SUMMARY INDICES IN EMPIRICAL RESEARCH: PRACTICES, PITFALLS, AND PROPOSALS

Danil Fedchenko

Department of Economics, Northwestern University

This Version: 5.0.1

This paper examines the common practice of combining multiple outcomes into summary indices to assess the effect of an intervention. The main finding is that none of the widely used procedures are entirely free of problems: some may reverse the sign of the estimated effect, while others may inflate it artificially — except for the simplest one: taking the simple average of different outcomes. In addition, I show that correct inference requires accounting for data-dependent weights when computing standard errors and illustrate, using the data from some existing studies, that applying this correction can increase reported standard errors substantially. I also discuss common data manipulations, such as imputing missing values in index components, and show that the resulting indices often differ from what they are intended to represent. It is therefore unclear how the results of studies using them should be interpreted. This study aims to provide guidance for researchers using summary indices, to promote more accurate application and producing results that are transparent and interpretable.

Keywords:

JEL Classification:

Danil Fedchenko: danilfedchenko@u.northwestern.edu

1. INTRODUCTION

Does living in a safer neighborhood improve mental well-being? Does a development program empower women? Does the arrival of social media harm students' mental health? To address such broad and important questions, empirical researchers often measure multiple outcomes and combine them into a single *summary index*. Two approaches to constructing such indices dominate current empirical practice.

In some studies, such as Kling, Liebman, and Katz (2007), Chetty et al. (2011), Casey, Glennerster, and Miguel (2012), Finkelstein et al. (2012), Banerjee et al. (2015), Hoynes, Schanzenbach, and Almond (2016), Bursztyn, González, and Yanagizawa-Drott (2020), Christensen et al. (2021), Levy (2021), Stantcheva (2021), Braghieri, Levy, and Makarin (2022), and Bhatt et al. (2024) the summary index is constructed by averaging control-group standardized outcomes. For instance, Kling, Liebman, and Katz (2007), whose approach has since become the standard reference in the literature, average standardized measures of depression, distress, sleep quality, and a few other metrics to construct an index of mental health. This index is used to study whether the Moving to Opportunity program – randomly offering households in distressed neighborhoods the chance to move to safer areas – improved mental health. I refer to thus constructed summary index as the *scale-normalized* (**SN**) summary index.

In other studies, such as Anderson (2008), Currie et al. (2015), Haushofer and Shapiro (2016), Cantoni et al. (2017), Cantoni et al. (2019), Chen and Yang (2019), Alfonsi et al. (2020), Allcott et al. (2020), Asher and Novosad (2020), Baranov et al. (2020), Allcott, Gentzkow, and Song (2022), Egger et al. (2022), Carranza et al. (2022), Beraja et al. (2023), Kinnan et al. (2024), and Baseler et al. (2025), the summary index is constructed by taking a weighted average of standardized outcomes where weights are determined by the inverse of the covariance matrix of these outcomes. I refer to thus constructed summary index as the *inverse-covariance* (**IC**) summary index.

Despite the widespread use of these two procedures in leading economics journals, they statistical properties remain largely unexamined. As a result, applied researchers may lack clear guidance on which procedure, if any, best fits their setting, and readers of those studies may struggle to interpret results based on the summary indices. The aim of this research is to examine when widely used summary indices yield reliable and interpretable answers – and when they may mislead. The main conclusion is that *the most reliable approach is also the simplest*: taking the simple average of outcomes, or, when outcomes are measured on different scales, their control-group–standardized average – the **SN** index.

To reach this conclusion, I first examine how estimators based on summary indices should be interpreted. In empirical practice, summary indices are often used in estimation in place of – or sometimes in addition to – the original data from which they are derived. I identify the class of estimators for which such a replacement yields a weighted sum of the estimates that would be obtained using the original outcomes. I refer to this class as linear causal estimators. This class encompasses commonly used estimators such as difference-in-means, ordinary least squares (OLS), and two-stage least squares (2SLS/IV/fuzzy RDD). Applying summary indices in estimators outside this class, however, does not ensure this straightforward interpretation and hence requires additional analysis.

Different summary indices assign different weights to the component-level treatment effects. I find that **SN** summary index always assigns positive weights, by contrast **IC** may assign negative weights, creating the possibility of *sign-reversal* – a situation in which all component level estimates are negative while the index estimate is positive (or vice versa). I also show that the common practice of standardizing the summary index so that its sample variance equals one hinders interpretation by a violation of an *equal-effects invariance* criterion. This criterion requires that if all component-level treatment effects – when normalized by its control-group standard deviation – are equal, then the summary index estimate should also equal that same normalized effect size. When this criterion is violated, the summary index estimate can artificially inflate the estimated treatment effect. Therefore, as displayed in the first two rows of Table 1, among the two commonly used summary indices – **SN** and **IC**, and their standardized versions, only the non-standardized **SN** index satisfies both no sign-reversal and equal-effects invariance.

Second, I examine the validity of standard inference procedures routinely applied to summary index estimates. I demonstrate that because the weights used in **IC** and **SN** summary indices are data-dependent, they contribute to the asymptotic variance of the summary index estimator – a factor overlooked by current empirical practices as the standard statistical packages implicitly treat the weights as fixed. As a result, reported standard errors are inconsistent and confidence intervals invalid. This issue, however, does not affect the validity of tests of the null treatment effect across all components of the summary index.

Since often summary indices aggregate outcomes measured on different scales, **SN** and **IC** summary indices are scale invariant as both include standardization of components by the control group standard deviations in the first step. However, when the components of a summary index are measured on the same scale, prior standardization by the control-group standard deviation becomes unnecessary. In this case, the simple average of the outcomes can serve as the summary index. The resulting weights are data-independent, ensuring that standard inference procedures remain valid. These results are illustrated in the last three rows in Table 1, where the last column reports the summary index computed as a simple average of the components.

	IC index	**IC** index standardized	**SN** index	**SN** index standardized	Simple aver. index
No sign-reversal	✗	✗	✓	✓	✓
Equal-effects invariance	✓	✗	✓	✗	✓
Valid test of null effect	✓	✓	✓	✓	✓
Default s.e. valid	✗	✗	✗	✗	✓
Scale invariant	✓	✓	✓	✓	✗

Table 1: Comparison of summary indices.

As can be concluded from Table 1, the best scale-invarant summary index among popularly employed is the non-standardized **SN** summary index. However, when confidence intervals are reported, the standard errors reported by statistical sofwares must be corrected. I do this, by deriving the correct asymptotic variance of the summary index estimator, and establish the joint asymptotic normality of multiple summary index estimates. These results enable the construction of jointly valid confidence intervals for multiple indices. This offers a statistically rigorous and visually intuitive alternative to the resampling-based methods currently popular in practice, such as Westfall and Young (1993).

Finally, I analyze the power of the tests typically conducted in practice, to evaluate the popular belief[3] that summary indices offer substantial power advantages. My findings are that these beliefs are not supported by theoretical results. I show that conventionally used t-statistics can be interpreted as projections of component-level treatment effect estimates onto a particular

[3]"[**SN** index] gives us **maximal power** to detect an effect on social outcomes, if such an effect is present." Banerjee et al. (2015) (p. 49)

Baranov et al. (2020) refer to **IC** as "**the most efficient** weighted average of a set of outcomes" (p. 838)

"An index [**SN**] is useful for **increasing power** to the extent that all underlying components move in the same direction…", Bhatt et al. (2024) (p. 20)

"In order to examine heterogeneous effects with **more statistical precision**, column 6 reports on an inverse covariance weighted index [**IC** index] of labor and spending", Kinnan et al. (2024) (p. 25)

"[**SN** index] **improves statistical power** to detect effects that go in the same direction within a domain" Kling, Liebman, and Katz (2007) (p. 89) ex-ante unknown direction determined by the chosen index construction procedure, and the true treatment effect. As a result, such tests exhibit especially high power against alternatives aligned with this direction. However, they may have low power against alternatives that are nearly orthogonal to it. Since the true direction of the treatment effect is unknown, there is no way to guard against this latter possibility. Therefore, ex-ante no summary index seems to dominate in terms of power. I also discuss common data manipulation when components of the summary index contain missing values. The main conclusion here is that these manipulations usually based on some imputations of the missing values often do not result in what they intend to represent, however, the simple fix to the problem is to exploit the linearity directly without appealing to imputations.

Thus, the central message of this paper is that when treatment effects can be expressed as linear functions of the outcomes, summary indices appear to be appropriate tools for testing the null of no treatment effect. In this setting, no procedure seems to dominate: all yield asymptotically valid tests, each most powerful against some a priori unknown alternative. However, if the size or direction of the effect also matters and is to be reported, the only procedure among those commonly used that satisfies natural criteria pertaining to the estimator interpretation appears to be the **SN** summary index without final standardization. While standard statistical software tends to understate the associated uncertainty, I provide formulas for correct standard errors for any linear causal estimator.

Literature Conceptually, this work relates to the literature on treatment evaluation with multiple outcomes. This literature dates back to Hotelling (1931), who introduced omnibus tests based on the T^2 statistic. Since such tests provide no information about the direction of departures from the null, their usefulness is limited for decision-making, as it is crucial to know whether the treatment benefits or harms recipients. To address this limitation, O'Brien (1984) proposed projection-based tests that target specific one-dimensional alternatives. The two statistics studied in O'Brien (1984) are linear combinations of sample mean differences across outcomes: one a simple average, the other weighted by the inverse variance–covariance matrix – the direct precursors of the **SN** and **IC** summary indices examined in this paper, respectively. These ideas later entered applied economics through (i) Kling, Liebman, and Katz (2007), who introduced the **SN** index to evaluate the Moving to Opportunity program, and (ii) Anderson (2008), who proposed the **IC** index to study the effects of early intervention programs.

Another popular approach to the treatment effect evaluation with multiple outcomes is multiple hypothesis testing, for which there is a large established literature (see Romano and Wolf (2005), List, Shaikh, and Xu (2019) and references therein). This approach is rarely applied directly to the components of a summary index, likely because constructing an index is simpler, and because in some applications researchers are interested only in a broad evaluation of intervention effects rather than in identifying which specific outcomes are affected. In this literature, my analysis complements recent work by Viviano, Wuthrich, and Niehaus (2021), who developed a decision-theoretic framework for when to use a summary index versus multiple testing. While their focus is on deriving an optimal index from first principles, my contribution is to study the properties of the popular off-the-shelf indices widely used in practice.

Notation Throughout the paper, I generally use standard Latin or Greek letters for both vectors and scalars, without relying on boldface. Boldface is reserved for cases where it is important to indicate that an object is a vector or a matrix (rather than a scalar). For any $\delta > 0$, $B_\delta(\nu)$ denotes the ball of radius δ centered at ν. For any integer k, $\vec{1}_k$ denotes k-dimensional vector of ones, and $I_k = (e_{1,k} \ \cdots \ e_{k,k})$ is the $k \times k$ identity matrix, while $I\{\cdot\}$ denotes the indicator function. For any vector a, $a \gneq 0$ means that all components of a are non-negative and $a \neq 0$. Inequalities and operators involving vectors are to be interpreted componentwise. All $o(\cdot)$ and $o_p(\cdot)$ terms are to be understood as referring to the limit $n \to \infty$, where n as usual denotes the sample size.

Structure of the paper This paper is structured as follows: Section 2 describes the statistical framework and formally introduces procedures for constructing summary indices; Section 3 contains the main takeaway of the paper, with each discussed in a dedicated subsection. All formal results and assumptions are presented in Section 5. Finally, Section 6 summarizes the results and concludes the article. Proofs are presented in Appendix A.

2. SETUP

Data description and index construction The observable data, $\{W_i\}_{i=1}^n := \{\boldsymbol{y}_i, x_i, D_i\}_{i=1}^n$ is assumed to be an i.i.d. sample from some distribution $P \in \boldsymbol{P}$. Here, $\boldsymbol{y}_i \in \mathbb{R}^p$ is a vector of outcomes (the dependent variable), $D_i \in \{0, 1\}$ is a binary cause (treatment, policy intervention, etc.,) whose effect on $\boldsymbol{y}_i$ is the main object of interest, and $x_i \in \mathbb{R}^{k-1}$ represents additional covariates that include a constant term as a first component. It will be convenient to group all independent variables into a single vector, so I denote $\tilde{x}_i = (D_i, x_i')' \in \mathbb{R}^k$. This summarized in Table 2.

Object	Dimension	Description
$\boldsymbol{y}$	$p \geq 1$	Outcome
D	$\{0, 1\}$	Cause (treatment)
x	$k - 1 \geq 1$	Covariates & 1
$\tilde{x}$	k	$(D, x)'$
W	$p + k$	$(\boldsymbol{y}, \tilde{x})$

Table 2: List of variables

The **SN** summary index, introduced by Kling, Liebman, and Katz (2007), is constructed as follows:

PROCEDURE 1: ***Scale** Normalized Summary Index*. First, let $G_\ell \subseteq \{1, ..., p\}$ denote the set of outcome indices that comprise summary index ℓ; G_ℓ sometimes is referred to as the *domain*. Denote by $\boldsymbol{y}_{G_\ell} := \left(y_j : j \in G_\ell\right)$ the corresponding outcomes, and let $p_\ell = |G_\ell|$ denote their number.

1. Switch signs in $\boldsymbol{y}_{G_\ell}$ if necessary so that positive direction indicates a "better" outcome;
2. Demean all outcomes in $\boldsymbol{y}_{G_\ell}$ by subtracting the control group mean and dividing each outcome by its control group standard deviation, calling the transformed vector of outcomes $\tilde{\boldsymbol{y}}_{G_\ell}$;
3. Summary index $\boldsymbol{s}_\ell := \left(s_{\ell,1}, ..., s_{\ell,n}\right)'$ is then computed as:
$$s_{\ell,i} = \left(\vec{1}'_{p_\ell}\vec{1}_{p_\ell}\right)^{-1}\vec{1}'_{p_\ell}\tilde{\boldsymbol{y}}_{G_\ell,i}$$
4. Sometimes $\boldsymbol{s}_\ell$ is standardized one more time, i.e., Step 2. is performed again for $\boldsymbol{s}_\ell$.

In the description of Procedure 1, $s_{\ell,i}$ is a scalar computed from $\boldsymbol{y}_{G_\ell}$, which aggregates the outcomes in domain G_ℓ. In applied work, it is common to construct multiple indices by specifying $q > 1$ domains of outcomes and applying Procedure 1 separately to each. Domains may also overlap, so that some outcomes appear in more than one domain, as in Casey, Glennerster, and Miguel (2012). This general case – where $q > 1$ **SN** summary indices are constructed from possibly overlapping domains – is what I will refer to throughout this work as the "**SN** summary index."

The **IC** summary index, introduced by Anderson (2008), is constructed as follows:

PROCEDURE 2: ***I**nverse **C**ovariance Summary Index.*

1. Switch signs in $\boldsymbol{y}_{G_\ell}$ if necessary so that positive direction indicates a "better" outcome;
2. Demean all outcomes[4], divide each outcome by its control group standard deviation, call the transformed outcome $\tilde{\boldsymbol{y}}_{G_\ell}$, and compute the variance-covariance matrix of $\tilde{\boldsymbol{y}}_{G_\ell}$, $\tilde{\boldsymbol{V}}_{\ell,n} := \widehat{\mathrm{Var}}\left[\tilde{\boldsymbol{y}}_{G_\ell}\right]$;
3. Summary index $\boldsymbol{s}_\ell := \left(s_{\ell,1}, \ldots, s_{\ell,n}\right)'$ is computed then as:

$$s_{\ell,i} = \left(\vec{1}'_{p_\ell} \tilde{\boldsymbol{V}}_{\ell,n}^{-1} \vec{1}_{p_\ell}\right)^{-1} \vec{1}'_{p_\ell} \tilde{\boldsymbol{V}}_{\ell,n}^{-1} \, \tilde{\boldsymbol{y}}_{G_\ell,i}$$

4. Sometimes $\boldsymbol{s}_\ell$ is standardized one more time, i.e., Step 2. is performed again for $\boldsymbol{s}_\ell$.

The main difference between Procedure 1 and Procedure 2 is that the latter incorporates the variance–covariance matrix of the outcomes, using its inverse to weight outcomes within a domain. Because this weighting scheme resembles the Generalized Least Squares approach, it is sometimes referred to as GLS weighting. Anderson (2008) motivates this approach as follows: "The GLS weighting procedure […] increases efficiency by ensuring that outcomes that are highly correlated with each other receive less weight, while outcomes that are uncorrelated and thus represent new information receive more weight.". Yet the label "GLS" is somewhat misleading: the crucial difference is that GLS estimator uses the inverse of the *error* variance–covariance matrix, not that of the outcomes.

In some applications, Procedure 1 or Procedure 2 are applied to outcomes that are themselves summary indices, a case also covered by the theory developed in this paper. For instance, Haushofer and Shapiro (2016) constructed a women's empowerment index by aggregating two sub-indices: one measuring experiences of abuse and the other capturing attitudes toward gender-based violence.

Other summary index construction procedures I now briefly describe other summary index procedures that appear in applied work:

(i) *PCA-weighted summary index*. The first principal component of the sample correlation matrix of $\boldsymbol{y}_{G_\ell}$ is taken as the summary index. This method is particularly common in

[4]In some studies, outcomes are demeaned by subtracting the control group mean as in Procedure **SN**. As shown later, this difference is immaterial for the results (see the discussion after Definition 3).

constructing "wealth" (or "consumption") indices from data on ownership of household goods, as introduced by Filmer and Pritchett (2001).

(ii) *Economically weighted index*. Although less common, some studies (e.g., Bhatt et al. (2024)) use weights based on presumed or estimated social costs or benefits of outcomes. For instance, Bhatt et al. (2024) constructed an index of violent crimes, weighting each crime by its estimated social cost.

Estimand and estimator In this paper, I consider a class of parameters that I term *linear causal estimands*: the expected value of $\boldsymbol{y}$ weighted by a known mean-zero function of the data $\omega(\cdot)$.

Definition 1: *Linear causal estimand*. Linear causal estimand induced by weighting function $\omega(\cdot,\cdot)$ and a functional $\nu(P)$ has the form:

$$\boldsymbol{\tau}(P) = E_P[\boldsymbol{y}\,\omega(\tilde{x},\nu(P))] \in \mathbb{R}^p, \tag{1}$$

where $\omega(\tilde{x},\nu(P))$ is a known scalar-valued measurable function[5] such that

$$E_P[\omega(\tilde{x},\nu(P))] = 0, \forall P \in \boldsymbol{P}^{6}, \tag{2}$$

and $\nu(\cdot)$ is a known functional of the data distribution.

The term causal in Definition 1 refers to the mean-zero property of the weighting function $\omega(\cdot)$, which ensures that if $\boldsymbol{y}$ is independent of $\tilde{x}$, the estimand equals zero. Since the true data distribution P is unknown it is replaced by the empirical distribution $P_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{W_i}$, where δ_{W_i} is the Dirac measure that places unit mass on the data point W_i. This gives rise to an estimator, which I refer to as the linear causal estimator, and define as follows:

Definition 2: *Linear causal estimator*. Linear causal estimator induced by weighting function $\omega(\cdot,\cdot)$ and a functional $\nu(P)$ has the form:

$$\hat{\boldsymbol{\tau}}_n = \boldsymbol{\tau}(P_n) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\,\omega(\tilde{x}_i,\nu(P_n)). \tag{3}$$

[5] None of the results change when $\omega(\cdot)$ becomes a row vector-valued, in this case $\boldsymbol{\tau}$ will become a matrix. This generalization is not considered here for clarity of exposition.

[6] It is implicitly assumed in the definition above that all involved expected values exist. If necessary, one can restrict the domain $\boldsymbol{P}$ to include only those distributions for which the functional $\nu(\cdot)$ is well-defined, and all expectations from the definition above exist.

REMARK 1: As will become clear later (see discussion after formula (16)), the key to the approach in this paper is to view both the parameter of interest and its estimator as functionals – that is, as the result of applying a transformation to the underlying data distribution. This motivates the notation $\boldsymbol{\tau}(P)$ and $\nu(P)$ in Definition 1. Many well-known statistics can be expressed in this way. For example, the sample average:

$$\hat{\mu}_n := \frac{1}{n}\sum_{i=1}^{n} W_i = \sum_{i=1}^{n} W_i \cdot \frac{1}{n} = \int W \,\mathrm{d}P_n = \mu(P_n),$$

where $\mu(\cdot)$ is a *functional* – a mapping from the set of distributions to real numbers – defined for any distribution P as $\mu(P) := \int W \,\mathrm{d}P = E_P[W]$ whenever the integral exists. Similarly, conditional means, variances, and functions thereof are also functionals.

