On-Line Appendix for "Asymptotically Robust Permutation-based Randomization Confidence Intervals for Parametric OLS Regression"

Alwyn Young, London School of Economics, November 2023

Table A1: Notation used in Appendix A (also reviewed as introduced in the appendix)

(1) Sequences of real vectors: $\mathbf{X}' = (\mathbf{x}_1,...,\mathbf{x}_N)$ & $\mathbf{D}' = (\mathbf{d}_1,...,\mathbf{d}_N)$, with $\mathbf{T}$ a row permutation of $\mathbf{X}$.

(2) Sample demeaned vector sequences: $\widetilde{\mathbf{T}} = \mathbf{OT}$, where $\mathbf{O} = \mathbf{I}_N - \mathbf{1}_N\mathbf{1}'_N/N$, with $\mathbf{I}_N$ the NxN identity matrix and $\mathbf{1}_N$ an Nx1 vector of ones.

(3) Sample standardized and orthogonalized vector sequences: $\breve{\mathbf{T}} = \widetilde{\mathbf{T}}(\widetilde{\mathbf{T}}'\widetilde{\mathbf{T}}/N)^{-½}$.

(4) Sample standardized vector sequences: $\widehat{\mathbf{T}} = \widetilde{\mathbf{T}}Dg(\widetilde{\mathbf{T}}'\widetilde{\mathbf{T}}/N)^{-½}$, where $Dg(\mathbf{Z})$ denotes a diagonal matrix with diagonal elements equal to those of $\mathbf{Z}$.

(5) Vector of pair-wise root-N correlations: $\mathbf{n}$, with elements n_{pq} denoting the correlation of the p^{th} and q^{th} columns of $\breve{\mathbf{T}}$ and $\breve{\mathbf{D}}$, as in $n_{pq} = \sum_{i=1}^{N} \breve{t}_{ip}\breve{d}_{iq}/N^{½}$.

(6) Expectation across row permutations $\mathbf{T}$ of $\mathbf{X}$: $E_{\mathbf{T}}()$, as in

$$E_{\mathbf{T}}(\breve{t}_{ip}) = \sum_{j_1=1}^{N} \breve{t}_{j_1p}/N = 0 \quad \& \quad E_{\mathbf{T}}(\breve{t}_{ip}^2) = \sum_{j_1=1}^{N} \breve{t}_{j_1p}^2/N = 1 \quad (\forall\, i \;\&\; p)\cdot$$

(7) Expectation across row permutations $\mathbf{T}$ of one of the τ^{th} joint moments of $\mathbf{n}$: $E_{\mathbf{T}}^{\tau}$, as in

$$E_{\mathbf{T}}^{\tau} = E_{\mathbf{T}}\left[\prod_{k=1}^{\tau} n_{p_kq_k}\right] = E_{\mathbf{T}}\left[N^{-\tau/2}\sum_{i_1=1}^{N}...\sum_{i_\tau=1}^{N}\breve{t}_{i_1p_1}\breve{d}_{i_1q_1}...\breve{t}_{i_\tau p_\tau}\breve{d}_{i_\tau q_\tau}\right],$$

where $p_1 ... p_\tau$ & $q_1 ... q_\tau$ denote columns of $\breve{\mathbf{T}}$ and $\breve{\mathbf{D}}$.

(8) Summation across m indices excluding ties between them: $\sum_{i_1,...,i_m}$, as in

$$\sum_{i_1,...,i_3=1}^{N} a_{i_1}b_{i_2}c_{i_3} = \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N} a_{i_1}b_{i_2}c_{i_3} - \sum_{i_1=1}^{N}\sum_{i_2=1}^{N} a_{i_1}b_{i_2}c_{i_2} - \sum_{i_1=1}^{N}\sum_{i_2=1}^{N} a_{i_1}b_{i_1}c_{i_2} - \sum_{i_1=1}^{N}\sum_{i_2=1}^{N} a_{i_1}b_{i_2}c_{i_1} + 2\sum_{i_1=1}^{N} a_{i_1}b_{i_1}c_{i_1}$$

(9) Partition of the τ n_{pq} used in $E_{\mathbf{T}}^{\tau}$ into m groupings that tie elements together through their i indices: $\{e_1\}, ... , \{e_m\}$, as in:

$$\{e_1\} = \{n_{p_1q_1}, n_{p_2q_2}\}\ , \{e_2\} = \{n_{p_3q_3}\},..., \{e_m\} = \{n_{p_{\tau-1}q_{\tau-1}}, n_{p_\tau q_\tau}\},\ \text{so that}$$

$$\breve{t}_{i_1}^{\{e_1\}} = \breve{t}_{i_1p_1}\breve{t}_{i_1p_2},\ \breve{t}_{i_2}^{\{e_2\}} = \breve{t}_{i_2p_3}, ..., \breve{t}_{i_m}^{\{e_m\}} = \breve{t}_{i_mp_{\tau-1}}\breve{t}_{i_mp_\tau} \quad \& \quad \breve{d}_{i_1}^{\{e_1\}} = \breve{d}_{i_1p_1}\breve{d}_{i_1p_2},\ \breve{d}_{i_2}^{\{e_2\}} = \breve{d}_{i_2p_3}, ..., \breve{d}_{i_m}^{\{e_m\}} = \breve{d}_{i_mp_{\tau-1}}\breve{d}_{i_mp_\tau}.$$

(10) Component of $E_{\mathbf{T}}^{\tau}$ based upon summation across unequal i indices for a given partition:

$$I(\tau, \{e_1\},..., \{e_m\}) = E_{\mathbf{T}}[N^{-\tau/2} \sum_{i_1,...,i_m=1}^{N} \breve{t}_{i_1}^{\{e_1\}}\breve{d}_{i_1}^{\{e_1\}}...\breve{t}_{i_m}^{\{e_m\}}\breve{d}_{i_m}^{\{e_m\}}]\cdot$$

(11) Summation across p and q indices which together cover all m elements in the partition:

$$J(\tau, p, q, \{e_1\},..., \{e_m\}) = \sum_{i_1=1}^{N}...\sum_{i_p=1}^{N}\sum_{j_1=1}^{N}...\sum_{j_q=1}^{N} \breve{t}_{j_{d_1}}^{\{e_1\}}\breve{d}_{i_{c_1}}^{\{e_1\}}...\breve{t}_{j_{d_m}}^{\{e_m\}}\breve{d}_{i_{c_m}}^{\{e_m\}}$$

$$\&\ 1 \le p \le m, 1 \le q \le m, 1 \le c_g \le m, 1 \le d_h \le m, (g,h = 1,...,m)$$

and at least one c_g (d_h) equal to every integer in 1...p (1...q) . The $2m$ indices c_g and d_h connect the $2m$ different elements to $p \le m$ and $q \le m$ counters in the summations.

(12) "mean" based upon τ^{th} absolute powers of absolute values of elements in n_{pq}:

$$\overline{M}(\tau, n_{pq}) = N^{-\frac{\tau}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp}^{\tau}\breve{d}_{iq}^{\tau}\right|.$$

(13) Asymptotically equal to: "~".

A. Multivariate Extension of Theorem I

Following the presentation in the paper, let $\mathbf{X}' = (\mathbf{x}_1,...,\mathbf{x}_N)$ and $\mathbf{D}' = (\mathbf{d}_1,...,\mathbf{d}_N)$ denote sequences of P x 1 and Q x 1 real vectors, respectively, and $\mathbf{O} = \mathbf{I}_N - \mathbf{1}_N\mathbf{1}_N'/N$ the centering matrix. We wish to show that across the row permutations $\mathbf{T}$ of $\mathbf{X}$:

$$\text{(A.1)}\quad \mathbf{n}(\mathbf{t}_i,\mathbf{d}_i) = \left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{N} \otimes \frac{\tilde{\mathbf{D}}'\tilde{\mathbf{D}}}{N}\right)^{-½} \frac{(\tilde{\mathbf{T}} \bullet \tilde{\mathbf{D}})'\mathbf{1}_N}{\sqrt{N}}$$

where $\tilde{\mathbf{H}} = \mathbf{OH}$, $\otimes$ denotes the Kronecker product and $\bullet$ the row-by-row Kronecker or face-splitting product, is asymptotically distributed multivariate iid standard normal if

$$\text{(A.2)}\quad \lim_{N\to\infty} \frac{N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}[x_{ip} - m(x_{ip})]^\tau \sum_{i=1}^{N}[d_{iq} - m(d_{iq})]^\tau}{\left(\sum_{i=1}^{N}[x_{ip} - m(x_{ip})]^2\right)^{\tau/2}\left(\sum_{i=1}^{N}[d_{iq} - m(d_{iq})]^2\right)^{\tau/2}} = 0$$

holds for all column combinations p and q of $\mathbf{X}$ and $\mathbf{D}$ and the matrices $\tilde{\mathbf{X}}'\tilde{\mathbf{X}}/N$ and $\tilde{\mathbf{D}}'\tilde{\mathbf{D}}/N$ are bounded with determinant > γ > 0 for all N sufficiently large. Hoeffding (1951) provides a proof for a broader, but univariate, permutation problem. The generalization to the multivariate case requires additional notation and consideration of cases, but otherwise I keep the presentation as close as possible to Hoeffding's so that the proof can be checked against his original contribution if desired.

Define

$$\text{(A.3)}\quad \breve{\mathbf{T}} = \mathbf{OT}\left(\frac{\mathbf{T}'\mathbf{OT}}{N}\right)^{-½} \;\&\; \breve{\mathbf{D}} = \mathbf{OD}\left(\frac{\mathbf{D}'\mathbf{OD}}{N}\right)^{-½},\ \text{so that}\ \mathbf{n}(\mathbf{t}_i,\mathbf{d}_i) = \frac{(\breve{\mathbf{T}} \bullet \breve{\mathbf{D}})'\mathbf{1}_N}{\sqrt{N}}.$$

For the element n_{pq} of the vector $\mathbf{n}$ based on the product of the p^{th} and q^{th} columns of $\mathbf{T}$ and $\mathbf{D}$

$$\text{(A.4)}\quad n_{pq} = \sum_{i=1}^{N}\frac{\breve{t}_{ip}\breve{d}_{iq}}{N^{½}},\ \text{where}\ \sum_{i=1}^{N}\breve{t}_{ip} = \sum_{i=1}^{N}\breve{d}_{iq} = 0,\ \sum_{i=1}^{N}\breve{t}_{ip}^2 = \sum_{i=1}^{N}\breve{d}_{iq}^2 = N,$$

$$\&\ \sum_{i=1}^{N}\breve{t}_{ip_1}\breve{t}_{ip_2} = \sum_{i=1}^{N}\breve{d}_{iq_1}\breve{d}_{iq_2} = 0\ \forall\ p_1 \neq p_2\ \&\ q_1 \neq q_2,$$

$$\text{as}\ \breve{\mathbf{T}}'\mathbf{1}_N = \mathbf{0}_p,\ \breve{\mathbf{D}}'\mathbf{1}_N = \mathbf{0}_q,\ \breve{\mathbf{T}}'\breve{\mathbf{T}} = N*\mathbf{I}_p\ \&\ \breve{\mathbf{D}}'\breve{\mathbf{D}} = N*\mathbf{I}_q.$$

We shall show that all of the moments of the vector $\mathbf{n}$ converge to those of the mean zero multivariate normal with identity covariance matrix.

We begin by showing how the moments of the permuted variables are calculated. As $\mathbf{T}$ is the row permutation of $\mathbf{X}$, $\breve{\mathbf{T}} = \tilde{\mathbf{T}}(\tilde{\mathbf{T}}'\tilde{\mathbf{T}}/N)^{-½}$ is simply the row permutation of $\breve{\mathbf{X}} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}/N)^{-½}$ and the sample moments of $\breve{\mathbf{T}}$ are the same as those of $\breve{\mathbf{X}}$. Since each of the N! permutations of the rows is equally likely, expectations across the row permutations $\mathbf{T}$ are given by

$$\text{(A.5)}\quad E_{\mathbf{T}}(\breve{t}_{ip}) = \sum_{j_1=1}^{N}\frac{\breve{t}_{j_1p}}{N} = 0\ \ \&\ \ E_{\mathbf{T}}(\breve{t}_{ip}^2) = \sum_{j_1=1}^{N}\frac{\breve{t}_{j_1p}^2}{N} = 1\ \ (\forall\ i\ \&\ p),$$

while if $i_1 \neq i_2$ we have

$$\text{(A.6)}\quad E_{\mathbf{T}}(\breve{t}_{i_1 p}\breve{t}_{i_2 p}) = \sum_{j_1, j_2 = 1}^{N} \frac{\breve{t}_{j_1 p}\breve{t}_{j_2 p}}{N(N-1)} = \sum_{j_1=1}^{N}\sum_{j_2=1}^{N} \frac{\breve{t}_{j_1 p}\breve{t}_{j_2 p}}{N(N-1)} - \sum_{j_1=1}^{N} \frac{\breve{t}_{j_1 p}^2}{N(N-1)} = 0 - \frac{1}{N-1} \;\; (\forall\; p),$$

where we use the notation $j_1, j_2, \ldots$ to denote summation across multiple indices, excluding ties between them. Similarly, if $p_1 \neq p_2$

$$\text{(A.7)}\quad E_{\mathbf{T}}(\breve{t}_{ip_1}\breve{t}_{ip_2}) = \sum_{j_1=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_1 p_2}}{N} = 0 \;\; (\forall\; i)$$

$$\&\;\; E_{\mathbf{T}}(\breve{t}_{i_1 p_1}\breve{t}_{i_2 p_2}) = \sum_{j_1=1}^{N}\sum_{j_2=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_2 p_2}}{N(N-1)} - \sum_{j_1=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_1 p_2}}{N(N-1)} = 0 \;\; (\forall\; i_1 \neq i_2).$$

Next, we compute the 1^{st} and 2^{nd} moments of the elements n_{pq} of the vector $\mathbf{n}$:

$$\text{(A.8)}\quad E_{\mathbf{T}}(n_{pq}) = \sum_{i_1=1}^{N} \frac{E_{\mathbf{T}}(\breve{t}_{i_1 p})\breve{d}_{i_1 q}}{N^{½}} = \sum_{i_1=1}^{N}\sum_{j_1=1}^{N} \frac{\breve{t}_{j_1 p}\breve{d}_{i_1 q}}{N^{3/2}} = 0$$

$$E_{\mathbf{T}}(n_{p_1 q_1} n_{p_2 q_2}) = \sum_{i_1=1}^{N}\sum_{i_2=1}^{N} \frac{E_{\mathbf{T}}(\breve{t}_{i_1 p_1}\breve{t}_{i_2 p_2})\breve{d}_{i_1 q_1}\breve{d}_{i_2 q_2}}{N} = \sum_{i_1=1}^{N} \frac{E_{\mathbf{T}}(\breve{t}_{i_1 p_1}\breve{t}_{i_1 p_2})\breve{d}_{i_1 q_1}\breve{d}_{i_1 q_2}}{N} + \sum_{i_1, i_2 = 1}^{N} \frac{E_{\mathbf{T}}(\breve{t}_{i_1 p_1}\breve{t}_{i_2 p_2})\breve{d}_{i_1 q_1}\breve{d}_{i_2 q_2}}{N}$$

$$= \sum_{i_1=1}^{N}\sum_{j_1=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_1 p_2}\breve{d}_{i_1 q_1}\breve{d}_{i_1 q_2}}{N^2} + \sum_{i_1, i_2 = 1}^{N}\sum_{j_1, j_2 = 1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_2 p_2}\breve{d}_{i_1 q_1}\breve{d}_{i_2 q_2}}{N^2(N-1)}$$

$$= \sum_{i_1=1}^{N}\sum_{j_1=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_1 p_2}\breve{d}_{i_1 q_1}\breve{d}_{i_1 q_2}}{N^2} + \underbrace{\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_2 p_2}\breve{d}_{i_1 q_1}\breve{d}_{i_2 q_2}}{N^2(N-1)}}_{=0}$$

$$\underbrace{-\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_1 p_2}\breve{d}_{i_1 q_1}\breve{d}_{i_2 q_2}}{N^2(N-1)}}_{=0} \underbrace{-\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_2 p_2}\breve{d}_{i_1 q_1}\breve{d}_{i_1 q_2}}{N^2(N-1)}}_{=0} + \sum_{i_1=1}^{N}\sum_{j_1=1}^{N} \frac{\breve{t}_{j_1 p_1}\breve{t}_{j_1 p_2}\breve{d}_{i_1 q_1}\breve{d}_{i_1 q_2}}{N^2(N-1)}$$

$$= 1 + \frac{1}{N-1} \;\; (\text{if } p_1 = p_2 \;\&\; q_1 = q_2) \;\text{ or } 0 \;\; (\text{otherwise}).$$

These examples illustrate, in a manner that hopefully makes the later exposition intelligible, how the calculation of expectations produces sums of summations, with those that are across unequal indices in turn expressible as further sums of summations. In the more immediate sense, (A.8) shows that the first moment of the vector $\mathbf{n}$ made up of PQ n_{pq} elements is $\mathbf{0}_{PQ}$, while its second moments asymptotically equal the identity matrix, as desired. The next few pages focus on the higher moments.

Let $E_{\mathbf{T}}^{\tau}$ denote one of the τ^{th} moments of the joint distribution of $\mathbf{n}$ across the row permutations $\mathbf{T}$

$$\text{(A.9)}\quad E_{\mathbf{T}}^{\tau} = E_{\mathbf{T}}\left[\prod_{k=1}^{\tau} n_{p_k q_k}\right] = E_{\mathbf{T}}\left[N^{-\tau/2}\sum_{i_1=1}^{N}\ldots\sum_{i_\tau=1}^{N} \breve{t}_{i_1 p_1}\breve{d}_{i_1 q_1}\ldots\breve{t}_{i_\tau p_\tau}\breve{d}_{i_\tau q_\tau}\right],$$

where the indices may reference the same columns of $\mathbf{T}$ and $\mathbf{D}$, i.e. $p_i = p_j$ or $q_i = q_j$ for some $i \neq j$, so that the moment is across combinations of powers of the n_{pq}. As can be seen from the second line of (A.8) earlier above, $E_{\mathbf{T}}^{\tau}$ needs to be separated into components based upon whether the i indices are identical or not, which leads to elements of the form

$$\text{(A.10)}\quad I(\tau,\{e_1\},...,\{e_m\}) = E_{\mathbf{T}}\left[N^{-\tau/2} \sum_{i_1,...,i_m=1}^{N} \breve{t}_{i_1}^{\{e_1\}} \breve{d}_{i_1}^{\{e_1\}} ... \breve{t}_{i_m}^{\{e_m\}} \breve{d}_{i_m}^{\{e_m\}} \right],\ \text{where}\ \sum_{i=1}^{m} e_i = \tau,\ e_i \geq 1\ \forall\ i,$$

and $\sum_{i_1,...,i_m}$ denotes the summation across each of m indices, excluding ties between the indices, the sets $\{e_1\}$, ... ,$\{e_m\}$ constitute a partition of the τ n_{pq} used in $E_{\mathbf{T}}^{\tau}$, with the notation e_i without {} denoting the number of elements in $\{e_i\}$, and the $\breve{t}^{\{e_i\}}$ and $\breve{d}^{\{e_i\}}$ denoting the product of the elements within each set $\{e_i\}$. The $\{e_i\}$ groupings tie elements together through their i indices. Thus, for example, we might have

$$\text{(A.11)}\quad \{e_1\} = \{n_{p_1q_1}, n_{p_2q_2}\}\ , \{e_2\} = \{n_{p_3q_3}\}, ..., \{e_m\} = \{n_{p_{\tau-1}q_{\tau-1}}, n_{p_\tau q_\tau}\}$$

$$\breve{t}_{i_1}^{\{e_1\}} = \breve{t}_{i_1p_1}\breve{t}_{i_1p_2},\ \breve{t}_{i_2}^{\{e_2\}} = \breve{t}_{i_2p_3},\ \ ...\ ,\ \breve{t}_{i_m}^{\{e_m\}} = \breve{t}_{i_mp_{\tau-1}}\breve{t}_{i_mp_\tau}$$

$$\text{and}\ \breve{d}_{i_1}^{\{e_1\}} = \breve{d}_{i_1p_1}\breve{d}_{i_1p_2},\ \breve{d}_{i_2}^{\{e_2\}} = \breve{d}_{i_2p_3},\ \ ...\ ,\ \breve{d}_{i_m}^{\{e_m\}} = \breve{d}_{i_mp_{\tau-1}}\breve{d}_{i_mp_\tau}$$

Since

$$\text{(A.12)}\quad E_{\mathbf{T}}\left[\breve{t}_{i_1}^{\{e_1\}} ... \breve{t}_{i_m}^{\{e_m\}}\right] = \frac{N-m!}{N!} \sum_{j_1,...,j_m}^{N} \breve{t}_{j_1}^{\{e_1\}} ... \breve{t}_{j_m}^{\{e_m\}}\ ,$$

we have

$$\text{(A.13)}\quad I(\tau,\{e_1\},...,\{e_m\}) = \underbrace{\frac{N-m!N^m}{N!}}_{\to 1} N^{-m-\frac{\tau}{2}} \sum_{i_1,...,i_m}^{N} \sum_{j_1,...,j_m}^{N} \breve{t}_{j_1}^{\{e_1\}} \breve{d}_{i_1}^{\{e_1\}} ... \breve{t}_{j_m}^{\{e_m\}} \breve{d}_{i_m}^{\{e_m\}}$$

$$\sim N^{-m-\frac{\tau}{2}} \sum_{i_1,...,i_m}^{N} \sum_{j_1,...,j_m}^{N} \breve{t}_{j_1}^{\{e_1\}} \breve{d}_{i_1}^{\{e_1\}} ... \breve{t}_{j_m}^{\{e_m\}} \breve{d}_{i_m}^{\{e_m\}},$$

which in turn can be expressed as the sum and difference of terms of the form

$$\text{(A.14)}\quad N^{-m-\frac{\tau}{2}} J(\tau,p,q,\{e_1\},...,\{e_m\}) = N^{-m-\frac{\tau}{2}} \sum_{i_1=1}^{N} ... \sum_{i_p=1}^{N} \sum_{j_1=1}^{N} ... \sum_{j_q=1}^{N} \breve{t}_{j_{d1}}^{\{e_1\}} \breve{d}_{i_{c1}}^{\{e_1\}} ... \breve{t}_{j_{dm}}^{\{e_m\}} \breve{d}_{i_{cm}}^{\{e_m\}}$$

$$\text{with}\ 1 \leq p \leq m,\ 1 \leq q \leq m,\ 1 \leq c_g \leq m,\ 1 \leq d_h \leq m,\ (g,h = 1,...,m)$$

and at least one c_g (d_h) equal to every integer in 1..p (1...q) . The $2m$ indices c_g and d_h connect the $2m$ different elements to $p \leq m$ and $q \leq m$ counters in the summations. The third line of (A.8) earlier provides an example of how expectations add summations across j to each $I(\tau,...)$, while the fourth and fifth lines show how the $I(\tau,...)$ are re-expressed as the sum of $J(\tau,...)$ forms.

Each J can be written as the product of subset J's

$$\text{(A.15)}\quad J(\tau,p,q,\{e_1\},...,\{e_m\}) = \prod_{k=1}^{s} J(\tau_k,p_k,q_k,\{e_{k1}\},...,\{e_{km_k}\})$$

where each $\{e_{ka}\}$ equals one of the original $\{e_b\}$, and the s $\{e_{k1}\},...,\{e_{km}\}$ cover $\{e_1\},...,\{e_m\}$ in its entirety, with

$$\text{(A.16)}\quad \sum_{i=1}^{m_k} e_{ki} = \tau_k,\ \sum_{k=1}^{s} \tau_k = \tau,\ \sum_{k=1}^{s} p_k = p,\ \sum_{k=1}^{s} q_k = q,\ \&\ \sum_{k=1}^{s} m_k = m.$$

We assume that each J is subdivided into the greatest possible number of factors. In the fourth line of (A.8) above, for example, we have:

$$\text{(A.17)}\quad J(\tau=2,p=2,q=2,\{n_{p_1q_1}\},\{n_{p_2q_2}\})=\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\breve{t}_{j_1p_1}\breve{t}_{j_2p_2}\breve{d}_{i_1q_1}\breve{d}_{i_2q_2}=$$

$$\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\breve{t}_{j_1p_1}\breve{d}_{i_1q_1}\sum_{i_2=1}^{N}\sum_{j_2=1}^{N}\breve{t}_{j_2p_2}\breve{d}_{i_2q_2}=J(\tau_1=1,p_1=1,q_1=1,\{n_{p_1q_1}\})J(\tau_2=1,p_2=1,q_2=1,\{n_{p_2q_2}\})$$

while all three terms in the fifth line are indivisible because the i, j counters for the $\breve{t}$ and $\breve{d}$ elements connect at least one element of $n_{p_1q_1}$ to $n_{p_2q_2}$. If $J(\tau_k,p_k,q_k,\{e_{k1}\},...,\{e_{km_k}\})$ is indivisible, it is because the $2m_k$ c_{kg} and d_{kh} subscript indices link across the m_k groups $\{e_{k1}\},...,\{e_{km_k}\}$. To do so, there must be at least m_k-1 equalities in these indices, i.e. at most m_k+1 distinct values. At the same time, these indices cover every one of the numbers in 1...p_k and 1...q_k, so we may conclude that

$$\text{(A.18)}\quad p_k+q_k\le m_k+1$$

We note that if $(c_{kg},d_{kg})=(c_{kh},d_{kh})$ for some $kg\neq kh$, we have more than the minimum m_k-1 equalities necessary for indivisibility and (A.18) holds with strict inequality. Summing across all s groups that make up $J(\tau,p,q,\{e_1\},...,\{e_m\})$,

$$\text{(A.19)}\quad p+q\le m+s$$

with strict inequality if $(c_{kg},d_{kg})=(c_{kh},d_{kh})$ ever holds.

