Appendix A Data

This Appendix describes construction of the data used in the application of Section 8.

A.1 Veneto Workers History

Our data come from the Veneto Workers History (VWH) file, which provides social security based earnings records on annual job spells for all workers employed in the Italian region of Veneto at any point between the years 1975 and 2001. Each job-year spell in the VWH lists a start date, an end date, the number of days worked that year, and the total wage compensation received by the employee in that year. The earnings records are not top-coded. We also observe the gender of each worker and several geographic variables indicating the location of each employer. See Card, Devicienti, and Maida (2014) and Serafinelli (2019) for additional discussion and analysis of the VWH.

We consider data from the years 1984–2001 as prior to that information on days worked tend to be of low quality. To construct the person-year panel used in our analysis, we follow the sample selection procedures described in Card, Heining, and Kline (2013). First, we drop employment spells in which the worker's age lies outside the range 18–64. The average worker in this sample has 1.21 jobs per year. To generate unique worker-firm assignments in each year, we restrict attention to spells associated with "dominant jobs" where the worker earned the most in each corresponding year. From this person-year file, we then exclude workers that (i) report a daily wage less than 5 real euros or have zero days worked (1.5% of remaining person-year observations) (ii) report a log daily wage change one year to the next that is greater than 1 in absolute value (6%) (iii) are employed in the public sector (10%) or (iv) have more than 10 jobs in any year or that have gender missing (0.1%).

Appendix B Computation

This Appendix describes the key computational aspects of the leave-out estimator $\hat{\theta}$, with an emphasis on the application to two-way fixed effects models with two time periods discussed in Example 4 and Section 8.

B.1 Leave-One-Out Connected Set

Existence of $\hat{\theta}$ requires $P_{ii} < 1$ (see Lemma 1) and the following describes an algorithm which prunes the data to ensure that $P_{ii} < 1$. In the two-way fixed effects model of Section 8.2, this condition requires that the bipartite network formed by worker-firm links remains connected when any one worker is removed. This boils down to finding workers that constitute cut vertices or *articulation points* in the corresponding bipartite network.

The algorithm below takes as input a connected bipartite network $\mathcal{G}$ where workers and firms are vertices. Edges between two vertices correspond to the realization of a match between a worker and a firm (see Jochmans and Weidner, 2016; Bonhomme, 2017, for discussion). In practice, one typically starts with a $\mathcal{G}$ corresponding to the *largest connected component* of a given bipartite network (see, e.g., Card et al., 2013). The output of the algorithm is a subset of $\mathcal{G}$ where removal of any given worker does not break the connectivity of the associated graph.

The algorithm relies on existing functions that efficiently finds articulation points and largest connected components. In MATLAB such functions are available in the *Boost Graph Library* and in R they are available in the *igraph* package.

Algorithm 1 Leave-One-Out Connected Set

```
function PruningNetwork(𝒢)    ▷ 𝒢 ≡ Connected bipartite network of firms and workers
    Construct 𝒢1 from 𝒢 by deleting all workers that are articulation points in 𝒢
    Let 𝒢 be the largest connected component of 𝒢1
    Return 𝒢
end function
```

The algorithm typically completes in less than a minute for datasets of the size considered in our application. Furthermore, the vast majority of firms removed using this algorithm are only associated with one mover.

B.2 Leave-Two-Out Connected Set

We also introduced a leave-two-out connected set, which is a subset of the original data such that removal of any *two* workers does not break the connectedness of the bipartite network formed by worker-firm links. The following algorithm proceeds by applying the idea in Algorithm 1 to each of the networks constructed by dropping one worker. A crucial difference from Algorithm 1 is that *two* workers who do not break connectedness in the input network may break connectedness when other workers have been removed. For this reason, the algorithm runs in an iterative fashion until it fails to remove any additional workers.

Algorithm 2 Leave-Two-Out Connected Set

function PRUNINGNETWORK2($\mathcal{G}$) ▷ $\mathcal{G} \equiv$ Leave-one-out connected bipartite network of firms and workers
    $a = 1$
    while $a > 0$ **do**
        $\mathcal{G}^{del} = \emptyset$
        for $g = 1, \ldots, N$ **do**
            Construct $\mathcal{G}_1$ from $\mathcal{G}$ by deleting worker g
            Add all workers that are articulation points in $\mathcal{G}_1$ to $\mathcal{G}^{del}$
        end for
        $a = |\mathcal{G}^{del}|$
        if $a > 0$ **then**
            Construct $\mathcal{G}_1$ from $\mathcal{G}$ by deleting all workers in $\mathcal{G}^{del}$
            Let $\mathcal{G}_2$ be the largest connected component of $\mathcal{G}_1$
            Let $\mathcal{G}$ be the output of applying Algorithm 1 to $\mathcal{G}_2$
        end if
    end while
    Return $\mathcal{G}$
end function

B.3 Computing $\hat{\theta}$

Our proposed leave-out estimator is a function of the $2n$ quadratic forms

$$P_{ii} = x_i' S_{xx}^{-1} x_i \qquad B_{ii} = x_i' S_{xx}^{-1} A S_{xx}^{-1} x_i \qquad \text{for } i = 1, ..., n.$$

The estimates reported in Section 8 of the paper rely on exact computation of these quantities. In our application, k is on the order of hundreds of thousands, making it infeasible to compute S_{xx}^{-1} directly. To circumvent this obstacle, we instead compute the k-dimensional vector $z_{i,exact} = S_{xx}^{-1} x_i$ separately for each $i = 1, .., n$. That is, we solve separately for each column of Z_{exact} in the system

$$\underset{k \times k}{S_{xx}} \underset{k \times n}{Z_{exact}} = \underset{k \times n}{X'}.$$

We then form $P_{ii} = x_i' z_{i,exact}$ and $B_{ii} = z_{i,exact}' A z_{i,exact}$. The solution $z_{i,exact}$ is computed via MATLAB's preconditioned conjugate gradient routine *pcg*. In computing this solution, we utilize the preconditioner developed by Koutis et al. (2011), which is optimized for diagonally dominant design matrices S_{xx}. These column-specific calculations are parallelized across different cores using MATLAB's *parfor* command.

B.3.1 Leaving a Cluster Out

Table 3 applies the leave-cluster-out estimator introduced in Remark 3 to estimate the variance of firm effects with more than two time periods and potential serial correlation. The estimator takes the form $\hat{\theta}_{cluster} = \sum_{i=1}^{n} y_i \tilde{x}_i' \hat{\beta}_{-c(i)}$ where $\hat{\beta}_{-c(i)}$ is the OLS estimator obtained after leaving out all observations in the cluster to which observation i belongs. A representation of $\hat{\theta}_{cluster}$ that is useful for computation takes the observations in the c-th cluster and collect their outcomes in y_c and their regressors in X_c. The leave-cluster-out estimator is then

$$\hat{\theta}_{cluster} = \hat{\beta}' A \hat{\beta} - \sum_{c=1}^{C} y_c' B_c (I - P_c)^{-1} (y_c - X_c \hat{\beta}),$$

where C denotes the total number of clusters, $P_c = X_c S_{xx}^{-1} X_c'$, and $B_c = X_c S_{xx}^{-1} A S_{xx}^{-1} X_c'$. Since the entries of P_c and B_c are of the form $P_{i\ell} = x_i' S_{xx}^{-1} x_\ell$ and $B_{i\ell} = x_i' S_{xx}^{-1} A S_{xx}^{-1} x_\ell$, computation can proceed in a similar fashion as described earlier for the leave-one-out estimator.

When defining the cluster as a worker-firm match, Table 3 applies $\hat{\theta}_{cluster}$ to the two-way fixed effects model in (6). When defining the cluster as a worker, the individual effects can not be estimated after leaving a cluster out. Table 3 therefore applies $\hat{\theta}_{cluster}$ after demeaning at the individual level. This transformation removes the individual effects so that the resulting model can be estimated after leaving a cluster out.

B.3.2 Johnson-Lindenstrauss Approximation

When n is on the order of hundreds of millions and k is on the order of tens of millions, the exact algorithm may no longer be tractable. The JLA simplifies computation of P_{ii} considerably by only requiring the solution of p systems of k linear equations. That is, one need only solve for the columns of Z_{JLA} in the system

$$\underset{k \times k}{S_{xx}} \underset{k \times p}{Z_{JLA}} = \underset{k \times p}{(R_P X)'},$$

which reduces computation time dramatically when p is small relative to n.

To compute B_{ii}, it is necessary to solve linear systems involving both A_1 and A_2, leading to $2p$

systems of equations when $A_1 \neq A_2$. However, for variance decompositions like the ones considered in Section 8.2, the same $2p$ systems can be reused for all three variance components, leading to a total of $3p$ systems of equations for the full variance decomposition. This is so because the three variance components use the matrices $A_\psi = A_f' A_f$, $A_{\alpha,\psi} = \frac{1}{2}(A_d' A_f + A_f' A_d)$, and $A_\alpha = A_d' A_d$ where

$$A_f' = \frac{1}{\sqrt{n}} \begin{bmatrix} 0 & 0 & 0 \\ f_1 - \bar{f} & \dots & f_n - \bar{f} \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad A_d' = \frac{1}{\sqrt{n}} \begin{bmatrix} d_1 - \bar{d} & \dots & d_n - \bar{d} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Based on these insights, Algorithm 3 below takes as inputs X, A_f, A_d, and p, and returns $\hat{P}_{ii}$ and three different $\hat{B}_{ii}$'s which are ultimately used to construct the corresponding variance component $\hat{\theta}_{JLA}$ as defined in Section 1.2.

Algorithm 3 Johnson-Lindenstrauss Approximation for Two-Way Fixed Effects Models

function JLA(X,A_f,A_d,p)
  Generate $R_B, R_P \in \mathbb{R}^{p\times n}$, where (R_B, R_P) are composed of mutually independent Rademacher entries
  Compute $(R_P X)'$, $(R_B A_f)'$, $(R_B A_d)' \in \mathbb{R}^{k\times p}$
  for $\kappa = 1, \dots, p$ **do**
    Let $r_{\kappa,0}$, $r_{\kappa,1}$, $r_{\kappa,2} \in \mathbb{R}^k$ be the κ-th columns of $(R_P X)'$, $(R_B A_f)'$, $(R_B A_d)'$
    Let $z_{\kappa,\ell} \in \mathbb{R}^k$ be the solution to $S_{xx} z = r_{\kappa,\ell}$ for $\ell = 0, 1, 2$
  end for
  Construct $Z_\ell = (z_{1,\ell}, \dots, z_{p,\ell}) \in \mathbb{R}^{k\times p}$ for $\ell = 0, 1, 2$
  Construct $\hat{P}_{ii} = \frac{1}{p}\left\|Z_0' x_i\right\|^2$, $\hat{B}_{ii,\psi} = \frac{1}{p}\left\|Z_1' x_i\right\|^2$, $\hat{B}_{ii,\alpha} = \frac{1}{p}\left\|Z_2' x_i\right\|^2$, $\hat{B}_{ii,\alpha\psi} = \frac{1}{p}(Z_1' x_i)'(Z_2' x_i)$ for $i = 1, \dots, n$
  Return $\{\hat{P}_{ii}, \hat{B}_{ii,\psi}, \hat{B}_{ii,\alpha}, \hat{B}_{ii,\alpha\psi}\}_{i=1}^n$
end function

B.3.3 Performance of the JLA

Figure B.1 evaluates the performance of the Johnson-Lindenstrauss approximation across 4 VWH samples that correspond to different (overlapping) time intervals (2000–2001; 1999–2001; 1998–2001; 1997–2001). The x-axis in Figure B.1 reports the total number of person and firm effects associated with a particular sample.

Figure B.1 shows that the computation time for exact computation of (B_{ii}, P_{ii}) increases rapidly as the number of parameters of the underlying AKM model grow; in the largest dataset considered – which involves more than a million worker and firm effects – exact computation takes about 8 hours. Computation of JLA complete in markedly shorter time: in the largest dataset considered computation time is less than 5 minutes when $p = 500$ and slightly over 6 minutes when $p = 2500$.

Notably, the JLA delivers estimates of the variance of firm effects almost identical to those computed via the exact method, with the quality of the approximation increasing for larger p. For instance, in the largest dataset, the exact estimate of variance of firm effects is 0.028883. By comparison, the JLA estimate equals 0.028765 when $p = 500$ and 0.0289022 when $p = 2500$.

In summary: for a sample with more than a million worker and firm effects, the JLA cuts computation time by a factor of 100 while introducing an approximation error of roughly 10^{-4}.

