Econometrica Supplementary Material

SUPPLEMENT TO "SAME ROOT DIFFERENT LEAVES: TIME SERIES AND CROSS-SECTIONAL METHODS IN PANEL DATA"

(*Econometrica*, Vol. 91, No. 6, November 2023, 2125–2154)

DENNIS SHEN
Department of Data Sciences & Operations, University of Southern California

PENG DING
Department of Statistics, University of California, Berkeley

JASJEET SEKHON
Departments of Statistics & Data Science and Political Science, Yale University

BIN YU
Departments of Statistics and EECS, University of California, Berkeley

THIS SUPPLEMENT IS STRUCTURED as follows. Appendix C complements Section 4.1.3 of the main body by presenting confidence intervals under heteroskedastic errors. Appendix D provides an extended discussion on principal component regression (PCR).

APPENDIX C: CONFIDENCE INTERVALS FOR HETEROSKEDASTIC NOISE

In Section 4.1.3, we presented confidence intervals under homoskedastic noise. Here, we present confidence intervals that are also motivated by Theorem 3 but for the heteroskedastic noise setting, that is, $(\boldsymbol{\Sigma}_T^{\text{hz}}, \boldsymbol{\Sigma}_N^{\text{vt}}, \boldsymbol{\Sigma}_T^{\text{dr}}, \boldsymbol{\Sigma}_N^{\text{dr}})$ are diagonal matrices whose nonzero entries are not necessarily identical. We construct our confidence intervals as in (18) and (19) of Section 4.1.3 using two popular strategies to estimate the covariance matrix.

Additional Notation. Recall $\boldsymbol{H}^u = \boldsymbol{U}\boldsymbol{U}'$ and $\boldsymbol{H}^v = \boldsymbol{V}\boldsymbol{V}'$. We define $\boldsymbol{H}_\perp^u = \boldsymbol{I} - \boldsymbol{H}^u$ and $\boldsymbol{H}_\perp^v = \boldsymbol{I} - \boldsymbol{H}^v$. With this notation, the HZ and VT in-sample errors can be written as $\boldsymbol{H}_\perp^u \boldsymbol{y}_T = \boldsymbol{y}_T - \boldsymbol{Y}_0 \widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{H}_\perp^v \boldsymbol{y}_N = \boldsymbol{y}_N - \boldsymbol{Y}_0' \widehat{\boldsymbol{\beta}}$, respectively.

C.1. *Jackknife Variance Estimation*

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

$$\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}} = \operatorname{diag}\big(\big[\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u \circ \boldsymbol{I}\big]^\dagger \big[\boldsymbol{H}_\perp^u \boldsymbol{y}_T \circ \boldsymbol{H}_\perp^u \boldsymbol{y}_T\big]\big), \tag{46}$$

$$\widehat{\boldsymbol{\Sigma}}_N^{\text{jack}} = \operatorname{diag}\big(\big[\boldsymbol{H}_\perp^v \circ \boldsymbol{H}_\perp^v \circ \boldsymbol{I}\big]^\dagger \big[\boldsymbol{H}_\perp^v \boldsymbol{y}_N \circ \boldsymbol{H}_\perp^v \boldsymbol{y}_N\big]\big). \tag{47}$$

Dennis Shen: dennis.shen@marshall.usc.edu
Peng Ding: pengdingpku@berkeley.edu
Jasjeet Sekhon: jasjeet.sekhon@yale.edu
Bin Yu: binyu@berkeley.edu

LEMMA 8:

(i) [HZ model] *Suppose Assumption* 1 *holds*. *If* $(\boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{I})$ *is nonsingular*, *then we have*

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

where $\Delta_{\ell\ell}^{\text{hz}} = \sum_{j\neq\ell} (\sigma_{jT}^{\text{hz}})^2 (H_{\ell j}^{u})^2 (1 - H_{\ell\ell}^{u})^{-2}$ *for* $\ell = 1, \ldots, N_0$.

(ii) [VT model] *Suppose Assumption* 2 *holds*. *If* $(\boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{I})$ *is nonsingular*, *then we have*

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

where $\Gamma_{\ell\ell}^{\text{vt}} = \sum_{j\neq\ell} (\sigma_{Nj}^{\text{vt}})^2 (H_{\ell j}^{v})^2 (1 - H_{\ell\ell}^{v})^{-2}$ *for* $\ell = 1, \ldots, T_0$.