Next, we take the absolute value, apply an inequality associated with that, and then apply Hölder's Inequality as well:

$$\text{(A.20)}\quad \left|J(\tau_k,p_k,q_k,\{e_{k1}\},...,\{e_{km_k}\})\right|\le\sum_{i_1=1}^{N}...\sum_{i_{p_k}=1}^{N}\sum_{j_1=1}^{N}...\sum_{j_{q_k}=1}^{N}\left|\breve{t}_{j_{d_1}}^{\{e_{k1}\}}\breve{d}_{i_{c_1}}^{\{e_{k1}\}}\right|...\left|\breve{t}_{j_{d_{m_k}}}^{\{e_{km_k}\}}\breve{d}_{i_{c_{m_k}}}^{\{e_{km_k}\}}\right|$$

$$\le\prod_{g=1}^{m_k}\left(\sum_{i_1=1}^{N}...\sum_{i_{p_k}=1}^{N}\sum_{j_1=1}^{N}...\sum_{j_{q_k}=1}^{N}\left|\breve{t}_{j_{d_g}}^{\{e_{kg}\}}\breve{d}_{i_{c_g}}^{\{e_{kg}\}}\right|^{\tau_k/e_{kg}}\right)^{e_{kg}/\tau_k}=\prod_{g=1}^{m_k}\left(N^{p_k+q_k-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{j}^{\{e_{kg}\}}\breve{d}_{i}^{\{e_{kg}\}}\right|^{\tau_k/e_{kg}}\right)^{e_{kg}/\tau_k}$$

where the reader is reminded that e_{kg} denotes the number of n_{pq} in $\{e_{kg}\}$, with $\sum e_{kg}=\tau_k$, allowing the application of Hölder's Inequality in the manner shown. We now decompose the set $\{e_{kg}\}$ into its constituent parts. Let 1..r, $r\le\tau$, index the unique n_{pq} variables across which the expectation $E_{\mathbf{T}}^{\tau}$ is taken, so that

$$\text{(A.21)}\quad E_{\mathbf{T}}^{\tau}=E_{\mathbf{T}}\left[\prod_{k=1}^{\tau}n_{p_kq_k}\right]=E_{\mathbf{T}}\left[\prod_{a=1}^{r}n_{p_aq_a}^{f_a}\right],$$

where, as earlier above, in the first product different values of k may reference the same n_{pq}, but in the second product each a references a unique n_{pq} and each f_a is > 0. Let $f_{1kg}...f_{rkg}$ (some of which are possibly 0) denote the power the unique n_{pq} in (A.21) are raised to in the grouping $\{e_{kg}\}$. We can then apply Hölder's Inequality once again[1]

[1]The use of Hölder's Inequality in the third line requires additional explanation. Holder's inequality states that for real numbers a_{cd} and $p_1...p_M$ all > 0, with $1/p_1+...+1/p_M=1$,

$$\sum_{c=1}^{N}\left|\prod_{d=1}^{M}a_{cd}\right|\le\prod_{d=1}^{M}\left(\sum_{c=1}^{N}|a_{cd}|^{p_d}\right)^{1/p_d}$$ (continued on next page)

$$\text{(A.22)}\quad \prod_{g=1}^{m_k}\left(N^{p_k+q_k-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_j^{\{e_{kg}\}}\breve{d}_i^{\{e_{kg}\}}\right|^{\tau_k/e_{kg}}\right)^{e_{kg}/\tau_k}$$

$$= N^{p_k+q_k-2}\prod_{g=1}^{m_k}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp_1}^{f_{1kg}}\ldots\breve{t}_{jp_r}^{f_{rkg}}\breve{d}_{iq_1}^{f_{1kg}}\ldots\breve{d}_{iq_r}^{f_{rkg}}\right|^{\tau_k/e_{kg}}\right)^{e_{kg}/\tau_k} \quad \text{where } \sum_{g=1}^{m_k}\sum_{h=1}^{r}f_{hkg}=\sum_{g=1}^{m_k}e_{kg}=\tau_k,$$

$$\le N^{p_k+q_k-2}\prod_{g=1}^{m_k}\left(\prod_{a=1}^{r}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp_a}^{\tau_k}\breve{d}_{iq_a}^{\tau_k}\right|\right)^{f_{akg}/e_{kg}}\right)^{e_{kg}/\tau_k} = N^{p_k+q_k-2}\prod_{g=1}^{m_k}\prod_{a=1}^{r}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp_a}^{\tau_k}\breve{d}_{iq_a}^{\tau_k}\right|\right)^{f_{akg}/\tau_k}$$

$$= N^{p_k+q_k-2}\prod_{a=1}^{r}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp_a}^{\tau_k}\breve{d}_{iq_a}^{\tau_k}\right|\right)^{f_{ak}/\tau_k} \quad \text{where } f_{ak}=\sum_{g=1}^{m_k}f_{akg} \;\&\; \sum_{a=1}^{r}f_{ak}=\tau_k.$$

$$= N^{p_k+q_k+\frac{\tau_k}{2}-1}\prod_{a=1}^{r}\overline{M}(\tau_k,n_{p_aq_a})^{f_{ak}/\tau_k} \text{ where } \overline{M}(\tau_k,n_{p_aq_a}) = N^{-\frac{\tau_k}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp_a}^{\tau_k}\breve{d}_{iq_a}^{\tau_k}\right|.$$

Applying the bound to each element on the right hand side of (A.15), we then have

$$\text{(A.23)}\quad N^{-m-\frac{\tau}{2}}\left|J(\tau,p,q,\{e_1\},\ldots,\{e_m\})\right| \le N^{p+q-s-m}\prod_{k=1}^{s}\prod_{a=1}^{r}\overline{M}(\tau_k,n_{p_aq_a})^{f_{ak}/\tau_k}.$$

Let us now assume (to be proven later) that assumption (A2) and the associated assumptions on $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N$ and $\widetilde{\mathbf{D}}'\widetilde{\mathbf{D}}/N$ earlier above are sufficient to guarantee that

$$\text{(A.24)}\quad N^{-\frac{\tau_k}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\breve{t}_{jp}^{\tau_k}\breve{d}_{iq}^{\tau_k} = N^{-\frac{\tau_k}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\breve{x}_{jp}^{\tau_k}\breve{d}_{iq}^{\tau_k} = o(1) \quad \text{for all } p \;\&\; q \text{ and } \tau_k = 3,4,\ldots$$

From this we see that if τ_k is even and greater than 2, then $\overline{M}(\tau_k,n_{pq})\to 0$. If τ_k is odd and greater than 1, we can apply the Cauchy-Schwarz inequality

$$\text{(A.25)}\quad \overline{M}(2\eta+1,n_{pq})^2 = \left(N^{-\frac{2\eta+1}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|n_{pq}\right|^{\eta}\left|n_{pq}\right|^{\eta+1}\right)^2$$

$$\le \left(N^{-\frac{2\eta}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|n_{pq}\right|^{2\eta}\right)\left(N^{-\frac{2\eta+2}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|n_{pq}\right|^{2\eta+2}\right)$$

$$= \left(N^{-\frac{2\eta}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}n_{pq}^{2\eta}\right)\left(N^{-\frac{2\eta+2}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}n_{pq}^{2\eta+2}\right) = o(1) \text{ for } \eta = 1,2,\ldots$$

To apply to (A.22), begin by defining $l = 1..N^2$, and using it to count through the ij indices in (A.22), so that $(i_1,j_1) = (1,1)$, $(i_2,j_2) = (1,2)$ ... $(i_{N+1},j_{N+1}) = (2,1)$ ... , and

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp_1}^{f_{1kg}}\ldots\breve{t}_{jp_r}^{f_{rkg}}\breve{d}_{iq_1}^{f_{1kg}}\ldots\breve{d}_{iq_r}^{f_{rkg}}\right|^{\tau_k/e_{kg}} = \sum_{l=1}^{N^2}\left|\breve{t}_{j_lp_1}^{f_{1kg}\tau_k/e_{kg}}\ldots\breve{t}_{j_lp_r}^{f_{rkg}\tau_k/e_{kg}}\breve{d}_{i_lq_1}^{f_{1kg}\tau_k/e_{kg}}\ldots\breve{d}_{i_lq_r}^{f_{rkg}\tau_k/e_{kg}}\right| = \sum_{l=1}^{N^2}\prod_{a\in\{a:f_{akg}>0\}}\left|\breve{t}_{j_lp_a}^{f_{akg}\tau_k/e_{kg}}\breve{d}_{i_lq_a}^{f_{akg}\tau_k/e_{kg}}\right|,$$

where in the last we drop multiplication by terms raised to a power $f_{akg} = 0$ (as those terms equal 1). We now apply the inequality

$$\sum_{l=1}^{N^2}\prod_{a\in\{a:f_{akg}>0\}}\left|\breve{t}_{j_lp_a}^{f_{akg}\tau_k/e_{kg}}\breve{d}_{i_lq_a}^{f_{akg}\tau_k/e_{kg}}\right| \le \prod_{a\in\{a:f_{akg}>0\}}\left(\sum_{l=1}^{N^2}\left|\breve{t}_{j_lp_a}^{\tau_k}\breve{d}_{i_lq_a}^{\tau_k}\right|\right)^{f_{akg}/e_{kg}} = \prod_{a=1}^{r}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp_a}^{\tau_k}\breve{d}_{iq_a}^{\tau_k}\right|\right)^{f_{akg}/e_{kg}},$$

where in the last we reintroduce multiplication by terms raised to a power $f_{akg} = 0$, as these equal 1. In sum, by removing and then bringing back in terms raised to the power 0, Hölder's inequality can be applied here (and in other instances below as well).

Finally, we have

$$(A.26)\quad \overline{M}(2,n_{pq}) = N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\breve{t}_{jp_a}^{2}\,\breve{d}_{iq_a}^{2}\right| = N^{-2}\sum_{i=1}^{N}\breve{d}_{iq_a}^{2}\sum_{j=1}^{N}\breve{t}_{jp_a}^{2} = 1.$$

Combining these results with (A.23), and the fact that $p+q \le s+m$, we see that if $\tau_k \ge 2$ for all k in 1..s and (a) $\tau_k > 2$ for any k or (b) $\tau_k = 2$ for all k and $p+q < s+m$, then $N^{-m-\tau/2}J(\tau...)$ asymptotically equals 0.

We now return to the equality in (A.15), expressing $J(\tau...)$ as the product of s $J(\tau_k...)$. If $\tau_k = 1$ we have $m_k = p_k = q_k = 1$, and $J(\tau_k...)$ is given by

$$(A.27)\quad \sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\breve{t}_{j_1p_1}\breve{d}_{i_1q_1} = \sum_{i_1=1}^{N}\breve{t}_{j_1p_1}\sum_{j_1=1}^{N}\breve{d}_{i_1q_1} = 0,$$

from which it follows that $N^{-m-\tau/2}J(\tau...) = 0$ for all N. Hence, the only case where $N^{-m-\tau/2}J(\tau...)$ may not be identically or asymptotically zero is where $\tau_k = 2$ for all k. This means that each $J(\tau_k,p_k,q_k,\{e_{k1}\},...,\{e_{km}\})$ involves two elements, $n_{p_1q_1}$ and $n_{p_2q_2}$, divided into $m_k = 1$ or 2 groups. If $m_k = 2$, then $p_k + q_k \le 3$. If $p_k + q_k = 3$, then $J(\tau_k...)$ is given by

$$(A.28)\quad \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N}\breve{t}_{j_1p_1}\breve{t}_{j_1p_2}\breve{d}_{i_1q_1}\breve{d}_{i_2q_2} \quad\text{or}\quad \sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\breve{t}_{j_1p_1}\breve{t}_{j_2p_2}\breve{d}_{i_1q_1}\breve{d}_{i_1q_2}$$

both of which are zero. If $p_k + q_k = 2$ for any k, then $p + q + s - m < 0$, and by the results of the previous paragraph $N^{-m-\tau/2}J(\tau...)$ is asymptotically zero.

From the above, we see that the only case where $N^{-m-\tau/2}J(\tau...)$ may not be identically or asymptotically zero is when for each subcomponent $J(\tau_k...)$ we have $\tau_k = 2$ and $m_k = p_k = q_k = 1$ (as $p_k \le m_k$, $q_k \le m_k$), i.e. there is only one grouping of two n_{pq}, summed across one index for i and one for j, i.e.

$$(A.29)\quad J(\tau_k = 2, p_k = 1, q_k = 1, \{n_{p_1q_1}, n_{p_2q_2}\}) = \sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\breve{t}_{j_1p_1}\breve{t}_{j_1p_2}\breve{d}_{i_1q_1}\breve{d}_{i_1q_2}$$

which equals N^{2} if $p_1 = p_2$ and $q_1 = q_2$ and 0 otherwise. Since $J(\tau...)$ is a product of $J(\tau_k...)$, we then know that the only form of $N^{-m-\tau/2}J(\tau...)$ that is not identically or asymptotically zero is:

$$(A.30)\quad N^{-m-\tau/2}J(\tau, p, q, \{n_{p_1q_1}, n_{p_1q_1}\},...,\{n_{p_mq_m}, n_{p_mq_m}\}) \text{ with } m = p = q = \tau/2$$

$$= N^{-\tau}\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}...\sum_{i_{\tau/2}=1}^{N}\sum_{j_{\tau/2}=1}^{N}\breve{t}_{j_1p_1}^{2}\breve{d}_{i_1q_1}^{2}...\breve{t}_{j_{\tau/2}p_{\tau/2}}^{2}\breve{d}_{i_{\tau/2}q_{\tau/2}}^{2} = N^{-\tau}N^{\tau} = 1.$$

As described earlier, $I(\tau,\{e_1\},...,\{e_m\})$ is made up of the sum and difference of $N^{-m-\tau/2}J(\tau...)$ terms, the only one of which is not identically or asymptotically zero is given in (A.30), i.e. when $I(\tau,...)$ involves powers of 2 of each n_{pq}. This implies that the only $I(\tau,...)$ that is not identically or asymptotically zero is that where τ is even, $I(\tau,...)$ must be positive, and

$$(A.31)\quad I(\tau,\{e_1\},...,\{e_m\}) \sim N^{-m-\frac{\tau}{2}}\sum_{i_1,...,i_m}^{N}\sum_{j_1,...,j_m}^{N}\breve{t}_{j_1}^{\{e_1\}}\breve{d}_{i_1}^{\{e_1\}}...\breve{t}_{j_m}^{\{e_m\}}\breve{d}_{i_m}^{\{e_m\}}$$

$$= N^{-\frac{\tau}{2}-\frac{\tau}{2}}J(\tau,\tau/2,\tau/2,\{e_1\},...,\{e_{\tau/2}\}) = N^{-\tau}N^{\tau} = 1.$$

$E_{\mathbf{T}}^{\tau}$ is made up of the sum of $I(\tau,...)$ which tie the τ n_{pq} elements (possibly repeating) into m groups through the indices i and j. To not be identically or asymptotically zero, the $I(\tau,...)$ must involve powers of 2 of each n_{pq}, so the only asymptotically non-zero $E_{\mathbf{T}}^{\tau}$ is that where the powers to which the r unique n_{pq} are raised, $f_1, ... ,f_r$, as well as $\tau = \sum f_a$, are all even. The number of ways in which f_a objects can be tied together in pairs is $(f_a$ -1)!! (where !! denotes the double factorial). Consequently, we have shown that for all $\tau > 2$

$$\text{(A.32)}\quad E_{\mathbf{T}}^{\tau} = E_{\mathbf{T}}\left[\prod_{a=1}^{r} n_{p_a q_a}^{f_a}\right] \rightarrow \left[\prod_{a=1}^{r}(f_a - 1)!!\right] \text{ (if all } f_a \text{ even)}, \quad = 0 \text{ (otherwise)},$$

which are the higher moments of a vector of independent mean zero standard normals!

All that remains is to show that assumption (A2) implies (A.24). Define

$$\text{(A.33)}\quad \widehat{x}_{ik} = \frac{x_{ip} - m(x_{ip})}{\left(\sum_{i=1}^{N}[x_{ip} - m(x_{ip})]^2\right)^{1/2}} \quad \& \quad \widehat{d}_{iq} = \frac{d_{iq} - m(d_{iq})}{\left(\sum_{i=1}^{N}[d_{iq} - m(d_{iq})]^2\right)^{1/2}}$$

so that assumption (A2) may be re-expressed as

$$\text{(A.34)}\quad \lim_{N\to\infty} N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\widehat{x}_{ip}^{\tau}\sum_{j=1}^{N}\widehat{d}_{iq}^{\tau} = 0 \;\; \forall \;\; p,\; q \;\; \& \;\; \forall\, \tau = 3, 4, ...$$

If τ is even, we can equivalently say that

$$\text{(A.34)}'\quad \lim_{N\to\infty} N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\left|\widehat{x}_{ip}^{\tau}\right|\sum_{j=1}^{N}\left|\widehat{d}_{iq}^{\tau}\right| = 0.$$

However, for any odd $\tau = 2\eta+1$, we note that by Hölder's inequality

$$\text{(A.35)}\quad N^{\frac{2\eta+1}{2}-1}\sum_{i=1}^{N}\left|\widehat{x}_{ip}^{2\eta+1}\right|\sum_{i=1}^{N}\left|\widehat{d}_{iq}^{2\eta+1}\right| \le \left(N^{\frac{2\eta+2}{2}-1}\sum_{i=1}^{N}\left|\widehat{x}_{ip}^{2\eta+2}\right|\sum_{i=1}^{N}\left|\widehat{d}_{iq}^{2\eta+2}\right|\right)^{½}\left(N^{\frac{2\eta}{2}-1}\sum_{i=1}^{N}\left|\widehat{x}_{ip}^{2\eta}\right|\sum_{i=1}^{N}\left|\widehat{d}_{iq}^{2\eta}\right|\right)^{½}$$

so (A.34)′ in fact applies for all τ = 3, 4, ...[2] We also note that

$$\text{(A.36)}\quad \breve{\mathbf{X}} = \widetilde{\mathbf{X}}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\right)^{-½} = N^{½}\widehat{\mathbf{X}}\mathbf{A}, \text{ where } \mathbf{A} = Dg(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{½}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-½}$$

$$\& \;\; \breve{\mathbf{D}} = \widetilde{\mathbf{D}}\left(\frac{\widetilde{\mathbf{D}}'\widetilde{\mathbf{D}}}{N}\right)^{-½} = N^{½}\widehat{\mathbf{D}}\mathbf{B}, \text{ where } \mathbf{B} = Dg(\widetilde{\mathbf{D}}'\widetilde{\mathbf{D}})^{½}(\widetilde{\mathbf{D}}'\widetilde{\mathbf{D}})^{-½}.$$

where $Dg(\mathbf{Z})$ denotes a diagonal matrix with diagonal elements equal to those of $\mathbf{Z}$. The elements of $\mathbf{A}$ and $\mathbf{B}$ are asymptotically bounded as for all N sufficiently large

$$\text{(A.37)}\quad trace(\mathbf{A}'\mathbf{A}) = trace((\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-½}\,Dg(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-½}) = trace((\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\,Dg(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})) < (\Delta P)^{P}/\gamma < \infty$$

where Δ and γ are the asymptotic upper bound on the diagonal elements and lower bound on the determinant of $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N$, respectively, with the same (with Q in place of P) in the case of $\mathbf{B}$. To see the

[2]When τ = 3 and η = 1, the second square root on the right-hand side of (A35) equals 1 while the first goes to 0; in all other cases both square roots on the right hand side go to zero.

last, note that the largest eigenvalue of $(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N)^{-1}$ is the inverse of the smallest eigenvalue of $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N$, which by the trace and determinant property of eigenvalues is greater than or equal to $\gamma/(\Delta P)^{P-1}$. The trace of $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N$ is bounded by ΔP. Using the fact that for real positive semi-definite matrices the trace of a matrix product is less than or equal to the maximum eigenvalue of one times the trace of the other (Fang, Loparo & Feng 1994), then gives the bound specified above.

With these results in mind, we complete the proof using properties of the absolute value and Hölder's inequality to show that

$$\text{(A.38)}\quad \left| N^{-\frac{\tau}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\breve{x}_{jp}^{\tau}\breve{d}_{iq}^{\tau}\right| = \left| N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{e=1}^{P}\hat{x}_{je}a_{ep}\right)^{\tau}\left(\sum_{f=1}^{Q}\hat{d}_{if}b_{fq}\right)^{\tau}\right|$$

$$= \left| N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\sum_{e=1}^{P}\sum_{f=1}^{Q}\hat{x}_{je}a_{ep}\hat{d}_{if}b_{fq}\right)^{\tau}\right| = \left| N^{\frac{\tau}{2}-1}\sum_{l=1}^{N^2}\left(\sum_{k=1}^{PQ}\hat{x}_{j_le_k}a_{e_kp}\hat{d}_{i_lf_k}b_{f_kq}\right)^{\tau}\right| =$$

$$\left| N^{\frac{\tau}{2}-1}\sum_{l=1}^{N^2}\sum_{g_1+\ldots+g_{PQ}=\tau}\frac{\tau!}{g_1!\cdots g_{PQ}!}\prod_{k=1}^{PQ}\hat{x}_{j_le_k}^{g_k}a_{e_kp}^{g_k}\hat{d}_{i_lf_k}^{g_k}b_{f_kq}^{g_k}\right| \le N^{\frac{\tau}{2}-1}\sum_{g_1+\ldots+g_{PQ}=\tau}\frac{\tau!}{g_1!\cdots g_{PQ}!}\sum_{l=1}^{N}\prod_{k=1}^{PQ}\left|\hat{x}_{j_le_k}^{g_k}a_{e_kp}^{g_k}\hat{d}_{i_lf_k}^{g_k}b_{f_kq}^{g_k}\right|$$

$$\le N^{\frac{\tau}{2}-1}\sum_{g_1+\ldots+g_{PQ}=\tau}\frac{\tau!}{g_1!\cdots g_{PQ}!}\prod_{k=1}^{PQ}\left(\sum_{l=1}^{N}\left|\hat{x}_{j_le_k}^{\tau}a_{e_kp}^{\tau}\hat{d}_{i_lf_k}^{\tau}b_{f_kq}^{\tau}\right|\right)^{g_k/\tau}$$

$$= \sum_{g_1+\ldots+g_{PQ}=\tau}\frac{\tau!}{g_1!\cdots g_{PQ}!}\prod_{k=1}^{PQ}\left(\left|a_{e_kp}^{\tau}b_{f_kq}^{\tau}\right|N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\hat{x}_{je_k}^{\tau}\hat{d}_{if_k}^{\tau}\right|\right)^{g_k/\tau} \to 0\ \text{[by (A34)′ above]},$$

where we use a and b to denote the elements of **A** and **B** as defined in (A36), in the second line we change double summations to single summations by introducing the subscripts $k = 1..PQ$ and $l = 1..N^2$ which we use on e & f and i & j to capture the movement through the original double summations of these,[3] in the third line we apply the multinomial expansion using the notation $\sum_{g_1+\ldots+g_{PQ}=\tau}$ to denote the summation across all sets of PQ non-negative integers that sum to τ, in the fourth line we apply Hölder's Inequality, and in the fifth line we make use of the boundedness of the elements a and b of **A** and **B**.

[3]That is, $(e_1,f_1) = (1,1)$, $(e_2,f_2) = (1,2)$ ... $(e_Q,f_Q) = (1,Q)$, $(e_{Q+1},f_{Q+1}) = (2,1)$ ... $(e_{PQ},f_{PQ}) = (P,Q)$, with a similar sequence for l.