B.3.4 Scaling to Very Large Datasets

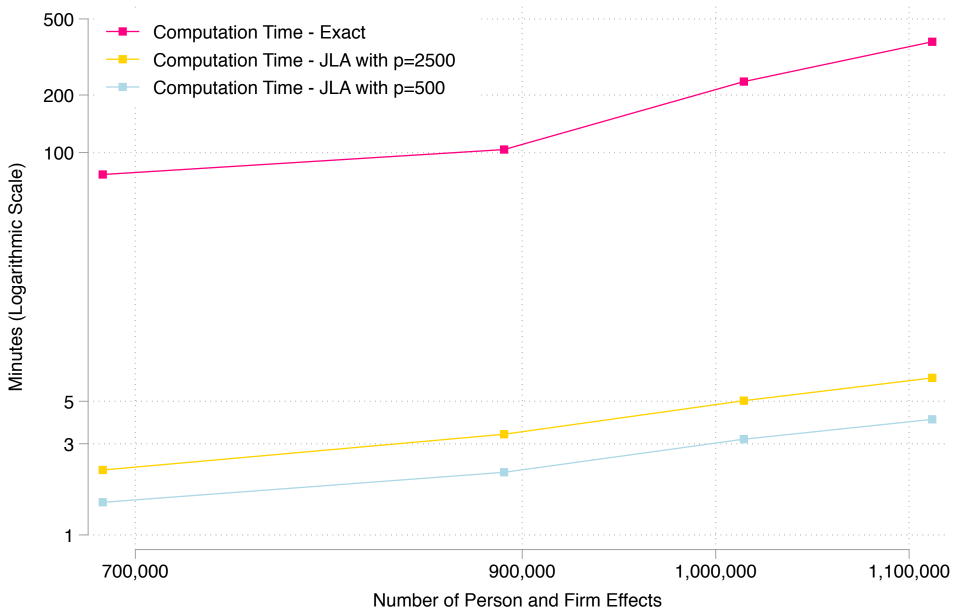
We now study how the JLA scales to much larger datasets of the dimension considered by Card et al. (2013) who fit models involving tens of millions of worker and firm effects to German social security records. To study the computational burden of a model of this scale, we rely on a synthetic dataset constructed from our original leave-one-out sample analyzed in Column 1 of Table 2, i.e., the pooled Veneto sample comprised of wage observations from the years 1999 and 2001. We scale the data by creating replicas of this base sample. To connect the replicas, we draw at random 10% of the movers and randomly exchange their period 1 firm assignments across replicas. By construction, this permutation maintains each (replicated) firm's size while ensuring leave-one-out connectedness of the resulting network.

Wage observations are drawn from a variant of the DGP described in Section 8.7 adapted to the levels formulation of the model. Specifically, each worker's wage is the sum of a rescaled person effect, a rescaled firm effect, and an error drawn independently in each period from a normal with variance $\frac{1}{2}\exp(\hat{a}_0 + \hat{a}_1 B_{gg} + \hat{a}_2 P_{gg} + \hat{a}_3 \ln L_{g2} + \hat{a}_4 \ln L_{g1})$. As highlighted by Figure B.1, computing the exact estimator in these datasets would be extremely costly. Drawing from a stable DGP allows us to instead benchmark the JLA estimator against the true value of the variance of firm effects.

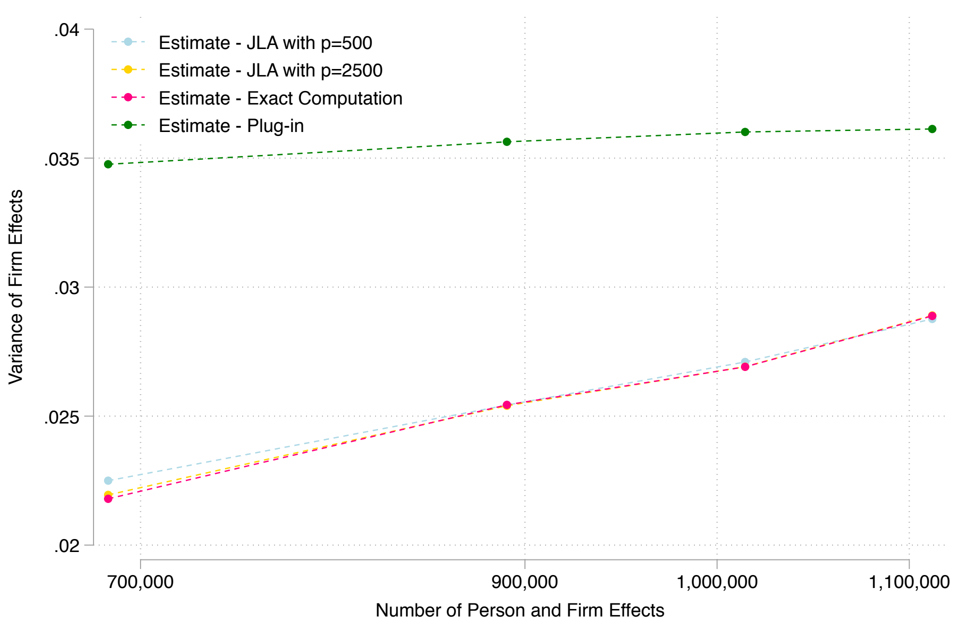
Figure B.2 displays the results. When setting $p = 250$, the JLA delivers a variance of firm effects remarkably close to the true variance of firm effects defined by our DGP. As expected, the distance between our approximation and the true variance component decreases with the sample size for a fixed p. Remarkably, we are able to compute the AKM variance decomposition in a dataset with approximately 15 million person and year effects in only 35 minutes. Increasing the number of simulated draws in the JLA to $p = 500$ delivers estimates of the variance of firm effects nearly indistinguishable from the true value. This is achieved in approximately one hour in the largest simulated dataset considered. The results of this exercise strongly suggest the leave-out estimator can be scaled to extremely large datasets involving the universe of administrative wage records in large countries such as Germany or the United States.

Figure B.1: Performance of the JLA Algorithm

(a) Computation Time



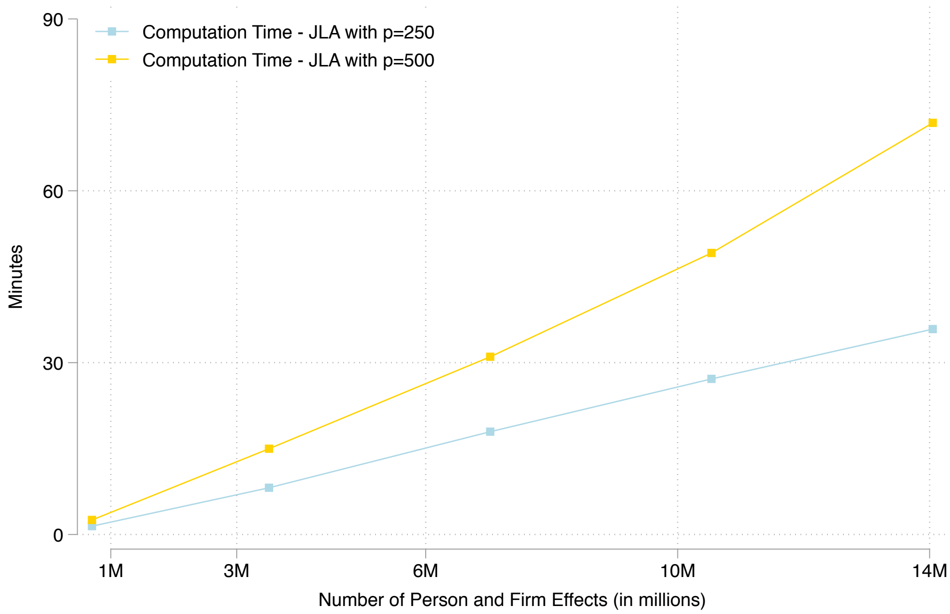
(b) Quality of the Approximation



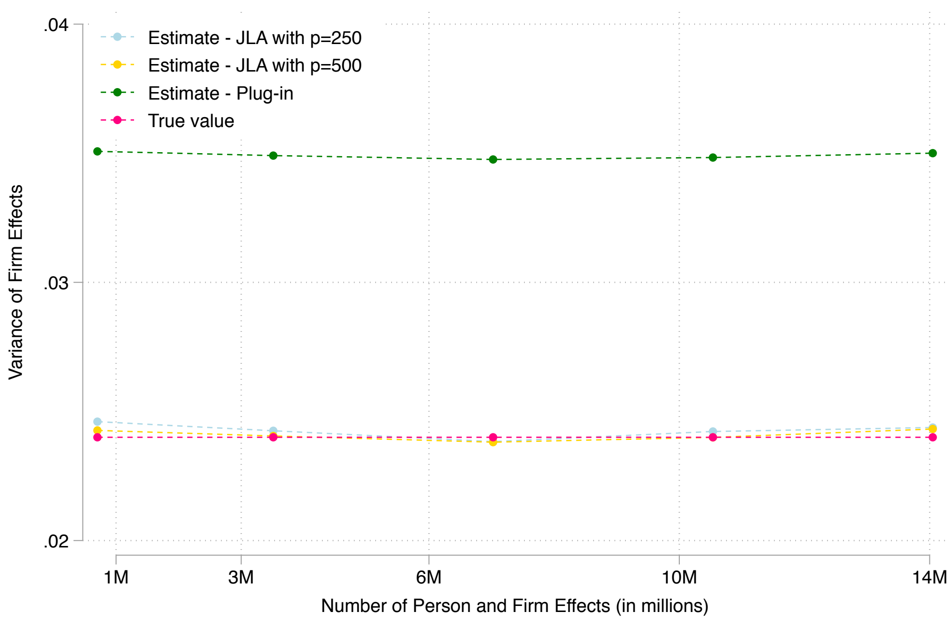
Note: Both panels consider 4 different samples of increasing length. The four samples contain data from the years 2000–2001, 1999–2001, 1998–2001, and 1997–2001, respectively. The x-axis reports the number of person and firm effects in each sample. Panel (a) shows the time to compute the KSS estimate when relying on either exact computation of $\{B_{ii}, P_{ii}\}_{i=1}^{n}$ or the Johnson-Lindenstrauss approximation (JLA) of these numbers using a p of either 500 or 2500. Panel (b) shows the resulting estimates and the plug-in estimate. Computations performed on a 32 core machine with 256 GB of dedicated memory. Source: VWH dataset.

Figure B.2: Scaling to Very Large Datasets

(a) Computation Time



(b) Quality of the Approximation



Note: Both panels consider synthetic datasets created from the pooled Veneto data in column 1 of Table 2 with $T = 2$. It considers $\{1, 5, 10, 15, 20\}$ replicas of this sample while generating random links across replicas such that firm size and T are kept fixed. Outcomes are generated from a DGP of the sort considered in Table 6. The x-axis reports the number of person and firm effects in each sample. Panel (a) shows the time to compute the Johnson-Lindenstrauss approximation $\hat{\theta}_{JLA}$ using a p of either 250 or 500. Panel (b) shows the resulting estimates, the plug-in estimate, and the true value of the variance of firm effects for the DGP. Computations performed on a 32 core machine with 256 GB of dedicated memory. Source: VWH dataset.

B.4 Split Sample Estimators

Sections 4.2 and 5.2 proposed standard error estimators predicated on being able to construct independent split sample estimators $\widehat{x_i'\beta}_{-i,1}$ and $\widehat{x_i'\beta}_{-i,2}$. This section describes an algorithm for construction of these split sample estimators in the two-way fixed effects model of Example 4. We restrict attention to the case with $T_g = 2$ and consider the model in first differences: $\Delta y_g = \Delta f_g' \psi + \Delta \varepsilon_g$ for $g = 1, \dots, N$. When worker g moves from firm j to j', we can estimate $\Delta f_g' \psi = \psi_{j'} - \psi_j$ without bias using OLS on any sub-sample where firms j and j' are connected, i.e., on any sample where there exist a path between firm j and j'. To construct two disjoint sub-samples where firms j and j' are connected we therefore use an algorithm to find disjoint paths between these firms and distribute them into two sub-samples which will be denoted $\mathcal{S}_1$ and $\mathcal{S}_2$. Because it can be computationally prohibitive to characterize all possible paths, we use a version of Dijkstra's algorithm to find many short paths.[10]

Our algorithm is based on a network where firms are vertices and two firms are connected by an edge if one or more workers moved between them. This view of the network is the same as the one taken in Section 7, but different from the one used in Sections B.1 and B.2 where both firms and workers were viewed as vertices. We use the adjacency matrix $\mathcal{A}$ to characterize the network in this section. To build the sub-samples $\mathcal{S}_1$ and $\mathcal{S}_2$, the algorithm successively drops workers from the network, so $\mathcal{A}_{-\mathcal{S}}$ will denote the adjacency matrix after dropping all workers in the set $\mathcal{S}$.

Given a network characterized by $\mathcal{A}$ and two connected firms j and j' in the network, we let $\dot{P}_{jj'}(\mathcal{A})$ denote the shortest path between them.[11] If j and j' are not connected $\dot{P}_{jj'}(\mathcal{A})$ is empty. Each edge in the path $\dot{P}_{jj'}(\mathcal{A})$ may have more than one worker associated with it. For each edge in $\dot{P}_{jj'}(\mathcal{A})$ the first step of the algorithm picks at random a single worker associated with that edge and places them in $\mathcal{S}_1$, while later steps place all workers associated with the shortest path in one of $\mathcal{S}_1$ and $\mathcal{S}_2$. This special first step ensures that the algorithm finds two independent unbiased estimators of $\Delta f_g' \psi$ whenever the network $\mathcal{A}$ is leave-two-out connected.

For a given worker g with firm assignments $j = j(g, 1), j' = j(g, 2)$ and a leave-two-out connected network $\mathcal{A}$ the algorithm returns the $\{P_{g\ell,1}, P_{g\ell,2}\}_{\ell=1}^N$ introduced in Section 4.2. Specifically, $\widehat{\Delta f_g' \psi}_{-g,1} = \sum_{\ell=1}^N P_{g\ell,1} \Delta y_\ell$ and $\widehat{\Delta f_g' \psi}_{-g,2} = \sum_{\ell=1}^N P_{g\ell,2} \Delta y_\ell$ are independent unbiased estimators of

[10] The algorithm presented below keeps running until it cannot find any additional paths. In our empirical implementation we stop the algorithm when it fails to find any new paths or as soon as one of the two sub-samples reach a size of at least 100 workers. We found that increasing this cap on the sub-sample size has virtually no effect on the estimated confidence intervals, but tends to increase computation time substantially.

