Supplement to "Leave-out estimation of variance components"

Patrick Kline, Raffaele Saggio, Mikkel Sølvsten

May 22, 2020

1 Proofs

Proof of Lemma 4. Define $\mathrm{B}_p = \frac{1}{p}\sum_{i=1}^n B_{ii}\sigma_i^2(1-P_{ii})^{-2}\big(2\sum_{\ell\neq i}^n P_{i\ell}^4 - P_{ii}^2(1-P_{ii})^2\big)$. Letting $(\hat{\theta}_{JLA}-\hat{\theta})_2$ be a second order approximation of $\hat{\theta}_{JLA}-\hat{\theta}$, we first show that $\mathbb{E}\big[(\hat{\theta}_{JLA}-\hat{\theta})_2\big] = \mathrm{B}_p$ and $\mathbb{V}[\hat{\theta}]{-1}\big(\mathbb{V}[(\hat{\theta}_{JLA}-\hat{\theta})_2]\big) = O(1/p)$. Then we finish the proof of the first claim by showing that the approximation error is ignorable. The bias bound follows immediately from the equality $\sum_{\ell\neq i}^n P_{i\ell}^2 = P_{ii}(1-P_{ii})$ which leads to $0 \leq \sum_{\ell\neq i}^n P_{i\ell}^4 \leq P_{ii}^2(1-P_{ii})^2$.

We have $\hat{\theta}_{JLA}-\hat{\theta} = (\hat{\theta}_{JLA}-\hat{\theta})_2 + AE_2$ where

$$(\hat{\theta}_{JLA}-\hat{\theta})_2 = \sum\nolimits_{i=1}^n \hat{\sigma}_i^2\left(B_{ii} - \hat{B}_{ii} - \hat{B}_{ii}\hat{a}_i - \hat{B}_{ii}\left(\hat{a}_i^2 - \frac{1}{p}\frac{3P_{ii}^3+P_{ii}^2}{1-P_{ii}}\right)\right)$$

for $\hat{a}_i = (1-P_{ii})^{-1}(\hat{P}_{ii}-P_{ii})$ and approximation error

$$AE_2 = \sum\nolimits_{i=1}^n \hat{\sigma}_i^2\hat{B}_{ii}\left(\frac{1}{p}\frac{3\hat{P}_{ii}^2+\hat{P}_{ii}^2-(3P_{ii}^2+P_{ii}^2)(1+\hat{a}_i)^2}{(1+\hat{a}_i)^2(1-P_{ii})} - \frac{\hat{a}_i^3}{1+\hat{a}_i}\right).$$

For the mean calculation involving $(\hat{\theta}_{JLA}-\hat{\theta})_2$ we use independence between $\hat{B}_{ii}$, $\hat{P}_{ii}$, and $\hat{\sigma}_i^2$, unbiasedness of $\hat{B}_{ii}$, $\hat{P}_{ii}$, and $\hat{\sigma}_i^2$, and the variance formula

$$\mathbb{V}[\hat{a}_i] = \frac{2}{p}\frac{P_{ii}^2-\sum_{\ell=1}^n P_{i\ell}^4}{(1-P_{ii})^2} = \frac{1}{p}\frac{3P_{ii}^3+P_{ii}^2}{1-P_{ii}} + \frac{P_{ii}^2(1-P_{ii})^2-2\sum_{\ell\neq i}^n P_{i\ell}^4}{p(1-P_{ii})^2}.$$

Taken together this implies that

$$\mathbb{E}\left[(\hat{\theta}_{JLA}-\hat{\theta})_2\right] = -\sum\nolimits_{i=1}^n B_{ii}\sigma_i^2\left(\mathbb{V}[\hat{\alpha}_i] - \frac{1}{p}\frac{3P_{ii}^3+P_{ii}^2}{1-P_{ii}}\right) = \mathrm{B}_p.$$

For the variance calculation we proceed term by term. We have for $y = (y_1, \ldots, y_n)'$ that

$$\begin{aligned}
\mathbb{V}\Big[\sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 (B_{ii} - \hat{B}_{ii})\Big] &= \mathbb{E}\Big[\mathbb{V}\Big[\sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \hat{B}_{ii} \mid y\Big]\Big] \leq 2p^{-1} \sum\nolimits_{i=1}^{n} \sum\nolimits_{\ell=1}^{n} B_{i\ell}^2 \mathbb{E}\big[\hat{\sigma}_i^2 \hat{\sigma}_\ell^2\big] \\
&= O\Big(p^{-1} \text{trace}(\tilde{A}^2)\Big), \\
\mathbb{V}\Big[\sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \hat{B}_{ii} \hat{a}_i\Big] &= \mathbb{E}\Big[\mathbb{V}\Big[\sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \hat{B}_{ii} \hat{a}_i \mid y, R_B\Big]\Big] \leq 2p^{-1} \sum\nolimits_{i=1}^{n} \sum\nolimits_{\ell=1}^{n} P_{i\ell}^2 \tfrac{\mathbb{E}\big[\hat{B}_{ii}\hat{B}_{\ell\ell}\big]\mathbb{E}\big[\hat{\sigma}_i^2\hat{\sigma}_\ell^2\big]}{(1-P_{ii})(1-P_{\ell\ell})} \\
&= O\Big(p^{-1} \text{trace}(\tilde{A}^2) + p^{-2} \text{trace}(\tilde{A}_1^2)^{1/2} \text{trace}(\tilde{A}_2^2)^{1/2}\Big)
\end{aligned}$$

where $\tilde{A}_\ell = S_{xx}^{-1/2} A_\ell' A_\ell S_{xx}^{-1/2}$ for $\ell = 1, 2$,

$$\begin{aligned}
\mathbb{V}\Big[\sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \hat{B}_{ii} \big(\hat{a}_i^2 - \mathbb{V}[\hat{a}_i]\big)\Big] &= \sum\nolimits_{i=1}^{n} \sum\nolimits_{\ell=1}^{n} \mathbb{E}\Big[\hat{B}_{ii}\hat{B}_{\ell\ell}\Big] \mathbb{E}\big[\hat{\sigma}_i^2 \hat{\sigma}_\ell^2\big] Cov\big(\hat{a}_i^2, \hat{a}_\ell^2\big) \\
&= O\Big(p^{-2} \text{trace}(\tilde{A}^2) + p^{-3} \text{trace}(\tilde{A}_1^2)^{1/2} \text{trace}(\tilde{A}_2^2)^{1/2}\Big) \\
\mathbb{V}\left[\sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \left(\hat{B}_{ii} - B_{ii}\right) \tfrac{2\sum_{\ell\neq i}^{n} P_{i\ell}^4 - P_{ii}^2 (1-P_{ii})^2}{p(1-P_{ii})^2}\right] &= O\left(p^{-3} \text{trace}(\tilde{A}^2)\right) \\
\mathbb{V}\left[\sum\nolimits_{i=1}^{n} B_{ii} \left(\hat{\sigma}_i^2 - \sigma_i^2\right) \tfrac{2\sum_{\ell\neq i}^{n} P_{i\ell}^4 - P_{ii}^2 (1-P_{ii})^2}{p(1-P_{ii})^2}\right] &= O\Big(p^{-2} \mathbb{V}[\hat{\theta}]\Big)
\end{aligned}$$

From these bounds it follows that $\mathbb{V}[\hat{\theta}]^{-1/2}\big((\hat{\theta}_{JLA} - \hat{\theta})_2 - \mathrm{B}_p\big) = o_p(1)$ since $\text{trace}(\tilde{A}^2) = O(\mathbb{V}[\hat{\theta}])$ and $p^{-4}\mathbb{V}[\hat{\theta}]^{-2}\mathbb{V}[\hat{\theta}_1]\mathbb{V}[\hat{\theta}_2] = o(1)$.

We now treat the approximation error while utilizing that $\mathbb{E}[\hat{a}_i^3] = O\left(1/p^2\right)$, $\mathbb{E}[\hat{a}_i^4] = O\left(1/p^2\right)$, and $\max_i |\hat{a}_i| = o_p(\log(n)/\sqrt{p})$ which follows from (Achlioptas, 2003, Theorem 1.1 and its proof). Proceeding term by term, we list the conclusions

$$\begin{aligned}
\sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \hat{B}_{ii} \hat{a}_i^3 + \sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \hat{B}_{ii} \hat{a}_i^4 &= p^{-2} O_p\Big(\mathbb{E}[\hat{\theta}_{1,\mathrm{PI}} - \theta_1] + \mathbb{E}[\hat{\theta}_{2,\mathrm{PI}} - \theta_2]\Big) \\
\sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \hat{B}_{ii} \frac{\hat{a}_i^5}{1 + \hat{a}_i} &= \tfrac{\log(n)}{p^{5/4}} O_p\Big(\mathbb{E}[\hat{\theta}_{1,\mathrm{PI}} - \theta_1] + \mathbb{E}[\hat{\theta}_{2,\mathrm{PI}} - \theta_2]\Big) \\
\tfrac{1}{p} \sum\nolimits_{i=1}^{n} \hat{\sigma}_i^2 \hat{B}_{ii} \tfrac{3\hat{P}_{ii}^2 + \hat{P}_{ii}^2 - (3P_{ii}^2 + P_{ii}^2)(1+\hat{a}_i)^2}{(1+\hat{a}_i)^2 (1-P_{ii})} &= \Big(p^{-2} + \tfrac{\log(n)}{p^{5/4}}\Big) O_p\Big(\mathbb{E}[\hat{\theta}_{1,\mathrm{PI}} - \theta_1] + \mathbb{E}[\hat{\theta}_{2,\mathrm{PI}} - \theta_2]\Big)
\end{aligned}$$

which finishes the proof. □

Proof of Lemma B.1. The proof of Lemma B.1 uses the notation and verifies the conditions of Lemmas A2.1 and A2.2 in Sølvsten (2020) referred to as SS2.1 and SS2.2, respectively. First, we show marginal convergence in distribution of $\mathcal{S}_n$ and $\mathcal{U}_n$. Then, we show joint convergence in distribution of $\mathcal{S}_n$ and $\mathcal{U}_n$. Let $V_n = (v_1, \ldots, v_n)$ where $\{v_i\}_i$ are as in the setup of Lemma B.1. Before starting we note that $\max_i \sigma_i^{-2} = O(1)$ and $2\sum_{i=1}^{n} \sum_{\ell\neq i} W_{i\ell}^2 \sigma_i^2 \sigma_\ell^2 = 1$

imply $\text{trace}(W^2) = \sum_{i=1}^n \sum_{\ell \neq i} W_{i\ell}^2 = O(1)$ so that $\lambda_{\max}(W^2) = o(1) \Leftrightarrow \text{trace}(W^4) = o(1)$.

We first consider the marginal distribution of $\mathcal{S}_n$.

Result 1.1. *If* $\max_i \mathbb{E}[v_i^4] + \sigma_i^{-2} = O(1)$, $\sum_{i=1}^n \dot{w}_i^2 \sigma_i^2 = 1$, *and* $\max_i \dot{w}_i^2 = o(1)$, *then* $\mathcal{S}_n \xrightarrow{d} \mathcal{N}(0,1)$.

In the notation of SS2.1 we have, $\Delta_i^0 \mathcal{S}_n = \dot{w}_i v_i$ and $\mathbb{E}[T_n \mid V_n] = 1 + \frac{1}{2} \sum_{i=1}^n \dot{w}_i^2 (v_i^2 - \sigma_i^2)$, so it follows from $\max_i \mathbb{E}[v_i^4] + \sigma_i^{-2} = O(1)$, $\sum_{i=1}^n \dot{w}_i^2 \sigma_i^2 = 1$, and Lemma B.1(i) that

$$\mathbb{E}[T_n \mid V_n] \xrightarrow{\mathcal{L}^1} 1, \quad \sum\nolimits_{i=1}^n \mathbb{E}[(\Delta_i^0 \mathcal{S}_n)^2] = 1, \qquad \sum\nolimits_{i=1}^n \mathbb{E}[(\Delta_i^0 \mathcal{S}_n)^4] \leq \max_i \tfrac{\mathbb{E}[v_i^4]}{\sigma_i^2} \dot{w}_i^2 = o(1),$$

so Result 1.1 follows from SS2.1.

Next we consider the marginal distribution of $\mathcal{U}_n$.

Result 1.2. *If* $\max_i \mathbb{E}[v_i^4] + \sigma_i^{-2} = O(1)$, $2\sum_{i=1}^n \sum_{\ell \neq i} W_{n,i\ell}^2 \sigma_{n,i}^2 \sigma_{n,\ell}^2 = 1$, *and trace*$(W^4) = o(1)$, *then* $\mathcal{U}_n \xrightarrow{d} \mathcal{N}(0,1)$.

In the notation of SS2.1 we have,

$$\Delta_i^0 \mathcal{U}_n = 2 v_i \sum\nolimits_{\ell \neq i} W_{i\ell} v_\ell \quad \text{and} \quad E[T_n \mid V_n] = \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} \sum\nolimits_{k \neq i} (v_i + \sigma_i^2) W_{i\ell} W_{ik} v_\ell v_k,$$

and

$$\sum\nolimits_{i=1}^n \mathbb{E}[(\Delta_i^0 \mathcal{U}_n)^2] = 2, \qquad \sum\nolimits_{i=1}^n \mathbb{E}[(\Delta_i^0 \mathcal{U}_n)^4] \leq 2^5 \max_i \mathbb{E}[v_i^4]^2 \max_i \sigma_i^{-4} \max_i \sum\nolimits_{\ell \neq i} W_{i\ell}^2,$$

where $\max_i \sum_{\ell \neq i} W_{i\ell}^2 \leq \text{trace}(W^4)^{1/2} = o(1)$. Now, split $\mathbb{E}[T_n \mid V_n] - 1$ into three terms

$$\begin{aligned}
a_n &= \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} \sigma_i^2 W_{i\ell}^2 (v_\ell + v_\ell^2 - \sigma_\ell^2) \\
b_n &= 2 \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} \sum\nolimits_{k \neq i,\ell} \sigma_k^2 W_{\ell k} W_{ik} v_i v_\ell + \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} W_{i\ell}^2 v_i (v_\ell^2 - \sigma_\ell^2) \\
c_n &= \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} \sum\nolimits_{k \neq i,\ell} W_{i\ell} W_{ik} (v_i^2 - \sigma_i^2) v_\ell v_k.
\end{aligned}$$

Convergence in $\mathcal{L}^1$ The random variables a_n, b_n, and c_n are a linear sum, a quadratic sum, and a cubic sum. We treat similar sums later, so we record sufficient conditions for their convergence in $\mathcal{L}^1$. For brevity, let $\sum_{i \neq \ell}^n = \sum_{i=1}^n \sum_{\ell \neq i}$, and $\sum_{i \neq \ell \neq k}^n = \sum_{i=1}^n \sum_{\ell \neq i} \sum_{k \neq i,\ell}$, etc. Use the notation $u_i = (v_{i1}, v_{i2}, v_{i3}, v_{i4}) \in \mathbb{R}^4$ to denote independent random vectors in order that the result applies to combinations of v_i and $v_i^2 - \sigma_i^2$ as in a_n, b_n, and c_n. For the inferential results we also treat quartic sums and provide the sufficient conditions here.

