)$ |
| Time and Unit | $\mathbb{E}[\widehat{Y}_{NT}(0)\|\boldsymbol{Y}_0] = \sum_{i \le N_0}\sum_{t \le T_0} \alpha_t^* \beta_i^* Y_{it}$ | $\mathbb{E}[\widehat{Y}(0)] = \frac{1}{NT}\sum_{i \le N}\sum_{t \le T} Y_{it}^*(0)$ |

Since the model-based and design-based frameworks attribute randomness differently, their expectations are taken over different probability measures. Nevertheless, the two estimators recover similar estimands for each source of randomness. With time sourced randomness, both estimands are weighted compositions of outcomes across time, which Bottmer et al. (2021) calls the "HZ" effect. With unit sourced randomness, both estimands are weighted compositions of outcomes across units, also called the "VT" effect. Finally, with time and unit sourced randomness, both estimands are weighted compositions of outcomes across time and units, which we coin the "DR" effect.

### 4.3. *From Insights to Practice*

Collectively, Theorem 3 and Table I illustrate that different sources of randomness lead to different estimands and different quantifications of uncertainty even for the same point estimate. The connection between assumptions of randomness and resulting estimands has also been highlighted in related contexts, e.g., Abadie et al. (2020), Bottmer et al. (2021), and Sekhon and Shem-Tov (2021). Translated to practice, these results stress that researchers' assumptions on the source of randomness matter for inference, as they usually do in observational research. As we demonstrate in the next section, these choices are substantively important in the three applications we consider.

## 5. ILLUSTRATIONS

This section illustrates key concepts developed in this article. Our report is based on three canonical synthetic controls studies: (i) terrorism in Basque Country, (ii) California's Proposition 99 (Abadie et al., 2010), and (iii) the reunification of West Germany (Abadie et al., 2015).

### 5.1. *Background on Case Studies*

**Basque study.** See Sections 1–2 for details.

**California study.** This study examines the effect of California's Proposition 99, an anti-tobacco legislation, on its tobacco consumption. The panel data contains per capita cigarette sales of $N = 39$ U.S. states over $T = 31$ years. There are $T_0 = 18$ pretreatment observations and $N_0 = 38$ control units. Our interest is to estimate California's cigarette sales in the absence of Proposition 99.

**West Germany study.** This study examines the economic impact of the 1990 reunification in West Germany. The panel data contains per capita GDP of $N = 17$ countries over $T = 44$ years. There are $T_0 = 30$ pretreatment observations and $N_0 = 16$ control units. Our interest is to estimate West Germany's GDP in the absence of reunification.

### 5.2. *Data-Inspired Simulation Studies*

In an attempt to document our analysis in a realistic environment, we calibrate our simulations to our three studies. As previewed in Section 4.1.3, we conduct a model-based analysis using the confidence intervals derived from Theorem 3. We reiterate that these intervals may not be practical for many real-world settings as they are rooted in our stylized assumptions and asymptotic analysis. Hence, the purpose of this section is not to advocate for their usage. Instead, these intervals are a vehicle to better understand the trade-offs in conducting inference under different sources of randomness.

#### 5.2.1. *Data Generating Process*

We consider the single treated unit and time period setting. Specifically, we consider the actual treated unit, e.g., Basque Country, and focus on the first post-treatment period, e.g., one year after the outset of terrorism; hence, $T = T_0 + 1$. Using the actual data, we generate the underlying regression models as

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{y}_T^* - \boldsymbol{Y}_0^*\boldsymbol{\alpha}\|_2^2 \quad \text{and} \quad \boldsymbol{\beta}^* = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{y}_N^* - (\boldsymbol{Y}_0^*)'\boldsymbol{\beta}\|_2^2,$$

where $\boldsymbol{y}_N^* = [Y_{Nt} : t \le T_0]$, $\boldsymbol{y}_T^* = [Y_{iT} : i \le N_0]$, and $\boldsymbol{Y}_0^* = [Y_{it} : i \le N_0, t \le T_0]$.

Observationally, we have access to $(\boldsymbol{y}_N, \boldsymbol{y}_T, \boldsymbol{Y}_0)$, which are defined as follows: Let $\boldsymbol{Y}_0$ be the rank $r$ approximation of $\boldsymbol{Y}_0^*$, where $r$ is chosen as the minimum number of singular values needed to capture at least 99.9% of $\boldsymbol{Y}_0^*$'s spectral energy. We sample $\boldsymbol{y}_T \sim \mathcal{N}(\boldsymbol{Y}_0\boldsymbol{\alpha}^*, (N_0 - r)^{-1}\|\boldsymbol{y}_T^* - \boldsymbol{Y}_0^*\boldsymbol{\alpha}^*\|_2^2\boldsymbol{I})$ and $\boldsymbol{y}_N \sim \mathcal{N}(\boldsymbol{Y}_0'\boldsymbol{\beta}^*, (T_0 - r)^{-1}\|\boldsymbol{y}_N^* - (\boldsymbol{Y}_0^*)'\boldsymbol{\beta}^*\|_2^2\boldsymbol{I})$. We then define three (latent) estimands: (i) $\mu_0^{\text{hz}} = \langle \boldsymbol{y}_N, \boldsymbol{H}^v\boldsymbol{\alpha}^* \rangle$, (ii) $\mu_0^{\text{vt}} = \langle \boldsymbol{y}_T, \boldsymbol{H}^u\boldsymbol{\beta}^* \rangle$, and (iii) $\mu_0^{\text{dr}} = \langle \boldsymbol{\alpha}^*, \boldsymbol{Y}_0'\boldsymbol{\beta}^* \rangle$, where $(\boldsymbol{H}^u, \boldsymbol{H}^v)$ are computed from $\boldsymbol{Y}_0$. These estimands correspond to Theorem 3 and Table I.

#### 5.2.2. *Simulation Results*

For the purposes of stability, we conduct 5000 replications of the above DGP for each study. In the $\ell$th simulation repeat, we learn the regression coefficients as

$$\widehat{\boldsymbol{\alpha}}^{(\ell)} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{y}_T^{(\ell)} - \boldsymbol{Y}_0\boldsymbol{\alpha}\|_2^2 \quad \text{and} \quad \widehat{\boldsymbol{\beta}}^{(\ell)} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{y}_N^{(\ell)} - \boldsymbol{Y}_0'\boldsymbol{\beta}\|_2^2.$$

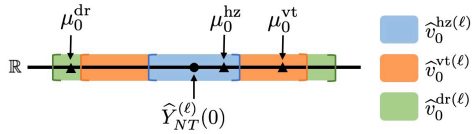



FIGURE 4.—Example illustration of Section 5.2.2 for one simulation repeat. Here, the HZ interval (blue) covers the HZ estimand but undercovers the VT and DR estimands. The VT interval (orange) covers the HZ and VT estimands but undercovers the DR estimand; however, relative to the HZ interval, the VT interval overcovers the HZ estimand. The DR interval (green) covers all estimands but overcovers the HZ and VT estimands relative to their respective intervals.

TABLE II

COVERAGE PROBABILITY FOR NOMINAL 95% CONFIDENCE INTERVALS ACROSS 5000 REPLICATIONS.

| Case study | $\widehat{v}_0^{\mathrm{hz}}$ | | | $\widehat{v}_0^{\mathrm{vt}}$ | | | $\widehat{v}_0^{\mathrm{dr}}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\mu_0^{\mathrm{hz}}$ | $\mu_0^{\mathrm{vt}}$ | $\mu_0^{\mathrm{dr}}$ | $\mu_0^{\mathrm{hz}}$ | $\mu_0^{\mathrm{vt}}$ | $\mu_0^{\mathrm{dr}}$ | $\mu_0^{\mathrm{hz}}$ | $\mu_0^{\mathrm{vt}}$ | $\mu_0^{\mathrm{dr}}$ |
| Basque | **0.92** | 0.74 | 0.66 | 0.99 | **0.92** | 0.87 | 1.00 | 0.97 | **0.94** |
| California | **0.94** | 1.00 | 0.92 | 0.66 | **0.93** | 0.61 | 0.96 | 1.00 | **0.95** |
| W. Germany | **0.93** | 1.00 | 0.91 | 0.48 | **0.94** | 0.46 | 0.95 | 1.00 | **0.93** |

TABLE III

AVERAGE COVERAGE LENGTH FOR NOMINAL 95% CONFIDENCE INTERVALS ACROSS 5000 REPLICATIONS. THE LENGTH IS NORMALIZED BY THE MAGNITUDE OF THE CORRESPONDING POINT ESTIMATE.

| Case study | $\widehat{v}_0^{\mathrm{hz}}$ | $\widehat{v}_0^{\mathrm{vt}}$ | $\widehat{v}_0^{\mathrm{dr}}$ |
|---|---|---|---|
| Basque | 0.02 | 0.03 | 0.04 |
| California | 0.07 | 0.03 | 0.08 |
| W. Germany | 0.03 | 0.01 | 0.03 |

The point estimate is $\widehat{Y}_{NT}^{(\ell)}(0) = \langle \boldsymbol{y}_N^{(\ell)}, \widehat{\boldsymbol{\alpha}}^{(\ell)} \rangle = \langle \boldsymbol{y}_T^{(\ell)}, \widehat{\boldsymbol{\beta}}^{(\ell)} \rangle$. We construct separate HZ, VT, and DR (homoskedastic) confidence intervals, denoted as $(\widehat{v}_0^{\mathrm{hz}(\ell)}, \widehat{v}_0^{\mathrm{vt}(\ell)}, \widehat{v}_0^{\mathrm{dr}(\ell)})$, centered around the point estimate. The estimands do not change in our replications.

In Tables II and III, we report the coverage probabilities (CP) and average lengths (AL), respectively, for each confidence interval with respect to each estimand at the 95% nominal mark across all simulation repeats; see Figure 4 for an illustration of one repeat. Across all three studies and with respect to $\mu_0^{\mathrm{hz}}$, the coverage of the HZ interval is closer to the nominal coverage than that of the VT and DR intervals as the latter two can substantially under- or over-cover. This storyline is consistent for the VT interval with respect to $\mu_0^{\mathrm{vt}}$ and the mixed interval with respect to $\mu_0^{\mathrm{dr}}$.

Collectively, our formal results and simulations demonstrate that (i) the choice of estimand directly affects the accuracy of inference; and (ii) the variance formulas developed for one estimand may not have the correct coverage for another estimand. Accordingly, researchers should carefully consider the source of randomness in their data as it can have a significant influence over their ability to conduct valid inference. These conclusions are in line with those drawn in Sekhon and Shem-Tov (2021), which analyzes the classical difference-in-means estimator with respect to standard estimands for randomized control trials.

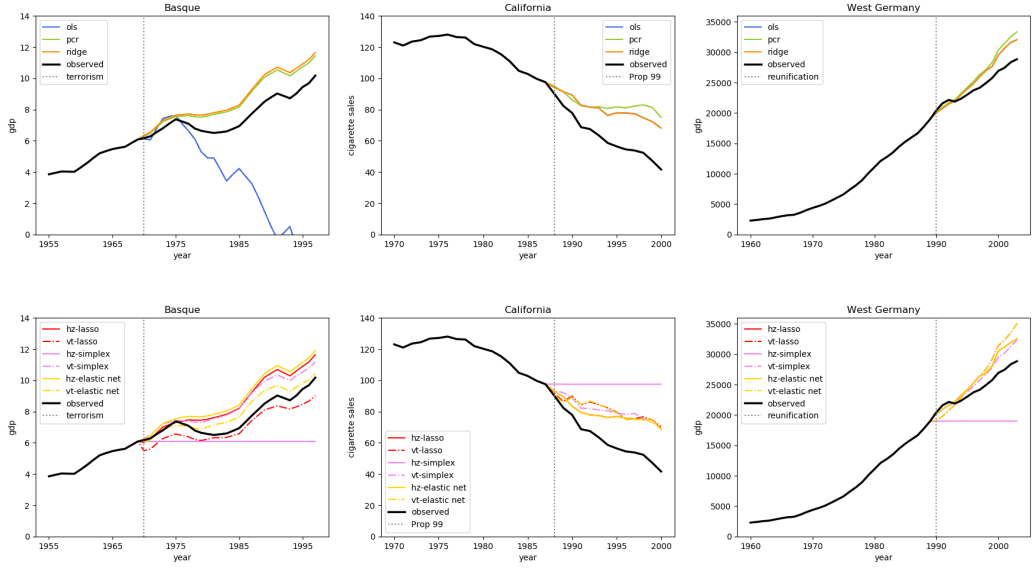



FIGURE 5.—Top and bottom figures correspond to symmetric and asymmetric estimators, respectively. From left to right, the figures are indexed by the Basque, California, and West Germany studies. Across all figures, the treated year is the dotted vertical line; the observed trajectory is in solid black; and the HZ and VT counterfactual trajectories are colored solid and dashed-dotted lines, respectively.

### 5.3. *Empirical Applications*

Next, we analyze our three case studies of interest. All estimators are trained on pretreatment data only, and the point and variance estimation formulas are separately applied for the treated unit at each post-treatment period $t > T_0$. We continue to use the confidence intervals from Section 4.1.3.

#### 5.3.1. *Implementation Details*

For ridge, lasso, and elastic net regressions, we use the default `scikit-learn` hyperparameters $(\lambda_1, \lambda_2)$. For PCR, we choose the number of principal components $k$ via the approach described in Section 5.2. This yields $k = 2$ for the Basque study, $k = 3$ for the California study, and $k = 4$ for the West Germany study. We implement simplex regression using the code made available at https://matheusfacure.github.io/python-causality-handbook/15-Synthetic-Control.html.

#### 5.3.2. *Point Estimation*

Figure 5 visualizes the counterfactual trajectories generated by the estimators in Section 3.1. Our findings reinforce Theorems 1 and 2. On a separate note, we observe that within the Basque study, the OLS estimates are wildly different from the others—likely due to overfitting—and HZ simplex regression reduces to the last observation carried forward (LOCF) estimator. In the California and West Germany studies, the estimates are all qualitatively similar with the exception of the HZ simplex regression, which again reduces to LOCF. In fact, the OLS and ridge estimates appear to overlap, as well as the lasso and elastic net estimates.

### 5.3.3. *Inference*

We present the OLS-based confidence intervals in Figure 6.[2] As a final reminder, the emphasis of Figure 6 is that the intervals associated with each model can vary in width; we do not put any stock in the specific magnitudes of these widths. With this in mind, consider the Basque study. The top row of plots shows that $\mu_0^{\text{vt}}$ is more accurately estimated than both $\mu_0^{\text{hz}}$ and $\mu_0^{\text{dr}}$. Put differently, there is less uncertainty about conducting inference on $\mu_0^{\text{vt}}$ relative to the other estimands. At the same time, these plots indicate that if $\mu_0^{\text{hz}}$ or $\mu_0^{\text{dr}}$ are the estimands of interest, then the VT confidence interval will undercover in both settings. Analogous statements apply to the remaining subfigures. As with our simulations, the large potential differences in coverage reinforce the importance of properly reasoning through the source of randomness in the data.