Table B1: Notation used in Appendices B & C (also reviewed as introduced in the appendices)

(1) Regression model: $\mathbf{y} = \mathbf{X_W}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ or $\mathbf{y} = \mathbf{Z}_+\boldsymbol{\gamma}_+ + \boldsymbol{\varepsilon}$ where $\mathbf{Z}_+ = (\mathbf{X_W},\mathbf{Z})$ and $\boldsymbol{\gamma}'_+ = (\boldsymbol{\beta}',\boldsymbol{\gamma}')$. Estimated parameters denoted by $\hat{}$. $\mathbf{X_W}$ is *NxPQ*, $\mathbf{Z}$ *NxK* and $\mathbf{Z}_+$ *NxK*$_+$.

(2) $\mathbf{A_B}$ and $\bullet$ denote the row by row Kronecker product, $\mathbf{A_B} = \mathbf{A} \bullet \mathbf{B}$. This appears in the form of *NxP* treatment variables $\mathbf{X}$ multiplied by *NxQ* interaction covariates $\mathbf{W}$ ($\mathbf{X_W}$) and the multiplication of $\mathbf{W}$ with errors $\boldsymbol{\varepsilon}$ ($\mathbf{W}_{\boldsymbol{\varepsilon}}$). $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product.

(3) The *N* observations are divided into *M* treatment groupings, with all observations in a treatment grouping receiving the same treatment value. We denote the *MxP* matrix of treatment values by $\boldsymbol{\mathcal{X}}$ and its row permutation by $\boldsymbol{\mathcal{T}}$, with $\mathbf{T}$ denoting the consequent *NxP* matrix of observation level treatment values. We use subscripted *m* or *i* to identify the m^{th} or i^{th} row of $\boldsymbol{\mathcal{X}}$ and $\mathbf{X}$, as in $\boldsymbol{\mathcal{x}}'_m$ and $\mathbf{x}'_i$. We use scripted notation as well for non-treatment matrices to distinguish those with *M* rows from those with *N*, as in $\mathcal{D}$ versus $\mathbf{D}$. Whereas in the treatment matrices the m^{th} element represents the common treatment given to all elements *i* in the m^{th} grouping, in the case of other variables the scripted matrix refers to the sum of all elements *i* in the m^{th} grouping, i.e. $\mathcal{d}_{mq} = \Sigma_{i \in m} d_{iq}$.

(4) $\mathbf{y}_{\mathbf{T},\boldsymbol{\beta}_0} = \mathbf{y} + (\mathbf{T_W} - \mathbf{X_W})\boldsymbol{\beta}_0$ denotes the counterfactual value of $\mathbf{y}$ under the null $\boldsymbol{\beta}_0$ following the row permutation $\boldsymbol{\mathcal{T}}$ of $\boldsymbol{\mathcal{X}}$ and the application of these values to observation groupings to form $\mathbf{T}$. $\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}$ are the associated parameter estimates.

(5) In addition to the $M \leq N$ treatment groupings, the practitioner divides the sample into $C \leq N$ cluster groups, within which the errors might be correlated, and $O \leq N$ other groupings in which other regressors might be correlated. $U \leq N$ denotes the largest number of groups the sample can be divided into such that the observations associated with each treatment *m*, cluster *c*, or regressor *o* grouping reside in at most one union grouping *u*. While errors and regressors within each union grouping may be arbitrarily correlated, they are independent across union groupings. We define intersection groupings *v* as the largest observational grouping such that all observations belong to at most one cluster grouping *c* and one treatment grouping *m*, with the number of such groupings $V \leq N$.

(6) Notation $\Sigma_{i \in u}$ denotes the sum across observations *i* in union grouping *u* (or cluster, treatment or intersection groupings *c*, *m* or *v*), and similarly $\Sigma_{m \subseteq u}$ denotes the sum across treatment groupings (or, similarly, cluster or intersection groupings *c* or *v*) in union grouping *u*.

(7) $\mathbf{d}'_i$ refers to the i^{th} row of matrix $\mathbf{D}$ and $\mathbf{D}_u$, $\mathbf{D}_c$, $\mathbf{D}_m$, & $\mathbf{D}_v$ to the rows of $\mathbf{D}$ associated with the subscripted union, cluster, treatment or intersection grouping. d_{ij} refers to the ij^{th} element of $\mathbf{D}$, while $\mathbf{d}_{uj}$ (or similarly *cj*, *mj* or *vj*) refers to the rows of the j^{th} column of $\mathbf{D}$ associated with the union grouping *u*.

(8) "Means" are calculated by dividing by *M*, and in addition to means across all observations, we also have means across subscripted groupings, defined as follows:

$$m(d_{i1}d_{i2}) = \sum_{i=1}^{N} \frac{d_{i1}d_{i2}}{M},\ m_g(d_{i1}d_{i2}) = \sum_{g=1}^{G}\sum_{i\in g} \frac{d_{i1}d_{i2}}{M}\ \&\ m_g(d_{i1},d_{j2}) = \sum_{g=1}^{G}\sum_{i\in g}\sum_{j\in g} \frac{d_{i1}d_{j2}}{M},$$

for $g = c$, u, or v and $G = C$, U or V. $\omega()$ denotes the mean across M elements, as in

$$\omega(x_{mp}) = \sum_{m=1}^{M} \frac{x_{mp}}{M}\ \&\ \omega(d_{mq}) = \sum_{m=1}^{M} \frac{d_{mq}}{M}.$$

(9) $\mathbf{1}_M$ & $\mathbf{0}_M$ denote $Mx1$ vectors of 1s & 0s, $\mathbf{0}_{QxQ}$ a QxQ matrix of 0s & $\mathbf{I}_M$ the MxM identity matrix.

(10) We use the ~ notation to denoted demeaned M matrices and their elements, as in $\tilde{X} = O X$ where $O = \mathbf{I}_M - \mathbf{1}_M\mathbf{1}'_M/M$.

(11) $\mathbf{t}_{\mathbf{W}k}$ and $\mathbf{x}_{\mathbf{W}l}$ denote the k^{th} and l^{th} columns of $\mathbf{T}_\mathbf{W}$ and $\mathbf{X}_\mathbf{W}$, with the i^{th} elements of these vectors given by $t_{ip(k)}w_{iq(k)}$ and $x_{ip(l)}w_{iq(l)}$, where $p(j)$ and $q(j)$ denote the columns of $\mathbf{T}$ (or $\mathbf{X}$) and $\mathbf{W}$ associated with the j^{th} column of $\mathbf{T}_\mathbf{W}$ (or $\mathbf{X}_\mathbf{W}$).

(12) $\xrightarrow{a.s.}$ denotes convergence almost surely across the probability law governing the data $\mathbf{D} = (\mathbf{X}_\mathbf{W},\mathbf{Z},\boldsymbol{\varepsilon})$. $\xrightarrow{p}$ & $\xrightarrow{d}$ denote convergence in probability and distribution across permutations T of X almost surely across the distribution of the data $\mathbf{D}$. $E()$ denotes the expectation across the data $\mathbf{D}$.

(13) $\mathbf{n}_{PQ}$ denotes the multivariate iid standard normal, indicated by $\mathbf{n}_{PQ} \sim \mathrm{N}(\mathbf{0}_{PQ},\mathbf{I}_{PQ})$.

B. Generalizing the Results to Allow for Clustering and Grouped Treatment

This appendix generalizes the results to include grouped treatment and standard errors which are homoskedastic, heteroskedasticity robust, or clustered at, below, above and across treatment groupings (i.e. at any level). As in the paper, we have

$$\text{(B.1)}\quad \mathbf{y} = \mathbf{X}_\mathbf{W}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

where $\mathbf{X}_\mathbf{W} = \mathbf{X} \bullet \mathbf{W}$ and • denotes the row by row Kronecker or "face-splitting" product of two matrices, while $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are N x 1 vectors of outcomes and residuals, $\mathbf{Z}$ and $\boldsymbol{\gamma}$ the N x K matrix of covariates and K x 1 vector of associated parameters, $\mathbf{X}$ an N x P matrix of treatment variables, $\mathbf{W}$ an N x Q matrix of interaction covariates, and $\boldsymbol{\beta}$ the PQ x 1 vector of parameters of interest. The sample is divided into $M \leq N$ groupings of observations, with all observations i in grouping m in $\mathbf{X}$ receiving the same treatment row vector. We use the matrix X to denote the M x P matrix of grouped treatments underlying $\mathbf{X}$, T any of the $M!$ equally likely row permutations of X, and $\mathbf{T}$ the N x P matrix of treatments associated with the allocation of T to the corresponding M observational groupings. Stratification is considered in a later appendix.

As before, we combine the treatment and non-treatment regressors into more compact notation, describing our regression model as

$$\text{(B.2)}\quad \mathbf{y} = \mathbf{Z}_+\boldsymbol{\gamma}_+ + \boldsymbol{\varepsilon}\ \text{ or (at the observation level) }\ y_i = \mathbf{z}'_{+i}\boldsymbol{\gamma}_+ + \varepsilon_i,$$

where $\mathbf{Z}_+ = (\mathbf{X}_\mathbf{W},\mathbf{Z})$ and $\boldsymbol{\gamma}'_+ = (\boldsymbol{\beta}',\boldsymbol{\gamma}')$ denote the full matrix of regressors and parameters and $\mathbf{z}'_{+i}$ the 1 x K_+ vector representing the i^{th} row of $\mathbf{Z}_+$. White (1980) assumes that $(\mathbf{z}'_{+i},\varepsilon_i)$ is a sequence of independent but

not necessarily identically distributed random vectors. To generalize his work, assume that in addition to the $M \leq N$ treatment groupings, the practitioner divides the sample into $C \leq N$ cluster groups, within which the errors might be correlated, and $O \leq N$ other groupings in which other regressors might be correlated, and let $U \leq N$ denote the largest number of groups the sample can be divided into such that the observations associated with each treatment m, cluster c, or regressor o grouping reside in at most one union grouping u. While errors and regressors within each union grouping may be arbitrarily correlated, they are independent across union groupings. Further below we will also have need to make use of intersection groupings v, defined as the largest observational grouping such that all observations belong to at most one cluster grouping c and one treatment grouping m, with the number of such groupings $V \leq N$. In the case where errors, treatment and other regressors are independent at the observation level, $C = M = U = V = N$.

We will need a substantial amount of additional notation. As before, $\mathbf{d}_i'$ refers to the i^{th} row of matrix $\mathbf{D}$, while we now use $\mathbf{D}_u$, $\mathbf{D}_c$, $\mathbf{D}_m$, and $\mathbf{D}_v$ to refer to the rows of $\mathbf{D}$ associated with the subscripted union, cluster, treatment or intersection grouping. d_{ij} refers to the ij^{th} element of $\mathbf{D}$, while $\mathbf{d}_{uj}$ (or similarly cj, mj or vj) refers to the rows of the j^{th} column of $\mathbf{D}$ associated with the union grouping u. We use the notation $\Sigma_{i\in u}$ to denote the sum across observations i in union grouping u (or cluster, treatment or intersection groupings c, m or v), and similarly $\Sigma_{m\subseteq u}$ denotes the sum across treatment groupings (or, similarly, cluster or intersection groupings c or v) in union grouping u. All "means" are calculated by dividing by M, and in addition to means across all observations, we also have means across subscripted groupings, defined as follows:

$$\text{(B.3)}\quad m(d_{i1}d_{i2}) = \sum_{i=1}^{N}\frac{d_{i1}d_{i2}}{M},\ \ m_g(d_{i1}d_{i2}) = \sum_{g=1}^{G}\sum_{i\in g}\frac{d_{i1}d_{i2}}{M},\ \ \&\ \ m_g(d_{i1},d_{j2}) = \sum_{g=1}^{G}\sum_{i\in g}\sum_{j\in g}\frac{d_{i1}d_{j2}}{M},$$

$$\text{where } m(d_{i1}d_{i2}) = m_g(d_{i1}d_{i2}) \text{ as } \sum_{i=1}^{N}\frac{d_{i1}d_{i2}}{M} = \sum_{g=1}^{G}\sum_{i\in g}\frac{d_{i1}d_{i2}}{M}$$

and where $g = c$, m or v and $G = C$, M or V. The reader will see that the m_c, m_m and m_v means for two variables are the divide-by-M means of the product of the variables summed at the c, m or v level (i.e. each "observation" for a variable equaling a sum across all observations i in a c, m or v grouping). As (B.3) shows, $m(d_{i1}d_{i2}) = m_g(d_{i1}d_{i2})$, but no such relation necessarily holds for summation across two groupings, as in $m_g(d_{i1},d_{j2})$. To distinguish matrices with M rows from those with N, we use scripted notation, as in the matrices $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{T}}$ versus $\mathbf{X}$ and $\mathbf{T}$ defined above. We further distinguish between the two by using subscripted m or i to identify the m^{th} or i^{th} row in each, as in $\boldsymbol{\mathcal{x}}_m'$ and $\mathbf{x}_i'$. We shall need to define other M matrices that are a function of the N matrices for non-treatment variables, as in $\boldsymbol{\mathcal{D}}$ versus $\mathbf{D}$. Whereas in the treatment matrices the m^{th} element represents the common treatment given to all elements i in the m^{th} grouping, in the case of other variables the scripted matrix refers to the sum of all elements i in the m^{th} grouping, i.e. $\mathcal{d}_{mq} = \Sigma_{i\in m}d_{iq}$. We use the notation $\omega()$ to denote the mean across M elements, as in

$$\text{(B.4)}\quad \omega(x_{mp}) = \sum_{m=1}^{M} \frac{x_{mp}}{M} \;\&\; \omega(d_{mq}) = \sum_{m=1}^{M} \frac{d_{mq}}{M}.$$

We note that for non treatment variables d, $\omega(d_{mq}) = m(d_{iq})$ & $\omega(x_{mp}d_{mq}) = m(x_{ip}d_{iq})$, as $x_{mp} = x_{ip}$ for all i in grouping m, but $\omega(x_{mp})$ does not necessarily equal $m(x_{ip})$. We shall use the ~ notation only with respect to demeaned M matrices and their elements, as in $\tilde{X} = O X$ where $O = \mathbf{I}_M - \mathbf{1}_M \mathbf{1}'_M / M$

Within this framework, we make the following White-type assumptions

(U1) (a) $(\mathbf{Z}_{+u}, \boldsymbol{\varepsilon}_u)$ is a sequence of independent but not necessarily identically distributed random matrices. (b) $E(\mathbf{z}_{+i}\varepsilon_i) = \mathbf{0}_{K+}$ for all i.

(U2) There exist positive finite constants δ, Δ and γ such that (a) for all u, $E(|\, \boldsymbol{\varepsilon}'_u \boldsymbol{\varepsilon}_u \,|^{1+\delta}) < \Delta$ and $E(|\, \mathbf{z}'_{+uj} \mathbf{z}_{+uk} \,|^{1+\delta}) < \Delta$ for all j, $k = 1...K_+$; (b) $\mathbf{M}_U = U^{-1} \sum_{u=1}^{U} E(\mathbf{Z}'_{+u} \mathbf{Z}_{+u})$ is non-singular for all U sufficiently large, with determinant $\mathbf{M}_U > \gamma > 0$.

(U3) There exist positive finite constants δ, Δ and γ such that (a) for all u, $E(|\, \mathbf{z}'_{+uj} \mathbf{z}_{+uj} \boldsymbol{\varepsilon}'_u \boldsymbol{\varepsilon}_u \,|^{1+\delta}) < \Delta$ for all $j = 1...K_+$; (b) $\mathbf{V}_U = U^{-1} \sum_{u=1}^{U} E(\mathbf{Z}'_{+u} \boldsymbol{\varepsilon}_u \boldsymbol{\varepsilon}'_u \mathbf{Z}_{+u})$ is non-singular for all U sufficiently large, with determinant $\mathbf{V}_U > \gamma > 0$.

(U4) There exist positive finite constants δ and Δ such that for all u & $j = 1...K_+$ $E(|\, (\mathbf{z}'_{+uj} \mathbf{z}_{+uj})^2 \,|^{1+\delta}) < \Delta$.

(U5) (a) If using the homoskedastic covariance estimate $\mathbf{V}_h(\hat{\boldsymbol{\gamma}}_+)$, the errors are iid with $E(\varepsilon_i^2 \,|\, \mathbf{z}_{+i}) = \sigma^2$ for all i & $E(\varepsilon_i \varepsilon_j \,|\, \mathbf{z}_{+i}, \mathbf{z}_{+j}) = 0$ for all $j \neq i$; (b) If using the heteroskedasticity (but not clustered) robust covariance estimate $\mathbf{V}_r(\hat{\boldsymbol{\gamma}}_+)$, the errors are independently but not necessarily identically distributed with $E(z_{+ik} \varepsilon_i \varepsilon_j z_{+jl}) = 0$ for all k & $l = 1...K_+$ if $j \neq i$; (c) If using the clustered robust covariance estimate $\mathbf{V}_{cl}(\hat{\boldsymbol{\gamma}}_+)$, the errors are independently distributed across clusters with $E(\mathbf{z}'_{+c_1 j} \boldsymbol{\varepsilon}_{c_1} \mathbf{z}'_{+c_2 k} \boldsymbol{\varepsilon}_{c_2}) = 0$ for all j, $k = 1...K_+$ if cluster $c_1 \neq c_2$.

U1 - U4 are a straightforward extension of White's work to allow for correlated regressors and errors across observation groupings and hence the following lemma is easily proven (in the next appendix):

Lemma B1: Assumptions U1 - U4 guarantee that for all U sufficiently large $\hat{\boldsymbol{\gamma}}_+$ exists, $\hat{\boldsymbol{\gamma}}_+ \xrightarrow{a.s.} \boldsymbol{\gamma}_+$, and $\sqrt{U}(\hat{\boldsymbol{\gamma}}_+ - \boldsymbol{\gamma}_+)$ is asymptotically (across the data generating process for the data sequence $\mathbf{Z}_+, \boldsymbol{\varepsilon}$) normally distributed with mean $\mathbf{0}_{K+}$ and covariance matrix $\mathbf{M}_U^{-1} \mathbf{V}_U \mathbf{M}_U^{-1}$. If U5a holds, $\hat{\mathbf{V}}_h(\hat{\boldsymbol{\gamma}}_+) \xrightarrow{a.s.} \mathbf{M}_U^{-1} \mathbf{V}_U \mathbf{M}_U^{-1}$; if U5b holds, $\hat{\mathbf{V}}_r(\hat{\boldsymbol{\gamma}}_+) \xrightarrow{a.s.} \mathbf{M}_U^{-1} \mathbf{V}_U \mathbf{M}_U^{-1}$; and if U5c holds, $\hat{\mathbf{V}}_{cl}(\hat{\boldsymbol{\gamma}}_+) \xrightarrow{a.s.} \mathbf{M}_U^{-1} \mathbf{V}_U \mathbf{M}_U^{-1}$.

Obviously, within the framework of the regression model, heteroskedasticity is just the case where the number of clusters $C = N$ and hence can be subsumed under clustering. However, in developing the randomization results further below, clusters that are larger than one observation generate additional complications in proofs (especially in Appendix C) as clustering can be below, above, at the same level as, or across treatment groupings. Consequently, I treat the heteroskedastic case as separate from clustering, even though at some points the exposition is redundant.

In addition to U1 - U5, we make four randomization inference specific assumptions

(A1) $\mathbf{G}_M = \sum_{m=1}^{M} E(\boldsymbol{\varkappa}_m \boldsymbol{\varkappa}_m')/M - \sum_{m=1}^{M} E(\boldsymbol{\varkappa}_m)/M \sum_{m=1}^{M} E(\boldsymbol{\varkappa}_m')/M$ is non-singular for all M sufficiently large with determinant $\mathbf{G}_M > \gamma > 0$.

(A2) Either the matrix $\mathbf{W}$ is part of $\mathbf{Z}$, i.e. the interactions with treatment in $\mathbf{X_W}$ are entered separately as covariates in the regression, or $E(\boldsymbol{\varkappa}_m) = \mathbf{0}_P$.

(A3) There exist positive finite constants θ, θ^*, Δ and γ, with $\theta(1+2\theta^*) > 1$, such that (a) for all m, q = 1...Q and p = 1...P, $E(|\mathbf{w}'_{mq}\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}'_m\mathbf{w}_{mq}|^{1+\theta}) < \Delta$ and $E(|\boldsymbol{\varkappa}_{mp}^4|^{1+\theta^*}) < \Delta$; (b) $\mathbf{W}_M = M^{-1}\sum_{m=1}^{M} E(\mathbf{W}'_m\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}'_m\mathbf{W}_m)$ is non-singular for all M sufficiently large, with determinant $\mathbf{W}_M > \gamma > 0$.

(A4) The maximum number of observations in a union grouping u, and by implication in a treatment grouping m or error correlation grouping c, is bounded from above by $\overline{N} < \infty$.

Assumptions A1, A2 and A3a are extensions of those given in the paper for observation level treatment to treatment groupings. As in the paper, we base the proofs on the version of A2 which states that $\mathbf{W}$ is part of $\mathbf{Z}$, as the alternative assumption is unlikely to hold. The condition A3b on $\mathbf{W}_M$ rules out cases where the average expectation of union grouped products as in $\mathbf{V}_U$ in U3b is positive definite but the average expectation of treatment grouped products as in A3b is not. A4 rules out asymptotically infinitely large cluster or treatment groupings, or overlaps across the two that generate infinite chains. It ensures that C, M, U and N are all of the same order, so that matrices that are positive definite when divided by one measure do not converge to matrices of 0s when divided by another and $\rightarrow \infty$ for one has the equivalent implication for the others. The formal asymptotics below are all stated as the units of treatment $M \rightarrow \infty$ as key elements are in terms of that measure.

With the assumptions given above, result (R1) in the text can be modified to read:

(R1) Given assumptions U1 - U4 and A1 - A4, for any $\boldsymbol{\beta}_0$ in a finite $\sqrt{M}$ neighbourhood of $\boldsymbol{\beta}$, i.e. such that $M(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'(\boldsymbol{\beta}-\boldsymbol{\beta}_0) < \Delta$ (a constant) $< \infty$, as $M \rightarrow \infty$ almost surely across the data generating process for $(\mathbf{Z}_+, \boldsymbol{\varepsilon})$ the distribution of $\sqrt{M}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0)$ across permutations $\boldsymbol{\mathcal{T}}$ of $\boldsymbol{\mathcal{X}}$ converges to that of the multivariate normal with mean $\mathbf{0}_{PQ}$ and almost surely bounded covariance matrix $\mathbf{C}_M$, while depending upon which of U5a, U5b or U5c hold, the homoskedastic, heteroskedasticity robust and clustered covariance estimates $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})$ converge in probability to $\mathbf{C}_M$, so that the Wald statistic $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ is asymptotically distributed chi-squared with PQ degrees of freedom and in probability converges to the value for the true null $\boldsymbol{\beta}_0 = \boldsymbol{\beta}$

$$\text{(B.5)}\quad \tau(\mathbf{T},\boldsymbol{\beta}_0) \overset{d(\mathbf{T})|a.s.(\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})}{\rightarrow} \chi^2_{PQ} \;\;\&\;\; \tau(\mathbf{T},\boldsymbol{\beta}_0) - \tau(\mathbf{T},\boldsymbol{\beta}) \overset{p(\mathbf{T})|a.s.(\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

Given Lemma B1, results (R2) - (R5) in the paper and its appendix then follow as before from (R1), as the Wald statistics for the original regression and for permutations $\boldsymbol{\mathcal{T}}$ of $\boldsymbol{\mathcal{X}}$ for tests of any subset or linear combination of parameters are both asymptotically distributed chi-squared with k degrees of freedom. The remainder of this appendix is dedicated to proving (R1). We begin by laying out some basic theorems and lemmas, and then examine the asymptotic distribution of coefficient estimates produced by

permutations $\mathcal{T}$ of $\mathcal{X}$ and the probability limit of their covariance estimates. Proofs of all lemmas used below are in Appendix C.