Result 1.3. *Let $S_{n1}=\sum_{i=1}^{n}\omega_i v_{i1}$, $S_{n2}=\sum_{i\neq\ell}^{n}\omega_{i\ell}v_{i1}v_{\ell 2}$, $S_{n3}=\sum_{i\neq\ell\neq k}^{n}\omega_{i\ell k}v_{i1}v_{\ell 2}v_{k3}$, and $S_{n4}=\sum_{i\neq\ell\neq k\neq m}^{n}\omega_{i\ell km}v_{i1}v_{\ell 2}v_{k3}v_{m4}$ where the weights ω_i, $\omega_{i\ell}$, $\omega_{i\ell k}$, and $\omega_{i\ell km}$ are non-random. Suppose that $\mathbb{E}[u_i]=0$, $\max_i \mathbb{E}[u_i'u_i]=O(1)$.*

1. *If $\sum_{i=1}^{n}\omega_i^2=o(1)$, then $S_{n1}\xrightarrow{\mathcal{L}^1}0$.*

2. *If $\sum_{i\neq\ell}^{n}\omega_{i\ell}^2=o(1)$, then $S_{n2}\xrightarrow{\mathcal{L}^1}0$.*

3. *If $\sum_{i\neq\ell\neq k}^{n}\omega_{i\ell k}^2=o(1)$, then $S_{n3}\xrightarrow{\mathcal{L}^1}0$.*

4. *If $\sum_{i\neq\ell\neq k\neq m}^{n}\omega_{i\ell km}^2=o(1)$, then $S_{n4}\xrightarrow{\mathcal{L}^1}0$.*

Consider S_{n3}, the other results follows from the same line of reasoning. In the notation of SS2.2 we have,

$$\Delta_i^0 S_{n3}=v_{i1}\sum_{\ell\neq i}\sum_{k\neq i,\ell}\omega_{i\ell k}v_{\ell 2}v_{k3}+v_{i2}\sum_{\ell\neq i}\sum_{k\neq i,\ell}\omega_{\ell ik}v_{\ell 1}v_{k3}+v_{i3}\sum_{\ell\neq i}\sum_{k\neq i,\ell}\omega_{\ell ki}v_{\ell 1}v_{k2}.$$

Focusing on the first term we have,

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left[\left(v_{i1}\sum_{\ell\neq i}\sum_{k\neq i,\ell}\omega_{i\ell k}v_{\ell 2}v_{k3}\right)^2\right]&\leq\max_i\mathbb{E}[u_i'u_i]^3\sum_{i\neq\ell\neq k}^{n}\left(\omega_{i\ell k}^2+\omega_{i\ell k}\omega_{ik\ell}\right)\\&\leq 2\max_i\mathbb{E}[u_i'u_i]^3\sum_{i\neq\ell\neq k}^{n}\omega_{i\ell k}^2,\end{aligned}$$

so the results follows from SS2.2, $\sum_{i\neq\ell\neq k}^{n}\omega_{i\ell k}^2=o(1)$, and the observation that the last bound also applies to the other two terms in $\Delta_i^0 S_{n3}$.

Returning to Result 1.2, we need to see how $a_n\xrightarrow{\mathcal{L}^1}0$, $b_n\xrightarrow{\mathcal{L}^1}0$ and $c_n\xrightarrow{\mathcal{L}^1}0$ follows from Result 1.3. Let $\bar{W}_{i\ell}=\sum_{k=1}^{n}W_{ik}W_{k\ell}$ and note that $\text{trace}(W^4)=\sum_{i=1}^{n}\sum_{\ell=1}^{n}\bar{W}_{i\ell}^2$. We have

$$\begin{gathered}\sum_{i=1}^{n}\left(\sum_{\ell\neq i}\sigma_\ell^2W_{i\ell}^2\right)^2\leq\max_i\sigma_i^4\sum_{i=1}^{n}\bar{W}_{ii}^2.\\ \sum_{i=1}^{n}\sum_{\ell\neq i}\left(\sum_{k\neq i,\ell}\sigma_k^2W_{\ell k}W_{ik}\right)^2\leq\max_i\sigma_i^4\sum_{i=1}^{n}\sum_{\ell=1}^{n}\bar{W}_{i\ell}^2\\ \sum_{i=1}^{n}\sum_{\ell\neq i}W_{i\ell}^4=O\left(\max_{i,\ell}W_{i\ell}^2\right)\\ \sum_{i=1}^{n}\sum_{\ell\neq i}\sum_{k\neq i,\ell}W_{i\ell}^2W_{ik}^2=O\left(\max_i\sum_{\ell\neq i}W_{i\ell}^2\right),\end{gathered}$$

all of which are $o(1)$ as $\text{trace}(W^4)=o(1)$.

Finally, we consider the joint distribution of $(\mathcal{S}_n, \mathcal{U}_n)'$. Let $(u_1, u_2)' \in R^2$ be given and non-random with $u_1^2 + u_2^2 = 1$. Define $\mathcal{W}_n = u_1 \mathcal{S}_n + u_2 \mathcal{U}_n$. Lemma B.1 follows if we show that $\mathcal{W}_n \xrightarrow{d} \mathcal{N}(0,1)$. In the notation of SS2.1 we have,

$$\Delta_i^0 \mathcal{W}_n = u_1 \dot{w}_i v_i + u_2 2 v_i \sum\nolimits_{\ell \neq i} W_{i\ell} v_\ell$$

and

$$\begin{aligned}\mathbb{E}[T_n \mid V_n] = u_1^2 \left(1 + \tfrac{1}{2} \sum\nolimits_{i=1}^n \dot{w}_i^2 (v_i^2 - \sigma_i^2)\right) + u_2^2 \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} \sum\nolimits_{k \neq i} (v_i + \sigma_i^2) W_{i\ell} W_{ik} v_\ell v_k \\ + u_1 u_2 3 \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} (v_i^2 + \sigma_i^2) \dot{w}_i W_{i\ell} v_j.\end{aligned}$$

The proofs of Results 1.1 and 1.2 lead to $\sum_{i=1}^n \mathbb{E}[(\Delta_i^0 \mathcal{W}_n)^2] = O(1)$, $\sum_{i=1}^n \mathbb{E}[(\Delta_i^0 \mathcal{W}_n)^4] = o(1)$, and that the first two terms of $\mathbb{E}[T_n \mid V_n]$ converge to $u_1^2 + u_2^2 = 1$. Thus the lemma follows if we show that the conditional covariance $3 \sum_{i=1}^n \sum_{\ell \neq i} (v_i^2 + \sigma_i^2) \dot{w}_i W_{i\ell} v_j$ converges to 0 in $\mathcal{L}^1$. This conditional covariance involves a linear and a quadratic sum:

$$\begin{aligned}\sum\nolimits_{i=1}^n \left(\sum\nolimits_{\ell \neq i} \sigma_\ell^2 w_\ell W_{i\ell}\right)^2 \leq \max_i \sigma_i^4 \max_\ell \lambda_\ell^2(W) \sum\nolimits_{i=1}^n \dot{w}_i^2 = O(\max_\ell \lambda_\ell^2(W)) \\ \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} \dot{w}_i^2 W_{i\ell}^2 \leq \sum\nolimits_{i=1}^n \sum\nolimits_{\ell \neq i} W_{i\ell}^2 \max_i \dot{w}_i^2 = O(\max_i \dot{w}_i^2)\end{aligned}$$

and Result 1.3 ends the proof. □

Proof of Lemma 5. The proof continues in two steps. First, it shows that $\hat{\mathbb{V}}[\hat{\theta}]$ has positive bias of smaller order than $\mathbb{V}[\hat{\theta}]$ when $|\mathcal{B}| = O(1)$. Second, it shows that $\hat{\mathbb{V}}[\hat{\theta}] - \mathbb{E}[\hat{\mathbb{V}}[\hat{\theta}]] = o_p(\mathbb{V}[\hat{\theta}])$. Combined with Theorem 2, these conclusions establish the claims of the lemma.
Bias of $\hat{\mathbb{V}}[\hat{\theta}]$ For the first term in $\hat{\mathbb{V}}[\hat{\theta}]$, a simple calculation shows that

$$\begin{aligned}\mathbb{E}\left[4 \sum_{i=1}^n \left(\sum\nolimits_{\ell \neq i} C_{i\ell} y_\ell\right)^2 \tilde{\sigma}_i^2\right] &= 4 \sum_{i=1}^n \left(\sum\nolimits_{\ell \neq i} C_{i\ell} x_\ell' \beta\right)^2 \sigma_i^2 + 4 \sum_{i=1}^n \sum_{\ell \neq i} C_{i\ell}^2 \sigma_i^2 \sigma_\ell^2 \\ &\quad + 4 \sum_{i=1}^n \sum_{\ell \neq i} \sum_{m=1}^n C_{mi} C_{m\ell} (P_{mi,1} P_{m\ell,2} + P_{mi,2} P_{m\ell,1}) \sigma_i^2 \sigma_\ell^2 \\ &= \mathbb{V}[\hat{\theta}] + 2 \sum_{i=1}^n \sum_{\ell \neq i} \tilde{C}_{i\ell} \sigma_i^2 \sigma_\ell^2.\end{aligned}$$

For the second term in $\hat{\mathbb{V}}[\hat{\theta}]$, we note that if $P_{ik,-\ell} P_{\ell k,-i} = 0$ for all k, then independence between error terms yield $\mathbb{E}[\widehat{\sigma_i^2 \sigma_\ell^2}] = \mathbb{E}[\hat{\sigma}_{i,-\ell}^2] \mathbb{E}[\hat{\sigma}_{\ell,-i}^2] = \sigma_i^2 \sigma_\ell^2$. Otherwise if $P_{i\ell,1} + P_{i\ell,2} = 0$, then

$$\mathbb{E}\big[\widehat{\sigma_i^2\sigma_\ell^2}\big] = \mathbb{E}\Big[\Big(\varepsilon_i - \sum\nolimits_{j\neq i} P_{ij,1}\varepsilon_j\Big)\Big(\varepsilon_i - \sum\nolimits_{k\neq i} P_{ik,2}\varepsilon_k\Big)(x_\ell'\beta + \varepsilon_\ell)\Big(\varepsilon_\ell - \sum\nolimits_{m\neq\ell} P_{\ell m,-i}\varepsilon_m\Big)\Big]$$
$$= \sigma_i^2\sigma_\ell^2 + x_\ell'\beta\mathbb{E}\left[\left(\varepsilon_i - \sum\nolimits_{j\neq i} P_{ij,1}\varepsilon_j\right)\left(\varepsilon_i - \sum\nolimits_{k\neq i} P_{ik,2}\varepsilon_k\right)\sum\nolimits_{m\neq\ell} P_{\ell m,-i}\varepsilon_m\right]$$

where the second term is zero since $P_{\ell i,-i} = 0$ and $P_{ij,1}P_{ij,2} = 0$ for all j. The same argument applies with the roles of i and ℓ reversed when $P_{\ell i,1} + P_{\ell i,2} = 0$.

Finally, when $(i,\ell) \in \mathcal{B}$ we have

$$\mathbb{E}\big[\widehat{\sigma_i^2\sigma_\ell^2}\big] = \left(\sigma_i^2\left(\sigma_\ell^2 + ((x_\ell - \bar{x})'\beta)^2\right) + O\left(n^{-1}\right)\right)1_{\{\tilde{C}_{i\ell}<0\}}$$

where the remainder is uniform in (i,ℓ) and stems from the use of $\bar{y}$ as an estimator of $\bar{x}'\beta$. Thus for sufficiently large n, $\mathbb{E}[\tilde{C}_{i\ell}\widehat{\sigma_i^2\sigma_\ell^2}]$ is smaller than $\tilde{C}_{i\ell}\sigma_i^2\sigma_\ell^2$ leading to a positive bias in $\hat{\mathbb{V}}[\hat{\theta}]$. This bias is

$$\sum\nolimits_{(i,\ell)\in\mathcal{B}} \tilde{C}_{i\ell}\sigma_i^2\left(\sigma_\ell^2 1_{\{\tilde{C}_{i\ell}>0\}} + ((x_\ell - \bar{x})'\beta)^2 1_{\{\tilde{C}_{i\ell}<0\}}\right) + O\left(\mathbb{V}[\hat{\theta}]/n\right)$$

which is ignorable when $|\mathcal{B}| = O(1)$.

Variability of $\hat{\mathbb{V}}[\hat{\theta}]$ Now, $\hat{\mathbb{V}}[\hat{\theta}] - \mathbb{E}[\hat{\mathbb{V}}[\hat{\theta}]]$ involves a number of terms all of which are linear, quadratic, cubic, or quartic sums. Result 1.3 provides sufficient conditions for their convergence in $\mathcal{L}^1$ and therefore in probability. We have already treated versions of linear, quadratic, and cubic terms carefully in the proof of Lemma *B*.1. Thus, we report here the calculations for the quartic terms (details for the remaining terms can be provided upon request) as they also highlight the role of the high-level condition $\lambda_{\max}(P_sP_s') = O(1)$ for $s = 1, 2$.