(iii) [DR model] *Suppose Assumption* 3 *holds*. *If* $(\boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{I})$ *and* $(\boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{I})$ *are nonsingular*, *then we have*

$$\begin{aligned}
\mathbb{E}\big[\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}} | \boldsymbol{Y}_0\big] &= \boldsymbol{\Sigma}_T^{\text{dr}} + \boldsymbol{\Delta}^{\text{dr}}, \\
\mathbb{E}\big[\widehat{\boldsymbol{\Sigma}}_N^{\text{jack}} | \boldsymbol{Y}_0\big] &= \boldsymbol{\Sigma}_N^{\text{dr}} + \boldsymbol{\Gamma}^{\text{dr}}, \\
\mathbb{E}\big[\widehat{v}^{\text{dr,jack}}(\boldsymbol{Y}_0) | \boldsymbol{Y}_0\big] &= v_0^{\text{dr}} + \big(\boldsymbol{H}^{u}\boldsymbol{\beta}^*\big)' \boldsymbol{\Delta}^{\text{dr}} \big(\boldsymbol{H}^{u}\boldsymbol{\beta}^*\big) + \big(\boldsymbol{H}^{v}\boldsymbol{\alpha}^*\big)' \boldsymbol{\Gamma}^{\text{dr}} \big(\boldsymbol{H}^{v}\boldsymbol{\alpha}^*\big) \\
&\quad + \operatorname{tr}\big(\boldsymbol{Y}_0^{\dagger} \boldsymbol{\Delta}^{\text{dr}} \big(\boldsymbol{Y}_0'\big)^{\dagger} \boldsymbol{\Gamma}^{\text{dr}}\big),
\end{aligned}$$

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

Lemma 8 establishes that the jackknife is conservative, provided $(\boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{I})$ and $(\boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{I})$ are nonsingular. Strictly speaking, the jackknife is well defined if these quantities are singular, as seen through the pseudoinverse in (46) and (47). Lemma 8 considers the nonsingular case for simplicity. We remark that $\max_\ell H_{\ell\ell}^{u} < 1$ and $\max_\ell H_{\ell\ell}^{v} < 1$ are sufficient conditions for invertibility.

C.2. *Hartley–Rao–Kiefer (HRK) Variance Estimation*

Next, we consider the covariance estimator proposed by Hartley, Rao, and Kiefer (1969). We index this estimator by the authors, Hartley–Rao–Kiefer (HRK):

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}_T^{\text{HRK}} &= \operatorname{diag}\big(\big[\boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{H}_{\perp}^{u}\big]^{-1} \big[\boldsymbol{H}_{\perp}^{u}\boldsymbol{y}_T \circ \boldsymbol{H}_{\perp}^{u}\boldsymbol{y}_T\big]\big), \\
\widehat{\boldsymbol{\Sigma}}_N^{\text{HRK}} &= \operatorname{diag}\big(\big[\boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{H}_{\perp}^{v}\big]^{-1} \big[\boldsymbol{H}_{\perp}^{v}\boldsymbol{y}_N \circ \boldsymbol{H}_{\perp}^{v}\boldsymbol{y}_N\big]\big).
\end{aligned}$$

LEMMA 9:

(i) [HZ model] *Suppose Assumption* 1 *holds*. *If* $(\boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{H}_{\perp}^{u})$ *is nonsingular*, *then we have*

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

(ii) [VT model] *Suppose Assumption* 2 *holds. If* $(\boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{H}_{\perp}^{v})$ *is nonsingular, then we have*

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

(iii) [DR model] *Suppose Assumption* 3 *holds. If* $(\boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{H}_{\perp}^{u})$ *and* $(\boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{H}_{\perp}^{v})$ *are nonsingular, then we have*

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

Lemma 9 establishes that the HRK estimator is unbiased, provided $(\boldsymbol{H}_{\perp}^{u} \circ \boldsymbol{H}_{\perp}^{u})$ and $(\boldsymbol{H}_{\perp}^{v} \circ \boldsymbol{H}_{\perp}^{v})$ are invertible. To discuss sufficient conditions for invertibility, consider $(\boldsymbol{H}^{u} \circ \boldsymbol{H}^{u})$. A sufficient condition is strict diagonal dominance (Varga, 1962): $(1 - H_{\ell\ell}^{u})^2 > \sum_{j \neq \ell} (H_{\ell j}^{u})^2$. Notice that $\boldsymbol{H}^{u}$ is an orthogonal projector and is thus idempotent, that is, $(\boldsymbol{H}^{u})^2 = \boldsymbol{H}^{u}$, and symmetric. Therefore,

$$H_{\ell\ell}^{u} = \big(H_{\ell\ell}^{u}\big)^2 + \sum_{j \neq \ell} \big(H_{\ell j}^{u}\big)^2 \quad \Longrightarrow \quad \sum_{j \neq \ell} \big(H_{\ell j}^{u}\big)^2 = H_{\ell\ell}^{u}\big(1 - H_{\ell\ell}^{u}\big),$$