[11] Many statistical software packages provide functions that can find shortest paths. In R they are available in the *igraph* package while in MATLAB a package that builds on the work of Yen (1971) is available at https://www.mathworks.com/matlabcentral/fileexchange/35397-k-shortest-paths-in-a-graph-represented-by-a-sparse-matrix-yen-s-algorithm?focused=3779015&tab=function.

$\Delta f_g'\psi$ that are also independent of Δy_g. If $\mathcal{A}$ is only leave-one-out connected then the algorithm may only find one path connecting j and j'. When this happens the algorithm sets $P_{g\ell,2} = 0$ for all ℓ as required in the formulation of the conservative standard errors proposed in Appendix C.5.1.

Algorithm 4 Split Sample Estimator for Inference

1: **function** SPLITSAMPLEESTIMATOR$(g, j, j', \mathcal{A})$ ▷ $\mathcal{A} \equiv$ Leave-one-out connected network
2: Let $\mathcal{S}_1 = \emptyset$ and $\mathcal{S}_2 = \emptyset$
3: For each edge in $\dot{P}_{jj'}(\mathcal{A}_{-g})$, pick at random one worker from $\mathcal{A}_{-g}$ who is associated with that edge and add that worker to $\mathcal{S}_1$
4: Add to $\mathcal{S}_2$ all workers from $\mathcal{A}_{-\{g,\mathcal{S}_1\}}$ who are associated with an edge in $\dot{P}_{jj'}(\mathcal{A}_{-\{g,\mathcal{S}_1\}})$
5: Add to $\mathcal{S}_1$ all workers from $\mathcal{A}_{-\{g,\mathcal{S}_1,\mathcal{S}_2\}}$ who are associated with an edge in $\dot{P}_{jj'}(\mathcal{A}_{-g})$
6: Let *stop* $= 1\{\dot{P}_{jj'}(\mathcal{A}_{-\{g,\mathcal{S}_1,\mathcal{S}_2\}}) = \emptyset\}$ and $s = 1$
7: **while** *stop* < 1 **do**
8: Add to $\mathcal{S}_s$ all workers from $\mathcal{A}_{-\{g,\mathcal{S}_1,\mathcal{S}_2\}}$ who are associated with an edge in $\dot{P}_{jj'}(\mathcal{A}_{-\{g,\mathcal{S}_1,\mathcal{S}_2\}})$
9: Let *stop* $= 1\{\dot{P}_{jj'}(\mathcal{A}_{-\{g,\mathcal{S}_1,\mathcal{S}_2\}}) = \emptyset\}$ and update s to $1 + 1\{s = 1\}$
10: **end while**
11: For $s = 1, 2$ and $\ell = 1, \ldots, N$, let $P_{g\ell,s} = 1\{\ell \in \mathcal{S}_s\}\Delta f_\ell'(\sum_{m\in\mathcal{S}_s} \Delta f_m \Delta f_m')^\dagger \Delta f_g$
12: Return $\{P_{g\ell,1}, P_{g\ell,2}\}_{\ell=1}^N$
13: **end function**

In line 5, all workers associated with the shortest path in line 3 are added to $\mathcal{S}_1$ if they were not added to $\mathcal{S}_2$ in line 4. This step ensures that all workers associated with $\dot{P}_{jj'}(\mathcal{A}_{-g})$ are used in the predictions. In line 11, $P_{g\ell,s}$ is constructed as the weight observation ℓ receives in the prediction $\Delta f_g'\hat{\psi}_s$ where $\hat{\psi}_s$ is the OLS estimator of ψ based on the sub-sample $\mathcal{S}_s$.

B.5 Test of Equal Firm Effects

This section describes computation and interpretation of the test of the hypothesis that firm effects for "younger" workers are equal to firm effects for the "older" workers which applies Remark 6 of the main text.

The hypothesis of interest corresponds to a restricted and unrestricted model which when written in matrix notation are

$$\Delta y = \Delta F \psi + \Delta\varepsilon \tag{9}$$

$$\Delta y = \Delta F_O \psi^O + \Delta F_Y \psi^Y + \Delta F_3 \psi_3 + \Delta\varepsilon = X\beta + \Delta\varepsilon \tag{10}$$

where Δy and ΔF collects the first differences Δy_g and Δf_g across g. ΔF_O represents ΔF for "doubly connected" firms present in each age group's leave-one-out connected set interacted with a dummy for whether the worker is "old"; ΔF_Y represents ΔF for doubly connected firms interacted with a dummy for young; ΔF_3 represents ΔF for firms that are associated with either younger movers or older movers but not both. Finally, we let $X = (\Delta F_O, \Delta F_Y, \Delta F_3)$, $\beta = (\psi^{O\prime}, \psi^{Y\prime}, \psi_3')'$, and $\psi = (\psi^{O\prime}, \psi_3')'$.

The hypothesis in question is $\psi^O - \psi^Y = 0$ or equivalently $R\beta = 0$ for $R = [I_r, -I_r, 0]$ and $r = |\mathcal{J}| = \dim(\psi^O)$. Thus we can create the numerator of our test statistic by applying Remark 6 to (10) yielding

$$\hat{\theta} = \hat{\beta}' A \hat{\beta} - \sum_{g=1}^{N} B_{gg} \hat{\sigma}_g^2 \tag{11}$$

where $A = \frac{1}{r} R'(RS_{xx}^{-1}R')^{-1}R$; B_{gg} and $\hat{\sigma}_g^2$ are defined as in Section 1.

Two insights help to simplify computation. First, since $\Delta F_O' \Delta F_Y = 0$, $\Delta F_O' \Delta F_3 = 0$ and $\Delta F_Y' \Delta F_3 = 0$, we can estimate equation (10) via two separate regressions, one on the leave-one-out connected set for younger workers and the other on the leave-one-out connected set for older workers. We normalize the firm effects so that the same firm is dropped in both leave-one-out samples.

Second, we note that $\hat{\beta}' A \hat{\beta} = y' B y$ where

$$B = X S_{xx}^{-1} A S_{xx}^{-1} X' = \frac{P_X - P_{\Delta F}}{r}, \tag{12}$$

$P_X = X S_{xx}^{-1} X'$, and $P_{\Delta F} = \Delta F (\Delta F' \Delta F)^{-1} \Delta F'$. Equation (12) therefore implies that B_{ii} in (11) is simply a scaled difference between two statistical leverages: the first one obtained in the unrestricted model (10), say $P_{X,gg}$, and the other on the restricted model of (9), say $P_{\Delta F, gg}$. Section B.3 describes how to efficiently compute these statistical leverages. To conduct inference on the quadratic form in (11) we apply the routine described in Section 4.2.

Appendix C Proofs

This Appendix contains all technical details and proofs that where left out of the paper. The material is primarily presented in the order it appears in the paper and under the same headings.

C.1 Unbiased Estimation of Variance Components

C.1.1 Estimator

Lemma C.1. *It follows from the Sherman-Morrison-Woodbury formula that the two representations of $\hat{\theta}$ given in (1) and (2) are numerically identical, i.e., that $\hat{\beta}' A\hat{\beta} - \sum_{i=1}^n B_{ii}\hat{\sigma}_i^2 = \sum_{i=1}^n y_i \tilde{x}_i' \hat{\beta}_{-i}$ whenever S_{xx} has full rank and $\max_i P_{ii} < 1$.*

Proof. The Sherman-Morrison-Woodbury formula states that if S_{xx} has full rank and $P_{ii} < 1$, then

$$S_{xx}^{-1} + \frac{S_{xx}^{-1} x_i x_i' S_{xx}^{-1}}{1 - x_i' S_{xx}^{-1} x_i} = \left(S_{xx} - x_i x_i'\right)^{-1}.$$

Furthermore, we have that $\tilde{x}_i' S_{xx}^{-1} x_i = x_i S_{xx}^{-1} A S_{xx}^{-1} x_i = B_{ii}$ so

$$\begin{aligned} y_i \tilde{x}_i' \hat{\beta}_{-i} &= y_i \tilde{x}_i' \left(S_{xx} - x_i x_i'\right)^{-1} \sum_{\ell \neq i} x_\ell y_\ell = y_i \tilde{x}_i' S_{xx}^{-1} \sum_{\ell \neq i} x_\ell y_\ell + \frac{y_i \tilde{x}_i' S_{xx}^{-1} x_i x_i' S_{xx}^{-1}}{1 - x_i' S_{xx}^{-1} x_i} \sum_{\ell \neq i} x_\ell y_\ell \\ &= y_i \tilde{x}_i' \hat{\beta} - B_{ii} y_i^2 + y_i B_{ii} \underbrace{x_i' \frac{S_{xx}^{-1}}{1 - x_i' S_{xx}^{-1} x_i} \sum_{\ell \neq i} x_\ell y_\ell}_{= x_i' \hat{\beta}_{-i}} = y_i \tilde{x}_i' \hat{\beta} - B_{ii} y_i (y_i - x_i' \hat{\beta}_{-i}) \end{aligned}$$

where the last expression equals $y_i \tilde{x}_i' \hat{\beta} - B_{ii} \hat{\sigma}_i^2$. This finishes the proof since $\hat{\beta}' A \hat{\beta} = \sum_{i=1}^n y_i \tilde{x}_i' \hat{\beta}$. In the above the Sherman-Morrison-Woodbury formula was also used to establish that

$$x_i' \hat{\beta}_{-i} = x_i' \left(S_{xx} - x_i x_i'\right)^{-1} \sum_{\ell \neq i} x_\ell y_\ell = x_i' \frac{S_{xx}^{-1}}{1 - x_i' S_{xx}^{-1} x_i} \sum_{\ell \neq i} x_\ell y_\ell,$$

and from this it follows that $y_i - x_i' \hat{\beta}_{-i} = \dfrac{y_i - x_i' \hat{\beta}}{1 - P_{ii}}$ as claimed in the paper. □

C.1.2 Large Scale Computation

All discussions of the computational aspects are collected in Appendix B.

C.1.3 Relation To Existing Approaches

Next we verify that the bias of $\hat{\theta}_{\text{HO}}$ is a function of the covariation between σ_i^2 and (B_{ii}, P_{ii}).

Lemma C.2. *The bias of* $\hat{\theta}_{HO}$ *is* $\sigma_{nB_{ii},\sigma_i^2} + S_B \frac{n}{n-k}\sigma_{P_{ii},\sigma_i^2}$ *where*

$$\sigma_{nB_{ii},\sigma_i^2} = \sum_{i=1}^{n} B_{ii}(\sigma_i^2 - \bar{\sigma}^2), \quad \bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2, \quad S_B = \sum_{i=1}^{n} B_{ii}, \quad \sigma_{P_{ii},\sigma_i^2} = \frac{1}{n}\sum_{i=1}^{n} P_{ii}(\sigma_i^2 - \bar{\sigma}^2).$$

Proof. Since $\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 = \frac{1}{n-k}\sum_{i=1}^{n}\sum_{\ell=1}^{n} M_{i\ell}\varepsilon_i\varepsilon_\ell$ we get that

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

□

Comparison to Jackknife Estimators

This subsection compares 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 2 and 3 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.,

$$\mathbb{E}[\hat{\theta}_{\text{PI}}] = \theta + \frac{\text{D}_1}{n} + \frac{\text{D}_2}{n^2}, \quad \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{\text{PI},-i}\right] = \theta + \frac{\text{D}_1}{n-1} + \frac{\text{D}_2}{(n-1)^2} \quad \text{for some } \text{D}_1, \text{D}_2. \tag{13}$$

Under (13), 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

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

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.,

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

From here it follows that the panel jackknife estimator $\hat{\theta}_{\mathrm{PJK}} = T\hat{\theta}_{\mathrm{PI}} - \frac{T-1}{T}\sum_{t=1}^{T}\hat{\theta}_{\mathrm{PI},-t}$ has a bias of $-\frac{\dot{\mathrm{D}}_2}{T(T-1)}$ and that the split panel jackknife estimator $\hat{\theta}_{\mathrm{SPJK}} = 2\hat{\theta}_{\mathrm{PI}} - \frac{1}{2}(\hat{\theta}_{\mathrm{PI},1} + \hat{\theta}_{\mathrm{PI},2})$ has a bias of $-\frac{2\dot{\mathrm{D}}_2}{T^2}$, both of which shrink faster to zero than $\frac{\dot{\mathrm{D}}_1}{T}$ 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 $\hat{\theta}_{\mathrm{PI}}$ 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 2 and 3 when (i) and/or (ii) do not hold. Finally we note that $\hat{\theta}_{\mathrm{PI}}$ (a quadratic function) need not be asymptotically linear as is evident from the non-normal asymptotic distribution of $\hat{\theta}$ derived in Theorem 3 of this paper.

Examples of Jackknife Failure

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

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

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:

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

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

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 we see that $1/T \to 0$ 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 in the next example which adds the "random coefficients" of Example 3.

Example 3 (Special case)**.** Consider the model

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

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:

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

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

$$(x_1, x_2, \dots, x_T) = (-1, 2, 0, \dots, 0, -1)$$

renders the panel jackknife corrections inconsistent for small or large T:

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

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 $\frac{\sigma^2}{6}$).