(a) Base Theorems and Lemmas

We restate Theorem's I and III in terms of the treatment groupings in which they will be applied:

Theorem I for Grouped Treatment:

Let $\boldsymbol{x}' = (x_1, \ldots, x_M)$ and $\boldsymbol{d}' = (d_1, \ldots, d_M)$ denote sequences of real numbers, not all equal, and $\boldsymbol{t}' = (t_1, \ldots, t_M)$ any of the $M!$ equally likely permutations of $\boldsymbol{x}$. Then as $M \to \infty$, the distribution of the random variable

$$\text{(Ia)}\quad n(t_m, d_m) = \sum_{i=1}^{M} \frac{[t_m - \omega(t_m)][d_m - \omega(d_m)]}{\left(\sum_{i=1}^{M} \frac{[x_m - \omega(x_m)]^2}{M} \sum_{i=1}^{M} \frac{[d_m - \omega(d_m)]^2}{M}\right)^{1/2}} M^{1/2}$$

as calculated across the realizations of $\boldsymbol{t}$ converges to that of the standard normal if for all integer $\tau > 2$

$$\text{(Ib)}\quad \lim_{M\to\infty} \frac{M^{\frac{\tau}{2}-1} \sum_{i=1}^{M} [x_m - \omega(x_m)]^\tau \sum_{i=1}^{M} [d_m - \omega(d_m)]^\tau}{\left(\sum_{i=1}^{M} [x_m - \omega(x_m)]^2\right)^{\tau/2} \left(\sum_{i=1}^{M} [d_m - \omega(d_m)]^2\right)^{\tau/2}} = 0.$$

If $\mathcal{X}' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$ and $\mathcal{D}' = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M)$ are sequences of vectors, and $\mathcal{T}$ any of the row permutations of $\mathcal{X}$, then the distribution of

$$\text{(Ic)}\quad \mathbf{n}(\boldsymbol{t}_m, \boldsymbol{d}_m) = \left(\frac{\tilde{\mathcal{X}}'\tilde{\mathcal{X}}}{M} \otimes \frac{\tilde{\mathcal{D}}'\tilde{\mathcal{D}}}{M}\right)^{-1/2} \frac{(\tilde{\mathcal{T}} \bullet \tilde{\mathcal{D}})' \mathbf{1}_M}{\sqrt{M}},$$

where $\otimes$ denotes the Kronecker product and $\bullet$ as above the row-by-row Kronecker or face-splitting product, is distributed multivariate iid standard normal if (Ib) holds for all pairwise combinations of the elements of the columns of $\mathcal{X}$ and $\mathcal{D}$ and the matrices $\tilde{\mathcal{X}}'\tilde{\mathcal{X}}/M$ and $\tilde{\mathcal{D}}'\tilde{\mathcal{D}}/M$ are bounded with determinant $> \gamma > 0$ for all M sufficiently large.

Theorem III for Grouped Treatment:

Let $\boldsymbol{x}' = (x_1, \ldots, x_M)$ and $\boldsymbol{d}' = (d_1, \ldots, d_M)$ denote sequences of real numbers, possibly all equal, and $\boldsymbol{t}' = (t_1, \ldots, t_M)$ any of the $M!$ equally likely permutations of $\boldsymbol{x}$. Then as $M \to \infty$, across the permutations $\boldsymbol{t}$ of $\boldsymbol{x}$ the random variable

$$\text{(IIIa)}\quad \omega(t_m d_m) - \omega(x_m)\omega(d_m) = \sum_{m=1}^{M} \frac{t_m d_m}{M} - \sum_{m=1}^{M} \frac{t_m}{M} \sum_{m=1}^{M} \frac{d_m}{M} = m(t_i d_i) - \omega(x_m) m(d_i) \xrightarrow{p} 0,$$

provided

$$\text{(IIIb)}\quad \lim_{M\to\infty} \frac{\sum_{m=1}^{M} \frac{[x_m - \omega(x_m)]^2}{M} \sum_{m=1}^{M} \frac{[d_m - \omega(d_m)]^2}{M}}{M} = 0.$$

If c_M is a sequence that converges to zero and the stronger condition

(IIIc) $\sum_{m=1}^{M}\frac{[x_m-\omega(x_m)]^2}{M}\sum_{m=1}^{M}\frac{[d_{qm}-\omega(d_{qm})]^2}{M}$ is asymptotically bounded

holds, then across the permutations t of x

(IIId) $\sqrt{M}[\omega(t_m d_{qm})-\omega(x_m)\omega(d_{qm})]c_M=\sqrt{M}[m(t_i d_i)-\omega(x_m)m(d_i)]c_M \xrightarrow{p} 0.$

The changes from the theorems given in the text are merely notational (using groupings of observations) and these theorems have already been proven in the paper and the appendix above. Theorem II in the paper also continues to hold, so that if conditions (Ib) and (IIIb) hold almost surely for the data sequence ($\mathbf{Z}_{+},\boldsymbol{\varepsilon}$), we can say that almost surely (across the data) $\mathbf{n}(t_m,d_{qm})$ converges in distribution and $\omega(t_m d_{qm})$ converges in probability across the row permutations T of X. As in the paper, with the exception of the clustered extension of White's result in Lemma B1 above, references to almost surely are with respect to the probability distribution governing the data sequence, whereas references to in distribution and in probability are with respect to the permutations T of X.

The following lemma will be useful below:

Lemma B2: Define W_{ε} as the M x Q matrix whose mq^{th} element $w_{mq\varepsilon}$ is the sum of the observational elements corresponding to the m^{th} treatment group in the N x Q matrix $\mathbf{W}_{\boldsymbol{\varepsilon}}$ (i.e. $w_{mq\varepsilon}=\sum_{i\epsilon m}w_{iq}\varepsilon_i$. If assumptions U1 - U4 and A1 - A4 hold, then:

(a) $\mathbf{Z}'\mathbf{Z}/M$, $\mathbf{W}'\mathbf{W}/M$, $\tilde{\mathbf{x}}'\tilde{\mathbf{x}}/M$ & $W_{\varepsilon}'W_{\varepsilon}/M$ are almost surely positive definite with determinant $>\gamma>0$ for all M sufficiently large, while $\tilde{W}_{\varepsilon}'\tilde{W}_{\varepsilon}/M-W_{\varepsilon}'W_{\varepsilon}/M\xrightarrow{a.s.}\mathbf{0}_{QxQ}$ and hence $\tilde{W}_{\varepsilon}'\tilde{W}_{\varepsilon}/M$ is also almost surely positive definite with determinant $>\gamma>0$ for all M sufficiently large.

(b) $\mathbf{Z}'\boldsymbol{\varepsilon}/M \xrightarrow{a.s.} \mathbf{0}_K$ & $\mathbf{X}_{\mathbf{W}}'\boldsymbol{\varepsilon}/M \xrightarrow{a.s.} \mathbf{0}_{PQ}$.

(c) Let $\prod_{k=1}^{n}x_{mk}$ denote the product of n = 1, 2, 3 or 4 elements of the columns of X, and $d_{i1}d_{i2}$ and $d_{i3}d_{i4}$ each the product of the elements of 2 columns of ($\mathbf{Z}_{+},\boldsymbol{\varepsilon}$) (with at most one in each being $\boldsymbol{\varepsilon}$). Then $|\omega(\prod_{k=1}^{n}x_{mk})|$, $|m(d_{i1}d_{i2})|$, $|m(d_{i1}d_{i2}d_{i3}d_{i4})|$, $|m_c(d_{i1}d_{i2},d_{i3}d_{i4})|$, $|m_m(d_{i1}d_{i2},d_{i3}d_{i4})|$ & $|m_v(d_{i1}d_{i2},d_{i3}d_{i4})|$ are all almost surely bounded, as are $(\mathbf{Z}'\mathbf{Z}/M)^{-1}$, $(\mathbf{W}'\mathbf{W}/M)^{-1}$, $(\tilde{\mathbf{x}}'\tilde{\mathbf{x}}/M)^{-1}$, $(W_{\varepsilon}'W_{\varepsilon}/M)^{-1}$ & $(\tilde{W}_{\varepsilon}'\tilde{W}_{\varepsilon}/M)^{-1}$.

(d) Let $\prod_{k=1}^{n}x_{mk}$ denote the product of $n>4$ elements from the columns of X and $d_{i1}d_{i2}$ and $d_{i3}d_{i4}$ each the product of the elements of two columns of ($\mathbf{Z}_{+},\boldsymbol{\varepsilon}$), with at most one in each case being $\boldsymbol{\varepsilon}$. Then for some a such that ½ > a > 0

$$M^{-a\left(\frac{n}{2}-2\right)}\omega(\prod_{k=1}^{n}x_{mk}) \xrightarrow{a.s.} 0,\ M^{-2a}m_c(d_{i1}^2d_{i2}^2,d_{i3}^2d_{i4}^2) \xrightarrow{a.s.} 0,$$

$$M^{-2a}m_m(d_{i1}^2d_{i2}^2,d_{i3}^2d_{i4}^2) \xrightarrow{a.s.} 0 \ \&\ M^{-2a}m_v(d_{i1}^2d_{i2}^2,d_{i3}^2d_{i4}^2) \xrightarrow{a.s.} 0.$$

(b) Asymptotic Distribution of Coefficient Estimates

The counterfactual outcome is given by $\mathbf{y}_{\mathbf{T},\boldsymbol{\beta}_0} = \mathbf{y} + (\mathbf{T} \bullet \mathbf{W} - \mathbf{X} \bullet \mathbf{W})\boldsymbol{\beta}_0$ and consequently, with $\mathbf{M} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ denoting the residual maker with respect to $\mathbf{Z}$, similar to the paper we have

$$\text{(B.6)}\quad \sqrt{M}\,(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0) = \left(\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}}{M}\right)^{-1}\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{X}_\mathbf{W}}{M}\mathbf{r} + \left(\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}}{M}\right)^{-1}\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\boldsymbol{\varepsilon}}{\sqrt{M}},$$

where $\mathbf{r} = \sqrt{M}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. This expression can be analyzed using the following lemma:

Lemma B3: Given assumptions U1 - U4 and A1 - A4:

(a) Condition IIIc of Theorem III almost surely holds for the mean of the product of the elements of one or two of the columns of $\mathbf{T}$ with the elements of two columns of $\mathbf{D} = (\mathbf{X}_\mathbf{W},\mathbf{Z},\boldsymbol{\varepsilon})$, no more than one of which is ε_i, so that in particular

$$m(t_{ip}d_{ij}d_{ik}) - \omega(x_{mp})m(d_{ij}d_{ik}) \xrightarrow{p} 0,\ m(t_{ip}t_{iq}d_{ij}d_{ik}) - \omega(x_{mp}x_{mq})m(d_{ij}d_{ik}) \xrightarrow{p} 0 \quad \&$$

$$\text{if } c_M \xrightarrow{a.s.} 0 \text{ then } \sqrt{M}\,[\,m(t_{ip}d_{ij}d_{ik}) - \omega(x_{mp})m(d_{ij}d_{ik})]c_M \xrightarrow{p} 0.$$

From Lemma B2c $\omega(x_{mp})$, $\omega(x_{mp}x_{mq})$ and $m(d_{ij}d_{ik})$ are known to be almost surely bounded.

(b) x_{mp} & $w_{mq\varepsilon}$ almost surely satisfy condition Ib of Theorem I for all column pairs of $\mathcal{X}$ and $\mathcal{W}_\varepsilon$, while $\tilde{\mathcal{X}}'\tilde{\mathcal{X}}/M$ & $\tilde{\mathcal{W}}_\varepsilon'\tilde{\mathcal{W}}_\varepsilon/M$ are almost surely bounded with determinant $> \gamma > 0$ for all M sufficiently large, so that across the row permutations $\mathcal{T}$ of $\mathcal{X}$ we have

$$\left(\frac{\tilde{\mathcal{X}}'\tilde{\mathcal{X}}}{M} \otimes \frac{\tilde{\mathcal{W}}_\varepsilon'\tilde{\mathcal{W}}_\varepsilon}{M}\right)^{-½}\frac{(\tilde{\mathcal{T}} \bullet \tilde{\mathcal{W}}_\varepsilon)'\mathbf{1}_{PQ}}{\sqrt{M}} \xrightarrow{d} \mathbf{n}_{PQ},\ \text{where } \mathbf{n}_{PQ} \sim \mathrm{N}(\mathbf{0}_{PQ},\mathbf{I}_{PQ}).$$

Moving forward, let $\mathbf{t}_{\mathbf{W}k}$ and $\mathbf{x}_{\mathbf{W}l}$ denote the k^{th} and l^{th} columns of $\mathbf{T}_\mathbf{W}$ and $\mathbf{X}_\mathbf{W}$, with the i^{th} elements of these vectors given by $t_{ip(k)}w_{iq(k)}$ and $x_{ip(l)}w_{iq(l)}$, where $p(j)$ and $q(j)$ denote the columns of $\mathbf{T}$ (or $\mathbf{X}$) and $\mathbf{W}$ associated with the j^{th} column of $\mathbf{T}_\mathbf{W}$ (or $\mathbf{X}_\mathbf{W}$). With this notation, we see that the kl^{th} element of $\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}/M$ can be expressed as

$$\text{(B.7)}\quad \frac{\mathbf{t}'_{\mathbf{W}k}\mathbf{M}\mathbf{t}_{\mathbf{W}l}}{M} = \frac{\mathbf{t}'_{\mathbf{W}k}[\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{t}_{\mathbf{W}l}}{M} = m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}) - \mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\mathbf{m}(t_{ip(l)}w_{iq(l)}\mathbf{z}_i),$$

$$\text{so}\quad \frac{\mathbf{t}'_{\mathbf{W}k}\mathbf{M}\mathbf{t}_{\mathbf{W}l}}{M} - [\omega(\boldsymbol{\varkappa}_{mp(k)}\boldsymbol{\varkappa}_{mp(l)}) - \omega(\boldsymbol{\varkappa}_{mp(k)})\omega(\boldsymbol{\varkappa}_{mp(l)})]m(w_{iq(k)}w_{iq(l)}) =$$

$$\underbrace{[m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}) - \omega(\boldsymbol{\varkappa}_{mp(k)}\boldsymbol{\varkappa}_{mp(l)})m(w_{iq(k)}w_{iq(l)})]}_{\xrightarrow{p}0\text{ (Lemma B3a)}} + \omega(\boldsymbol{\varkappa}_{mp(k)})\omega(\boldsymbol{\varkappa}_{mp(l)})[m(w_{iq(k)}w_{iq(l)}) - \underbrace{\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\mathbf{m}(w_{iq(l)}\mathbf{z}_i)}_{=m(w_{iq(k)}w_{iq(l)})\text{ for }M\text{ sufficiently large (Lemma B2a)}}\ ]$$

$$-\underbrace{[\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) - \omega(\boldsymbol{\varkappa}_{mp(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)]}_{\xrightarrow{p}\mathbf{0}'_K\text{ (Lemma B3a)}}\ \underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}}_{\text{bounded (Lemma B2c)}}\ \underbrace{[\mathbf{m}(t_{ip(l)}w_{iq(l)}\mathbf{z}_i) - \omega(\boldsymbol{\varkappa}_{mp(l)})\mathbf{m}(w_{iq(l)}\mathbf{z}_i)]}_{\xrightarrow{p}\mathbf{0}_K\text{ (Lemma B3a)}}$$

$$-\underbrace{\omega(\boldsymbol{\varkappa}_{mp(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}}_{\text{bounded (Lemma B2c)}}\ \underbrace{[\mathbf{m}(t_{ip(l)}w_{iq(l)}\mathbf{z}_i) - \omega(\boldsymbol{\varkappa}_{mp(l)})\mathbf{m}(w_{iq(l)}\mathbf{z}_i)]}_{\xrightarrow{p}\mathbf{0}_K\text{ (Lemma B3a)}}$$

$$-\underbrace{[\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) - \omega(\boldsymbol{\varkappa}_{mp(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)]}_{\xrightarrow{p}\mathbf{0}'_K\text{ (Lemma B3a)}}\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\omega(\boldsymbol{\varkappa}_{mp(l)})\mathbf{m}(w_{iq(l)}\mathbf{z}_i)}_{\text{bounded (Lemma B2c)}} \xrightarrow{p} 0$$

where $\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) = (m(t_{ip(k)}w_{iq(k)}z_{i1}), \ldots, m(t_{ip(k)}w_{iq(k)}z_{iK}))$ and we use the fact that as $\mathbf{m}(w_{iq(k)}\mathbf{z}'_i) = \mathbf{w}'_{q(k)}\mathbf{Z}/N$, where $\mathbf{w}_{q(k)}$ is the $q(k)^{th}$ column of $\mathbf{W}$ which is included in the covariates $\mathbf{Z}$ (assumption A2), so for all N sufficiently large that $\mathbf{Z}'\mathbf{Z}/M$ is guaranteed to be invertible $\mathbf{w}'_{q(k)}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ is a row vector of zeros with a 1 in the column corresponding to the position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$. Similarly, the kl^{th} element of $\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}/M$ can be expressed as[4]

$$\text{(B.8)}\quad \frac{\mathbf{t}'_{\mathbf{W}k}\mathbf{M}\mathbf{x}_{\mathbf{W}l}}{M} = \frac{\mathbf{t}'_{\mathbf{W}k}[\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{x}_{\mathbf{W}l}}{M} = m(t_{ip(k)}w_{iq(k)}w_{iq(l)}x_{ip(l)}) - \mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\mathbf{x}_{\mathbf{W}l}}{M}$$

$$= \underbrace{[m(t_{ip(k)}w_{iq(k)}w_{iq(l)}x_{ip(l)}) - \omega(\boldsymbol{\varkappa}_{mp(k)})m(w_{iq(k)}w_{iq(l)}x_{ip(l)})]}_{\xrightarrow{p}0\text{ (Lemma B3a)}} - \underbrace{[\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) - \omega(\boldsymbol{\varkappa}_{mp(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)]}_{\xrightarrow{p}\mathbf{0}'_K\text{ (Lemma B3a)}}\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\mathbf{x}_{\mathbf{W}l}}{M}}_{\text{bounded (Lemma B2c)}}$$

$$+\omega(\boldsymbol{\varkappa}_{mp(k)})m(w_{iq(k)}w_{iq(l)}x_{ip(l)}) - \omega(\boldsymbol{\varkappa}_{mp(k)})\ \underbrace{\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\mathbf{x}_{\mathbf{W}l}}{M}}_{=m(w_{iq(k)}w_{iq(l)}x_{ip(l)})\text{ for }M\text{ sufficiently large (Lemma B2a)}}\ \xrightarrow{p}\ 0,$$

where we again make use of the assumption that $\mathbf{W}$ is included in $\mathbf{Z}$. Combining these results, we have:

$$\text{(B.9)}\quad \frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{M} - \frac{\widetilde{\boldsymbol{\varkappa}}'\widetilde{\boldsymbol{\varkappa}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\xrightarrow{p}\mathbf{0}_{PQxPQ}\quad\&\quad\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}}{M}\xrightarrow{p}\mathbf{0}_{PQx\,PQ}.$$

Finite values of $\mathbf{r} = \sqrt{M}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, multiplied by $\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}/M$, asymptotically have no influence in (B.6).

The remaining part of (B.6) is the vector $\mathbf{T}'_{\mathbf{W}}\mathbf{M}\boldsymbol{\varepsilon}/\sqrt{M}$, the k^{th} term of which equals:

[4]In applying the Lemmas, keep in mind that $w_{iq(l)}x_{ip(l)}$ are the elements of one column of $\mathbf{X}_{\mathbf{W}}$.

$$\text{(B.10)}\quad \frac{\mathbf{t}'_{\mathbf{W}k}\mathbf{M}\boldsymbol{\varepsilon}}{\sqrt{M}} = \sqrt{M}\,m(t_{ip(k)}w_{iq(k)}\varepsilon_i) - \sqrt{M}\,\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{M} =$$

$$\underbrace{\sqrt{M}\left[\begin{matrix} m(t_{ip(k)}w_{iq(k)}\varepsilon_i) - \\ \omega(t_{mp(k)})m(w_{iq(k)}\varepsilon_i)\end{matrix}\right]}_{v_k} - \underbrace{\sqrt{M}\left[\begin{matrix}\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) - \\ \omega(t_{mp(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\end{matrix}\right]\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{M}}_{\xrightarrow{a.s.}\mathbf{0}_K\ \text{(Lemmas B2b,B2c)}}}_{\xrightarrow{p}0\ \text{(Lemma B3a)}} + \sqrt{M}\left[\begin{matrix}\omega(t_{mp(k)})m(w_{iq(k)}\varepsilon_i) \\ -\omega(t_{mp(k)})\underbrace{\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{M}}_{=m(w_{iq(k)}\varepsilon_i)\ \text{for all M sufficiently large (Lemma B2a)}}\end{matrix}\right],$$

so the only term that asymptotically is non-zero is v_k, which equals

$$\text{(B.11)}\quad v_k = \sqrt{M}\left(\sum_{i=1}^{N}\frac{t_{ip(k)}w_{iq(k)}\varepsilon_i}{M} - \omega(t_{mp(k)})m(w_{iq(k)}\varepsilon_i)\right) = \sqrt{M}\left(\sum_{i=1}^{M}\frac{t_{mp(k)}\sum_{i\in m}w_{iq(k)}\varepsilon_i}{M} - \omega(t_{mp(k)})m(w_{iq(k)}\varepsilon_i)\right)$$

$$= \sqrt{M}\left(\sum_{i=1}^{M}\frac{t_{mp(k)}w_{mq(k)\varepsilon}}{M} - \omega(t_{mp(k)})\omega(w_{mq(k)\varepsilon})\right) = \sum_{m=1}^{M}\frac{[t_{mp(k)} - \omega(t_{mp(k)})][w_{mq(k)\varepsilon} - \omega(w_{mq(k)\varepsilon})]}{\sqrt{M}},$$

which is the k^{th} element of $(\tilde{\boldsymbol{\mathcal{T}}} \bullet \tilde{\boldsymbol{\mathcal{W}}}_{\boldsymbol{\varepsilon}})'\mathbf{1}_{PQ}/\sqrt{M}$. Applying Lemma B3b we then see that

$$\text{(B.12)}\quad \left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\tilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{M}\right)^{-½}\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\boldsymbol{\varepsilon}}{\sqrt{M}} \xrightarrow{d} \mathbf{n}_{PQ},\ \text{where}\ \mathbf{n}_{PQ} \sim \mathrm{N}(\mathbf{0}_{PQ}, \mathbf{I}_{PQ}).$$

Combining the preceding results, we see that for finite $\mathbf{r} = \sqrt{M}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ we have

$$\text{(B.13)}\quad \left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\tilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{M}\right)^{-½}\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)\sqrt{M}\,(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0) =$$

$$\underbrace{\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\tilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{M}\right)^{-½}\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)}_{\text{a.s. bounded positive definite matrices (Lemmas B2a, B2c)}}\underbrace{\left(\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{M}\right)^{-1}}_{\xrightarrow{p}\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)^{-1}}\underbrace{\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}}{M}}_{\xrightarrow{p}\mathbf{0}_{PQxPQ}}\underbrace{\mathbf{r}}_{\text{bounded}} +$$

$$\underbrace{\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\tilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{M}\right)^{-½}\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)\left(\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{M}\right)^{-1}\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\tilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{M}\right)^{½}\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{M}\otimes\frac{\tilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\tilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{M}\right)^{-½}\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\boldsymbol{\varepsilon}}{\sqrt{M}}}_{\xrightarrow{p}\mathbf{I}_{PQ}\qquad\qquad\qquad\qquad\qquad\qquad\xrightarrow{d}\mathbf{n}_{PQ}} \xrightarrow{d} \mathbf{n}_{PQ}.$$

(c) Probability Limit of the Homoskedastic Covariance Estimate

The estimated residuals are given by

$$\text{(B.14)}\quad \hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\boldsymbol{\beta}_0} = \mathbf{M}\mathbf{y}_{\mathbf{T},\boldsymbol{\beta}_0} - \mathbf{M}\mathbf{T}_{\mathbf{W}}\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} = \mathbf{M}\mathbf{X}_{\mathbf{W}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \mathbf{M}\boldsymbol{\varepsilon} - \mathbf{M}\mathbf{T}_{\mathbf{W}}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0),$$

so using the fact that **MM** = **M** we see that the average squared residual equals

$$\text{(B.15)}\quad \frac{\hat{\boldsymbol{\varepsilon}}'_{\mathbf{T},\boldsymbol{\beta}_0}\hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\boldsymbol{\beta}_0}}{N} = \frac{\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}}{N} + \frac{M}{N}\left(\mathbf{r}'\frac{\mathbf{X}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}}{M^2}\mathbf{r} + \hat{\mathbf{r}}'\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{M^2}\hat{\mathbf{r}} + 2\frac{\boldsymbol{\varepsilon}'\mathbf{M}\mathbf{X}_{\mathbf{W}}}{M^{3/2}}\mathbf{r} - 2\hat{\mathbf{r}}'\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}}{M^2}\mathbf{r} - 2\hat{\mathbf{r}}'\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\boldsymbol{\varepsilon}}{M^{3/2}}\right),$$

where $\mathbf{r} = \sqrt{M}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ and $\hat{\mathbf{r}} = \sqrt{M}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0)$. From Lemma B2, (B.9) & (B.13) above, and the fact that $M/N \leq 1$, we have

$$\text{(B.16)}\quad \frac{\mathbf{X}_\mathbf{W}'\mathbf{M}\mathbf{X}_\mathbf{W}}{M} = \frac{\mathbf{X}_\mathbf{W}'\mathbf{X}_\mathbf{W}}{M} - \frac{\mathbf{X}_\mathbf{W}'\mathbf{Z}}{M}\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\mathbf{X}_\mathbf{W}}{M}\left[\begin{array}{c}\text{almost surely}\\ \text{bounded}\end{array}\right],$$

$$\frac{\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}}{N} - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{N} = -\frac{M}{N}\left[\frac{\boldsymbol{\varepsilon}'\mathbf{Z}}{M}\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{M}\right] \xrightarrow{a.s.} 0,\ \frac{\boldsymbol{\varepsilon}'\mathbf{M}\mathbf{X}_\mathbf{W}}{M} = \frac{\boldsymbol{\varepsilon}'\mathbf{X}_\mathbf{W}}{M} - \frac{\boldsymbol{\varepsilon}'\mathbf{Z}}{M}\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\frac{\mathbf{Z}'\mathbf{X}_\mathbf{W}}{M} \xrightarrow{a.s.} \mathbf{0}'_{PQ},$$

$$\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}}{M} - \underbrace{\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M} \otimes \frac{\mathbf{W}'\mathbf{W}}{M}}_{\text{almost surely bounded}} \xrightarrow{p.} \mathbf{0}_{PQxPQ},\ \frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{X}_\mathbf{W}}{M} \xrightarrow{p.} \mathbf{0}_{PQ\,x\,PQ},\ \&\ \frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\boldsymbol{\varepsilon}}{M} \xrightarrow{p.} \mathbf{0}_{PQ}.$$

Consequently, for the homoskedastic covariance estimate $\hat{\mathbf{V}}_h(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}) = (\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W})^{-1}(\hat{\boldsymbol{\varepsilon}}'_{\mathbf{T},\boldsymbol{\beta}_0}\hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\boldsymbol{\beta}_0}/N - K_+)$:

$$\text{(B.17)}\quad M\hat{\mathbf{V}}_h(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}) - \left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M} \otimes \frac{\mathbf{W}'\mathbf{W}}{M}\right)^{p}{}^{-1}\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{N} \to \mathbf{0}_{PQxPQ}\ .$$

We now note the following lemma:

Lemma B4: If as in U5a the errors are iid and homoskedastic, with $E(\varepsilon_i^2 \mid \mathbf{z}_{\perp i}) = \sigma^2$ & $E(\varepsilon_i\varepsilon_j \mid \mathbf{z}_{+i}, \mathbf{z}_{+j}) = 0$ for all i and $j \neq i$, then $\tilde{\boldsymbol{\mathcal{W}}}_\varepsilon'\tilde{\boldsymbol{\mathcal{W}}}_\varepsilon/M - (\mathbf{W}'\mathbf{W}/M)(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/N) \xrightarrow{a.s.} \mathbf{0}_{QxQ}$.