The quartic term in $4\sum_{i=1}^n \left(\sum_{\ell\neq i} C_{i\ell}y_\ell\right)^2 \tilde{\sigma}_i^2$ is $\sum_{i\neq\ell\neq m\neq k}^n \omega_{i\ell mk}\varepsilon_i\varepsilon_\ell\varepsilon_m\varepsilon_k$ where

$$\omega_{i\ell mk} = \sum_{j=1}^n C_{ji}C_{j\ell}M_{jm,1}M_{jk,2} \quad \text{and} \quad M_{i\ell,s} = \begin{cases} 1, & \text{if } i = \ell, \\ -P_{i\ell,s}, & \text{if } i \neq \ell. \end{cases}$$

Letting $\odot$ denote Hadamard (element-wise) product and $M_s = I_n - P_s$, we have

$$
\begin{aligned}
\sum_{i \neq \ell \neq m \neq k}^{n} \omega_{i\ell mk}^2 \leq \sum_{i,\ell,m,k}^{n} \omega_{i\ell mk}^2 &= \sum_{j,j'} (C^2)_{jj'}^2 (M_1 M_1')_{jj'} (M_2 M_2')_{jj'} \\
&= \operatorname{trace}\left((C^2 \odot C^2)(M_1 M_1' \odot M_2 M_2')\right) \\
&\leq \lambda_{\max}\left(M_1 M_1' \odot M_2 M_2'\right) \operatorname{trace}\left(C^2 \odot C^2\right) = O\left(\operatorname{trace}\left(C^4\right)\right) = o\left(\mathbb{V}[\hat{\theta}]^2\right)
\end{aligned}
$$

where $\lambda_{\max}\left(M_1 M_1' \odot M_2 M_2'\right) = O(1)$ follows from $\lambda_{\max}(P_s P_s') = O(1)$ and we established the last equality in the proof of Theorem 2. The quartic term involved in $2\sum_{i=1}^n \sum_{\ell \neq i} \tilde{C}_{i\ell} \widehat{\sigma_i^2 \sigma_\ell^2}$ has variability of the same order as $\sum_{i \neq \ell \neq m \neq k}^n \omega_{i\ell mk} \varepsilon_i \varepsilon_\ell \varepsilon_m \varepsilon_k$ where

$$
\omega_{i\ell mk} = \tilde{C}_{i\ell} M_{im,1} M_{lk,1} + \sum_{j=1}^n \tilde{C}_{ij} M_{im,1} M_{jk,1} M_{j\ell,2}.
$$

Letting $\tilde{C} = (\tilde{C}_{i\ell})_{i,\ell}$, we find that

$$
\begin{aligned}
\sum_{i \neq \ell \neq m \neq k}^{n} \omega_{i\ell mk}^2 &\leq 2\sum_{i,\ell}^{n} \tilde{C}_{i\ell}^2 (M_1 M_1')_{ii} (M_2 M_2')_{\ell\ell} + 2\sum_{j,j'} \sum_{i}^{n} \tilde{C}_{ij} \tilde{C}_{ij'} (M_1 M_1')_{ii} (M_1 M_1')_{jj'} (M_2 M_2')_{jj'} \\
&= O\left(\sum_{i,\ell}^n \tilde{C}_{i\ell}^2 + \operatorname{trace}\left((\tilde{C}^2 \odot M_1 M_1')(M_1 M_1' \odot M_2 M_2')\right)\right) = O\left(\operatorname{trace}\left(\tilde{C}^2\right)\right).
\end{aligned}
$$

We have $\tilde{C} = C \odot C + 2(C \odot P_1)'(C \odot P_2) + 2(C \odot P_2)'(C \odot P_1)$, from which we obtain that

$$
\operatorname{trace}(\tilde{C}^2) = O\left(\left(\max_{i,\ell} C_{i\ell}^2 + \lambda_{\max}(C^2)\right) \operatorname{trace}(C^2)\right) = o\left(\mathbb{V}[\hat{\theta}]^2\right)
$$

where we established the last equality in the proof of Theorem 2. $\square$

Section 6.2 proposed standard errors for the case of $q > 0$, but omitted a few definition as they were analogous to those for the case of $q = 0$. Those definitions are $\tilde{C}_{i\ell q} = C_{i\ell q}^2 + 2\sum_{m=1}^n C_{miq} C_{m\ell q} (P_{mi,1} P_{m\ell,2} + P_{mi,2} P_{m\ell,1})$ where $C_{i\ell q} = B_{i\ell q} - 2^{-1} M_{i\ell}\left(M_{ii}^{-1} B_{iiq} + M_{\ell\ell}^{-1} B_{\ell\ell q}\right)$ for $B_{i\ell q} = B_{i\ell} - \sum_{s=1}^q \lambda_s w_{is} w_{\ell s}$. Furthermore,

$$
\widetilde{\sigma_i^2 \sigma_\ell^2} = \begin{cases}
\hat{\sigma}_{i,-\ell}^2 \cdot \hat{\sigma}_{\ell,-i}^2, & \text{if } P_{ik,-\ell} P_{\ell k,-i} = 0 \text{ for all } k, \\
\tilde{\sigma}_i^2 \cdot \hat{\sigma}_{\ell,-i}^2, & \text{else if } P_{i\ell,1} + P_{i\ell,2} = 0, \\
\hat{\sigma}_{i,-\ell}^2 \cdot \tilde{\sigma}_\ell^2, & \text{else if } P_{\ell i,1} + P_{\ell i,2} = 0, \\
\hat{\sigma}_{i,-\ell}^2 \cdot (y_\ell - \bar{y})^2 \cdot 1_{\{\tilde{C}_{i\ell q} < 0\}}, & \text{otherwise.}
\end{cases}
$$

Proof of Lemma 6. The statements $\mathbb{V}[\hat{\mathsf{b}}_q]^{-1}\hat{\mathbb{V}}[\hat{\mathsf{b}}_q] \xrightarrow{p} I_q$ and $\mathbb{V}[\hat{\theta}_q]^{-1}\hat{\mathbb{V}}[\hat{\theta}_q] \xrightarrow{p} 1$ follow by applying the arguments in the proofs of Theorem 1 and Lemma 5. Thus we focus on the remaining claim that

$$\delta(v) := \frac{\hat{\mathcal{C}}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q] - \mathcal{C}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q]}{\mathbb{V}[v'\hat{\mathsf{b}}_q]^{1/2}\mathbb{V}[\hat{\theta}_q]^{1/2}} \xrightarrow{p} 0 \quad \text{where} \quad \hat{\mathcal{C}}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q] = 2\sum_{i=1}^{n} v'\mathsf{w}_{iq}\left(\sum_{\ell \neq i} C_{i\ell q} y_\ell\right)\tilde{\sigma}_i^2$$

for all non-random $v \in \mathbb{R}^q$ with $v'v = 1$.

Unbiasedness of $\hat{\mathcal{C}}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q]$ Since $\tilde{\sigma}_i^2$ is unbiased for σ_i^2, it follows that

$$\mathbb{E}\left[\hat{\mathcal{C}}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q]\right] = 2\sum_{i=1}^{n} v'\mathsf{w}_{iq}\left(\sum_{\ell \neq i} C_{i\ell q} x_\ell'\beta\right)\sigma_i^2 + 2\sum_{i=1}^{n} v'\mathsf{w}_{iq}\left(\sum_{\ell \neq i} C_{i\ell q}\mathbb{E}[\varepsilon_\ell \tilde{\sigma}_i^2]\right) = \mathcal{C}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q]$$

as split sampling ensures that $\mathbb{E}[\varepsilon_\ell \tilde{\sigma}_i^2] = 0$ for $\ell \neq i$.

Variability of $\hat{\mathcal{C}}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q]$ Now, $\hat{\mathcal{C}}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q] - \mathcal{C}[v'\hat{\mathsf{b}}_q, \hat{\theta}_q]$ is composed of the following linear, quadratic, and quartic sums:

$$\sum_{i=1}^{n} v'\mathsf{w}_{iq}\left[\left(\varepsilon_i^2 - \sigma_i^2\right)\sum_{\ell \neq i} C_{i\ell q} x_\ell'\beta + \sigma_i^2 \sum_{\ell \neq i} C_{i\ell q}\varepsilon_\ell + \sum_{\ell \neq i} C_{i\ell q}\sigma_\ell^2 \sum_{k \neq \ell}\left(M_{i\ell,1}M_{ik,2} + M_{i\ell,2}M_{ik,1}\right)\varepsilon_k\right]$$

$$\begin{aligned}\sum_{i=1}^{n} v'\mathsf{w}_{iq}\Bigg[&\sum_{\ell \neq i} C_{i\ell q} x_\ell'\beta \sum_{m}\sum_{k \neq m} M_{im,1}M_{ik,2}\varepsilon_m\varepsilon_k + \sum_{\ell \neq i} C_{i\ell q}\varepsilon_\ell\left(\varepsilon_i^2 - \sigma_i^2\right) \\ &+ \sum_{\ell \neq i} C_{i\ell q}\sum_{k \neq \ell}\left(M_{i\ell,1}M_{ik,2} + M_{i\ell,2}M_{ik,1}\right)\varepsilon_k\left(\varepsilon_\ell^2 - \sigma_\ell^2\right)\Bigg]\end{aligned}$$

$$\sum_{i=1}^{n} v'\mathsf{w}_{iq}\sum_{\ell \neq i} C_{i\ell q}\sum_{m \neq \ell}\sum_{k \neq m,\ell} M_{im,1}M_{ik,2}\varepsilon_\ell\varepsilon_m\varepsilon_k$$

These seven terms are $o_p(\mathbb{V}[v'\hat{\mathsf{b}}_q]^{1/2}\mathbb{V}[\hat{\theta}_q]^{1/2})$ by Result 1.3 as outlined in the following.

$$\sum_{i=1}^{n}(v'\mathsf{w}_{iq})^2\left(\sum_{\ell\neq i}C_{i\ell q}x_\ell'\beta\right)^2 = O(\max_i \mathsf{w}_{iq}'\mathsf{w}_{iq}\mathbb{V}[\hat{\theta}_q]) = o(\mathbb{V}[v'\hat{\mathsf{b}}_q]\mathbb{V}[\hat{\theta}_q])$$

$$\sum_{\ell=1}^{n}\left(\sum_{i=1}^{n}v'\mathsf{w}_{iq}C_{i\ell q}\right)^2 = O(\lambda_{\max}(C_q^2)\mathbb{V}[v'\hat{\mathsf{b}}_q]) = O(\lambda_{q+1}^2\mathbb{V}[v'\hat{\mathsf{b}}_q]) = o(\mathbb{V}[v'\hat{\mathsf{b}}_q]\mathbb{V}[\hat{\theta}_q])$$

$$\sum_{k=1}^{n}\left(\sum_{i=1}^{n}v'\mathsf{w}_{iq}\sum_{\ell}C_{i\ell q}M_{i\ell,1}M_{ik,2}\right)^2 = O(\max_i \mathsf{w}_{iq}'\mathsf{w}_{iq}\text{trace}(C_qM_1\odot C_qM_1)) = o(\mathbb{V}[v'\hat{\mathsf{b}}_q]\mathbb{V}[\hat{\theta}_q])$$

$$\sum_{m=1}^{n}\sum_{k=1}^{n}\left(\sum_{i=1}^{n}v'\mathsf{w}_{iq}\sum_{\ell\neq i}C_{i\ell q}x_\ell'\beta M_{im,1}M_{ik,2}\right)^2 = O\left(\sum_{i=1}^{n}(v'\mathsf{w}_{iq})^2\left(\sum_{\ell\neq i}C_{i\ell q}x_\ell'\beta\right)^2\right)$$

$$\sum_{i=1}^{n}\sum_{\ell\neq i}C_{i\ell q}^2(v'\mathsf{w}_{iq})^2 = O(\max_i \mathsf{w}_{iq}'\mathsf{w}_{iq}\mathbb{V}[\hat{\theta}_q])$$

$$\sum_{k=1}^{n}\sum_{\ell=1}^{n}\left(\sum_{i=1}^{n}v'\mathsf{w}_{iq}C_{i\ell q}M_{i\ell,1}M_{ik,2}\right)^2 = O\left(\mathbb{V}[v'\hat{\mathsf{b}}_q]\lambda_{\max}((C_q\odot M_1)(C_q\odot M_1)')\right) = o(\mathbb{V}[v'\hat{\mathsf{b}}_q]\mathbb{V}[\hat{\theta}_q])$$

$$\sum_{\ell=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{n}\left(\sum_{i=1}^{n}v'\mathsf{w}_{iq}C_{i\ell q}M_{im,1}M_{ik,2}\right)^2 = O\left(\mathbb{V}[v'\hat{\mathsf{b}}_q]\lambda_{\max}(C_q^2)\right) \qquad \square$$

Before turning to a proof of Lemma 7, we give precise definitions of the curvature and critical value used in the construction of our proposed confidence interval. The curvature as introduced for the general problem considered by Andrews and Mikusheva (2016) does not have a closed-form representation, but we show that it does in the special case considered here. For implementation, the a closed form solution circumvents numerical approximation.

Critical value function For a given curvature $\kappa > 0$ and confidence level $1-\alpha$, the critical value function $z_{\alpha,\kappa}$ is the $(1-\alpha)$'th quantile of

$$\rho\left(\chi_q, \chi_1, \kappa\right) = \sqrt{\chi_q^2 + \left(\chi_1 + \frac{1}{\kappa}\right)^2} - \frac{1}{\kappa}$$

where χ_q^2 and χ_1^2 are independently distributed random variables from the χ-squared distribution with q and 1 degrees of freedom, respectively. $\rho\left(\chi_q, \chi_1, \kappa\right)$ is the Euclidean distance from (χ_q, χ_1) to the circle with center $(0, -\frac{1}{\kappa})$ and radius $\frac{1}{\kappa}$. The critical value function at $\kappa = 0$ is the limit of $z_{\alpha,\kappa}$ as $\kappa \downarrow 0$, which is the $(1-\alpha)$'th quantile of a central χ_1^2 random variable. See Andrews and Mikusheva (2016) for additional details.

Curvature For generic $\tilde{\Sigma}_q$, our proposed confidence interval $C_\alpha^\theta(\tilde{\Sigma}_q)$ inverts hypotheses of the type $H_0 : \theta = c$ versus $H_1 : \theta \neq c$ based on the value of the test statistic

$$\min_{\mathsf{b}_q,\theta_q : g(\mathsf{b}_q,\theta_q,c)=0} \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix}' \tilde{\Sigma}_q^{-1} \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix}$$

where $g(\mathsf{b}_q, \theta_q, c) = \sum_{\ell=1}^{q} \lambda_\ell \dot{b}_\ell^2 + \theta_q - c$ and $\mathsf{b}_q = (\dot{b}_1, \ldots, \dot{b}_q)'$. This testing problem depends on the manifold $S = \{x = \tilde{\Sigma}_q^{-1/2}(\mathsf{b}_q, \theta_q)' : g(\mathsf{b}_q, \theta_q, c) = 0\}$ for which we need an upper bound on the maximal curvature. We derive this upper bound using the parameterization $\mathbf{x}(\dot{y}) = \tilde{\Sigma}_q^{-1/2}(\dot{y}_1, \ldots, \dot{y}_q, c - \sum_{\ell=1}^{q} \lambda_\ell \dot{y}_\ell^2)'$ which maps from $\mathbb{R}^q$ to S, is a homeomorphism, and has a Jacobian of full rank:

$$d\mathbf{x}(\dot{y}) = \tilde{\Sigma}_q^{-1/2} \begin{bmatrix} \operatorname{diag}(1, \ldots, 1) \\ -2\lambda_1 \dot{y}_1, \ldots, -2\lambda_q \dot{y}_q \end{bmatrix}$$

The maximal curvature of S, $\kappa(\tilde{\Sigma}_q)$, is then given as $\kappa(\tilde{\Sigma}_q) = \max_{\dot{y} \in \mathbb{R}^q} \kappa_{\dot{y}}$ where

$$\kappa_{\dot{y}} = \sup_{u \in \mathbb{R}^q} \frac{\|(I - P_{\dot{y}})V(u \odot u)\|}{\|d\mathbf{x}(\dot{y})u\|^2}, \qquad V = \tilde{\Sigma}_q^{-1/2} \begin{bmatrix} 0 \\ -2\lambda_1, \ldots, -2\lambda_q \end{bmatrix},$$

and $P_{\dot{y}} = d\mathbf{x}(\dot{y})(d\mathbf{x}(\dot{y})' d\mathbf{x}(\dot{y}))^{-1} d\mathbf{x}(\dot{y})'$.