Computations For this special case of example 2 we have that $A = \frac{I_N}{N}$ and $S_{xx} = TI_N$ so that $\tilde{A} = \frac{I_N}{NT}$ and $\mathrm{trace}(\tilde{A}^2) = \frac{1}{NT^2} = o(1)$ which implies consistency of $\hat{\theta}$. Similarly we have that the bias of $\tilde{\theta}$ is

$$\frac{1}{n}\sum_{g=1}^{N} T_g \mathbb{V}[\hat{\alpha}_g] = \frac{1}{n}\sum_{g=1}^{N}\sigma^2 = \frac{\sigma^2}{T} \qquad \text{where } \hat{\alpha}_g = \frac{1}{T_g}\sum_{t=1}^{T_g} y_{gt}.$$

The same types of calculations lead to the other biases reported in the paper.

For this special case of example 3 we have that $A = \begin{bmatrix} 0 & 0 \\ 0 & \frac{I_N}{N} \end{bmatrix}$ and $S_{xx} = \begin{bmatrix} TI_N & 0 \\ 0 & I_N \sum_{t=1}^{T} x_t^2 \end{bmatrix}$ which implies that $\text{trace}(\tilde{A}^2) = \frac{1}{N\left(\sum_{t=1}^{T} x_t^2\right)^2} = o(1)$ and therefore consistency of $\hat{\theta}$. Similarly we have that the bias of $\tilde{\theta}$ is

$$\frac{1}{n}\sum_{g=1}^{N} T_g \mathbb{V}[\hat{\delta}_g] = \frac{\sigma^2}{\sum_{t=1}^{T} x_t^2} \qquad \text{where } \hat{\delta}_g = \frac{\sum_{t=1}^{T_g} x_t y_{gt}}{\sum_{t=1}^{T} x_t^2}.$$

The same types of calculations lead to the other biases reported above. Now for the numerical example $(x_1, x_2, \ldots, x_T) = (-1, 2, 0, \ldots, 0, -1)$ we have $\sum_{t=1}^{T} x_t^2 = 6$, $\sum_{t=T/2+1}^{T}(x_t - \bar{x}_2)^2 = 1 - \frac{2}{T}$, $\sum_{t=1}^{T/2}(x_t - \bar{x}_1)^2 = 2\sum_{t=1}^{T/2} x_t^2 - T\bar{x}_1^2 = 5 - \frac{2}{T}$, and

$$\sum_{s\neq t}(x_s - \bar{x}_{-t})^2 = \begin{cases} 2 - \frac{4}{T-1} & \text{if } t = 2, \\ 5 - \frac{1}{T-1} & \text{if } t \in \{1, T\}, \\ 6 & \text{otherwise,} \end{cases}$$

Thus

$$\begin{aligned} \mathbb{E}[\hat{\theta}_{\text{PJK}}] - \theta &= \frac{T\sigma^2}{\sum_{t=1}^{T} x_t^2} - \sigma^2 \frac{(T-1)}{T} \sum_{t=1}^{T} \frac{1}{\sum_{s\neq t}(x_s - \bar{x}_{-t})^2} \\ &= \sigma^2 \frac{T}{6} - \sigma^2 \frac{T-1}{T}\left(\frac{2}{5 - \frac{1}{T-1}} + \frac{1}{2 - \frac{4}{T-1}} + \frac{T-3}{6}\right) \\ &= \sigma^2\left(\frac{2}{3} - \frac{4}{6T} - \frac{T-1}{T}\frac{2}{5 - \frac{1}{T-1}} - \frac{T-1}{T}\frac{1}{2 - \frac{4}{T-1}}\right) = -\frac{7}{30}\sigma^2 + O\left(\frac{1}{T}\right) \\ \text{and} \quad \mathbb{E}[\hat{\theta}_{\text{SPJK}}] - \theta &= \frac{2\sigma^2}{\sum_{t=1}^{T} x_t^2} - \frac{\sigma^2}{2\sum_{t=1}^{T/2}(x_t - \bar{x}_1)^2} + \frac{\sigma^2}{2\sum_{t=T/2+1}^{T}(x_t - \bar{x}_2)^2} \\ &= \sigma^2\left(\frac{1}{3} - \frac{1}{10 - \frac{4}{T}} - \frac{1}{2 - \frac{4}{T}}\right) = -\frac{8}{30}\sigma^2 + O\left(\frac{1}{T}\right). \end{aligned}$$

C.1.4 Finite Sample Properties

Here we provide a restatement and proof of Lemmas 1 and 2 together with a characterization of the finite sample distribution of $\hat{\theta}$ which was excluded from the main text.

Lemma C.3. *Recall that* $\theta^* = \hat{\beta}' A \hat{\beta} - \sum_{i=1}^n B_{ii}\sigma_i^2$.

1. *If* $\max_i P_{ii} < 1$, *then* $\mathbb{E}[\hat{\theta}] = \theta$.
2. *Unbiased estimators of* $\theta = \beta' A \beta$ *exist for all* A *if and only if* $\max_i P_{ii} < 1$.
3. *If* $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, *then* $\theta^* = \sum_{\ell=1}^r \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell]\right)$ *and* $\hat{b} \sim \mathcal{N}\left(b, \mathbb{V}[\hat{b}]\right)$.
4. *If* $\max_i P_{ii} < 1$ *and* $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, *then* $\hat{\theta} = \sum_{\ell=1}^{r_C} \lambda_\ell(C)\left(\hat{y}_\ell^2 - V_{\ell\ell}\right)$ *where* $\hat{y} \sim \mathcal{N}(\mu, V)$, $\mu = Q_C' X\beta$, $V = Q_C' \Sigma Q_C$, $C = (C_{i\ell})_{i,\ell}$, $\Sigma = diag(\sigma_1^2, \ldots, \sigma_n^2)$, *and* $C = Q_C D_C Q_C'$ *is a spectral decomposition of* C *such that* $D_C = diag(\lambda_1(C), \ldots, \lambda_{r_C}(C)$ *and* r_C *is the rank of* C.

Proof. First note that $\hat{\beta}' A \hat{\beta} = \sum_{i=1}^n \sum_{\ell=1}^n B_{i\ell} y_i y_\ell$ and $\hat{\sigma}_i^2 = y_i(y_i - x_i'\hat{\beta}_{-i}) = y_i M_{ii}^{-1} \sum_{\ell=1}^n M_{i\ell} y_\ell$, so

$$\begin{aligned}\hat{\theta} &= \sum_{i=1}^n \sum_{\ell=1}^n B_{i\ell} y_i y_\ell - B_{ii} M_{ii}^{-1} M_{i\ell} y_i y_\ell \\ &= \sum_{i=1}^n \sum_{\ell=1}^n \left(B_{i\ell} - 2^{-1} M_{i\ell}\left(B_{ii} M_{ii}^{-1} + B_{\ell\ell} M_{\ell\ell}^{-1}\right)\right) y_i y_\ell = \sum_{i=1}^n \sum_{\ell\neq i} C_{i\ell} y_i y_\ell.\end{aligned}$$

The errors are mean zero and uncorrelated across observations, so

$$\mathbb{E}[\hat{\theta}] = \sum_{i=1}^n \sum_{\ell\neq i} C_{i\ell} x_i'\beta x_\ell'\beta = \sum_{i=1}^n \sum_{\ell=1}^n B_{i\ell} x_i'\beta x_\ell'\beta - B_{ii} M_{ii}^{-1} M_{i\ell} x_i'\beta x_\ell'\beta = \theta,$$

since $\sum_{i=1}^n \sum_{\ell=1}^n B_{i\ell} x_i x_\ell' = A$ and $\sum_{\ell=1}^n M_{i\ell} x_\ell = 0$. This shows the first claim of the lemma.

It suffices to show that no unbiased estimator of $\beta' S_{xx}\beta$ exist when $\max_i P_{ii} = 1$. Any potential unbiased estimator must have the representation $y'Dy + U$ where $\mathbb{E}[U] = 0$ and $D = (D_{i\ell})_{i,\ell}$ satisfies (i) $D_{ii} = 0$ for all i and (ii) $X'DX = S_{xx}$ for $X = (x_1, \ldots, x_n)'$. (ii) implies that D must be $D = I + P\tilde{D}M + M\tilde{D}P + M\tilde{D}M$ for some $\tilde{D}$ where $P = (P_{i\ell})_{i,\ell}$ and $M = (M_{i\ell})_{i,\ell}$. If the exist a i with $P_{ii} = 1$, then $\sum_{\ell=1}^n P_{i\ell}^2 = P_{ii}$ yields $M_{i\ell} = 0$ for all ℓ which implies that D_{ii} must equal 1 to satisfy (ii). However, this makes it impossible to satisfy (i). This shows the second claim.

Recall the spectral decomposition $\tilde{A} = QDQ'$ and definition of $\hat{b} = Q' S_{xx}^{1/2}\hat{\beta}$ which satisfies that $\hat{b} \sim \mathcal{N}(b, \mathbb{V}[\hat{b}])$ when $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$. We have that $\theta^* = \sum_{\ell=1}^r \lambda_\ell \left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell]\right)$ since

$$\hat{\beta}' A \hat{\beta} = \hat{\beta}' S_{xx}^{1/2} \tilde{A} S_{xx}^{1/2} \hat{\beta} = \hat{b}' D \hat{b} = \sum_{\ell=1}^r \lambda_\ell \hat{b}_\ell^2,$$

and

$$\sum_{i=1}^{n} B_{ii}\sigma_i^2 = \text{trace}(B\Sigma) = \text{trace}(A\mathbb{V}[\hat{\beta}]) = \text{trace}(D\mathbb{V}[\hat{b}]) = \sum_{\ell=1}^{r} \lambda_\ell \mathbb{V}[\hat{b}_\ell].$$

where $B = (B_{i\ell})_{i,\ell}$. This shows the third claim.

The matrix C is is well-defined as $\max_i P_{ii} < 1$. Define $\hat{y} = Q_C'(y_1, \dots, y_n)'$ which satisfies that $\hat{y} \sim \mathcal{N}(\mu, V)$ when $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$. Furthermore,

$$\hat{\theta} = y'Cy = \hat{y}'D_C\hat{y} = \sum_{\ell=1}^{r_C} \lambda_\ell(C)\hat{y}_\ell^2,$$

and $C_{ii} = 0$ for all i, so that $\sum_\ell \lambda_\ell(C) V_{\ell\ell} = \text{trace}(C\Sigma) = 0$. This shows the last claim. □

C.1.5 Consistency

The next result provides a restatement and proof of Lemma 3.

Lemma C.4. *If Assumption 1 and one of the following conditions hold, then* $\hat{\theta} - \theta \xrightarrow{p} 0$.

(i) A *is positive semi-definite,* $\theta = \beta' A \beta = O(1)$*, and* $\text{trace}(\tilde{A}^2) = \sum_{\ell=1}^{r} \lambda_\ell^2 = o(1)$.

(ii) $A = \frac{1}{2}(A_1'A_2 + A_2'A_1)$ *where* $\theta_1 = \beta' A_1' A_1 \beta$, $\theta_2 = \beta' A_2' A_2 \beta$ *satisfy (i).*

Proof. Suppose that A is positive semi-definite. The difference between $\hat{\theta}$ and θ is

$$\hat{\theta} - \theta = 2\sum_{i=1}^{n}\sum_{\ell=1}^{n} B_{i\ell} x_\ell' \beta \varepsilon_i + \sum_{i=1}^{n}\sum_{\ell \neq i} B_{i\ell}\varepsilon_i\varepsilon_\ell + \sum_{i=1}^{n} B_{ii}(\varepsilon_i^2 - \hat{\sigma}_i^2),$$

and each term has mean zero so we show that their variances are small in large samples. The variance of the first term is

$$4\sum_{i=1}^{n}\left(\sum_{\ell=1}^{n} B_{i\ell} x_\ell'\beta\right)^2 \sigma_i^2 \leq 4\max_i \sigma_i^2 \beta' X' B^2 X \beta = 4 \max_i \sigma_i^2 \beta' A S_{xx}^{-1} A \beta \leq 4 \max_i \sigma_i^2 \theta \lambda_1 = o(1)$$

where $B = (B_{i\ell})_{i,\ell}$, the last inequality follows from positive semi-definiteness of A, and the last equality follows from $\theta = O(1)$ and $\lambda_1 \leq \text{trace}(\tilde{A}^2)^{1/2} = o(1)$. The variance of the second term is

$$2\sum_{i=1}^{n}\sum_{\ell \neq i} B_{i\ell}^2 \sigma_i^2 \sigma_\ell^2 \leq 2\max_i \sigma_i^4 \sum_{i=1}^{n}\sum_{\ell=1}^{n} B_{i\ell}^2 = 2\max_i \sigma_i^4 \text{trace}(\tilde{A}^2) = o(1).$$

Finally, the variance of the third term is

$$\sum_{i=1}^{n}\left(\sum_{\ell=1}^{n} M_{ll}^{-1} B_{\ell\ell} M_{i\ell} x_\ell' \beta\right)^2 \sigma_i^2 + 2\sum_{i=1}^{n}\sum_{\ell\neq i} M_{ii}^{-2} B_{ii}^2 M_{i\ell}^2 \sigma_i^2 \sigma_\ell^2$$
$$\leq \frac{1}{c^2}\max_i \sigma_i^2 \max_i (x_i'\beta)^2 \sum_{i=1}^{n} B_{ii}^2 + \frac{2}{c}\max_i \sigma_i^4 \sum_{i=1}^{n} B_{ii}^2 = o(1)$$

where $\min_i M_{ii} \geq c > 0$ and $\sum_{i=1}^n B_{ii}^2 \leq \text{trace}(\tilde{A}^2) = o(1)$. This shows the first claim of the lemma.