In practice, researchers often apply these estimators to summary indices in place of, or in addition to, applying them to the original outcomes $\boldsymbol{y}$. This yields a lower-, q-dimensional linear estimator:

$$\hat{\boldsymbol{\beta}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{s}_{i,n}\, \omega(\tilde{x}_i, \nu(P_n)), \tag{4}$$

where $\boldsymbol{s}_{i,n}$ denotes the summary index value of $\boldsymbol{y}_i$. In what follows I sometimes refer to $\hat{\boldsymbol{\beta}}_n$ as the *summary index estimator*. Each component $\hat{\beta}_\ell$, corresponds to the summary index constructed for the outcome domain G_ℓ, $\ell \in [q]$, as defined in Procedure 1.

EXAMPLE 1: *Least squares with vector-valued dependent variable*. Suppose $\nu(P) = E_P[\tilde{x}\tilde{x}']^{-1}$, and

$$\omega(\tilde{x}, \nu(P)) = \tilde{x}'\nu(P)e_{1,k} \text{ (recall } e_{1,k} = (1, 0, ..., 0)' \text{)},$$

then the coefficient $\boldsymbol{\tau}$ from the linear projection:

$$\boldsymbol{y} = \boldsymbol{\tau}D + \Gamma x + u, E_P[u\tilde{x}] = 0$$

is a linear estimand:

$$\boldsymbol{\tau} = E_P\left[\boldsymbol{y}\,\tilde{x}'\nu(P)e_{1,k}\right],$$

and since the second element of $\tilde{x}$ is 1,

$$e_{2,k}'E_P[\tilde{x}\tilde{x}'] = E_P[\tilde{x}'] \Rightarrow e_{2,k}' = E_P[\tilde{x}'](E_P[\tilde{x}\tilde{x}'])^{-1} \Rightarrow$$

$$\Rightarrow E_P[\omega(\tilde{x}, \nu(P))] = E_P\left[\tilde{x}'(E_P[\tilde{x}\tilde{x}'])^{-1}e_{1,k}\right] = e_{2,k}'e_{1,k} = 0;$$

the corresponding linear estimator which is the usual least squares estimator is:

$$\hat{\boldsymbol{\tau}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y_i}\tilde{x}_i'\nu(P_n)e_{1,k} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y_i}\tilde{x}_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i\tilde{x}_i'\right)^{-1} e_{1,k}.$$

3. TAKEAWAYS

In this section, I present the main takeaways of the paper. Sections 3.1 - 3.3 examine the relationship between the summary index estimator $\hat{\boldsymbol{\beta}}_n$ and the full-data linear estimator $\hat{\boldsymbol{\tau}}_n$. Sections 3.4 - Section 3.8 analyze the large-sample properties of $\hat{\boldsymbol{\beta}}_n$, deriving its asymptotic distribution and assessing how the choice of summary index procedure shapes the properties of statistical tests. Finally Sections 3.9 - 3.10 study common approaches to handling missing data in the components of summary indices, highlight their pitfalls, and propose simple corrections.

3.1. FOR LINEAR ESTIMATORS, $\hat{\boldsymbol{\beta}}_n$ – LINEAR TRANSFORMATION OF $\hat{\boldsymbol{\tau}}_n$.

The first important observation is that *all summary indices introduced in Section 2 belong to a class of affine dimension-reduction transformations of the original outcome vector* $\boldsymbol{y}$, which I refer to as linear summary index. The following discussion illustrates this point.

DEFINITION 3: *Linear summary index.* Let $\boldsymbol{y} \in \mathbb{R}^p$. A vector $\boldsymbol{s} \in \mathbb{R}^q$ is called a linear summary index of $\boldsymbol{y}$, induced by a matrix $A \in \mathbb{R}^{q\times p}$, with $q < p$, if:

$$\boldsymbol{s} = A\boldsymbol{y} + b, \tag{5}$$

for some $b \in \mathbb{R}^q$.

Notice that because the weighting function $\omega(\cdot)$ in the linear estimator definition has mean zero (condition (2)), any constant vector added to s cancels out in the estimator expression:

$$\hat{\boldsymbol{\beta}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{s}_{i,n}\, \omega(\tilde{x}_i, \nu(P_n)).$$

Therefore, the term b from the linear summary index definition (5) is irrelevant to the results. Hence, as far as the linear estimator is concerned, *it is immaterial whether the outcomes are demeaned or not – and, if demeaned, whether this is done using the control group mean or the* *full sample mean – at Step 2 of Procedure 1 and Procedure 2.* For this reason, in Definition 3 only the matrix A is said to induce the index, despite the presence of b in its expression.

To see that Procedures 1 and 2 described in Section 2 yield affine transformations of $\boldsymbol{y}$, note that each step in these procedures – subtraction, scalar and matrix multiplication, addition – is itself affine. Successive application of affine operations guarantees that the overall transformation takes the affine form (5). Similarly, the first principal component is a weighted linear combination of outcomes, with weights given by the entries of a specific eigenvector. Hence, the PCA-based summary index also takes the affine form (5). In all of the examples above, the matrix A that induces the linear summary index depends on the data. To reflect this dependence, I will henceforth denote it by A_n , with the subscript n indicating its data-driven nature.

The linearity of both the summary index transformation and the estimator implies a straightforward conclusion:

$$\hat{\boldsymbol{\beta}}_n = A_n \hat{\boldsymbol{\tau}}_n,$$

the linear estimator $\hat{\boldsymbol{\beta}}_n$, computed using the summary index, is itself a linear transformation of $\hat{\boldsymbol{\tau}}_n$ – the linear estimator based on the original outcome vector $\boldsymbol{y}$. The coefficients in this transformation are entries of the matrix A_n that induces the summary index.[7] For this result to hold, two ingredients are crucial: (i) the linearity of the summary index, and (ii) the linearity of the estimator with respect to the outcome vector $\boldsymbol{y}$. Thus, this result delineates the natural class of estimators for which the use of summary indices admits the intuitive interpretation of a weighted sum of the component-level effects. Employing summary indices outside of this class requires further analysis to ensure that the resulting estimates are interpreted correctly.

For linear estimators, the only distinction between different summary index construction procedures lies in the weights – the matrix A_n – that each assigns to the component estimates. In the next two sections, I introduce two criteria for these weights to yield a meaningful summary of treatment effects and show that only the specific version of the **SN** summary index (Procedure 1) satisfies both.

3.2. WEIGHTS CAN BE NEGATIVE. NO SIGN-REVERSAL CRITERION.

Since one of the purposes of a summary index is, as the name suggests, to provide a meaningful summary of the treatment effects on its component outcomes, it seems reasonable to propose the following *no sign-reversal criterion*: if all outcomes within a domain are affected by a cause

[7]Kling, Liebman, and Katz (2007) make the same observation but only for the difference in means estimator and **SN** summary index (see their footnote 10, p. 89) . In fact, the relation holds much more broadly: for the entire class of linear estimators and entire class of linear summary indices.

in the same direction, then the estimator based on the summary index should also reflect that same direction. Formally, for any summary index $\ell = 1, \ldots, q$:

$$\hat{\boldsymbol{\tau}}_{G_\ell} \gneq (\lneq, =)\, 0 \Rightarrow \hat{\beta}_\ell > (<, =)\, 0,$$

where $\hat{\boldsymbol{\tau}}_{G_\ell} = \left\{ \hat{\tau}_{n,j} : j \in G_\ell \right\}$ – components of $\hat{\boldsymbol{\tau}}_n$ that correspond to the summary index domain G_ℓ, and $a \gneq 0 \Longleftrightarrow a \geq 0, a \neq 0$.

This criterion is conceptually related to a large literature on the ability of lower-dimensional estimates to aggregate higher-dimensional parameters, typically in the form of weighted averages. Examples include De Chaisemartin and d'Haultfoeuille (2020), who analyze the two-way fixed effects estimator's ability to aggregate heterogeneous treatment effects (see also Goodman-Bacon (2021), Callaway and Sant'Anna (2021), Sun and Abraham (2021), and Borusyak, Jaravel, and Spiess (2024)); Goldsmith-Pinkham, Hull, and Kolesár (2024), who study how regression estimates capture treatment effects for different values of other covariates when treatment has multiple arms; and Bugni, Canay, and McBride (2023), who analyze how regression estimates capture the effects for different values of post-treatment, observable actions.

For the linear summary index,

$$\hat{\boldsymbol{\beta}}_n = A_n \hat{\boldsymbol{\tau}}_n \Rightarrow \text{each component } \hat{\beta}_\ell = a_\ell' \, \hat{\boldsymbol{\tau}}_{G_\ell},$$

where a_ℓ is $p_\ell -$ dimensional non-zero vector. Hence, for the linear summary index no sign-reversal holds if all entries of a_ℓ are non-negative for every summary index $\ell = 1, \ldots, q$.

In Section 5 I show that:

(i) For **SN** summary index :

$$a_\ell' \underset{+}{\propto} \vec{1}_{p_\ell}' \, \hat{S}_{\ell,0}^{-1}, \tag{6}$$

where $\underset{+}{\propto}$ denotes proportional with a positive coefficient, $\hat{\boldsymbol{S}}_{\ell,0} \in \mathbb{R}^{p_\ell \times p_\ell}$ is the diagonal matrix with the control group sample standard deviations of $\boldsymbol{y}_{G_\ell}$ on the diagonal.

(ii) For **IC** summary index :

$$a_\ell' \underset{+}{\propto} \vec{1}_{p_\ell}' \hat{\boldsymbol{S}}_{\ell,0} \hat{\boldsymbol{\Sigma}}_{\ell,n}^{-1}, \tag{7}$$

where $\hat{\boldsymbol{\Sigma}}_{\ell,n}$ is the sample variance-covariance matrix of $\boldsymbol{y}_{G_\ell}$.

(iii) For PCA-weighted summary index:

$$a_\ell' = \mathbf{v}' \operatorname{Diag}\left(\hat{\Sigma}_{\ell,n}\right)^{-\frac{1}{2}}, \tag{8}$$

where $\mathbf{v}$ is the eigenvector of the sample correlation matrix of $\boldsymbol{y}_{G_\ell}$ that corresponds to the largest eigenvalue.

Thus, ***SN** procedure satisfies the no sign-reversal criterion*, since the inverses of the control-group standard deviations are always positive. By contrast, *the **IC** procedure fails the no sign-reversal criterion*, because, as formula (7) demonstrates, it relies on the inverse variance–covariance matrix that might have negative entries. A concrete example of this failure, which leads to sign-reversal, is demonstrated in Example 2. Similarly, the PCA-weighting procedure fails the no sign-reversal criterion since the components of eigenvectors can be negative. Thus, *if a researcher uses the **IC** or PCA-weighted summary indices and perform linear estimation, it is advisable to verify that sign reversal does not occur*, as it is theoretically possible.

3.3. IF FINAL INDEX IS STANDARDIZED WEIGHTS DO NOT SUM TO ONE.

Another natural requirement for a summary index is that if all component treatment effects (normalized, for instance, by their standard deviation in the control group) are equal, then the index estimate should also equal that same value. I term this property *equal-effects invariance* criterion. Formally, a summary index procedure satisfies equal-effect invariance criterion if and only if for any summary index $\ell = 1, ..., q$ and for any $c \in \mathbb{R} \setminus \{0\}$:

$$\hat{S}_{\ell,0}^{-1}\,\hat{\boldsymbol{\tau}}_{G_\ell} = c\,\vec{1}_{p_\ell} \Rightarrow \hat{\beta}_\ell = c.$$

This requirement is *satisfied for **SN** and **IC** summary index procedures if and only if the final index is not standardized* (i.e., step 4. in Procedure 1, Procedure 2 is ignored). I show in Section 5 that if step 4 is performed for Procedure 1 then:

$$\hat{\beta}_\ell = \left(\vec{1}_{p_\ell}'\hat{S}_{\ell,0}^{-1}\hat{\Sigma}_{\ell,0}\hat{S}_{\ell,0}^{-1}\vec{1}_{p_\ell}\right)^{-\frac{1}{2}}\vec{1}_{p_\ell}'\hat{S}_{\ell,0}^{-1}\,\hat{\boldsymbol{\tau}}_{G_\ell},$$

and consequently,

$$\hat{\boldsymbol{\tau}}_{G_\ell} = c\,\hat{S}_{\ell,0}\,\vec{1}_{p_\ell} \Rightarrow \hat{\beta}_\ell = c\,\frac{p_\ell}{\sqrt{\vec{1}_{p_\ell}'\hat{S}_{\ell,0}^{-1}\hat{\Sigma}_{\ell,0}\hat{S}_{\ell,0}^{-1}\vec{1}_{p_\ell}}} \neq c,$$

if $c \neq 0$. For example, suppose the (within-domain) covariance matrix has equicorrelation structure:

$$\hat{\Sigma}_{\ell,0} = (1-\rho)I_{p_\ell} + \rho\,\vec{1}_{p_\ell}\vec{1}_{p_\ell}' = \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \dots & \dots & 1 \end{pmatrix}$$

for some $\rho \in \left(-\frac{1}{p_\ell - 1}, 1\right)$, which ensures positive definiteness. Then, under equal (normalized) component effects $\hat{S}_{\ell,0}^{-1}\,\hat{\boldsymbol{\tau}}_{G_\ell} = c\,\vec{1}_{p_\ell}$:

$$\frac{\hat{\beta}_\ell}{c} = \sqrt{\frac{p_\ell}{1 + (p_\ell - 1)\rho}} > 1,$$

with equality only if $\rho = 1$. Thus, the index inflates the effect when components are correlated but not perfectly collinear. Similarly, if step 4 is performed for Procedure 2 then:

$$\hat{\beta}_\ell = \left(\vec{1}'_{p_\ell}\hat{S}_{\ell,0}\hat{\Sigma}_\ell^{-1}\hat{\Sigma}_{\ell,0}\hat{\Sigma}_\ell^{-1}\hat{S}_{\ell,0}\vec{1}_{p_\ell}\right)^{-\frac{1}{2}}\vec{1}'_{p_\ell}\hat{S}_{\ell,0}\hat{\Sigma}_\ell^{-1}\,\hat{\boldsymbol{\tau}}_{G_\ell},$$

and so

$$\hat{\beta}_\ell = c\,\frac{\vec{1}'_{p_\ell}\hat{S}_{\ell,0}\hat{\Sigma}_\ell^{-1}\hat{S}_{\ell,0}\vec{1}'_{p_\ell}}{\sqrt{\vec{1}'_{p_\ell}\hat{S}_{\ell,0}\hat{\Sigma}_\ell^{-1}\hat{\Sigma}_{\ell,0}\hat{\Sigma}_\ell^{-1}\hat{S}_{\ell,0}\vec{1}_{p_\ell}}} \neq c.$$

Therefore, to satisfy the equal-effects invariance criterion, the final standardization step of an index construction procedure should be avoided.

3.4. STOCHASTIC A_n MIGHT AFFECT THE ASYMPTOTIC VARIANCE OF $\hat{\boldsymbol{\beta}}_n$.

Under some regularity conditions specified in Section 5, the relation between estimators $\hat{\boldsymbol{\beta}}_n = A_n\hat{\boldsymbol{\tau}}_n$ carries over to the population-level relation:

$$\operatorname{plim}_{n\to\infty}\hat{\boldsymbol{\beta}}_n := \boldsymbol{\beta} = A\,\boldsymbol{\tau},$$

where the matrix A is the probability limit of A_n. In practice, standard errors and confidence intervals for components of $\hat{\boldsymbol{\beta}}_n$ are routinely reported using standard statistical packages, which implicitly treat the summary index sample $\{s_{i,n}, i = 1, ..., n\}$ as an i.i.d. sample. But as formulas (6) - (8) demonstrate for commonly used summary indices the weighting matrix A_n is data-dependent, and therefore the resulting $s_{i,n}$ are not i.i.d. As a result, *the standard errors and confidence intervals computed under the i.i.d. assumption are in general inconsistent*.

To illustrate this point, in Section 5 I show that under mild regularity conditions $\hat{\boldsymbol{\tau}}_n$ can be expressed as:

$$\hat{\boldsymbol{\tau}}_n = \boldsymbol{\tau} + \frac{1}{n}\sum_{i=1}^{n}\vartheta_P(W_i) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{9}$$

for some known function $\vartheta_P(\cdot)$ called the *influence function.* If A_n were non-stochatic, so $A_n = A$ then

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) = A\,\sqrt{n}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}) =$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}A\vartheta_P(W_i) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}_P[A\vartheta_P(W)]) \tag{10}$$

However, when A_n is data-dependent and admits similar representation as (9):

$$A_n = A + \frac{1}{n}\sum_{i=1}^{n}\psi_P(W_i) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{11}$$

then,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (A \vartheta_P(W_i) + \psi_P(W_i)\boldsymbol{\tau}) + o_p(1), \tag{12}$$

therefore

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}_P[A\, \vartheta_P(W) + \psi_P(W)\, \boldsymbol{\tau}]\,). \tag{13}$$

Comparing this limit with (10), which applies under the assumption that A_n is fixed, shows that whether A_n is stochastic generally affects the asymptotic variance of $\hat{\boldsymbol{\beta}}_n$. As demonstrated in Section 5, conventional sofware-reported estimates are consistent for (10), but not for (13). In general, this results in *invalid confidence intervals for components of* $\hat{\boldsymbol{\beta}}_n$, even though such intervals are routinely reported in practice. That said, (13) shows one important exception: when $\boldsymbol{\tau} = 0$, conventional software packages do deliver correct variance estimates. In this case, the software-reported t-test of the null hypothesis $\boldsymbol{\tau} = 0$ is asymptotically valid. Thus, *if the sole objective is testing this hypothesis – and no confidence intervals for* $\boldsymbol{\beta}$ *are required – no further corrections are needed*. Notably, this conclusion holds regardless of the summary index used.

So, to obtain consistent standard errors, one must use the asymptotic variance formula (13), which requires computing the influence functions $\vartheta_P(\cdot)$ and $\psi_P(\cdot)$. In the next section, I demonstrate how this can be done.

3.5. COMPUTING THE CORRECTION TERM IN THE ASYMPTOTIC VARIANCE

When confidence intervals for components of $\boldsymbol{\beta}$ are of interest and the summary index relies on a stochastic, data-dependent weighting matrix, one must use the asymptotic variance formula (13). This formula incorporates the influence functions $\vartheta_P(\cdot)$ and $\psi_P(\cdot)$ arising from the decompositions of the linear estimator $\hat{\boldsymbol{\tau}}_n$ and the stochastic weighting matrix A_n, respectively. I show that for linear estimands and commonly used weighting matrices, the influence functions can be derived by differentiating a suitable real-valued function. This requires that the estimator be represented as a *functional of the distribution* – precisely the reason it was introduced in this form in Section 2.

To build intuition, consider a generic functional $\theta(P)$, if $\theta(\cdot)$ were instead a differentiable function defined on $\mathbb{R}^k$, then for any $x_0, h \in \mathbb{R}^k$:

$$\theta(x_0 + h) = \theta(x_0) + \langle \theta'(x_0), h \rangle + o(\|h\|), \tag{14}$$

and so for points x and x_0 within neighborhood of size $\|h\|$ of each other :

$$\theta(x) = \theta(x_0) + \langle \theta'(x_0), x - x_0 \rangle + o(\|h\|),$$

where $\theta'(x_0)$ is the gradient, and $\langle \cdot, \cdot \rangle$ is the usual scalar product in $\mathbb{R}^k$. The idea is to extend this relation to functionals defined on the space of distributions. For sufficiently large sample sizes n, the empirical distribution $P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{W_i}$ lies, roughly speaking, in a neighborhood of size $\frac{1}{\sqrt{n}}$ of the true distribution P. Therefore, if $\theta(\cdot)$ is a *differentiable functional*, a similar relation can be expected to hold in a probabilistic sense:

$$\theta(P_n) = \theta(P) + \left\langle \theta'(P), \frac{1}{n}\sum_{i=1}^{n} \delta_{W_i} - P \right\rangle + o_p\left(\frac{1}{\sqrt{n}}\right) \Rightarrow \text{by linearity of } \langle\cdot,\cdot\rangle$$

$$\Rightarrow \theta(P_n) = \theta(P) + \frac{1}{n}\sum_{i=1}^{n} \langle \theta'(P), \delta_{W_i} - P \rangle + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{15}$$

which is exactly the form used in the decompositions (9), (11). Most importantly, (15) reveals the form of the *influence function*:

$$\langle \theta'(P), \delta_{W_i} - P \rangle,$$

which can be obtained as an ordinary derivative. Defining a real-valued function $f(t)$ as:

$$f(t) := \theta\big(P + t\big(\delta_{W_i} - P\big)\big),$$

by applying the chain rule we have:

$$f'(0) = \langle \theta'(P), \delta_{W_i} - P \rangle. \tag{16}$$

Formula (16) demonstrates why it was important to view the linear estimand $\boldsymbol{\tau}(P)$ as a functional of the data distribution: doing so makes it possible to define the function $f(t)$ above, whose derivative at 0 equals the influence function. This, in turn, yields the decompositions (9) and (11) that ultimately lead to the correct asymptotic variance formula for $\hat{\boldsymbol{\beta}}_n$ in (13).