Curvature when $q = 1$ In this case the maximization over u drops out and we have

$$\kappa(\tilde{\Sigma}_1) = \max_{\dot{y} \in \mathbb{R}} \frac{\sqrt{V'V - \frac{(v'V)^2}{v'v}}}{v'v} \qquad \text{where } v = \tilde{\Sigma}_1^{-1/2}(1, -2\lambda_1 \dot{y})'$$

and $V = \tilde{\Sigma}_1^{-1/2}(0, -2\lambda_1)$. The value $\dot{y}^* = -\frac{\tilde{\rho}\tilde{\mathbb{V}}[\hat{\theta}_q]}{2\lambda_1 \tilde{\mathbb{V}}[\hat{b}_1]}$ for $\tilde{\rho} = \frac{\tilde{\mathbb{C}}[\hat{b}_1, \hat{\theta}_q]}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2} \tilde{\mathbb{V}}[\hat{\theta}_q]^{1/2}}$ is both a minimizer of $v'v$ and $(v'V)^2$, so we obtain that $\kappa(\tilde{\Sigma}_1) = \frac{2|\lambda_1| \tilde{\mathbb{V}}[\hat{b}_1]}{\tilde{\mathbb{V}}[\hat{\theta}_q]^{1/2}(1 - \tilde{\rho}^2)^{1/2}}$.

Curvature when $q > 1$ In this case we first maximize over $\dot{y}$ and then over u. For a fixed u we want to find

$$\max_{\dot{y} \in \mathbb{R}^q} \frac{\sqrt{V_u' V_u - V_u' P_{\dot{y}} V_u}}{v_{u,\dot{y}}' v_{u,\dot{y}}} 5 \text{ where } V_u = \tilde{\Sigma}_q^{-1/2}(0, -2\sum_{\ell=1}^{q} \lambda_\ell u_\ell^2),\ v_{u,\dot{y}} = \tilde{\Sigma}_q^{-1/2}(u', -2u' D_q \dot{y})',$$

and $D_q = \operatorname{diag}(\lambda_1, \ldots, \lambda_q)$. The value for $\dot{y}$ that solves $-2D_q \dot{y} = \tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{-1} \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q, \hat{\theta}_q]$ sets

$P_{\dot{y}} V_u = 0$ and minimizes $v_{u,\dot{y}}' v_{u,\dot{y}}$. Thus we obtain

$$\kappa(\tilde{\Sigma}_q) = \frac{2 \max_{u \in \mathbb{R}^q} \frac{|u' D_q u|}{u' \tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{-1} u}}{\left(\tilde{\mathbb{V}}[\hat{\theta}_q] - \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q, \hat{\theta}_q]' \tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{-1} \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q, \hat{\theta}_q]\right)^{1/2}} = \frac{2 |\dot{\dot{\lambda}}_1 (\tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{1/2} D_q \tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{1/2})|}{\left(\tilde{\mathbb{V}}[\hat{\theta}_q] - \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q, \hat{\theta}_q]' \tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{-1} \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q, \hat{\theta}_q]\right)^{1/2}}$$

where $\dot{\dot{\lambda}}_1(\cdot)$ is the eigenvalue of largest magnitude. This formula simplifies to the one derived above when $q = 1$.

Proof of Lemma 7. The following two conditions are the inputs to the proof of Theorem 2 in Andrews and Mikusheva (2016), from which it follows that

$$\begin{aligned} \liminf_{n \to \infty} \mathbb{P}\left(\theta \in \hat{C}^{\theta}_{\alpha,q}\right) &= \liminf_{n \to \infty} \mathbb{P}\left(\min_{(\mathsf{b}_q', \theta_q)' : g(\mathsf{b}_q, \theta_q, \theta) = 0} \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix}' \hat{\Sigma}_q^{-1} \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix} \leq z^2_{\alpha, \hat{\kappa}_q} \right) \\ &\geq 1 - \alpha \end{aligned}$$

where $g(\mathsf{b}_q, \theta_q, \theta) = \sum_{\ell=1}^{q} \lambda_\ell \dot{b}_\ell^2 + \theta_q - \theta$ and $\mathsf{b}_q = (\dot{b}_1, \ldots, \dot{b}_q)'$.

Condition (i) requires that $\hat{\Sigma}_q^{-1/2} \left((\hat{\mathsf{b}}_q', \hat{\theta}_q)' - \mathbb{E}[(\hat{\mathsf{b}}_q', \hat{\theta}_q)'] \right) \xrightarrow{d} \mathcal{N}\left(0, I_{q+1}\right)$, which follows from Theorem 3 and $\Sigma_q^{-1} \hat{\Sigma}_q \xrightarrow{p} I_{q+1}$.

Condition (ii) is satisfied if the conditions of Lemma 1 in Andrews and Mikusheva (2016) are satisfied. To verify this, take the manifold

$$\tilde{S} = \left\{ \dot{x} \in \mathbb{R}^{q+1} : \tilde{g}(\dot{x}) = 0 \right\}$$

for

$$\tilde{g}(\dot{x}) = \dot{x}' \hat{\Sigma}_q^{1/2} \begin{bmatrix} D_q & 0 \\ 0 & 0 \end{bmatrix} \hat{\Sigma}_q^{1/2} \dot{x} + (2\mathbb{E}[\hat{\mathsf{b}}_q]', 1) \begin{bmatrix} D_q & 0 \\ 0 & 1 \end{bmatrix} \hat{\Sigma}_q^{1/2} \dot{x}.$$

The curvature of $\tilde{S}$ is $\hat{\kappa}$, $\tilde{g}(0) = 0$, and $\tilde{g}$ is continuously differentiable with a Jacobian of rank 1. These are the conditions of Lemma 1 in Andrews and Mikusheva (2016). □

2 Calculation of $C_\alpha^\theta(\tilde{\Sigma}_1)$ in practice

To calculate our proposed confidence interval one can rely on an implicit representation of $C_\alpha^\theta(\tilde{\Sigma}_1)$ which is $C_\alpha^\theta(\tilde{\Sigma}_1) = \left[\lambda_1 b_{1,-}^2 + \theta_{1,-}, \lambda_1 b_{1,+}^2 + \theta_{1,+}\right]$ where $b_{1,\pm}$ and $\theta_{1,\pm}$ are solutions to

$$b_{1,\pm} = \hat{b}_1 \pm z_{\alpha,\kappa(\tilde{\Sigma}_1)} \left(\tilde{\mathbb{V}}[\hat{b}_1](1 - \tilde{a}(b_{1,\pm}))\right)^{1/2}, \tag{1}$$

$$\theta_{1,\pm} = \hat{\theta}_1 - \tilde{\rho}\frac{\tilde{\mathbb{V}}[\hat{\theta}_1]^{1/2}}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}}(\hat{b}_1 - b_{1,\pm}) \pm z_{\alpha,\kappa(\tilde{\Sigma}_1)} \left(\tilde{\mathbb{V}}[\hat{\theta}_1](1 - \tilde{\rho}^2)\tilde{a}(b_{1,\pm})\right)^{1/2} \tag{2}$$

for $\tilde{a}(\dot{b}_1) = \left(1 + \left(\text{sgn}(\lambda_1)\kappa(\tilde{\Sigma}_1)\dot{b}_1\tilde{\mathbb{V}}[\hat{b}_1]^{-1/2} + \tilde{\rho}/\sqrt{1-\tilde{\rho}^2}\right)^2\right)^{-1}$.

This construction is fairly intuitive. When $\hat{\rho} = 0$, the interval has endpoints that combine

$$\lambda_1 \left(\hat{b}_1 \pm z_{\alpha,\kappa(\tilde{\Sigma}_1)} \left(\tilde{\mathbb{V}}[\hat{b}_1](1 - \tilde{a}(b_{1,\pm}))\right)^{1/2}\right)^2 \quad \text{and} \quad \hat{\theta}_q \pm z_{\alpha,\kappa(\tilde{\Sigma}_1)} \left(\tilde{\mathbb{V}}[\hat{\theta}_q]a(b_{1,\pm})\right)^{1/2}$$

where $a(\dot{b}_1)$ estimates the fraction of $\mathbb{V}[\hat{\theta}]$ that stems from $\hat{\theta}_1$ when $\mathbb{E}[\hat{b}_1] = \dot{b}_1$. When $\hat{\rho}$ is non-zero, $C_\alpha^\theta(\tilde{\Sigma}_1)$ involves an additional rotation of $(\hat{b}_1, \hat{\theta}_1)'$. This representation of $C_\alpha^\theta(\tilde{\Sigma}_1)$ is however not unique as (1),(2) can have multiple solutions. Thus we derive the representation above together with an additional side condition that ensures uniqueness and represents $b_{1,\pm}$ and $\theta_{1,\pm}$ as solutions to a fourth order polynomial.

Derivation The upper end of $C_\alpha^\theta(\tilde{\Sigma}_1)$ is found by noting that maximization over a linear function in θ_1 implies that the constraint must bind at the maximum. Thus we can reformulate the bivariate problem as a univariate problem

$$\max_{(\dot{b}_1,\dot{\theta}_1)\in \mathsf{E}_\alpha(\tilde{\Sigma}_1)} \lambda_1\dot{b}_1^2 + \dot{\theta}_1 = \max_{\dot{b}_1} \lambda_1\dot{b}_1^2 + \hat{\theta}_1 - \tilde{\rho}\frac{\tilde{\mathbb{V}}[\hat{\theta}_1]^{1/2}}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}}(\hat{b}_1 - \dot{b}_1) + \sqrt{\tilde{\mathbb{V}}[\hat{\theta}_1](1-\tilde{\rho}^2)\left(z^2_{\alpha,\kappa(\tilde{\Sigma}_1)} - \frac{(\hat{b}_1-\dot{b}_1)^2}{\tilde{\mathbb{V}}[\hat{b}_1]}\right)}$$

where we are implicitly enforcing the constraint on $\dot{b}_1$ that the term under the square-root is non-negative. Thus we will find a global maximum in $\dot{b}_1$ and note that it satisfies this constraint. The first order condition for a maximum is

$$2\lambda_1\dot{b}_1 + \tilde{\rho}\frac{\tilde{\mathbb{V}}[\hat{\theta}_1]^{1/2}}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}} + \frac{\hat{b}_1 - \dot{b}_1}{\tilde{\mathbb{V}}[\hat{b}_1]}\sqrt{\frac{\tilde{\mathbb{V}}[\hat{\theta}_1](1-\tilde{\rho}^2)}{z^2_{\alpha,\kappa(\tilde{\Sigma}_1)} - \frac{(\hat{b}_1-\dot{b}_1)^2}{\tilde{\mathbb{V}}[\hat{b}_1]}}} = 0$$

which after a rearrangement and squaring of both sides yields $\frac{(\hat{b}_1-\dot{b}_1)^2}{\tilde{\mathbb{V}}[\hat{b}_1]} = (1 - a(\dot{b}))z^2_{\alpha,\kappa(\tilde{\Sigma}_1)}$. This in turn leads to the representation of $b_{1,\pm}$ given in (1). All solutions to this equation satisfies the implicit non-negativity constraint since any solution $\dot{b}$ satisfies

$$z^2_{\alpha,\kappa(\tilde{\Sigma}_1)} - \frac{(\hat{b}_1 - \dot{b}_1)^2}{\tilde{\mathbb{V}}[\hat{b}_1]} = a(\dot{b}_1) z^2_{\alpha,\kappa(\tilde{\Sigma}_1)} > 0.$$

A slightly different arrangement of the first order condition reveals the equivalent quartic condition

$$\frac{(\hat{b}_1-\dot{b}_1)^2}{\tilde{\mathbb{V}}[\hat{b}_1]} \left(1 + \left(\frac{\operatorname{sgn}(\lambda_1)\kappa(\tilde{\Sigma}_1)\dot{b}_1}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}} + \frac{\tilde{\rho}}{\sqrt{1-\tilde{\rho}^2}}\right)^2\right) = \left(\frac{\operatorname{sgn}(\lambda_1)\kappa(\tilde{\Sigma}_1)\dot{b}_1}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}} + \frac{\tilde{\rho}}{\sqrt{1-\tilde{\rho}^2}}\right)^2 z^2_{\alpha,\kappa(\tilde{\Sigma}_1)} \tag{3}$$

which has at most four solutions that are given on closed form. Thus the solution $b_{1,+}$ can be found as the maximizer of

$$\lambda_1 \dot{b}_1^2 + \hat{\theta}_1 - \tilde{\rho}\frac{\tilde{\mathbb{V}}[\hat{\theta}_1]^{1/2}}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}}(\hat{b}_1 - \dot{b}_1) + z_{\alpha,\kappa(\tilde{\Sigma}_1)}\left(\tilde{\mathbb{V}}[\hat{\theta}_q]a(\dot{b}_1)\right)^{1/2}$$

among the at most four solutions to (3). More importantly, the maximum is the upper end of $C^\theta_\alpha(\tilde{\Sigma}_1)$. Now, for the minimization problem we instead have

$$\min_{(\dot{b}_1,\dot{\theta}_1)\in \mathsf{E}_\alpha(\tilde{\Sigma}_1)} \lambda_1 \dot{b}_1^2 + \dot{\theta}_1 = \min_{\dot{b}_1} \lambda_1 \dot{b}_1^2 + \hat{\theta}_1 - \tilde{\rho}\frac{\tilde{\mathbb{V}}[\hat{\theta}_1]^{1/2}}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}}(\hat{b}_1 - \dot{b}_1) - \sqrt{\tilde{\mathbb{V}}[\hat{\theta}_1](1-\tilde{\rho}^2)\left(z^2_{\alpha,\kappa(\tilde{\Sigma}_1)} - \frac{(\hat{b}_1-\dot{b}_1)^2}{\tilde{\mathbb{V}}[\hat{b}_1]}\right)}$$

which when rearranging and squaring the first order condition again leads to (3) as a necessary condition for a minimum. Thus $b_{1,-}$ and the lower end of $C^\theta_\alpha(\tilde{\Sigma}_1)$ can be found by minimizing

$$\lambda_1 \dot{b}_1^2 + \hat{\theta}_1 - \tilde{\rho}\frac{\tilde{\mathbb{V}}[\hat{\theta}_1]^{1/2}}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}}(\hat{b}_1 - \dot{b}_1) - z_{\alpha,\kappa(\tilde{\Sigma}_1)}\left(\tilde{\mathbb{V}}[\hat{\theta}_q]a(\dot{b}_1)\right)^{1/2}$$

over the at most four solutions to (3).