From Lemma B4, (B.13) & (B.17) we see that when the errors are iid, $M\hat{\mathbf{V}}_h(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})$ converges in probability to the asymptotic covariance matrix of normally distributed $\sqrt{M}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0)$, so the Wald statistic is asymptotically distributed chi-squared with PQ degrees of freedom. Moreover, every appearance of $\mathbf{r}$ in the Wald statistic $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ is multiplied by a term that almost surely across $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ in probability across permutations $\mathbf{T}$ converges to 0, so that in probability $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ converges to $\tau(\mathbf{T},\boldsymbol{\beta})$, as stated in (R1).

(d) Probability Limit of the Heteroskedasticity Robust Covariance Estimate

For the heteroskedasticity robust covariance estimate we have

$$\text{(B.18)}\ M\mathbf{V}_r(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}) = \left(\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}}{M}\right)^{-1}\mathbf{A}\left(\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}}{M}\right)^{-1}, \text{where } \mathbf{A} = \frac{(\mathbf{M}\mathbf{T}_\mathbf{W} \bullet \hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\boldsymbol{\beta}_0})'(\mathbf{M}\mathbf{T}_\mathbf{W} \bullet \hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\boldsymbol{\beta}_0})}{M},$$

with kl^{th} term given by

$$\text{(B.19)}\quad \mathbf{A}_{kl} = \frac{1}{M}\sum_{i=1}^{N}(t_{ip(k)}w_{iq(k)} - \sum_{a=1}^{K} z_{ia}\hat{\delta}_{ak})(t_{ip(l)}w_{iq(l)} - \sum_{b=1}^{K} z_{ib}\hat{\delta}_{bl})\hat{\varepsilon}^2_{\mathbf{T},\boldsymbol{\beta}_0 i},$$

$$\text{where } \hat{\varepsilon}_{\mathbf{T},\boldsymbol{\beta}_0 i} = \varepsilon_i - \sum_{c=1}^{K} z_{ic}\hat{\eta}_c + \sum_{d=1}^{PQ}(x_{ip(d)}w_{iq(d)} - \sum_{e=1}^{K} z_{ie}\hat{\tau}_{ed})\frac{r_d}{\sqrt{M}} - \sum_{f=1}^{PQ}(t_{ip(f)}w_{iq(f)} - \sum_{g=1}^{K} z_{ig}\hat{\delta}_{gf})\frac{\hat{r}_f}{\sqrt{M}},$$

using the formula for $\hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\boldsymbol{\beta}_0}$ from (B.14) earlier with $\mathbf{r} = \sqrt{M}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, $\hat{\mathbf{r}} = \sqrt{M}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0)$, $\hat{\boldsymbol{\delta}}_\iota = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{t}_{\mathbf{W}k}$, $\hat{\boldsymbol{\tau}}_k = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}_{\mathbf{W}k}$ and $\hat{\boldsymbol{\eta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}$. From Lemmas B2b and B2c we have $\hat{\boldsymbol{\eta}} \xrightarrow{a.s.} \mathbf{0}_K$ and from B2c know that the limit of $\hat{\boldsymbol{\tau}}_k$ is almost surely bounded. As for $\hat{\boldsymbol{\delta}}_k$ its plim across the distribution of $\mathbf{T}$ is bounded as

$$\text{(B.20)}\ \hat{\boldsymbol{\delta}}_k - \underbrace{\omega(\boldsymbol{\varkappa}_{mp(k)})\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\mathbf{m}(w_{iq(k)}\mathbf{z}_i)}_{\text{bounded (Lemma B2c)}} = \underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}}_{\text{bounded (Lemma B2c)}} \underbrace{[\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}_i) - \omega(\boldsymbol{\varkappa}_{mp(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}_i)]}_{\xrightarrow{p}\mathbf{0}_K\ \text{(Lemma B3a)}} \xrightarrow{p} 0 \cdot$$

As for all N sufficiently large $(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{w}_{q(k)}$ is a vector of zeros with a 1 in the row corresponding to the column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$ and $\omega(\boldsymbol{\varkappa}_{mp(k)})$ is known to be bounded by Lemma B2c, plim $\hat{\delta}_{ak} = 0$ unless a is the column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$, in which case plim $\hat{\delta}_{ak} - \omega(\boldsymbol{\varkappa}_{mp(k)}) = 0$. The elements of $\mathbf{r}$ are finite and of $\hat{\mathbf{r}}$ are asymptotically multivariate normal, so when divided by any positive power of M have a probability limit of zero.

When the terms in (B.19) are multiplied out, most involve a product with an element of $\mathbf{r}/\sqrt{M}$, $\hat{\mathbf{r}}/\sqrt{M}$, or $\hat{\boldsymbol{\eta}}$ that has a plim of zero, parameters $\hat{\tau}$ and $\hat{\delta}$ with bounded probability limits, and the mean of the product of the elements of 0 to 4 columns of $\mathbf{T}$ and the elements of 4 columns of $\mathbf{D} = (\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})$ (no more than two of which are ε_i). The following lemma allows us to conclude that the plim of all such terms is zero:

Lemma B5: Assumptions U1 - U4 and the additional A1 - A4 ensure that for some a in (0,½) condition IIIb of Theorem III almost surely holds for the mean of the product of the elements of n = 1, 2, 3, or 4 columns of $\mathbf{T}$ divided by $N^{a\max(n-2,0)}$ with the elements of four columns of $\mathbf{D} = (\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})$, no more than two of which are ε, so that across the permutations $\boldsymbol{\mathcal{T}}$ of $\boldsymbol{\mathcal{X}}$

$$m(M^{-a\max(n-2,0)}(\prod_{o=1}^{n} t_{ip(o)})d_{ij}d_{ik}d_{il}d_{im}) - \omega(M^{-a\max(n-2,0)}\prod_{o=1}^{n}\boldsymbol{\varkappa}_{mp(o)})m(d_{ij}d_{ik}d_{il}d_{im}) \xrightarrow{p} 0\ \cdot$$

From Lemma B2c we know that the sample means of the product of the elements of one through four columns of $\mathbf{X}$ or four columns of $\mathbf{D}$ are almost surely bounded, so the probability limit in Lemma B5 is bounded when n = 1 or 2 and 0 when n = 3 or 4. Consequently, in (B.19) every term that involves the product of an element of $\mathbf{r}/\sqrt{M}$, $\hat{\mathbf{r}}/\sqrt{M}$, or $\hat{\boldsymbol{\eta}}$ that has a plim of zero with the mean of the product of four columns of $\mathbf{D}$ with zero, one or two columns of $\mathbf{T}$ has a probability limit of zero. Every term in (B.19) that involves the product of n = 3 or 4 columns of $\mathbf{T}$ with four columns of $\mathbf{D}$ also includes at least n - 2 $\hat{\mathbf{r}}/\sqrt{M}$ terms which can be re-expressed as $(\hat{\mathbf{r}}/M^{½-a})(1/M^{a})$ for some a in (0,½). The $1/M^{a}$ parts can be used to satisfy Lemma B5, while the $\hat{\mathbf{r}}/M^{½-a}$ part converges in probability to 0. Thus, all such terms also have a plim of 0.

The above only leaves terms in (B.19) that involve the product of two or less columns of $\mathbf{T}$ and do not include an element of $\mathbf{r}/\sqrt{M}$, $\hat{\mathbf{r}}/\sqrt{M}$, or $\hat{\boldsymbol{\eta}}$, namely

$$\text{(B.21)}\ \sum_{i=1}^{N}\frac{t_{ip(k)}w_{iq(k)}t_{ip(l)}w_{iq(l)}\varepsilon_i^2}{M}-\sum_{a=1}^{K}\hat{\delta}_{ak}\sum_{i=1}^{N}\frac{t_{ip(l)}w_{iq(l)}z_{ia}\varepsilon_i^2}{M}-\sum_{b=1}^{K}\hat{\delta}_{bl}\sum_{i=1}^{N}\frac{t_{ip(k)}w_{iq(k)}z_{ib}\varepsilon_i^2}{M}+\sum_{a=1}^{K}\sum_{b=1}^{K}\hat{\delta}_{ak}\hat{\delta}_{bl}\sum_{i=1}^{N}\frac{z_{ia}z_{ib}\varepsilon_i^2}{M}$$

$$=m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}\varepsilon_i^2)-\sum_{a=1}^{K}\hat{\delta}_{ak}m(t_{ip(l)}w_{iq(l)}z_{ia}\varepsilon_i^2)-\sum_{b=1}^{K}\hat{\delta}_{bl}m(t_{ip(k)}w_{iq(k)}z_{ib}\varepsilon_i^2)+\sum_{a=1}^{K}\sum_{b=1}^{K}\hat{\delta}_{ak}\hat{\delta}_{bl}m(z_{ia}z_{ib}\varepsilon_i^2)$$

$$\text{where}\ \ m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}\varepsilon_i^2)-\omega(\boldsymbol{x}_{mp(k)}\boldsymbol{x}_{mp(l)})m(w_{iq(k)}w_{iq(l)}\varepsilon_i^2)\underbrace{\xrightarrow{p}0}_{\text{Lemma B5}}$$

$$\&\ \ m(t_{ip(l)}w_{iq(l)}z_{ia}\varepsilon_i^2)-\omega(\boldsymbol{x}_{mp(l)})m(w_{iq(l)}z_{ia}\varepsilon_i^2)\underbrace{\xrightarrow{p}0}_{\text{Lemma B5}},$$

$$\text{so}\ \mathbf{A}_{kl}-[\omega(\boldsymbol{x}_{mp(k)}x_{mp(l)})-\omega(\boldsymbol{x}_{mp(k)})\omega(\boldsymbol{x}_{mp(l)})]m(w_{iq(k)}w_{iq(l)}\varepsilon_i^2)\xrightarrow{p}0,$$

where we use the boundedness of means of products of up to four terms (Lemma B2c) and the fact noted above that plim $\hat{\delta}_{ak}=0$ unless a is the column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$, in which case plim $\hat{\delta}_{ak}-\omega(\boldsymbol{x}_{mp(k)})=0$ and $z_{ia}=w_{iq(k)}$. This allows us to state that

$$\text{(B.22)}\ \ \mathbf{A}-\frac{\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}}}{M}\otimes\frac{\mathbf{W}_{\boldsymbol{\varepsilon}}'\mathbf{W}_{\boldsymbol{\varepsilon}}}{M}\xrightarrow{p}\mathbf{0}_{PQxPQ},$$

and consequently for the heteroskedasticity robust covariance estimate we have

$$\text{(B.23)}\ M\mathbf{V}_r(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})-\left(\frac{\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)^{-1}\left(\frac{\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}}}{M}\otimes\frac{\mathbf{W}_{\boldsymbol{\varepsilon}}'\mathbf{W}_{\boldsymbol{\varepsilon}}}{M}\right)\left(\frac{\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)^{-1}\xrightarrow{p}\mathbf{0}_{PQxPQ}.$$

We now note the following lemma:

Lemma B6: If as in U5b the errors are independently but not identically distributed, with $E(z_{+ik}\varepsilon_i\varepsilon_j z_{+jl})=0$ for all k & $l=1...K_+$ if $j\neq i$, then $\tilde{\boldsymbol{w}}_{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{w}}_{\boldsymbol{\varepsilon}}/M-(\mathbf{W}_{\boldsymbol{\varepsilon}}'\mathbf{W}_{\boldsymbol{\varepsilon}}/M)\xrightarrow{a.s.}\mathbf{0}_{QxQ}$.

From Lemma B6, (B.13) & (B.23) we see that when the errors are heteroskedastic but not clustered, $M\hat{\mathbf{V}}_r(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})$ converges in probability to the asymptotic covariance matrix of normally distributed $\sqrt{M}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}-\boldsymbol{\beta}_0)$, so the Wald statistic is asymptotically distributed chi-squared with PQ degrees of freedom. Moreover, every appearance of $\mathbf{r}$ in the equation for the Wald statistic $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ is multiplied by a term that almost surely across $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ in probability across permutations $\mathbf{T}$ converges to 0, so that in probability $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ converges to $\tau(\mathbf{T},\boldsymbol{\beta})$, as stated in (R1).

(e) Probability Limit of the Clustered Robust Covariance Estimate

For the clustered robust covariance estimate we again use the sandwich formula (B.18), but this time with the kl^{th} element of $\mathbf{A}$ given by

$$\text{(B.24)}\ \ \mathbf{A}_{kl}=\frac{1}{M}\sum_{c=1}^{C}\left(\sum_{i\in c}(t_{ip(k)}w_{iq(k)}-\sum_{a=1}^{K}z_{ia}\hat{\delta}_{ak})\hat{\varepsilon}_{\mathbf{T},\boldsymbol{\beta}_0 i}\right)\left(\sum_{j\in c}(t_{jp(l)}w_{jq(l)}-\sum_{b=1}^{K}z_{jb}\hat{\delta}_{bl})\hat{\varepsilon}_{\mathbf{T},\boldsymbol{\beta}_0 j}\right)$$

$$\text{where}\ \ \hat{\varepsilon}_{\mathbf{T},\boldsymbol{\beta}_0 i}=\varepsilon_i-\sum_{c=1}^{K}z_{ic}\hat{\eta}_c+\sum_{d=1}^{PQ}(x_{ip(d)}w_{iq(d)}-\sum_{e=1}^{K}z_{ie}\hat{\tau}_{ed})\frac{r_d}{\sqrt{M}}-\sum_{f=1}^{PQ}(t_{ip(f)}w_{iq(f)}-\sum_{g=1}^{K}z_{ig}\hat{\delta}_{gf})\frac{\hat{r}_f}{\sqrt{M}},$$

and where, as before, the notation $i \in c$ denotes the summation across the set of observations i in cluster c. The following lemma is useful (the reader is reminded that subscript v in m_v denotes the treatment cross cluster intersection grouping, as described earlier above):

Lemma B7: Let $t_{i1}...t_{i4}$ denote columns of $\mathbf{T}$ and v_{i1} and v_{j2} each the product of the elements of two columns of $\mathbf{D} = (\mathbf{Z}_{+},\boldsymbol{\varepsilon})$, no more than one of which in each case is $\boldsymbol{\varepsilon}$. Given assumptions U1 - U4 and A1- A4, for some a in (0,½)

$$\text{(L7a)}\quad m_c(t_{i1}v_{i1},v_{j2}) - \omega(x_{m1})m_c(v_{i1},v_{j2}) \xrightarrow{p} 0 \;\; \& \; m_c(t_{i1}t_{i2}v_{i1},v_{j2}) - \omega(x_{m1}x_{m2})m_c(v_{i1},v_{j2}) \xrightarrow{p} 0$$

$$\text{(L7b)}\quad m_c(t_{i1}v_{i1},t_{j2}v_{j2}) - \left([\omega(x_{m1}x_{m2}) - \omega(x_{m1})\omega(x_{m2})]m_v(v_{i1},v_{j2}) + \omega(x_{m1})\omega(x_{m2})m_c(v_{i1},v_{j2})\right) \xrightarrow{p} 0,$$

$$\text{(L7c)}\quad M^{-a}m_c(t_{i1}t_{i2}v_{i1},t_{j3}v_{j2}) \xrightarrow{p} 0 \;\; \& \;\; M^{-2a}m_c(t_{i1}t_{i2}v_{i1},t_{j3}t_{j4}v_{j2}) \xrightarrow{p} 0.$$

$\omega(\prod_{k=1}^{n} x_{mk})$ for n = 1 to 4, $m_c(v_{i1},v_{j2})$ & $m_v(v_{i1},v_{j2})$ are almost surely bounded (Lemma B2c).

As in the case of the robust covariance calculation earlier above, when multiplied out most of the terms in (B.24) involve an element of $\mathbf{r}/\sqrt{M}$, $\hat{\mathbf{r}}/\sqrt{M}$, or $\hat{\boldsymbol{\eta}}$ whose plim is zero, multiplied possibly by a parameter from $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\delta}}$ whose plim is bounded, times one of the m_c terms described in Lemma 7 or simply $m_c(v_{i1},v_{j2})$. From Lemma B2c we know that the $\omega()$ sample means of the product of the elements of n = one through four columns of $\mathbf{X}$ and the $m_c()$ sample means of the product of the elements of four columns of $(\mathbf{Z}_{+},\boldsymbol{\varepsilon})$ (no more than two of which are $\boldsymbol{\varepsilon}$) are almost surely bounded. Consequently, in (B.24) every term that involves the product of an element of $\mathbf{r}/\sqrt{M}$, $\hat{\mathbf{r}}/\sqrt{M}$, or $\hat{\boldsymbol{\eta}}$ that has a plim of zero with the mean of the product of four columns of $\mathbf{D}$ with zero, one or two columns of $\mathbf{T}$ has a probability limit of zero. Every term in (B.24) that involves the product of n = three or four columns of $\mathbf{T}$ with four columns of $\mathbf{D}$ also includes at least n - 2 $\hat{\mathbf{r}}/\sqrt{M}$ terms which can be re-expressed as $(\hat{\mathbf{r}}/M^{½-a})(1/M^{a})$ for some a in (0,½). The $1/M^{a}$ parts can be used to satisfy Lemma B7c, while the $\hat{\mathbf{r}}/M^{½-a}$ part converges in probability to 0. Consequently, the plim of all such terms is 0 and we need only focus on terms in (B.24) which do not involve an element of $\mathbf{r}/\sqrt{M}$, $\hat{\mathbf{r}}/\sqrt{M}$, or $\hat{\boldsymbol{\eta}}$. These are

$$\text{(B.25)}\quad \underbrace{\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{t_{ip(k)}w_{iq(k)}\varepsilon_i t_{jp(l)}w_{jq(l)}\varepsilon_j}{M}}_{m_c(t_{ip(k)}w_{iq(k)}\varepsilon_i,t_{jp(l)}w_{jq(l)}\varepsilon_j)} - \sum_{b=1}^{K}\hat{\delta}_{bl}\underbrace{\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{t_{ip(k)}w_{iq(k)}\varepsilon_i z_{jb}\varepsilon_j}{M}}_{m_c(t_{ip(k)}w_{iq(k)}\varepsilon_i,z_{jb}\varepsilon_j)}$$
$$-\sum_{a=1}^{K}\hat{\delta}_{ak}\underbrace{\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{z_{ia}\varepsilon_i t_{jp(l)}w_{jq(l)}\varepsilon_j}{M}}_{m_c(t_{ip(l)}w_{iq(l)}\varepsilon_i,z_{ja}\varepsilon_j)} + \sum_{a=1}^{K}\sum_{b=1}^{K}\hat{\delta}_{ak}\hat{\delta}_{bl}\underbrace{\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{z_{ia}\varepsilon_i z_{jb}\varepsilon_j}{M}}_{m_c(z_{ia}\varepsilon_i,z_{jb}\varepsilon_j)},$$

so using Lemma B7 we see that

$$\text{(B.26)}\quad \mathbf{A}_{kl} - [\omega(x_{mp(k)}x_{mp(l)}) - \omega(x_{mp(k)})\omega(x_{mp(l)})]m_v(w_{iq(k)}\varepsilon_i, w_{jq(l)}\varepsilon_j) \xrightarrow{p} 0,$$

where we once again use the fact that plim $\hat{\delta}_{ak}$ is only non-zero in the "a" column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$, with plim $\hat{\delta}_{ak} = \omega(x_{mp(k)})$ and $z_{ia} = w_{iq(k)}$. Consequently, for the clustered robust covariance estimate

$$\text{(B.27)}\quad \mathbf{A} - \frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M} \otimes \frac{\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon}'\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon}}{M} \xrightarrow{p} \mathbf{0}_{PQxPQ} \quad \left[\text{where } \frac{\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon}'\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon}}{M} = \sum_{v=1}^{V}\sum_{i\in v}\sum_{j\in v} \frac{\mathbf{w}_i \varepsilon_i \varepsilon_j \mathbf{w}'_j}{M} = \sum_{v=1}^{V} \frac{\mathbf{W}_v' \boldsymbol{\varepsilon}_v \boldsymbol{\varepsilon}_v' \mathbf{W}_v}{M} \right]$$

$$\Rightarrow M\mathbf{V}_{cl}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}) - \left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M} \otimes \frac{\mathbf{W}'\mathbf{W}}{M} \right)^{-1} \left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M} \otimes \frac{\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon}'\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon}}{M} \right) \left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M} \otimes \frac{\mathbf{W}'\mathbf{W}}{M} \right)^{-1} \xrightarrow{p} \mathbf{0}_{PQxPQ}.$$

We now note the following lemma:

Lemma B8: If as in U5c $E(\mathbf{z}'_{+c_1,i}\boldsymbol{\varepsilon}_{c_1}\mathbf{z}'_{+c_2,k}\boldsymbol{\varepsilon}_{c_2}) = 0$ for all j, $k = 1...K_+$ if cluster $c_1 \neq c_2$,
then $\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon}'\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon} / M - \tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon}'\tilde{\boldsymbol{\mathcal{W}}}_{\varepsilon} / M \xrightarrow[a.s.]{} \mathbf{0}_{QxQ}$.

From Lemma B8, (B.13) & (B.27) we see that when the errors are clustered, $M\hat{\mathbf{V}}_{cl}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})$ converges in probability to the asymptotic covariance matrix of normally distributed $\sqrt{M}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0)$, so the Wald statistic is asymptotically distributed chi-squared with PQ degrees of freedom. Moreover, every appearance of $\mathbf{r}$ in the equation for the Wald statistic $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ is multiplied by a term that almost surely across $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ in probability across $\mathbf{T}$ converges to 0, so that in probability $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ converges to $\tau(\mathbf{T},\boldsymbol{\beta})$, as stated in (R1).