## 6. CONCLUSION

This article contributes to panel data analysis in two primary ways: (i) we prove that HZ, VT, and DR approaches—while seemingly very different—all yield algebraically identical point estimates for several standard settings, i.e., these approaches can be equivalent ways of looking at the same problem in the absence of any additional considerations; (ii) further, we demonstrate that even though these approaches may share the same point estimate, the source of randomness assumed by each approach leads to different estimands and different quantifications of uncertainty.

Our results show that assumptions made by researchers that may appear arbitrary about the source of randomness result in different inferences. This is expected in observational work because no randomization was actually implemented. Nevertheless, it is important to check the sensitivity of reported results to these randomness source assumptions. A potentially fruitful path forward is to build upon the principles of predictability, computability, and stability (PCS) (Yu and Kumbier, 2020) to create measures that incorporate our uncertainty over randomness source assumptions. We leave it to future work to formalize a treatment of this problem.

## REFERENCES


Abadie, Alberto (2021): "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, 59, 391–425. [7]

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge (2020): "Sampling-Based versus Design-Based Uncertainty in Regression Analysis," *Econometrica*, 88, 265–296. [14]

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of Californiaâs Tobacco Control Program," *Journal of the American Statistical Association*, 105. [7, 14]

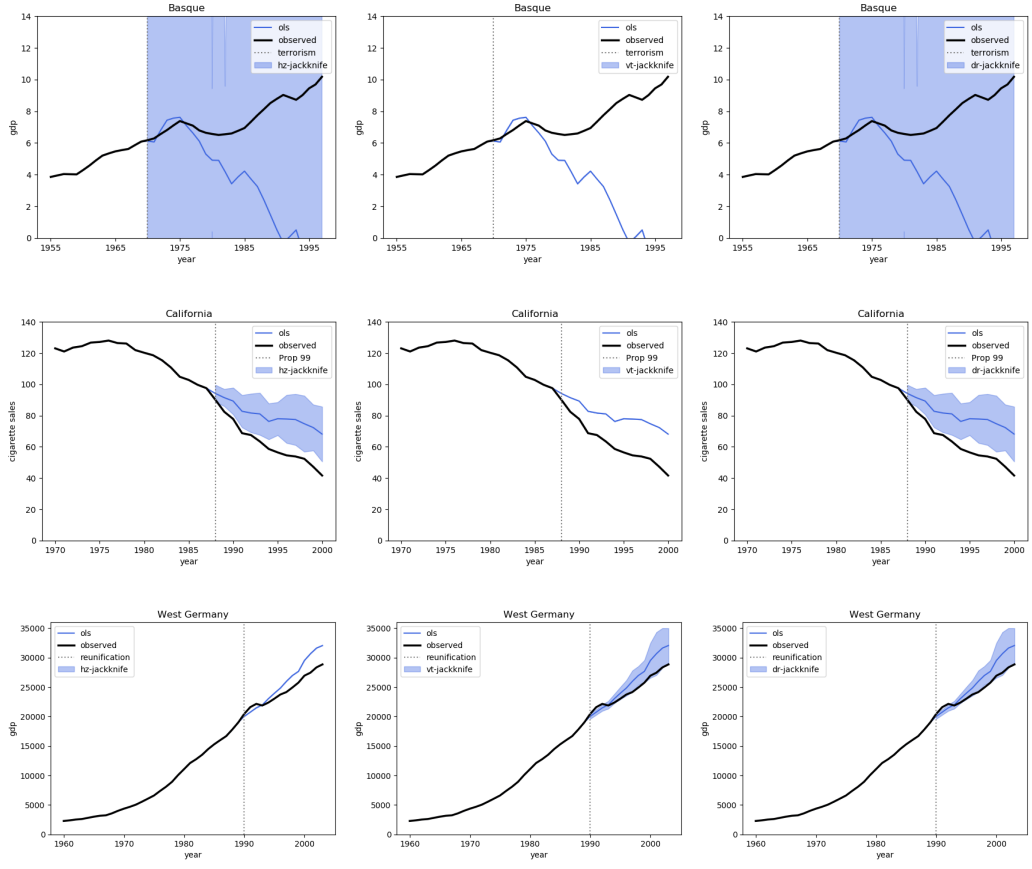
——— (2015): "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510. [5, 7, 14]

Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132. [1, 2, 4, 7]

Abadie, Alberto and Jérémy L'Hour (2021): "A Penalized Synthetic Control Estimator for Disaggregated Data," *Journal of the American Statistical Association*, 116, 1817–1834. [7, 9]

Agarwal, Anish, Devavrat Shah, and Dennis Shen (2021): "Synthetic Interventions," *arXiv preprint arXiv:2006.07691*. [6, 28, 29, 35]


[2] As Figure 6 shows, OLS-based confidence intervals can suffer from degeneracy due to zero in-sample residuals (overfitting). One rectification is to substitute OLS with PCR. In Appendix E, we discuss advantages of PCR over OLS and present inferential results for PCR.




(a) HZ model. (b) VT model. (c) DR model.

FIGURE 6.—OLS estimates with jackknife confidence intervals. From top to bottom, the rows are indexed by the Basque, California, and West Germany studies. From left to right, the columns are indexed by the HZ, VT, and DR models. The VT confidence intervals for Basque and California, and HZ confidence interval for West Germany, are degenerate due to zero in-sample residuals from overfitting.

AMJAD, MUHAMMAD, VISHAL MISRA, DEVAVRAT SHAH, AND DENNIS SHEN (2019): "MRSC: Multi-Dimensional Robust Synthetic Control," *Proc. ACM Meas. Anal. Comput. Syst.*, 3. [6]

AMJAD, MUHAMMAD, DEVAVRAT SHAH, AND DENNIS SHEN (2018): "Robust Synthetic Control," *Journal of Machine Learning Research*, 19, 1–51. [6]

ARKHANGELSKY, DMITRY, SUSAN ATHEY, DAVID A. HIRSHBERG, GUIDO W. IMBENS, AND STEFAN WAGER (2021): "Synthetic Difference-in-Differences," *American Economic Review*, 111, 4088–4118. [8, 9]

ASHENFELTER, ORLEY (1978): "Estimating the Effect of Training Programs on Earnings," *The Review of Economics and Statistics*, 60, 47–57. [8]

ATHEY, SUSAN, MOHSEN BAYATI, NIKOLAY DOUDCHENKO, GUIDO IMBENS, AND KHASHAYAR KHOSRAVI (2021): "Matrix Completion Methods for Causal Panel Data Models," *Journal of the American Statistical Association*, 116, 1716–1730. [2, 5]

ATHEY, SUSAN AND GUIDO W. IMBENS (2017): "The State of Applied Econometrics: Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31, 3–32. [2]

BEN-MICHAEL, ELI, AVI FELLER, AND JESSE ROTHSTEIN (2021): "The Augmented Synthetic Control Method," *Journal of the American Statistical Association*, 116, 1789–1803. [6, 7, 9]

BOTTMER, LEA, GUIDO IMBENS, JANN SPIESS, AND MERRILL WARNICK (2021): "A Design-Based Perspective on Synthetic Control Methods," . [13, 14]

Carvalho, Carlos, Ricardo Masini, and Marcelo C. Medeiros (2018): "ArCo: An artificial counterfactual approach for high-dimensional panel time-series data," *Journal of Econometrics*, 207, 352–380. [6]

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu (2021): "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," *Journal of the American Statistical Association*, 116, 1849–1864. [6]

Cline, Randall E. (1965): "Representations for the Generalized Inverse of Sums of Matrices," *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis*, 2, 99–114. [32]

Doudchenko, Nikolay and Guido W Imbens (2016): "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis," Working Paper 22791, National Bureau of Economic Research. [5, 6]

Gunasekar, Suriya, Blake Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nathan Srebro (2017): "Implicit Regularization in Matrix Factorization," in *Advances in Neural Information Processing Systems*. [8]

Hartley, H. O., J. N. K. Rao, and Grace Kiefer (1969): "Variance Estimation with One Unit per Stratum," *Journal of the American Statistical Association*, 64, 841–851. [29, 34]

Hoff, Peter D. (2017): "Lasso, fractional norm and structured sparse estimation using a Hadamard product parametrization," *Computational Statistics & Data Analysis*, 115, 186–198. [21, 24]

Hsiao, Cheng, H. Steve Ching, and Shui Ki Wan (2012): "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong with Mainland China," *Journal of Applied Econometrics*, 27, 705–740. [6]

Imbens, Guido W. and Jeffrey M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47, 5–86. [2]

Lehmann, E.L. (2000): "Elements of Large-Sample Theory," . [25]

Li, Kathleen T. (2020): "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods," *Journal of the American Statistical Association*, 115, 2068–2083. [6]

Li, Kathleen T. and David R. Bell (2017): "Estimation of average treatment effects with panel data: Asymptotic theory and implementation," *Journal of Econometrics*, 197, 65–75. [5, 6]

Meyer, Carl D. (1973): "Generalized Inversion of Modified Matrices," *SIAM Journal on Applied Mathematics*, 24, 315–323. [32]

Neyshabur, Behnam, Ryota Tomioka, and Nathan Srebro (2015): "In Search of the Real Inductive Bias: On the Role of Implicit Regularization in Deep Learning," in *International Conference on Learning Representations*. [8]

Robins, James, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky (2007): "Comment: Performance of Double-Robust Estimators When "Inverse Probability" Weights Are Highly Variable," *Statistical Science*, 22, 544 – 559. [9]

Rosenbaum, PAUL and Donald Rubin (1983): "The Central Role of the Propensity Score in Observational Studies For Causal Effects," *Biometrika*, 70, 41–55. [2]

Sekhon, Jasjeet S. and Yotam Shem-Tov (2021): "Inference on a New Class of Sample Average Treatment Effects," *Journal of the American Statistical Association*, 116, 798–804. [14, 16]

Styan, George P.H. (1973): "Hadamard products and multivariate statistical analysis," *Linear Algebra and its Applications*, 6, 217–240. [24]

Tibshirani, Ryan J. (2013): "The lasso problem and uniqueness," *Electronic Journal of Statistics*, 7, 1456 – 1490. [6]

Varga, Richard S. (1962): *Matrix Iterative Analysis*, Prentice-Hall Series in Automatic Computation, Englewood Cliffs: Prentice-Hall. [30]

Yu, Bin and Karl Kumbier (2020): "Veridical data science," *Proceedings of the National Academy of Sciences*, 117, 3920–3929. [18]

## APPENDIX A: Proofs for Point Estimation

### A.1. *Proof of Theorem 1*

Proof: (i) [OLS] Consider HZ regression. The optimality conditions state

$$\nabla_{\boldsymbol{\alpha}} \|\boldsymbol{y}_T - \boldsymbol{Y}_0 \boldsymbol{\alpha}\|_2^2 = 0.$$

Solving for $\boldsymbol{\alpha}$, we derive $\boldsymbol{Y}_0' \boldsymbol{Y}_0 \boldsymbol{\alpha} = \boldsymbol{Y}_0' \boldsymbol{y}_T$. Using the pseudoinverse, we obtain $\widehat{\boldsymbol{\alpha}} = (\boldsymbol{Y}_0' \boldsymbol{Y}_0)^\dagger \boldsymbol{Y}_0' \boldsymbol{y}_T = \boldsymbol{Y}_0^\dagger \boldsymbol{y}_T$. Thus, the HZ prediction is given by $\widehat{Y}_{NT}^{\text{hz}}(0) = \langle \boldsymbol{y}_N, \boldsymbol{Y}_0^\dagger \boldsymbol{y}_T \rangle$.

Following the arguments above for VT regression, we obtain $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Y}'_0)^\dagger \boldsymbol{y}_N$ and $\widehat{Y}^{\text{vt}}_{NT}(0) = \langle \boldsymbol{y}_T, (\boldsymbol{Y}'_0)^\dagger \boldsymbol{y}_N \rangle$. Given that $(\boldsymbol{Y}'_0)^\dagger = (\boldsymbol{Y}^\dagger_0)'$, the proof for OLS is complete.

(ii) [PCR] The proof follows that of OLS with $\boldsymbol{Y}^{(k)}_0$ in place of $\boldsymbol{Y}_0$.

(iii) [Ridge] Following the proof for OLS, we obtain $\widehat{Y}^{\text{hz}}_{NT}(0) = \langle \boldsymbol{y}_N, (\boldsymbol{Y}'_0\boldsymbol{Y}_0 + \lambda_2 \boldsymbol{I})^{-1}\boldsymbol{Y}'_0\boldsymbol{y}_T \rangle$ and $\widehat{Y}^{\text{vt}}_{NT}(0) = \langle \boldsymbol{y}_T, (\boldsymbol{Y}_0\boldsymbol{Y}'_0 + \lambda_2 \boldsymbol{I})^{-1}\boldsymbol{Y}_0\boldsymbol{y}_N \rangle$. Observing $(\boldsymbol{Y}'_0\boldsymbol{Y}_0 + \lambda \boldsymbol{I})^{-1}\boldsymbol{Y}'_0 = \boldsymbol{Y}'_0(\boldsymbol{Y}_0\boldsymbol{Y}'_0 + \lambda \boldsymbol{I})^{-1}$ completes the proof. *Q.E.D.*

## A.2. *Proof of Theorem 2*

To prove Theorem 2, we first establish our results for lasso and elastic net in Appendix A.2.1 and simplex regression in Appendix A.2.2, and then assemble them together in Appendix A.2.3.

### A.2.1. *Lasso & Elastic Net Regressions*

We first establish a general result in Lemma 1 for $\ell_p$-penalties, where $p = 2/K$ and $K$ is an integer $\geq 1$, based on the contributions of Hoff (2017). More formally, consider

(a) HZ regression: for $K \geq 1$ and $\lambda > 0$,

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\text{argmin}} \, \|\boldsymbol{y}_T - \boldsymbol{Y}_0\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_p^p, \quad \text{with} \quad \widehat{Y}^{\text{hz}}_{NT}(0) = \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}} \rangle. \tag{17}$$

(b) VT regression: for $K \geq 1$ and $\lambda > 0$,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \, \|\boldsymbol{y}_N - \boldsymbol{Y}'_0\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_p^p, \quad \text{with} \quad \widehat{Y}^{\text{vt}}_{NT}(0) = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}} \rangle. \tag{18}$$

Note that $K = 2$ yields lasso regression while $K > 2$ yields non-convex penalties. As such, we will use Lemma 1 to establish our results for lasso and elastic net regressions. We relegate the proof of Lemma 1 to Appendix A.2.4.