3 Inference with non-existing split sample estimators

The standard error estimators considered in Lemmas 5 and 6 relies on existence of the independent and unbiased estimators $\widehat{x_i'\beta}_{-i,1}$ and $\widehat{x_i'\beta}_{-i,2}$. Here, we propose an adjustment for observations where these estimators do not exist. The adjustment ensures that one can obtain valid inference as stated in the lemma at the end of the section.

For observations where it is not possible to create $\widehat{x_i'\beta}_{-i,1}$ and $\widehat{x_i'\beta}_{-i,2}$, we construct $\widehat{x_i'\beta}_{-i,1}$ to satisfy the requirements in Lemma 6 and set $P_{i\ell,2} = 0$ for all ℓ so that $\widehat{x_i'\beta}_{-i,2} = 0$. Then we define $\mathcal{Q}_i = 1_{\{\max_\ell P_{i\ell,2}^2=0\}}$ as an indicator that $\widehat{x_i'\beta}_{-i,2}$ could not be constructed as an unbiased estimator.

Based on this we let

$$\hat{\mathbb{V}}_2[\hat{\theta}] = 4\sum_{i=1}^{n}\left(\sum\nolimits_{\ell\neq i} C_{i\ell}y_\ell\right)^2 \tilde{\sigma}_{i,2}^2 - 2\sum_{i=1}^{n}\sum_{\ell\neq i}\tilde{C}_{i\ell}\widehat{\sigma_i^2\sigma_\ell^2}_2$$

where $\tilde{\sigma}_{i,2}^2 = (1-\mathcal{Q}_i)\tilde{\sigma}_i^2 + \mathcal{Q}_i(y_i-\bar{y})^2$ and

$$\widehat{\sigma_i^2\sigma_\ell^2}_2 = \begin{cases} \hat{\sigma}_{i,-\ell}^2\cdot\hat{\sigma}_{\ell,-i}^2, & \text{if } P_{ik,-\ell}P_{\ell k,-i}=0 \text{ for all } k \text{ and } \mathcal{Q}_{i\ell}=\mathcal{Q}_{\ell i}=0 \\ \tilde{\sigma}_i^2\cdot\hat{\sigma}_{\ell,-i}^2, & \text{else if } P_{i\ell,1}+P_{i\ell,2}=0 \text{ and } \mathcal{Q}_i=\mathcal{Q}_{\ell i}=0, \\ \hat{\sigma}_{i,-\ell}^2\cdot\tilde{\sigma}_\ell^2, & \text{else if } P_{\ell i,1}+P_{\ell i,2}=0 \text{ and } \mathcal{Q}_\ell=\mathcal{Q}_{i\ell}=0, \\ \hat{\sigma}_{i,-\ell}^2\cdot(y_\ell-\bar{y})^2\cdot 1_{\{\tilde{C}_{i\ell}<0\}}, & \text{else if } \mathcal{Q}_{i\ell}=0, \\ (y_i-\bar{y})^2\cdot\hat{\sigma}_{\ell,-i}^2\cdot 1_{\{\tilde{C}_{i\ell}<0\}}, & \text{else if } \mathcal{Q}_{\ell i}=0, \\ (y_i-\bar{y})^2\cdot(y_\ell-\bar{y})^2\cdot 1_{\{\tilde{C}_{i\ell}<0\}}, & \text{otherwise} \end{cases}$$

where we let $\mathcal{Q}_{i\ell} = 1_{\{P_{i\ell,1}\neq 0\neq \mathcal{Q}_i\}}$. The definition of $\hat{\mathbb{V}}_2[\hat{\theta}]$ is such that $\hat{\mathbb{V}}_2[\hat{\theta}] = \hat{\mathbb{V}}[\hat{\theta}]$ when two independent unbiased estimators of $x_i'\beta$ can be formed for all observations, i.e., when $\mathcal{Q}_i=0$ for all i.

Similarly, we let

$$\hat{\Sigma}_{q,2} = \sum_{i=1}^{n}\begin{bmatrix} \mathsf{w}_{iq}\mathsf{w}_{iq}'\hat{\sigma}_{i,2}^2 & 2\mathsf{w}_{iq}\left(\sum\nolimits_{\ell\neq i}C_{i\ell q}y_\ell\right)\tilde{\sigma}_{i,2}^2 \\ 2\mathsf{w}_{iq}'\left(\sum\nolimits_{\ell\neq i}C_{i\ell q}y_\ell\right)\tilde{\sigma}_{i,2}^2 & 4\left(\sum\nolimits_{\ell\neq i}C_{i\ell q}y_\ell\right)^2\tilde{\sigma}_i^2 - 2\sum\nolimits_{\ell\neq i}\tilde{C}_{i\ell q}^2\widetilde{\sigma_i^2\sigma_\ell^2}_2 \end{bmatrix}$$

where $\hat{\sigma}_{i,2}^2 = (1-\mathcal{Q}_i)\hat{\sigma}_i^2 + \mathcal{Q}_i(y_i-\bar{y})^2$ and $\widetilde{\sigma_i^2\sigma_\ell^2}_2$ is defined as $\widehat{\sigma_i^2\sigma_\ell^2}_2$ but using $\tilde{C}_{i\ell q}$ instead of $\tilde{C}_{i\ell}$.

The following lemma shows that these estimators of the asymptotic variance leads to valid inference when coupled with the confidence intervals proposed in Sections 5 and 7.

Lemma 3.1. *Suppose that* $\sum_{\ell\neq i}^n P_{i\ell,1}x_\ell'\beta = x_i'\beta$, *either* $\sum_{\ell\neq i}^n P_{i\ell,2}x_\ell'\beta = x_i'\beta$ *or* $\max_\ell P_{i\ell,2}^2=0$, $P_{i\ell,1}P_{i\ell,2}=0$ *for all* ℓ, *and* $\lambda_{\max}(P_sP_s') = O(1)$ *where* $P_s = (P_{i\ell,s})_{i,\ell}$.

1. *If the conditions of Theorem 2 hold, then* $\liminf_{n\to\infty}\mathbb{P}\left(\theta\in\left[\hat{\theta}\pm z_\alpha\hat{\mathbb{V}}_2[\hat{\theta}]^{1/2}\right]\right)\geq 1-\alpha$.
2. *If the conditions of Theorem 3 hold, then* $\liminf_{n\to\infty}\mathbb{P}\left(\theta\in C_\alpha^\theta(\hat{\Sigma}_{q,2})\right)\geq 1-\alpha$.

Proof of Lemma 3.1. As in the proof of Lemma 5 it suffices for the first claim to show that $\hat{\mathbb{V}}_2[\hat{\theta}]$ has a positive bias in large samples and that $\hat{\mathbb{V}}_2[\hat{\theta}] - \mathbb{E}[\hat{\mathbb{V}}_2[\hat{\theta}]]$ is $o_p(\mathbb{V}[\hat{\theta}])$. The second claim involves no new arguments relative to the proof of Lemma 5 and is therefore omitted. Thus we briefly report the positive bias in $\hat{\mathbb{V}}_2[\hat{\theta}]$.

We have that

$$\begin{aligned}
\mathbb{E}\left[\hat{\mathbb{V}}_2[\hat{\theta}]\right] = \mathbb{V}[\hat{\theta}] &+ 4\sum_{i:\mathcal{Q}_i=1}\left(\sum\nolimits_{\ell\neq i} C_{i\ell}x_\ell'\beta\right)^2((x_i-\bar{x})'\beta)^2 \\
&+ 2\sum_{(i,\ell)\in\mathcal{B}_1}\tilde{C}_{i\ell}\sigma_i^2\left(\sigma_\ell^2 1_{\{\tilde{C}_{i\ell}>0\}} + ((x_\ell-\bar{x})'\beta)^2 1_{\{\tilde{C}_{i\ell}<0\}}\right) \\
&+ 2\sum_{(i,\ell)\in\mathcal{B}_2}\tilde{C}_{i\ell}\sigma_\ell^2\left(\sigma_i^2 1_{\{\tilde{C}_{i\ell}>0\}} + ((x_i-\bar{x})'\beta)^2 1_{\{\tilde{C}_{i\ell}<0\}}\right) \\
&+ 2\sum_{(i,\ell)\in\mathcal{B}_3}\tilde{C}_{i\ell}\left(\sigma_i^2\sigma_\ell^2 1_{\{\tilde{C}_{i\ell}>0\}} + \left(2\sigma_i^2((x_\ell-\bar{x})'\beta)^2 + ((x_i-\bar{x})'\beta(x_\ell-\bar{x})'\beta)^2\right)1_{\{\tilde{C}_{i\ell}<0\}}\right) \\
&+ O\left(\mathbb{V}[\hat{\theta}]/n\right)
\end{aligned}$$

where the remainder stems from estimation of $\bar{y}$ and $\mathcal{B}_1,\ \mathcal{B}_2,\ \mathcal{B}_3$ refers to pairs of observations that fall in each of the three last cases in the definition of $\widehat{\sigma_i^2\sigma_\ell^2}_2$.

The proof of the second claim contains two main parts. One part is to establish that the bias $\hat{\Sigma}_{q,2}$ is positive semidefinite in large samples, and that $\mathbb{E}[\hat{\Sigma}_{q,2}]^{-1}\hat{\Sigma}_{q,2} - I_{q+1}$ is $o_p(1)$. These arguments are only sketched as they are analogues to those presented in the proof of Lemma 5 and the first part of this lemma. The other part is to show that this positive semidefinite asymptotic bias in the variance estimator does not alter the validity of the confidence interval based on it.

Validity First, we let QDQ' be the spectral decomposition of $\mathbb{E}[\hat{\Sigma}_{q,2}]^{-1/2}\Sigma_q\mathbb{E}[\hat{\Sigma}_{q,2}]^{-1/2}$. Here, $\mathsf{QQ}' = \mathsf{Q}'\mathsf{Q} = I_{q+1}$ and all diagonal entries in the diagonal matrix D belongs to $(0,1]$ in large samples. Now,

$$\mathbb{P}\left(\theta \in C_\alpha^\theta(\hat{\Sigma}_{q,2})\right) = \mathbb{P}\left(\min_{(\mathsf{b}_q',\theta_q)':g(\mathsf{b}_q,\theta_q,\theta)=0}\begin{pmatrix}\hat{\mathsf{b}}_q-\mathsf{b}_q\\ \hat{\theta}_q-\theta_q\end{pmatrix}'\mathbb{E}[\hat{\Sigma}_{q,2}]^{-1}\begin{pmatrix}\hat{\mathsf{b}}_q-\mathsf{b}_q\\ \hat{\theta}_q-\theta_q\end{pmatrix} \le z^2_{\alpha,\kappa(\mathbb{E}[\hat{\Sigma}_{q,2}])}\right) + o(1)$$

where the minimum distance statistic above satisfies

$$\min_{(\mathsf{b}_q',\theta_q)':g(\mathsf{b}_q,\theta_q,\theta)=0}\begin{pmatrix}\hat{\mathsf{b}}_q-\mathsf{b}_q\\ \hat{\theta}_q-\theta_q\end{pmatrix}'\mathbb{E}[\hat{\Sigma}_{q,2}]^{-1}\begin{pmatrix}\hat{\mathsf{b}}_q-\mathsf{b}_q\\ \hat{\theta}_q-\theta_q\end{pmatrix} = \min_{x\in S_2}(\xi-x)'(\xi-x)$$

where $S_2 = \{x : x = \mathsf{Q}'\mathbb{E}[\hat{\Sigma}_{q,2}]^{-1/2}\left((\mathsf{b}_q',\theta_q)' - \mathbb{E}[(\hat{\mathsf{b}}_q',\hat{\theta}_q)']\right), g(\mathsf{b}_q,\theta_q,\theta)=0\}$ and the random vector $\xi = \mathsf{Q}'\mathbb{E}[\hat{\Sigma}_{q,2}]^{-1/2}\left((\hat{\mathsf{b}}_q',\hat{\theta}_q)' - \mathbb{E}[(\hat{\mathsf{b}}_q',\hat{\theta}_q)']\right)$ has the property that $\mathsf{D}^{-1/2}\xi \xrightarrow{d} \mathcal{N}(0,I_{q+1})$. From the geometric consideration in Andrews and Mikusheva (2016) it follows that S_2 has curvature of $\kappa(\mathbb{E}[\hat{\varSigma}_{q,2}])$ since curvature is invariant to rotations. Furthermore,

$$\begin{aligned}\min_{x\in S_2}(\xi-x)'(\xi-x) &\leq \rho^2\left(\|\xi_{-1}\|,|\xi_1|,\kappa(\mathbb{E}[\hat{\varSigma}_{q,2}])\right)\\ &\leq \rho^2\left(\|(\mathsf{D}^{-1/2}\xi)_{-1}\|,|(\mathsf{D}^{-1/2}\xi)_1|,\kappa(\mathbb{E}[\hat{\varSigma}_{q,2}])\right)\end{aligned}$$

where $\xi=(\xi_1,\xi_{-1}')'$ and $\mathsf{D}^{-1/2}\xi=((\mathsf{D}^{-1/2}\xi)_1,(\mathsf{D}^{-1/2}\xi)_{-1}')$ and the first inequality follows from the proof of Theorem 1 in Andrews and Mikusheva (2016). Thus

$$\begin{aligned}\liminf_{n\to\infty}\mathbb{P}\left(\theta\in C_\alpha^\theta(\hat{\varSigma}_{q,2})\right) &= \liminf_{n\to\infty}\mathbb{P}\left(\min_{x\in S_2}(\xi-x)'(\xi-x)\leq z^2_{\alpha,\kappa(\mathbb{E}[\hat{\varSigma}_{q,2}])}\right)\\ &\geq \liminf_{n\to\infty}\mathbb{P}\left(\rho^2\left(\chi_q,\chi_1,\kappa(\mathbb{E}[\hat{\varSigma}_{q,2}])\right)\leq z^2_{\alpha,\kappa(\mathbb{E}[\hat{\varSigma}_{q,2}])}\right)=1-\alpha\end{aligned}$$

since $(\|\xi_{-1}\|,|\xi_1|)\xrightarrow{d}(\chi_q,\chi_1)$.