C. Proofs of Lemmas used in Appendix B

We make use below of the corollaries to Markov's Law of Large Numbers, the Continuous Mapping Theorem and the Borel-Cantelli Lemma given in Appendix C in the paper. In applying the corollaries, we can treat the random variables generated by vector inner products as a single observation, as in $\boldsymbol{\varepsilon}_u'\boldsymbol{\varepsilon}_u$ or $\boldsymbol{\varepsilon}_m'\boldsymbol{\varepsilon}_m$. An issue that arises, however, is that the Markov Corollary is stated in terms of independent random variables. Given assumption U1 above, this is always true for observations made by the inner product of union groupings *u*, but it need not necessarily be true for observations made by the inner product of groupings based upon *m*, *c* or *v*, as these are subsets of *u*. We can, however, apply a Corollary to a Law of Large Numbers for Heterogeneous Dependent Sequences (hereafter, LLNHDS) given by White (1984):

LLNHDS Corollary: For the Borel field generated by the random variable $d_i(\omega)$, $i = n \,..\, n + m$ & ω in Ω, let $B_n^{n+m} = \sigma(d_n,...,d_{n+m})$ be the smallest σ - field of subsets of Ω with respect to which $d_i(\omega)$, $i = n \,..\, n + m$, are measureable. Let $B_{-\infty}^{n} = \sigma(...,i_n)$ be the smallest collection of subsets of Ω that contains the union of the σ - fields B_a^n as $a \to -\infty$ and $B_{n+m}^{\infty} = \sigma(i_{n+m},...)$ be the smallest collection of subsets of Ω that contains the union of the σ - fields B_{n+m}^{a} as $a \to \infty$. Let $\mathcal{G}$ and $\mathcal{H}$ be σ-fields and define $\phi(\mathcal{G},\mathcal{H}) \equiv \sup_{\{G\in\mathcal{G},H\in\mathcal{H},P(G)>0\}} |P(H \mid G) - P(H)|$. Define the mixing coefficients $\phi(m) \equiv \sup_n \phi(B_{-\infty}^{n}, B_{n+m}^{\infty})$. Let $\{d_i\}$ be a sequence with $\phi(m) = O(m^{-\lambda})$ for $\lambda > r/(2r\text{-}1)$, *r* a real number with $1 \le r \le \infty$, such that $E(|d_i|^{r+\delta}) < \Delta < \infty$ for some $\delta > 0$ and all *i*. Then

$$\sum_{i=1}^{N}\frac{d_i}{N} - \sum_{i=1}^{N}\frac{E(d_i)}{N} \xrightarrow{a.s.} 0.$$

In the case of variables which are γ independent, i.e. d_i is independent of $d_{i-\tau}$ for all $\tau > \gamma$ and all *i*, $\phi(m) = 0$ for all $m > \gamma$, *r* is 1, and the requirement for the result becomes $E(|d_i|^{1+\delta}) < \Delta < \infty$ for some $\delta > 0$ and all *i*. In this case, the requirement for the Strong Law is the same as used in the Markov Corollary in the paper. Since by assumption U1 $(\mathbf{Z}_{+u}, \boldsymbol{\varepsilon}_u)$ is a sequence of independent random matrices, and each union grouping has no more than $\overline{N}$ observations, we see that all variables are $\overline{N}$ independent. This allows us to apply the LLNHDS Corollary with $r = 1$ to the means of groupings below the union level, such as $\boldsymbol{\varepsilon}_m'\boldsymbol{\varepsilon}_m$.

Lemma B1: The proof of this lemma follows proofs given in White (1980), with notation and cases adapted to our specific framework. All variables are independent across union groupings, so assumption U2a and the Markov Corollary guarantee that

$$\text{(C1.1)}\quad \frac{\mathbf{Z}_+'\mathbf{Z}_+}{U} - \mathbf{M}_U = \sum_{u=1}^{U}\frac{\mathbf{Z}_{+u}'\mathbf{Z}_{+u}}{U} - \sum_{u=1}^{U}\frac{E(\mathbf{Z}_{+u}'\mathbf{Z}_{+u})}{U} \xrightarrow{a.s.} \mathbf{0}_{K_+xK_+}.$$

$\mathbf{M}_U$ is almost surely positive definite for all *U* sufficiently large with determinant greater than some $\gamma > 0$ (assumption U2b), so by the Continuous Mapping Theorem Corollary $\mathbf{Z}_+'\mathbf{Z}_+/U$ is almost surely positive definite for all *U* sufficiently large with determinant greater than $\gamma > 0$. Consequently, $\hat{\boldsymbol{\gamma}}_+$ exists for all *U* sufficiently large. U3a implies that

$$\text{(C1.2) } E(|\,\mathbf{z}'_{+uj}\boldsymbol{\varepsilon}_u\boldsymbol{\varepsilon}'_u\mathbf{z}_{+uk}\,|^{1+\delta}) \overbrace{\leq \sqrt{\prod_{m=j,k} E(|\,\mathbf{z}'_{+um}\boldsymbol{\varepsilon}_u\,|^{2(1+\delta)})}}^{\text{Hölder's Inequality}} \overbrace{\leq \sqrt{\prod_{m=j,k} E(|\,\mathbf{z}'_{+um}\mathbf{z}_{+um}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u\,|^{(1+\delta)})}}^{\text{Cauchy-Schwarz Inequality}} \overbrace{< \Delta}^{\text{U3a}},$$

for all j, $k = 1...K_+$. Using this and Jensen's Inequality then ensures that for all u and j

$$\text{(C1.3) } E(|\,\mathbf{z}'_{+uj}\boldsymbol{\varepsilon}_u\,|^{1+\delta}) \leq E(|\,\mathbf{z}'_{+uj}\boldsymbol{\varepsilon}_u\boldsymbol{\varepsilon}'_u\mathbf{z}_{+uj}\,|^{1+\delta})^{1/2} < \Delta^{1/2},$$

which along with the Markov Corollary and U1b indicates that

$$\text{(C1.4) } \sum_{u=1}^{U}\frac{\mathbf{Z}'_{+u}\boldsymbol{\varepsilon}_u}{U} - \sum_{u=1}^{U}\frac{E(\mathbf{Z}'_{+u}\boldsymbol{\varepsilon}_u)}{U} \xrightarrow{a.s.} \mathbf{0}_{K_+}$$

$$\Rightarrow \sum_{u=1}^{U}\frac{\mathbf{Z}'_{+u}\boldsymbol{\varepsilon}_u}{U} = \frac{\mathbf{Z}'_{+}\boldsymbol{\varepsilon}}{U} \xrightarrow{a.s.} \mathbf{0}_{K_+} \text{ as } \sum_{u=1}^{U}\frac{E(\mathbf{Z}'_{+u}\boldsymbol{\varepsilon}_u)}{U} = \sum_{i=1}^{N}\frac{E(\mathbf{z}_{+i}\varepsilon_i)}{U} = \sum_{i=1}^{N}\frac{\mathbf{0}_{K+}}{U}.$$

This guarantees that

$$\text{(C1.5) } \hat{\boldsymbol{\gamma}}_+ = (\mathbf{Z}'_+\mathbf{Z}_+)^{-1}\mathbf{Z}'_+(\mathbf{Z}'_+\boldsymbol{\gamma}_+ + \boldsymbol{\varepsilon}) = \boldsymbol{\gamma}_+ + \left(\frac{\mathbf{Z}'_+\mathbf{Z}_+}{U}\right)^{-1}\frac{\mathbf{Z}'_+\boldsymbol{\varepsilon}}{U} \xrightarrow{a.s.} \boldsymbol{\gamma}_+.$$

$\mathbf{Z}'_+\boldsymbol{\varepsilon}/\sqrt{U}$ is a vector with expectation and variance:

$$\text{(C1.6) } E\left(\frac{\mathbf{Z}'_+\boldsymbol{\varepsilon}}{\sqrt{U}}\right) = \sum_{i=1}^{N}\frac{E(\mathbf{z}_{+i}\varepsilon_i)}{\sqrt{U}} = \mathbf{0}_{K+},\ E\left(\frac{\mathbf{Z}'_+\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{Z}_+}{U}\right) = \sum_{u=1}^{U}\frac{E(\mathbf{Z}'_{+u}\boldsymbol{\varepsilon}_u\boldsymbol{\varepsilon}'_u\mathbf{Z}_{+u})}{U} = \mathbf{V}_U,$$

where we make use of the fact that by U1 observations are independent across u with $E(\mathbf{Z}'_{+u}\boldsymbol{\varepsilon}_u) = \mathbf{0}_{K+}$. Since $\mathbf{V}_U$ is positive definite for all U sufficiently large (assumption U3b), let $\mathbf{E}$ denote its eigenvectors and $\boldsymbol{\Lambda}$ its diagonal matrix of positive eigenvalues, and define $\mathbf{V}_U^{-1/2} = \mathbf{E}\boldsymbol{\Lambda}^{-1/2}\mathbf{E}'$, where $\boldsymbol{\Lambda}^{-1/2}$ is a diagonal matrix with entries equal to the inverse of the square root of the eigenvalues in $\boldsymbol{\Lambda}$. The largest eigenvalue of $\mathbf{V}_U^{-1} = \mathbf{V}_U^{-1/2}\mathbf{V}_U^{-1/2}$ is the inverse of the smallest eigenvalue of $\mathbf{V}_U$. Since the determinant of $\mathbf{V}_U$ is $> \gamma > 0$ for all U sufficiently large, and by Jensen's Inequality and (C1.2) its diagonal elements are bounded by $\Delta^{1/(1+\delta)}$, by the trace and determinant property of eigenvalues we know that the smallest eigenvalue of $\mathbf{V}_U$, $\lambda_{\min}$, is greater than $\gamma/(K_+\Delta^{1/(1+\delta)})\text{^}(K_+\text{-}1)$.

As noted in White (1980, p. 829 - see also White 1980a and Hoadley 1971), given (C1.6) a multivariate Liapounov Central Limit theorem implies $\mathbf{V}_U^{-1/2}\mathbf{Z}'_+\boldsymbol{\varepsilon}/\sqrt{U}$ is asymptotically distributed $N(\mathbf{0}_{K+},\mathbf{I}_{K+})$ provided that for all $\boldsymbol{\kappa}$ in $\mathbb{R}^{K+}$ and some $\delta > 0$

$$\text{(C1.7) } \sum_{u=1}^{U} E(|\,\boldsymbol{\kappa}'\mathbf{V}_U^{-1/2}\mathbf{Z}'_u\boldsymbol{\varepsilon}_u\,|^{2+\delta})/U^{(2+\delta)/2} \to 0.$$

Define $\boldsymbol{\varphi} = \boldsymbol{\kappa}'\mathbf{V}_U^{-1/2}$, and note that by the properties of the Rayleigh quotient $\boldsymbol{\varphi}'\boldsymbol{\varphi} \leq \boldsymbol{\kappa}'\boldsymbol{\kappa}/\lambda_{\min}$, i.e. for any given $\boldsymbol{\kappa}$ the elements φ_k of $\boldsymbol{\varphi}$ are bounded. By Minkowski's Inequality and (C1.2)

$$\text{(C1.8) } E(|\,\boldsymbol{\kappa}'\mathbf{V}_U^{-1/2}\mathbf{Z}'_u\boldsymbol{\varepsilon}_u\,|^{2+\delta}) = E(|\sum_{k=1}^{K+}\varphi_k\mathbf{z}'_{+uk}\boldsymbol{\varepsilon}_u\,|^{2+\delta})$$

$$\leq \left(\sum_{k=1}^{K+}(E\,|\,\varphi_k\mathbf{z}'_{+uk}\boldsymbol{\varepsilon}_u\,|^{2+\delta})^{\frac{1}{(2+\delta)}}\right)^{2+\delta} < \left(\sum_{k=1}^{K+}(|\,\varphi_k\,|^{2+\delta}\ \Delta)^{\frac{1}{(2+\delta)}}\right)^{2+\delta} < \infty,$$

so (C1.7) holds and using (C1.1), we can say that $\mathbf{V}_U^{-½}\mathbf{M}_U\sqrt{U}(\hat{\boldsymbol{\gamma}}_+ - \boldsymbol{\gamma}_+)$ is asymptotically distributed $N(\mathbf{0}_{K+},\mathbf{I}_{K+})$.

We now show that $U\hat{\mathbf{V}}(\hat{\boldsymbol{\gamma}}_+)$, for the three covariance estimates considered in U5, converges almost surely to the asymptotic covariance matrix $\mathbf{M}_U^{-1}\mathbf{V}_U\mathbf{M}_U^{-1}$ of $\sqrt{U}(\hat{\boldsymbol{\gamma}}_+ - \boldsymbol{\gamma}_+)$. Depending upon whether assumption U5a, U5b or U5c hold, we have

$$\text{(C1.9a) } \mathbf{V}_U = \sigma^2 \sum_{i=1}^{N} \frac{E(\mathbf{z}_{+i}\mathbf{z}'_{+i})}{U} = \sigma^2 \mathbf{M}_U \text{ (if U5a);}$$

$$\text{(C1.9b) } \mathbf{V}_U = \sum_{i=1}^{N} \frac{E(\mathbf{z}_{+i}\varepsilon_i^2\mathbf{z}'_{+i})}{U} \text{ (if U5b);}$$

$$\text{(C1.9c) } \mathbf{V}_U = \sum_{c=1}^{C} \frac{E(\mathbf{Z}'_{+c}\boldsymbol{\varepsilon}_c\boldsymbol{\varepsilon}'_c\mathbf{Z}_{+c})}{U} \text{ (if U5c).}$$

All of these expressions are positive definite for sufficiently large U (by U3b). They are also bounded, as examining the diagonal terms we see

$$\text{(C1.10a) } \sum_{i=1}^{N} \frac{E(z_{+ij}^2)}{U} = \sum_{u=1}^{U} \frac{E(\mathbf{z}'_{+uj}\mathbf{z}'_{+uj})}{U} \overbrace{< \Delta^{1/(1+\delta)}}^{\text{U2a \& Jensen's Inequality}};$$

$$\text{(C1.10b) } \sum_{i=1}^{N} \frac{E(z_{+ij}^2\varepsilon_i^2)}{U} \overbrace{\leq \sum_{u=1}^{U} \frac{E(\mathbf{z}'_{+uj}\mathbf{z}_{+uj}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u)}{U}}^{i\in u} \overbrace{< \Delta^{1/(1+\delta)}}^{\text{U3a \& Jensen's Inequality}};$$

$$\text{(C1.10c) } \sum_{c=1}^{C} \frac{E(\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c)}{U} \overbrace{\leq \sum_{c=1}^{C} \frac{E(\mathbf{z}'_{+cj}\mathbf{z}_{+cj}\boldsymbol{\varepsilon}'_c\boldsymbol{\varepsilon}_c)}{U}}^{\text{Cauchy-Schwarz Inequality}} \overbrace{\leq \sum_{u=1}^{U} \frac{E(\mathbf{z}'_{+uj}\mathbf{z}_{+uj}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u)}{U}}^{c\subseteq u} \overbrace{< \Delta^{1/(1+\delta)}}^{\text{U3a \& Jensen's Inequality}}.$$

For the homoskedastic covariance estimate, assuming U5a and using $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon} + \mathbf{Z}_+(\boldsymbol{\gamma}_+ - \hat{\boldsymbol{\gamma}}_+)$ we have

$$\text{(C1.11) } U\hat{\mathbf{V}}_h(\hat{\boldsymbol{\gamma}}_+) = \left(\frac{\mathbf{Z}'_+\mathbf{Z}_+}{U}\right)^{-1} \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{N} \frac{N}{N-K_+}, \text{ where}$$

$$\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{N} = \underbrace{\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{N}}_{\substack{a.s.\\ \to \sigma^2 \text{ by}\\ \text{Markov's Strong Law}}} + 2\underbrace{\frac{U}{N}}_{\leq 1}\underbrace{\frac{\boldsymbol{\varepsilon}'\mathbf{Z}_+}{U}}_{\substack{a.s.\\ \to \mathbf{0}'_{K+}\\ \text{by (C1.4)}}}\underbrace{(\boldsymbol{\gamma}_+ - \hat{\boldsymbol{\gamma}}_+)}_{\substack{a.s.\\ \to \mathbf{0}_{K+}\\ \text{by (C1.5)}}} + \underbrace{\frac{U}{N}}_{\leq 1}\underbrace{(\boldsymbol{\gamma}_+ - \hat{\boldsymbol{\gamma}}_+)'}_{\substack{a.s.\\ \to \mathbf{0}'_{K+}\\ \text{by (C1.5)}}}\underbrace{\frac{\mathbf{Z}'_+\mathbf{Z}_+}{U}}_{\substack{\text{a.s. bounded)}\\ \text{by (C1.1) \& (C1.10a)}}}\underbrace{(\boldsymbol{\gamma}_+ - \hat{\boldsymbol{\gamma}}_+)}_{\substack{a.s.\\ \to \mathbf{0}_{K+}\\ \text{by (C1.5)}}} \overset{a.s.}{\to} \sigma^2,$$

$$\text{so } U\hat{\mathbf{V}}_h(\hat{\boldsymbol{\gamma}}_+) - \mathbf{M}_U^{-1}\sigma^2 \overset{a.s.}{\to} \mathbf{0}_{K_+xK_+}, \text{ where } \mathbf{M}_U^{-1}\mathbf{V}_U\mathbf{M}_U^{-1} = \mathbf{M}_U^{-1}\sigma^2 \text{ (if U5a holds),}$$

and where we use the fact that (C1.1) and the Continuous Mapping Theorem Corollary imply that $(\mathbf{Z}'_+\mathbf{Z}_+/U)^{-1} - \mathbf{M}_U^{-1} \overset{a.s.}{\to} \mathbf{0}_{K+xK+}$. For the heteroskedasticity robust covariance estimate we have

$$\text{(C1.12) } U\hat{\mathbf{V}}_r(\hat{\boldsymbol{\gamma}}_+) = \left(\frac{\mathbf{Z}'_+\mathbf{Z}_+}{U}\right)^{-1} \mathbf{A} \left(\frac{\mathbf{Z}'_+\mathbf{Z}_+}{U}\right)^{-1}, \text{ where}$$

$$\mathbf{A}_{jk} = \sum_{i=1}^{N} \frac{z_{+ij}z_{+ik}\varepsilon_i^2}{U} + 2\sum_{l=1}^{K+}(\gamma_{+l} - \hat{\gamma}_{+l})\sum_{i=1}^{N} \frac{z_{+ij}z_{+ik}z_{+il}\varepsilon_i}{U} + \sum_{l=1}^{K+}\sum_{m=1}^{K+}(\gamma_{+l} - \hat{\gamma}_{+l})(\gamma_{+m} - \hat{\gamma}_{+m})\sum_{i=1}^{N} \frac{z_{+ij}z_{+ik}z_{+il}z_{+im}}{U}.$$

The last two terms of $\mathbf{A}_{jk}$ converge almost surely to 0 as $\boldsymbol{\gamma}_+ - \hat{\boldsymbol{\gamma}}_+$, which converges almost surely to zero, is multiplied by terms which are asymptotically almost surely bounded:

$$\text{(C1.13)}\left|\sum_{i=1}^{N}\frac{z_{+ij}z_{+ik}z_{+il}z_{+im}}{U}\right| \le \overbrace{\sqrt{\sum_{i=1}^{N}\frac{z_{+ij}^2 z_{+ik}^2}{U}\sum_{i=1}^{N}\frac{z_{+il}^2 z_{+im}^2}{U}}}^{\text{Cauchy–Schwarz Inequality}} \le \overbrace{\sqrt[4]{\prod_{n=j,k,l,m}\sum_{i=1}^{N}\frac{z_{+in}^4}{U}}}^{\text{Cauchy–Schwarz Inequality}} \le \overbrace{\sqrt[4]{\prod_{n=j,k,l,m}\sum_{u=1}^{U}\frac{(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}}}^{i\in u}$$

$$\text{where } \underbrace{\sum_{u=1}^{U}\frac{(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U} - \sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U} \xrightarrow{a.s.} 0 \;\;\&\;\; \sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U} < \Delta^{1/(1+\delta)}}_{\text{Markov Corollary, Jensen's Inequality \& U4}},$$

$$\left|\sum_{i=1}^{N}\frac{z_{+ij}z_{+ik}z_{+il}\varepsilon_i}{U}\right| \le \overbrace{\sqrt[4]{\left(\prod_{n=j,k}\sum_{i=1}^{N}\frac{z_{+in}^4}{U}\right)\left(\sum_{i=1}^{N}\frac{z_{+il}^2\varepsilon_i^2}{U}\right)^2}}^{\text{Cauchy–Schwarz Inequality}} \le \overbrace{\sqrt[4]{\left(\prod_{n=j,k}\sum_{u=1}^{U}\frac{(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}\right)\left(\sum_{u=1}^{U}\frac{\mathbf{z}'_{+ul}\mathbf{z}_{+ul}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u}{U}\right)^2}}^{i\in u}$$

$$\text{where } \underbrace{\sum_{u=1}^{U}\frac{(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U} - \sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U} \xrightarrow{a.s.} 0 \;\&\; \sum_{u=1}^{U}\frac{\mathbf{z}'_{+ul}\mathbf{z}_{+ul}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u}{U} - \sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+ul}\mathbf{z}_{+ul}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u)}{U} \xrightarrow{a.s.} 0}_{\text{Markov Corollary, U3 \& U4}}$$

$$\text{and } \underbrace{\sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U} < \Delta^{1/(1+\delta)} \;\&\; \sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+ul}\mathbf{z}_{+ul}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u)}{U} < \Delta^{1/(1+\delta)}}_{\text{Jensen's Inequality, U3 \& U4}}.$$

The expectation of $|\sum_{i\in u} z_{+ij}z_{+ik}\varepsilon_i^2|^{1+\delta}$ is bounded as

$$\text{(C1.14)}\; E(|\sum_{i\in u} z_{+ij}z_{+ik}\varepsilon_i^2|^{1+\delta}) \overbrace{\le E(|\prod_{n=j,k}\sum_{i\in u} z_{+in}^2\varepsilon_i^2|^{1/2(1+\delta)})}^{\text{Cauchy-Schwarz Inequality}} \overbrace{\le \sqrt{\prod_{n=j,k}E(|\sum_{i\in u} z_{+in}^2\varepsilon_i^2|^{1+\delta})}}^{\text{Hölder's Inequality}} \overbrace{\le \sqrt{\prod_{n=j,k}E(|\mathbf{z}'_{+u_i n}\mathbf{z}_{+u_i n}\boldsymbol{\varepsilon}'_{u_i}\boldsymbol{\varepsilon}_{u_i}|^{1+\delta})}}^{i\in u} \overbrace{< \Delta}^{\text{U3a}},$$

where u_i denotes the union grouping to which observation i belongs. Consequently, given U5b, by the Markov Corollary

$$\text{(C1.15)}\; \sum_{u=1}^{U}\frac{1}{U}\sum_{i\in u} z_{+ij}z_{+ik}\varepsilon_i^2 - \sum_{u=1}^{U}\frac{1}{U}E\left(\sum_{i\in u} z_{+ij}z_{+ik}\varepsilon_i^2\right) = \sum_{i=1}^{N}\frac{z_{+ij}z_{+ik}\varepsilon_i^2}{U} - \sum_{i=1}^{N}\frac{E(z_{+ij}z_{+ik}\varepsilon_i^2)}{U} \xrightarrow{a.s.} 0,$$

$$\mathbf{A}_{jk} - \sum_{i=1}^{N}\frac{E(z_{+ij}z_{+ik}\varepsilon_i^2)}{U} \xrightarrow{a,s,} 0 \text{ and } U\hat{\mathbf{V}}_r(\hat{\boldsymbol{\gamma}}_+) - \mathbf{M}_U^{-1}\mathbf{V}_U\mathbf{M}_U^{-1} \xrightarrow{a,s,} \mathbf{0}_{K_+xK_+} \text{ (if U5b holds).}$$

For the clustered robust covariance estimate we have

$$\text{(C1.16)}\; U\hat{\mathbf{V}}_{cl}(\hat{\boldsymbol{\gamma}}_+) = \left(\frac{\mathbf{Z}'_+\mathbf{Z}_+}{U}\right)^{-1}\mathbf{A}\left(\frac{\mathbf{Z}'_+\mathbf{Z}_+}{U}\right)^{-1}, \text{ where } \mathbf{A}_{jk} = \sum_{c=1}^{C}\frac{\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c}{U} +$$

$$\sum_{l=1}^{K+}(\gamma_{+l}-\hat{\gamma}_{+l})\left(\sum_{c=1}^{C}\frac{\mathbf{z}'_{+ck}\mathbf{z}_{+cl}\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c}{U} + \sum_{c=1}^{C}\frac{\mathbf{z}'_{+cj}\mathbf{z}_{+cl}\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c}{U}\right) + \sum_{l=1}^{K+}\sum_{m=1}^{K+}(\gamma_{+l}-\hat{\gamma}_{+l})(\gamma_{+m}-\hat{\gamma}_{+m})\sum_{c=1}^{C}\frac{\mathbf{z}'_{+cj}\mathbf{z}_{+cl}\mathbf{z}'_{+ck}\mathbf{z}_{+cm}}{U}.$$