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

C.3. *Discussion*

We highlight that Lemmas 1 (from Section 4.1.3), 8, and 9 only hold in expectation. For any particular realization, $\widehat{v}_0^{\mathrm{dr}}$ may exhibit unexpected properties. For instance, if $\operatorname{tr}(\boldsymbol{Y}_0^{\dagger} \widehat{\boldsymbol{\Sigma}}_T (\boldsymbol{Y}_0')^{\dagger} \widehat{\boldsymbol{\Sigma}}_N) > \max\{\widehat{v}_0^{\mathrm{hz}}, \widehat{v}_0^{\mathrm{vt}}\}$, then $\widehat{v}_0^{\mathrm{dr}} < \min\{\widehat{v}_0^{\mathrm{hz}}, \widehat{v}_0^{\mathrm{vt}}\}$; thus, the mixed coverage will be smaller than both HZ and VT coverages. In fact, $\widehat{v}_0^{\mathrm{dr}}$ can be negative if $\operatorname{tr}(\boldsymbol{Y}_0^{\dagger} \widehat{\boldsymbol{\Sigma}}_T (\boldsymbol{Y}_0')^{\dagger} \widehat{\boldsymbol{\Sigma}}_N) > \widehat{v}_0^{\mathrm{hz}} + \widehat{v}_0^{\mathrm{vt}}$, which may occur if both HZ and VT in-sample errors are "too large." For these scenarios, one naïve solution is to modify $\widehat{v}_0^{\mathrm{dr}}$ as $\widehat{v}_0^{\mathrm{dr}} = \widehat{v}_0^{\mathrm{hz}} + \widehat{v}_0^{\mathrm{vt}}$, which is conservative by Lemmas 1, 8, and 9. However, this case is arguably better resolved with a different point estimator altogether.

C.4. *Empirical Applications—Extended*

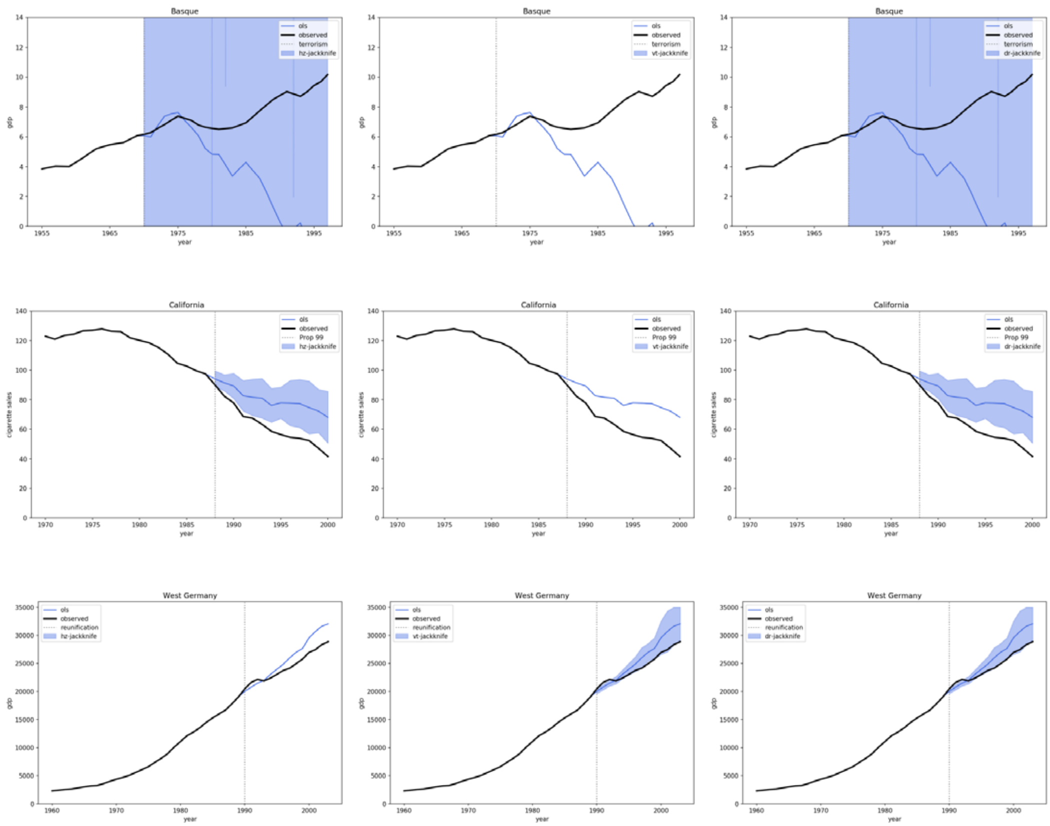
We extend our analysis in Section 5.3.3 to include results with the heteroskedastic confidence intervals. Figure C.1 presents the jackknife-based confidence intervals for our three case studies. We underscore that the conclusions drawn in Section 5.3.3 hold here as well. We remark that the conditions necessary for the HRK-based confidence intervals do not hold for OLS.