Bias and variability in $\hat{\varSigma}_{q,2}$ We finish by reporting the positive semidefinite bias in $\hat{\varSigma}_{q,2}$. We have that

$$\mathbb{E}\left[\hat{\varSigma}_{q,2}\right]=\varSigma_q+\sum_{i:\mathcal{Q}_i=1}\sigma_i^2\begin{pmatrix}\mathsf{w}_{iq}\\ 2\sum_{\ell\neq i}C_{i\ell}x_\ell'\beta\end{pmatrix}\begin{pmatrix}\mathsf{w}_{iq}\\ 2\sum_{\ell\neq i}C_{i\ell}x_\ell'\beta\end{pmatrix}'+\begin{bmatrix}0&0\\0&\mathcal{B}\end{bmatrix}+O\left(\mathbb{V}[\hat{\theta}]/n\right)$$

where

$$\begin{aligned}\mathcal{B}=2&\sum_{(i,\ell)\in\mathcal{B}_1}\tilde{C}_{i\ell q}\sigma_i^2\left(\sigma_\ell^2 1_{\{\tilde{C}_{i\ell q}>0\}}+((x_\ell-\bar{x})'\beta)^2 1_{\{\tilde{C}_{i\ell q}<0\}}\right)\\ &+2\sum_{(i,\ell)\in\mathcal{B}_2}\tilde{C}_{i\ell q}\sigma_\ell^2\left(\sigma_i^2 1_{\{\tilde{C}_{i\ell q}>0\}}+((x_i-\bar{x})'\beta)^2 1_{\{\tilde{C}_{i\ell q}<0\}}\right)\\ &+2\sum_{(i,\ell)\in\mathcal{B}_3}\tilde{C}_{i\ell q}\left(\sigma_i^2\sigma_\ell^2 1_{\{\tilde{C}_{i\ell q}>0\}}+\left(2\sigma_i^2((x_\ell-\bar{x})'\beta)^2+((x_i-\bar{x})'\beta(x_\ell-\bar{x})'\beta)^2\right)1_{\{\tilde{C}_{i\ell q}<0\}}\right)\end{aligned}$$

for $\mathcal{B}_1,\ \mathcal{B}_2,\ \mathcal{B}_3$ referring to pairs of observations that fall in each of the three last cases in the definition of $\widetilde{\sigma_i^2\sigma_\ell^2}_2$. □

4 Verifying Conditions

This section fills in details omitted from the discussion of Examples 1–3 in Sections 2 and 8.

Example 1. We first derive the representations of $\hat{\sigma}_\alpha^2$ given in section 2. When there are no common regressors, the representation in (4) follows from $B_{ii} = \frac{1}{nT_{g(i)}}\left(1 - T_{g(i)}/n\right)$ and

$$\hat{\sigma}_g^2 = \tfrac{1}{T_g}\sum\nolimits_{t=1}^{T_g} y_{gt}\left(y_{gt} - \tfrac{1}{T_g-1}\sum\nolimits_{s\neq t} y_{gs}\right) = \tfrac{1}{T_g}\sum\nolimits_{i:g(i)=g}\hat{\sigma}_i^2$$

which yields that

$$\sum\nolimits_{i=1}^{n} B_{ii}\hat{\sigma}_i^2 = \tfrac{1}{n}\sum\nolimits_{g=1}^{N}\left(1 - \tfrac{T_g}{n}\right)\hat{\sigma}_g^2.$$

With common regressors, it follows from the formula for block inversion of matrices that

$$\begin{aligned}\tilde{X}' &= AS_{xx}^{-1}(D,X)' = \tfrac{1}{n}\left(\left(I - (I-P_D)X\left(X'(I-P_D)X'\right)^{-1}X'\right)\left(D - \bar{d}\mathbf{1}_n\right), 0\right)' \\ &= \tfrac{1}{n}\left(D - \bar{d}\mathbf{1}_n - (I-P_D)X\hat{\Gamma}, 0\right)'\end{aligned}$$

where $D = (d_1, \dots, d_n)'$, $X = (x_{g(1)t(1)}, \dots, x_{g(n)t(n)})'$, $P_D = DS_{dd}^{-1}D'$, $\mathbf{1}_n = (1, \dots, 1)'$, and $S_{dd} = D'D$. Thus it follows that

$$\tilde{x}_i = \tfrac{1}{n}\begin{pmatrix} d_i - \bar{d} - \hat{\Gamma}'(x_{g(i)t(i)} - \bar{x}_{g(i)}) \\ 0 \end{pmatrix}.$$

The no common regressors claims are immediate. With common regressors we have

$$P_{i\ell} = T_{g(i)}^{-1}\mathbf{1}_{\{g(i)=g(\ell)\}} + n^{-1}(x_{g(i)t(i)} - \bar{x}_{g(i)})'W^{-1}(x_{g(\ell)t(\ell)} - \bar{x}_{g(\ell)}) = T_{g(i)}^{-1}\mathbf{1}_{\{i=\ell\}} + O(n^{-1})$$

where $W = \frac{1}{n}\sum_{g=1}^{N}\sum_{t=1}^{T}(x_{gt} - \bar{x}_g)(x_{gt} - \bar{x}_g)'$ so $P_{ii} \leq C < 1$ in large samples. The eigenvalues of $\tilde{A}$ are equal to the eigenvalues of

$$\tfrac{1}{n}\left(I_N - nS_{dd}^{-1/2}\bar{d}\bar{d}'S_{dd}^{-1/2}\right)\left(I_N + \tfrac{1}{n}S_{dd}^{1/2}D'XW^{-1}X'DS_{dd}^{-1/2}\right)$$

which in turn satisfies that $\frac{c_1}{n} \leq \lambda_\ell \leq \frac{c_2}{n}$ for $\ell = 1, \dots, N-1$ and $c_2 \geq c_1 > 0$ not depending on n. $w_i'w_i = O(P_{ii})$ so Theorem 1 applies when N is fixed and $\min_g T_g \to \infty$. Finally,

$$\begin{aligned}\max_i \mathbb{V}[\hat{\theta}]^{-1}(\tilde{x}_i'\beta)^2 &= N^{-1}O\left(\max\nolimits_{g,t}\alpha_g^2 + \|x_{gt}\|^2\tfrac{1}{n}\sum\nolimits_{i=1}^{n}\|x_{g(i)t(i)}\|^2\sigma_\alpha^2\right) \\ \max_i \mathbb{V}[\hat{\theta}]^{-1}(\check{x}_i'\beta)^2 &= N^{-1}O\left(\max\nolimits_{i,j}(x_j'\beta)^2\left(\sum\nolimits_{\ell=1}^{n}|M_{i\ell}|\right)^2\right)\end{aligned}$$

and $\sum_{\ell=1}^{n}|M_{i\ell}| = O(1)$ so Theorem 2 applies when $N \to \infty$.

Example 2. $\tilde{A}$ is diagonal with N diagonal entries of $\frac{1}{n}\frac{T_g}{S_{zz,g}}$, so $\lambda_g = \frac{1}{n}\frac{T_g}{S_{zz,g}}$ for $g =$

$1, \ldots, N$. $\mathrm{trace}(\tilde{A}^2) \leq \frac{\lambda_1}{\min_g S_{zz,g}} \frac{1}{n} \sum_{g=1}^N T_g = O(\lambda_1)$. $\max_i w_i' w_i = \max_{g,t} \frac{(z_{gt} - \bar{z}_g)^2}{S_{zz,g}} = o(1)$ when $\min_g S_{zz,g} \to \infty$. Furthermore, $\mathbb{V}[\hat{\theta}]^{-1} = O(\frac{n^2}{N})$, so

$$\mathbb{V}[\hat{\theta}]^{-1} \max_i (\tilde{x}_i' \beta)^2 = O\left(\max_{g,t} \frac{z_{gt}^2 \delta_g^2}{N S_{zz,g}}\right) = o(1),$$

and $M_{i\ell} = 0$ if $g(i) \neq g(\ell)$ so

$$\mathbb{V}[\hat{\theta}]^{-1} \max_i (\check{x}_i' \beta)^2 = n N^{-1/2} O\left(\max_g \left(\sum_{i:g(i)=g} B_{ii}\right)^2\right) = O\left(\max_g \left(\frac{T_g}{\sqrt{N} S_{xx,g}}\right)^2\right) = o(1)$$

both under the condition that $N \to \infty$ and $\frac{\sqrt{N} S_{xx,1}}{T_1} \to \infty$. Used above:

$$P_{i\ell} = T_{g(i)}^{-1} \mathbf{1}_{\{g(i)=g(\ell)\}} + \frac{(z_{g(i)t(i)} - \bar{z}_{g(i)})(z_{g(i)t(\ell)} - \bar{z}_{g(i)})}{S_{zz,g(i)}} \mathbf{1}_{\{g(i)=g(\ell)\}} \qquad B_{ii} = \frac{1}{n} \frac{z_{g(i)t(i)} - \bar{z}_{g(i)}}{S_{zz,g(i)}} \frac{T_{g(i)}}{S_{zz,g(i)}}.$$

Finally,

$$\begin{aligned} \max_i \mathsf{w}_{iq}' \mathsf{w}_{iq} &= \max_t \frac{(z_{1t} - \bar{z}_1)^2}{S_{zz,1}} = o(1) \\ \mathbb{V}[\hat{\theta}_q]^{-1} \max_i (\tilde{x}_{iq}' \beta)^2 &= O\left(\max_{g \geq 2, t} \frac{z_{gt}^2 \delta_g^2}{N S_{zz,g}}\right) = o(1), \\ \mathbb{V}[\hat{\theta}_q]^{-1} \max_i (\check{x}_{iq}' \beta)^2 &= O\left(\max_{g \geq 2} \left(\frac{T_g}{\sqrt{N} S_{xx,g}}\right)^2\right) = o(1) \end{aligned}$$

under the conditions that $\frac{\sqrt{N}}{T_2} S_{zz,2} \to \infty$ and $S_{zz,1} \to \infty$. Thus, Theorem 3 applies when $\frac{\sqrt{N}}{T_1} S_{zz,1} = O(1)$.

Example 3. Let $\dot{f}_i = (\mathbf{1}_{\{j(g,t)=0\}}, f_i')' = (\mathbf{1}_{\{j(g,t)=0\}}, \mathbf{1}_{\{j(g,t)=1\}}, \ldots, \mathbf{1}_{\{j(g,t)=J\}})'$ and define the following partial design matrices with and without dropping ψ_0 from the model:

$$S_{ff} = \sum_{i=1}^n f_i f_i', \qquad S_{\dot{f}\dot{f}} = \sum_{i=1}^n \dot{f}_i \dot{f}_i', \qquad S_{\Delta f \Delta f} = \sum_{g=1}^N \Delta f_g \Delta f_g', \qquad S_{\Delta \dot{f} \Delta \dot{f}} = \sum_{g=1}^N \Delta \dot{f}_g \Delta \dot{f}_g',$$

where $\Delta \dot{f}_g = \dot{f}_{i(g,2)} - \dot{f}_{i(g,1)}$. Letting $\dot{D}$ be a diagonal matrix that holds the diagonal of $S_{\Delta \dot{f} \Delta \dot{f}}$ we have that

$$E = \dot{D} S_{\dot{f}\dot{f}}^{-1} \quad \text{and} \quad \mathcal{L} = \dot{D}^{-1/2} S_{\Delta \dot{f} \Delta \dot{f}} \dot{D}^{-1/2}.$$

$S_{\Delta \dot{f} \Delta \dot{f}}$ is rank deficient with $S_{\Delta \dot{f} \Delta \dot{f}} \mathbf{1}_{J+1} = 0$ from which it follows that the non-zero eigenvalues of $E^{1/2} \mathcal{L} E^{1/2}$ (which are the non-zero eigenvalues of $S_{\dot{f}\dot{f}}^{-1} S_{\Delta \dot{f} \Delta \dot{f}}$) are also the eigenvalues of $S_{\Delta f \Delta f}(S_{ff}^{-1} + \frac{\mathbf{1}_J \mathbf{1}_J'}{S_{\dot{f}\dot{f},11}})$. Finally, from the Woodbury formula we have that A_{ff} is invertible with

$$A_{ff}^{-1} = n(S_{ff} - n\bar{f}\bar{f}')^{-1} = n\left(S_{ff}^{-1} + n\frac{S_{ff}^{-1}\bar{f}\bar{f}'S_{ff}^{-1}}{1-n\bar{f}'S_{ff}^{-1}\bar{f}}\right) = n\left(S_{ff}^{-1} + \frac{\mathbf{1}_J \mathbf{1}_J'}{S_{\dot{f}\dot{f},11}}\right),$$

so

$$\lambda_\ell = \lambda_\ell(A_{ff}S_{\Delta f \Delta f}^{-1}) = \frac{1}{\lambda_{J+1-\ell}(S_{\Delta f \Delta f}A_{ff}^{-1})} = \frac{1}{n\lambda_{J+1-\ell}(E^{1/2}\mathcal{L}E^{1/2})}.$$

With $E_{jj} = 1$ for all j, we have that

$$\frac{\lambda_1^2}{\sum_{\ell=1}^J \lambda_\ell^2} = \frac{\dot{\lambda}_J^{-2}}{\sum_{\ell=1}^J \dot{\lambda}_\ell^{-2}} \leq \frac{4}{(\sqrt{J}\dot{\lambda}_J)^2}$$

since $\dot{\lambda}_\ell \leq 2$ (Chung, 1997, Lemma 1.7). An algebraic definition of Cheeger's constant $\mathcal{C}$ is

$$\mathcal{C} = \min_{X \subseteq \{0,\ldots,J\}: \sum_{j\in X} \dot{D}_{jj} \leq \frac{1}{2}\sum_{j=0}^J \dot{D}_{jj}} \frac{-\sum_{j\in X}\sum_{k\notin X} S_{\Delta\dot{f}\Delta\dot{f},jk}}{\sum_{j\in X}\dot{D}_{jj}}$$

and it follows from the Cheeger inequality $\dot{\lambda}_J \geq 1 - \sqrt{1-\mathcal{C}^2}$ (Chung, 1997, Theorem 2.3) that $\sqrt{J}\dot{\lambda}_J \to \infty$ if $\sqrt{J}\mathcal{C} \to \infty$.