LEMMA 1: *For any $K \geq 1$ and $\lambda > 0$, a* HZ *and* VT *regression solution is*

$$\widehat{Y}^{\text{hz}}_{NT}(0) = \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}}_1 \circ \cdots \circ \widehat{\boldsymbol{\alpha}}_K \rangle, \quad \textit{and} \quad \widehat{Y}^{\text{vt}}_{NT}(0) = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}_1 \circ \cdots \circ \widehat{\boldsymbol{\beta}}_K \rangle,$$

*where for every $k \leq K$,*

$$\widehat{\boldsymbol{\alpha}}_k = \left( \boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})\boldsymbol{Y}'_0\boldsymbol{Y}_0\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k}) + \frac{\lambda}{K}\boldsymbol{I} \right)^{-1} \boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})\boldsymbol{Y}'_0\boldsymbol{y}_T,$$

$$\widehat{\boldsymbol{\beta}}_k = \left( \boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{\sim k})\boldsymbol{Y}_0\boldsymbol{Y}'_0\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{\sim k}) + \frac{\lambda}{K}\boldsymbol{I} \right)^{-1} \boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{\sim k})\boldsymbol{Y}_0\boldsymbol{y}_N,$$

*$\widehat{\boldsymbol{\alpha}}_{\sim k} = \widehat{\boldsymbol{\alpha}}_1 \circ \cdots \circ \widehat{\boldsymbol{\alpha}}_{k-1} \circ \widehat{\boldsymbol{\alpha}}_{k+1} \circ \cdots \circ \widehat{\boldsymbol{\alpha}}_K$, $\widehat{\boldsymbol{\beta}}_{\sim k} = \widehat{\boldsymbol{\beta}}_1 \circ \cdots \circ \widehat{\boldsymbol{\beta}}_{k-1} \circ \widehat{\boldsymbol{\beta}}_{k+1} \circ \cdots \circ \widehat{\boldsymbol{\beta}}_K$, and $\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})$ and $\boldsymbol{D}(\widehat{\boldsymbol{\beta}} \sim k)$ are diagonal matrices formed from $\widehat{\boldsymbol{\alpha}}_{\sim k}$ and $\widehat{\boldsymbol{\beta}}_{\sim k}$, respectively.*

LEMMA 2—Lasso & Elastic Net Regressions: HZ $\neq$ VT *for (i) lasso and (ii) elastic net.*

PROOF: (i) [Lasso] By Lemma 1 for $K=2$ and $\lambda=\lambda_1>0$, a HZ regression solution is

$$\widehat{Y}^{\text{hz}}_{NT}(0)=\langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}}_1\circ\widehat{\boldsymbol{\alpha}}_2\rangle, \tag{19}$$

where $\widehat{\boldsymbol{\alpha}}_{1+k}=\left(\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{2-k})\boldsymbol{Y}_0'\boldsymbol{Y}_0\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{2-k})+(\lambda_1/2)\boldsymbol{I}\right)^{-1}\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{2-k})\boldsymbol{Y}_0'\boldsymbol{y}_T$ for $k\in\{0,1\}$.

Similarly, a VT regression solution is given by

$$\widehat{Y}^{\text{vt}}_{NT}(0)=\langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}_1\circ\widehat{\boldsymbol{\beta}}_2\rangle, \tag{20}$$

where $\widehat{\boldsymbol{\beta}}_{1+k}=(\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{2-k})\boldsymbol{Y}_0\boldsymbol{Y}_0'\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{2-k})+(\lambda_1/2)\boldsymbol{I})^{-1}\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{2-k})\boldsymbol{Y}_0\boldsymbol{y}_N$ for $k\in\{0,1\}$.

From (19) and (20), we see that the HZ regression solution can be linear in $y$ and at least quadratic in $q$. On the other hand, the VT regression solution can be linear in $q$ and at least quadratic in $y$. Since the lasso solution is unique under the assumption the entries of $\boldsymbol{Y}_0$ are drawn from a continuous distribution, this implies that HZ and VT regressions do not yield matching estimates in general.

(ii) [Elastic net] Consider HZ regression. We rewrite (2) in a lasso formulation:

$$\widehat{\boldsymbol{\alpha}}^*=\underset{\boldsymbol{\alpha}^*}{\text{argmin}}\,\|\boldsymbol{y}_T^*-\boldsymbol{Y}_0^*\boldsymbol{\alpha}^*\|_2^2+\lambda^*\|\boldsymbol{\alpha}^*\|_1, \tag{21}$$

where $q^*=(\boldsymbol{y}_T',\boldsymbol{0}')'$, $\boldsymbol{Y}_0^*=(1+\lambda_2)^{-1/2}(\boldsymbol{Y}_0',\sqrt{\lambda_2}\boldsymbol{I})'$, $\lambda^*=(1+\lambda_2)^{-1/2}\lambda_1$, and $\boldsymbol{\alpha}^*=(1+\lambda_2)^{1/2}\boldsymbol{\alpha}$. We apply Lemma 1 to (21) with $K=2$ and $\lambda=\lambda^*>0$ to obtain

$$\widehat{Y}^{\text{hz}}_{NT}(0)=\frac{\langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}}_1^*\circ\widehat{\boldsymbol{\alpha}}_2^*\rangle}{\sqrt{1+\lambda_2}}, \tag{22}$$

where $\widehat{\boldsymbol{\alpha}}^*_{1+k}=\left((1+\lambda_2)^{-1/2}\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}^*_{2-k})(\boldsymbol{Y}_0'\boldsymbol{Y}_0+\lambda_2\boldsymbol{I})\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}^*_{2-k})+\frac{\lambda_1}{2}\boldsymbol{I}\right)^{-1}\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}^*_{2-k})\boldsymbol{Y}_0'\boldsymbol{y}_T$ for $k\in\{0,1\}$.

Similarly, for VT regression, we proceed as above to obtain

$$\widehat{Y}^{\text{vt}}_{NT}(0)=\frac{\langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}_1^*\circ\widehat{\boldsymbol{\beta}}_2^*\rangle}{\sqrt{1+\lambda_2}}, \tag{23}$$

where $\widehat{\boldsymbol{\beta}}^*_{1+k}=\left((1+\lambda_2)^{-1/2}\boldsymbol{D}(\widehat{\boldsymbol{\beta}}^*_{2-k})(\boldsymbol{Y}_0\boldsymbol{Y}_0'+\lambda_2\boldsymbol{I})\boldsymbol{D}(\widehat{\boldsymbol{\beta}}^*_{2-k})+\frac{\lambda_1}{2}\boldsymbol{I}\right)^{-1}\boldsymbol{D}(\widehat{\boldsymbol{\beta}}^*_{2-k})\boldsymbol{Y}_0\boldsymbol{y}_N$ for $k\in\{0,1\}$.

From (22) and (23), we see that the HZ solution can be linear in $y$ and at least quadratic in $q$. On the other hand, the VT solution can be linear in $q$ and at least quadratic in $y$. Since the elastic net regression solution is unique, this implies that HZ and VT regressions do not yield matching estimates in general. *Q.E.D.*

A.2.2. *Simplex Regression*

LEMMA 3—Simplex regression: HZ $\neq$ VT *for simplex regression.*

PROOF: Consider HZ regression. We write the Lagrangian of (4) as

$$\widehat{\boldsymbol{\alpha}}=\underset{\boldsymbol{\alpha}}{\text{argmin}}\;\|\boldsymbol{y}_T-\boldsymbol{Y}_0\boldsymbol{\alpha}\|_2^2+\lambda\|\boldsymbol{\alpha}\|_2^2-(\boldsymbol{\theta}^{\text{hz}})'\boldsymbol{\alpha}+\nu^{\text{hz}}(\boldsymbol{1}'\boldsymbol{\alpha}-1),$$

where $\boldsymbol{\theta}^{\text{hz}} \in \mathbb{R}^{T_0}$ and $\nu^{\text{hz}} \in \mathbb{R}$. By the KKT conditions, optimality is achieved if the following are satisfied: (i) $\widehat{\boldsymbol{\alpha}} \succeq \mathbf{0}$ and $\mathbf{1}'\widehat{\boldsymbol{\alpha}} = 1$; (ii) $\widehat{\boldsymbol{\theta}}^{\text{hz}} \succeq \mathbf{0}$; (iii) $\widehat{\theta}_i^{\text{hz}}\widehat{\alpha}_i = 0$ for $i = 1, \ldots, T_0$; (iv) $\widehat{\boldsymbol{\alpha}} = (\boldsymbol{Y}_0'\boldsymbol{Y}_0 + \lambda \boldsymbol{I})^{-1}(\boldsymbol{Y}_0'\boldsymbol{y}_T + (1/2)\widehat{\boldsymbol{\theta}}^{\text{hz}} - (\widehat{\nu}^{\text{hz}}/2)\mathbf{1})$. Thus, given primal and dual feasible variables $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\theta}}^{\text{hz}}, \widehat{\nu}^{\text{hz}})$, we can write the final HZ prediction as

$$\widehat{Y}_{NT}^{\text{hz}}(0) = \widehat{Y}_{NT}^{\text{hz,ols}}(0) + (1/2)\boldsymbol{y}_N'(\boldsymbol{Y}_0'\boldsymbol{Y}_0 + \lambda \boldsymbol{I})^{-1}(\widehat{\boldsymbol{\theta}}^{\text{hz}} - \widehat{\nu}^{\text{hz}}\mathbf{1}),$$

where $\widehat{Y}_{NT}^{\text{hz,ols}}(0) = \boldsymbol{y}_N'(\boldsymbol{Y}_0'\boldsymbol{Y}_0 + \lambda \boldsymbol{I})^{-1}\boldsymbol{Y}_0'\boldsymbol{y}_T$ converges to the prediction corresponding to the OLS solution with minimum $\ell_2$-norm as $\lambda \to 0^+$. Similarly, for VT regression, the KKT conditions are (i) $\widehat{\boldsymbol{\beta}} \succeq \mathbf{0}$ and $\mathbf{1}'\widehat{\boldsymbol{\beta}} = 1$; (ii) $\widehat{\boldsymbol{\theta}}^{\text{vt}} \succeq \mathbf{0}$; (iii) $\widehat{\theta}_i^{\text{vt}}\widehat{\beta}_i = 0$ for $i = 1, \ldots, N_0$; (iv) $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Y}_0\boldsymbol{Y}_0' + \lambda \boldsymbol{I})^{-1}(\boldsymbol{Y}_0\boldsymbol{y}_N + (1/2)\widehat{\boldsymbol{\theta}}^{\text{vt}} - (\widehat{\nu}^{\text{vt}}/2)\mathbf{1})$. For primal and dual feasible variables $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}^{\text{vt}}, \widehat{\nu}^{\text{vt}})$, this yields

$$\widehat{Y}_{NT}^{\text{vt}}(0) = \widehat{Y}_{NT}^{\text{vt,ols}}(0) + (1/2)\boldsymbol{y}_T'(\boldsymbol{Y}_0\boldsymbol{Y}_0' + \lambda \boldsymbol{I})^{-1}(\widehat{\boldsymbol{\theta}}^{\text{vt}} - \widehat{\nu}^{\text{vt}}\mathbf{1}),$$

where $\widehat{Y}_{NT}^{\text{vt,ols}}(0) = \boldsymbol{y}_T'(\boldsymbol{Y}_0\boldsymbol{Y}_0' + \lambda \boldsymbol{I})^{-1}\boldsymbol{Y}_0\boldsymbol{y}_N$ converges to the prediction corresponding to the OLS solution with minimum $\ell_2$-norm as $\lambda \to 0^+$. Notably, as per Theorem 1, $\widehat{Y}_{NT}^{\text{hz,ols}}(0) = \widehat{Y}_{NT}^{\text{vt,ols}}(0) = \widehat{Y}_{NT}^{\text{ols}}(0)$ for any $\lambda \geq 0$. As a result,

$$\widehat{Y}_{NT}^{\text{hz}}(0) = \widehat{Y}_{NT}^{\text{ols}}(0) + (1/2)\boldsymbol{y}_N'(\boldsymbol{Y}_0'\boldsymbol{Y}_0 + \lambda \boldsymbol{I})^{-1}(\widehat{\boldsymbol{\theta}}^{\text{hz}} - \widehat{\nu}^{\text{hz}}\mathbf{1}) \tag{24}$$

$$\widehat{Y}_{NT}^{\text{vt}}(0) = \widehat{Y}_{NT}^{\text{ols}}(0) + (1/2)\boldsymbol{y}_T'(\boldsymbol{Y}_0\boldsymbol{Y}_0' + \lambda \boldsymbol{I})^{-1}(\widehat{\boldsymbol{\theta}}^{\text{vt}} - \widehat{\nu}^{\text{vt}}\mathbf{1}). \tag{25}$$

As seen from (24) and (25), the leading terms in the HZ and VT simplex regression predictions are identical. The remaining terms, however, can differ from one another. As an example, consider $N = T$ with $\boldsymbol{Y}_0 = \boldsymbol{I}$, $\boldsymbol{y}_N = \mathbf{0}$, $\boldsymbol{y}_T = (1+\lambda)(\widehat{\boldsymbol{\theta}}^{\text{vt}} - \widehat{\nu}^{\text{vt}}\mathbf{1})$. By construction, observe that $\widehat{\boldsymbol{\beta}} = (2(1+\lambda))^{-1}(\widehat{\boldsymbol{\theta}}^{\text{vt}} - \widehat{\nu}^{\text{vt}}\mathbf{1})$. Recall from the KKT conditions for VT regression that $\widehat{\boldsymbol{\beta}} \succeq \mathbf{0}$ and $\mathbf{1}'\widehat{\boldsymbol{\beta}} = 1$. Therefore, at least one entry of $(\widehat{\boldsymbol{\theta}}^{\text{vt}} - \widehat{\nu}^{\text{vt}}\mathbf{1})$ must be strictly positive. This yields

$$(1+\lambda)^{-1}\boldsymbol{y}_T'(\widehat{\boldsymbol{\theta}}^{\text{vt}} - \widehat{\nu}^{\text{vt}}\mathbf{1}) = (\widehat{\boldsymbol{\theta}}^{\text{vt}} - \widehat{\nu}^{\text{vt}}\mathbf{1})'(\widehat{\boldsymbol{\theta}}^{\text{vt}} - \widehat{\nu}^{\text{vt}}\mathbf{1}) > 0.$$

Combining the above, we obtain $\widehat{Y}_{NT}^{\text{hz}}(0) = 0$ and $\widehat{Y}_{NT}^{\text{vt}}(0) > 0$. *Q.E.D.*

A.2.3. *Putting Everything Together—Proof of Theorem 2*

PROOF: The proof is immediate from Lemmas 2 and 3. *Q.E.D.*

A.2.4. *Proof of Lemma 1: $\ell_p$-penalties*

PROOF: We recall the Hadamard product parametrization (HPP): for any vector $\boldsymbol{z}$ and integer $K \geq 1$, $\|\boldsymbol{z}\|_p^p = \min_{\boldsymbol{z}_1 \circ \cdots \circ \boldsymbol{z}_K = \boldsymbol{z}} (1/K)\sum_{k=1}^K \|\boldsymbol{z}_k\|_2^2$, where $\circ$ denotes the Hadamard (componentwise) product. We rewrite our subclass of $\ell_p$-penalties, i.e., (17) and (18), as sums of $\ell_2$-penalties via the HPP technique:

$$(\widehat{\boldsymbol{\alpha}}_1, \ldots, \widehat{\boldsymbol{\alpha}}_K) = \underset{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K}{\operatorname{argmin}} \|\boldsymbol{y}_T - \boldsymbol{Y}_0(\boldsymbol{\alpha}_1 \circ \cdots \circ \boldsymbol{\alpha}_K)\|_2^2 + \frac{\lambda}{K}\sum_{k=1}^K \|\boldsymbol{\alpha}_k\|_2^2 \tag{26}$$

$$(\widehat{\boldsymbol{\beta}}_1,\ldots,\widehat{\boldsymbol{\beta}}_K) = \underset{\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_K}{\operatorname{argmin}} \|\boldsymbol{y}_N - \boldsymbol{Y}'_0(\boldsymbol{\beta}_1 \circ \cdots \circ \boldsymbol{\beta}_K)\|_2^2 \;+\; \frac{\lambda}{K}\sum_{k=1}^{K}\|\boldsymbol{\beta}_k\|_2^2, \tag{27}$$

where $\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\alpha}}_1 \circ \cdots \circ \widehat{\boldsymbol{\alpha}}_K$ and $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_1 \circ \cdots \circ \widehat{\boldsymbol{\beta}}_K$. Below, we leverage the results of Hoff (2017), which provides an alternating ridge regression algorithm to solve for (26)–(27).