C.5. *Deferred Proofs From This Section*

We present the proofs for this section.

C.5.1. *Proof of Lemma 8*

PROOF: Before we establish the biases of $(\widehat{\boldsymbol{\Sigma}}_T^{\mathrm{jack}}, \widehat{\boldsymbol{\Sigma}}_N^{\mathrm{jack}})$, we first justify their forms. Jackknife is a popular approach to estimate the covariances of $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}})$. Below, we follow the



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

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

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

To describe the jackknife, we define $\widehat{\boldsymbol{\alpha}}_{\sim i}$ as the minimum ℓ_2-norm solution to (2), where $\lambda_1 = \lambda_2 = 0$, without the ith observation, that is,

$$\widehat{\boldsymbol{\alpha}}_{\sim i} = \left(\boldsymbol{Y}_{0,\sim i}' \boldsymbol{Y}_{0,\sim i}\right)^{\dagger} \boldsymbol{Y}_{0,\sim i}' \boldsymbol{y}_{T,\sim i}, \tag{48}$$

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

$$\widehat{\boldsymbol{V}}^{\text{jack}} = \frac{1}{(T_0 - 1)^2} \sum_{i \le N_0} (\tilde{\boldsymbol{\alpha}}_i - \widehat{\boldsymbol{\alpha}})(\tilde{\boldsymbol{\alpha}}_i - \widehat{\boldsymbol{\alpha}})'. \tag{49}$$

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

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

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

$$\left(\boldsymbol{Y}_{0,\sim i}'\boldsymbol{Y}_{0,\sim i}\right)^{\dagger} = \left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger} + \left(1 - H_{ii}^u\right)^{-1}\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\boldsymbol{y}_i\boldsymbol{y}_i'\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}. \tag{50}$$

Recall $\widehat{\boldsymbol{\alpha}} = (\boldsymbol{Y}_0'\boldsymbol{Y}_0)^{\dagger}\boldsymbol{Y}_0'\boldsymbol{y}_T$ and note $Y_{iT} - \boldsymbol{y}_i'\widehat{\boldsymbol{\alpha}}$ is the ith element of $\widehat{\boldsymbol{\varepsilon}}_T = \boldsymbol{H}_{\perp}^u\boldsymbol{y}_T$. Using these facts, we plug (50) into (48) to yield

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_{\sim i} &= \left[\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger} + \left(1 - H_{ii}^u\right)^{-1}\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\boldsymbol{y}_i\boldsymbol{y}_i'\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\right]\left(\boldsymbol{Y}_0'\boldsymbol{y}_T - \boldsymbol{y}_i Y_{iT}\right) \\
&= \widehat{\boldsymbol{\alpha}} - \left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\boldsymbol{y}_i Y_{iT} + \left(1 - H_{ii}^u\right)^{-1}\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\boldsymbol{y}_i\boldsymbol{y}_i'\widehat{\boldsymbol{\alpha}} \\
&\quad - H_{ii}^u\left(1 - H_{ii}^u\right)^{-1}\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\boldsymbol{y}_i Y_{iT} \\
&= \widehat{\boldsymbol{\alpha}} - \left(1 - H_{ii}^u\right)^{-1}\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\boldsymbol{y}_i\widehat{\varepsilon}_{iT}.
\end{aligned} \tag{51}$$