Again, the last two terms of $\mathbf{A}_{jk}$ converge almost surely to 0 as $\boldsymbol{\gamma}_+ - \hat{\boldsymbol{\gamma}}_+$, which converges almost surely to zero, is multiplied by terms which are asymptotically almost surely bounded:

$$\text{(C1.17)}\ \left|\sum_{c=1}^{C}\frac{\mathbf{z}'_{+cj}\mathbf{z}_{+cl}\mathbf{z}'_{+ck}\mathbf{z}_{+cm}}{U}\right| \overbrace{\leq \sqrt{\sum_{c=1}^{C}\frac{(\mathbf{z}'_{+cj}\mathbf{z}_{+cl})^2}{U}\sum_{c=1}^{C}\frac{(\mathbf{z}'_{+ck}\mathbf{z}_{+cm})^2}{U}} \leq \sqrt{\sum_{c=1}^{C}\frac{(\mathbf{z}'_{+cj}\mathbf{z}_{+cj}\mathbf{z}'_{+cl}\mathbf{z}_{+cl})}{U}\sum_{c=1}^{C}\frac{(\mathbf{z}'_{+ck}\mathbf{z}_{+ck}\mathbf{z}'_{+cm}\mathbf{z}_{+cm})}{U}}}^{\text{Cauchy–Schwarz Inequality}}$$

$$\overbrace{\leq \sqrt[4]{\prod_{n=j,k,l,m}\sum_{c=1}^{C}\frac{(\mathbf{z}'_{+cn}\mathbf{z}_{+cn})^2}{U}}}^{\text{Cauchy–Schwarz Inequality}} \overbrace{\leq \sqrt[4]{\prod_{n=j,k,l,m}\sum_{u=1}^{U}\frac{(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}}}^{c\subseteq u}$$

$$\text{where } \underbrace{\sum_{u=1}^{U}\frac{(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}-\sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}\xrightarrow{a.s.}0 \ \ \&\ \ \sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}<\Delta^{1/(1+\delta)}}_{\text{Markov Corollary, Jensen's Inequality \& U4}}$$

$$\left|\sum_{c=1}^{C}\frac{\mathbf{z}'_{+cj}\mathbf{z}_{+cl}\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c}{U}\right| \overbrace{\leq \sqrt{\sum_{c=1}^{C}\frac{(\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c)^2}{U}\sum_{c=1}^{C}\frac{(\mathbf{z}'_{+cj}\mathbf{z}_{+cl})^2}{U}} \leq \sqrt[4]{\left(\sum_{c=1}^{C}\frac{\mathbf{z}'_{+ck}\mathbf{z}_{+ck}\boldsymbol{\varepsilon}'_c\boldsymbol{\varepsilon}_c}{U}\right)^2\prod_{n=j,l}\sum_{c=1}^{C}\frac{(\mathbf{z}'_{+cn}\mathbf{z}_{+cn})^2}{U}}}^{\text{Cauchy–Schwarz Inequality}}$$

$$\overbrace{\leq \sqrt[4]{\left(\sum_{u=1}^{U}\frac{\mathbf{z}'_{+uk}\mathbf{z}_{+uk}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u}{U}\right)^2\prod_{n=j,l}\sum_{u=1}^{U}\frac{(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}}}^{c\subseteq u}$$

$$\text{where } \underbrace{\sum_{u=1}^{U}\frac{(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}-\sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}\xrightarrow{a.s.}0 \ \&\ \sum_{u=1}^{U}\frac{\mathbf{z}'_{+uk}\mathbf{z}_{+uk}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u}{U}-\sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+uk}\mathbf{z}_{+uk}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u)}{U}\xrightarrow{a.s.}0}_{\text{Markov Corollary, U3 \& U4}}$$

$$\text{and } \underbrace{\sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+un}\mathbf{z}_{+un})^2}{U}<\Delta^{1/(1+\delta)} \ \&\ \sum_{u=1}^{U}\frac{E(\mathbf{z}'_{+uk}\mathbf{z}_{+uk}\boldsymbol{\varepsilon}'_u\boldsymbol{\varepsilon}_u)}{U}<\Delta^{1/(1+\delta)}}_{\text{Jensen's Inequality, U3 \& U4}}.$$

The expectation of $|\sum_{c\subseteq u}\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c|^{1+\delta}$ is also bounded as

$$\text{(C1.18)}\ E(|\sum_{c\subseteq u}\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c|^{1+\delta}) \overbrace{\leq E(|\prod_{n=j,k}\sum_{c\subseteq u}(\mathbf{z}'_{+cn}\boldsymbol{\varepsilon}_c)^2|^{½(1+\delta)})}^{\text{Cauchy-Schwarz Inequality}}$$

$$\overbrace{\leq \sqrt{\prod_{n=j,k}E(|\sum_{c\subseteq u}(\mathbf{z}'_{+cn}\boldsymbol{\varepsilon}_c)^2|^{1+\delta})}}^{\text{Hölder's Inequality}} \overbrace{\leq \sqrt{\prod_{n=j,k}E(|\sum_{c\subseteq u}\mathbf{z}'_{+cn}\mathbf{z}_{+cn}\boldsymbol{\varepsilon}'_c\boldsymbol{\varepsilon}_c|^{1+\delta})}}^{\text{Cauchy-Schwarz Inequality}} \overbrace{\leq \sqrt{\prod_{n=j,k}E(|\mathbf{z}'_{+u_cn}\mathbf{z}_{+u_cn}\boldsymbol{\varepsilon}'_{u_c}\boldsymbol{\varepsilon}_{u_c}|^{1+\delta})}}^{c\subseteq u} \overbrace{<\Delta}^{\text{U3a}},$$

where u_c denotes the union grouping to which cluster group c belongs. Consequently, given U5c, by the Markov Corollary

$$\text{(C1.19)}\ \sum_{u=1}^{U}\frac{1}{U}\sum_{c\subseteq u}\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c-\sum_{u=1}^{U}\frac{1}{U}E\left(\sum_{c\subseteq u}\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c\right)=\sum_{c=1}^{C}\frac{\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c}{U}-\sum_{c=1}^{C}\frac{E(\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c)}{U}\xrightarrow{a.s.}0,$$

$$\mathbf{A}_{jk}-\sum_{c=1}^{C}\frac{E(\mathbf{z}'_{+cj}\boldsymbol{\varepsilon}_c\mathbf{z}'_{+ck}\boldsymbol{\varepsilon}_c)}{U}\xrightarrow{a,s,}0 \text{ and } U\hat{\mathbf{V}}_{cl}(\hat{\boldsymbol{\gamma}}_+)-\mathbf{M}_U^{-1}\mathbf{V}_U\mathbf{M}_U^{-1}\xrightarrow{a,s,}\mathbf{0}_{K_+xK_+} \text{ (if U5c holds)},$$

which completes the proof of the lemma.

Lemma B2a: As U/M is bounded from above by 1, (C1.1) above allows us to state that

$$\text{(C2.1)}\ \frac{\mathbf{Z}'_+\mathbf{Z}_+}{M}-\frac{U}{M}\mathbf{M}_U\xrightarrow{a.s.}\mathbf{0}_{K_+xK_+}.$$

As $\mathbf{M}_U$ is almost surely positive definite for all U (equivalently, by A4, M) sufficiently large with determinant greater than some $\gamma > 0$ (U2b), and U/M is bounded from below by $\overline{N}^{-1}$ (assumption A4), by the Continuous Mapping Theorem Corollary $\mathbf{Z}_+'\mathbf{Z}_+/M$ is almost surely positive definite for all M sufficiently large with determinant greater than $\overline{N}^{-1}\gamma > 0$. Since the trace of $\mathbf{M}_U$ is bounded from above by $\Delta^{1/(1+\delta)}K_+$ (C1.10a above), by the trace and determinant property of eigenvalues we know that for sufficiently large N its smallest eigenvalue is almost surely greater than $\lambda = \gamma/(K_+\Delta^{1/(1+\delta)})\text{^}(K_+\text{-}1) > 0$. It follows that almost surely for all M sufficiently large the smallest eigenvalue of $\mathbf{Z}_+'\mathbf{Z}_+/M$ is greater than $\overline{N}^{-1}\lambda$. Since $\mathbf{Z}$ and $\mathbf{W}$ are part of $\mathbf{Z}_+$ (assumption A2), it follows from the properties of the Rayleigh quotient that the minimum eigenvalues of the sub-matrices $\mathbf{Z}'\mathbf{Z}/M$ and $\mathbf{W}'\mathbf{W}/M$ are greater than or equal to that of $\mathbf{Z}_+'\mathbf{Z}_+/M$. Consequently, for sufficiently large M both matrices are almost surely positive definite with determinants > $\overline{N}^{-K}\lambda^K$ and $\overline{N}^{-Q}\lambda^Q$, respectively. By Jensen's Inequality the assumption $E(|\boldsymbol{x}_{ip}^4|^{1+\theta^*}) < \Delta$ in A3a implies that $E(|\boldsymbol{x}_{ip}^n|^{1+\delta}) < \Delta^{n/4}$ for n = 1...3, so by the LLNHDS Corollary above

$$\text{(C2.2)}\quad \frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M} - \mathbf{G}_M = \left(\frac{\boldsymbol{\mathcal{X}}'\boldsymbol{\mathcal{X}}}{M} - \frac{\boldsymbol{\mathcal{X}}'\mathbf{1}_M}{M}\frac{\mathbf{1}_M'\boldsymbol{\mathcal{X}}}{M}\right) - \left(\sum_{m=1}^{M}\frac{E(\boldsymbol{x}_m\boldsymbol{x}_m')}{M} - \sum_{i=1}^{M}\frac{E(\boldsymbol{x}_m)}{M}\sum_{i=1}^{M}\frac{E(\boldsymbol{x}_m')}{M}\right) \xrightarrow{a.s.} \mathbf{0}_{PxP},$$

and from A1a for all M sufficiently large the determinant of $\mathbf{G}_M$ is almost surely greater than $\gamma > 0$, so using the Continuous Mapping Theorem Corollary the same can be said of $\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}/M$.

By Hölder's Inequality and A3a, $E(|\mathbf{w}_{mq}'\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m'\mathbf{w}_{mr}|^{1+\theta}) \le E(|\mathbf{w}_{mq}'\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m'\mathbf{w}_{mq}|^{1+\theta})^{½}E(|\mathbf{w}_{mr}'\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m'\mathbf{w}_{mr}|^{1+\theta})^{½} < \Delta$ for all q & r = 1...Q. Consequently, using the LLNHDS Corollary once again

$$\text{(C2.3)}\quad \frac{\boldsymbol{\mathcal{W}}_\varepsilon'\boldsymbol{\mathcal{W}}_\varepsilon}{M} - \mathbf{W}_M = \sum_{m=1}^{M}\frac{\mathbf{W}_m'\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m'\mathbf{W}_m}{M} - \sum_{m=1}^{M}\frac{E(\mathbf{W}_m'\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m'\mathbf{W}_m)}{M} \xrightarrow{a.s.} \mathbf{0}_{QxQ},$$

where by A3b $\mathbf{W}_M$ is positive definite for all M sufficiently large with determinant > $\gamma > 0$, and so by the Continuous Mapping Theorem Corollary the same is true of $\boldsymbol{\mathcal{W}}_\varepsilon'\boldsymbol{\mathcal{W}}_\varepsilon/M$. As U/M is bounded, from (C1.4) we know that:

$$\text{(C2.4)}\quad \frac{\mathbf{Z}_+'\boldsymbol{\varepsilon}}{M} = \frac{U}{M}\sum_{u=1}^{U}\frac{\mathbf{Z}_{+u}'\boldsymbol{\varepsilon}_u}{U} \xrightarrow{a.s.} \mathbf{0}_{K+}.$$

Since $\mathbf{W}$ is included in $\mathbf{Z}_+$ and $\mathbf{1}_M'\boldsymbol{\mathcal{W}}_\varepsilon = \boldsymbol{\varepsilon}'\mathbf{W}$, the implication, as indicated in the lemma, is that

$$\text{(C2.5)}\quad \frac{\tilde{\boldsymbol{\mathcal{W}}}_\varepsilon'\tilde{\boldsymbol{\mathcal{W}}}_\varepsilon}{M} - \frac{\boldsymbol{\mathcal{W}}_\varepsilon'\boldsymbol{\mathcal{W}}_\varepsilon}{M} = \frac{\boldsymbol{\mathcal{W}}_\varepsilon'\boldsymbol{\mathcal{W}}_\varepsilon}{M} - \frac{\mathbf{W}'\boldsymbol{\varepsilon}}{M}\frac{\boldsymbol{\varepsilon}'\mathbf{W}}{M} - \frac{\boldsymbol{\mathcal{W}}_\varepsilon'\boldsymbol{\mathcal{W}}_\varepsilon}{M} \xrightarrow{a.s.} \mathbf{0}_{QxQ}.$$

Lemma B2b: As $\mathbf{Z}$ and $\mathbf{X}_\mathbf{W}$ are part of $\mathbf{Z}_+$, (C1.4) and (C2.4) above establish this lemma.

Lemma B2c: With regards to the matrix inverses in the Lemma, for an invertible positive definite matrix $\mathbf{A}$ the largest eigenvalue of $\mathbf{A}^{-1}$ is equal to the inverse of the smallest eigenvalue of $\mathbf{A}$. From the proof of B2a we know that each of the matrices in Lemma B2c is invertible with a smallest eigenvalue almost surely greater than some $\lambda > 0$ for all N sufficiently large, so it follows that the elements of their inverses are all almost surely bounded. For the means of products of columns of $\boldsymbol{\mathcal{X}}$, by Jensen's Inequality the assumption $E(|\boldsymbol{x}_{mp}^4|^{1+\theta^*}) < \Delta$ in A3a implies that $E(|\boldsymbol{x}_{mp}^n|^{1+\theta^*}) < \Delta^{n/4}$ and $E(|\boldsymbol{x}_{mp}^n|) < \Delta^{n/4(1+\theta^*)}$ for n = 1, 2, 3 or 4, and by the LLNHDS Corollary $\omega(\boldsymbol{x}_{mp}^n) - \omega(E(\boldsymbol{x}_{mp}^n)) \xrightarrow{a.s.} 0$, which tells us the means of powers up

to 4 of $\boldsymbol{\varkappa}$ are bounded. The sample mean of products of 2, 3 or 4 different columns of $\boldsymbol{\mathcal{X}}$ can then be bounded by repeated application of the Cauchy-Schwarz Inequality:

$$\text{(C2.6)}\ \left|\omega(\boldsymbol{\varkappa}_{mp}\boldsymbol{\varkappa}_{mq})\right| = \left|\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mp}\boldsymbol{\varkappa}_{mq}}{M}\right| \le \sqrt{\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mp}^2}{M}\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mq}^2}{M}},\ \left|\omega(\boldsymbol{\varkappa}_{mp}\boldsymbol{\varkappa}_{mq}\boldsymbol{\varkappa}_{mr})\right| \le \sqrt[4]{\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mp}^4}{M}\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mq}^4}{M}}\sqrt{\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mr}^2}{M}}$$

$$\&\ \left|\omega(\boldsymbol{\varkappa}_{mp}\boldsymbol{\varkappa}_{mq}\boldsymbol{\varkappa}_{mr}\boldsymbol{\varkappa}_{ms})\right| \le \sqrt[4]{\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mp}^4}{M}\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mq}^4}{M}\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{mr}^4}{M}\sum_{m=1}^{M}\frac{\boldsymbol{\varkappa}_{ms}^4}{M}}\ .$$

Turning to $|m(d_{i1}d_{i2})|$, where $d_{i1}d_{i2}$ is the product of the elements of two columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$, let $\mathbf{d}_{uj}$ denote the group u observations of the j^{th} column of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ and d_{ij} the i^{th} observation of that column. Then

$$\text{(C2.7)}\ |m(\prod_{j=1}^{2}d_{ij})| \le \underbrace{\sqrt{\prod_{j=k,l}\sum_{i=1}^{N}\frac{d_{ij}^2}{M}}}_{\text{C-Schwarz Inequality}} \le \underbrace{\sqrt{\prod_{j=k,l}\sum_{u=1}^{U}\frac{\mathbf{d}_{uj}'\mathbf{d}_{uj}}{U}}}_{M\ge U}$$

$$\&\ \underbrace{\sum_{u=1}^{U}\frac{\mathbf{d}_{uj}'\mathbf{d}_{uj}}{U}-\sum_{u=1}^{U}\frac{E(\mathbf{d}_{uj}'\mathbf{d}_{uj})}{U}\xrightarrow{a.s.}0}_{\text{Markov Corollary \& U2a}},\ \text{while}\ \underbrace{E(\mathbf{d}_{uj}'\mathbf{d}_{uj})<\Delta^{1/(1+\delta)}}_{\text{Jensen's Inequality \& U2a}},$$

so the mean of the product of 2 elements from the columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ is also almost surely bounded. Next, we note that U3a and U4 imply that for any $\mathbf{d}_{uj}$ and $\mathbf{d}_{uk}$ denoting u group column elements of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$, with no more than one of these referring to $\boldsymbol{\varepsilon}$, there exist positive finite constants δ and Δ such that for all u

$$\text{(C2.8)}\ \ E(|\mathbf{d}_{uj}'\mathbf{d}_{uj}\mathbf{d}_{uk}'\mathbf{d}_{uk}|^{1+\delta})<\Delta.$$

When one of the $\mathbf{d}_{uj}$ denotes elements of $\boldsymbol{\varepsilon}$, (C2.8) is merely a restatement of U3a. When both $\mathbf{d}_{uj}$ denote elements of $\mathbf{Z}_+$, we use Hölder's Inequality & U4 to show that :

$$\text{(C2.9)}\ \ E(|\mathbf{z}_{+uj}'\mathbf{z}_{+uj}\mathbf{z}_{+uk}'\mathbf{z}_{+uk}|^{1+\delta}) \overbrace{\le \sqrt{E(|(\mathbf{z}_{+uj}'\mathbf{z}_{+uj})^2|^{1+\delta})E(|(\mathbf{z}_{+uk}'\mathbf{z}_{+uk})^2|^{1+\delta})}}^{\text{Hölder's Inequality}} \overbrace{<\Delta}^{\text{U4}}.$$

We then have

$$\text{(C2.10)}\ |m(d_{i1}d_{i2}d_{i3}d_{i4})| \overbrace{\le \sqrt{\sum_{u=1}^{U}\sum_{i\in u}\frac{d_{i1}^2d_{i2}^2}{M}\sum_{u=1}^{U}\sum_{i\in u}\frac{d_{i3}^2d_{i4}^2}{M}}}^{\text{Cauchy-Schwarz Inequality}} \overbrace{\le \sqrt{\sum_{u=1}^{U}\frac{\mathbf{d}_{u1}'\mathbf{d}_{u1}\mathbf{d}_{u2}'\mathbf{d}_{u2}}{U}\sum_{u=1}^{U}\frac{\mathbf{d}_{u3}'\mathbf{d}_{u3}\mathbf{d}_{u4}'\mathbf{d}_{u4}}{U}}}^{i\in u,M\ge U}$$

$$\&\ \underbrace{\sum_{u=1}^{U}\frac{\mathbf{d}_{uj}'\mathbf{d}_{uj}\mathbf{d}_{uk}'\mathbf{d}_{uk}}{U}-\sum_{u=1}^{U}\frac{E(\mathbf{d}_{uj}'\mathbf{d}_{uj}\mathbf{d}_{uk}'\mathbf{d}_{uk})}{U}\xrightarrow{a.s.}0}_{\text{Markov Corollary \& (C2.8)}},\ \text{while}\ \underbrace{E(\mathbf{d}_{uj}'\mathbf{d}_{uj}\mathbf{d}_{uk}'\mathbf{d}_{uk})<\Delta^{1/(1+\delta)}}_{\text{Jensen's Inequality \& (C2.8)}}.$$

Similarly, for m_c, m_m and m_v, let the letter g (G) denote either c, m or v (C, M or V), so

$$\text{(C2.11)}\quad |m_g(d_{i1}d_{i2},d_{i3}d_{i4})| = \left|\sum_{u=1}^{U}\frac{1}{M}\sum_{g\subseteq u}\mathbf{d}'_{g1}\mathbf{d}_{g2}\mathbf{d}'_{g3}\mathbf{d}_{g4}\right| \overset{\text{Cauchy-Schwarz Inequality}}{\leq \sqrt{\sum_{u=1}^{U}\frac{1}{M}\sum_{g\subseteq u}(\mathbf{d}'_{g1}\mathbf{d}_{g2})^2\sum_{u=1}^{U}\frac{1}{M}\sum_{g\subseteq u}(\mathbf{d}'_{g3}\mathbf{d}_{g4})^2}}$$

$$\overset{\text{Cauchy-Schwarz Inequality}}{\leq \sqrt{\sum_{u=1}^{U}\frac{1}{M}\sum_{g\subseteq u}\mathbf{d}'_{g1}\mathbf{d}_{g1}\mathbf{d}'_{g2}\mathbf{d}_{g2}\sum_{u=1}^{U}\frac{1}{M}\sum_{g\subseteq u}\mathbf{d}'_{g3}\mathbf{d}_{g3}\mathbf{d}'_{g4}\mathbf{d}_{g4}}} \overset{g\subseteq u,\,M\geq U}{\leq \sqrt{\sum_{u=1}^{U}\frac{1}{U}\mathbf{d}'_{u1}\mathbf{d}_{u1}\mathbf{d}'_{u2}\mathbf{d}_{u2}\sum_{u=1}^{U}\frac{1}{U}\mathbf{d}'_{u3}\mathbf{d}_{u3}\mathbf{d}'_{u4}\mathbf{d}_{u4}}},$$

$$\text{while } \underbrace{\sum_{u=1}^{U}\frac{1}{U}\mathbf{d}'_{uj}\mathbf{d}_{uj}\mathbf{d}'_{uk}\mathbf{d}_{uk} - \sum_{u=1}^{U}\frac{1}{U}E(\mathbf{d}'_{uj}\mathbf{d}_{uj}\mathbf{d}'_{uk}\mathbf{d}_{uk}) \overset{a.s.}{\to} 0}_{\text{Markov Corollary \& (C2.8)}} \;\&\; \underbrace{E(\mathbf{d}'_{uj}\mathbf{d}_{uj}\mathbf{d}'_{uk}\mathbf{d}_{uk}) < \Delta^{1/(1+\delta)}}_{\text{Jensen's Inequality \& (C2.8)}}.$$

Together (C2.10) & (2.11) establish that $m(d_{i1}d_{i2}d_{i3}d_{i4})$, $m_c(d_{i1}d_{i2},d_{j3}d_{j4})$, $m_m(d_{i1}d_{i2},d_{j3}d_{j4})$ & $m_v(d_{i1}d_{i2},d_{j3}d_{j4})$ are all almost surely bounded, which completes the proof of the lemma.