Consider HZ regression. Let us solve for $\boldsymbol{\alpha}_k$ by fixing $\boldsymbol{\alpha}_{k'}$ for $k' \neq k$. To begin, observe that $(\boldsymbol{\alpha}_1 \circ \cdots \circ \boldsymbol{\alpha}_K)'\boldsymbol{Y}'_0\boldsymbol{Y}_0(\boldsymbol{\alpha}_1 \circ \cdots \circ \boldsymbol{\alpha}_K) = \boldsymbol{\alpha}'_k(\boldsymbol{Y}'_0\boldsymbol{Y}_0 \circ \boldsymbol{\alpha}_{\sim k}\boldsymbol{\alpha}'_{\sim k})\boldsymbol{\alpha}_k$ and $(\boldsymbol{\alpha}_1 \circ \cdots \circ \boldsymbol{\alpha}_K)'\boldsymbol{Y}'_0\boldsymbol{y}_T = \boldsymbol{\alpha}'_k(\boldsymbol{\alpha}_{\sim k} \circ \boldsymbol{Y}'_0\boldsymbol{y}_T)$, where $\boldsymbol{\alpha}_{\sim k} = \boldsymbol{\alpha}_1 \circ \cdots \circ \boldsymbol{\alpha}_{k-1} \circ \boldsymbol{\alpha}_{k+1} \circ \cdots \circ \boldsymbol{\alpha}_K$. This allows us to write the optimality conditions as

$$\nabla_{\boldsymbol{\alpha}_k}\left\{\boldsymbol{\alpha}'_k\left(\boldsymbol{Y}'_0\boldsymbol{Y}_0 \circ \boldsymbol{\alpha}_{\sim k}\boldsymbol{\alpha}'_{\sim k} + (\lambda/K)\boldsymbol{I}\right)\boldsymbol{\alpha}_k - 2\boldsymbol{\alpha}'_k(\boldsymbol{\alpha}_{\sim k} \circ \boldsymbol{Y}'_0\boldsymbol{y}_T)\right\} = 0.$$

This is quadratic in $\boldsymbol{\alpha}_k$ for fixed $\boldsymbol{\alpha}_{\sim k}$. Thus, the unique minimizer at convergence is

$$\widehat{\boldsymbol{\alpha}}_k = \left(\boldsymbol{Y}'_0\boldsymbol{Y}_0 \circ \widehat{\boldsymbol{\alpha}}_{\sim k}\widehat{\boldsymbol{\alpha}}'_{\sim k} + (\lambda/K)\boldsymbol{I}\right)^{-1}(\widehat{\boldsymbol{\alpha}}_{\sim k} \circ \boldsymbol{Y}'_0\boldsymbol{y}_T),$$

where $\widehat{\boldsymbol{\alpha}}_{\sim k} = \widehat{\boldsymbol{\alpha}}_1 \circ \cdots \circ \widehat{\boldsymbol{\alpha}}_{k-1} \circ \widehat{\boldsymbol{\alpha}}_{k+1} \circ \cdots \circ \widehat{\boldsymbol{\alpha}}_K$. Leveraging properties of the Hadamard product in Styan (1973), we rewrite $\boldsymbol{Y}'_0\boldsymbol{Y}_0 \circ \widehat{\boldsymbol{\alpha}}_{\sim k}\widehat{\boldsymbol{\alpha}}'_{\sim k} = \boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})\boldsymbol{Y}'_0\boldsymbol{Y}_0\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})$ and $\boldsymbol{Y}'_0\boldsymbol{y}_T \circ \widehat{\boldsymbol{\alpha}}_{\sim k} = \boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})\boldsymbol{Y}'_0\boldsymbol{y}_T$, where $\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})$ is the diagonal matrix formed from $\widehat{\boldsymbol{\alpha}}_{\sim k}$. Thus, $\widehat{\boldsymbol{\alpha}}_k = (\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})\boldsymbol{Y}'_0\boldsymbol{Y}_0\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k}) + (\lambda/K)\boldsymbol{I})^{-1}\boldsymbol{D}(\widehat{\boldsymbol{\alpha}}_{\sim k})\boldsymbol{Y}'_0\boldsymbol{y}_T$. Turning to VT regression, we have $\widehat{\boldsymbol{\beta}}_k = (\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{\sim k})\boldsymbol{Y}_0\boldsymbol{Y}'_0\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{\sim k}) + (\lambda/K)\boldsymbol{I})^{-1}\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{\sim k})\boldsymbol{Y}_0\boldsymbol{y}_N$, where $\widehat{\boldsymbol{\beta}}_{\sim k} = \widehat{\boldsymbol{\beta}}_1 \circ \cdots \circ \widehat{\boldsymbol{\beta}}_{k-1} \circ \widehat{\boldsymbol{\beta}}_{k+1} \circ \cdots \circ \widehat{\boldsymbol{\beta}}_K$ and $\boldsymbol{D}(\widehat{\boldsymbol{\beta}}_{\sim k})$ is the diagonal matrix formed from $\widehat{\boldsymbol{\beta}}_{\sim k}$. This completes the proof. *Q.E.D.*

### A.3. *Proof of Corollary 1*

Proof: Consider OLS. Recall that $\widehat{Y}^{\text{hz}}_{NT}(0) = \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}}\rangle$ with $\widehat{\boldsymbol{\alpha}} = \boldsymbol{Y}_0^{\dagger}\boldsymbol{y}_T$ and $\widehat{Y}^{\text{vt}}_{NT}(0) = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}\rangle$ with $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Y}'_0)^{\dagger}\boldsymbol{y}_N$. By Theorem 1, $\widehat{Y}^{\text{hz}}_{NT}(0) = \widehat{Y}^{\text{vt}}_{NT}(0)$. Thus,

$$\widehat{Y}^{\text{dr}}_{NT}(0) = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}\rangle + \langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}}\rangle - \langle \widehat{\boldsymbol{\alpha}}, \boldsymbol{Y}'_0\widehat{\boldsymbol{\beta}}\rangle = 2\langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}\rangle - \langle \widehat{\boldsymbol{\alpha}}, \boldsymbol{Y}'_0\widehat{\boldsymbol{\beta}}\rangle. \tag{28}$$

Since $(\boldsymbol{Y}'_0)^{\dagger} = (\boldsymbol{Y}_0^{\dagger})'$, we have

$$\langle \widehat{\boldsymbol{\alpha}}, \boldsymbol{Y}'_0\widehat{\boldsymbol{\beta}}\rangle = \boldsymbol{y}'_T(\boldsymbol{Y}'_0)^{\dagger}\boldsymbol{Y}'_0(\boldsymbol{Y}'_0)^{\dagger}\boldsymbol{y}_N = \boldsymbol{y}'_T(\boldsymbol{Y}'_0)^{\dagger}\boldsymbol{y}_N = \boldsymbol{y}'_T\widehat{\boldsymbol{\beta}}. \tag{29}$$

Plugging (29) into (28), we conclude $\widehat{Y}^{\text{dr}}_{NT}(0) = 2\langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}\rangle - \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}\rangle = \widehat{Y}^{\text{vt}}_{NT}(0) = \widehat{Y}^{\text{hz}}_{NT}(0)$. Now, observe that the same arguments above hold when $\boldsymbol{Y}_0^{(k)}$ takes the place of $\boldsymbol{Y}_0$ for any $k < R$. Therefore, the same reduction can be derived for PCR. *Q.E.D.*

### A.4. *Proof of Corollary 2*

Proof: Let $\boldsymbol{Y}_0^{\text{hz}} = [\boldsymbol{1}, \boldsymbol{Y}_0]$ and $\boldsymbol{Y}_0^{\text{vt}} = [\boldsymbol{1}, \boldsymbol{Y}'_0]$. The proof is immediate from Theorem 1 by noting that $(\boldsymbol{Y}_0^{\text{hz}})' \neq \boldsymbol{Y}_0^{\text{vt}}$. *Q.E.D.*

### A.5. *Proof of Corollary 3*

Proof: Consider HZ ridge regression. The optimality conditions give

$$\nabla_{(\alpha_0,\alpha_1,\boldsymbol{\alpha})}\left\{\|\boldsymbol{y}_T - \boldsymbol{Y}_0\boldsymbol{\alpha} - \alpha_0\boldsymbol{1}\|_2^2 + \|\boldsymbol{y}_N - \boldsymbol{\alpha}_1\boldsymbol{1}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2\right\} = 0.$$

Solving for $\alpha_0$, we have $\widehat{\alpha}_0 = (1/N_0)\langle \boldsymbol{y}_T, \mathbf{1}\rangle$, which uses $\boldsymbol{Y}_0'\mathbf{1} = \mathbf{0}$. Solving for $\alpha_1$, we have $\widehat{\alpha}_1 = (1/T_0)\langle \boldsymbol{y}_N, \mathbf{1}\rangle$. Finally, solving for $\boldsymbol{\alpha}$, we have $\widehat{\boldsymbol{\alpha}} = (\boldsymbol{Y}_0'\boldsymbol{Y}_0 + \lambda \boldsymbol{I})^{-1}\boldsymbol{Y}_0'\boldsymbol{y}_T$.

Identical arguments for VT regression yield $\widehat{\beta}_0 = (1/T_0)\langle \boldsymbol{y}_N, \mathbf{1}\rangle$, $\widehat{\beta}_1 = (1/N_0)\langle \boldsymbol{y}_T, \mathbf{1}\rangle$, and $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Y}_0\boldsymbol{Y}_0' + \lambda \boldsymbol{I})\boldsymbol{Y}_0\boldsymbol{y}_N$.

Further, Theorem 1 gives $\langle \boldsymbol{y}_N, \widehat{\boldsymbol{\alpha}}\rangle = \langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}}\rangle$. Observing that $\widehat{\alpha}_0 = \widehat{\beta}_1$ and $\widehat{\beta}_0 = \widehat{\alpha}_1$, proves our ridge result. Setting $\lambda = 0$ and using the pseudoinverse, we have our OLS result. The result for PCR is established from OLS by substituting $\boldsymbol{Y}_0^{(k)}$ for $\boldsymbol{Y}_0$. This completes the proof. *Q.E.D.*

## APPENDIX B: PROOFS FOR INFERENCE

### B.1. *Proof of Theorem 3*

To establish Theorem 3, we first state a few useful results.

LEMMA 4—Theorem 2.7.1 of Lehmann (2000): *Let $X_i$ for $i = 1, \ldots, n$ be independently distributed with means $\mathbb{E}[X_i] = \zeta_i$ and variances $\sigma_i^2$, and with finite third moments. Let $\bar{X} = (1/n)\sum_{i=1}^n X_i$. Then* $\mathrm{Var}(\bar{X})^{-1/2} \cdot (\bar{X} - \mathbb{E}[\bar{X}]) \xrightarrow{d} \mathcal{N}(0,1)$, *provided*

$$\left(\sum_{i=1}^{n} \mathbb{E}\left[|X_i - \zeta_i|^3\right]\right)^2 = o\left(\left(\sum_{i=1}^{n} \sigma_i^2\right)^3\right).$$

LEMMA 5: *Consider a random vector $\boldsymbol{x}$ and random matrix $\boldsymbol{A}$. Let $\mathbb{E}[\boldsymbol{x}|\boldsymbol{A}] = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{x}|\boldsymbol{A}) = \boldsymbol{\Sigma}$. Then $\mathbb{E}[\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}|\boldsymbol{A}] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})$.*

PROOF: (i) [HZ model] Let Assumption 1 hold. By (39), Lemma 4 yields

$$\mathrm{Var}(\widehat{Y}_{NT}(0)|\boldsymbol{y}_N, \boldsymbol{Y}_0)^{-1/2} \cdot (\widehat{Y}_{NT}(0) - \mathbb{E}[\widehat{Y}_{NT}(0)|\boldsymbol{y}_N, \boldsymbol{Y}_0]) \xrightarrow{d} \mathcal{N}(0,1).$$

To evaluate $\mathbb{E}[\widehat{Y}_{NT}(0)|\boldsymbol{y}_N, \boldsymbol{Y}_0]$, we first observe that

$$\mathbb{E}[\widehat{Y}_{NT}(0)|\boldsymbol{y}_N, \boldsymbol{Y}_0] = \mathbb{E}[\langle \boldsymbol{y}_N, \boldsymbol{Y}_0^{\dagger}\boldsymbol{y}_T\rangle|\boldsymbol{y}_N, \boldsymbol{Y}_0] = \boldsymbol{y}_N'\boldsymbol{Y}_0^{\dagger}\boldsymbol{Y}_0\boldsymbol{\alpha}^* = \boldsymbol{y}_N'\boldsymbol{H}^v\boldsymbol{\alpha}^*. \tag{30}$$

Moving to the variance term, we note that

$$\mathrm{Var}(\widehat{Y}_{NT}(0)|\boldsymbol{y}_N, \boldsymbol{Y}_0) = \boldsymbol{y}_N' \,\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}|\boldsymbol{y}_N, \boldsymbol{Y}_0)\boldsymbol{y}_N. \tag{31}$$

Towards evaluating the above, we note that

$$\mathrm{Cov}(\widehat{\boldsymbol{\alpha}}|\boldsymbol{y}_N, \boldsymbol{Y}_0) = \boldsymbol{Y}_0^{\dagger}\,\mathrm{Cov}(\boldsymbol{\varepsilon}_T|\boldsymbol{y}_N, \boldsymbol{Y}_0)(\boldsymbol{Y}_0')^{\dagger} = \boldsymbol{Y}_0^{\dagger}\boldsymbol{\Sigma}_T^{\mathrm{hz}}(\boldsymbol{Y}_0')^{\dagger}. \tag{32}$$

Plugging (32) into (31), we obtain

$$\mathrm{Var}(\widehat{Y}_{NT}(0)|\boldsymbol{y}_N, \boldsymbol{Y}_0) = \boldsymbol{y}_N'\boldsymbol{Y}_0^{\dagger}\boldsymbol{\Sigma}_T^{\mathrm{hz}}(\boldsymbol{Y}_0')^{\dagger}\boldsymbol{y}_N = \widehat{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_T^{\mathrm{hz}}\widehat{\boldsymbol{\beta}}, \tag{33}$$

where we recall that $\widehat{\boldsymbol{\beta}} = (\boldsymbol{Y}_0')^{\dagger}\boldsymbol{y}_N$. Putting it all together, we conclude

$$(\widehat{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_T^{\mathrm{hz}}\widehat{\boldsymbol{\beta}})^{-1/2} \cdot (\widehat{Y}_{NT}(0) - \langle \boldsymbol{y}_N, \boldsymbol{H}^v\boldsymbol{\alpha}^*\rangle) \xrightarrow{d} \mathcal{N}(0,1).$$