Inserting (51) into our pseudo-estimate, we have

$$\begin{aligned}
\tilde{\boldsymbol{\alpha}}_i &= T_0\widehat{\boldsymbol{\alpha}} - (T_0 - 1)\left(\widehat{\boldsymbol{\alpha}} - \left(1 - H_{ii}^u\right)^{-1}\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\boldsymbol{y}_i\widehat{\varepsilon}_{iT}\right) \\
&= \widehat{\boldsymbol{\alpha}} + (T_0 - 1)\left(1 - H_{ii}^u\right)^{-1}\left(\boldsymbol{Y}_0'\boldsymbol{Y}_0\right)^{\dagger}\boldsymbol{y}_i\widehat{\varepsilon}_{iT}.
\end{aligned} \tag{52}$$

Inserting (52) into (49), we have

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

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

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

To arrive at (46), we define $\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}} = \boldsymbol{\Omega}$. This corresponds to the EHW estimator with the jackknife correction. We derive (47) for $\widehat{\boldsymbol{\beta}}$ by applying the same arguments above. Now, we will evaluate the biases of $(\widehat{\boldsymbol{\Sigma}}_T^{\text{jack}}, \widehat{\boldsymbol{\Sigma}}_N^{\text{jack}})$.

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

$$\mathbb{E}\left[\left(\boldsymbol{H}_{\perp}^u \circ \boldsymbol{H}_{\perp}^u \circ \boldsymbol{I}\right)^{\dagger}(\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T)|\boldsymbol{y}_N, \boldsymbol{Y}_0\right] = \left(\boldsymbol{H}_{\perp}^u \circ \boldsymbol{H}_{\perp}^u \circ \boldsymbol{I}\right)^{\dagger}\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T|\boldsymbol{y}_N, \boldsymbol{Y}_0]. \tag{53}$$

To evaluate (53), we follow the derivations of (35) and (37) to obtain

$$\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \boldsymbol{H}_{\perp}^u \boldsymbol{Y}_0 \boldsymbol{\alpha}^* = \boldsymbol{0}, \tag{54}$$

$$\mathrm{Cov}(\widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0) = \boldsymbol{H}_{\perp}^u \boldsymbol{\Sigma}_T^{\mathrm{hz}} \boldsymbol{H}_{\perp}^u. \tag{55}$$

Recall that $\mathbb{E}[X^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2$ for any random variable X. Thus, combining (54) with (55) gives

$$\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0] = \big(\boldsymbol{H}_{\perp}^u \boldsymbol{\Sigma}_T^{\mathrm{hz}} \boldsymbol{H}_{\perp}^u \circ \boldsymbol{I}\big)\boldsymbol{1}. \tag{56}$$

Let $\widehat{\boldsymbol{\gamma}} = \mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0]$. By (56), the ℓth entry of $\widehat{\boldsymbol{\gamma}}$ can be written as

$$\widehat{\gamma}_\ell = \sum_{j \neq \ell} \big(H_{j\ell}^u\big)^2 \big(\sigma_{jT}^{\mathrm{hz}}\big)^2 + \big(1 - H_{\ell\ell}^u\big)^2 \big(\sigma_{\ell T}^{\mathrm{hz}}\big)^2,$$

where $H_{j\ell}^u$ is the (j, ℓ)th entry of $\boldsymbol{H}^u$. In turn, this allows us to rewrite (56) as

$$\widehat{\boldsymbol{\gamma}} = \big(\boldsymbol{H}_{\perp}^u \circ \boldsymbol{H}_{\perp}^u\big) \boldsymbol{\Sigma}_T^{\mathrm{hz}} \boldsymbol{1}. \tag{57}$$

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

$$\widehat{\zeta}_\ell = \big(\sigma_{\ell T}^{\mathrm{hz}}\big)^2 + \sum_{j \neq \ell} \frac{\big(H_{\ell j}^u\big)^2}{\big(1 - H_{\ell\ell}^u\big)^2} \big(\sigma_{jT}^{\mathrm{hz}}\big)^2.$$

Therefore, $\mathrm{diag}(\widehat{\boldsymbol{\zeta}}) = \boldsymbol{\Sigma}_T^{\mathrm{hz}} + \boldsymbol{\Delta}^{\mathrm{hz}}$, where $\Delta_{\ell\ell}^{\mathrm{hz}} = \sum_{j \neq \ell} (\sigma_{jT}^{\mathrm{hz}})^2 (H_{\ell j}^u)^2 (1 - H_{\ell\ell}^u)^{-2}$ for $\ell = 1, \dots, N_0$. Notice if $\max_\ell H_{\ell\ell}^u < 1$, then $(\boldsymbol{H}_{\perp}^u \circ \boldsymbol{H}_{\perp}^u \circ \boldsymbol{I})$ is nonsingular, that is, the pseudo-inverse is precisely the inverse. In this situation, plugging the above into (53) gives