When A is non-definite, we write $A = \frac{1}{2}\left(A_1'A_2 + A_2'A_1\right)$ and note that

$$\beta' A S_{xx}^{-1} A\beta \leq \frac{1}{2}\left(\Theta_1 \lambda_{\max}(\tilde{A}_2) + \Theta_2 \lambda_{\max}(\tilde{A}_1)\right) \quad \text{and} \quad \text{trace}(\tilde{A}^2) \leq \text{trace}(\tilde{A}_1^2)^{1/2}\text{trace}(\tilde{A}_2^2)^{1/2}$$

where $\tilde{A}_\ell = S_{xx}^{-1/2} A_k' A_k S_{xx}^{-1/2}$ for $\ell = 1, 2$ and $\lambda_{\max}(\tilde{A}_2)$ is the largest eigenvalue of $\tilde{A}_2$. Thus consistency of $\hat{\theta}$ follows from $\Theta_1 = O(1)$, $\Theta_2 = O(1)$, $\text{trace}(\tilde{A}_1^2) = o(1)$, and $\text{trace}(\tilde{A}_2^2) = o(1)$. □

The next result provides a restatement and proof of Lemma 4.

Lemma C.5. *If Assumption 1, $n/p^4 = o(1)$, $\mathbb{V}[\hat{\theta}]^{-1} = O(n)$, and one of the following conditions hold, then $\mathbb{V}[\hat{\theta}]^{-1/2}(\hat{\theta}_{JLA} - \hat{\theta} - \mathrm{B}_p) = o_p(1)$ where $|\mathrm{B}_p| \leq \frac{1}{p}\sum_{i=1}^n P_{ii}^2 |B_{ii}| \sigma_i^2$.*

(i) A is positive semi-definite and $\mathbb{E}[\hat{\beta}' A \hat{\beta}] - \theta = O(1)$.

(ii) $A = \frac{1}{2}(A_1'A_2 + A_2'A_1)$ where $\theta_1 = \beta' A_1' A_1 \beta$, $\theta_2 = \beta' A_2' A_2 \beta$ satisfy (i) and $\frac{\mathbb{V}[\hat{\theta}_1]\mathbb{V}[\hat{\theta}_2]}{n\mathbb{V}[\hat{\theta}]^2} = O(1)$.

Proof. Define $\mathrm{B}_p = \frac{1}{p}\sum_{i=1}^n B_{ii}\sigma_i^2 \frac{2\sum_{\ell\neq i}^n P_{i\ell}^4 - P_{ii}^2(1-P_{ii})^2}{(1-P_{ii})^2}$. Letting $(\hat{\theta}_{JLA} - \hat{\theta})_2$ be a second order approximation of $\hat{\theta}_{JLA} - \hat{\theta}$, we first show that $\mathbb{E}\left[(\hat{\theta}_{JLA} - \hat{\theta})_2\right] = \mathrm{B}_p$ and $\frac{\mathbb{V}[(\hat{\theta}_{JLA} - \hat{\theta})_2]}{\mathbb{V}[\hat{\theta}]} = O(\frac{1}{p})$. 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

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

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

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

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

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

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

$$\mathbb{V}\left[\sum_{i=1}^n \hat{\sigma}_i^2(B_{ii} - \hat{B}_{ii})\right] = \mathbb{E}\left[\mathbb{V}\left[\sum_{i=1}^n \hat{\sigma}_i^2 \hat{B}_{ii} \,|\, y\right]\right] \leq \frac{2}{p}\sum_{i=1}^n\sum_{\ell=1}^n B_{i\ell}^2 \mathbb{E}\left[\hat{\sigma}_i^2\hat{\sigma}_\ell^2\right] = O\left(\frac{\mathrm{trace}(\tilde{A}^2)}{p}\right),$$

$$\begin{aligned}
\mathbb{V}\left[\sum_{i=1}^n \hat{\sigma}_i^2 \hat{B}_{ii}\hat{a}_i\right] &= \mathbb{E}\left[\mathbb{V}\left[\sum_{i=1}^n \hat{\sigma}_i^2 \hat{B}_{ii}\hat{a}_i \,|\, y, R_B\right]\right] \leq \frac{2}{p}\sum_{i=1}^n\sum_{\ell=1}^n P_{i\ell}^2 \tfrac{\mathbb{E}\left[\hat{B}_{ii}\hat{B}_{\ell\ell}\right]\mathbb{E}\left[\hat{\sigma}_i^2\hat{\sigma}_\ell^2\right]}{(1-P_{ii})(1-P_{\ell\ell})} \\
&= O\left(\frac{\mathrm{trace}(\tilde{A}^2)}{p} + \frac{\mathrm{trace}(\tilde{A}_1^2)^{1/2}\mathrm{trace}(\tilde{A}_2^2)^{1/2}}{p^2}\right)
\end{aligned}$$

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

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

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

We now treat the approximation error while utilizing that $\mathbb{E}[\hat{a}_i^3] = O\left(\frac{1}{p^2}\right)$, $\mathbb{E}[\hat{a}_i^4] = O\left(\frac{1}{p^2}\right)$, 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

$$\sum_{i=1}^n \hat\sigma_i^2 \hat B_{ii} \hat a_i^3 + \sum_{i=1}^n \hat\sigma_i^2 \hat B_{ii} \hat a_i^4 = O_p\left(\frac{\mathbb{E}[\hat\Theta_{1,\text{PI}} - \Theta_1] + \mathbb{E}[\hat\Theta_{2,\text{PI}} - \Theta_2]}{p^2}\right)$$

$$\sum_{i=1}^n \hat\sigma_i^2 \hat B_{ii} \frac{\hat a_i^5}{1+\hat a_i} = O_p\left(\frac{\log(n)}{\sqrt{p}} \frac{\mathbb{E}[\hat\Theta_{1,\text{PI}} - \Theta_1] + \mathbb{E}[\hat\Theta_{2,\text{PI}} - \Theta_2]}{p^2}\right)$$

$$\frac{1}{p}\sum_{i=1}^n \hat\sigma_i^2 \hat B_{ii} \frac{3\hat P_{ii}^2 + \hat P_{ii}^2 - (3P_{ii}^2 + P_{ii}^2)(1+\hat a_i)^2}{(1+\hat a_i)^2(1-P_{ii})} = O_p\left(\left(1 + \frac{\log(n)}{\sqrt{p}}\right) \frac{\mathbb{E}[\hat\Theta_{1,\text{PI}} - \Theta_1] + \mathbb{E}[\hat\Theta_{2,\text{PI}} - \Theta_2]}{p^2}\right)$$

which finishes the proof. □

C.2 Examples

All mathematical discussions of the examples are collected in Appendix C.7.

C.3 Quadratic Forms of Fixed Rank

The next result provides a restatement and proof of Theorem 1.

Theorem C.1. *If Assumption 1 holds, r is fixed, and $\max_i w_i' w_i = o(1)$, then*

1. $\mathbb{V}[\hat b]^{-1/2}(\hat b - b) \xrightarrow{d} \mathcal{N}(0, I_r)$ *where* $b = Q' S_{xx}^{1/2} \beta$,
2. $\mathbb{V}[\hat b]^{-1} \hat{\mathbb{V}}[\hat b] \xrightarrow{p} I_r$,
3. $\hat\theta = \sum_{\ell=1}^r \lambda_\ell \left(\hat b_\ell^2 - \hat{\mathbb{V}}[\hat b_\ell]\right) + o_p(\mathbb{V}[\hat\theta]^{1/2})$,

Proof. The proof has two steps: First, we write $\hat\theta$ as $\sum_{\ell=1}^r \lambda_\ell \left(\hat b_\ell^2 - \hat{\mathbb{V}}[\hat b_\ell]\right)$ plus an approximation error which is of smaller order than $\mathbb{V}[\hat\theta]$. This argument establishes the last two claims of the lemma. Second, we use Lyapounov's CLT to show that $\hat b \in \mathbb{R}^r$ is jointly asymptotically normal.
Decomposition and Approximation From the proof of Lemma 2 it follows that

$$\hat\theta = \sum_{\ell=1}^r \lambda_\ell \left(\hat b_\ell^2 - \hat{\mathbb{V}}[\hat b_\ell]\right) + \sum_{i=1}^n B_{ii}(\sigma_i^2 - \hat\sigma_i^2)$$

where we now show that the mean zero random variable $\sum_{i=1}^n B_{ii}(\sigma_i^2 - \hat\sigma_i^2)$ is $o_p(\mathbb{V}[\hat\theta]^{1/2})$.

We have

$$\sum_{i=1}^n B_{ii}(\hat\sigma_i^2 - \sigma_i^2) = \sum_{i=1}^n B_{ii} \sum_{\ell=1}^n M_{ii}^{-1} x_i'\beta M_{i\ell}\varepsilon_\ell + \sum_{i=1}^n B_{ii}(\varepsilon_i^2 - \sigma_i^2) + \sum_{i=1}^n B_{ii} \sum_{\ell\neq i} M_{ii}^{-1} M_{i\ell}\varepsilon_i\varepsilon_\ell.$$

The variances of these three terms are

$$\sum_{\ell=1}^{n}\sigma_\ell^2\left(\sum_{i=1}^{n}M_{i\ell}B_{ii}M_{ii}^{-1}x_i'\beta\right)^2 \le \max_i \sigma_i^2\sum_{i=1}^{n}B_{ii}^2M_{ii}^{-2}(x_i'\beta)^2 \le \max_i\sigma_i^2\max_i(x_i'\beta)^2M_{ii}^{-2}\times\sum_{i=1}^{n}B_{ii}^2,$$

$$\sum_{i=1}^{n}B_{ii}^2\mathbb{V}[\varepsilon_i^2] \le \max_i\mathbb{E}[\varepsilon_i^4]\times\sum_{i=1}^{n}B_{ii}^2,$$

$$\sum_{i=1}^{n}\sum_{\ell\neq i}\left(B_{ii}^2M_{ii}^{-2}+B_{ii}M_{ii}^{-1}B_{\ell\ell}M_{\ell\ell}^{-1}\right)M_{i\ell}^2\sigma_i^2\sigma_\ell^2 \le 2\max_i\sigma_i^4M_{ii}^{-2}\times\sum_{i=1}^{n}B_{ii}^2.$$

Furthermore, we have that

$$\mathbb{V}[\hat\theta]^{-1}\sum_{i=1}^{n}B_{ii}^2 \le \max_i w_i'w_i\mathbb{V}[\hat\theta]^{-1}\sum_{l=1}^{r}\lambda_l^2(\tilde A) \le \max_i w_i'w_i\max_i\sigma_i^{-4}=o(1),$$

so each of the three variances are of smaller order than $\mathbb{V}[\hat\theta]$.

For the second claim it suffices to show that $\delta(v) := \frac{\hat{\mathbb{V}}[v'\hat b]-\mathbb{V}[v'\hat b]}{\mathbb{V}[v'\hat b]} = o_p(1)$ for all nonrandom $v\in\mathbb{R}^r$ with $v'v=1$. Let $v\in\mathbb{R}^r$ be nonrandom with $v'v=1$. As above we have that $\delta(v) = \sum_{i=1}^{n}w_i(v)(\hat\sigma_i^2-\sigma_i^2)$ is a mean zero variable which is $o_p(1)$ if $\sum_{i=1}^{n}w_i(v)^4=o(1)$ where $w_i(v) = \frac{(v'w_i)^2}{\sum_{i=1}^{n}\sigma_i^2(v'w_i)^2}$. But this follows from

$$\sum_{i=1}^{n}w_i(v)^4 \le \max_i\sigma_i^{-4}\max_i w_i'w_i = o(1)$$

where the inequality is implied by $\max_i w_i'w_i = o(1)$, $v'v=1$, and $\sum_{i=1}^{n}w_iw_i' = I_r$.

Asymptotic Normality Next we show that all linear combinations of $\hat b$ are asymptotically normal. Let $v\in\mathbb{R}^r$ be a non-random vector with $v'v=1$. Lyapunov's CLT implies that $\mathbb{V}[v'\hat b]^{-1/2}v'(\hat b-b)\xrightarrow{d} N(0,1)$ if

$$\mathbb{V}[v'\hat b]^{-2}\sum_{i=1}^{n}\mathbb{E}[\varepsilon_i^4](v'Q'S_{xx}^{-1/2}x_i)^4 = \mathbb{V}[v'\tilde\beta]^{-2}\sum_{i=1}^{n}\mathbb{E}[\varepsilon_i^4](v'w_i)^4 = o(1). \tag{14}$$

We have that $\max_i w_i'w_i = o(1)$ implies (14) since $\max_i(v'w_i)^2\le\max_i w_i'w_i$ and

$$\sum_{i=1}^{n}(v'w_i)^2 = 1, \qquad \mathbb{V}[v'\tilde\beta]^{-1}\le\max_i\sigma_i^{-2}=O(1), \qquad \max_i\mathbb{E}[\varepsilon_i^4]=O(1),$$

by definition of w_i and Assumption 1. □

C.4 Quadratic Forms of Growing Rank

This appendix provides restatements and proofs of Theorems 2 and 3. The proofs relies on an auxiliary lemma which extends a central limit theorem given in Sølvsten (2019).