3.6. DEMONSTRATION FOR THE LINEAR ESTIMAND.

For the linear estimand $\boldsymbol{\tau}(P) = E_P[\boldsymbol{y}\,\omega(\tilde{x}, \nu(P))]$ we have:

$$f(t) = \tau\Big(P + t\big(\delta_{W_i} - P\big)\Big) = E_P\Big[\boldsymbol{y}\,\omega\Big(\tilde{x}, \nu\Big(P + t\big(\delta_{W_i} - P\big)\Big)\Big)\Big] +$$

$$+ t\Big(\boldsymbol{y}_i\,\omega\Big(\tilde{x}_i, \nu\Big(P + t\big(\delta_{W_i} - P\big)\Big)\Big) - E_P\Big[\boldsymbol{y}\,\omega\Big(\tilde{x}, \nu\Big(P + t\big(\delta_{W_i} - P\big)\Big)\Big)\Big]\Big),$$

taking the derivative with respect to t and evaluating it at $t = 0$ yields the influence function:

$$\vartheta_P(W_i) := f'(0) = E_P\left[\boldsymbol{y}\left\langle \frac{\partial \omega(\tilde{x}, \nu)}{\partial \nu}\bigg|_{\nu=\nu(P)}, \langle \nu'(P), \delta_{W_i} - P \rangle \right\rangle\right] +$$

$$+ \boldsymbol{y}_i \omega(\tilde{x}_i, \nu(P)) - E_P[\boldsymbol{y}\,\omega(\tilde{x}, \nu(P))] \tag{17}$$

The first term in (17) involves $\langle \nu'(P), \delta_{W_i} - P \rangle$ – the influence function of an anxiliary functional $\nu(P)$, and its scalar product with the derivative of $\omega(\tilde{x}, \nu)$. The form of this scalar product depends on the spaces on which the functionals operate. For example, in the least-squares estimator (Example 1), the functional $\nu(P) = E_P[\tilde{x}\tilde{x}']^{-1}$ is $k \times k$ matrix. This implies: (i) its influence function is also $k \times k$ dimensional matrix; and (ii) the derivative $\frac{\partial \omega(\tilde{x}, \nu)}{\partial \nu}|_{\nu=\nu(P)}$

of a scalar-valued function is likewise $k \times k$ matrix. Hence, in this case the scalar product in (17) is the Frobenius inner product, defined as:

$$\langle \boldsymbol{M}_1, \boldsymbol{M_2} \rangle = \operatorname{tr}(\boldsymbol{M}_1' \boldsymbol{M}_2), \forall\, \boldsymbol{M}_1, \boldsymbol{M}_2 \in \mathbb{R}^{k \times k}.$$

Hence *the general recipe of finidng the influence function of a linear estimand* $\boldsymbol{\tau}(P)$ is:

1. Compute $\langle \nu'(P), \delta_{W_i} - P \rangle$ – the influence function of $\nu(P)$ via (16),
2. Compute the derivative $\frac{\partial \omega(\tilde{x}, \nu)}{\partial \nu}|_{\nu = \nu(P)}$ and its scalar product with $\langle \nu'(P), \delta_{W_i} - P \rangle$,
3. Compute the influence function of $\boldsymbol{\tau}(P)$ via (17).

The following example illustrates this procedure.

EXAMPLE 1: *Least squares with vector-valued dependent variable.* (Continue: the influence function of $\boldsymbol{\tau}(P)$): As shown in Example 1 above,

$$\nu(P) = E_P[\tilde{x}\tilde{x}']^{-1}, \omega(\tilde{x}, \nu(P)) = \tilde{x}'\nu(P)e_{1,k}.$$

1. The influence function of $\nu(P)$ via (16) :

$$\langle \nu'(P), \delta_{W_i} - P \rangle = \frac{\mathrm{d}}{\mathrm{d}t}[E_P[\tilde{x}\tilde{x}'] + t(\tilde{x}_i\tilde{x}_i' - E_P[\tilde{x}\tilde{x}'])]^{-1}\bigg|_{t=0} =$$

$$= -E_P[\tilde{x}\tilde{x}']^{-1}(\tilde{x}_i\tilde{x}_i' - E_P[\tilde{x}\tilde{x}'])E_P[\tilde{x}\tilde{x}']^{-1},$$

   where I used the basic identity for the derivative of the inverse of a square matrix $\boldsymbol{M}$:

$$\frac{\mathrm{d}\boldsymbol{M}^{-1}}{\mathrm{d}t} = -\boldsymbol{M}^{-1}\frac{\mathrm{d}\boldsymbol{M}}{\mathrm{d}t}\boldsymbol{M}^{-1},$$

2. The derivative and the scalar product:

$$\frac{\partial \omega(\tilde{x}, \nu)}{\partial \nu} = \tilde{x}e_{1,k}' \Rightarrow \left\langle \frac{\partial \omega(\tilde{x}, \nu)}{\partial \nu}, \langle \nu'(P), \delta_{W_i} - P \rangle \right\rangle =$$

$$= -\operatorname{tr}\Big(e_{1,k}\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}(\tilde{x}_i\tilde{x}_i' - E_P[\tilde{x}\tilde{x}'])E_P[\tilde{x}\tilde{x}']^{-1}\Big) =$$

$$= -\Big(\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}\tilde{x}_i\Big)\omega(\tilde{x}_i, \nu(P)) + \omega(\tilde{x}, \nu(P)),$$

   where I used the cyclic property of the trace operator to obtan the last equality.

3. The influence function of $\boldsymbol{\tau}(P)$ via (17):

$$\vartheta_P(W_i) = -E_P\Big[\boldsymbol{y}\,\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}\Big]\tilde{x}_i\,\omega(\tilde{x}_i, \nu(P)) + E_P[\boldsymbol{y}\,\omega(\tilde{x}, \nu(P))] +$$

$$+\,\boldsymbol{y}_i\,\omega(\tilde{x}_i, \nu(P)) - E_P[\boldsymbol{y}\,\omega(\tilde{x}, \nu(P))] = \boldsymbol{u}_i\,\omega(\tilde{x}_i, \nu(P)),$$

   where I used the fact that $E_P\Big[\boldsymbol{y}\,\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}\Big]\tilde{x}_i = \boldsymbol{y}_i - \boldsymbol{u}_i$

To obtain $\psi_P(\cdot)$, the influence function of the weighting matrix A_n, note that for the **SN** and **IC** summary indices, A_n is a differentiable function of conditional and unconditional variances, which themselves can be expressed as functionals of the distribution. Hence, A_n is itself a differentiable functional, and its influence function $\psi_P(\cdot)$ can be derived by following the same steps as above, together with the chain rule. In Section 5, I derive $\psi_P(\cdot)$ explicitly for the **SN** summary index without final standardization.

3.7. DO SUMMARY INDICES IMPROVE EFFICIENCY?

The asymptotic distribution:

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}_P[A\,\vartheta_P(W) + \psi_P(W)\,\boldsymbol{\tau}]\,)$$

given in (13) provides a framework for assessing whether summary indices deliver substantial efficiency gains, as is often claimed (see, e.g., Footnote 3). The summary index procedure enters the asymptotic variance through two channels: (i) the probability limit A and its product with $\vartheta_P(\cdot)$, the influence function of $\hat{\boldsymbol{\tau}}_n$; and (ii) the random fluctuations of A_n, $\psi_P(\cdot)$, which matter only when $\boldsymbol{\tau} \neq 0$. Thus, minimizing the asymptotic variance of $\hat{\boldsymbol{\beta}}_n$ requires considering both channels jointly, and the choice of index procedure should be guided by their combined effect. It is clear, however, that neither **SN** nor **IC** summary indices achieve this: even if $\boldsymbol{\tau}$ is assumed to be zero, their constructions do not take into account the variation in the estimation of $\hat{\boldsymbol{\tau}}_n$ – that is, $\mathrm{Var}_P[\vartheta_P(W)]$. As a result, in general, none of the commonly used procedures provide systematic power advantages.

To refine the analysis, I turn to the standard local alternative framework and show that t-tests based on summary indices each target a specific local alternative, achieving maximal power against it, while their performance against other alternatives depends on the alternative and can be close to the size of the test for some alternatives.

3.8. LOCAL POWER OF t-TEST DEPENDS ON THE SUMMARY INDEX. NONE IS UNIFORMLY BEST.

I adopt the standard framework of a sequence of local alternatives. In the present multidimensional setting, this should be interpreted as a sequence of local directions in $\mathbb{R}^p$. Specifically, assume that $\boldsymbol{\tau} = \frac{\boldsymbol{\delta}}{\sqrt{n}}$, so that the estimator admits the representation:

$$\hat{\boldsymbol{\tau}}_n = \frac{\boldsymbol{\delta}}{\sqrt{n}} + \frac{1}{n}\sum_{i=1}^{n} \vartheta_P(W_i) + o_p\left(\frac{1}{n}\right), \tag{18}$$

where $E_P[\vartheta_P(W)] = 0, E_P[\vartheta_P(W)\vartheta_P'(W)] = V$ – a positive definite matrix, and $\boldsymbol{\delta} \in \mathbb{R}^p \setminus \{\mathbf{0}\}$ – local alternative direction. The representation in (18), and in particular the scaling $\boldsymbol{\tau} = \frac{\boldsymbol{\delta}}{\sqrt{n}}$, formalizes the idea that the "true" direction $\boldsymbol{\tau}$ is small in magnitude relative to the sample size. For illustrative purposes, suppose $q = 1$, meaning that only a single summary index is constructed. This simplification is motivated by the observation that inference procedures in empirical studies are typically based on the marginal normality of individual indices, although some adjustments for multiple testing are often applied. In this case, the scalar linear estimator $\hat{\beta}_n$ is computed, and let $\widetilde{\text{s.e.}}\left(\hat{\beta}\right)$ denote its standard error estimator. Consider the corresponding t-statistic, $\hat{T}_n$. Using (18) simple calculations give:

$$\hat{T}_n = \frac{\sqrt{n}\hat{\beta}_n}{\widetilde{\text{s.e.}}\left(\hat{\beta}\right)} \xrightarrow{d} \underbrace{\frac{A\boldsymbol{\delta}}{\sqrt{AVA'}}}_{:=\Delta} + \mathcal{N}(0,1), \tag{19}$$

and so the rejection probability:

$$P\left(\left|\hat{T}_n\right| > z_{1-\frac{\alpha}{2}}\right) \to 1 - \Phi\left(z_{1-\frac{\alpha}{2}} - |\Delta|\right) + \Phi\left(-z_{1-\frac{\alpha}{2}} - |\Delta|\right).$$

By Cauchy-Schwartz inequality:

$$|\Delta| = \frac{\|A\boldsymbol{\delta}\|}{\sqrt{AVA'}} \leq \sqrt{\boldsymbol{\delta}'V^{-1}\boldsymbol{\delta}},$$

where the upper bound on the right hand side is achieved at

$$A^* \propto \boldsymbol{\delta}'V^{-1}. \tag{20}$$

Hence, if a linear index is induced by some A_n that is consistent for $A^* \propto \boldsymbol{\delta}'V^{-1}$ then Δ would achieve its largest possible value, and hence the t-test of the null $H_0 : \boldsymbol{\tau} = 0$ based on thus constructed summary index will have the largest asymptotic local power among any other summary index construction procedures.

For **SN** and **IC** summary indices, as shown in (6) and (7), $A^{\mathbf{SN}} \propto \vec{1}'S_0^{-1}$, $A^{\mathbf{IC}} \propto \vec{1}'S_0\Sigma^{-1}$, where S_0 is a diagonal matrix with control group standard deviations on the diagonal, and Σ is the variance covariance matrix of $\boldsymbol{y}$. Thus, **SN** attains the maximal local power if $\boldsymbol{\delta}^*_{\mathbf{SN}} \propto VS_0^{-1}\vec{1}$; while **IC** does so if $\boldsymbol{\delta}^*_{\mathbf{IC}} \propto V\Sigma^{-1}S_0\vec{1}$ – a subset of possible directions that is somewhat arbitrary and, more importantly, ex ante unknown. Consequently, these widely used *summary index procedures can deliver high-powered tests (in the local power sense) only for a limited set of the ex-ante unknown directions*. In general, little can be said about power, since $|\Delta|$ may be arbitrarily close to zero, making the power equal to the size. Overall, broad claims about the power advantages of summary indices appear overly optimistic and lack theoretical support.

Remark 2: Notice that if one attempts to estimate the optimal direction $\boldsymbol{\delta}^*$ from the data and then construct the corresponding optimal weighting matrix A_n^* by (20), using the same data to construct an estimator of V, the t-statistics $\hat{T}_n$, will no longer be asymptotically normal, instead it will have $\sqrt{\chi_p^2}$ asymptotic distribution; and the test will become equivalent to the omnibus Hotelling (1931)'s T^2 test. In a recent work, Anderson and Magruder (2022) propose using a hold-out sample to estimate the optimal direction and derive the corresponding weighting matrix, and then employing another independent sample for testing.

3.9. Missing outcomes: imputations distort linearity.

It is not uncommon in practice to encounter missing values for some components of $\boldsymbol{y}$. For example, if there are 3 outcomes, a dataset might look as displayed in Table 3.

$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	$\boldsymbol{y_3}$	$\boldsymbol{D}$	$\boldsymbol{x}$
N/A	$y_{2,1}$	N/A	1	x_1
$y_{1,2}$	N/A	$y_{3,2}$	1	x_2
$y_{1,3}$	$y_{2,3}$	N/A	1	x_3
$y_{1,4}$	N/A	$y_{3,4}$	1	x_4
N/A	$y_{2,5}$	N/A	0	x_5
$y_{1,6}$	N/A	$y_{3,6}$	0	x_6
$y_{1,7}$	$y_{2,7}$	N/A	0	x_7
$y_{1,8}$	$y_{2,8}$	$y_{3,8}$	0	x_8

Table 3: Observed data with missing values denoted by N/A.

Missing data generally pose a threat to identification, often leaving many parameters only partially identified. Point identification can be restored by imposing assumptions on the missingness mechanism. For example, the popular missing-at-random assumption requires that the distribution of observed data coincides with the distribution of the unobserved, missing data.

Under this assumption, observations with missing values can simply be dropped, and the estimator based on the remaining complete data will be consistent. However, dropping an entire row when only a few components are missing can be inefficient, so researchers often try to use all available data. Specifically, when some components of the summary index are missing, as in Table 3, instead of dropping first 7 rows, applied researchers typically adopt one of the following two approaches.

For the **SN** procedure, Kling, Liebman, and Katz (2007) proposed imputing missing components with the treatment-group mean. For the dataset in Table 3, the resulting dataset under this imputation scheme is shown below:

$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	$\boldsymbol{y_3}$	$\boldsymbol{D}$	$\boldsymbol{x}$
$(y_{1,2}+y_{1,3}+y_{1,4})/3$	$y_{2,1}$	$(y_{3,2}+y_{3,4})/2$	1	x_1
$y_{1,2}$	$(y_{2,1}+y_{2,3})/2$	$y_{3,2}$	1	x_2
$y_{1,3}$	$y_{2,3}$	$(y_{3,2}+y_{3,4})/2$	1	x_3
$y_{1,4}$	$(y_{2,1}+y_{2,3})/2$	$y_{3,4}$	1	x_4
$(y_{1,6}+y_{1,7}+y_{1,8})/3$	$y_{2,5}$	$(y_{3,6}+y_{3,8})/2$	0	x_5
$y_{1,6}$	$(y_{2,5}+y_{2,7}+y_{2,8})/3$	$y_{3,6}$	0	x_6
$y_{1,7}$	$y_{2,7}$	$(y_{3,6}+y_{3,8})/2$	0	x_7
$y_{1,8}$	$y_{2,8}$	$y_{3,8}$	0	x_8

Table 4: Dataset when missing values are imputed with the treatment assignment group mean as suggested by Kling, Liebman, and Katz (2007)

When the estimand of interest is the simple difference in means , imputing missing observations in this way yields numerically the same estimator for each component as simply dropping the missing observations. For example, using Table 4, it is straightforward to see that:

$$\hat{\tau}_n^1 = \bar{y}_{1,1} - \bar{y}_{1,0} = \frac{y_{1,2}+y_{1,3}+y_{1,4}}{3} - \frac{y_{1,6}+y_{1,7}+y_{1,8}}{3} \equiv$$

$$\equiv \frac{1}{4}\left(y_{1,2}+y_{1,3}+y_{1,4}+\frac{y_{1,2}+y_{1,3}+y_{1,4}}{3}\right) - \frac{1}{4}\left(y_{1,6}+y_{1,7}+y_{1,8}+\frac{y_{1,6}+y_{1,7}+y_{1,8}}{3}\right).$$

As argued above, under the missing-at-random assumption, this estimator that just ignores missing data is consistent for the true difference in means, $\hat{\boldsymbol{\tau}}_n \xrightarrow{p} \boldsymbol{\tau}$. However, with this imputation scheme, when **SN** summary index is constructed, the control-group standard deviations that are used in the weighting matrix are underestimated: imputing the mean does not add to the squared deviations from the mean, but it does increase the number of observations (the denominator in the variance estimator). As a result, even under missing-at-random, $A_n \overset{p}{\nrightarrow} A$. Hence, in general, under this imputation scheme, even when the assumptions required for point identification with missing data hold,

$$\text{plim}_{n\to\infty} \hat{\boldsymbol{\beta}}_n \neq A\boldsymbol{\tau}.$$

For the **IC** procedure, Anderson (2008) proposed ignoring missing values when the summary index is constructed – effectively treating them as zeros when computing the weighted average – claiming that the procedure "uses all of the available data, but it weights outcomes with fewer missing values more heavily.". In reality, under this imputation scheme, the weight assigned to an individual observation for a given outcome component can vary depending on how many observations for other components are missing for the same individual. Moreover, the weights in the **IC** procedure depend on the covariance matrix of the outcomes and are therefore not determined solely by the number of missing values. As Table 5 demonstrates, the claim does not hold even on average across components, nor even when the covariance structure is ignored and equal weights are imposed.

$\boldsymbol{y_1}$	$\boldsymbol{y_2}$	$\boldsymbol{y_3}$
N/A	$y_{2,1}$	N/A
$y_{1,2}$	N/A	$y_{3,2}$
$y_{1,3}$	$y_{2,3}$	N/A
$y_{1,4}$	N/A	$y_{3,4}$
N/A	$y_{2,5}$	N/A
$y_{1,6}$	N/A	$y_{3,6}$
$y_{1,7}$	$y_{2,7}$	N/A
$y_{1,8}$	$y_{2,8}$	$y_{3,8}$

weight 1	weight 2	weight 3
0	1	0
$\frac{1}{2}$	0	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	0
$\frac{1}{2}$	0	$\frac{1}{2}$
0	1	0
$\frac{1}{2}$	0	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	0
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
17/48	**20/48**	**11/48**

Table 5: Observed data with missing values denoted by N/A (left) and corresponding weights used in constructing the equally weighted index (right). The final row reports in boldface the average weight assigned to each outcome.