$$\begin{aligned}
\mathbb{E}\big[\widehat{\boldsymbol{\Sigma}}_T^{\mathrm{jack}} | \boldsymbol{y}_N, \boldsymbol{Y}_0\big] &= \mathrm{diag}\big(\big(\boldsymbol{H}_{\perp}^u \circ \boldsymbol{H}_{\perp}^u \circ \boldsymbol{I}\big)^{-1} \mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T | \boldsymbol{y}_N, \boldsymbol{Y}_0]\big) \\
&= \mathrm{diag}\big(\big(\boldsymbol{H}_{\perp}^u \circ \boldsymbol{H}_{\perp}^u \circ \boldsymbol{I}\big)^{-1} \widehat{\boldsymbol{\gamma}}\big) \\
&= \mathrm{diag}(\widehat{\boldsymbol{\zeta}}) \\
&= \boldsymbol{\Sigma}_T^{\mathrm{hz}} + \boldsymbol{\Delta}^{\mathrm{hz}}.
\end{aligned} \tag{58}$$

From this, we conclude that

$$\begin{aligned}
\mathbb{E}\big[\widehat{v}_0^{\mathrm{hz,jack}} | \boldsymbol{y}_N, \boldsymbol{Y}_0\big] &= \widehat{\boldsymbol{\beta}}' \mathbb{E}\big[\widehat{\boldsymbol{\Sigma}}_T^{\mathrm{jack}} | \boldsymbol{y}_N, \boldsymbol{Y}_0\big] \widehat{\boldsymbol{\beta}} \\
&= \widehat{\boldsymbol{\beta}}' \big(\boldsymbol{\Sigma}_T^{\mathrm{hz}} + \boldsymbol{\Delta}^{\mathrm{hz}}\big) \widehat{\boldsymbol{\beta}} \\
&= v_0^{\mathrm{hz}} + \widehat{\boldsymbol{\beta}}' \boldsymbol{\Delta}^{\mathrm{hz}} \widehat{\boldsymbol{\beta}},
\end{aligned}$$

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

(ii) [VT model] Let Assumption 2 hold. Following the arguments above, we conclude $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\mathrm{jack}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\mathrm{vt}} + \boldsymbol{\Gamma}^{\mathrm{vt}}$, where $\Gamma_{\ell\ell}^{\mathrm{vt}} = \sum_{j \neq \ell} (\sigma_{Nj}^{\mathrm{vt}})^2 (H_{\ell j}^v)^2 (1 - H_{\ell\ell}^v)^{-2}$ for $\ell = 1, \dots, T_0$. Thus, $\mathbb{E}[\widehat{v}_0^{\mathrm{vt,jack}} | \boldsymbol{y}_T, \boldsymbol{Y}_0] = v_0^{\mathrm{vt}} + \widehat{\boldsymbol{\alpha}}' \boldsymbol{\Gamma}^{\mathrm{vt}} \widehat{\boldsymbol{\alpha}}$, where we note that $\widehat{\boldsymbol{\alpha}}' \boldsymbol{\Gamma}^{\mathrm{vt}} \widehat{\boldsymbol{\alpha}} \geq 0$.

(iii) [DR model] Let Assumption 3 hold. We define $(\sigma_{iT}^{\mathrm{dr}})^2 = \mathrm{Var}(\varepsilon_{iT}|\boldsymbol{Y}_0)$ for $i = 1, \dots, N_0$ and $(\sigma_{Nt}^{\mathrm{dr}})^2 = \mathrm{Var}(\varepsilon_{Nt}|\boldsymbol{Y}_0)$ for $t = 1, \dots, T_0$. Following the arguments that led to (58), we obtain $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\mathrm{jack}}|\boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\mathrm{dr}} + \boldsymbol{\Delta}^{\mathrm{dr}}$, where $\Delta_{\ell\ell}^{\mathrm{dr}} = \sum_{j\neq\ell}(\sigma_{jT}^{\mathrm{dr}})^2(H_{\ell j}^u)^2(1 - H_{\ell\ell}^u)^{-2}$ for $\ell = 1, \dots, N_0$. Similarly, we obtain $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\mathrm{jack}}|\boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\mathrm{dr}} + \boldsymbol{\Gamma}^{\mathrm{dr}}$, where $\Gamma_{\ell\ell}^{\mathrm{dr}} = \sum_{j\neq\ell}(\sigma_{Nj}^{\mathrm{dr}})^2(H_{\ell j}^v)^2(1 - H_{\ell\ell}^v)^{-2}$ for $\ell = 1, \dots, T_0$. Applying Lemma 7 then gives