(ii) [VT model] Let Assumption 2 hold. Following the arguments above, we have

$$(\widehat{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}_N^{\mathrm{vt}}\widehat{\boldsymbol{\alpha}})^{-1/2}\cdot(\widehat{Y}_{NT}(0)-\langle \boldsymbol{y}_T,\boldsymbol{H}^u\boldsymbol{\beta}^*\rangle)\xrightarrow{d}\mathcal{N}(0,1).$$

(iii) [DR model] Let Assumption 3 hold. We write

$$\widehat{Y}_{NT}(0)=\boldsymbol{y}_N'\boldsymbol{Y}_0^{\dagger}\boldsymbol{y}_T=\sum_{i\le N_0}\sum_{t\le T_0}(\boldsymbol{Y}_0^{\dagger})_{it}Y_{iT}Y_{Nt}. \tag{34}$$

Observe that (34) is a sum of independent random variables with $\mathbb{E}[Y_{iT}Y_{Nt}|\boldsymbol{Y}_0]=\mathbb{E}[Y_{iT}|\boldsymbol{Y}_0]\mathbb{E}[Y_{Nt}|\boldsymbol{Y}_0]$ and $\mathrm{Var}(Y_{iT}Y_{Nt}|\boldsymbol{Y}_0)=\mathbb{E}[Y_{iT}|\boldsymbol{Y}_0]^2\sigma_{Nt}^2+\mathbb{E}[Y_{Nt}|\boldsymbol{Y}_0]^2\sigma_{iT}^2+\sigma_{iT}^2\sigma_{Nt}^2$. Lemma 4 then establishes that

$$\mathrm{Var}(\widehat{Y}_{NT}(0)|\boldsymbol{Y}_0)^{-1/2}\cdot(\widehat{Y}_{NT}(0)-\mathbb{E}[\widehat{Y}_{NT}(0)|\boldsymbol{Y}_0])\xrightarrow{d}\mathcal{N}(0,1).$$

Our aim is to evaluate $\mathbb{E}[\widehat{Y}_{NT}(0)|\boldsymbol{Y}_0]$ and $\mathrm{Var}(\widehat{Y}_{NT}(0)|\boldsymbol{Y}_0)$. Towards the former, we use Assumption 3 with the law of total expectation to obtain

$$\begin{aligned}\mathbb{E}[\widehat{Y}_{NT}(0)|\boldsymbol{Y}_0]&=\mathbb{E}\left[\mathbb{E}[\boldsymbol{y}_N'\boldsymbol{Y}_0^{\dagger}(\boldsymbol{Y}_0\boldsymbol{\alpha}^*+\boldsymbol{\varepsilon}_T)|\boldsymbol{\varepsilon}_N,\boldsymbol{Y}_0]|\boldsymbol{Y}_0\right]\\&=\mathbb{E}\left[(\boldsymbol{Y}_0'\boldsymbol{\beta}^*+\boldsymbol{\varepsilon}_N)'\boldsymbol{Y}_0^{\dagger}\boldsymbol{Y}_0\boldsymbol{\alpha}^*|\boldsymbol{Y}_0\right]=\langle\boldsymbol{\beta}^*,\boldsymbol{Y}_0\boldsymbol{\alpha}^*\rangle.\end{aligned}$$

Note that we have used the fact that $\boldsymbol{y}_N$ is deterministic given $(\boldsymbol{\varepsilon}_N,\boldsymbol{Y}_0)$. Similarly, by the law of total variance,

$$\mathrm{Var}(\widehat{Y}_{NT}(0)|\boldsymbol{Y}_0)=\mathbb{E}[\mathrm{Var}(\widehat{Y}_{NT}(0)|\boldsymbol{\varepsilon}_N,\boldsymbol{Y}_0)|\boldsymbol{Y}_0]+\mathrm{Var}(\mathbb{E}[\widehat{Y}_{NT}(0)|\boldsymbol{\varepsilon}_N,\boldsymbol{Y}_0]|\boldsymbol{Y}_0). \tag{35}$$

Following the derivation of (33), we have

$$\mathbb{E}[\mathrm{Var}(\widehat{Y}_{NT}(0)|\boldsymbol{\varepsilon}_N,\boldsymbol{Y}_0)|\boldsymbol{Y}_0]=(\boldsymbol{Y}_0'\boldsymbol{\beta}^*)'\boldsymbol{A}(\boldsymbol{Y}_0'\boldsymbol{\beta}^*)+\mathbb{E}[\boldsymbol{\varepsilon}_N'\boldsymbol{A}\boldsymbol{\varepsilon}_N|\boldsymbol{Y}_0]+2\mathbb{E}[\boldsymbol{\varepsilon}_N'\boldsymbol{Y}_0'\boldsymbol{\beta}^*|\boldsymbol{Y}_0], \tag{36}$$

where $\boldsymbol{A}=\boldsymbol{Y}_0^{\dagger}\boldsymbol{\Sigma}_T^{\mathrm{dr}}(\boldsymbol{Y}_0')^{\dagger}$. Notice that Assumption 3 gives $\mathbb{E}[\boldsymbol{\varepsilon}_N'\boldsymbol{Y}_0'\boldsymbol{\beta}^*|\boldsymbol{Y}_0]=0$. Since $\boldsymbol{A}$ is deterministic given $\boldsymbol{Y}_0$, Lemma 5 yields

$$\mathbb{E}[\boldsymbol{\varepsilon}_N'\boldsymbol{A}\boldsymbol{\varepsilon}_N|\boldsymbol{Y}_0]=\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}_N^{\mathrm{dr}}). \tag{37}$$

Following the arguments that led to the derivation of (30), we have

$$\mathrm{Var}(\mathbb{E}[\widehat{Y}_{NT}(0)|\boldsymbol{\varepsilon}_N,\boldsymbol{Y}_0]|\boldsymbol{Y}_0)=\mathrm{Var}(\boldsymbol{y}_N'\boldsymbol{H}^v\boldsymbol{\alpha}^*|\boldsymbol{Y}_0)=(\boldsymbol{H}^v\boldsymbol{\alpha}^*)'\boldsymbol{\Sigma}_N^{\mathrm{dr}}(\boldsymbol{H}^v\boldsymbol{\alpha}^*). \tag{38}$$

Plugging (36), (37), and (38) into (35), we arrive at our desired result. *Q.E.D.*

## APPENDIX C: Inference

### C.1. *Model-Based Asymptotic Results on Inference*

We state the moment conditions of Theorem 3. Let $(\sigma_{iT}^{\mathrm{hz}})^2 = \mathrm{Var}(\varepsilon_{iT}|\boldsymbol{y}_N, \boldsymbol{Y}_0)$ for $i = 1,\dots,N_0$. Define $\{(\sigma_{Nt}^{\mathrm{vt}})^2, (\sigma_{iT}^{\mathrm{dr}})^2, (\sigma_{Nt}^{\mathrm{dr}})^2\}$ for $i = 1,\dots,N_0$ and $t = 1,\dots,T_0$ with respect to $(\boldsymbol{\Sigma}_N^{\mathrm{vt}}, \boldsymbol{\Sigma}_T^{\mathrm{dr}}, \boldsymbol{\Sigma}_N^{\mathrm{dr}})$ analogously. These are the conditions:

**I: HZ condition.**

$$\left(\sum_{i\le N_0} \mathbb{E}\left[|\widehat{\beta}_i \varepsilon_{iT}|^3 |\ \boldsymbol{y}_N, \boldsymbol{Y}_0\right]\right)^2 = o\left(\left(\sum_{i\le N_0} \widehat{\beta}_i^2 (\sigma_{iT}^{\mathrm{hz}})^2\right)^3\right). \tag{39}$$

**II: VT condition.**

$$\left(\sum_{t\le T_0} \mathbb{E}\left[|\widehat{\alpha}_t \varepsilon_{Nt}|^3 |\ \boldsymbol{y}_T, \boldsymbol{Y}_0\right]\right)^2 = o\left(\left(\sum_{t\le T_0} \widehat{\alpha}_t^2 (\sigma_{Nt}^{\mathrm{vt}})^2\right)^3\right). \tag{40}$$

**III: DR condition.**

$$\begin{aligned}&\left(\sum_{i\le N_0}\sum_{t\le T_0} \mathbb{E}\Big[|(\boldsymbol{Y}_0^\dagger)_{it}\left\{\mathbb{E}[Y_{iT}|\boldsymbol{Y}_0]\varepsilon_{Nt} + \mathbb{E}[Y_{Nt}|\boldsymbol{Y}_0]\varepsilon_{iT} + \varepsilon_{iT}\varepsilon_{Nt}\right\}|^3 \mid \boldsymbol{Y}_0\Big]\right)^2 \\ &= o\left(\left(\sum_{i\le N_0}\sum_{t\le T_0} (\boldsymbol{Y}_0^\dagger)_{it}^2 \left\{\mathbb{E}[Y_{iT}|\boldsymbol{Y}_0]^2 (\sigma_{Nt}^{\mathrm{dr}})^2 + \mathbb{E}[Y_{Nt}|\boldsymbol{Y}_0]^2 (\sigma_{iT}^{\mathrm{dr}})^2 + (\sigma_{iT}^{\mathrm{dr}})^2 (\sigma_{Nt}^{\mathrm{dr}})^2\right\}\right)^3\right)\end{aligned} \tag{41}$$

If $(\varepsilon_{iT}, (\sigma_{iT}^{\mathrm{hz}})^2)$ are bounded, then (39) translates to $\sum_{i\le N_0} |\widehat{\beta}_i|^3 = o(\|\widehat{\boldsymbol{\beta}}\|_2^3)$, which rules out outlier coefficients; a similar interpretation can be derived for (40). Similarly, if $(\varepsilon_{iT}, \varepsilon_{Nt})$ and $(\sigma_{iT}^2, \sigma_{Nt}^2)$ are bounded for all $(i,t)$, then (41) loosely translates to

$$\sum_{i\le N_0} |\widehat{\beta}_i|^3 + \sum_{t\le T_0} |\widehat{\alpha}_t|^3 + \sum_{i\le N_0}\sum_{t\le T_0} |(\boldsymbol{Y}_0^\dagger)_{it}|^3 = o\left(\|\widehat{\boldsymbol{\beta}}\|_2^3 + \|\widehat{\boldsymbol{\alpha}}\|_2^3 + \|\boldsymbol{Y}_0^\dagger\|_F^3\right),$$

which effectively bounds the magnitudes of the HZ and VT OLS coefficients and pseudoinverse matrix entries.

### C.2. *Model-Based Confidence Intervals*

We present the confidence intervals previewed in Section 4.1.3. Theorem 3 motivates separate HZ, VT, and DR confidence intervals: for $\theta \in (0,1)$,

$$\begin{aligned}\mu_0^{\mathrm{hz}} &\in \left[\widehat{Y}_{NT}(0) \ \pm\ z_{\frac{\theta}{2}} \sqrt{\widehat{v}_0^{\mathrm{hz}}}\right], \\ \mu_0^{\mathrm{vt}} &\in \left[\widehat{Y}_{NT}(0) \ \pm\ z_{\frac{\theta}{2}} \sqrt{\widehat{v}_0^{\mathrm{vt}}}\right], \\ \mu_0^{\mathrm{dr}} &\in \left[\widehat{Y}_{NT}(0) \ \pm\ z_{\frac{\theta}{2}} \sqrt{\widehat{v}_0^{\mathrm{dr}}}\right],\end{aligned}$$

where $z_{\frac{\theta}{2}}$ is the upper $\theta/2$ quantile of $\mathcal{N}(0,1)$, and $(\widehat{v}_0^{\text{hz}}, \widehat{v}_0^{\text{vt}}, \widehat{v}_0^{\text{dr}})$ are the estimators of $(v_0^{\text{hz}}, v_0^{\text{vt}}, v_0^{\text{dr}})$. We construct

$$\widehat{v}_0^{\text{hz}} = \widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{\Sigma}}_T \widehat{\boldsymbol{\beta}}, \quad \widehat{v}_0^{\text{vt}} = \widehat{\boldsymbol{\alpha}}' \widehat{\boldsymbol{\Sigma}}_N \widehat{\boldsymbol{\alpha}}, \quad \widehat{v}_0^{\text{dr}} = \widehat{v}_0^{\text{hz}} + \widehat{v}_0^{\text{vt}} - \text{tr}(\boldsymbol{Y}_0^{\dagger} \widehat{\boldsymbol{\Sigma}}_T (\boldsymbol{Y}_0')^{\dagger} \widehat{\boldsymbol{\Sigma}}_N), \tag{42}$$

where $\widehat{\boldsymbol{\Sigma}}_T$ and $\widehat{\boldsymbol{\Sigma}}_N$ are the estimators of $(\boldsymbol{\Sigma}_T^{\text{hz}}, \boldsymbol{\Sigma}_T^{\text{dr}})$ and $(\boldsymbol{\Sigma}_N^{\text{vt}}, \boldsymbol{\Sigma}_N^{\text{dr}})$, respectively. We precisely define them under homoskedastic and heteroskedastic errors below. To reduce ambiguity, we index $(\widehat{v}_0^{\text{hz}}, \widehat{v}_0^{\text{vt}}, \widehat{v}_0^{\text{dr}})$ by the covariance estimator. Recall $\boldsymbol{H}^u = \boldsymbol{U}\boldsymbol{U}'$ and $\boldsymbol{H}^v = \boldsymbol{V}\boldsymbol{V}'$. We define $\boldsymbol{H}_{\perp}^u = \boldsymbol{I} - \boldsymbol{H}^u$ and $\boldsymbol{H}_{\perp}^v = \boldsymbol{I} - \boldsymbol{H}^v$. With this notation, the HZ and VT in-sample errors are $\boldsymbol{H}_{\perp}^u \boldsymbol{y}_T = \boldsymbol{y}_T - \boldsymbol{Y}_0 \widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{H}_{\perp}^v \boldsymbol{y}_N = \boldsymbol{y}_N - \boldsymbol{Y}_0' \widehat{\boldsymbol{\beta}}$, respectively.

It is clear from (42) that $(\widehat{v}_0^{\text{hz}}, \widehat{v}_0^{\text{vt}})$ are plug-in estimators for $(v_0^{\text{hz}}, v_0^{\text{vt}})$. As such, we discuss $\widehat{v}_0^{\text{dr}}$ with respect to $v_0^{\text{dr}}$. Recall $\widehat{\boldsymbol{\alpha}} = \boldsymbol{H}^v \widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}} = \boldsymbol{H}^u \widehat{\boldsymbol{\beta}}$ by construction. To justify the negative trace in $\widehat{v}_0^{\text{dr}}$, note that $\widehat{v}_0^{\text{hz}}$ is a quadratic involving $(\boldsymbol{y}_N, \boldsymbol{y}_T)$. Since both quantities are random, the expectation of $\widehat{v}_0^{\text{hz}}$ induces an additional term that precisely corresponds to the trace term in $v_0^{\text{dr}}$. The same property holds for $\widehat{v}_0^{\text{vt}}$. Thus, $\widehat{v}_0^{\text{dr}}$ corrects for this bias via the negative trace.