As can be seen, although the first outcome has the fewest missing values, the average weight assigned to it is $\frac{17}{48}$, which is not the largest. In fact, this procedure introduces complex dependence by assigning different weights to different observations depending on how many other outcomes for this observations are missing. *Thus, commonly used schemes for handling missing data fail to achieve their intended purposes, even under assumptions on the missing-data mechanism that would otherwise resolve issues of point identification.*

3.10. MISSING OUTCOMES: EXPLOITING LINEARITY WITHOUT IMPUTATIONS.

Complications described above can be avoided by using the relation $\hat{\boldsymbol{\beta}}_n = A_n \hat{\boldsymbol{\tau}}_n$ and the influence function approach outlined above. This also provides a formal way to address identification problems caused by missing data and to articulate assumptions about the missing-data mechanism. If the mechanism ensures that both $\boldsymbol{\tau}(P)$ and A are point-identified and can be estimated consistently, then $\boldsymbol{\beta} = A\boldsymbol{\tau}$ is likewise point-identified and consistently estimable by the product $A_n\hat{\boldsymbol{\tau}}_n$. For inference, correct standard errors can be obtained by the same steps as above, with influence functions adjusted to account for the treatment of missing values.

For example, with the data from Table 3, under missing-at-random assumption, datapoints $i = 2, 3, 4, 6, 7, 8$ are used to compute $\hat{\tau}_n^1$, datapoints $i = 1, 3, 5, 7, 8$ are used to compute $\hat{\tau}_n^2$; and datapoints $i = 2, 4, 6, 8$ are used to compute $\hat{\tau}_n^3$. Similarly, observations $i = 6, 7, 8$ are used to compute the control group standard deviation for y_1, $i = 5, 7, 8$ for y_2, and $i = 6, 8$ for y_3[8] This approach makes use of all available data, but because different components of $\hat{\boldsymbol{\tau}}_n$ are based on different subsamples, it differs from treating missing outcomes as zeros when constructing the index, as in Anderson (2008).

The influence functions derived above must be adjusted for the case with missing values. Let $\{\boldsymbol{z}_i\}_{i=1}^n$ be binary indicator vectors denoting which components of $\boldsymbol{y}_i$ are missing, with $z_{\ell,i} = 1 \iff y_{\ell,i}$ is observed. It can be shown that, *under the missing-at-random assumption, the influence functions from the complete-data case should be multiplied by* $\frac{\boldsymbol{z}_i}{E_P[\boldsymbol{z}]}$, where the division of two vectors is understood component-wise (see Chapter 7 in Tsiatis (2006)). Then the asymptotic distribution of $\sqrt{n}\left(\hat{\beta}_n - \beta\right)$ can be derived as shown above.

4. EMPIRICAL ILLUSTRATION

	Original study (**IC** index)	**SN** index (default s.e.)	**SN** index (correct s.e.)	Simple aver. index
New road	0.410	0.182	0.182	0.060
	(0.187)	(0.101)	(0.180)	(0.031)
p-value	0.03	0.07	0.31	0.06

Table 6: Impact of new road on the transportation index (Asher and Novosad (2020))

5. FORMAL RESULTS

This section provides the formal statements of the results summarized in Section 3 and develops the theoretical foundations of the linear causal estimator. In particular, I set out the assumptions

[8] Note that this is not how STATA's `correlate` command handles missing values by default; instead, it drops all rows with any missing values.

under which this class of estimators is consistent and admits an asymptotically linear representation; I introduce the appropriate notion of functional differentiability that validates the informal influence function arguments in Section 3.5; and I establish the consistency of variance estimators obtained by plugging in sample analogues for unknown population parameters. While the technical tools employed here are standard, the main contribution lies in the unified treatment of the linear causal estimator, which makes the application of the methodology developed above straightforward and systematic.

5.1. INTERPRETATION OF SUMMARY INDEX ESTIMATES

The following theorem establishes the link between the linear causal estimator $\hat{\boldsymbol{\beta}}_n$ based on a summary index and $\hat{\boldsymbol{\tau}}_n$, the corresponding estimator computed from the original outcome vector. As discussed in Section 3, for linear summary indices – i.e., affine transformations of $\boldsymbol{y}$ – the estimator $\hat{\boldsymbol{\beta}}_n$ can be interpreted as a linear combination of the components of $\hat{\boldsymbol{\tau}}_n$, with weights identical to those used in constructing the index.

THEOREM 5.1 *Let $\hat{\boldsymbol{\beta}}_n$ be a linear causal estimator (Definition 2) of the form*

$$\hat{\boldsymbol{\beta}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{s}_{i,n}\omega(\tilde{x}_i, \nu(P_n)),$$

computed with $\boldsymbol{s}_{i,n}$ – the linear summary index of $\boldsymbol{y}_i$ induced by A_n (Definition 3) . Then

$$\hat{\boldsymbol{\beta}}_n = A_n \hat{\boldsymbol{\tau}}_n,$$

where $\hat{\boldsymbol{\tau}}_n$ is a linear causal estimator computed with the original data $\boldsymbol{y}$:

$$\hat{\boldsymbol{\tau}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i\, \omega(\tilde{x}_i, \nu(P_n)).$$

PROOF *In Appendix A* *Q.E.D.*

The result above is purely algebraic and requires no assumptions on the data-generating process. It hinges on two features: (i) the affine (linear) structure of the summary index and (ii) the linear form of the estimator, together with the fact that the weights $\omega(\cdot)$ have mean zero, i.e., $\frac{1}{n}\sum_{i=1}^{n} \omega(\tilde{x}_i, \nu(P_n)) = 0$. If either of the first two conditions fails, the result does not hold. If only the mean-zero property is violated, however, the constant term b_n from the affine transformation – which normally cancels out, as discussed in Section 3 – reappears in the expression for the linear estimator. This is illustrated in the following corollary, which adapts Theorem 5.1 to the case of least squares with a vector-valued dependent variable.

COROLLARY 5.1 *Let* $\left(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n\right) \in \mathbb{R}^{q\times k}$ *be least squares estimates from the regression:*

$$\boldsymbol{s}_{i,n} = \hat{\boldsymbol{\beta}}_n D_i + \hat{\Lambda}_n x_i + \hat{\varepsilon}_i,$$

where $\boldsymbol{s}_{i,n} = A_n \boldsymbol{y}_i + b_n$*. Then*

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_n' \\ \hat{\Lambda}_n' \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\tau}}_n' \\ \hat{\Gamma}_n' \end{pmatrix} A_n' + \begin{pmatrix} 0_q' \\ b_n' \\ 0_{(k-2)\times q} \end{pmatrix}, \tag{21}$$

where $\left(\hat{\boldsymbol{\tau}}_n, \hat{\Gamma}_n\right) \in \mathbb{R}^{p\times k}$ *are least squares estimates from the regression:*

$$\boldsymbol{y}_i = \hat{\boldsymbol{\tau}}_n D_i + \hat{\Gamma}_n x_i + \hat{\boldsymbol{u}}_i.$$

PROOF *In Appendix A* *Q.E.D.*

To see that the least-squares estimator of the intercept does not satisfy the mean-zero weight condition, note that it can be written as:

$$\hat{\lambda}_{n,1} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{s}_{i,n}\tilde{x}_i' \left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1} e_{2,k} = A_n \frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i\, \tilde{x}_i' \left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1} e_{2,k} +$$

$$+ b_n \frac{1}{n}\sum_{i=1}^n \tilde{x}_i' \left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1} e_{2,k} = A_n \hat{\gamma}_{n,1} + b_n,$$

where the last equality follows from:

$$\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\, e_{2,k} = \frac{1}{n}\sum_{i=1}^n \tilde{x}_i \cdot 1 \Rightarrow e_{2,k} = \left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n \tilde{x}_i \Rightarrow$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^n \tilde{x}_i' \left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1} e_{2,k} = 1 \neq 0.$$

In the following propositions, I demonstrate that both the **SN** and **IC** procedures are special cases of linear summary indices, as defined in Definition 3, with their respective weighting matrices derived below.

PROPOSITION 1 — **SN** is linear summary index:

*(i) If standardization at step 4 of Procedure 1 is not performed then **SN** is a linear summary index induced by:*

$$A^{\mathbf{SN}}_{\ell,n} = \frac{1}{p_\ell}\,\vec{1}'_{p_\ell}\hat{S}^{-1}_{\ell,0}, \tag{22}$$

where $\hat{S}_{\ell,0}$ is a diagonal matrix with control group sample standard deviations of $\boldsymbol{y}_{G_\ell}$ on the diagonal, i.e., the outcomes within domain G_ℓ.

*(ii) If standardization at step 4 of Procedure 1 is performed then **SN** is a linear summary index induced by:*

$$A^{\mathbf{SN}\,+\,ST}_{\ell,n} = \left(\vec{1}'_{p_\ell}\,\hat{S}^{-1}_{\ell,0}\,\hat{\Sigma}_{\ell,0}\,\hat{S}^{-1}_{\ell,0}\,\vec{1}_{p_\ell}\right)^{-\frac{1}{2}}\vec{1}'_{p_\ell}\,\hat{S}^{-1}_{\ell,0}, \tag{23}$$

where $\hat{\Sigma}_{\ell,0}$ is the control group sample variance–covariance matrix of $\boldsymbol{y}_{G_\ell}$.

PROOF *In Appendix A.* *Q.E.D.*

PROPOSITION 2 — **IC** is linear summary index:

*(i) If standardization at step 4 of Procedure 2 is not performed then **IC** is a linear summary index induced by:*

$$A^{\mathbf{IC}}_{\ell,n} = \left(\vec{1}'_{p_\ell}\,\hat{S}_{\ell,0}\,\hat{\Sigma}^{-1}_{\ell}\,\hat{S}_{\ell,0}\,\vec{1}_{p_\ell}\right)^{-1}\vec{1}'_{p_\ell}\,\hat{S}_{\ell,0}\,\hat{\Sigma}^{-1}_{\ell}, \tag{24}$$

where $\hat{\Sigma}^{-1}_{\ell}$ is sample variance-covariance matrix of $\boldsymbol{y}_{G_\ell}$.

*(ii) If standardization at step 4 of Procedure 2 is performed then **IC** is a linear summary index induced by:*

$$A^{\mathbf{IC}\,+\,ST}_{\ell,n} = \left(\vec{1}'_{p_\ell}\,\hat{S}_{\ell,0}\,\hat{\Sigma}^{-1}_{\ell}\,\hat{\Sigma}_{\ell,0}\,\hat{\Sigma}^{-1}_{\ell}\,\hat{S}_{\ell,0}\,\vec{1}_{p_\ell}\right)^{-\frac{1}{2}}\vec{1}'_{p_\ell}\,\hat{S}_{\ell,0}\,\hat{\Sigma}^{-1}_{\ell}, \tag{25}$$

PROOF *In Appendix A.* *Q.E.D.*

Expressions (22) and (23) show that the **SN** summary index satisfies the no sign-reversal criterion, since all entries of its weighting matrix are non-negative. By contrast, as (24) and (25) indicate, the **IC** procedure violates the no sign-reversal criterion because the entries of $\hat{\Sigma}_{\ell}^{-1}$ can be negative. As a result, it is possible for every component of the index to display a negative treatment effect, while the summary index itself reflects a positive effect. The following example illustrates this phenomenon at the population level.

EXAMPLE 2: *Illustration of no sign-reversal violation by the **IC** summary index.*

Let $p = 3, q = 1$, consider the following data-generating process:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \frac{1}{10}\begin{pmatrix} 1 \\ 1 \\ 14 \end{pmatrix} D + \varepsilon, E[\varepsilon] = \mathbf{0}, \mathrm{Var}[\varepsilon] = \begin{pmatrix} 1 & \frac{2}{5} & \frac{12}{5} \\ \frac{2}{5} & 1 & \frac{7}{2} \\ \frac{12}{5} & \frac{7}{2} & \frac{72}{5} \end{pmatrix}, D \sim \mathrm{Bern}\left(\frac{1}{2}\right), D \perp \varepsilon.$$

As derived above,

$$A^{\mathbf{IC}} = \left(\vec{1}_p' \, S_0 \, \Sigma^{-1} \, S_0 \, \vec{1}_p\right)^{-1} \vec{1}_p' \, S_0 \, \Sigma^{-1} \approx (0.73 \;\; 1.17 \;\; -0.24).$$

However,

$$\boldsymbol{\tau} = \left(\frac{1}{10} \frac{1}{10} \frac{14}{10}\right)' > 0, \text{ but } \beta = A^{\mathbf{IC}} \, \boldsymbol{\tau} \approx -0.15 < 0.$$

5.2. LARGE-SAMPLE PROPERTIES

In this section, I derive the large-sample properties of the linear causal estimator. In Section 5.2.1, I establish conditions for consistency. In Section 5.2.2, I introduce the notion of Hadamard differentiability of the functional and show that, under suitable conditions, the linear causal estimand is Hadamard differentiable, which yields asymptotic normality. Finally, I provide conditions under which the sample analogue of the asymptotic variance–covariance matrix is consistent. This, in turn, allows researchers to use the formulas to compute and report correct standard errors for any linear causal estimator and for any linear summary index admitting an asymptotic linear representation (such as **SN** or **IC**).

In this section I provide conditions udner which the linear causal estimator is consistent and asymptotically linear, that is, it admits the represenation (9). I impose the following assumption on the data-generating process.

ASSUMPTION 1: P is such that: (i) $E_P[\|\boldsymbol{y}\|^2] < \infty$; (ii) $\mathrm{Var}_P[\boldsymbol{y}]$ is invertible.

Condition (i) is a standard moment requirement, while condition (ii) rules out linear dependence among the components of $\boldsymbol{y}$, i.e., no component is an exact linear combination of the others. I impose the following regularity conditions on the nuisance functional $\nu(P)$ and the weighting function $\omega(\cdot)$.

ASSUMPTION 2:

(i) $\nu(\cdot)$ is weakly continuous i.e., for any $P \in \boldsymbol{P}$, $\nu(H_n) \to \nu(P)$ whenever $H_n \in \boldsymbol{P}$ weakly converges to P, as $n \to \infty$;

(ii) $\forall \nu, \forall \varepsilon > 0, \forall \eta > 0, \exists\, \delta > 0, N_0$ such that for all $n \geq N_0$:

$$P\left(\sup_{\tilde{\nu} \in B_\delta(\nu)} \frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{y}_i(\omega(\tilde{x}_i, \nu) - \omega(\tilde{x}_i, \tilde{\nu}))\| > \eta \right) < \varepsilon.$$

Part (i) of Assumption 2, requires that the values of ν are close whenever the underlying distributions are close. This is a common condition to ensure that plugging in the empirical distribution yields a consistent estimator. Part (ii) of Assumption 2 is stochastic equicontinuity assumption that requires that the random (in $\tilde{x}$) difference $\|\boldsymbol{y}(\omega(\tilde{x}, \nu) - \omega(\tilde{x}, \tilde{\nu}))\|$ is small in a stochastic sense whenever ν and $\tilde{\nu}$ are close to each other. I impose the following assumption on the linear summary index procedure.

ASSUMPTION 3: *Consistency of weighting in the linear summary index.* Let the linear summary index be induced by A_n such that:

$$\mathrm{plim}_{n \to \infty} A_n = A < \infty,$$

where $A \in \mathbb{R}^{q \times p}$ is a fixed matrix.

Under Assumption 1, Assumption 2, and Assumption 3, the algebraic relation $\hat{\boldsymbol{\beta}}_n = A_n \hat{\boldsymbol{\tau}}_n$ between the summary index estimator $\hat{\boldsymbol{\beta}}_n$ and the full-data linear estimator $\hat{\boldsymbol{\tau}}_n$ extends to the population level. This result is formalized in the following theorem.

THEOREM 5.2 *Let $\hat{\boldsymbol{\beta}}_n$ be the linear estimator:*

$$\hat{\boldsymbol{\beta}}_n := \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{s}_{i,n}\, \omega(\tilde{x}_i, \nu(P_n))$$

computed using $\boldsymbol{s}$ – linear summary index of $\boldsymbol{y}$ induced by A_n, which satisfies Assumption 3 for some A. Then under Assumption 1 and Assumption 2 it holds that:

$$\hat{\boldsymbol{\beta}}_n \xrightarrow{p} A\boldsymbol{\tau}, n \to \infty,$$

where $\boldsymbol{\tau}$ is the linear causal estimand:

$$\boldsymbol{\tau} = E_P[\boldsymbol{y}\, \omega(\tilde{x}, \nu(P))].$$

PROOF *In Appendix A* *Q.E.D.*

5.2.2. ASYMPTOTIC NORMALITY

Next I formally introduce the notion of functional differentiability that is necessary to establish the asymptotic linearity of the linear estimator (representation (9)) that eventually leads to the asymptotic distribution of the summary index estimator $\hat{\boldsymbol{\beta}}_n$ provided in (13). Consider a generic functional of the data distribution $\theta(P)$.

CONVENTION 1: A functional $\theta(\cdot) : \mathbb{D} \mapsto \mathbb{E}$ maps a normed space $\mathbb{D}$ equipped with a (semi-)norm $\|\cdot\|_{\mathbb{D}}$ to a Euclidean space $\mathbb{E}$ equipped with the standard inner products for $\mathbb{R}^n$, and with Frobenius inner product for $\mathbb{R}^{n\times m}$: $\langle A, B\rangle := \mathrm{tr}(A'B)$. The norm induced by $\langle \cdot, \cdot \rangle$ denoted by $\|\cdot\|$.

DEFINITION 4: *Hadamard differentiability*. A mapping $\theta(\cdot) : \mathbb{D} \mapsto \mathbb{E}$ is Hadamard differentiable at $x \in \mathbb{D}$ if there exists a continuous linear functional $\theta'_x(h) : \mathbb{D}_0 \mapsto \mathbb{E}$, with $\mathbb{D}_0 \subseteq \mathbb{D}$, such that:

$$\lim_{t \to 0} \left\| \frac{\theta(x + th_t) - \theta(x)}{t} - \theta'_x(h) \right\| \to 0, h_t \to h, \tag{26}$$

$\forall h_t : x + th_t \in \mathbb{D}$.

This notion of functional differentiability leads to asymptotic linearity of $\theta(P_n)$. Indeed, take the sequences of numbers $t(n) = \frac{1}{\sqrt{n}}$ and the induced by it sequence of functionals $h_{t(n)} = \sqrt{n}(P_n - P) = \sqrt{n}\, \frac{1}{n} \sum_{i=1}^{n} \left(\delta_{W_i} - P\right)$, then apply the delta method (Theorem 20.8 in van der Vaart (2000)) to establish:

$$\frac{\theta\left(P + \frac{1}{\sqrt{n}}\sqrt{n}(P_n - P)\right) - \theta(P)}{\frac{1}{\sqrt{n}}} - \theta'_P\left(\sqrt{n}(P_n - P)\right) = o_p(1) \Rightarrow$$

$$\Rightarrow \theta(P_n) - \theta(P) - \frac{1}{\sqrt{n}}\theta'_P\left(\sqrt{n}(P_n - P)\right) = o_p\left(\frac{1}{\sqrt{n}}\right),$$

finally the linearity of $\theta'_P(\cdot)$ leads to the representation:

$$\theta(P_n) - \theta(P) = \frac{1}{n}\sum_{i=1}^{n} \theta'_P\left(\delta_{W_i} - P\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{27}$$

To obtain the functional form of the influence function $\theta'_P\left(\delta_{W_i} - P\right)$, note that if the functional is Hadamard differentiable, then (26) holds for any sequence $h_t \to h$, and in particular for the constant sequence $h_t = h$. In this case, (26) reduces to

$$\lim_{t \to 0} \left\| \frac{\theta(P + th) - \theta(P)}{t} - \theta'_P(h) \right\| = 0, \tag{28}$$

which corresponds to the weaker notion of functional differentiability known as Gateaux differentiability. As (28) demonstrates, the Gateaux differentiability reduces the functional differentiability to the differentiability of the real-valued function:

$$f(t; P, h) := \theta(P + th)$$

at $t = 0$, with the derivative $f'(0; P, h) = \theta'_P(h)$. Therefore, the influence function can be derived using standard calculus as:

$$\theta'_P\left(\delta_{W_i} - P\right) = \frac{\mathrm{d}}{\mathrm{d}t}\theta\left(P + t\left(\delta_{W_i} - P\right)\right)\bigg|_{t=0}. \tag{29}$$

In practice, one typically computes a candidate influence function using (29), and then establishes the conditions ensuring that the remainder term is sufficiently small for the representation (27) to hold. In case of the linear causal estimator, I make the following assumptions.