$$\mathbb{E}\big[\widehat{v}_0^{\mathrm{dr,jack}}|\boldsymbol{Y}_0\big] = v_0^{\mathrm{dr}} + \big(\boldsymbol{H}^u\boldsymbol{\beta}^*\big)'\boldsymbol{\Delta}^{\mathrm{dr}}\big(\boldsymbol{H}^u\boldsymbol{\beta}^*\big) + \big(\boldsymbol{H}^v\boldsymbol{\alpha}^*\big)'\boldsymbol{\Gamma}^{\mathrm{dr}}\big(\boldsymbol{H}^v\boldsymbol{\alpha}^*\big) + \mathrm{tr}\big(\boldsymbol{Y}_0^\dagger\boldsymbol{\Delta}^{\mathrm{dr}}\big(\boldsymbol{Y}_0'\big)^\dagger\boldsymbol{\Gamma}^{\mathrm{dr}}\big).$$

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

C.5.2. *Proof of Lemma 9*

PROOF: We adopt the strategy of Hartley, Rao, and Kiefer (1969) to prove our desired result.

(i) [HZ model] Let Assumption 1 hold. As in the proof of Lemma 8, we define $\widehat{\boldsymbol{\varepsilon}}_T = \boldsymbol{H}_\perp^u\boldsymbol{y}_T$. Observe

$$\mathbb{E}\big[\big(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u\big)^{-1}(\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T)|\boldsymbol{y}_N, \boldsymbol{Y}_0\big] = \big(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u\big)^{-1}\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T|\boldsymbol{y}_N, \boldsymbol{Y}_0]. \tag{59}$$

To evaluate (59), we plug in (57) to obtain

$$\mathbb{E}\big[\big(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u\big)^{-1}(\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T)|\boldsymbol{y}_N, \boldsymbol{Y}_0\big] = \big(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u\big)^{-1}\big(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u\big)\boldsymbol{\Sigma}_T^{\mathrm{hz}}\mathbf{1} = \boldsymbol{\Sigma}_T^{\mathrm{hz}}\mathbf{1}. \tag{60}$$

Plugging (60) into (59) yields

$$\mathbb{E}\big[\widehat{\boldsymbol{\Sigma}}_T^{\mathrm{HRK}}|\boldsymbol{y}_N, \boldsymbol{Y}_0\big] = \mathrm{diag}\big(\big(\boldsymbol{H}_\perp^u \circ \boldsymbol{H}_\perp^u\big)^{-1}\mathbb{E}[\widehat{\boldsymbol{\varepsilon}}_T \circ \widehat{\boldsymbol{\varepsilon}}_T|\boldsymbol{y}_N, \boldsymbol{Y}_0]\big) = \boldsymbol{\Sigma}_T^{\mathrm{hz}}. \tag{61}$$

It then follows that $\mathbb{E}[\widehat{v}_0^{\mathrm{hz,HRK}}|\boldsymbol{y}_N, \boldsymbol{Y}_0] = v_0^{\mathrm{hz}}$.