C.4.1 A Central Limit Theorem

The proofs of Theorem 2 and Theorem 3 is based on the following lemma. Let $\{v_{n,i}\}_{i,n}$ be a triangular array of row-wise independent random variables with $\mathbb{E}[v_{n,i}] = 0$ and $\mathbb{V}[v_{n,i}] = \sigma_{n,i}^2$, let $\{\dot{w}_{n,i}\}_{i,n}$ be a triangular array of non-random weights that satisfy $\sum_{i=1}^n \dot{w}_{n,i}^2 \sigma_{n,i}^2 = 1$ for all n, and let $(W_n)_n$ be a sequence of symmetric non-random matrices in $\mathbb{R}^{n\times n}$ with zeroes on the diagonal that satisfy $2\sum_{i=1}^n \sum_{\ell\neq i} W_{n,i\ell}^2 \sigma_{n,i}^2 \sigma_{n,\ell}^2 = 1$. For simplicity, we drop the subscript n on $v_{n,i}$, $\sigma_{n,i}^2$, $\dot{w}_{n,i}$ and W_n. Define

$$\mathcal{S}_n = \sum_{i=1}^n \dot{w}_i v_i \quad \text{and} \quad \mathcal{U}_n = \sum_{i=1}^n \sum_{\ell\neq i} W_{i\ell} v_i v_\ell.$$

Lemma C.6. *If* $\max_i \mathbb{E}[v_i^4] + \sigma_i^{-2} = O(1)$,

$$(i)\ \max_i \dot{w}_i^2 = o(1), \qquad (ii)\ \lambda_{\max}(W^2) = o(1),$$

then $(\mathcal{S}_n, \mathcal{U}_n)' \xrightarrow{d} \mathcal{N}(0, I_2)$.

This lemma extends the main result of Appendix A2 in Sølvsten (2019) to allow for $\{v_i\}_i$ to be an array of non-identically distributed variables and presents the conclusion in a way that is tailored to the application in this paper. The proof requires no substantially new ideas compared to Sølvsten (2019), but we give it at the end of the next section for completeness.

C.4.2 Limit Distributions

Theorem C.2. *If*

$$(i)\ \mathbb{V}[\hat{\theta}]^{-1} \max_i \left((\tilde{x}_i'\beta)^2 + (\check{x}_i'\beta)^2 \right) = o(1), \quad (ii)\ \frac{\lambda_1^2}{\sum_{\ell=1}^r \lambda_\ell^2} = o(1),$$

and Assumption 1 holds, then $\mathbb{V}[\hat{\theta}]^{-1/2}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, 1)$.

Proof. The proof involves two steps: First, we decompose $\hat{\theta}$ into a weighted sum of two terms of the type described in Lemma C.6. Second, we use Lemma C.6 to show joint asymptotic normality of the two terms. The conclusion that $\hat{\theta}$ is asymptotically normal is immediate from there.

Decomposition The difference between $\hat{\theta}$ and θ is

$$\hat{\theta} - \theta = \sum_{i=1}^{n} \left(2\tilde{x}_i'\beta - \check{x}_i'\beta\right) \varepsilon_i + \sum_{i=1}^{n} \sum_{\ell \neq i} C_{i\ell}\varepsilon_i\varepsilon_\ell,$$

where these two terms are uncorrelated and have variances

$$V_{\mathcal{S}} = \sum_{i=1}^{n} (2\tilde{x}_i'\beta - \check{x}_i'\beta)^2 \sigma_i^2 \quad \text{and} \quad V_{\mathcal{U}} = 2\sum_{i=1}^{n} \sum_{\ell \neq i} C_{i\ell}^2 \sigma_i^2 \sigma_\ell^2.$$

Thus we write $\mathbb{V}[\hat{\theta}]^{-1/2}(\hat{\theta} - \theta) = \omega_1 \mathcal{S}_n + \omega_2 \mathcal{U}_n$ where

$$\mathcal{S}_n = V_{\mathcal{S}}^{-1/2} \sum_{i=1}^{n} \left(2\tilde{x}_i'\beta - \check{x}_i'\beta\right) \varepsilon_i, \qquad \mathcal{U}_n = V_{\mathcal{U}}^{-1/2} \sum_{i=1}^{n} \sum_{\ell \neq i} C_{i\ell}\varepsilon_i\varepsilon_\ell,$$

$$\omega_1 = \sqrt{V_{\mathcal{S}}/\mathbb{V}[\hat{\theta}]}, \qquad \omega_2 = \sqrt{V_{\mathcal{U}}/\mathbb{V}[\hat{\theta}]}.$$

Asymptotic Normality We will argue along converging subsequences. Move to a subsequence where ω_1 converges. If the limit is zero, then $\mathbb{V}[\hat{\theta}]^{-1/2}(\hat{\theta} - \theta) = \omega_2 \mathcal{U}_n + o_p(1)$ and so it follows from Result $C.2$ below and Theorem 2(ii) that $\hat{\theta}$ is asymptotically normal. Thus we consider the case where the limit of ω_1 is nonzero.

In the notation of Lemma $C.6$ we have

$$\dot{w}_i = \frac{\left(2\tilde{x}_i'\beta - \check{x}_i'\beta\right)}{V_{\mathcal{S}}^{1/2}} \quad \text{and} \quad W_{i\ell} = \frac{C_{i\ell}}{V_{\mathcal{U}}^{1/2}}.$$

For Lemma $C.6$(i) we have

$$\max_i \dot{w}_i^2 \leq 4\omega_1^{-1} \max_i \frac{(\tilde{x}_i'\beta)^2 + (\check{x}_i'\beta)^2}{\mathbb{V}[\hat{\theta}]} = o(1),$$

where the last equality follows from Theorem 2(i) and the nonzero limit of ω_1.

For Lemma $C.6$(ii) we show instead that $\text{trace}(W^4) = o(1)$. It can be shown that for all n, $\text{trace}(C^4) \leq c_U \cdot \text{trace}(B^4) = c_U \cdot \text{trace}(\tilde{A}^4) \leq c_U \lambda_1^2 \cdot \text{trace}(\tilde{A}^2)$ and $V_{\mathcal{U}} \geq c_L \min_i \sigma_i^4 \cdot \text{trace}(\tilde{A})$, where the finite and nonzero constants c_U and c_L do not depend on n (but depend on $\min_i M_{ii}$ which is bounded away from zero). Thus, Assumption 1 implies that

$$\text{trace}(W^4) \leq \frac{c_U \lambda_1^2 \cdot \text{trace}(\tilde{A}^2)}{(c_L \min_i \sigma_i^4 \cdot \text{trace}(\tilde{A}^2))^2} = O\left(\frac{\lambda_1^2}{\text{trace}(\tilde{A}^2)}\right) = o(1)$$

where the last equality follows from Theorem 2(ii). $\square$

Theorem C.3. *If* $\max_i \mathsf{w}_{iq}'\mathsf{w}_{iq} = o(1)$, $\mathbb{V}[\hat{\theta}_q]^{-1}\max_i\left((\tilde{x}_{iq}'\beta)^2 + (\check{x}_{iq}'\beta)^2\right) = o(1)$, *and Assumptions* 1 *and* 2 *holds, then*

1. $\mathbb{V}[(\hat{\mathsf{b}}_q', \hat{\theta}_q)']^{-1/2}\left((\hat{\mathsf{b}}_q', \hat{\theta}_q)' - \mathbb{E}[(\hat{\mathsf{b}}_q', \hat{\theta}_q)']\right) \xrightarrow{d} \mathcal{N}\left(0, I_{q+1}\right)$
2. $\hat{\theta} = \sum_{\ell=1}^{q}\lambda_\ell\left(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell]\right) + \hat{\theta}_q + o_p(\mathbb{V}[\hat{\theta}]^{1/2})$

for

$$\mathbb{V}[(\hat{\mathsf{b}}_q', \hat{\theta}_q)'] = \sum_{i=1}^{n}\begin{bmatrix} \mathsf{w}_{iq}\mathsf{w}_{iq}'\sigma_i^2 & 2\mathsf{w}_{iq}\left(\sum_{\ell\neq i} C_{i\ell q}x_\ell'\beta\right)\sigma_i^2 \\ 2\mathsf{w}_{iq}'\left(\sum_{\ell\neq i} C_{i\ell q}x_\ell'\beta\right)\sigma_i^2 & 4\left(\sum_{\ell\neq i} C_{i\ell q}x_\ell'\beta\right)^2\sigma_i^2 + 2\sum_{\ell\neq i} C_{i\ell q}^2\sigma_i^2\sigma_\ell^2 \end{bmatrix},$$

$C_{i\ell q} = B_{i\ell q} - 2^{-1}M_{i\ell}\left(M_{ii}^{-1}B_{iiq} + M_{\ell\ell}^{-1}B_{\ell\ell q}\right)$, $B_{i\ell q} = x_i'S_{xx}^{-1/2}\tilde{A}_qS_{xx}^{-1/2}x_\ell$, $\tilde{A}_q = \sum_{\ell=q+1}^{r}\lambda_\ell q_\ell q_\ell'$, $\tilde{x}_{iq} = \sum_{\ell=1}^{n}B_{i\ell q}x_\ell$, *and* $\check{x}_{iq} = \sum_{\ell=1}^{n}M_{i\ell}M_{\ell\ell}^{-1}B_{\ell\ell q}x_\ell$.

Proof. The proof involves two steps: First, we write $\hat{\theta}$ as the sum of (1a) a quadratic function applied to $\hat{\mathsf{b}}_q$, (1b) an approximation error which is of smaller order than $\mathbb{V}[\hat{\theta}]$, and (2) a weighted sum of two terms, $\mathcal{S}_n$ and $\mathcal{U}_n$, of the type described in Lemma *C*.6. Second, we use Lemma *C*.6 to show that $(\hat{\mathsf{b}}_q', \mathcal{S}_n, \mathcal{U}_n)' \in \mathbb{R}^{q+2}$ is jointly asymptotically normal.

Decomposition and Approximation We have that

$$\hat{\theta} = \sum_{\ell=1}^{q}\lambda_\ell(\hat{b}_\ell^2 - \mathbb{V}[\hat{b}_\ell]) + \hat{\theta}_q + o_p(\mathbb{V}[\hat{\theta}]^{1/2}) \quad \text{for} \quad \hat{\theta}_q = \sum_{i=1}^{n}\sum_{\ell\neq i} C_{i\ell q}y_iy_\ell$$

since

$$\hat{\beta}'A\hat{\beta} = \sum_{\ell=1}^{q}\lambda_\ell\hat{b}_\ell^2 + \sum_{i=1}^{n}\sum_{\ell=1}^{n} B_{i\ell q}y_iy_\ell$$

and

$$\begin{aligned}\sum_{i=1}^{n} B_{ii}\hat{\sigma}_i^2 &= \sum_{i=1}^{n} B_{ii\mathbf{1}}\sigma_i^2 + \sum_{i=1}^{n} B_{iiq}\hat{\sigma}_i^2 + \sum_{i=1}^{n} B_{ii,-q}(\hat{\sigma}_i^2 - \sigma_i^2) \\ &= \sum_{\ell=1}^{q}\lambda_\ell\mathbb{V}[\hat{b}_\ell] + \sum_{i=1}^{n} B_{iiq}\hat{\sigma}_i^2 + o_p(\mathbb{V}[\hat{\theta}]^{1/2})\end{aligned}$$

where $B_{ii,-q} = B_{ii} - B_{iiq}$ and it follows from $\max_i \mathsf{w}_{iq}'\mathsf{w}_{iq} = o(1)$ and the calculations in the proof of Theorem 1 that the mean zero random variable $\sum_{i=1}^{n} B_{ii,-q}(\hat{\sigma}_i^2 - \sigma_i^2)$ is $o_p(\mathbb{V}[\hat{\theta}]^{1/2})$.

We will further center and rescale $\hat{\theta}_q$ by writing

$$\mathbb{V}[\hat{\theta}_q]^{-1/2}\left(\hat{\theta}_q - \mathbb{E}[\hat{\theta}_q]\right) = \omega_1\mathcal{S}_n + \omega_2\mathcal{U}_n$$

where

$$\mathcal{S}_n = V_{\mathcal{S}}^{-1/2} \sum_{i=1}^{n} \left(2\tilde{x}_{iq}'\beta - \check{x}_{iq}'\beta\right) \varepsilon_i, \qquad \mathcal{U}_n = V_{\mathcal{U}}^{-1/2} \sum_{i=1}^{n} \sum_{\ell \neq i} C_{i\ell q} \varepsilon_i \varepsilon_\ell,$$

$$V_{\mathcal{S}} = \sum_{i=1}^{n} \left(2\tilde{x}_{iq}'\beta - \check{x}_{iq}'\beta\right)^2 \sigma_i^2, \qquad V_{\mathcal{U}} = 2 \sum_{i=1}^{n} \sum_{\ell \neq i} C_{i\ell q}^2 \sigma_i^2 \sigma_\ell^2,$$

$$\omega_1 = \sqrt{V_{\mathcal{S}} / \mathbb{V}[\hat{\theta}_q]}, \qquad \omega_2 = \sqrt{V_{\mathcal{U}} / \mathbb{V}[\hat{\theta}_q]},$$

and $\mathcal{U}_n$ is uncorrelated with both $\mathcal{S}_n$ and $\hat{\mathsf{b}}_q$.