ASSUMPTION 4:

(i) $\nu(\cdot)$ is Hadamard-differentiable on its entire domain, and its derivative $\nu'_P(\cdot)$ is defined and continuous on the entire domain, with $E_P[\nu'_P(\delta_W - P)] = 0$ and $E_P\left[\|\nu'_P(\delta_W - P)\|^2\right] < \infty$;

(ii) $\omega(\tilde{x}, \nu)$ is differentiable in the second argument for any $\nu, \tilde{x}$; furthermore for any ν_0, any $\varepsilon > 0$, and any $\eta > 0$ there exist $\delta > 0, N_0 > 0$ such that for all $n > N_0$:

$$P\left(\sup_{\tilde{\nu} \in B_\delta(\nu_0)} \frac{1}{n} \sum_{i=1}^{n} \left\| \boldsymbol{y}_i \, \frac{\omega(\tilde{x}_i, \tilde{\nu}) - \omega(\tilde{x}_i, \nu_0) - \left\langle \frac{\partial \omega(\tilde{x}_i, \nu_0)}{\partial \boldsymbol{\nu}}, \tilde{\nu} - \nu_0 \right\rangle}{\frac{1}{\sqrt{n}}} \right\| > \eta \right) < \varepsilon;$$

(iii) $\forall s, j \in [p]$:

$$E_P\left[y_s \, y_j \, \omega^2(\tilde{x}, \nu(P))\right] < \infty, E_P\left[\left\| y_s \, \frac{\partial \omega(\tilde{x}, \boldsymbol{\nu})}{\partial \nu}\right|_{\nu = \nu(P)} \right\|\right] < \infty.$$

Part (ii) of Assumption 4 is a stochastic equi-differentiability condition on $\boldsymbol{y} \, \omega(\tilde{x}, \nu)$. It ensures that the linear approximation based on the derivative remains sufficiently accurate in probability when averaged over the sample $\{W_i : i = 1, ..., n\}$. Together with the Hadamard differentiability of $\nu(P)$ (part (i)) and the finiteness of certain moments (part (iii)), it implies the Hadamard differentiability of the linear causal estimand $\boldsymbol{\tau}(P)$. Note that condition (ii) in Assumption 4 holds trivially when $\omega(\tilde{x}, \nu)$ is linear in ν, as in the least-squares estimand (Example 1). Under this assumption asymptotic linearity of the linear estimator can be established.

LEMMA 5.1 *Under Assumption 1 and Assumption 4 the linear estimator $\hat{\boldsymbol{\tau}}_n$ defined in (3) admits the following representation:*

$$\hat{\boldsymbol{\tau}}_n = \boldsymbol{\tau} + \frac{1}{n}\sum_{i=1}^{n} \vartheta_P(W_i) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where $\vartheta_P(W) \in \mathbb{R}^p$ is the influence function whose ℓ −th component equals:

$$\vartheta_P^{\ell}(W) = y_\ell\, \omega(\tilde{x}, \nu(P)) - E_P[y_\ell\, \omega(\tilde{x}, \nu(P))] + \\ + \left\langle E_P\left[y_\ell \frac{\partial \omega(\tilde{x}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}}\bigg|_{\boldsymbol{\nu}=\nu(P)}\right], \nu_P'(\delta_W - P)\right\rangle.$$

In addition,

$$E_P[\vartheta_P(W)] = 0, E_P[\vartheta_P(W)\, \vartheta_P'(W)] < \infty.$$

PROOF *In Appendix A* *Q.E.D.*

I now establish a similar asymptotic linearity representation for the non-standardized **SN** summary index procedure – the only one among the commonly used approaches that satisfies both the no sign-reversal and equal-effects invariance criteria discussed above.

THEOREM 5.3 — Asymptotic linearity of **SN**: *Under Assumption 1, the weighting matrix associated with **SN** summary index without final standardization, $A_n^{\mathrm{SN}} \in \mathbb{R}^{q\times p}$ admits the representation:*

$$A_n^{\mathrm{SN}} = A^{\mathrm{SN}} + \frac{1}{n}\sum_{i=1}^{n} \psi_P^{\mathrm{SN}}(W_i) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

with

$$A^{\mathrm{SN}} = \left(\frac{I\{j \in G_\ell\}}{|G_\ell|} \frac{1}{\sqrt{\mathrm{Var}_P\left[y_j | D = 0\right]}} \right)_{\ell \in [q],\, j \in [p]},$$

and the influence functions $\psi_P^{\mathrm{SN}}(W)$ whose components are given by:

$$\psi_P^{\mathrm{SN}}(W_i)_{\ell,j} =$$

$$= -\frac{1}{2}\frac{I\{j \in G_\ell\}}{|G_\ell|}\frac{\frac{1-D_i}{P(D=0)}\Big(\big(y_{j,i} - E_P\big[y_j|D=0\big]\big)^2 - \mathrm{Var}_P\big[y_j|D=0\big]\Big)}{\big(\mathrm{Var}_P\big[y_j|D=0\big]\big)^{\frac{3}{2}}}. \quad (30)$$

The asymptotic linear representations of $\hat{\boldsymbol{\tau}}_n$ (Lemma 5.1) and A_n^{SN} (Theorem 5.3) allow us to derive the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$. This extends straightforwardly to any linear summary index procedure that admits a similar representation. Accordingly, I impose the following assumption.

ASSUMPTION 5: *Asymptotic linearity of weights in the linear summary index.* Let the linear summary index be induced by A_n such that:

$$A_n = A + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\psi}_P(W_i) + o_P\left(\frac{1}{\sqrt{n}}\right), \text{ as } n \to \infty,$$

where $\boldsymbol{\psi}_P(W) : \mathbb{R}^{p+k} \mapsto \mathbb{R}^{q\times p}$ is such that $E_P[\psi_P(W)] = 0, E_P[\psi_P(W)\psi_P'(W)] < \infty$.

Then the asymptotic normality of $\hat{\boldsymbol{\beta}}_n$ follows.

THEOREM 5.4 — Asymptotic normality of $\hat{\boldsymbol{\beta}}_n$: *Let*

(i) $\boldsymbol{\tau}$ *be the linear estimand defined in (1)*

(ii) $\boldsymbol{s}_{i,n}$ *be the linear summary index of* $\boldsymbol{y}_i$ *induced by* A_n*, which satisfies Assumption 5 for some A and the influence function* $\psi_P(\cdot)$*;*

(iii) $\hat{\boldsymbol{\beta}}_n$ *be the linear estimator computed with* $\boldsymbol{s}$*.*

Then under Assumption 1 and Assumption 4, $\hat{\boldsymbol{\beta}}_n$ *is asymptotically normal,*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \varsigma), \quad (31)$$

where $\boldsymbol{\beta} = A\boldsymbol{\tau}$, with the variance-covariance matrix:

$$\varsigma = E_P\left[(A\vartheta_P(W) + \psi_P(W)\boldsymbol{\tau})(A\vartheta_P(W) + \psi_P(W)\boldsymbol{\tau})'\right], \tag{32}$$

where $\vartheta_P(W) \in \mathbb{R}^p$ is the influence function of $\boldsymbol{\tau}$.

Proof *In Appendix A* *Q.E.D.*

As discussed in Section 3, standard statistical software reports standard errors that ignore the variability of the weighting matrix A_n. The theorem below formalizes this point for the least-squares estimator, showing that the t-statistic does not have a standard normal asymptotic distribution unless $\boldsymbol{\tau} = 0$.

Assumption 6: *Linear model*. Let

(i)

$$\boldsymbol{y} = D\boldsymbol{\tau} + \Gamma x + \boldsymbol{u}, E\left[\boldsymbol{u}\begin{pmatrix} D \\ x \end{pmatrix}'\right] = 0, E_P[\tilde{x}\tilde{x}']^{-1} < \infty;$$

(ii) $E[\|\boldsymbol{u}\|^2] < \infty, E_P[\|\tilde{x}\|^2] < \infty, E_P\left[\left\|\tilde{x}u_j\right\|^2\right] < \infty, \forall j = 1, ..., p.$

Theorem 5.5 *Let*

(i) Assumption 6 holds

(ii) $\boldsymbol{s}_{i,n}$ be the linear summary index of $\boldsymbol{y}$ induced by A_n, which satisfies Assumption 5 for some A and influence functions $\psi_P(\cdot)$

(iii) $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_1, ..., \hat{\beta}_q\right)' \in \mathbb{R}^q$ be the least squares estimates from the regressions:

$$s_j = \hat{\beta}_j D + \hat{\lambda}_j' x + \hat{\varepsilon}_j, j = 1, ..., q; \tag{33}$$

(iv) $\widetilde{\text{s.e.}}\left(\hat{\beta}_j\right)$ be the standard error of $\hat{\beta}_j$ reported from the same regressions, e.g., if heteroskedasticity consistent (HC1) option is chosen then

$$\widetilde{\text{s.e.}}\left(\hat{\beta}_j\right) = \frac{1}{\sqrt{n}}\sqrt{e'_{1,k}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\hat{\varepsilon}^2_{i,j}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\right)^{-1}\right)e_{1,k}}.$$

Then the t-statistics have the following marginal asymptotic distributions:

$$\frac{\hat{\beta}_j - \beta_j}{\widetilde{\text{s.e.}}\left(\hat{\beta}_j\right)} \xrightarrow{d} \mathcal{N}\left(0, \sigma_j^2\right), \forall j = 1, ..., q,$$

where

$$\sigma_j^2 = 1 + \frac{E_P\left[\left(e'_{j,q}\psi_P(W)\boldsymbol{\tau}\right)^2\right] + 2E_P\left[\left(e'_{j,q}\psi_P(W)\boldsymbol{\tau}\right)\left(e'_{1,k}E_P[\tilde{x}\tilde{x}']^{-1}\tilde{x}\right)\left(\boldsymbol{u}'A'e_{j,q}\right)\right]}{E\left[\left(e'_{j,q}A\boldsymbol{u}\right)^2\left(\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}e_{1,k}\right)^2\right]}.$$

PROOF *In Appendix A* *Q.E.D.*

To conclude this section, I demonstrate that the asymptotic variance from Theorem 5.4 can be consistently estimated if corresponding influence functions $\psi_P(\cdot)$ and $\vartheta_P(\cdot)$ are estimated sufficiently well. For that, I make the following additional assumptions.

ASSUMPTION 7: There exists an estimator $\hat{\psi}(\cdot)$ that satisfies:

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{\psi}(W_i) - \psi_P(W_i)\right\|^2 \xrightarrow{p} 0.$$

ASSUMPTION 8:

(i)

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\nu_P'\left(\delta_{W_i}-P\right)-\nu_{P_n}'\left(\delta_{W_i}-P_n\right)\right\|^2=o_p(1);$$

(ii) $\forall\eta>0,\forall\varepsilon>0,\forall\nu_0>0,\exists\delta>0,\exists N_1$ such that $\forall n>N_1$:

$$P\left(\sup_{\tilde{\nu}\in B_\delta(\nu_0)}\frac{1}{n}\sum_{i=1}^{n}\left\|y_i^\ell\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}}-y_i^\ell\frac{\partial\omega(\tilde{x}_i,\tilde{\nu})}{\partial\boldsymbol{\nu}}\right\|>\eta\right)<\varepsilon,\forall\,\ell\in[p];$$

(iii) $\forall\ell\in[p]$:

$$E_P\left[\left\|y^\ell\frac{\partial\omega(\tilde{x},\nu(P))}{\partial\boldsymbol{\nu}}\right\|^2\right]<\infty$$

LEMMA 5.2 *Under Assumption 1, Assumption 4, Assumption 5, Assumption 7, Assumption 8, the asymptotic variance - covariance matrix ς defined in (32) can be consistently estimated as follows:*

$$\hat{\varsigma}_n:=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\psi}(W_i)\hat{\boldsymbol{\tau}}_n+A_n\boldsymbol{\vartheta}_{P_n}(W_i)\right)\left(\hat{\psi}(W_i)\hat{\boldsymbol{\tau}}_n+A_n\boldsymbol{\vartheta}_{P_n}(W_i)\right)'\xrightarrow{p}\varsigma.$$

PROOF *In Appendix A* *Q.E.D.*

6. SUMMARY AND CONCLUSION

This paper studies widely used summary index procedures. Because these indices are affine transformations of the outcome vector, any linear estimator computed with a summary index is itself a linear transformation of the estimator based on the original data. I show that the scale-normalized (**SN**) index satisfies both no sign-reversal and equal-effects invariance, while the inverse-covariance (**IC**) and PCA-based indices do not. In practice, this means that researchers using **IC** or PCA indices must guard against sign reversals, and that final standardization of any index should be avoided, as it violates equal-effects invariance.

On inference, I demonstrate that the data dependence of weights implies that conventional software reports inconsistent standard errors and invalid confidence intervals. These problems disappear only when the null hypothesis of no effect holds, so t-tests of the null remain valid across all procedures. However, if confidence intervals or effect sizes are reported, corrections are required.

Claims of power advantages for summary index tests are also overstated: I show that each procedure is highly powered only against a narrow set of alternatives, never known a priori. Finally, I show that common imputation-based strategies for handling missing outcomes fail even under missing-at-random, whereas directly exploiting the linear structure preserves identification.

Taken together, these results suggest that while summary indices are suitable for testing the null of no effect – where the choice of procedure does not matter – only the non-standardized **SN** index satisfies natural interpretability criteria when effect sizes themselves are of interest.

REFERENCES

ALFONSI, L., O. BANDIERA, V. BASSI, R. BURGESS, I. RASUL, M. SULAIMAN, AND A. VITALI. (2020): “Tackling youth unemployment: Evidence from a labor market experiment in Uganda,” *Econometrica,* 88, 2369–2414.

ALLCOTT, H., L. BRAGHIERI, S. EICHMEYER, AND M. GENTZKOW. (2020): “The welfare effects of social media,” *American economic review,* 110, 629–76.

ALLCOTT, H., M. GENTZKOW, AND L. SONG. (2022): “Digital addiction,” *American Economic Review,* 112, 2424–63.

ANDERSON, M. L. (2008): “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects,” *Journal of the American statistical Association,* 103, 1481–95.

ANDERSON, M. L., AND J. MAGRUDER. (2022): Highly Powered Analysis Plans,.

ASHER, S., AND P. NOVOSAD. (2020): “Rural roads and local economic development,” *American economic review,* 110, 797–823.

BANERJEE, A., E. DUFLO, R. GLENNERSTER, AND C. KINNAN. (2015): “The miracle of microfinance? Evidence from a randomized evaluation,” *American economic journal: Applied economics,* 7, 22–53.

BARANOV, V., S. BHALOTRA, P. BIROLI, AND J. MASELKO. (2020): “Maternal depression, women’s empowerment, and parental investment: Evidence from a randomized controlled trial,” *American economic review,* 110, 824–59.

BASELER, T., T. GINN, R. HAKIZA, H. OGUDE-CHAMBERT, AND O. WOLDEMIKAEL. (2025): “Can redistribution change policy views? Aid and attitudes toward refugees,” *Journal of Political Economy,* 133, 0.

BERAJA, M., A. KAO, D. Y. YANG, AND N. YUCHTMAN. (2023): “AI-tocracy,” *The Quarterly Journal of Economics,* 138, 1349–1402.

BHATT, M. P., S. B. HELLER, M. KAPUSTIN, M. BERTRAND, AND C. BLATTMAN. (2024): “Predicting and preventing gun violence: An experimental evaluation of READI Chicago,” *The quarterly journal of economics,* 139, 1–56.

BORUSYAK, K., X. JARAVEL, AND J. SPIESS. (2024): “Revisiting event-study designs: robust and efficient estimation,” *Review of Economic Studies,* 91, 3253–85.

BRAGHIERI, L., R. LEVY, AND A. MAKARIN. (2022): “Social media and mental health,” *American Economic Review,* 112, 3660–93.

BUGNI, F. A., I. A. CANAY, AND S. MCBRIDE. (2023): “Decomposition and interpretation of treatment effects in settings with delayed outcomes,” *arXiv preprint arXiv:2302.11505,*.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott. (2020): "Misperceived social norms: Women working outside the home in Saudi Arabia," *American economic review,* 110, 2997–3029.

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

Cantoni, D., Y. Chen, D. Y. Yang, N. Yuchtman, and Y. J. Zhang. (2017): "Curriculum and ideology," *Journal of political economy,* 125, 338–92.

Cantoni, D., D. Y. Yang, N. Yuchtman, and Y. J. Zhang. (2019): "Protests as strategic games: experimental evidence from Hong Kong's antiauthoritarian movement," *The Quarterly Journal of Economics,* 134, 1021–77.

Carranza, E., R. Garlick, K. Orkin, and N. Rankin. (2022): "Job search and hiring with limited information about workseekers' skills," *American Economic Review,* 112, 3547–83.

Casey, K., R. Glennerster, and E. Miguel. (2012): "Reshaping institutions: Evidence on aid impacts using a preanalysis plan," *The Quarterly Journal of Economics,* 127, 1755–1812.

Chaisemartin, C. De, and X. d'Haultfoeuille. (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American economic review,* 110, 2964–96.

Chen, Y., and D. Y. Yang. (2019): "The impact of media censorship: 1984 or brave new world?," *American Economic Review,* 109, 2294–2332.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. (2011): "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly journal of economics,* 126, 1593–1660.

Christensen, D., O. Dube, J. Haushofer, B. Siddiqi, and M. Voors. (2021): "Building resilient health systems: experimental evidence from Sierra Leone and the 2014 Ebola outbreak," *The Quarterly Journal of Economics,* 136, 1145–98.

Currie, J., L. Davis, M. Greenstone, and R. Walker. (2015): "Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings," *American Economic Review,* 105, 678–709.

Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. Walker. (2022): "General equilibrium effects of cash transfers: experimental evidence from Kenya," *Econometrica,* 90, 2603–43.

Filmer, D., and L. H. Pritchett. (2001): "Estimating wealth effects without expenditure data —or tears: an application to educational enrollments in states of India," *Demography,* 38, 115–32.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker, and t. Oregon Health Study Group. (2012): "The Oregon health insurance experiment: evidence from the first year," *The Quarterly journal of economics,* 127, 1057–1106.

GOLDSMITH-PINKHAM, P., P. HULL, AND M. KOLESÁR. (2024): "Contamination bias in linear regressions," *American Economic Review,* 114, 4015–51.

GOODMAN-BACON, A. (2021): "Difference-in-differences with variation in treatment timing," *Journal of econometrics,* 225, 254–77.

HAUSHOFER, J., AND J. SHAPIRO. (2016): "The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya," *The Quarterly Journal of Economics,* 131, 1973–2042.

HOTELLING, H. (1931): "The generalization of Student's ratio,".

HOYNES, H., D. W. SCHANZENBACH, AND D. ALMOND. (2016): "Long-run impacts of childhood access to the safety net," *American Economic Review,* 106, 903–34.

KINNAN, C., K. SAMPHANTHARAK, R. TOWNSEND, AND D. VERA-COSSIO. (2024): "Propagation and insurance in village networks," *American Economic Review,* 114, 252–84.

KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ. (2007): "Experimental analysis of neighborhood effects," *Econometrica,* 75, 83–119.

LEVY, R. (2021): "Social media, news consumption, and polarization: Evidence from a field experiment," *American economic review,* 111, 831–70.

LIST, J. A., A. M. SHAIKH, AND Y. XU. (2019): "Multiple hypothesis testing in experimental economics," *Experimental Economics,* 22, 773–93.

O'BRIEN, P. C. (1984): "Procedures for comparing samples with multiple endpoints," *Biometrics,* , 1079–87.

ROMANO, J. P., AND M. WOLF. (2005): "Exact and approximate stepdown methods for multiple hypothesis testing," *Journal of the American Statistical Association,* 100, 94–108.

STANTCHEVA, S. (2021): "Understanding tax policy: How do people reason?," *The Quarterly Journal of Economics,* 136, 2309–69.

SUN, L., AND S. ABRAHAM. (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of econometrics,* 225, 175–99.

TSIATIS, A. A. (2006): *Semiparametric Theory and Missing Data,* Springer.

VAART, A. W. VAN DER. (2000): *Asymptotic Statistics,* Cambridge university press.

VIVIANO, D., K. WUTHRICH, AND P. NIEHAUS. (2021): "A model of multiple hypothesis testing," *arXiv preprint arXiv:2104.13367,*.

WESTFALL, P. H., AND S. S. YOUNG. (1993): *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment,* John Wiley & Sons.