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

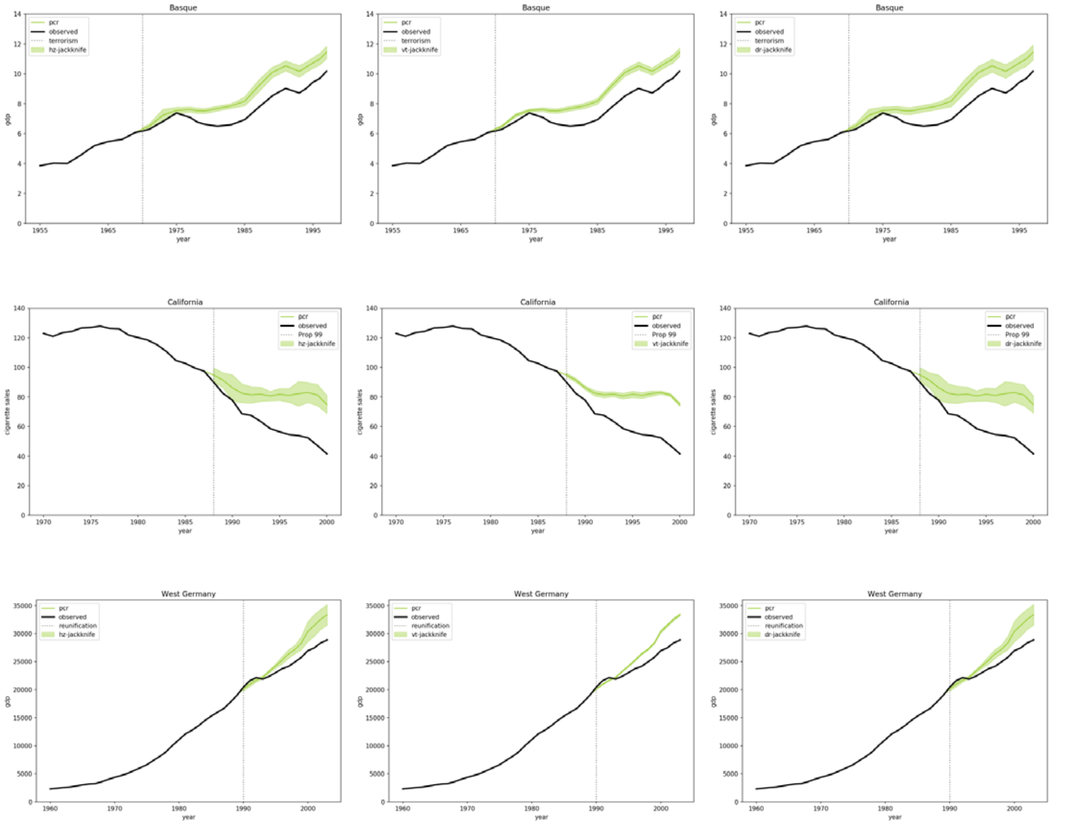
(iii) [DR model] Let Assumption 3 hold. Following the arguments that led to (61), we obtain $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_T^{\mathrm{HRK}}|\boldsymbol{Y}_0] = \boldsymbol{\Sigma}_T^{\mathrm{dr}}$ and $\mathbb{E}[\widehat{\boldsymbol{\Sigma}}_N^{\mathrm{HRK}}|\boldsymbol{Y}_0] = \boldsymbol{\Sigma}_N^{\mathrm{dr}}$. Applying Lemma 7 then gives $\mathbb{E}[\widehat{v}_0^{\mathrm{dr,HRK}}|\boldsymbol{Y}_0] = v_0^{\mathrm{dr}}$. The proof is complete. *Q.E.D.*

APPENDIX D: PRINCIPAL COMPONENT REGRESSION

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

D.1. *Comparing PCR to OLS*

Intuitively, PCR-based models postulate that the data are inherently low-dimensional. We comment on several benefits of PCR over OLS. To begin, the HZ and VT OLS variance estimators constructed in Section 4.1.3 can suffer from degeneracy when N and T are of different sizes. That is, if $N < T$, then the HZ in-sample error is zero (assuming full column rank), which causes the HZ coverage to collapse on the point estimate; analogous statements hold for the VT coverage when $N > T$. The PCR-based variance estimators, on the other hand, can avoid degeneracy through the number of chosen principal components k (regularization). On a related note, the nonsingularity conditions required for the



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

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

jackknife and HRK variance estimators can also be controlled by k. See Agarwal, Shah, and Shen (2021) for various methods on choosing k.

D.2. *Empirical Applications—Extended*

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

REFERENCES

AGARWAL, ANISH, DEVAVRAT SHAH, AND DENNIS SHEN (2021): "Synthetic Interventions,". arXiv preprint arXiv:2006.07691. [8]

CLINE, RANDALL E. (1965): "Representations for the Generalized Inverse of Sums of Matrices," *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis*, 2 (1), 99–114. [5]

HARTLEY, HERMAN O., J. N. K. RAO, AND GRACE KIEFER (1969): "Variance Estimation With One Unit per Stratum," *Journal of the American Statistical Association*, 64 (327), 841–851. [2,7]

MEYER, CARL D. (1973): "Generalized Inversion of Modified Matrices," *SIAM Journal on Applied Mathematics*, 24 (3), 315–323. [5]

VARGA, RICHARD S. (1962): *Matrix Iterative Analysis*. Prentice-Hall Series in Automatic Computation. Englewood Cliffs: Prentice-Hall. [3]

Co-editor Guido W. Imbens handled this manuscript.

Manuscript received 10 October, 2022; final version accepted 12 October, 2023; available online 12 October, 2023.