### C.2.1. *Homoskedastic Errors*

Consider $\boldsymbol{\Sigma}_T^{\text{hz}}$ with identical diagonal elements, i.e., $\boldsymbol{\Sigma}_T^{\text{hz}} = (\sigma_T^{\text{hz}})^2 \boldsymbol{I}$, where $(\sigma_T^{\text{hz}})^2 = \text{Var}(\varepsilon_{iT} | \boldsymbol{y}_N, \boldsymbol{Y}_0)$ for $i = 1, \ldots, N_0$. Let $(\boldsymbol{\Sigma}_N^{\text{vt}}, \boldsymbol{\Sigma}_T^{\text{dr}}, \boldsymbol{\Sigma}_N^{\text{dr}})$ be defined analogously. We use the standard variance estimators

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_T^{\text{homo}} &= \frac{1}{N_0 - R} \|\boldsymbol{H}_{\perp}^u \boldsymbol{y}_T\|_2^2 \boldsymbol{I}, \\ \widehat{\boldsymbol{\Sigma}}_N^{\text{homo}} &= \frac{1}{T_0 - R} \|\boldsymbol{H}_{\perp}^v \boldsymbol{y}_N\|_2^2 \boldsymbol{I}, \end{aligned} \tag{43}$$

where $R = \text{rank}(\boldsymbol{Y}_0)$, which can be computed as $R = \text{tr}(\boldsymbol{H}^u) = \text{tr}(\boldsymbol{H}^v)$.

LEMMA 6: *Consider homoskedastic errors. (i)* [HZ model] *Under Assumption 1, we have*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{homo}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\text{hz}} \quad \textit{and} \quad \mathbb{E}[\widehat{v}_0^{\text{hz,homo}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] = v_0^{\text{hz}}.$$

*(ii)* [VT model] *Under Assumption 2, we have*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\text{homo}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\text{vt}} \quad \textit{and} \quad \mathbb{E}[\widehat{v}_0^{\text{vt,homo}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = v_0^{\text{vt}}.$$

*(iii)* [DR model] *Under Assumption 3, we have*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{homo}} | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\text{dr}}, \quad \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\text{homo}} | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\text{dr}}, \quad \textit{and} \quad \mathbb{E}[\widehat{v}_0^{\text{dr,homo}} | \boldsymbol{Y}_0] = v_0^{\text{dr}}.$$

Lemma 6 is a well known result within the OLS literature, albeit it is typically formalized under the stricter full column rank assumption. As a comparison with the synthetic controls literature, we take note of the recent work of Agarwal et al. (2021). Agarwal et al. (2021) propose a VT PCR estimator under the homoskedastic setting and provide a similar confidence interval to that of (43) via large-sample approximations. Under a closely related VT model, they propose $\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{\Sigma}}_N^{\text{homo}} \widehat{\boldsymbol{\beta}}$ in place of $\widehat{\boldsymbol{\alpha}}' \widehat{\boldsymbol{\Sigma}}_N^{\text{homo}} \widehat{\boldsymbol{\alpha}}$. While

the point estimate of Agarwal et al. (2021) also takes the form $\langle \boldsymbol{y}_T, \widehat{\boldsymbol{\beta}} \rangle$, their variance estimator only depends on $(\boldsymbol{y}_N, \boldsymbol{Y}_0)$; in comparison, ours depends on $(\boldsymbol{y}_N, \boldsymbol{y}_T, \boldsymbol{Y}_0)$. Hence, the confidence interval as per Agarwal et al. (2021) is numerically identical for every post-treatment point estimate while ours can vary across the post-treatment periods, which may be favorable.

C.2.2. *Heteroskedastic Errors*

We adopt two strategies for the heteroskedastic setting.

*I: Jackknife.* The first estimator is based on the jackknife. Traditionally, the jackknife estimates the covariance of the regression coefficients $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$. By analyzing said estimates, we derive the following:

$$\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}} = \operatorname{diag}\left( \left[ \boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \circ \boldsymbol{I} \right]^\dagger \left[ \boldsymbol{H}_\perp^u \boldsymbol{y}_T \circ \boldsymbol{H}_\perp^u \boldsymbol{y}_T \right] \right) \tag{44}$$

$$\widehat{\boldsymbol{\Sigma}}_N^{\text{jack}} = \operatorname{diag}\left( \left[ \boldsymbol{H}_\perp^v \circ \boldsymbol{H}_\perp^v \circ \boldsymbol{I} \right]^\dagger \left[ \boldsymbol{H}_\perp^v \boldsymbol{y}_N \circ \boldsymbol{H}_\perp^v \boldsymbol{y}_N \right] \right). \tag{45}$$

LEMMA 7: *Consider heteroskedastic errors. (i)* [HZ model] *Let Assumption 1 hold. If* $(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \circ \boldsymbol{I})$ *is nonsingular, then*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\text{hz}} + \boldsymbol{\Delta}^{\text{hz}} \quad \textit{and} \quad \mathbb{E}[\widehat{v}_0^{\text{hz,jack}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] = v_0^{\text{hz}} + \widehat{\boldsymbol{\alpha}}' \boldsymbol{\Delta}^{\text{hz}} \widehat{\boldsymbol{\alpha}},$$

*where* $\Delta_{\ell\ell}^{\text{hz}} = \sum_{j \neq \ell} (\sigma_{jT}^{\text{hz}})^2 (H_{\ell j}^u)^2 (1 - H_{\ell\ell}^u)^{-2}$ *for* $\ell = 1, \ldots, N_0$. *(ii)* [VT model] *Let Assumption 2 hold. If* $(\boldsymbol{H}_\perp^v \circ \boldsymbol{H}_\perp^v \circ \boldsymbol{I})$ *is nonsingular, then*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\text{jack}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\text{vt}} + \boldsymbol{\Gamma}^{\text{vt}} \quad \textit{and} \quad \mathbb{E}[\widehat{v}_0^{\text{vt,jack}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = v_0^{\text{vt}} + \widehat{\boldsymbol{\beta}}' \boldsymbol{\Gamma}^{\text{vt}} \widehat{\boldsymbol{\beta}},$$

*where* $\Gamma_{\ell\ell}^{\text{vt}} = \sum_{j \neq \ell} (\sigma_{Nj}^{\text{vt}})^2 (H_{\ell j}^v)^2 (1 - H_{\ell\ell}^v)^{-2}$ *for* $\ell = 1, \ldots, T_0$. *(iii)* [DR model] *Let Assumption 3 hold. If* $(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \circ \boldsymbol{I})$ *and* $(\boldsymbol{H}_\perp^v \circ \boldsymbol{H}_\perp^v \circ \boldsymbol{I})$ *are nonsingular, then*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}} | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\text{dr}} + \boldsymbol{\Delta}^{\text{dr}}, \quad \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\text{jack}} | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\text{dr}} + \boldsymbol{\Gamma}^{\text{dr}},$$

$$\mathbb{E}[\widehat{v}^{\text{dr,jack}}(\boldsymbol{Y}_0) | \boldsymbol{Y}_0]$$

$$= v_0^{\text{dr}} + (\boldsymbol{H}^u \boldsymbol{\beta}^*)' \boldsymbol{\Delta}^{\text{dr}} (\boldsymbol{H}^u \boldsymbol{\beta}^*) + (\boldsymbol{H}^v \boldsymbol{\alpha}^*)' \boldsymbol{\Gamma}^{\text{dr}} (\boldsymbol{H}^v \boldsymbol{\alpha}^*) + \operatorname{tr}(\boldsymbol{Y}_0^\dagger \boldsymbol{\Delta}^{\text{dr}} (\boldsymbol{Y}_0')^\dagger \boldsymbol{\Gamma}^{\text{dr}}),$$

*where* $\Delta_{\ell\ell}^{\text{dr}}$ *and* $\Gamma_{\ell\ell}^{\text{dr}}$ *are defined analogously to* $\Delta_{\ell\ell}^{\text{hz}}$ *and* $\Gamma_{\ell\ell}^{\text{vt}}$, *respectively, with* $(\sigma_{jT}^{\text{dr}})^2$ *and* $(\sigma_{Nj}^{\text{dr}})^2$ *in place of* $(\sigma_{jT}^{\text{hz}})^2$ *and* $(\sigma_{Nj}^{\text{vt}})^2$, *respectively.*

Lemma 7 establishes that the jackknife is conservative, provided $(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \circ \boldsymbol{I})$ and $(\boldsymbol{H}_\perp^v \circ \boldsymbol{H}_\perp^v \circ \boldsymbol{I})$ are nonsingular. Strictly speaking, the jackknife is well defined if these quantities are singular, as seen through the pseudoinverse in (44) and (45). Lemma 7 considers the nonsingular case for simplicity. We remark that $\max_\ell H_{\ell\ell}^u < 1$ and $\max_\ell H_{\ell\ell}^v < 1$ are sufficient conditions for invertibility.

*II: HRK-estimator.* Next, we consider the covariance estimator proposed by Hartley et al. (1969). We index this estimator by the authors, Hartley-Rao-Kiefer:

$$\widehat{\boldsymbol{\Sigma}}_T^{\text{HRK}} = \operatorname{diag}\left( \left[ \boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \right]^{-1} \left[ \boldsymbol{H}_\perp^u \boldsymbol{y}_T \circ \boldsymbol{H}_\perp^u \boldsymbol{y}_T \right] \right)$$

$$\widehat{\boldsymbol{\Sigma}}_N^{\text{HRK}} = \operatorname{diag}\left( \left[ \boldsymbol{H}_\perp^v \circ \boldsymbol{H}_\perp^v \right]^{-1} \left[ \boldsymbol{H}_\perp^v \boldsymbol{y}_N \circ \boldsymbol{H}_\perp^v \boldsymbol{y}_N \right] \right).$$

LEMMA 8: *Consider heteroskedastic errors. (i)* [HZ model] *Let Assumption 1 hold. If* $(\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)$ *is nonsingular, then*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{HRK}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\text{hz}} \quad \textit{and} \quad \mathbb{E}[\widehat{v}_0^{\text{hz,HRK}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] = v_0^{\text{hz}}.$$

*(ii)* [VT model] *Let Assumption 2 hold. If* $(\boldsymbol{H}^v_\perp \circ \boldsymbol{H}^v_\perp)$ *is nonsingular, then*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\text{HRK}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\text{vt}} \quad \textit{and} \quad \mathbb{E}[\widehat{v}_0^{\text{vt,HRK}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = v_0^{\text{vt}}.$$

*(iii)* [DR model] *Let Assumption 3 hold. If* $(\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)$ *and* $(\boldsymbol{H}^v_\perp \circ \boldsymbol{H}^v_\perp)$ *are nonsingular, then*

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{HRK}} | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\text{dr}}, \quad \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\text{HRK}} | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\text{dr}}, \quad \textit{and} \quad \mathbb{E}[\widehat{v}_0^{\text{dr,HRK}} | \boldsymbol{Y}_0] = v_0^{\text{dr}}.$$

Lemma 8 establishes that the HRK estimator is unbiased, provided $(\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)$ and $(\boldsymbol{H}^v_\perp \circ \boldsymbol{H}^v_\perp)$ are invertible. To discuss sufficient conditions for invertibility, consider $(\boldsymbol{H}^u \circ \boldsymbol{H}^u)$. A sufficient condition is strict diagonal dominance (Varga, 1962): $(1 - H^u_{\ell\ell})^2 > \sum_{j \neq \ell} (H^u_{\ell j})^2$. Notice that $\boldsymbol{H}^u$ is an orthogonal projector and is thus idempotent, i.e., $(\boldsymbol{H}^u)^2 = \boldsymbol{H}^u$, and symmetric. Therefore,

$$H^u_{\ell\ell} = (H^u_{\ell\ell})^2 + \sum_{j \neq \ell} (H^u_{\ell j})^2 \implies \sum_{j \neq \ell} (H^u_{\ell j})^2 = H^u_{\ell\ell}(1 - H^u_{\ell\ell}),$$

which allows us to simplify the condition as $(1 - H^u_{\ell\ell})^2 > H^u_{\ell\ell} - (H^u_{\ell\ell})^2$. Thus, $\max_\ell H^u_{\ell\ell} < 1/2$ is a sufficient condition for invertibility. Since $\text{tr}(\boldsymbol{H}^u) = R$, this restricts $R < N_0/2$. The same arguments apply for $(\boldsymbol{H}^v \circ \boldsymbol{H}^v)$.

### C.2.3. *Discussion*

We highlight that Lemmas 6–8 only hold in expectation. For any particular realization, $\widehat{v}_0^{\text{dr}}$ may exhibit unexpected properties. For instance, if $\text{tr}(\boldsymbol{Y}_0^\dagger \widehat{\boldsymbol{\Sigma}}_T (\boldsymbol{Y}_0')^\dagger \widehat{\boldsymbol{\Sigma}}_N) > \max\{\widehat{v}_0^{\text{hz}}, \widehat{v}_0^{\text{vt}}\}$, then $\widehat{v}_0^{\text{dr}} < \min\{\widehat{v}_0^{\text{hz}}, \widehat{v}_0^{\text{vt}}\}$; thus, the mixed coverage will be smaller than both HZ and VT coverages. In fact, $\widehat{v}_0^{\text{dr}}$ can be negative if $\text{tr}(\boldsymbol{Y}_0^\dagger \widehat{\boldsymbol{\Sigma}}_T (\boldsymbol{Y}_0')^\dagger \widehat{\boldsymbol{\Sigma}}_N) > \widehat{v}_0^{\text{hz}} + \widehat{v}_0^{\text{vt}}$, which may occur if both HZ and VT in-sample errors are "too large". For these scenarios, one naïve solution is to modify $\widehat{v}_0^{\text{dr}}$ as $\widehat{v}_0^{\text{dr}} = \widehat{v}_0^{\text{hz}} + \widehat{v}_0^{\text{vt}}$, which is conservative by Lemmas 6–8. However, this case is arguably better resolved with a different point estimator altogether.

## APPENDIX D: PROOFS FOR MODEL-BASED CONFIDENCE INTERVALS

We first state a useful lemma to prove Lemmas 6–8.