Asymptotic Normality As in the proof of Theorem 2, we will argue along converging subsequences and therefore move to a subsequence where ω_1 converges. If the limit is zero, then the conclusion of the theorem follows from Lemma *C*.6 applied to $(\mathbb{V}[v'\hat{\mathsf{b}}_q]^{-1/2}(v'\hat{\mathsf{b}}_q - \mathbb{E}[v'\hat{\mathsf{b}}_q]), \mathcal{U}_n)'$ for $v \in \mathbb{R}^q$ with $v'v = 1$. Thus we consider the case where the limit of ω_1 is nonzero.

Next we use Lemma *C*.6 to show that

$$\left(\frac{v'\hat{\mathsf{b}}_q - \mathbb{E}[v'\hat{\mathsf{b}}_q] + u\mathcal{S}_n}{\mathbb{V}[\hat{\mathsf{b}}_q + u\mathcal{S}_n]^{1/2}}, \mathcal{U}_n \right)' \xrightarrow{d} \mathcal{N}(0, I_2)$$

for any non-random $(v', u)' \in \mathbb{R}^{q+1}$ with $v'v + u^2 = 1$. In the notation of Lemma *C*.6 we have

$$\dot{w}_i = \frac{v'\mathsf{w}_{iq} + uV_{\mathcal{S}}^{-1/2}\left(2\tilde{x}_{iq}'\beta - \check{x}_{iq}'\beta\right)}{\mathbb{V}[\hat{\mathsf{b}}_q + u\mathcal{S}_n]^{1/2}} \quad \text{and} \quad W_{i\ell} = \frac{C_{i\ell q}}{V_{\mathcal{U}}^{1/2}}.$$

A simple calculation shows that $\mathbb{V}[v'\hat{\mathsf{b}}_q + u\mathcal{S}_n] \geq \min_i \sigma_i^2 \gg 0$, so $\max_i \dot{w}_i^2 = o(1)$ follows from Theorem 3(i), Theorem 3(ii), and ω_1 being bounded away from zero.

Similarly, we have as in the proof of Theorem 2 that

$$\text{trace}(C_q^4) \leq c\text{trace}(B_q^4) \leq c\lambda_{q+1}^2 \sum_{\ell=q+1}^{r} \lambda_\ell^2 \quad \text{and} \quad V_{\mathcal{U}}^2 \geq \omega_2^{-4} \min_i \sigma_i^8 \text{trace}(\tilde{A}^2)^2$$

for $C_q = (C_{i\ell q})_{i,\ell}$ and $B_q = (B_{i\ell q})_{i,\ell}$, so Assumptions 1 and 2 yield $\text{trace}(W^4) = o(1)$. □

C.4.3 Proof of a Central Limit Theorem

The proof of Lemma *C*.6 uses the notation and verifies the conditions of Lemmas A2.1 and A2.2 in Sølvsten (2019) 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 *C*.6.

Before starting we note that $\max_i \sigma_i^{-2} = O(1)$ and $2\sum_{i=1}^n \sum_{\ell \neq i} W_{i\ell}^2 \sigma_i^2 \sigma_\ell^2 = 1$ implies that

$\text{trace}(W^2) = \sum_{i=1}^n \sum_{\ell \neq i} W_{i\ell}^2 = O(1)$ and therefore that

$$\lambda_{\max}(W^2) = o(1) \Leftrightarrow \text{trace}(W^4) = o(1).$$

Marginal Distributions

Result C.1. $\max_i \mathbb{E}[v_i^4] + \sigma_i^{-2} = O(1)$, $\sum_{i=1}^n \dot{w}_i^2 \sigma_i^2 = 1$, *and Lemma C.6(i) implies that* $\mathcal{S}_n \xrightarrow{d} \mathcal{N}(0,1)$.

In the notation of SS2.1 we have,

$$\Delta_i^0 \mathcal{S}_n = \dot{w}_i v_i \quad \text{and} \quad E[T_n \,|\, V_n] = 1 + \tfrac{1}{2} \sum_{i=1}^n \dot{w}_i^2 (v_i^2 - \sigma_i^2),$$

and 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 C.6(i) that

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

so Result C.1 follows from SS2.1.

Result C.2. $\max_i \mathbb{E}[v_i^4] + \sigma_i^{-2} = O(1)$, $2\sum_{i=1}^n \sum_{\ell \neq i} W_{n,i\ell}^2 \sigma_{n,i}^2 \sigma_{n,\ell}^2 = 1$, *and Lemma C.6(ii) implies that* $\mathcal{U}_n \xrightarrow{d} \mathcal{N}(0,1)$.

In the notation of SS2.1 we have,

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

and

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

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

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

Interlude: Convergence in $\mathcal{L}^1$

a_n, b_n, and c_n are a linear sum, a quadratic sum, and a cubic sum. We will need to treat similar sums later, so we record some simple sufficient conditions for their convergence. 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. We 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 above. For the inferential results we will also treat quartic sums, so we provide the sufficient conditions here.

Result C.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,

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

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 $\Delta_i^0 S_{n3}$.

Marginal Distributions, Continued

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 C.3, let $\bar{W}_{i\ell}=\sum_{k=1}^{n}W_{ik}W_{k\ell}$ and note that $\mathrm{trace}(W^4)=\sum_{i=1}^{n}\sum_{\ell=1}^{n}\bar{W}_{i\ell}^2$. We have

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

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

Joint Distribution

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 C.6 follows if we show that $\mathcal{W}_n\xrightarrow{d}\mathcal{N}(0,1)$. In the notation of SS2.1 we have,

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

and

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

$$+ u_1 u_2 3 \sum_{i=1}^n \sum_{\ell \neq i} (v_i^2 + \sigma_i^2) \dot{w}_i W_{i\ell} v_j.$$

The proofs of Result $C.1$ and Result $C.2$ showed that

$$\sum_{i=1}^n \mathbb{E}[(\Delta_i^0 \mathcal{W}_n)^2] = O(1), \; \sum_{i=1}^n \mathbb{E}[(\Delta_i^0 \mathcal{W}_n)^4] = o(1)$$

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 so

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

$$\sum_{i=1}^n \sum_{\ell \neq i} \dot{w}_i^2 W_{i\ell}^2 \leq \sum_{i=1}^n \sum_{\ell \neq i} W_{i\ell}^2 \max_i \dot{w}_i^2 = O(\max_i \dot{w}_i^2)$$

ends the proof.

C.5 Asymptotic Variance Estimation

This appendix provides restatements and proofs of Lemmas 5 and 6 which establish consistency of the proposes standard error estimators that rely on sample splitting. Furthermore, it gives adjustments to those standard errors that guarantee existence whenever two independent unbiased estimators of $x_i'\beta$ cannot be formed. However, these adjustments may provide a somewhat conservative assessment of the uncertainty in $\hat{\theta}$ as further investigated in the simulations of Section 8.7.

Lemma C.7. *For $s = 1, 2$, suppose that $\widehat{x_i'\beta}_{-i,s} = \sum_{\ell \neq i}^n P_{i\ell,s} y_\ell$ satisfies $\sum_{\ell \neq i}^n P_{i\ell,s} x_\ell' \beta = x_i'\beta$, $P_{i\ell,1} P_{i\ell,2} = 0$ for all ℓ, and $\lambda_{\max}(P_s P_s') = O(1)$.*

1. *If the conditions of Theorem 2 hold and $|\mathcal{B}| = O(1)$, then $\frac{\hat{\theta} - \theta}{\hat{\mathbb{V}}[\hat{\theta}]^{1/2}} \xrightarrow{d} \mathcal{N}(0, 1)$.*

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

Proof. The proof continues in two steps: First, we show that $\hat{\mathbb{V}}[\hat{\theta}]$ has a positive bias which is of smaller order than $\mathbb{V}[\hat{\theta}]$ when $|\mathcal{B}| = O(1)$. Second, we show that $\hat{\mathbb{V}}[\hat{\theta}] - \mathbb{E}[\hat{\mathbb{V}}[\hat{\theta}]] = o_p(\mathbb{V}[\hat{\theta}])$. When combined with Theorem 2, these conclusions imply the two 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

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

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

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

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

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

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

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

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 C.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 C.6. 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 $\lambda_{\max}(P_s P_s') = O(1)$ 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

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

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

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

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

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

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

Section 5.2 proposed standard errors for the case of $q > 0$, but left a few details to the appendix since the definitions were completely analogous to the previous lemma. 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 $C_{i\ell q}$ was introduced in the proof of Theorem 3 and is of the form $C_{i\ell q} = B_{i\ell q} - 2^{-1}M_{i\ell}\left(M_{ii}^{-1}B_{iiq} + M_{\ell\ell}^{-1}B_{\ell\ell q}\right)$ for $B_{i\ell q} = B_{i\ell} - \sum_{s=1}^q \lambda_s w_{is}w_{\ell s}$.

Furthermore, the proposed standard error estimator relies on

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

Lemma C.8. *For $s = 1, 2$, suppose that $\widehat{x_i'\beta}_{-i,s}$ satisfies $\sum_{\ell\neq i}^n P_{i\ell,s}x_\ell'\beta = x_i'\beta$, $P_{i\ell,1}P_{i\ell,2} = 0$ 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 3 hold and $|\mathcal{B}| = O(1)$, then $\Sigma_q^{-1}\hat{\Sigma}_q \xrightarrow{p} I_{q+1}$.*
2. *If the conditions of Theorem 3 hold, then $\liminf_{n\to\infty} \mathbb{P}\left(\theta \in \hat{C}_{\alpha,q}^\theta\right) \geq 1 - \alpha$.*

The following provides a proof of the first claim of this lemma, while we postpone a proof of the second claim to the end of Appendix C.6.

Proof. 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 Theorem *C*.1 and Lemma *C*.7. 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 $\mathbb{E}[\varepsilon_\ell\tilde{\sigma}_i^2]$ 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:

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

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

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

These seven terms are $o_p(\mathbb{V}[v'\hat{\mathsf{b}}_q]^{1/2}\mathbb{V}[\hat{\theta}_q]^{1/2})$ by Result C.3 as outlined in the following.

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

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

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

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

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

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

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

C.5.1 Conservative Variance Estimation

The standard error estimators considered in the preceding two lemmas relied on existence of the independent and unbiased estimators $\widehat{x_i'\beta}_{-i,1}$ and $\widehat{x_i'\beta}_{-i,2}$. This part of the appendix creates 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 subsection.

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

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

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 defintion 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

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

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 4 and 6.

Lemma C.9. *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* $\liminf_{n\to\infty}\mathbb{P}\left(\theta \in C_\alpha^\theta(\hat{\Sigma}_{q,2})\right) \geq 1-\alpha$.

The following provides a proof of the first claim of this lemma, while we postpone a proof of the second claim to the end of Appendix C.6.

Proof. As in the proof of Lemma 5 it suffices 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

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

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 $\widehat{\sigma_i^2 \sigma_\ell^2}_2$. ☐

C.6 Inference with Nuisance Parameters

This Appendix starts by defining curvature and accompanying critical value for a given curvature as introduced in Section 6. Then it derives the closed form representation of $C_\alpha^\theta(\tilde{\Sigma}_1)$ for any variance matrix $\tilde{\Sigma}_1 \in \mathbb{R}^{2\times 2}$ where for general q we have

$$C_\alpha^\theta(\tilde{\Sigma}_q) = \left[\min_{(\dot{b}_1,\ldots,\dot{b}_q,\dot{\theta}_q)' \in \mathsf{E}_\alpha(\tilde{\Sigma}_q)} \sum_{\ell=1}^q \lambda_\ell \dot{b}_\ell^2 + \dot{\theta}_q, \max_{(\dot{b}_1,\ldots,\dot{b}_q,\dot{\theta}_q)' \in \mathsf{E}_\alpha(\tilde{\Sigma}_q)} \sum_{\ell=1}^q \lambda_\ell \dot{b}_\ell^2 + \dot{\theta}_q \right]$$

and

$$\mathsf{E}_\alpha(\tilde{\Sigma}_q) = \left\{ (\mathsf{b}_q', \theta_q)' \in \mathbb{R}^{q+1} : \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix}' \tilde{\Sigma}_q^{-1} \begin{pmatrix} \hat{\mathsf{b}}_q - \mathsf{b}_q \\ \hat{\theta}_q - \theta_q \end{pmatrix} \leq z^2_{\alpha,\kappa(\tilde{\Sigma}_q)} \right\}.$$

Finally, it proofs validity of $\hat{C}_{\alpha,q}^\theta = C_\alpha^\theta(\hat{\Sigma}_q)$ and $C_\alpha^\theta(\hat{\Sigma}_{q,2})$ for any fixed q. As for $\hat{\Sigma}_q$ and $\hat{\Sigma}_{q,2}$, we partition $\tilde{\Sigma}_q$ into $\tilde{\Sigma}_q = \begin{bmatrix} \tilde{\mathbb{V}}[\hat{\mathsf{b}}_q] & \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q, \hat{\theta}_q]' \\ \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q, \hat{\theta}_q] & \tilde{\mathbb{V}}[\hat{\theta}_q] \end{bmatrix}$ with $\tilde{\mathbb{V}}[\hat{\theta}_q] \in \mathbb{R}$. In Section 6, $\hat{C}_{\alpha,q}^\theta = C_\alpha^\theta(\hat{\Sigma}_q)$, $\hat{\mathsf{E}}_{\alpha,q} = \mathsf{E}_\alpha(\hat{\Sigma}_q)$, and $\hat{\kappa}_q = \kappa(\hat{\Sigma}_q)$.