PROPOSITION 3 ***IC** is a linear index*

PROOF *Suppose that on Step 2 the control group sample mean is subtracted. By definition* $\tilde{\boldsymbol{y}}_i = \hat{\boldsymbol{S}}_0^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_0) \Rightarrow \bar{\tilde{\boldsymbol{y}}} = \hat{\boldsymbol{S}}_0^{-1}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{y}}_0)$

$$\tilde{\boldsymbol{V}}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\boldsymbol{y}}_i - \bar{\tilde{\boldsymbol{y}}}\right)\left(\tilde{\boldsymbol{y}}_i - \bar{\tilde{\boldsymbol{y}}}\right)' = \hat{\boldsymbol{S}}_0^{-1}\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})'\hat{\boldsymbol{S}}_0^{-1} = \hat{\boldsymbol{S}}_0^{-1}\hat{\boldsymbol{\Sigma}}_n\hat{\boldsymbol{S}}_0^{-1},$$

$$s_i := \left(\vec{1}_p'\tilde{\boldsymbol{V}}_n^{-1}\vec{1}_p\right)^{-1}\vec{1}_p'\tilde{\boldsymbol{V}}_n^{-1}\tilde{\boldsymbol{y}}_i = \left(\vec{1}_p'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\Sigma}}_n^{-1}\hat{\boldsymbol{S}}_0\vec{1}_p\right)^{-1}\vec{1}_p'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\Sigma}}_n^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_0) = A\boldsymbol{y}_i + b,$$

where

$$A = \left(\vec{1}_p'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\Sigma}}_n^{-1}\hat{\boldsymbol{S}}_0\vec{1}_p\right)^{-1}\vec{1}_p'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\Sigma}}_n^{-1},$$

$$b = -\left(\vec{1}_p'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\Sigma}}_n^{-1}\hat{\boldsymbol{S}}_0\vec{1}_p\right)^{-1}\vec{1}_p'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\Sigma}}_n^{-1}\bar{\boldsymbol{y}}_0.$$

If the full sample mean is subtracted on Step 2 then

$$\tilde{\boldsymbol{y}}_i = \hat{\boldsymbol{S}}_0^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}) \Rightarrow \bar{\tilde{\boldsymbol{y}}} = 0,$$

$$\Rightarrow \tilde{\boldsymbol{V}}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{\boldsymbol{y}}_i - \bar{\tilde{\boldsymbol{y}}}\right)\left(\tilde{\boldsymbol{y}}_i - \bar{\tilde{\boldsymbol{y}}}\right)' = \hat{\boldsymbol{S}}_0^{-1}\hat{\boldsymbol{\Sigma}}_n\hat{\boldsymbol{S}}_0^{-1} \Rightarrow$$

$$\Rightarrow s_i := \left(\vec{1}_p'\tilde{\boldsymbol{V}}_n^{-1}\vec{1}_p\right)^{-1}\vec{1}_p'\tilde{\boldsymbol{V}}_n^{-1}\tilde{\boldsymbol{y}}_i = \left(\vec{1}_p'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\Sigma}}_n^{-1}\hat{\boldsymbol{S}}_0\vec{1}_p\right)^{-1}\vec{1}_p'\hat{\boldsymbol{S}}_0\hat{\boldsymbol{\Sigma}}_n^{-1}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}) =$$

$$= A\boldsymbol{y}_i + b$$

Q.E.D.

PROOF *of Theorem 5.1:*

$$\hat{\beta}_n := \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{s}_i\omega(\tilde{x}_i, \nu(P_n)) = \frac{1}{n}\sum_{i=1}^{n}(A_n\boldsymbol{y}_i + b_n)\omega(\tilde{x}_i, \nu(P_n)) =$$

$$= A_n\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\omega(\tilde{x}_i, \nu(P_n)) + b_n\frac{1}{n}\sum_{i=1}^{n}\omega(\tilde{x}_i, \nu(P_n)) = A_n\hat{\boldsymbol{\tau}}_n.$$

The last inequality follows from $E_P[\omega(\tilde{x}, \nu(P))] = 0$ *for any* P*, including empirical distribution* P_n*.* *Q.E.D.*

PROOF *of Corollary 5.1:*

Denote $\tilde{x}_i = \binom{D_i}{x_i} \in \mathbb{R}^k$*, then*

$$\begin{pmatrix}\hat{\beta}' \\ \hat{\Lambda}'\end{pmatrix} = \left(\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1}\left(\sum_{i=1}^n \tilde{x}_i\boldsymbol{s}_i'\right) =$$

$$= \left(\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1}\left(\sum_{i=1}^n \tilde{x}_i\boldsymbol{y}_i'\right)A_n' + \left(\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1}\left(\sum_{i=1}^n \tilde{x}_i b_n'\right) =$$

$$= \begin{pmatrix}\hat{\boldsymbol{\tau}}' \\ \hat{\Gamma}'\end{pmatrix}A_n' + \begin{pmatrix}\mathbf{0}_q' \\ b_n' \\ \mathbf{0}_{(k-2)\times q}\end{pmatrix},$$

where the last equality can be obtained by noticing that since the second component of $\tilde{\boldsymbol{x}}_i$ *is 1, then*

$$\tilde{x}_i'\begin{pmatrix}\mathbf{0}_q' \\ b_n' \\ \mathbf{0}_{(k-2)\times q}\end{pmatrix} = b_n' \Rightarrow \left(\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)\begin{pmatrix}\mathbf{0}_q' \\ b_n' \\ \mathbf{0}_{(k-2)\times q}\end{pmatrix} = \sum_{i=1}^n \tilde{x}_i' b_n.$$

One can use the result of Theorem 5.1 directly: let $\boldsymbol{\Omega}_{i,n} := \tilde{x}_i'\left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1}$*, then notice that*

$$\begin{pmatrix}\hat{\boldsymbol{\tau}} \\ \hat{\Gamma}\end{pmatrix} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i\boldsymbol{\Omega}_{i,n},$$

and further,

$$e_{2,k}'\left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right) = \frac{1}{n}\sum_{i=1}^n (e_{2,k}'\tilde{x}_i)\tilde{x}_i' = \frac{1}{n}\sum_{i=1}^n \tilde{x}_i' \Rightarrow \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Omega}_{i,n} = e_{2,k}',$$

and so the result follows from Theorem 5.1. *Q.E.D.*

PROOF *of Theorem 5.2: Denote* $\nu(P) = \nu_0$*, then*

$$\|\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}\| = \left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i\omega(\tilde{x}_i, \hat{\nu}_n) - E_P[\boldsymbol{y}\omega(\tilde{x}, \nu_0)]\right\| = \left\|\frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i[\omega(\tilde{x}_i, \hat{\nu}_n) - \omega(\tilde{x}_i, \nu_0)] + \right.$$

$$+\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\omega(\tilde{x}_i,\nu_0)-E_P[\boldsymbol{y}\omega(\tilde{x},\nu_0)]\Bigg\| \leq \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)]\|+$$

$$+\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\omega(\tilde{x}_i,\nu_0)-E_P[\boldsymbol{y}\omega(\tilde{x},\nu_0)]\right\| \leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}|\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)|^2}+$$

$$+\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\omega(\tilde{x}_i,\nu_0)-E_P[\boldsymbol{y}\omega(\tilde{x},\nu_0)]\right\|.$$

By weak law of large numbers,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i\,\omega(\tilde{x}_i,\nu_0)-E_P[\boldsymbol{y}\,\omega(\tilde{x},\nu_0)]\right\| \xrightarrow{p} 0,$$

$$\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i\|^2 \xrightarrow{p} E_P[\|\boldsymbol{y}\|^2]<\infty \text{ by Assumption 1;}$$

Next I show that $\forall\,\varepsilon>0,\exists\,N$ *such that for all* $n>N$*:*

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)]\|>\varepsilon\right)<\varepsilon,$$

and so

$$\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)]\|=o_p(1).$$

By Part (ii) of Assumption 2 for any $\varepsilon>0$ *there exist* $\delta=\delta(\varepsilon,\nu_0)>0, N_0=N_0(\varepsilon,\nu_0)>0$ *such that for all* $n>N_0(\varepsilon,\nu_0)$*:*

$$P\left(\sup_{\tilde{\nu}\,\in\,B_\delta(\nu_0)}\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\nu_0)-\omega(\tilde{x}_i,\tilde{\nu})]\|>\varepsilon\right)<\frac{\varepsilon}{2}.$$

Further,

$$\left\{\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)]\|>\varepsilon\right\}\subseteq\{\|\hat{\nu}_n-\nu_0\|>\delta\}\bigcup$$

$$\bigcup\left\{\sup_{\tilde{\nu}\,\in\,B_\delta(\nu_0)}\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\tilde{\nu})-\omega(\tilde{x}_i,\nu_0)]\|>\varepsilon\right\}\Rightarrow$$

$$\Rightarrow P\left(\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)]\|>\varepsilon\right)\leq P(\|\hat{\nu}_n-\nu_0\|>\delta)+$$

$$+P\left(\sup_{\tilde{\nu}\,\in\, B_\delta(\nu_0)}\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)]\|>\varepsilon\right).$$

By part (i) of Assumption 2, $\|\hat{\nu}_n-\nu_0\|\xrightarrow{p}0$, *let* $N_1=N_1(\varepsilon,\nu_0)$ *be such that for all* $n\geq N_1$*:*

$$P(\|\hat{\nu}_n-\nu_0\|>\delta)<\frac{\varepsilon}{2},$$

hence for all $n>N=N(\varepsilon,\nu_0):=\max\{N_0,N_1\}$*:*

$$P\left(\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{y}_i[\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)]\|>\varepsilon\right)<\varepsilon.$$

Q.E.D.

PROOF *of Theorem 5.3: Consider an element* (ℓ,j) *of matrix* A_n *that is non zero. Denote* $\theta_1=E\left[y_j|D=0\right],\theta_2=E\left[y_j^2|D=0\right]$*, and their sample analogs as* $\hat{\theta}_{1,n},\hat{\theta}_{2,n}$*. Let* $f(x_1,x_2):=\frac{1}{|\mathcal{G}_\ell|}\frac{1}{\sqrt{x_2-x_1^2}}$*, by the mean-value theorem:*

$$a_{n,\ell,j}=\frac{1}{|\mathcal{G}_\ell|}\frac{1}{\sqrt{\hat{\theta}_{2,n}-\hat{\theta}_{1,n}^2}}=\frac{1}{|\mathcal{G}_\ell|}\frac{1}{\sqrt{\theta_2-\theta_1^2}}-$$

$$-\frac{1}{2}\frac{1}{|\mathcal{G}_\ell|}\frac{1}{\left(\theta_2-\theta_1^2\right)^{\frac{3}{2}}}\left(\hat{\theta}_{1,n}-\theta_1\ \ \hat{\theta}_{2,n}-\theta_2\right)\begin{pmatrix}-2\theta_1\\1\end{pmatrix}+$$

$$+\frac{1}{2}\left(\hat{\theta}_{1,n}-\theta_1\ \ \hat{\theta}_{2,n}-\theta_2\right)\nabla^2 f\left(\xi_{1,n},\xi_{2,n}\right)\begin{pmatrix}\hat{\theta}_{1,n}-\theta_1\\\hat{\theta}_{2,n}-\theta_2\end{pmatrix},$$

where $\left(\xi_{1,n},\xi_{2,n}\right)\in\left\{(1-\lambda)\left(\hat{\theta}_{1,n},\hat{\theta}_{2,n}\right)+\lambda(\theta_1,\theta_2):\lambda\in[0,1]\right\}$*. Rewrite the second term in the expansion:*

$$-\frac{1}{2}\frac{1}{|\mathcal{G}_\ell|}\frac{1}{\left(\theta_2-\theta_1^2\right)^{\frac{3}{2}}}\left(\hat{\theta}_{2,n}-\theta_2-2\theta_1\left(\hat{\theta}_{1,n}-\theta_1\right)\right)=$$

$$= -\frac{1}{2}\frac{1}{|\mathcal{G}_\ell|}\frac{1}{\left(\mathrm{Var}\left[y_j|D=0\right]\right)^{\frac{3}{2}}}\frac{1}{n}\sum_{i=1}^{n}\frac{1-D_i}{P(D=0)}\left(\left(y_{j,i}-\theta_1\right)^2-\left(\theta_2-\theta_1^2\right)\right)+$$

$$+\left(2\theta_1\left(\theta_1-\hat{\theta}_{1,n}\right)+\hat{\theta}_{2,n}-\theta_2\right)\left(1-\frac{1}{n}\sum_{i=1}^{n}\frac{1-D_i}{P(D=0)}\right)=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\psi(W_i)_{\ell,j}+\left(2\theta_1\left(\theta_1-\hat{\theta}_{1,n}\right)+\hat{\theta}_{2,n}-\theta_2\right)\left(1-\frac{1}{n}\sum_{i=1}^{n}\frac{1-D_i}{P(D=0)}\right).$$

So,

$$\sqrt{n}\left(a_{n,\ell,j}-a_{\ell,j}\right)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}(W_i)_{\ell,j}=$$

$$=\left(2\theta_1\left(\theta_1-\hat{\theta}_{1,n}\right)+\hat{\theta}_{2,n}-\theta_2\right)\sqrt{n}\left(1-\frac{1}{n}\sum_{i=1}^{n}\frac{1-D_i}{P(D=0)}\right)+$$

$$+\frac{1}{2}\left(\hat{\theta}_{1,n}-\theta_1\ \ \hat{\theta}_{2,n}-\theta_2\right)\nabla^2 f\left(\xi_{1,n},\xi_{2,n}\right)\sqrt{n}\begin{pmatrix}\hat{\theta}_{1,n}-\theta_1\\ \hat{\theta}_{2,n}-\theta_2\end{pmatrix},$$

where $\left(\xi_{1,n},\xi_{2,n}\right)\in\left\{(1-\lambda)\left(\hat{\theta}_{1,n},\hat{\theta}_{2,n}\right)+\lambda(\theta_1,\theta_2):\lambda\in[0,1]\right\}$*. By the weak law of large numbers* $\left(\hat{\theta}_{1,n},\hat{\theta}_{2,n}\right)\xrightarrow{p}(\theta_1,\theta_2)$*, hence* $\left(2\theta_1\left(\theta_1-\hat{\theta}_{1,n}\right)+\hat{\theta}_{2,n}-\theta_2\right)=o_p(1)$*; further* $\left(\xi_{1,n},\xi_{2,n}\right)\xrightarrow{p}(\theta_1,\theta_2)$*, and since* $\nabla^2 f(x_1,x_2)$ *is continuous on its domain, continuous mapping theorem implies that* $\nabla^2 f\left(\xi_{1,n},\xi_{2,n}\right)=o_p(1)$*. By Lindberg-Levy central limit theorem,* $\sqrt{n}\left(1-\frac{1}{n}\sum_{i=1}^{n}\frac{1-D_i}{P(D=0)}\right)=O_p(1)$*, and* $\sqrt{n}\begin{pmatrix}\hat{\theta}_{1,n}-\theta_1\\ \hat{\theta}_{2,n}-\theta_2\end{pmatrix}=O_p(1)$*, hence-*

$$\sqrt{n}\left(a_{n,\ell,j}-a_{\ell,j}\right)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}(\boldsymbol{y}_i,\tilde{\boldsymbol{x}}_i)_{\ell,j}=o_p(1).$$

Q.E.D.

Proof *of Lemma 5.1: In what follows I denote* $\nu(P)=\nu_0$*,* $\frac{\partial\omega(\tilde{x},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}}\Big|_{\boldsymbol{\nu}=\nu_0}=\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}$*, and* $\nu(P_n)=\hat{\nu}_n$*. Consider* $\ell-th$ *component of* $\hat{\boldsymbol{\tau}}_n$*:*

$$\hat{\boldsymbol{\tau}}_n^\ell=\frac{1}{n}\sum_{i=1}^{n}y_i^\ell\omega(\tilde{x}_i,\hat{\nu}_n)=\boldsymbol{\tau}^\ell+\frac{1}{n}\sum_{i=1}^{n}y_i^\ell(\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0))+$$

$$+\frac{1}{n}\sum_{i=1}^{n}\big(y_i^\ell \omega(\tilde{x}_i,\nu_0)-E_P\big[y^\ell\omega(\tilde{x},\nu_0)\big]\big),$$

$$\sqrt{n}\big(\hat{\boldsymbol{\tau}}_n^\ell-\boldsymbol{\tau}^\ell\big)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}y_i^\ell(\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0))+$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(y_i^\ell \omega(\tilde{x}_i,\nu_0)-E_P\big[y^\ell\omega(\tilde{x},\nu_0)\big]\big)$$

Consider the first term:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}y_i^\ell(\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0))=$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}y_i^\ell\left(\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle\right)+$$

$$+\left\langle\frac{1}{n}\sum_{i=1}^{n}y_i^\ell\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\sqrt{n}(\hat{\nu}_n-\nu_0)-\nu_P'\big(\sqrt{n}(P_n-P)\big)\right\rangle+$$

$$+\left\langle\frac{1}{n}\sum_{i=1}^{n}\left(y_i^\ell\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}}-E\left[y^\ell\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right]\right),\nu_P'\big(\sqrt{n}(P_n-P)\big)\right\rangle+$$

$$+\left\langle E\left[y^\ell\frac{\partial\omega(\tilde{x},\nu(P))}{\partial\boldsymbol{\nu}}\right],\nu_P'\big(\sqrt{n}(P_n-P)\big)\right\rangle.$$

Next I consider the first three terms in the expression above and show that they all are of order $o_p(1)$ as $n\to\infty$.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}y_i^\ell\left(\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle\right)=$$

$$=\frac{1}{n}\sum_{i=1}^{n}y_i^\ell\left(\frac{\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}\right)$$

1. *By part (ii) of Assumption 4, for any $\varepsilon>0$ there exist $\delta=\delta(\varepsilon,\nu_0)>0, N_0=N_0(\varepsilon,\nu_0)>0$ such that for all $n>N_0$:*

$$P\left(\sup_{\tilde{\nu}\,\in\,B_\delta(\nu_0)}\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{y}_i\frac{\omega(\tilde{x}_i,\tilde{\nu})-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\tilde{\nu}-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}\right\|>\varepsilon\right)<\frac{\varepsilon}{2}.$$

   By the property of the L_2 norm:

$$\left\{ \frac{1}{n}\sum_{i=1}^{n} \left| y_i^\ell \left(\frac{\omega(\tilde{x}_i, \hat{\nu}_n) - \omega(\tilde{x}_i, \nu_0) - \left\langle \frac{\partial \omega(\tilde{x}_i,\nu_0)}{\partial \boldsymbol{\nu}}, \hat{\nu}_n - \nu_0 \right\rangle}{\frac{1}{\sqrt{n}}} \right) \right| > \varepsilon \right\} \subseteq$$

$$\subseteq \left\{ \frac{1}{n}\sum_{i=1}^{n} \left\| \boldsymbol{y}_i \frac{\omega(\tilde{x}_i, \hat{\nu}_n) - \omega(\tilde{x}_i, \nu_0) - \left\langle \frac{\partial \omega(\tilde{x}_i,\nu_0)}{\partial \boldsymbol{\nu}}, \hat{\nu}_n - \nu_0 \right\rangle}{\frac{1}{\sqrt{n}}} \right\| > \varepsilon \right\} \subseteq$$

$$\subseteq \{\|\hat{\nu}_n - \nu_0\| > \delta\} \bigcup \left\{ \sup_{\tilde{\nu} \in B_\delta(\nu_0)} \frac{1}{n}\sum_{i=1}^{n} \left\| \boldsymbol{y}_i \frac{\omega(\tilde{x}_i, \tilde{\nu}) - \omega(\tilde{x}_i, \nu_0) - \left\langle \frac{\partial \omega(\tilde{x}_i,\nu_0)}{\partial \boldsymbol{\nu}}, \tilde{\nu} - \nu_0 \right\rangle}{\frac{1}{\sqrt{n}}} \right\| > \varepsilon \right\}.$$

Since there exists some $N_1 = N_1(\varepsilon, \nu_0)$ such that for all $n > N_1$, $P(\|\hat{\nu}_n - \nu_0\| > \delta) < \frac{\varepsilon}{2}$ which follows from continuity of $\nu(\cdot)$ (which in turn follows from Part (i) of Assumption 4) and weak convergence of $P_n \Rightarrow P$, then for all $n > \max\{N_1, N_0\}$:

$$P\left(\frac{1}{n}\sum_{i=1}^{n} \left| y_i^\ell \left(\frac{\omega(\tilde{x}_i, \hat{\nu}_n) - \omega(\tilde{x}_i, \nu_0) - \left\langle \frac{\partial \omega(\tilde{x}_i,\nu_0)}{\partial \boldsymbol{\nu}}, \hat{\nu}_n - \nu_0 \right\rangle}{\frac{1}{\sqrt{n}}} \right) \right| > \varepsilon \right) \leq \varepsilon$$

2. *By the Cauchy-Schwartz inequality:*

$$\left| \left\langle \frac{1}{n}\sum_{i=1}^{n} y_i^\ell \frac{\partial \omega(\tilde{x}_i, \nu_0)}{\partial \boldsymbol{\nu}}, \sqrt{n}(\hat{\nu}_n - \nu_0) - \nu'_P\big(\sqrt{n}(P_n - P)\big) \right\rangle \right| \leq$$

$$\leq \left\| \frac{1}{n}\sum_{i=1}^{n} y_i^\ell \frac{\partial \omega(\tilde{x}_i, \nu_0)}{\partial \boldsymbol{\nu}} \right\| \left\| \frac{\nu\Big(P + \frac{1}{\sqrt{n}}\sqrt{n}(P_n - P)\Big) - \nu(P)}{\frac{1}{\sqrt{n}}} - \nu'_P\big(\sqrt{n}(P_n - P)\big) \right\| = o_p(1),$$

since

$$\left\| \frac{\nu\Big(P + \frac{1}{\sqrt{n}}\sqrt{n}(P_n - P)\Big) - \nu(P)}{\frac{1}{\sqrt{n}}} - \nu'_P\big(\sqrt{n}(P_n - P)\big) \right\| = o_p(1)$$

which follows from the second assertion of Theorem 20.8 in van der Vaart (2000) if $\nu'_P(\cdot)$ exists and continuous on its entire domain which is assumed by Part (i) of Assumption 4, and $\left\| \frac{1}{n}\sum_{i=1}^{n} y_i^\ell \frac{\partial \omega(\tilde{x}_i,\nu_0)}{\partial \boldsymbol{\nu}} \right\| \leq \frac{1}{n}\sum_{i=1}^{n} \left\| y_i^\ell \frac{\partial \omega(\tilde{x}_i,\nu_0)}{\partial \boldsymbol{\nu}} \right\| \xrightarrow{p} E_P\left[\left\| y^\ell \frac{\partial \omega(\tilde{x},\nu_0)}{\partial \boldsymbol{\nu}} \right\| \right] < \infty$ by part (iii) of Assumption 4.