LEMMA 9: [DR model] *Let Assumption 3 hold. Then,*

$$\begin{aligned}
\mathbb{E}[\widehat{v}_0^{\text{dr}} | \boldsymbol{Y}_0] = {} & (\boldsymbol{H}^u \boldsymbol{\beta}^*)' \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T | \boldsymbol{Y}_0] (\boldsymbol{H}^u \boldsymbol{\beta}^*) + \text{tr}(\boldsymbol{Y}_0^\dagger \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T | \boldsymbol{Y}_0] (\boldsymbol{Y}_0')^\dagger \boldsymbol{\Sigma}_N^{\text{dr}}) \\
& + (\boldsymbol{H}^v \boldsymbol{\alpha}^*)' \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N | \boldsymbol{Y}_0] (\boldsymbol{H}^v \boldsymbol{\alpha}^*) + \text{tr}(\boldsymbol{Y}_0^\dagger \boldsymbol{\Sigma}_T^{\text{dr}} (\boldsymbol{Y}_0')^\dagger \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N | \boldsymbol{Y}_0]) \\
& - \text{tr}(\boldsymbol{Y}_0^\dagger \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T | \boldsymbol{Y}_0] (\boldsymbol{Y}_0')^\dagger \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N | \boldsymbol{Y}_0]).
\end{aligned}$$

D.0.1. *Proof of Lemma 6*

PROOF: (i) [HZ model] Let Assumption 1 hold. Taking note that $\boldsymbol{H}^u_\perp \boldsymbol{Y}_0 = \boldsymbol{0}$,

$$\begin{aligned}\|\boldsymbol{H}^u_\perp \boldsymbol{y}_T\|_2^2 &= \boldsymbol{y}_T' \boldsymbol{H}^u_\perp \boldsymbol{y}_T \\ &= (\boldsymbol{Y}_0\boldsymbol{\alpha} + \boldsymbol{\varepsilon}_T)' \boldsymbol{H}^u_\perp (\boldsymbol{Y}_0\boldsymbol{\alpha} + \boldsymbol{\varepsilon}_T) \\ &= \boldsymbol{\varepsilon}_T' \boldsymbol{H}^u_\perp \boldsymbol{\varepsilon}_T.\end{aligned}$$

Applying Lemma 5 then gives

$$\mathbb{E}[\boldsymbol{\varepsilon}_T' \boldsymbol{H}^u_\perp \boldsymbol{\varepsilon}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \operatorname{tr}(\boldsymbol{H}^u_\perp)(\sigma_T^{\text{hz}})^2 = (N_0 - R)(\sigma_T^{\text{hz}})^2, \tag{46}$$

where the final equality follows because the trace of a projection matrix equals its rank. Taken altogether, we have $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{homo}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\text{hz}}$. Therefore,

$$\mathbb{E}[\widehat{v}_0^{\text{hz,homo}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \widehat{\boldsymbol{\beta}}' \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{homo}} | \boldsymbol{y}_N, \boldsymbol{Y}_0] \widehat{\boldsymbol{\beta}} = v_0^{\text{hz}}.$$

(ii) [VT model] Let Assumption 2 hold. Following the arguments above, we conclude that $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\text{homo}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\text{vt}}$ and $\mathbb{E}[\widehat{v}_0^{\text{vt,homo}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = v_0^{\text{vt}}$.

(iii) [DR model] Let Assumption 3 hold. Following the arguments that led to (46), we obtain $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\text{homo}} | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\text{dr}}$ and $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\text{homo}} | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\text{dr}}$. Applying Lemma 9 then gives $\mathbb{E}[\widehat{v}_0^{\text{dr,homo}} | \boldsymbol{Y}_0] = v_0^{\text{dr}}$. The proof is complete. *Q.E.D.*

## D.1. *Proof of Lemma 7*

PROOF: Before we establish the biases of $(\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}}, \widehat{\boldsymbol{\Sigma}}_N^{\text{jack}})$, we first justify their forms. Jackknife is a popular approach to estimate the covariances of $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$. Below, we follow the standard techniques to derive the jackknife estimate of these objects, which will then be used to derive $(\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}}, \widehat{\boldsymbol{\Sigma}}_N^{\text{jack}})$. Without loss of generality, we begin with $\widehat{\boldsymbol{\alpha}}$. Notably, while standard derivations consider $\boldsymbol{Y}_0$ with full column rank, we consider a general matrix $\boldsymbol{Y}_0$ that may be rank deficient. This difference is subtle so the following proof is by no means novel. We provide it simply for completeness.

To describe the jackknife, we define $\widehat{\boldsymbol{\alpha}}_{\sim i}$ as the minimum $\ell_2$-norm solution to (2), where $\lambda_1 = \lambda_2 = 0$, without the $i$th observation, i.e.,

$$\widehat{\boldsymbol{\alpha}}_{\sim i} = (\boldsymbol{Y}'_{0,\sim i} \boldsymbol{Y}_{0,\sim i})^\dagger \boldsymbol{Y}'_{0,\sim i} \boldsymbol{y}_{T,\sim i}, \tag{47}$$

where $\boldsymbol{Y}_{0,\sim i}$ and $\boldsymbol{y}_{T,\sim i}$ correspond to $\boldsymbol{Y}_0$ and $\boldsymbol{y}_T$ without the $i$th observation. We define the pseudo-estimator as $\tilde{\boldsymbol{\alpha}}_i = T_0 \widehat{\boldsymbol{\alpha}} - (T_0 - 1)\widehat{\boldsymbol{\alpha}}_{\sim i}$. With these quantities defined, we write the jackknife variance estimator as

$$\widehat{\boldsymbol{V}}^{\text{jack}} = \frac{1}{(T_0 - 1)^2} \sum_{i \le N_0} (\tilde{\boldsymbol{\alpha}}_i - \widehat{\boldsymbol{\alpha}})(\tilde{\boldsymbol{\alpha}}_i - \widehat{\boldsymbol{\alpha}})'. \tag{48}$$

To evaluate this quantity, we will rewrite $\widehat{\boldsymbol{\alpha}}_{\sim i}$ in a more convenient form. In particular,

$$\begin{aligned}\boldsymbol{Y}'_{0,\sim i} \boldsymbol{Y}_{0,\sim i} &= \boldsymbol{Y}'_0 \boldsymbol{Y}'_0 - \boldsymbol{y}_i \boldsymbol{y}_i' \\ \boldsymbol{Y}'_{0,\sim i} \boldsymbol{y}_{T,\sim i} &= \boldsymbol{Y}'_0 \boldsymbol{y}_T - \boldsymbol{y}_i Y_{iT},\end{aligned}$$

where $\boldsymbol{y}_i = [Y_{it} : t \leq T_0]$ is the $i$th row of $\boldsymbol{Y}_0$. We do not assume that $\boldsymbol{Y}_0'\boldsymbol{Y}_0$ is nonsingular. As such, we use a generalized form of the Sherman-Morrison formula (Cline, 1965, Meyer, 1973) to obtain

$$(\boldsymbol{Y}_{0,\sim i}'\boldsymbol{Y}_{0,\sim i})^\dagger = (\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger + (1 - H_{ii}^u)^{-1}(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{y}_i \boldsymbol{y}_i'(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger. \tag{49}$$

Recall $\widehat{\boldsymbol{\alpha}} = (\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{Y}_0'\boldsymbol{y}_T$ and note $Y_{iT} - \boldsymbol{y}_i'\widehat{\boldsymbol{\alpha}}$ is the $i$th element of $\widehat{\boldsymbol{\varepsilon}}_T = \boldsymbol{H}_\perp^u \boldsymbol{y}_T$. Using these facts, we plug (49) into (47) to yield

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_{\sim i} &= \left[(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger + (1 - H_{ii}^u)^{-1}(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{y}_i \boldsymbol{y}_i'(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger\right](\boldsymbol{Y}_0'\boldsymbol{y}_T - \boldsymbol{y}_i Y_{iT}) \\
&= \widehat{\boldsymbol{\alpha}} - (\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{y}_i Y_{iT} + (1 - H_{ii}^u)^{-1}(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{y}_i \boldsymbol{y}_i'\widehat{\boldsymbol{\alpha}} - H_{ii}^u(1 - H_{ii}^u)^{-1}(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{y}_i Y_{iT} \\
&= \widehat{\boldsymbol{\alpha}} - (1 - H_{ii}^u)^{-1}(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{y}_i \widehat{\varepsilon}_{iT}.
\end{aligned} \tag{50}$$

Inserting (50) into our pseudo-estimate, we have

$$\begin{aligned}
\tilde{\boldsymbol{\alpha}}_i &= T_0\widehat{\boldsymbol{\alpha}} - (T_0 - 1)\left(\widehat{\boldsymbol{\alpha}} - (1 - H_{ii}^u)^{-1}(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{y}_i \widehat{\varepsilon}_{iT}\right) \\
&= \widehat{\boldsymbol{\alpha}} + (T_0 - 1)(1 - H_{ii}^u)^{-1}(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{y}_i \widehat{\varepsilon}_{iT}.
\end{aligned} \tag{51}$$

Inserting (51) into (48), we have

$$\begin{aligned}
\widehat{\boldsymbol{V}}^{\text{jack}} &= (\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \left(\sum_{i \leq N_0} \frac{\widehat{\varepsilon}_{iT}^2}{(1 - H_{ii}^u)^2} \boldsymbol{y}_i \boldsymbol{y}_i'\right)(\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \\
&= (\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger \boldsymbol{Y}_0' \boldsymbol{\Omega} \boldsymbol{Y}_0 (\boldsymbol{Y}_0'\boldsymbol{Y}_0)^\dagger,
\end{aligned}$$

where $\boldsymbol{\Omega}$ is a diagonal matrix with $\Omega_{ii} = \widehat{\varepsilon}_{iT}^2(1 - H_{ii}^u)^{-2}$. Equivalently, $\boldsymbol{\Omega} = \text{diag}([\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \circ \boldsymbol{I}]^\dagger [\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T])$. It then follows that

$$\boldsymbol{y}_N' \widehat{\boldsymbol{V}}^{\text{jack}} \boldsymbol{y}_N = \widehat{\boldsymbol{\beta}}' \boldsymbol{\Omega} \widehat{\boldsymbol{\beta}}.$$

To arrive at (44), we define $\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}} = \boldsymbol{\Omega}$. This corresponds to the EHW estimator with the jackknife correction. We derive (45) for $\widehat{\boldsymbol{\beta}}$ by applying the same arguments above. Now, we will evaluate the biases of $(\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}}, \widehat{\boldsymbol{\Sigma}}_N^{\text{jack}})$.

(i) [HZ model] Let Assumption 1 hold. We define $(\sigma_{iT}^{\text{hz}})^2 = \text{Var}(\varepsilon_{iT} | \boldsymbol{y}_N, \boldsymbol{Y}_0)$ for $i = 1, \ldots, N_0$. Observe that

$$\mathbb{E}[(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \circ \boldsymbol{I})^\dagger (\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T) | \boldsymbol{y}_N, \boldsymbol{Y}_0] = (\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \circ \boldsymbol{I})^\dagger \mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0]. \tag{52}$$

To evaluate (52), we follow the derivations of (30) and (32) to obtain

$$\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \boldsymbol{H}_\perp^u \boldsymbol{Y}_0 \boldsymbol{\alpha}^* = \boldsymbol{0} \tag{53}$$

$$\text{Cov}(\widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0) = \boldsymbol{H}_\perp^u \boldsymbol{\Sigma}_T^{\text{hz}} \boldsymbol{H}_\perp^u. \tag{54}$$

Recall that $\mathbb{E}[X^2] = \text{Var}(X) + \mathbb{E}[X]^2$ for any random variable $X$. Thus, combining (53) with (54) gives

$$\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0] = (\boldsymbol{H}_\perp^u \boldsymbol{\Sigma}_T^{\text{hz}} \boldsymbol{H}_\perp^u \circ \boldsymbol{I})\boldsymbol{1}. \tag{55}$$

Let $\widehat{\boldsymbol{\gamma}} = \mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0]$. By (55), the $\ell$th entry of $\widehat{\boldsymbol{\gamma}}$ can be written as

$$\widehat{\gamma}_\ell = \sum_{j \neq \ell} (H^u_{j\ell})^2 (\sigma^{\mathrm{hz}}_{jT})^2 + (1 - H^u_{\ell\ell})^2 (\sigma^{\mathrm{hz}}_{\ell T})^2,$$

where $H^u_{j\ell}$ is the $(j,\ell)$th entry of $\boldsymbol{H}^u$. In turn, this allows us to rewrite (55) as

$$\widehat{\boldsymbol{\gamma}} = (\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp) \boldsymbol{\Sigma}^{\mathrm{hz}}_T \mathbf{1}. \tag{56}$$

Next, let $\widehat{\boldsymbol{\zeta}} = (\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp \circ \boldsymbol{I})^{-1} \widehat{\boldsymbol{\gamma}}$. Notice that the $\ell$th entry of $\widehat{\boldsymbol{\zeta}}$ is given by

$$\widehat{\zeta}_\ell = (\sigma^{\mathrm{hz}}_{\ell T})^2 + \sum_{j \neq \ell} \frac{(H^u_{\ell j})^2}{(1 - H^u_{\ell\ell})^2} (\sigma^{\mathrm{hz}}_{jT})^2.$$

Therefore, $\mathrm{diag}(\widehat{\boldsymbol{\zeta}}) = \boldsymbol{\Sigma}^{\mathrm{hz}}_T + \boldsymbol{\Delta}^{\mathrm{hz}}$, where $\Delta^{\mathrm{hz}}_{\ell\ell} = \sum_{j \neq \ell} (\sigma^{\mathrm{hz}}_{jT})^2 (H^u_{\ell j})^2 (1 - H^u_{\ell\ell})^{-2}$ for $\ell = 1, \ldots, N_0$. Notice if $\max_\ell H^u_{\ell\ell} < 1$, then $(\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp \circ \boldsymbol{I})$ is nonsingular, i.e., the pseudo-inverse is precisely the inverse. In this situation, plugging the above into (52) gives

$$\begin{aligned}
\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\mathrm{jack}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0] &= \mathrm{diag}\left( (\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp \circ \boldsymbol{I})^{-1} \mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0] \right) \\
&= \mathrm{diag}\left( (\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp \circ \boldsymbol{I})^{-1} \widehat{\boldsymbol{\gamma}} \right) \\
&= \mathrm{diag}(\widehat{\boldsymbol{\zeta}}) \\
&= \boldsymbol{\Sigma}^{\mathrm{hz}}_T + \boldsymbol{\Delta}^{\mathrm{hz}}.
\end{aligned} \tag{57}$$

From this, we conclude that

$$\begin{aligned}
\mathbb{E}[\widehat{v}^{\mathrm{hz,jack}}_0 | \boldsymbol{y}_N, \boldsymbol{Y}_0] &= \widehat{\boldsymbol{\beta}}' \mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\mathrm{jack}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0] \widehat{\boldsymbol{\beta}} \\
&= \widehat{\boldsymbol{\beta}}' (\boldsymbol{\Sigma}^{\mathrm{hz}}_T + \boldsymbol{\Delta}^{\mathrm{hz}}) \widehat{\boldsymbol{\beta}} \\
&= v^{\mathrm{hz}}_0 + \widehat{\boldsymbol{\beta}}' \boldsymbol{\Delta}^{\mathrm{hz}} \widehat{\boldsymbol{\beta}},
\end{aligned}$$

where we note that $\widehat{\boldsymbol{\beta}}' \boldsymbol{\Delta}^{\mathrm{hz}} \widehat{\boldsymbol{\beta}} \geq 0$.