C.6.1 Preliminaries

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 variates 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 The 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:

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

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 $P_{\dot{y}} = d\mathbf{x}(\dot{y})(d\mathbf{x}(\dot{y})' d\mathbf{x}(\dot{y}))^{-1} d\mathbf{x}(\dot{y})'$. See Andrews and Mikusheva (2016) for additional details.

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

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

and $D_q = \text{diag}(\lambda_1, \ldots, \lambda_q)$. 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

$$\kappa(\tilde{\Sigma}_q) = \frac{2\max_{u\in\mathbb{R}^q}\frac{|u'D_qu|}{u'\tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{-1}u}}{\left(\tilde{\mathbb{V}}[\hat{\theta}_q] - \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q,\hat{\theta}_q]'\tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{-1}\tilde{\mathbb{C}}[\hat{\mathsf{b}}_q,\hat{\theta}_q]\right)^{1/2}} = \frac{2|\dot{\vdots}\!\lambda_1(\tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{1/2}D_q\tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{1/2})|}{\left(\tilde{\mathbb{V}}[\hat{\theta}_q] - \tilde{\mathbb{C}}[\hat{\mathsf{b}}_q,\hat{\theta}_q]'\tilde{\mathbb{V}}[\hat{\mathsf{b}}_q]^{-1}\tilde{\mathbb{C}}[\hat{\mathsf{b}}_q,\hat{\theta}_q]\right)^{1/2}}$$

where $\dot{\vdots}\!\lambda_1(\cdot)$ is the eigenvalue of largest magnitude. This formula simplifies to the one derived above when $q = 1$.

C.6.2 Closed Form Representation of $C_\alpha^\theta(\tilde{\Sigma}_1)$

An implicit representation of $C_\alpha^\theta(\tilde{\Sigma}_1)$ 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

$$b_{1,\pm} = \hat{b}_1 \pm z_{\alpha,\kappa(\tilde{\Sigma}_1)}\left(\tilde{\mathbb{V}}[\hat{b}_1](1 - \tilde{a}(b_{1,\pm}))\right)^{1/2}, \tag{15}$$

$$\theta_{1,\pm} = \hat{\theta}_1 - \tilde{\rho}\frac{\tilde{\mathbb{V}}[\hat{\theta}_1]^{1/2}}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}}(\hat{b}_1 - b_{1,\pm}) \pm z_{\alpha,\kappa(\tilde{\Sigma}_1)}\left(\tilde{\mathbb{V}}[\hat{\theta}_1](1 - \tilde{\rho}^2)\tilde{a}(b_{1,\pm})\right)^{1/2} \tag{16}$$

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

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 (15),(16) 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 $C^{\theta}_{\alpha}(\tilde{\Sigma}_1)$ 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 (15). 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

$$\frac{(\hat{b}_1-\dot{b}_1)^2}{\tilde{\mathbb{V}}[\hat{b}_1]}\left(1+\left(\frac{\operatorname{sgn}(\lambda_1)\kappa(\tilde{\Sigma}_1)\dot{b}_1}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}} + \frac{\tilde{\rho}}{\sqrt{1-\tilde{\rho}^2}}\right)^2\right) = \left(\frac{\operatorname{sgn}(\lambda_1)\kappa(\tilde{\Sigma}_1)\dot{b}_1}{\tilde{\mathbb{V}}[\hat{b}_1]^{1/2}} + \frac{\tilde{\rho}}{\sqrt{1-\tilde{\rho}^2}}\right)^2 z^2_{\alpha,\kappa(\tilde{\Sigma}_1)} \quad (17)$$

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 (17). More importantly, the maximum is the upper end of

$C_\alpha^\theta(\tilde{\Sigma}_1)$. 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 (17) as a necessary condition for a minimum. Thus $b_{1,-}$ and the lower end of $C_\alpha^\theta(\tilde{\Sigma}_1)$ 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 (17).

C.6.3 Asymptotic Validity

Lemma C.10. *If $\Sigma_q^{-1}\hat{\Sigma}_q \xrightarrow{p} I_{q+1}$ and the conditions of Theorem 3 hold, then*

$$\liminf_{n\to\infty}\mathbb{P}\left(\theta\in\hat{C}^\theta_{\alpha,q}\right)\geq 1-\alpha.$$

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

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

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). □

Proof of the second claims in Lemmas C.8 and C.9. The proof contains two main parts. One part is to establish that the biases of $\hat{\Sigma}_q$ and $\hat{\Sigma}_{q,2}$ are positive semidefinite in large samples, and that $\mathbb{E}[\hat{\Sigma}_q]^{-1}\hat{\Sigma}_q - I_{q+1}$ and $\mathbb{E}[\hat{\Sigma}_{q,2}]^{-1}\hat{\Sigma}_{q,2} - I_{q+1}$ are $o_p(1)$. These arguments are analogues to those presented in the proofs of Lemmas C.8 and C.9 and are therefore only sketched. 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. We only cover $\hat{\Sigma}_{q,2}$ as that estimator simplifies to $\hat{\Sigma}_q$ when the design is sufficiently well-behaved.

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 $\kappa(\mathbb{E}[\hat{\Sigma}_{q,2}])$ since curvature is invariant to rotations. Furthermore,

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

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

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

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

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

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

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 $\widetilde{\sigma_i^2 \sigma_\ell^2}$. □

C.7 Verifying Conditions

Example 1. The only non-immediate conclusions are that:

$$\begin{aligned}
\mathbb{V}[\hat{\theta}]^{-1} \max_i (\tilde{x}_i'\beta)^2 &= O\left(\frac{\max_i (x_i'\beta)^2/n^2}{\min_i \sigma_i^2 \operatorname{trace}(\tilde{A}^2)}\right) = O\left(\frac{\max_i (x_i'\beta)^2}{r}\right) \\
\mathbb{V}[\hat{\theta}]^{-1} \max_i (\check{x}_i'\beta)^2 &= O\left(\frac{\max_{i,j} M_{jj}^{-2} \left(P_{jj} - \frac{1}{n}\right)^2 (x_j'\beta)^2 \left(\sum_{\ell=1}^n |M_{i\ell}|\right)^2 /n^2}{\min_i \sigma_i^2 \operatorname{trace}(\tilde{A}^2)}\right) \\
&= O\left(\frac{\max_{i,j} (x_j'\beta)^2 \left(\sum_{\ell=1}^n |M_{i\ell}|\right)^2}{r}\right).
\end{aligned}$$

Example 2. We first derive the representations of $\hat{\sigma}_\alpha^2$ given in section 2. When there are no common regressors, the representation in (5) follows from $B_{ii} = \frac{1}{nT_{g(i)}}\left(1 - \frac{T_{g(i)}}{n}\right)$ and

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

which yields that

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

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

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

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

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

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 $P_{ii} \le C < 1$ in large samples. The eigenvalues of $\tilde{A}$ are equal to the eigenvalues of

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

which in turn satisfies that $\frac{c_1}{n} \le \lambda_\ell \le \frac{c_2}{n}$ for $\ell = 1, \dots, N-1$ and $c_2 \ge c_1 > 0$ 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,

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

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

We finish this example with a setup where an unbalanced panel leads to a bias and inconsistency in $\hat{\theta}_{\text{HO}}$. Consider

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

where N is even, $(T_g = 2, \mathbb{E}[\varepsilon_{gt}^2] = 2\sigma^2)$ for $g \le N/2$ and $(T_g = 3, \mathbb{E}[\varepsilon_{gt}^2] = \sigma^2)$ for $g > N/2$, and the estimand is,

$$\theta = \frac{1}{n}\sum_{g=1}^{N} T_g \alpha_g^2 \qquad \text{where } n = \sum_{g=1}^{N} T_g = \frac{5N}{2}.$$

Here we have that $\tilde{A} = I_N/n$ and $\text{trace}(\tilde{A}^2) = N/n^2 = o(1)$ as $n \to \infty$ so the leave-out estimator is consistent. Furthermore,

$$nB_{ii} = P_{ii} = \begin{cases} \frac{1}{2}, & \text{if } i \le N, \\ \frac{1}{3}, & \text{otherwise,} \end{cases} \qquad \sigma_i^2 = \begin{cases} 2\sigma^2, & \text{if } i \le N, \\ \sigma^2, & \text{otherwise,} \end{cases}$$

so

$$\mathbb{E}[\tilde{\theta}] - \theta = \sum_{i=1}^{n} B_{ii}\sigma_i^2 = \frac{\sigma^2}{n}\left(N + \frac{N}{2}\right) = \frac{3\sigma^2}{5},$$

$$\mathbb{E}[\hat{\theta}_{\text{HO}}] - \theta = \sigma_{nB_{ii},\sigma_i^2} + S_B \frac{n}{n-N}\sigma_{P_{ii},\sigma_i^2} = \frac{2\sigma^2}{50} + \frac{2}{3} \times \frac{2\sigma^2}{50} = \frac{\sigma^2}{15}.$$

Example 3. $\tilde{A}$ 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 $g = 1, \ldots, N$. $\text{trace}(\tilde{A}^2) \le \frac{\lambda_1}{\min_g S_{zz,g}} \frac{1}{n}\sum_{g=1}^{N} T_g = O(\lambda_1)$. $\max_i w_i' w_i = \max_{g,t} \frac{(z_{gt} - \bar{z}_g)^2}{S_{zz,g}} = o(1)$ 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

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

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

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

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

Finally,

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

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 4. 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

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

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 C.11. *Suppose that* $\frac{\log(J)}{np_b} + \frac{J \log(J)}{n} \to 0$ *as* $n \to \infty$ *and* $J \to \infty$*. Then*

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

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

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

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

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

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

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

it follows from (Oliveira, 2009, Corollary 7.1) that $\mathbb{P}(\|S\| \geq t) \leq 2(J+1)e^{-\frac{t^2(\frac{J}{n}+\frac{1}{np_b})}{c(8+4t)}}$ 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

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

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,

$$\max_g \underline{\mathrm{w}}_{g1}^2 = \max_g \left(\frac{c'(\underline{\mathcal{L}}^\dagger)^{1/2}\Delta \dot{f}_g}{\sqrt{n}} \right)^2 = \left(\frac{2}{\sqrt{2p_b n}} \right)^2 = \frac{2}{np_n} = o(1),$$

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

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

$$\max_g (\underline{\tilde{x}}_{g1}'\beta)^2 = \max_g \left(\frac{1}{n}\psi' \Delta \dot{f}_g \right)^2 = O\left(\frac{1}{n^2}\right) = o\left(\mathbb{V}[\hat{\theta}]\right).$$

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

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

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.

$$\begin{aligned} \max_g |P_{gg} - \underline{P}_{gg}| &= \max_g |\Delta \dot{f}_g' (S_{\Delta \dot{f} \Delta \dot{f}}^\dagger - \tfrac{J+1}{n}\underline{\mathcal{L}}^\dagger)\Delta \dot{f}_g| \\ &= \tfrac{J+1}{n} \max_g \left| \Delta \dot{f}_g' (\underline{\mathcal{L}}^\dagger)^{1/2} \left(\underline{\mathcal{L}}^{1/2} \tfrac{n}{J+1} S_{\Delta \dot{f}\Delta \dot{f}}^\dagger \underline{\mathcal{L}}^{1/2} - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} \right) (\underline{\mathcal{L}}^\dagger)^{1/2} \Delta \dot{f}_g \right| \\ &= O\left(\left\| \underline{\mathcal{L}}^\dagger \tfrac{J+1}{n} S_{\Delta \dot{f}\Delta \dot{f}} - I_{J+1} + \tfrac{\mathbf{1}_{J+1}\mathbf{1}_{J+1}'}{J+1} \right\| \right) \max_g \underline{P}_{gg} = o\left(\max_g \underline{P}_{gg} \right) \\ \left| \text{trace}(\tilde{A}^2 - \underline{\tilde{A}}^2) \right| &= \left| \sum_{\ell=1}^J \frac{1}{n^2 \dot{\lambda}_\ell^2} - \frac{1}{n^2 \dot{\underline{\lambda}}_\ell^2} \right| = \text{trace}(\underline{\tilde{A}}^2) O\left(\max_\ell \left| \frac{\dot{\lambda}_\ell - \dot{\underline{\lambda}}_\ell}{\dot{\underline{\lambda}}_\ell} \right| \right) = o_p\left(\text{trace}(\underline{\tilde{A}}^2)\right) \\ \left| \frac{\lambda_1^2}{\sum_{\ell=1}^J \lambda_\ell^2} - \frac{\underline{\lambda}_1^2}{\sum_{\ell=1}^J \underline{\lambda}_\ell^2} \right| &= \frac{\underline{\lambda}_1^2}{\sum_{\ell=1}^J \underline{\lambda}_\ell^2} O\left(\frac{|\dot{\lambda}_J - \dot{\underline{\lambda}}_J|}{\dot{\underline{\lambda}}_J} + \frac{\left|\text{trace}(\underline{\tilde{A}}^2 - \tilde{A}^2)\right|}{\text{trace}(\underline{\tilde{A}}^2)} \right) = o_p(1) \end{aligned}$$

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

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

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,

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