3. *By Cauchy-Schwartz:*

$$\left|\left\langle \frac{1}{n}\sum_{i=1}^{n}\left(y_i^{\ell}\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}} - E_P\left[y^{\ell}\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right]\right), \nu'_P\big(\sqrt{n}(P_n - P)\big)\right\rangle\right| \leq$$

$$\leq \left\|\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{\ell}\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}} - E_P\left[y^{\ell}\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right]\right)\right\| \left\|\nu'_P\big(\sqrt{n}(P_n - P)\big)\right\| = o_p(1),$$

since by the weak law of large numbers $\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{\ell}\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}} - E_P\left[y^{\ell}\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right]\right) = o_p(1)$*, where* $E_P\left[y^{\ell}\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right] < \infty$ *by part (iii) of Assumption 4; further by continuity of* $\nu'_P(\cdot)$ *and weak convergence* $\sqrt{n}(P_n - P) = O_P(1)$*, the continuous mapping theorem implies* $\left\|\nu'_P\big(\sqrt{n}(P_n - P)\big)\right\| = O_p(1)$.

Thus,

$$\sqrt{n}\big(\hat{\boldsymbol{\tau}}_n^{\ell} - \boldsymbol{\tau}^{\ell}(P)\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} y_i^{\ell}(\omega(\tilde{x}_i,\hat{\nu}_n) - \omega(\tilde{x}_i,\nu(P))) +$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(y_i^{\ell}\omega(\tilde{x}_i,\nu(P)) - E_P\big[y^{\ell}\omega(\tilde{x},\nu(P))\big]\big) =$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg(y_i^{\ell}\omega(\tilde{x}_i,\nu(P)) - E_P\big[y^{\ell}\omega(\tilde{x},\nu(P))\big] +$$

$$+\left\langle E_P\left[y^{\ell}\frac{\partial\omega(\tilde{x},\nu(P))}{\partial\boldsymbol{\nu}}\right], \nu'_P\big(\delta_{\tilde{x}_i} - P\big)\right\rangle\Bigg) + o_p(1).$$

Additionally, $E_P[\nu'_P(\delta_W - P)] = 0 \Rightarrow E_P[\vartheta_P(W)] = 0$*, and* $E_P\big[y_\ell y_j \omega^2(\tilde{x},\nu(P))\big] < \infty$ *and* $E_P\left[\|\nu'_P(\delta_W - P)\|^2\right] < \infty$ *further imply that* $E[\vartheta_P(W)\vartheta'_P(W)] < \infty$. *Q.E.D.*

PROOF *of Theorem 5.4: From Lemma 5.1 it follows that*

$$\sqrt{n}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\vartheta}_P(W_i) + o_p(1);$$

since A_n *is asymptotically linear,*

$$\sqrt{n}(A_n - A) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_P(W_i) + o_p(1);$$

By Theorem 5.1, weak law of large numbers, and Lindberg-Levy central limit theorem:

$$\hat{\boldsymbol{\beta}}_n = A_n\hat{\boldsymbol{\tau}}_n \Rightarrow$$

$$\Rightarrow \sqrt{n}\Big(\hat{\boldsymbol{\beta}}_n - A\boldsymbol{\tau}\Big) = \sqrt{n}(A_n - A)(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}) + \sqrt{n}(A_n - A)\boldsymbol{\tau} + A\sqrt{n}(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}) =$$

$$= O_p(1)o_p(1) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_P(W_i)\boldsymbol{\tau} + A\boldsymbol{\vartheta}_P(W_i)) + o_p(1) =$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_P(W_i)\boldsymbol{\tau} + A\boldsymbol{\vartheta}_P(W_i)) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \varsigma).$$

Q.E.D.

Proof *of Lemma 5.2: First, I will show that*

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\vartheta}_{P_n}(W_i)\boldsymbol{\vartheta}'_{P_n}(W_i) \xrightarrow{p} E_P[\boldsymbol{\vartheta}_P(W)\boldsymbol{\vartheta}'_P(W)].$$

Second, I will show that

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\hat{\psi}(W_i) \otimes \hat{\psi}(W_i)\Big) \xrightarrow{p} E_P[\psi(W) \otimes \psi(W)],$$

from that it will follow:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\hat{\psi}(W_i) \otimes \hat{\psi}(W_i)\Big)\operatorname{vec}(\hat{\boldsymbol{\tau}}_n\hat{\boldsymbol{\tau}}'_n) \xrightarrow{p} E_P[\psi_P(W_i) \otimes \psi_P(W_i)]\operatorname{vec}(\boldsymbol{\tau}\boldsymbol{\tau}') \Rightarrow$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\operatorname{vec}\Big(\hat{\psi}(W_i)\hat{\boldsymbol{\tau}}_n\hat{\boldsymbol{\tau}}'_n\hat{\psi}'(W_i)\Big) \xrightarrow{p} E_P[\operatorname{vec}(\psi_P(W)\boldsymbol{\tau}\boldsymbol{\tau}'\psi'_P(W))] \Rightarrow$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\Big(\hat{\psi}(W_i)\hat{\boldsymbol{\tau}}_n\hat{\boldsymbol{\tau}}'_n\hat{\psi}'(W)\Big) \xrightarrow{p} E_P[\psi_P(W)\boldsymbol{\tau}\boldsymbol{\tau}'\psi'_P(W)].$$

Finally I will show that

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\vartheta_{P_n}(W_i) \otimes \hat{\psi}(W_i)\Big) \xrightarrow{p} E_P[\vartheta_P(W) \otimes \psi_P(W)],$$

which by similar logic as above will imply:

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\vartheta_{P_n}(W_i) \otimes \hat{\psi}(W_i)\Big) \xrightarrow{p} E_P[\vartheta_P(W) \otimes \psi_P(W)] \Rightarrow$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}(W_i)\hat{\boldsymbol{\tau}}_n\vartheta'_{P_n}(W_i)A'_n \xrightarrow{p} E_P[\psi_P(W)\boldsymbol{\tau}\vartheta'_P(W)A'].$$

1. To show:

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\vartheta}_{P_n}(W_i)\boldsymbol{\vartheta}'_{P_n}(W_i)\xrightarrow{p}E_P[\boldsymbol{\vartheta}_P(W)\boldsymbol{\vartheta}'_P(W)].$$

I need:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\vartheta}_{P_n}(W_i)\boldsymbol{\vartheta}'_{P_n}(W_i)-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\vartheta}_P(W_i)\boldsymbol{\vartheta}'_P(W_i)\right\|\xrightarrow{p}0,$$

which will imply the result when combined with the weak law of large numbers and the triangular inequality. For the ease of notation let $\nu_0 := \nu(P), \hat{\nu}_n := \nu(P_n)$*,* $\vartheta_i :=$ $\boldsymbol{\vartheta}_P(W_i)$*,* $\hat{\vartheta}_i := \boldsymbol{\vartheta}_{P_n}(W_i)$*. For* $\ell = 1, ..., p$ *define the collection of functionals:*

$$\chi_\ell(P)=E_P\left[y_\ell\frac{\partial\omega(\tilde{x},\nu(P))}{\partial\boldsymbol{\nu}}\right].$$

Then

$$\hat{\vartheta}_{i,\ell}=y_i^\ell\omega(\tilde{x}_i,\hat{\nu}_n)-\hat{\tau}_n^\ell+\left\langle\chi_\ell(P_n),\nu'_{P_n}\left(\delta_{W_i}-P_n\right)\right\rangle,$$

and so for any $\ell, s\in[p]$ *:*

$$\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\vartheta}_{i,\ell}\hat{\vartheta}_{i,s}-\vartheta_{i,\ell}\vartheta_{i,s}\right|=\frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}+\vartheta_{i,\ell}\right)\left(\hat{\vartheta}_{i,s}-\vartheta_{i,s}+\vartheta_{i,s}\right)-\vartheta_{i,\ell}\vartheta_{i,s}\right|=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right)\left(\hat{\vartheta}_{i,s}-\vartheta_{i,s}\right)+\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right)\vartheta_{i,s}+\vartheta_{i,\ell}\left(\hat{\vartheta}_{i,s}-\vartheta_{i,s}\right)\right|\leq$$

$$\leq\frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right)\left(\hat{\vartheta}_{i,s}-\vartheta_{i,s}\right)\right|+\frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right)\vartheta_{i,s}\right|+\frac{1}{n}\sum_{i=1}^{n}\left|\vartheta_{i,\ell}\left(\hat{\vartheta}_{i,s}-\vartheta_{i,s}\right)\right|;$$

$$\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}=\frac{1}{\sqrt{n}}\,y_i^\ell\frac{\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}+y_i^\ell\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle$$

$$-\left(\hat{\tau}_n^\ell-\tau^\ell\right)+\left\langle\chi_\ell(P_n)-\chi_\ell(P),\nu'_{P_n}\left(\delta_{W_i}-P_n\right)\right\rangle+\left\langle\chi_\ell(P),\nu'_{P_n}\left(\delta_{W_i}-P_n\right)-\nu'_P\left(\delta_{W_i}-P\right)\right\rangle.$$

Then there exist numbers $\lambda_0, ..., \lambda_4$ *such that:*

$$\left|\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right|^2=(\frac{1}{\sqrt{n}}\,y_i^\ell\frac{\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}+y_i^\ell\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle$$

$$-\left(\hat{\tau}_n^\ell-\tau^\ell\right)+\left\langle\chi_\ell(P_n)-\chi_\ell(P),\nu'_{P_n}\left(\delta_{W_i}-P_n\right)\right\rangle+$$

$$+\left\langle\chi_\ell(P),\nu'_{P_n}\left(\delta_{W_i}-P_n\right)-\nu'_P\left(\delta_{W_i}-P\right)\right\rangle)^2\leq\lambda_0\,\frac{1}{n}\left|y_i^\ell\frac{\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}\right|^2+$$

$$\lambda_1\left|\left\langle y_i^\ell\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle\right|^2+\lambda_2\left(\hat{\tau}_n^\ell-\tau^\ell\right)^2+\lambda_3\left\langle\chi_\ell(P_n)-\chi_\ell(P),\nu'_{P_n}\left(\delta_{W_i}-P_n\right)\right\rangle^2+$$

$$\lambda_4\left\langle\chi_\ell(P),\nu'_{P_n}\left(\delta_{W_i}-P_n\right)-\nu'_P\left(\delta_{W_i}-P\right)\right\rangle^2\leq\lambda_0\,\frac{1}{n}\left|y_i^\ell\frac{\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}\right|^2+$$

$$+\lambda_1\left\|y_i^\ell\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}}\right\|^2\|\hat{\nu}_n-\nu_0\|^2+\lambda_2\left(\hat{\tau}_n^\ell-\tau^\ell\right)^2+$$

$$+\lambda_3\,\|\chi_\ell(P_n)-\chi_\ell(P)\|^2\left\|\nu'_{P_n}\left(\delta_{W_i}-P_n\right)\right\|^2+\lambda_4\,\|\chi_\ell(P)\|^2\left\|\nu'_{P_n}\left(\delta_{W_i}-P_n\right)-\nu'_P\left(\delta_{W_i}-P\right)\right\|^2.$$

And so,

$$\frac{1}{n}\sum_{i=1}^n\left|\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right|^2\leq\lambda_0\frac{1}{n}\frac{1}{n}\sum_{i=1}^n\left|y_i^\ell\frac{\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}\right|^2+$$

$$+\lambda_1\|\hat{\nu}_n-\nu_0\|^2\frac{1}{n}\sum_{i=1}^n\left\|y_i^\ell\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}}\right\|^2+\lambda_2\left(\hat{\tau}_n^\ell-\tau^\ell\right)^2+$$

$$+\lambda_3\,\|\chi_\ell(P_n)-\chi_\ell(P)\|^2\frac{1}{n}\sum_{i=1}^n\left\|\nu'_{P_n}\left(\delta_{W_i}-P_n\right)\right\|^2+$$

$$+\lambda_4\,\|\chi_\ell(P)\|^2\frac{1}{n}\sum_{i=1}^n\left\|\nu'_{P_n}\left(\delta_{W_i}-P_n\right)-\nu'_P\left(\delta_{W_i}-P\right)\right\|^2.$$

Then notice that: (i)

$$\frac{1}{n}\frac{1}{n}\sum_{i=1}^n\left|y_i^\ell\frac{\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}\right|^2\leq$$

$$\leq\left(\frac{1}{n}\sum_{i=1}^n\left|y_i^\ell\frac{\omega(\tilde{x}_i,\hat{\nu}_n)-\omega(\tilde{x}_i,\nu_0)-\left\langle\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}},\hat{\nu}_n-\nu_0\right\rangle}{\frac{1}{\sqrt{n}}}\right|\right)^2=o_p(1),$$

which follows from part (ii) of Assumption 4 and was shown in the proof of Lemma 5.1. (ii) $\|\hat{\nu}_n - \nu_0\|^2 = o_p(1)$, as $\nu(\cdot)$ is Hadamard differentiable and hence weakly continuous, and

$$\frac{1}{n}\sum_{i=1}^{n}\left\|y_i^{\ell}\frac{\partial\omega(\tilde{x}_i,\nu_0)}{\partial\boldsymbol{\nu}}\right\|^2 \xrightarrow{p} E_P\left[\left\|y^{\ell}\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right\|^2\right] < \infty$$

by part (iii) of Assumption 8. (iii) $\left(\hat{\tau}_n^{\ell} - \tau^{\ell}\right)^2 = o_p(1)$ which follows from Lemma 5.1; (iv)

$$\|\chi_{\ell}(P_n) - \chi_{\ell}(P)\| = \left\|\frac{1}{n}\sum_{i=1}^{n} y_i^{\ell}\frac{\partial\omega(\tilde{x},\hat{\nu}_n)}{\partial\boldsymbol{\nu}} - E_P\left[y_{\ell}\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right]\right\| \leq$$

$$\leq \left\|\frac{1}{n}\sum_{i=1}^{n} y_i^{\ell}\left(\frac{\partial\omega(\tilde{x},\hat{\nu}_n)}{\partial\boldsymbol{\nu}} - \frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right)\right\| + \left\|\frac{1}{n}\sum_{i=1}^{n} y_i^{\ell}\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}} - E_P\left[y_{\ell}\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}}\right]\right\| = o_p(1),$$

which follows from the weak law of large numbers and part (ii) of Assumption 8; (v) $\frac{1}{n}\sum_{i=1}^{n}\left\|\nu'_{P_n}\left(\delta_{W_i} - P_n\right)\right\|^2 \xrightarrow{p} E_P\|\nu'_P(\delta_W - P)\|^2$ by part (i) of Assumption 8. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\vartheta}_{i,\ell} - \vartheta_{i,\ell}\right|^2 = o_p(1),$$

and Cauchy-Schwartz implies:

$$\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\vartheta}_{i,\ell} - \vartheta_{i,\ell}\right|\left|\hat{\vartheta}_{i,s} - \vartheta_{i,s}\right| \leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\vartheta}_{i,\ell} - \vartheta_{i,\ell}\right|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\vartheta}_{i,s} - \vartheta_{i,s}\right|^2} = o_p(1).$$

2. *Similarly, Assumption 7 and Assumption 5 imply that*

$$\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\psi}^{\ell,k}(W_i)\hat{\psi}^{s,j}(W_i) - \psi_P^{\ell,k}(W_i)\psi_P^{s,j}(W_i)\right| = o_p(1), \forall\, \ell, s = 1, ..., q; s, j = 1, ...p \Rightarrow$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\psi}(W_i)\otimes\hat{\psi}(W_i)\right)\operatorname{vec}(\hat{\boldsymbol{\tau}}_n\hat{\boldsymbol{\tau}}_n') \xrightarrow{p} E_P[\psi(W)\otimes\psi(W)]\operatorname{vec}(\boldsymbol{\tau}\boldsymbol{\tau}') \Rightarrow$$

$$\Rightarrow \operatorname{vec}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\psi}(W_i)\hat{\boldsymbol{\tau}}_n\hat{\boldsymbol{\tau}}_n'\hat{\psi}'(W_i)\right) \xrightarrow{p} \operatorname{vec}(E_P[\psi_P(W)\boldsymbol{\tau}\boldsymbol{\tau}'\psi_P'(W)]) \Rightarrow$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}(W_i)\hat{\boldsymbol{\tau}}_n\hat{\boldsymbol{\tau}}_n'\hat{\psi}'(W_i) \xrightarrow{p} E_P[\psi_P(W)\boldsymbol{\tau}\boldsymbol{\tau}'\psi_P'(W)].$$

3. *Similarly,*

$$\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\vartheta}_{i,\ell}\hat{\psi}_{s,j}(W_i)-\vartheta_{i,\ell}\psi_P^{s,j}(W_i)\right|=$$

$$\frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}+\vartheta_{i,\ell}\right)\left(\hat{\psi}_{s,j}(W_i)-\psi_P^{s,j}(W_i)+\psi_P^{s,j}(W_i)\right)-\vartheta_{i,\ell}\psi_P^{s,j}(W_i)\right|=$$

$$\frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right)\left(\hat{\psi}_{s,j}(W_i)-\psi_P^{s,j}(W_i)\right)+\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right)\psi_P^{s,j}(W_i)+\vartheta_{i,\ell}\left(\hat{\psi}_{s,j}(W_i)-\psi_P^{s,j}(W_i)\right)\right|\leq$$

$$\leq\frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right)\left(\hat{\psi}_{s,j}(W_i)-\psi_P^{s,j}(W_i)\right)\right|+\frac{1}{n}\sum_{i=1}^{n}\left|\left(\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right)\psi_P^{s,j}(W_i)\right|+$$

$$\frac{1}{n}\sum_{i=1}^{n}\left|\vartheta_{i,\ell}\left(\hat{\psi}_{s,j}(W_i)-\psi_P^{s,j}(W_i)\right)\right|=o_p(1),\forall s=1,...,q;\ell,j=1,...,p,$$

which follows from $\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\vartheta}_{i,\ell}-\vartheta_{i,\ell}\right|^2=o_p(1)$ *shown above,* $\frac{1}{n}\sum_{i=1}^{n}\left|\hat{\psi}_{s,j}(W_i)-\psi_P^{s,j}(W_i)\right|^2=o_p(1)$ *that follows from Assumption 7. This completes the proof.* *Q.E.D.*