(ii) [VT model] Let Assumption 2 hold. Following the arguments above, we conclude $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\mathrm{jack}}_N | \boldsymbol{y}_T, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}^{\mathrm{vt}}_N + \boldsymbol{\Gamma}^{\mathrm{vt}}$, where $\Gamma^{\mathrm{vt}}_{\ell\ell} = \sum_{j \neq \ell} (\sigma^{\mathrm{vt}}_{Nj})^2 (H^v_{\ell j})^2 (1 - H^v_{\ell\ell})^{-2}$ for $\ell = 1, \ldots, T_0$. Thus, $\mathbb{E}[\widehat{v}^{\mathrm{vt,jack}}_0 | \boldsymbol{y}_T, \boldsymbol{Y}_0] = v^{\mathrm{vt}}_0 + \widehat{\boldsymbol{\alpha}}' \boldsymbol{\Gamma}^{\mathrm{vt}} \widehat{\boldsymbol{\alpha}}$, where we note that $\widehat{\boldsymbol{\alpha}}' \boldsymbol{\Gamma}^{\mathrm{vt}} \widehat{\boldsymbol{\alpha}} \geq 0$.

(ii) [DR model] Let Assumption 3 hold. We define $(\sigma^{\mathrm{dr}}_{iT})^2 = \mathrm{Var}(\varepsilon_{iT} | \boldsymbol{Y}_0)$ for $i = 1, \ldots, N_0$ and $(\sigma^{\mathrm{dr}}_{Nt})^2 = \mathrm{Var}(\varepsilon_{Nt} | \boldsymbol{Y}_0)$ for $t = 1, \ldots, T_0$. Following the arguments that led to (57), we obtain $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\mathrm{jack}}_T | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}^{\mathrm{dr}}_T + \boldsymbol{\Delta}^{\mathrm{dr}}$, where $\Delta^{\mathrm{dr}}_{\ell\ell} = \sum_{j \neq \ell} (\sigma^{\mathrm{dr}}_{jT})^2 (H^u_{\ell j})^2 (1 - H^u_{\ell\ell})^{-2}$ for $\ell = 1, \ldots, N_0$. Similarly, we obtain $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\mathrm{jack}}_N | \boldsymbol{Y}_0] = \boldsymbol{\Sigma}^{\mathrm{dr}}_N + \boldsymbol{\Gamma}^{\mathrm{dr}}$, where $\Gamma^{\mathrm{dr}}_{\ell\ell} = \sum_{j \neq \ell} (\sigma^{\mathrm{dr}}_{Nj})^2 (H^v_{\ell j})^2 (1 - H^v_{\ell\ell})^{-2}$ for $\ell = 1, \ldots, T_0$. Applying Lemma 9 then gives

$$\begin{aligned}
&\mathbb{E}[\widehat{v}^{\mathrm{dr,jack}}_0 | \boldsymbol{Y}_0] \\
&= v^{\mathrm{dr}}_0 + (\boldsymbol{H}^u \boldsymbol{\beta}^*)' \boldsymbol{\Delta}^{\mathrm{dr}} (\boldsymbol{H}^u \boldsymbol{\beta}^*) + (\boldsymbol{H}^v \boldsymbol{\alpha}^*)' \boldsymbol{\Gamma}^{\mathrm{dr}} (\boldsymbol{H}^v \boldsymbol{\alpha}^*) + \mathrm{tr}(\boldsymbol{Y}^\dagger_0 \boldsymbol{\Delta}^{\mathrm{dr}} (\boldsymbol{Y}'_0)^\dagger \boldsymbol{\Gamma}^{\mathrm{dr}}).
\end{aligned}$$

The proof is complete. *Q.E.D.*

### D.2. *Proof of Lemma 8*

PROOF: We adopt the strategy of Hartley et al. (1969) to prove our desired result.

(ii) [HZ model] Let Assumption 1 hold. As in the proof of Lemma 7, we define $\widehat{\boldsymbol{\varepsilon}}_T = \boldsymbol{H}^u_\perp \boldsymbol{y}_T$. Observe

$$\mathbb{E}[(\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)^{-1}(\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T)|\boldsymbol{y}_N, \boldsymbol{Y}_0] = (\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)^{-1}\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T|\boldsymbol{y}_N, \boldsymbol{Y}_0]. \tag{58}$$

To evaluate (58), we plug in (56) to obtain

$$\mathbb{E}[(\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)^{-1}(\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T)|\boldsymbol{y}_N, \boldsymbol{Y}_0] = (\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)^{-1}(\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)\boldsymbol{\Sigma}^{\text{hz}}_T \mathbf{1} = \boldsymbol{\Sigma}^{\text{hz}}_T \mathbf{1}. \tag{59}$$

Plugging (59) into (58) yields

$$\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\text{HRK}}_T|\boldsymbol{y}_N, \boldsymbol{Y}_0] = \text{diag}\left((\boldsymbol{H}^u_\perp \circ \boldsymbol{H}^u_\perp)^{-1}\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T|\boldsymbol{y}_N, \boldsymbol{Y}_0]\right) = \boldsymbol{\Sigma}^{\text{hz}}_T. \tag{60}$$

It then follows that $\mathbb{E}[\widehat{v}^{\text{hz,HRK}}_0|\boldsymbol{y}_N, \boldsymbol{Y}_0] = v^{\text{hz}}_0$.

(ii) [VT model] Let Assumption 2 hold. Following the same arguments as above, we conclude $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\text{HRK}}_N|\boldsymbol{y}_T, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}^{\text{vt}}_N$ and $\mathbb{E}[\widehat{v}^{\text{vt,HRK}}_0|\boldsymbol{y}_T, \boldsymbol{Y}_0] = v^{\text{vt}}_0$.

(ii) [DR model] Let Assumption 3 hold. Following the arguments that led to (60), we obtain $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\text{HRK}}_T|\boldsymbol{Y}_0] = \boldsymbol{\Sigma}^{\text{dr}}_T$ and $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}^{\text{HRK}}_N|\boldsymbol{Y}_0] = \boldsymbol{\Sigma}^{\text{dr}}_N$. Applying Lemma 9 then gives $\mathbb{E}[\widehat{v}^{\text{dr,HRK}}_0|\boldsymbol{Y}_0] = v^{\text{dr}}_0$. The proof is complete. *Q.E.D.*

### D.3. *Proof of Lemma 9*

PROOF: By linearity of expectations,

$$\mathbb{E}[\widehat{v}^{\text{dr}}_0|\boldsymbol{Y}_0] = \mathbb{E}[\widehat{v}^{\text{hz}}_0|\boldsymbol{Y}_0] + \mathbb{E}[\widehat{v}^{\text{vt}}_0|\boldsymbol{Y}_0] - \mathbb{E}[\text{tr}(\boldsymbol{Y}^\dagger_0 \widehat{\boldsymbol{\Sigma}}_T (\boldsymbol{Y}'_0)^\dagger \widehat{\boldsymbol{\Sigma}}_N)|\boldsymbol{Y}_0]. \tag{61}$$

We evaluate each term in (61).

Beginning with the first term, note that the randomness in $\widehat{\boldsymbol{\Sigma}}_T$ stems from $\boldsymbol{\varepsilon}_T$ and $\widehat{\boldsymbol{\beta}}$ is deterministic given $(\boldsymbol{\varepsilon}_N, \boldsymbol{Y}_0)$. As such, Assumption 3 with Lemma 5 gives

$$\begin{aligned}
\mathbb{E}[\widehat{v}^{\text{hz}}_0|\boldsymbol{Y}_0] &= \mathbb{E}[\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{\Sigma}}_T \widehat{\boldsymbol{\beta}}|\boldsymbol{Y}_0] \\
&= \mathbb{E}\left[\mathbb{E}[\widehat{\boldsymbol{\beta}}' \widehat{\boldsymbol{\Sigma}}_T \widehat{\boldsymbol{\beta}}|\boldsymbol{\varepsilon}_N, \boldsymbol{Y}_0]|\boldsymbol{Y}_0\right] \\
&= \mathbb{E}\left[\boldsymbol{y}'_N \boldsymbol{Y}^\dagger_0 \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T|\boldsymbol{Y}_0](\boldsymbol{Y}'_0)^\dagger \boldsymbol{y}_N|\boldsymbol{Y}_0\right] \\
&= \mathbb{E}\left[(\boldsymbol{Y}'_0\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}_N)\boldsymbol{Y}^\dagger_0 \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T|\boldsymbol{Y}_0](\boldsymbol{Y}'_0)^\dagger(\boldsymbol{Y}'_0\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}_N)|\boldsymbol{Y}_0\right] \\
&= (\boldsymbol{H}^u\boldsymbol{\beta}^*)'\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T|\boldsymbol{Y}_0](\boldsymbol{H}^u\boldsymbol{\beta}^*) + \text{tr}(\boldsymbol{Y}^\dagger_0 \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T|\boldsymbol{Y}_0](\boldsymbol{Y}'_0)^\dagger \boldsymbol{\Sigma}^{\text{dr}}_N).
\end{aligned}$$

By an analogous argument, we derive

$$\mathbb{E}[\widehat{v}^{\text{vt}}_0|\boldsymbol{Y}_0] = (\boldsymbol{H}^v\boldsymbol{\alpha}^*)'\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N|\boldsymbol{Y}_0](\boldsymbol{H}^v\boldsymbol{\alpha}^*) + \text{tr}(\boldsymbol{Y}^\dagger_0 \boldsymbol{\Sigma}^{\text{dr}}_T (\boldsymbol{Y}'_0)^\dagger \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N|\boldsymbol{Y}_0]).$$

Finally, we use the linearity of the trace operator with Assumption 3 to obtain

$$\begin{aligned}\mathbb{E}[\mathrm{tr}(\boldsymbol{Y}_0^\dagger \widehat{\boldsymbol{\Sigma}}_T (\boldsymbol{Y}_0')^\dagger \widehat{\boldsymbol{\Sigma}}_N)|\boldsymbol{Y}_0] &= \mathbb{E}\left[\mathbb{E}[\mathrm{tr}(\boldsymbol{Y}_0^\dagger \widehat{\boldsymbol{\Sigma}}_T (\boldsymbol{Y}_0')^\dagger \widehat{\boldsymbol{\Sigma}}_N)|\boldsymbol{\varepsilon}_N, \boldsymbol{Y}_0]|\boldsymbol{Y}_0\right] \\ &= \mathbb{E}\left[\mathrm{tr}(\boldsymbol{Y}_0^\dagger \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T|\boldsymbol{Y}_0](\boldsymbol{Y}_0')^\dagger \widehat{\boldsymbol{\Sigma}}_N)|\boldsymbol{Y}_0\right] \\ &= \mathrm{tr}(\boldsymbol{Y}_0^\dagger \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T|\boldsymbol{Y}_0](\boldsymbol{Y}_0')^\dagger \mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N|\boldsymbol{Y}_0]).\end{aligned}$$

Putting everything together completes the proof. *Q.E.D.*

## APPENDIX E: Principal Component Regression

The results in Section 4, which are stated for OLS, immediately extend to PCR by replacing $\boldsymbol{Y}_0$ with $\boldsymbol{Y}_0^{(k)}$ for any $k < R$. See Section 3 for details of the PCR method.

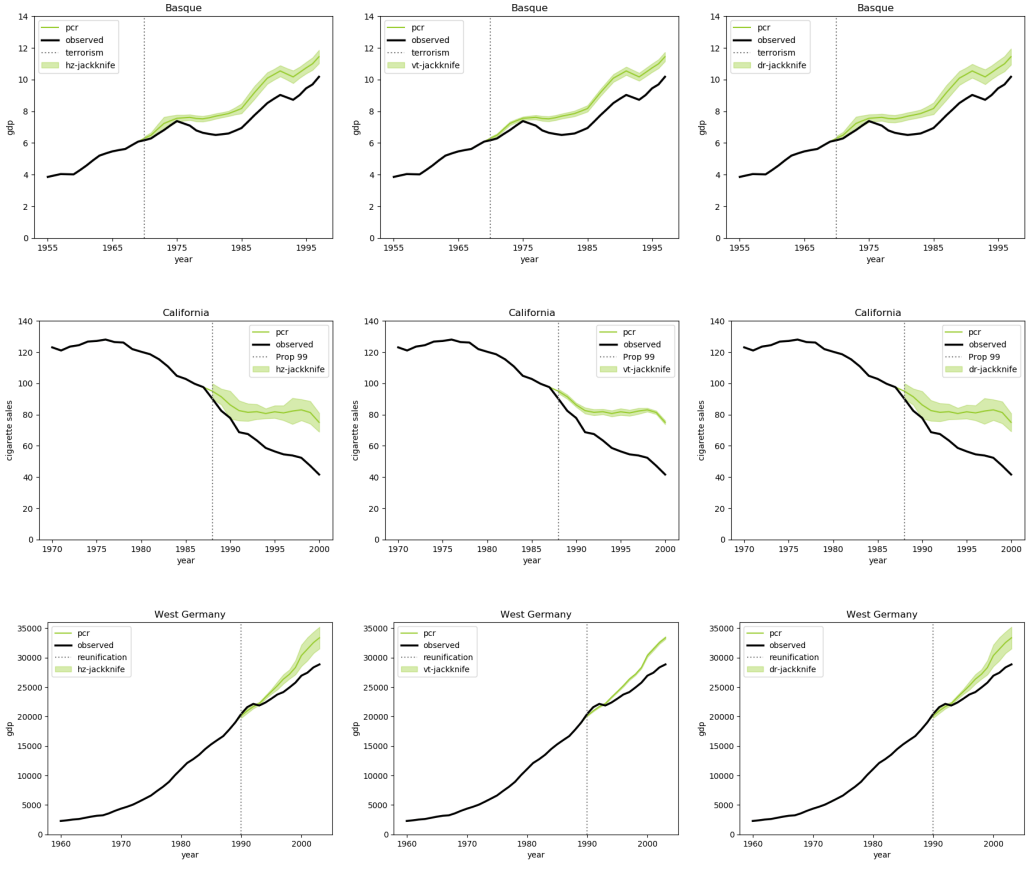
### E.1. *Comparing PCR to OLS*

Intuitively, PCR-based models operate under the belief that the data is inherently low-dimensional. We comment on several benefits of PCR over OLS. To begin, the HZ and VT OLS variance estimators constructed in Section C.2 can suffer from degeneracy when $N$ and $T$ are of different sizes. That is, if $N < T$, then the HZ in-sample error is likely zero (otherwise known as overfitting), which causes the HZ coverage to collapse on the point estimate; analogous statements hold for the VT coverage when $N > T$. The PCR-based variance estimators, on the other hand, can avoid degeneracy through the number of chosen principal components $k$ (regularization). On a related note, the nonsingularity conditions required for the jackknife and HRK variance estimators can also be by controlled by $k$. See Agarwal et al. (2021) for various methods on choosing $k$.

### E.2. *Empirical Applications—Extended*

Here, we extend our analysis in Section 5.3.3 to include results for PCR. We present the PCR-based confidence intervals for our three case studies in Figure E.1. For visualization ease, we only plot the jackknife intervals. Notably, the same conclusions drawn for OLS hold for PCR as well.

*Co-editor [Name Surname; will be inserted later] handled this manuscript.*




(a) HZ model. (b) VT model. (c) DR model.

FIGURE E.1.—PCR estimates with jackknife confidence intervals. From top to bottom, the rows are indexed by the Basque, California, and West Germany studies. From left to right, the columns are indexed by the HZ, VT, and DR models.