For the stochastic block model we consider J odd and order the firms so that the first $(J+1)/2$ firms belongs to the first block, and the remaining firms belong to the second block. We assume that $\Delta\dot{f}_g$ is generated *i.i.d.* across g according to

$$\Delta\dot{f} = \mathrm{W}(1-\mathrm{D}) + \mathrm{BD}$$

where $(\mathrm{W}, \mathrm{B}, \mathrm{D})$ are mutually independent, $P(\mathrm{D} = 1) = 1 - P(\mathrm{D} = 0) = p_b \leq \frac{1}{2}$, W is uniformly distributed on $\{v \in \mathbb{R}^{J+1} : v'\mathbf{1}_{J+1} = 0, v'v = 2, \max_j v_j = 1, v'c = 0\}$, and B is uniformly distributed on $\{v \in \mathbb{R}^{J+1} : v'\mathbf{1}_{J+1} = 0, v'v = 2, \max_j v_j = 1, (v'c)^2 = 4\}$ for $c = (\mathbf{1}_{(J+1)/2}', -\mathbf{1}_{(J+1)/2}')'$. In this model $E_{jj} = 1$ for all j. The following lemma characterizes the large sample behavior of $S_{\Delta\dot{f}\Delta\dot{f}}$ and $\mathcal{L}$. Based on this lemma it is relatively straightforward (but tedious) to verify the high-level conditions imposed in the paper.

Lemma 4.1. *Suppose that* $\frac{\log(J)}{np_b} + \frac{J\log(J)}{n} \to 0$ *as* $n \to \infty$ *and* $J \to \infty$. *Then*

$$\left\|\underline{\mathcal{L}}^{\dagger}\frac{J+1}{n}S_{\Delta\dot{f}\Delta\dot{f}} - I_{J+1} + \frac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1}\right\| = o_p(1) \quad \text{and} \quad \left\|\underline{\mathcal{L}}^{\dagger}\mathcal{L} - I_{J+1} + \frac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1}\right\| = o_p(1)$$

where $\underline{\mathcal{L}} = I_{J+1} - \frac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} - (1-2p_b)\frac{cc'}{J+1}$ *and* $\|\cdot\|$ *returns the largest singular value of its* *argument. Additionally,* $\max_\ell \underline{\dot\lambda}_\ell^{-1}\left|\dot\lambda_\ell - \underline{\dot\lambda}_\ell\right| = o_p(1)$ *where* $\underline{\dot\lambda}_1 \geq \cdots \geq \underline{\dot\lambda}_J$ *are the non-zero eigenvalues of* $\underline{\mathcal{L}}^\dagger$*.*

Proof. First note that

$$\tfrac{J+1}{n}\mathbb{E}[S_{\Delta\dot f\Delta\dot f}] - \underline{\mathcal{L}} = \tfrac{2+2p_b}{J-1}\left(I_{J+1} - \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} - \tfrac{cc'}{J+1}\right) + \tfrac{4p_b}{J-1}\tfrac{cc'}{J+1},$$

and $\underline{\mathcal{L}}^\dagger = I_{J+1} - \frac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} - \left(1 - \frac{1}{2p_b}\right)\frac{cc'}{J+1}$, so

$$\left\|\underline{\mathcal{L}}^\dagger \tfrac{J+1}{n}\mathbb{E}[S_{\Delta\dot f\Delta\dot f}] - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1}\right\| = \left\|\tfrac{2+2p_b}{J-1}\left(I_{J+1} - \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} - \tfrac{cc'}{J+1}\right) + \tfrac{2}{J-1}\tfrac{cc'}{J+1}\right\| = \tfrac{2+2p_b}{J-1}$$

Therefore, we can instead show that $\|S\| = o_p(1)$ for the zero mean random matrix

$$S = (\underline{\mathcal{L}}^\dagger)^{1/2}\tfrac{J+1}{n}\left(S_{\Delta\dot f\Delta\dot f} - \mathbb{E}[S_{\Delta\dot f\Delta\dot f}]\right)(\underline{\mathcal{L}}^\dagger)^{1/2} = \sum\nolimits_{g=1}^N s_g s_g' - \mathbb{E}[s_g s_g']$$

where $s_g = \sqrt{\frac{J+1}{n}}\Delta\dot f_g - \frac{\sqrt{2p_b}-1}{\sqrt{2p_b n}}\Delta\dot f_g' c\frac{c}{\sqrt{J+1}}$. Now since

$$s_g's_g = O\left(\tfrac{J}{n} + \tfrac{1}{np_b}\right) \quad \text{and} \quad \left\|\sum\nolimits_{g=1}^N \mathbb{E}[s_g s_g' s_g s_g']\right\| = O\left(\tfrac{J}{n} + \tfrac{1}{np_b}\right)$$

(Oliveira, 2009, Corollary 7.1) yields that $\mathbb{P}(\|S\| \geq t) \leq 2(J+1)exp\left(-t^2(\frac{J}{n} + \frac{1}{np_b})/(8c + 4ct)\right)$ for some constant c not depending on n. Letting $t \propto \sqrt{\frac{\log(J/\delta_n)}{np_b} + \frac{J\log(J/\delta_n)}{n}}$ for δ_n that approaches zero slowly enough that $\frac{\log(J/\delta_n)}{np_b} + \frac{J\log(J/\delta_n)}{n} \to 0$ yields the conclusion that $\|S\| = o_p(1)$.

Since $\mathcal{L} = \dot D^{-1/2} S_{\Delta\dot f\Delta\dot f}\dot D^{-1/2}$ the second conclusion follows from the first if $\|\frac{J+1}{n}\dot D - I_{J+1}\| = o_p(1)$. We have $\frac{J+1}{n}\mathbb{E}[\dot D] = I_{J+1}$ and $\frac{J+1}{n}\dot D_{jj} = \frac{J+1}{n}\sum_{g=1}^N(\Delta\dot f_g' e_j)^2$ where e_j is the j-th basis vector in $\mathbb{R}^{J+1}$ and $\mathbb{P}((\Delta\dot f_g' e_j)^2 = 1) = 1 - \mathbb{P}((\Delta\dot f_g' e_j)^2 = 0) = \frac{2}{J+1}$. Thus it follows from $\mathbb{V}(\frac{J+1}{n}\dot D_{jj}) \leq 2\frac{J+1}{n}$ and standard exponential inequalities that $\|\frac{J+1}{n}\dot D - I_{J+1}\| = \max_j|\frac{J+1}{n}\dot D_{jj} - 1| = o_p(1)$ since $\frac{J\log(J)}{n} \to 0$.

Finally, we note that $\left\|\underline{\mathcal{L}}^\dagger\mathcal{L} - I_{J+1} + \frac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1}\right\| \leq \epsilon$ implies

$$v'\underline{\mathcal{L}}v(1-\epsilon) \leq v'\mathcal{L}v \leq v'\underline{\mathcal{L}}v(1+\epsilon)$$

which together with the Courant-Fischer min-max principle yields $(1-\epsilon) \leq \frac{\dot\lambda_j}{\underline{\dot\lambda}_j} \leq (1+\epsilon)$. □

Next, we will verify the high-level conditions of the paper in a model that uses $\frac{n}{J+1}\underline{\mathcal{L}}$ in place of $S_{\Delta\dot{f}\Delta\dot{f}}$ and $\frac{1}{n}\underline{\mathcal{L}}^{\dagger}$ in place of $\tilde{A}$ and $\frac{n}{J+1}I_{J+1}$ in place of $\dot{D}$. Using an underscore to denote objects from this model we have

$$\begin{aligned}
\max_g \underline{P}_{gg} &= \max_g \tfrac{J+1}{n}\Delta\dot{f}_g'\underline{\mathcal{L}}^{\dagger}\Delta\dot{f}_g = 2\tfrac{J+1}{n} + 2\tfrac{(1-2p_b)}{np_b} = o(1),\\
\text{trace}(\underline{\tilde{A}}^2) &= \tfrac{\text{trace}((\underline{\mathcal{L}}^{\dagger})^2)}{n^2} = \tfrac{J-1}{n^2} + \tfrac{1}{4(np_b)^2} = o(1),\\
\tfrac{\underline{\lambda}_1^2}{\sum_{\ell=1}^J \underline{\lambda}_\ell^2} &= \tfrac{1}{\dot{\underline{\lambda}}_J^2\text{trace}((\underline{\mathcal{L}}^{\dagger})^2)} = \tfrac{1}{(J-1)4p_b^2+1}
\end{aligned}$$

which is $o(1)$ if and only if $\sqrt{J}p_b \to \infty$, and $\frac{\underline{\lambda}_2^2}{\sum_{\ell=1}^J \underline{\lambda}_\ell^2} \leq \frac{1}{J}$. Furthermore,

$$\begin{aligned}
\max_g \underline{\mathsf{w}}_{g1}^2 &= n^{-1}\max_g \left(c'(\underline{\mathcal{L}}^{\dagger})^{1/2}\Delta\dot{f}_g\right)^2 = \left(\tfrac{2}{\sqrt{2p_b n}}\right)^2 = \tfrac{2}{np_n} = o(1),\\
\max_g(\underline{\tilde{x}}_g'\beta)^2 &= n^{-2}\max_g\left(\psi'\underline{\mathcal{L}}^{\dagger}\Delta\dot{f}_g\right)^2 \leq 2n^{-2}\left[\max_g(\Delta\dot{f}_g'\psi)^2 + \left(1-\tfrac{1}{2p_b}\right)^2(\bar{\psi}_{cl,1}-\bar{\psi}_{cl,2})^2\right]\\
&= O\left(n^{-2}+(np_b)^{-2}\right)
\end{aligned}$$

which is $o\left(\mathbb{V}[\hat{\theta}]\right)$ if $\sqrt{J}p_b \to \infty$ as $\text{trace}(\underline{\tilde{A}}^2) = O(\mathbb{V}[\hat{\theta}])$ and

$$\max_g(\underline{\tilde{x}}_{g1}'\beta)^2 = n^{-2}\max_g\left(\psi'\Delta\dot{f}_g\right)^2 = O\left(n^{-2}\right) = o\left(\mathbb{V}[\hat{\theta}]\right).$$

Finally,

$$\max_g(\check{x}_g'\beta)^2 = O\left(\sum_{g=1}^N B_{gg}^2\right) = O\left(\max_g B_{gg}\text{trace}(\tilde{A})\right)$$

where

$$\begin{aligned}
\max_g \underline{B}_{gg} &= \max_g \Delta\dot{f}_g'\tfrac{J+1}{n^2}(\underline{\mathcal{L}}^{\dagger})^2\Delta\dot{f}_g = 2\tfrac{J+1}{n^2} + \tfrac{1-4p_b^2}{(np_b)^2} = O\left(\text{trace}(\underline{\tilde{A}}^2)\right)\\
\text{trace}(\underline{\tilde{A}}) &= \tfrac{J-1}{n} + \tfrac{1}{2p_b n} = o(1)
\end{aligned}$$

so $\max_g \underline{B}_{gg}\text{trace}(\underline{\tilde{A}}) = O(\text{trace}(\underline{\tilde{A}}^2))o(1)$.

Finally, we use the previous lemma to transfer the above results to their relevant sample analogues.

$$\begin{aligned}
\max_g |P_{gg} - \underline{P}_{gg}| &= \max_g |\Delta \dot{f}_g'(S^\dagger_{\Delta \dot{f} \Delta \dot{f}} - \tfrac{J+1}{n}\underline{\mathcal{L}}^\dagger)\Delta \dot{f}_g| \\
&= \tfrac{J+1}{n} \max_g \left| \Delta \dot{f}_g'(\underline{\mathcal{L}}^\dagger)^{1/2} \left(\underline{\mathcal{L}}^{1/2} \tfrac{n}{J+1} S^\dagger_{\Delta \dot{f} \Delta \dot{f}} \underline{\mathcal{L}}^{1/2} - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} \right) (\underline{\mathcal{L}}^\dagger)^{1/2} \Delta \dot{f}_g \right| \\
&= O\left(\left\| \underline{\mathcal{L}}^\dagger \tfrac{J+1}{n} S_{\Delta \dot{f} \Delta \dot{f}} - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} \right\| \right) \max_g \underline{P}_{gg} = o\left(\max_g \underline{P}_{gg} \right)
\end{aligned}$$

$$\left| \mathrm{trace}(\tilde{A}^2 - \underline{\tilde{A}}^2) \right| = \left| \sum_{\ell=1}^J \tfrac{1}{n^2 \dot{\lambda}_\ell^2} - \tfrac{1}{n^2 \underline{\dot{\lambda}}_\ell^2} \right| = \mathrm{trace}(\underline{\tilde{A}}^2) O\left(\max_\ell \left| \tfrac{\dot{\lambda}_\ell - \underline{\dot{\lambda}}_\ell}{\underline{\dot{\lambda}}_\ell} \right| \right) = o_p\left(\mathrm{trace}(\underline{\tilde{A}}^2) \right)$$

$$\left| \tfrac{\lambda_1^2}{\sum_{\ell=1}^J \lambda_\ell^2} - \tfrac{\underline{\lambda}_1^2}{\sum_{\ell=1}^J \underline{\lambda}_\ell^2} \right| = \tfrac{\underline{\lambda}_1^2}{\sum_{\ell=1}^J \underline{\lambda}_\ell^2} O\left(\tfrac{|\dot{\lambda}_J - \underline{\dot{\lambda}}_J|}{\underline{\dot{\lambda}}_J} + \tfrac{\left|\mathrm{trace}(\underline{\tilde{A}}^2 - \tilde{A}^2)\right|}{\mathrm{trace}(\underline{\tilde{A}}^2)} \right) = o_p(1)$$

with a similar argument applying to $\frac{\lambda_2^2}{\sum_{\ell=1}^J \lambda_\ell^2} - \frac{\underline{\lambda}_2^2}{\sum_{\ell=1}^J \underline{\lambda}_\ell^2}$. Furthermore,

$$\max_g \mathsf{w}_{g1}^2 = \max_g \left(\Delta \dot{f}_g (\tfrac{J+1}{n} \underline{\mathcal{L}}^\dagger)^{1/2} (\underline{\mathcal{L}} \tfrac{n}{J+1} S^\dagger_{\Delta \dot{f} \Delta \dot{f}})^{1/2} q_1 \right)^2 \leq \|(\underline{\mathcal{L}} \tfrac{n}{J+1} S^\dagger_{\Delta \dot{f} \Delta \dot{f}})^{1/2}\| \max_g \underline{P}_{gg} = o_p(1)$$

and $\max_g |(\tilde{x}_g'\beta)^2 - (\underline{\tilde{x}}_g'\beta)^2| = o_p(\mathrm{trace}(\underline{\tilde{A}}^2))$ since