PROOF *of Theorem 5.5: As shown in Lemma 5.1:*

$$\boldsymbol{\vartheta}_P(W):=\boldsymbol{u}\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}e_{1,k}.$$

Functional $\nu(P)=E_P[\tilde{x}\tilde{x}']^{-1}$ *is Hadamard differentiable with Hadamard derivative :*

$$\nu_P'\left(\delta_{\tilde{x}_i}-P\right)=E_P[\tilde{x}\tilde{x}']^{-1}(E_P[\tilde{x}\tilde{x}']-\tilde{x}_i\tilde{x}_i')E_P[\tilde{x}\tilde{x}']^{-1},$$

$$E\left[\left\|\nu_P'\left(\delta_{\tilde{x}_i}-P\right)\right\|^2\right]<\infty,\text{since } E_P[\|\tilde{x}\|^2]<\infty$$

So part (i) of Assumption 4 and $E_P\left[\|\nu_P'(\delta_W-P)\|^2\right]<\infty$ *are satisfied. Also, notice that*

$$\frac{\partial\omega(\tilde{x},\nu)}{\partial\boldsymbol{\nu}}=\tilde{x}e_{1,k}'\Rightarrow\omega(\tilde{x},\tilde{\nu})-\omega(\tilde{x},\nu_0)-\left\langle\frac{\partial\omega(\tilde{x},\nu_0)}{\partial\boldsymbol{\nu}},\tilde{\nu}-\nu_0\right\rangle=$$

$$=\tilde{x}(\tilde{\nu}-\nu_0)e_{1,k}-\mathrm{tr}\left(e_{1,k}\tilde{x}'(\tilde{\nu}-\nu_0)\right)=0$$

so part (ii) of Assumption 4 vacuously holds.

$$E\left[\left\|\frac{\partial\omega(\tilde{x},\boldsymbol{\nu})}{\partial\boldsymbol{\nu}}\Big|_{\boldsymbol{\nu}=\boldsymbol{\nu}_0}\right\|^2\right]=E[\|\tilde{x}\|^2]<\infty,$$

so part (iii) of Assumption 4 holds as well. Furthermore, $E_P\left[\|u_j\tilde{x}\|^2\right] < \infty \Rightarrow E_P[\boldsymbol{\vartheta}_P(W)\boldsymbol{\vartheta}'_P(W)] < \infty$*, so we can invoke Theorem 5.4. Then it follows from Theorem 5.4*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}(0, \varsigma),$$

$$\varsigma = E[\psi\boldsymbol{\tau}\boldsymbol{\tau}'\psi'] + E\left[\psi\boldsymbol{\tau}\left(\boldsymbol{u}\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}e_{1,k}\right)'\right]A' + AE\left[\boldsymbol{u}\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}e_{1,k}\boldsymbol{\tau}'\psi'\right] +$$

$$+AE\left[\boldsymbol{u}\boldsymbol{u}'\left(\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}e_{1,k}\right)^2\right]A',$$

$$\sqrt{n}\left(\hat{\beta}_j - \beta_j\right) \xrightarrow{d} \mathcal{N}\left(0, \varsigma_j\right),$$

$$\varsigma_j := e'_{j,q}\varsigma e_{j,q} = E\left[\left(e'_{j,q}\psi\boldsymbol{\tau}\right)^2\right] + 2E\left[\left(e'_{j,q}\psi\boldsymbol{\tau}\right)\left(\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}e_{1,k}\right)\left(\boldsymbol{u}'A'e_{j,q}\right)\right] +$$

$$+E\left[\left(e'_{j,q}A\boldsymbol{u}\right)^2\left(\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1}e_{1,k}\right)^2\right];$$

Let

$$\hat{\varrho}_n = \left(\left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1} \otimes \boldsymbol{I}_p\right)\left(\frac{1}{n}\sum_{i=1}^n\left(\mathrm{vec}[\hat{\boldsymbol{u}}_i\tilde{x}_i']\,\mathrm{vec}\,[\hat{\boldsymbol{u}}_i\tilde{x}_i']'\right)\right) \times$$

$$\times\left(\left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1} \otimes \boldsymbol{I}_p\right),$$

each element of the matrix $\frac{1}{n}\sum_{i=1}^n\left(\mathrm{vec}[\hat{\boldsymbol{u}}_i\tilde{x}_i']\,\mathrm{vec}\,[\hat{\boldsymbol{u}}_i\tilde{x}_i']'\right)$*, which has the following form:*

$$\frac{1}{n}\sum_{i=1}^n \tilde{x}_{i,\ell}\hat{u}_{i,h}\hat{u}_{i,j}\tilde{x}_{i,s}.$$

Let $\boldsymbol{\theta}_j := e'_{j,p}E_P[\boldsymbol{y}\tilde{x}']E_P[\tilde{x}\tilde{x}']^{-1}$ *be the* $j-th$ *row of matrix* $(\boldsymbol{\tau}\ \Gamma)$ *and* $\hat{\boldsymbol{\theta}}_j := e'_{j,p}\left(\frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i\tilde{x}_i'\right)\left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_i\tilde{x}_i'\right)^{-1}$ *be the* $j-th$ *row of matrix* $(\hat{\boldsymbol{\tau}}\ \hat{\Gamma})$ *for* $j = 1, ..., p$*. Then*

$$\frac{1}{n}\sum_{i=1}^n \tilde{x}_{i,\ell}\tilde{x}_{i,s}\hat{u}_{i,j}\hat{u}_{i,h} = \frac{1}{n}\sum_{i=1}^n \tilde{x}_{i,\ell}\tilde{x}_{i,s}u_{i,j}u_{i,h} +$$

$$+\frac{1}{n}\sum_{i=1}^n \tilde{x}_{i,\ell}\tilde{x}_{i,s}\left(\hat{u}_{i,j}\hat{u}_{i,h} - u_{i,j}u_{i,h}\right).$$

Further,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i,\ell}\tilde{x}_{i,s}\big(\hat{u}_{i,j}\hat{u}_{i,h}-u_{i,j}u_{i,h}\big)\right| \leq \left|\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i,\ell}\tilde{x}_{i,s}\right| \max_{1\leq i\leq n}\left|\hat{u}_{i,j}\hat{u}_{i,h}-u_{i,j}u_{i,h}\right| =$$

$$= \left|\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i,\ell}\tilde{x}_{i,s}\right| \max_{1\leq i\leq n}\left|\Big(\big(\boldsymbol{\theta}_j-\hat{\boldsymbol{\theta}}_j\big)'\tilde{x}_i+u_{i,j}\Big)\Big(\big(\boldsymbol{\theta}_h-\hat{\boldsymbol{\theta}}_h\big)'\tilde{x}_i+u_{i,h}\Big)-u_{i,j}u_{i,h}\right| =$$

$$= \left|\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i,\ell}\tilde{x}_{i,s}\right| \max_{1\leq i\leq n}\left|\big(\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\big)'\tilde{x}_i\tilde{x}_i'\big(\hat{\boldsymbol{\theta}}_h-\boldsymbol{\theta}_h\big)-\big(\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\big)'\tilde{x}_iu_{i,j}-\big(\hat{\boldsymbol{\theta}}_h-\boldsymbol{\theta}_h\big)'\tilde{x}_iu_{i,h}\right| \leq$$

$$\leq \left|\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i,\ell}\tilde{x}_{i,s}\right|\Big(\big\|\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\big\|\big\|\hat{\boldsymbol{\theta}}_h-\boldsymbol{\theta}_h\big\|\max_{1\leq i\leq n}\|\tilde{x}_i\|+\big\|\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\big\|\max_{1\leq i\leq n}\|\tilde{x}_iu_{i,j}\|+$$

$$+\big\|\hat{\boldsymbol{\theta}}_h-\boldsymbol{\theta}_h\big\|\max_{1\leq i\leq n}\|\tilde{x}_iu_{i,h}\|\Big) \leq \left|\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i,\ell}\tilde{x}_{i,s}-E[\tilde{x}_\ell\tilde{x}_s]\right|\Big(\big\|\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\big\|\big\|\hat{\boldsymbol{\theta}}_h-\boldsymbol{\theta}_h\big\|\max_{1\leq i\leq n}\|\tilde{\boldsymbol{x}}_i\|+$$

$$+\big\|\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\big\|\max_{1\leq i\leq n}\|\tilde{\boldsymbol{x}}_iu_{i,j}\|+\big\|\hat{\boldsymbol{\theta}}_h-\boldsymbol{\theta}_h\big\|\max_{1\leq i\leq n}\|\tilde{\boldsymbol{x}}_iu_{i,h}\|\Big)+$$

$$+|E[\tilde{x}_\ell\tilde{x}_s]|\Big(\big\|\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\big\|\big\|\hat{\boldsymbol{\theta}}_h-\boldsymbol{\theta}_h\big\|\max_{1\leq i\leq n}\|\tilde{\boldsymbol{x}}_i\|+$$

$$+\big\|\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\big\|\max_{1\leq i\leq n}\|\tilde{\boldsymbol{x}}_iu_{i,j}\|+\big\|\hat{\boldsymbol{\theta}}_h-\boldsymbol{\theta}_h\big\|\max_{1\leq i\leq n}\|\tilde{\boldsymbol{x}}_iu_{i,h}\|\Big).$$

Further,

$$P\left(\frac{1}{\sqrt{n}}\max_{1\leq i\leq n}\|\tilde{\boldsymbol{x}}_i\|>\varepsilon\right)=P\big(\cup_{1\leq i\leq n}\left\{\|\tilde{\boldsymbol{x}}_i\|>\sqrt{n}\varepsilon\right\}\big)\leq\sum_{i=1}^{n}P\Big(\|\tilde{\boldsymbol{x}}_i\|^2>n\varepsilon^2\Big)\leq$$

$$\leq n\frac{E\Big[\|\tilde{\boldsymbol{x}}_i\|^2 I\big\{\|\tilde{\boldsymbol{x}}_i\|^2>n\varepsilon^2\big\}\Big]}{n\varepsilon^2}\to 0, n\to\infty,$$

where the last convergence to zero follows from $E\big[\|\tilde{\boldsymbol{x}}\|^2\big]<\infty$. *For the same reason,*

$$\frac{1}{\sqrt{n}}\max_{1\leq i\leq n}\big\|\tilde{x}_iu_{i,j}\big\|\overset{p}{\to}0,$$

$$\frac{1}{\sqrt{n}}\max_{1\leq i\leq n}\big\|\tilde{x}_iu_{i,h}\big\|\overset{p}{\to}0,$$

which follows from $E\Big[\big\|\tilde{x}u_j\big\|^2\Big], E\Big[\|\tilde{x}u_h\|^2\Big]<\infty$. *Thus,*

$$\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i,\ell}\tilde{x}_{i,s}\big(\hat{u}_{i,j}\hat{u}_{i,h}-u_{i,j}u_{i,h}\big)\right|\leq$$

$$\leq o_{p(1)}\left(O_p(1)O_p(1)o_p\left(\frac{1}{\sqrt{n}}\right)+O_p(1)o_p(1)+o_p(1)o_p(1)\right)=o_p(1).$$

This implies that

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_{i,\ell}\tilde{x}_{i,s}\hat{u}_{i,j}\hat{u}_{i,h}\xrightarrow{p}E\left[\tilde{x}_\ell\tilde{x}_s u_j u_h\right],$$

and

$$(e'_{1,k}\otimes\boldsymbol{I}_p)\hat{\varrho}_n(e_{1,k}\otimes\boldsymbol{I}_p)\xrightarrow{p}(e'_{1,k}\otimes\boldsymbol{I}_p)\varrho(e'_{1,k}\otimes\boldsymbol{I}_p)=$$

$$=E_P\left[\boldsymbol{u}\boldsymbol{u}'\left(\tilde{x}'E_P[\tilde{x}\tilde{x}']^{-1'}e_{1,k}\right)^2\right]$$

Furthermore,

$$\hat{\varsigma}=A_n(e'_{1,k}\otimes I_p)\hat{\varrho}_n(e_{1,k}\otimes I_p)A'_n=$$

$$=A_n(e'_{1,k}\otimes I_p)\left(\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\right)\right)^{-1}\otimes I_p)\left(\frac{1}{n}\sum_{i=1}^{n}(\mathrm{vec}[\hat{\boldsymbol{u}}_i\tilde{x}'_i]\,\mathrm{vec}\,[\hat{\boldsymbol{u}}_i\tilde{x}'_i]')\right)\times$$

$$\times\left(\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\right)^{-1}\otimes\boldsymbol{I}_p\right)(e_{1,k}\otimes I_p)A'_n=A_n\left(e'_{1,k}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\right)^{-1}\otimes I_p\right)\times$$

$$\times\left(\frac{1}{n}\sum_{i=1}^{n}(\tilde{\boldsymbol{x}}_i\otimes I_p)\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}'_i(\tilde{\boldsymbol{x}}'_i\otimes I_p)\right)\left(\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\right)^{-1}e_{1,k}\otimes\boldsymbol{I}_p\right)A'_n=$$

$$=A_n\left(\frac{1}{n}\sum_{i=1}^{n}\left(e'_{1,k}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\right)^{-1}\tilde{\boldsymbol{x}}_i\otimes\boldsymbol{I}_p\right)\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}'_i\left(\tilde{\boldsymbol{x}}'_i\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\right)^{-1}e_{1,k}\otimes\boldsymbol{I}_p\right)\right)A'_n=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(e'_{1,k}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_{\boldsymbol{i}}'\right)^{-1}\tilde{\boldsymbol{x}}_i\right)^2A_n\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}'_iA'_n.$$

Also, it follows from Corollary 5.1 that $A_n\hat{\boldsymbol{u}}_i=\hat{\boldsymbol{\varepsilon}}_i$, hence for any $j=1,...,q$:

$$\hat{\varsigma}_{jj}=e'_{j,q}\hat{\varsigma}_n e_{j,q}=\frac{1}{n}\sum_{i=1}^{n}\left(e'_{1,k}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i\tilde{x}'_i\right)^{-1}\tilde{x}_i\right)^2e'_{j,q}\hat{\boldsymbol{\varepsilon}}_i\hat{\boldsymbol{\varepsilon}}'_ie_{j,q}=$$

$$=e'_{1,k}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\hat{\varepsilon}^2_{i,j}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}'_i\right)^{-1}e_{1,k}.$$

hence

$$\sqrt{n}\ \widetilde{\text{s.e.}}^2\left(\hat{\beta}_j\right) = e'_{j,q} A_n \left(e'_{1,k} \otimes \boldsymbol{I}_p\right) \hat{\varrho}_n \left(e_{1,k} \otimes \boldsymbol{I}_p\right) A'_n e_{j,q}$$

by continuous mapping theorem:

$$\sqrt{n}\ \widetilde{\text{s.e.}}^2\left(\hat{\beta}_j\right) \xrightarrow{p} e'_{j,q} A E_P\left[\boldsymbol{u}\boldsymbol{u}'\left(\tilde{x}' E_P[\tilde{x}\tilde{x}']^{-1'} e_{1,k}\right)^2\right] A' e_{j,q} =$$

$$= E_P\left[\left(e'_{j,q} A\boldsymbol{u}\right)^2 \left(\tilde{x}' E_P[\tilde{x}\tilde{x}]^{-1} e_{1,k}\right)^2\right] = \mathrm{Var}_P\left[e'_{j,q} A\boldsymbol{u}\tilde{x}' E_P[\tilde{x}\tilde{x}]^{-1} e_{1,k}\right].$$

so the result follows from Slutskiy theorem. *Q.E.D.*

B. DERIVATION OF INFLUENCE FUNCTION FOR THE LEAST SQUARES ESTIMATOR

As shown in Example 1,

$$\boldsymbol{\tau}(P) = E_P\left[\boldsymbol{y}\,\tilde{x}' E_P[\tilde{x}\tilde{x}']^{-1} e_{1,k}\right], \nu(P) = E_P[\tilde{x}\tilde{x}']^{-1}, \omega(\tilde{x}, \nu(P)) = \tilde{x}'\nu(P) e_{1,k}.$$

$$\nu'_P\left(\delta_{\tilde{x}_i} - P\right) = E_P[\tilde{x}\tilde{x}']^{-1}(E_P[\tilde{x}\tilde{x}'] - \tilde{x}_i\tilde{x}'_i) E_P[\tilde{x}\tilde{x}']^{-1},$$

$$\frac{\partial\omega(\tilde{x}, \nu(P))}{\partial\nu} = \tilde{x} e'_{1,k} \Rightarrow \left\langle E_{\boldsymbol{y},\tilde{x}}\left[y^\ell \frac{\partial\omega(\tilde{x}, \nu(P))}{\partial\boldsymbol{\nu}}\right], \nu'_P\left(\delta_{\tilde{x}_i} - P\right)\right\rangle =$$

$$= \mathrm{tr}\left(e_{1,k} E_P\left[y^\ell\tilde{x}'\right] E_P[\tilde{x}\tilde{x}']^{-1}(E_P[\tilde{x}\tilde{x}'] - \tilde{x}_i\tilde{x}'_i) E_P[\tilde{x}\tilde{x}']^{-1}\right) = \text{using cyclic prop. of tr} =$$

$$= E_P\left[y^\ell\tilde{x}'\right] E_P[\tilde{x}\tilde{x}']^{-1}(E_P[\tilde{x}\tilde{x}'] - \tilde{x}_i\tilde{x}'_i) E_P[\tilde{x}\tilde{x}']^{-1} e_{1,k} =$$

$$= \tau^\ell - E_P\left[y^\ell\tilde{x}'\right] E_P[\tilde{x}\tilde{x}']^{-1}\tilde{x}_i\tilde{x}'_i E_P[\tilde{x}\tilde{x}']^{-1} e_{1,k};$$

so by (15): $\sqrt{n}\left(\hat{\tau}^\ell_n - \tau^\ell\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(y^\ell_i - E_P\left[y^\ell\tilde{x}\right] E_P[\tilde{x}\tilde{x}']^{-1}\tilde{x}_i\right)\tilde{x}'_i E_P[\tilde{x}\tilde{x}']^{-1} e_{1,k} + o_p(1) =$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n u^\ell_i \tilde{x}'_i E_P[\tilde{x}\tilde{x}']^{-1} e_{1,k} + o_p(1).$$

Also, as shown in Theorem 5.3,

$$A^{\mathrm{SN}}_n = A + \frac{1}{n}\sum_{i=1}^n \psi^{\mathrm{SN}}_P(W_i) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

and as shown in Theorem 5.1, the estimate from (33) satisfies: $\hat{\boldsymbol{\beta}}_n = A^{\mathrm{SN}}_n \hat{\boldsymbol{\tau}}_n$. Thus, by ((12)):

$$\sqrt{n}\ \ \hat{\boldsymbol{\beta}}_n = \sqrt{n}\ \ A^{\mathrm{SN}}_n\hat{\boldsymbol{\tau}}_n = \sqrt{n}\left(A^{\mathrm{SN}}_n - A^{\mathrm{SN}} + A^{\mathrm{SN}}\right)(\boldsymbol{\tau} + \hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}) =$$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi^{\mathrm{SN}}_P(W_i)\boldsymbol{\tau} + \sqrt{n}\left(A^{\mathrm{SN}}_n - A^{\mathrm{SN}}\right)(\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}) + \sqrt{n}A^{\mathrm{SN}}\boldsymbol{\tau} +$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} A^{\mathrm{SN}}\boldsymbol{u}_i\tilde{x}_i' E_P[\tilde{x}\tilde{x}']^{-1} e_{1,k} + o_p(1) \Rightarrow$$

$$\Rightarrow \sqrt{n}\ \left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi_P^{\mathrm{SN}}(W_i)\boldsymbol{\tau} + A^{\mathrm{SN}}\boldsymbol{u}_i\tilde{x}_i' E_P[\tilde{x}\tilde{x}']^{-1} e_{1,k}\right) + o_p(1),$$