$$\begin{aligned}
\max_g (\tilde{x}_g'\beta - \underline{\tilde{x}}_g'\beta)^2 &= \tfrac{J+1}{n^2} \max_g \left(\Delta \dot{f}_g' \underline{\mathcal{L}}^\dagger \left(\underline{\mathcal{L}} S_{\Delta \dot{f} \Delta \dot{f}} \dot{D} - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} \right) \tfrac{\psi}{\sqrt{J+1}} \right)^2 \\
&\leq \left\| \underline{\mathcal{L}} S_{\Delta \dot{f} \Delta \dot{f}} \dot{D} - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} \right\| \max_g \underline{B}_{gg} \tfrac{\|\psi\|^2}{J+1} = o_p(\mathrm{trace}(\underline{\tilde{A}}^2))
\end{aligned}$$

and this also handles $\max_i |(\tilde{x}_{g1}'\beta)^2 - (\underline{\tilde{x}}_{g1}'\beta)^2| = o_p(1)$ as the previous result does not depend on the behavior of $\sqrt{J}p_b$. Finally,

$$\begin{aligned}
\max_g |B_{gg} - \underline{B}_{gg}| &= \tfrac{J+1}{n^2} \max_g \left| \Delta \dot{f}_g' \underline{\mathcal{L}}^\dagger \left(\tfrac{n}{J+1} \underline{\mathcal{L}} S^\dagger_{\Delta \dot{f} \Delta \dot{f}} \dot{D} S^\dagger_{\Delta \dot{f} \Delta \dot{f}} \underline{\mathcal{L}} - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} \right) \underline{\mathcal{L}}^\dagger \Delta \dot{f}_g \right| \\
&\leq \left\| \tfrac{n}{J+1} \underline{\mathcal{L}} S^\dagger_{\Delta \dot{f} \Delta \dot{f}} \tfrac{J+1}{n} \dot{D} \tfrac{n}{J+1} S^\dagger_{\Delta \dot{f} \Delta \dot{f}} \underline{\mathcal{L}} - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} \right\| \max_g \underline{B}_{gg} = o_p(\max_g \underline{B}_{gg})
\end{aligned}$$

$$\left| \mathrm{trace}(\underline{\tilde{A}} - \tilde{A}) \right| = \left| \sum_{\ell=1}^J \tfrac{1}{n \dot{\lambda}_\ell} - \tfrac{1}{n \underline{\dot{\lambda}}_\ell} \right| = \mathrm{trace}(\underline{\tilde{A}}) O\left(\max_\ell \left| \tfrac{\dot{\lambda}_\ell - \underline{\dot{\lambda}}_\ell}{\underline{\dot{\lambda}}_\ell} \right| \right) = o_p\left(\mathrm{trace}(\underline{\tilde{A}}) \right)$$

5 Relation To Existing Approaches

Next we verify that the bias of $\hat{\theta}_{\mathrm{HO}}$ is a function of the covariation between σ_i^2 and (B_{ii}, P_{ii}). Specifically, the bias of $\hat{\theta}_{\mathrm{HO}}$ is $\sigma_{nB_{ii},\sigma_i^2} + S_B \frac{n}{n-k} \sigma_{P_{ii},\sigma_i^2}$ where $\sigma_{nB_{ii},\sigma_i^2} = \sum_{i=1}^n B_{ii}(\sigma_i^2 - \bar{\sigma}^2)$, $\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2$, $S_B = \sum_{i=1}^n B_{ii}$, $\sigma_{P_{ii},\sigma_i^2} = \frac{1}{n}\sum_{i=1}^n P_{ii}(\sigma_i^2 - \bar{\sigma}^2)$. This is so since $\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^n (y_i - x_i'\hat{\beta})^2 = \frac{1}{n-k}\sum_{i=1}^n \sum_{\ell=1}^n M_{i\ell}\varepsilon_i\varepsilon_\ell$ from which we get that

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_{\mathrm{HO}}] - \theta &= \sum_{i=1}^n B_{ii}\sigma_i^2 - \left(\sum_{i=1}^n B_{ii} \right) \tfrac{1}{n-k} \sum_{i=1}^n M_{ii}\sigma_i^2 \\
&= \sum_{i=1}^n B_{ii}(\sigma_i^2 - \bar{\sigma}^2) - S_B \tfrac{1}{n-k} \sum_{i=1}^n M_{ii}(\sigma_i^2 - \bar{\sigma}^2) = \sigma_{nB_{ii},\sigma_i^2} + S_B \tfrac{n}{n-k} \sigma_{P_{ii},\sigma_i^2}.
\end{aligned}$$

From this formula and the discussion of Example 1, it immediately follows that the homoscedasticity-only estimator $\hat{\theta}_{\text{HO}}$ is first order biased in unbalanced panels with heteroscedasticity

Comparison to Jackknife Estimators

We finish by comparing the leave-out estimator $\hat{\theta}$ to estimators predicated on jackknife bias corrections. We start by introducing some of the high-level assumptions that are typically used to motivate jackknife estimators. We then consider some variants of Examples 1 and 2 where these high-level conditions fail to hold and establish that the jackknife estimators have first order biases while the leave-out estimator retains consistency.

High-level Conditions Jackknife bias corrections are typically motivated by the high-level assumption that the bias of a plug-in estimator $\hat{\theta}_{\text{PI}}$ shrinks with the sample size in a known way and that the bias of $\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{\text{PI},-i}$ depends on sample size in an identical way, i.e.,

$$\mathbb{E}[\hat{\theta}_{\text{PI}}] = \theta + \tfrac{\text{D}_1}{n} + \tfrac{\text{D}_2}{n^2}, \quad \mathbb{E}\left[\tfrac{1}{n}\sum\nolimits_{i=1}^{n}\hat{\theta}_{\text{PI},-i}\right] = \theta + \tfrac{\text{D}_1}{n-1} + \tfrac{\text{D}_2}{(n-1)^2} \quad \text{for some } \text{D}_1, \text{D}_2. \tag{4}$$

Under (4), the jackknife estimator $\hat{\theta}_{\text{JK}} = n\hat{\theta}_{\text{PI}} - \frac{n-1}{n}\sum_{i=1}^{n}\hat{\theta}_{\text{PI},-i}$ has a bias of $-\frac{\text{D}_2}{n(n-1)}$.

For some long panel settings the bias in $\hat{\theta}_{\text{PI}}$ is shrinking in the number of time periods T such that

$$\mathbb{E}[\hat{\theta}_{\text{PI}}] = \theta + \tfrac{\dot{\text{D}}_1}{T} + \tfrac{\dot{\text{D}}_2}{T^2} \quad \text{for some } \dot{\text{D}}_1, \dot{\text{D}}_2.$$

In such settings, it may be that the biases of $\frac{1}{T}\sum_{t=1}^{T}\hat{\theta}_{\text{PI},-t}$ and $\frac{1}{2}(\hat{\theta}_{\text{PI},1} + \hat{\theta}_{\text{PI},2})$ depend on T in an identical way, i.e.,

$$\mathbb{E}\left[\tfrac{1}{T}\sum\nolimits_{t=1}^{T}\hat{\theta}_{\text{PI},-t}\right] = \theta + \tfrac{\dot{\text{D}}_1}{T-1} + \tfrac{\dot{\text{D}}_2}{(T-1)^2} \quad \text{and} \quad \mathbb{E}\left[\tfrac{1}{2}(\hat{\theta}_{\text{PI},1} + \hat{\theta}_{\text{PI},2})\right] = \theta + \tfrac{2\dot{\text{D}}_1}{T} + \tfrac{4\dot{\text{D}}_2}{T^2}.$$

From here it follows that the panel jackknife estimator $\hat{\theta}_{\text{PJK}} = T\hat{\theta}_{\text{PI}} - \frac{T-1}{T}\sum_{t=1}^{T}\hat{\theta}_{\text{PI},-t}$ has a bias of $-\frac{\dot{\text{D}}_2}{T(T-1)}$ and that the split panel jackknife estimator $\hat{\theta}_{\text{SPJK}} = 2\hat{\theta}_{\text{PI}} - \frac{1}{2}(\hat{\theta}_{\text{PI},1} + \hat{\theta}_{\text{PI},2})$ has a bias of $-\frac{2\dot{\text{D}}_2}{T^2}$, both of which shrink faster to zero than $\frac{\dot{\text{D}}_1}{T}$ if $T \to \infty$. Typical sufficient conditions for bias-representations of this kind to hold (to second order) are that (i) $T \to \infty$, (ii) the design is stationary over time, and (iii) that $\hat{\theta}_{\text{PI}}$ is asymptotically linear (see, e.g., Hahn and Newey, 2004; Dhaene and Jochmans, 2015). Below we illustrate that jackknife corrections can be inconsistent in Examples 1 and 2 when (i) and/or (ii) do not hold.

Examples of Jackknife Failure

Example 1 (Special case)**.** Consider the model

$$y_{gt} = \alpha_g + \varepsilon_{gt} \qquad\qquad (g = 1, \dots, N,\ t = 1, \dots, T \geq 2),$$

where $\sigma_{gt}^2 = \sigma^2$ and suppose the parameter of interest is $\theta = \frac{1}{N}\sum_{g=1}^{N} \alpha_g^2$. For T even, we have the following bias calculations:

$$\mathbb{E}[\hat{\theta}_{\mathrm{PI}}] = \theta + \tfrac{\sigma^2}{T}, \qquad \mathbb{E}\left[\tfrac{1}{n}\sum\nolimits_{i=1}^{n} \hat{\theta}_{\mathrm{PI},-i}\right] = \theta + \tfrac{\sigma^2}{T} + \tfrac{\sigma^2}{n(T-1)},$$
$$\mathbb{E}\left[\tfrac{1}{T}\sum\nolimits_{t=1}^{T} \hat{\theta}_{\mathrm{PI},-t}\right] = \theta + \tfrac{\sigma^2}{T-1}, \qquad \mathbb{E}\left[\tfrac{1}{2}(\hat{\theta}_{\mathrm{PI},1} + \hat{\theta}_{\mathrm{PI},2})\right] = \theta + \tfrac{2\sigma^2}{T}.$$

The jackknife estimator $\hat{\theta}_{JK}$ has a first order bias of $-\frac{\sigma^2}{T(T-1)}$, which when $T = 2$ is as large as that of $\hat{\theta}_{\mathrm{PI}}$ but of opposite sign. By contrast, both of the panel jackknife estimators, $\hat{\theta}_{PJK}$ and the leave-out estimator are exactly unbiased and consistent as $n \to \infty$ when T is fixed.

This example shows that the jackknife estimator can fail when applied to a setting where the number of regressors is large relative to sample size. Here the number of regressors is N and the sample size is NT, yielding a ratio of $1/T$ and $1/T \to 0$ is necessary for consistency of $\hat{\theta}_{\mathrm{JK}}$. While the panel jackknife corrections appear to handle the presence of many regressors, this property disappears when adding the "random" coefficients of Example 2.

Example 2 (Special case)**.** Consider the model

$$y_{gt} = \alpha_g + x_{gt}\delta_g + \varepsilon_{gt} \qquad\qquad (g = 1, \dots, N,\ t = 1, \dots, T \geq 3)$$

where $\sigma_{gt}^2 = \sigma^2$ and $\theta = \frac{1}{N}\sum_{g=1}^{N} \delta_g^2$.

An analytically convenient example arises when the regressor design is "balanced" across groups as follows: $(x_{g1}, x_{g2}, \dots, x_{gT}) = (x_1, x_2, \dots, x_T)$, where x_1, x_2, x_3 take distinct values and $\sum_{t=1}^{T} x_t = 0$. The leave-out estimator is unbiased and consistent for any $T \geq 3$, whereas for even $T \geq 4$ we have the following bias calculations:

$$\mathbb{E}[\hat{\theta}_{\mathrm{PI}}] = \theta + \tfrac{\sigma^2}{\sum_{t=1}^{T} x_t^2},$$
$$\mathbb{E}\left[\tfrac{1}{T}\sum\nolimits_{t=1}^{T} \hat{\theta}_{\mathrm{PI},-t}\right] = \theta + \tfrac{\sigma^2}{T}\sum\nolimits_{t=1}^{T} \tfrac{1}{\sum_{s\neq t}(x_s - \bar{x}_{-t})^2},$$
$$\mathbb{E}\left[(\hat{\theta}_{\mathrm{PI},1} + \hat{\theta}_{\mathrm{PI},2})/2\right] = \theta + \tfrac{\sigma^2}{2\sum_{t=1}^{T/2}(x_t - \bar{x}_1)^2} + \tfrac{\sigma^2}{2\sum_{t=T/2+1}^{T}(x_t - \bar{x}_2)^2},$$

where $\bar{x}_{-t} = \frac{1}{T-1}\sum_{s\neq t} x_s$, $\bar{x}_1 = \frac{2}{T}\sum_{t=1}^{T/2} x_t$, and $\bar{x}_2 = \frac{2}{T}\sum_{t=T/2+1}^{T} x_t$.

The calculations above reveal that non-stationarity in either the level or variability of x_t over time can lead to a negative bias in panel jackknife approaches, e.g.,

$$\mathbb{E}\left[\hat{\theta}_{\mathrm{SPJK}}\right] - \theta \leq \frac{2\sigma^2}{\sum_{t=1}^{T} x_t^2} - \frac{\sigma^2}{2\sum_{t=1}^{T/2} x_t^2} - \frac{\sigma^2}{2\sum_{t=T/2+1}^{T} x_t^2} \leq 0$$

where the first inequality is strict if $\bar{x}_1 \neq \bar{x}_2$ and the second if $\sum_{t=1}^{T/2} x_t^2 \neq \sum_{t=T/2+1}^{T} x_t^2$. In fact, the following example $(x_1, x_2, \ldots, x_T) = (-1, 2, 0, \ldots, 0, -1)$ renders the panel jackknife corrections inconsistent for small or large T:

$$\mathbb{E}[\hat{\theta}_{\mathrm{PJK}}] = \theta - \frac{7/5}{6}\sigma^2 + O\left(\tfrac{1}{T}\right) \quad \text{and} \quad \mathbb{E}[\hat{\theta}_{\mathrm{SPJK}}] = \theta - \frac{8/5}{6}\sigma^2 + O\left(\tfrac{1}{T}\right).$$

Inconsistency results here from biases of first order that are negative and larger in magnitude than the original bias of $\hat{\theta}_{\mathrm{PI}}$ (which is $\sigma^2/6$).

References

Achlioptas, D. (2003). Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of computer and System Sciences 66*(4), 671–687.

Andrews, I. and A. Mikusheva (2016). A geometric approach to nonlinear econometric models. *Econometrica 84*(3), 1249–1264.

Chung, F. R. (1997). *Spectral graph theory.* Number 92. American Mathematical Soc.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies 82*(3), 991–1030.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*(4), 1295–1319.

Oliveira, R. I. (2009). Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*.

Sølvsten, M. (2020). Robust estimation with many instruments. *Journal of Econometrics 214*(2), 495–512.