Lemma B2d: When looking at the product of $n > 4$ column elements of $\boldsymbol{\mathcal{X}}$, we have

$$\text{(C2.12)}\quad \frac{|\omega(\prod_{p=1}^{n}\boldsymbol{\mathcal{x}}_{mp})|}{M^{a(\frac{n}{2}-2)}} \leq \frac{\sum_{m=1}^{M}\frac{|\prod_{p=1}^{n}\boldsymbol{\mathcal{x}}_{mp}|}{M}}{M^{a(\frac{n}{2}-2)}} \leq \frac{\sqrt[n]{\prod_{p=1}^{n}\sum_{m=1}^{M}\frac{|\boldsymbol{\mathcal{x}}_{mp}^{n}|}{M}}}{M^{a(\frac{n}{2}-2)}} \leq \sqrt[n]{\prod_{p=1}^{n}\left[\underset{m\leq M}{\text{Max}}\frac{\boldsymbol{\mathcal{x}}_{mp}^{2}}{M^{a}}\right]^{\frac{n}{2}-2}\sum_{m=1}^{M}\frac{\boldsymbol{\mathcal{x}}_{mp}^{4}}{M}},$$

where we make use of Hölder's Inequality. As for the m_c, m_m and m_v mentioned in the lemma, with g (G) denoting c, m or v (C, M, or V), we can say

$$\text{(C2.13)}\quad \frac{m_g(d_{i1}^2d_{i2}^2,d_{j3}^2d_{j4}^2)}{M^{2a}} = \sum_{g=1}^{G}\sum_{i\in g}\sum_{j\in g}\frac{d_{i1}^2d_{i2}^2d_{j3}^2d_{j4}^2}{M^{1+2a}} \leq \underset{i\leq N}{\text{Max}}\frac{d_{i1}^2d_{i2}^2}{M^{2a}}\sum_{g=1}^{G}\sum_{i\in g}\sum_{j\in g}\frac{d_{j3}^2d_{j4}^2}{M}$$

$$\leq \underset{i\leq N}{\text{Max}}\frac{d_{i1}^2d_{i2}^2}{M^{2a}}\overline{N}\sum_{g=1}^{G}\sum_{j\in g}\frac{d_{j3}^2d_{j4}^2}{M} \leq \underset{i\leq N}{\text{Max}}\frac{d_{i1}^2d_{i2}^2}{M^{2a}}\frac{\overline{N}U}{M}\sum_{u=1}^{U}\frac{\mathbf{d}'_{u3}\mathbf{d}_{u3}\mathbf{d}'_{u4}\mathbf{d}_{u4}}{U},$$

where $\overline{N}$ from A4 is the maximum union grouping size. Since Lemma B2c showed that $\omega(\prod_{k=1}^{4}\boldsymbol{\mathcal{x}}_{mk})$ is bounded, while the average of $\mathbf{d}'_{u3}\mathbf{d}_{u3}\mathbf{d}'_{u4}\mathbf{d}_{u4}$ is as seen in (C2.11) similarly bounded, we need only show

$$\text{(C2.14)}\ \exists a < ½ \text{ such that: } \underset{m\leq M}{\text{Max}}\ \boldsymbol{\mathcal{x}}_{mp}^{2}/M^{a} \overset{a.s.}{\to} 0\ \forall\ p \text{ and } \underset{i\leq N}{\text{Max}}\ d_{ij}^2d_{ik}^2/M^{2a} \overset{a.s.}{\to} 0\ \forall\ j,k.$$

From A3 and (C2.8) above, we know that for all m, u, j, k & p there exist finite positive constants δ and Δ such that $E(|\boldsymbol{\mathcal{x}}_{mp}^{4}|^{1+\delta}) < \Delta$ and $E(|\mathbf{d}'_{uj}\mathbf{d}_{uj}\mathbf{d}'_{uk}\mathbf{d}_{uk}|^{1+\delta}) < \Delta$. Applying Markov's Inequality

$$\text{(C2.15)}\quad \sum_{M=1}^{\infty}P(x_{Mp}^2 \geq M^{\gamma}) = \sum_{M=1}^{\infty}P(x_{Mp}^4 \geq M^{2\gamma}) \leq \sum_{M=1}^{\infty}\frac{\Delta}{M^{2\gamma(1+\delta)}} < \infty \text{ if } 2\gamma(1+\delta) > 1$$

$$\sum_{U=1}^{\infty}P(\mathbf{d}'_{Uj}\mathbf{d}_{Uj}\mathbf{d}'_{Uk}\mathbf{d}_{Uk} \geq U^{2\gamma}) \leq \sum_{U=1}^{\infty}\frac{\Delta}{U^{2\gamma(1+\delta)}} < \infty \text{ if } 2\gamma(1+\delta) > 1.$$

So, following the Borel-Cantelli Corollary, both $\text{Max}_{m\leq M}\boldsymbol{\mathcal{x}}_{mp}^{2}/M^{\gamma}$ and $\text{Max}_{i\leq N}\ d_{ij}^2d_{ik}^2/M^{2\gamma}$ $\leq \text{Max}_{u\leq U}\mathbf{d}'_{uj}\mathbf{d}_{uj}\mathbf{d}'_{uk}\mathbf{d}_{uk}/U^{2\gamma}$ are almost surely bounded for $\gamma > ½(1+\delta)^{-1}$, and consequently for all a in $(\gamma,½)$ (C2.14) holds, proving the lemma.

Lemma B3a: For $n = 1$ or 2 we need to show that the following is bounded:

$$\text{(C3.1)}\ \sum_{m=1}^{M}\frac{[\prod_{k=1}^{n}x_{mk}-\omega(\prod_{k=1}^{n}x_{mk})]^2}{M}\sum_{m=1}^{M}\frac{[\sum_{i\in m}d_{i1}d_{i2}-\sum_{m=1}^{M}\sum_{i\in m}d_{i1}d_{i2}/M]^2}{M}$$

$$=[\omega(\prod_{k=1}^{n}x_{mk}^2)-\omega(\prod_{k=1}^{n}x_{mk})^2][m_m(d_{i1}d_{i2},d_{j1}d_{j2})-m(d_{i1}d_{i2})^2]$$

All of these were shown to be bounded in the proof of Lemma B2c.

Lemma B3b: Lemmas B2a and B2c already established that $\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}}/M$ & $\tilde{\boldsymbol{w}}_{\varepsilon}'\tilde{\boldsymbol{w}}_{\varepsilon}/M$ are almost surely bounded with determinant > $\gamma > 0$ for all M sufficiently large, so all that remains is condition Ib. Our objective is to prove that for all integer $\tau > 2$ and all p and q

$$\text{(C3.2)}\ M^{\frac{\tau}{2}-1}\sum_{m=1}^{M}\tilde{x}_{mp}^{\tau}\sum_{m=1}^{M}\tilde{w}_{mq\varepsilon}^{\tau}\Big/\left(\sum_{m=1}^{M}\tilde{x}_{mp}^2\right)^{\tau/2}\left(\sum_{m=1}^{M}\tilde{w}_{mq\varepsilon}^2\right)^{\tau/2}\xrightarrow{a.s.}0,$$

where $w_{mq\varepsilon}=\sum_{i\in m}w_{iq}\varepsilon_i$. We begin by noting that:

$$\text{(C3.3)}\ \left|\frac{M^{\frac{\tau}{2}-1}\sum_{m=1}^{M}\tilde{x}_{mp}^{\tau}\sum_{m=1}^{M}\tilde{w}_{mq\varepsilon}^{\tau}}{\left(\sum_{m=1}^{M}\tilde{x}_{mp}^2\sum_{m=1}^{M}\tilde{w}_{mq\varepsilon}^2\right)^{\tau/2}}\right|\le\left|\frac{M^{\frac{\tau}{2}-1}\left(\underset{m\le M}{\text{Max}}\,\tilde{x}_{mp}^2\,\underset{m\le M}{\text{Max}}\,\tilde{w}_{mq\varepsilon}^2\right)^{\frac{\tau}{2}-1}\sum_{m=1}^{M}\tilde{x}_{mp}^2\sum_{m=1}^{M}\tilde{w}_{mq\varepsilon}^2}{\left(\sum_{m=1}^{M}\tilde{x}_{mp}^2\sum_{m=1}^{M}\tilde{w}_{mq\varepsilon}^2\right)^{\tau/2}}\right|=\left|\frac{\frac{\underset{m\le M}{\text{Max}}\,\tilde{x}_{mp}^2\,\underset{m\le M}{\text{Max}}\,\tilde{w}_{mq\varepsilon}^2}{M}}{\sum_{m=1}^{M}\frac{\tilde{x}_{mp}^2}{M}\sum_{m=1}^{M}\frac{\tilde{w}_{mq\varepsilon}^2}{M}}\right|^{\frac{\tau}{2}-1}.$$

With regards to the denominator in the last expression, by the Schur-Horn Theorem the smallest diagonal element of a symmetric real matrix is greater than or equal to its smallest eigenvalue. Lemmas B2a and B2c established that $\tilde{\boldsymbol{x}}'\tilde{\boldsymbol{x}}/M$ & $\tilde{\boldsymbol{w}}_{\varepsilon}'\tilde{\boldsymbol{w}}_{\varepsilon}/M$ are almost surely bounded with determinant > $\gamma > 0$ for all M sufficiently large. For a K x K matrix with determinant > $\gamma > 0$ and non-negative diagonal elements bounded from above by Δ', by the trace and determinant property of eigenvalues the smallest eigenvalue is bounded from below by $\lambda(K)= \gamma/(K\Delta')\text{^}(K\text{-}1)$. Consequently, the denominator in the last expression is almost surely > $\lambda(P)\lambda(Q) > 0$ for all N sufficiently large.

Turning to the numerator, since for $d_m^* = x_{mp}$ or $w_{mq\varepsilon}$

$$\text{(C3.4)}\ \underset{m\le M}{\text{Max}}\,\tilde{d}_m^{*2}\le\underset{m\le M}{\text{Max}}\,d_m^{*2}+2\sqrt{\underset{m\le M}{\text{Max}}(d_m^{*2})}\left|\omega(d_m^*)\right|+\omega(d_m^*)^2$$

and Lemma B2c showed that both $\omega(x_{mp})$ & $\omega(w_{mq\varepsilon}) = m(w_{iq}\varepsilon_i)$ converge almost surely to the bounded mean values of $E(x_{mp})$ & $E(w_{mq\varepsilon})$, respectively, to prove (Ib) all that remains is to show that $\text{Max}_{m\le M}x_{mp}^2\text{Max}_{m\le M}w_{mq\varepsilon}^2/M\xrightarrow{a.s.}0$. From assumption A3a we know that there exist finite positive constants θ, θ^* and Δ, with $\theta(1+2\theta^*) > 1$, such that $E(|x_{mp}^4|^{1+\theta^*})<\Delta$ and $E(|w_{mq\varepsilon}^2|^{1+\theta^*})<\Delta$. Consequently, applying Markov's Inequality

$$\text{(C3.5)}\ \sum_{M=1}^{\infty}P(x_{Mp}^2\ge M^a)=\sum_{M=1}^{\infty}P(x_{Mp}^4\ge M^{2a})\le\sum_{M=1}^{\infty}\frac{\Delta}{M^{2a(1+\theta^*)}}<\infty\ \text{ if }\ 2a(1+\theta^*)>1$$

$$\&\ \sum_{M=1}^{\infty}P(w_{Mq\varepsilon}^2\ge M^b)\le\sum_{M=1}^{\infty}\frac{\Delta}{M^{b(1+\theta)}}<\infty\ \text{ if }\ b(1+\theta)>1.$$

Both conditions can be met with $a > 0$, $b > 0$ and $a + b < 1$ if $\theta(1+2\theta^*) > 1$ as

$$\text{(C3.6) } 1 > a+b > \frac{1}{2(1+\theta^*)} + \frac{1}{1+\theta} = 1 - \frac{\theta(1+2\theta^*)-1}{2(1+\theta^*)(1+\theta)}$$

poses no contradiction. We then see that $\text{Max}_{m\le M}\boldsymbol{\chi}_{mp}^2 \text{Max}_{m\le M}\boldsymbol{\omega}_{mq\varepsilon}^2 / M^{a+b}$ is by the Borel-Cantelli Lemma Corollary almost surely bounded by 1, so $\text{Max}_{m\le M}\boldsymbol{\chi}_{mp}^2 \text{Max}_{m\le M}\boldsymbol{\omega}_{mq\varepsilon}^2 / M \xrightarrow{a.s.} 0$.

Lemma B4: From the proof of Lemma B2a we know that $\boldsymbol{\omega}_\varepsilon'\boldsymbol{\omega}_\varepsilon / M - \mathbf{W}_M \xrightarrow{a.s.} \mathbf{0}_{QxQ}$ & $\tilde{\boldsymbol{\omega}}_\varepsilon'\tilde{\boldsymbol{\omega}}_\varepsilon / M - \boldsymbol{\omega}_\varepsilon'\boldsymbol{\omega}_\varepsilon / M \xrightarrow{a.s.} \mathbf{0}_{QxQ}$, so $\tilde{\boldsymbol{\omega}}_\varepsilon'\tilde{\boldsymbol{\omega}}_\varepsilon / M - \mathbf{W}_M \xrightarrow{a.s.} \mathbf{0}_{QxQ}$. When using the homoskedastic covariance estimate, from U5a we have

$$\text{(C4.1) } \mathbf{W}_M = \sum_{m=1}^{M} \frac{E(\mathbf{W}_m' \boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}_m' \mathbf{W}_m)}{M} = \sum_{m=1}^{M}\sum_{i\in m}\sum_{j\in m} \frac{E(\mathbf{w}_i \varepsilon_i \mathbf{w}_j' \varepsilon_j)}{M} = \sum_{m=1}^{M}\sum_{i\in m} \frac{\sigma^2 E(\mathbf{w}_i \mathbf{w}_i')}{M} = \sigma^2 \frac{U}{M}\sum_{u=1}^{U} \frac{E(\mathbf{W}_u'\mathbf{W}_u)}{U}.$$

where we recall that we use the notation $\mathbf{w}_i'$ to denote the i^{th} row of $\mathbf{W}$. As $\mathbf{W}$ is part of $\mathbf{Z}_+$ (assumption A2), from U2a and the Markov Corollary we have:

$$\text{(C4.2) } \frac{\mathbf{W}'\mathbf{W}}{U} - \sum_{u=1}^{U} \frac{E(\mathbf{W}_u'\mathbf{W}_u)}{U} = \sum_{u=1}^{U} \frac{\mathbf{W}_u'\mathbf{W}_u}{U} - \sum_{u=1}^{U} \frac{E(\mathbf{W}_u'\mathbf{W}_u)}{U} \xrightarrow{a.s.} \mathbf{0}_{QxQ},$$

while, as ε_i is iid, by the Strong Law of Large Numbers,

$$\text{(C4.3) } \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{N} = \sum_{i=1}^{N} \frac{\varepsilon_i^2}{N} \xrightarrow{a.s.} \sigma^2,$$

so, as U/M is bounded between $(1, \bar{N}^{-1})$, $\tilde{\boldsymbol{\omega}}_\varepsilon'\tilde{\boldsymbol{\omega}}_\varepsilon / M - (\mathbf{W}'\mathbf{W}/M)(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/N) \xrightarrow{a.s.} \mathbf{0}_{QxQ}$, as stated in the lemma.

Lemma B5: We need to show that for n = 1...4, $d_{i1}d_{i2}$ and $d_{j3}d_{j4}$ each equal to the product of the elements of two columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ with no more than one in each case equal to $\boldsymbol{\varepsilon}$, and some a in (0,½)

$$\text{(C5.1) } M^{-1}\sum_{m=1}^{M} \frac{[M^{-a\max(n-2,0)}\prod_{k=1}^{n}\boldsymbol{\chi}_{mk} - \omega(M^{-a\max(n-2,0)}\prod_{k=1}^{n}\boldsymbol{\chi}_{mk})]^2}{M} \sum_{m=1}^{M} \frac{[\sum_{i\in m} d_{i1}d_{i2}d_{i3}d_{i4} - \sum_{m=1}^{M}\sum_{i\in m} d_{i1}d_{i2}d_{i3}d_{i4}/M]^2}{M}$$

$$= [M^{-2a\max(n-2,0)}\omega(\prod_{k=1}^{n}\boldsymbol{\chi}_{mk}^2) - \left(M^{-a\max(n-2,0)}\omega(\prod_{k=1}^{n}\boldsymbol{\chi}_{mk})\right)^2] \frac{\left[m_m(d_{i1}d_{i2}d_{i3}d_{i4}, d_{j1}d_{j2}d_{j3}d_{j4}) - m(d_{i1}d_{i2}d_{i3}d_{i4})^2\right]}{M} \xrightarrow{a.s.} 0.$$

The $m()$ means of any 4 columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ (no more than two of which are $\boldsymbol{\varepsilon}$) have already been shown to be almost surely bounded (Lemma B2c), while from Lemmas B2c and B2d we know that the $\omega()$ mean of n = 1..4 columns of $\boldsymbol{\chi}$ times $M^{-a\max(n-2,0)}$ and the $\omega()$ mean of $2n$ = 2, 4, 6, or 8 columns of $\boldsymbol{\chi}$ times $M^{-2a\max(n-2,0)}$ are all almost surely bounded. So all that is needed to establish (C5.1) is to show that:

$$\text{(C5.2) } \frac{m_m(d_{i1}d_{i2}d_{i3}d_{i4}, d_{j1}d_{j2}d_{j3}d_{j4})}{M} = M^{-2}\sum_{m=1}^{M}\left(\sum_{i\in m} d_{i1}d_{i2}d_{i3}d_{i4}\right)^2 \le M^{-2}\sum_{m=1}^{M}\left(\sum_{i\in m} d_{i1}^2 d_{i2}^2 \sum_{j\in m} d_{j3}^2 d_{j4}^2\right)$$

$$\le \underset{i\le N}{Max}\frac{d_{i1}^2 d_{i2}^2}{M^2}\sum_{m=1}^{M}\sum_{i\in m}\sum_{j\in m} d_{j3}^2 d_{j4}^2 \le \underset{i\le N}{Max}\frac{d_{i1}^2 d_{i2}^2}{N}\frac{\bar{N}N}{M}\sum_{m=1}^{M}\sum_{j\in m}\frac{d_{j3}^2 d_{j4}^2}{M} = \underset{i\le N}{Max}\frac{d_{i1}^2 d_{i2}^2}{N}\frac{\bar{N}N}{M} m(d_{j3}^2 d_{j4}^2) \xrightarrow{a.s.} 0.$$

$\bar{N}N/M$ is bounded by $\bar{N}^2$ and $m(d_{j3}^2 d_{j4}^2)$ is almost surely bounded by Lemma B2c. From (C2.8) earlier we know that $E(|d_{ij}^2 d_{ik}^2|^{1+\delta}) < \Delta < \infty$ for some $\delta > 0$, so applying Markov's Inequality

$$\text{(C5.3)}\ \sum_{N=1}^{\infty} P(d_{Nj}^2 d_{Nk}^2 \geq N^a) \leq \sum_{N=1}^{\infty} \frac{\Delta}{N^{a(1+\delta)}} < \infty \ \text{ if } \ a(1+\delta) > 1.$$

As $\delta > 0$, we know that there exists an $a < 1$ such that (C5.3) holds, so by the Borel-Cantelli Lemma Corollary $\max_{i \leq N} d_{ij}^2 d_{ik}^2 / N^a$ is almost surely bounded by 1, while $\max_{i \leq N} d_{ij}^2 d_{ik}^2 / N \xrightarrow{a.s.} 0$, as desired.

Lemma B6: From the proof of Lemma B2a we know that $\tilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}' \tilde{\mathbf{W}}_{\boldsymbol{\varepsilon}} / M - \mathbf{W}_M \xrightarrow{a.s.} \mathbf{0}_{QxQ}$. When using the heteroskedasticity robust covariance estimate, from U5b we have

$$\text{(C6.1)}\ \mathbf{W}_M = \sum_{m=1}^{M} \frac{E(\mathbf{W}_m' \boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}_m' \mathbf{W}_m)}{M} = \sum_{m=1}^{M} \sum_{i \in m} \sum_{j \in m} \frac{E(\mathbf{w}_i \varepsilon_i \varepsilon_j \mathbf{w}_j')}{M} = \sum_{m=1}^{M} \sum_{i \in m} \frac{E(\mathbf{w}_i \varepsilon_i^2 \mathbf{w}_i')}{M} = \sum_{i=1}^{N} \frac{E(\mathbf{w}_i \varepsilon_i^2 \mathbf{w}_i')}{M}.$$

From (C1.14) earlier we know that $E(|\, z_{+ij} z_{+ik} \varepsilon_i^2 \,|^{1+\delta}) < \Delta < \infty$. Since $\mathbf{W}$ is a part of $\mathbf{Z}_+$ (assumption A2), applying the LLNHDS Corollary and using the fact that N/M is bounded between 1 and $\overline{N}$

$$\text{(C6.2)}\ \sum_{i=1}^{N} \frac{\mathbf{w}_i \varepsilon_i^2 \mathbf{w}_i'}{N} - \sum_{i=1}^{N} \frac{E(\mathbf{w}_i \varepsilon_i^2 \mathbf{w}_i')}{N} = \frac{M}{N} \left(\frac{\mathbf{W}_{\boldsymbol{\varepsilon}}' \mathbf{W}_{\boldsymbol{\varepsilon}}}{M} - \mathbf{W}_M \right) \xrightarrow{a.s.} \mathbf{0}_{QxQ},$$

thereby proving Lemma B6.

Lemma B7: We cannot directly invoke Theorem III, as the expressions in the lemma involve cluster groupings and generate somewhat different results, but the method of proof will be similar. To minimize clutter, for the purposes of this proof we change notation and let t_{m1}, t_{m2}, ... (or generically t_{mk}) each denote the product of 1 *or* 2 columns of $\boldsymbol{\tau}$, with x_{m1}, x_{m2} ... (x_{mk}) and t_{i1}, t_{i2}, ... (t_{ik}) denoting the products of corresponding columns of $\mathbf{X}$ and $\mathbf{T}$. We also let v_{i1} and v_{j2} each denote the product of two columns of ($\mathbf{Z}_+$,$\boldsymbol{\varepsilon}$). In addition, define m_i as the group m associated with observation i, and $I_{m_g = m_i}$ a (0,1) indicator function for whether $m_g = m_i$, and note that, since for all m t_{mk} has the same expectation across the row permutations $\boldsymbol{\tau}$ of $\mathbf{X}$, $E_{\boldsymbol{\tau}}(t_{m_i k}) = E_{\boldsymbol{\tau}}(t_{mk})$, with similar results for higher moments. We begin by using the symmetry and equal likelihood of permutations to calculate the expectation of t_{mk} and products of t_{mk} across the row permutations $\boldsymbol{\tau}$ of $\mathbf{X}$:

$$\text{(C7.1)}\ \ E_{\boldsymbol{\tau}}(t_{mk}) = \sum_{m=1}^{M} \frac{x_{mk}}{M} = \omega(x_{mk}),\ E_{\boldsymbol{\tau}}(t_{mk}^2) = \sum_{m=1}^{M} \frac{x_{mk}^2}{M} = \omega(x_{mk}^2)$$

$$\& \ E_{\boldsymbol{\tau}}(t_{mk} t_{n(\neq m)k}) = \sum_{m=1}^{M} \sum_{n=1}^{M} \frac{x_{mk} x_{nk}}{M(M-1)} - \sum_{m=1}^{M} \frac{x_{mk}^2}{M(M-1)} = \frac{\omega(x_{mk})^2 M}{M-1} - \frac{\omega(x_{mk}^2)}{M-1}.$$

We then use these to calculate the expectation across row permutations $\boldsymbol{\tau}$ of the expression in (L7a):

$$\text{(C7.2)}\quad E_{\boldsymbol{\tau}}(m_c(t_{i1}v_{i1},v_{j2}))=\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{E_{\boldsymbol{\tau}}(t_{m_i1})v_{i1}v_{j2}}{M}=\omega(\boldsymbol{x}_{m1})m_c(v_{i1},v_{j2})$$

$$E_{\boldsymbol{\tau}}(m_c(t_{i1}v_{i1},v_{i2})^2)=\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{E_{\boldsymbol{\tau}}(t_{m_i1}^2)v_{g1}v_{h2}v_{i1}v_{j2}}{M^2}I_{m_g=m_i}$$

$$+\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{E_{\boldsymbol{\tau}}(t_{m_g1}t_{m_i1})v_{g1}v_{h2}v_{i1}v_{j2}}{M^2}I_{m_g\neq m_i}$$

$$=\omega(\boldsymbol{x}_{m1}^2)\frac{\sum I_{m_g=m_i}}{M}+\left[\frac{\omega(\boldsymbol{x}_{m1})^2M}{M-1}-\frac{\omega(\boldsymbol{x}_{m1}^2)}{M-1}\right]\left(m_c(v_{i1},v_{j2})^2-\frac{\sum I_{m_g=m_i}}{M}\right)$$

$$\text{where}\ \sum I_{m_g=m_i}=\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{g1}v_{h2}v_{i1}v_{j2}}{M}I_{m_g=m_i}$$

$$\Rightarrow E_{\boldsymbol{\tau}}([m_c(t_{i1}v_{i1},v_{i2})-\omega(\boldsymbol{x}_{m1})m_c(v_{i1},v_{j2})]^2)=$$

$$\frac{\sum I_{m_g=m_i}-m_c(v_{i1},v_{j2})^2}{M-1}[\omega(\boldsymbol{x}_{m1}^2)-\omega(\boldsymbol{x}_{m1})^2]\xrightarrow{a.s.}0,$$

where the last follows by the fact that $\boldsymbol{x}_{m1}$ is at most the product of elements of two columns of $\boldsymbol{\mathcal{X}}$, while v_{i1} and v_{j2} are always the product of elements of two columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$, so $\omega(\boldsymbol{x}_{m1})$, $\omega(\boldsymbol{x}_{m1}^2)$, & $m_c(v_{i1},v_{j2})$ are by Lemma B2c bounded, while using the Cauchy-Schwarz Inequality and the bound $\overline{N}$ on the size of clusters we see that

$$\text{(C7.3)}\quad \left|\sum I_{m_g=m_i}\right|\leq\sqrt{\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{g1}^2v_{h2}^2}{M}I_{m_g=m_i}\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{i1}^2v_{j2}^2}{M}I_{m_g=m_i}}$$

$$=\sum_{c=1}^{C}\sum_{g\in c}\sum_{h\in c}\frac{v_{g1}^2v_{h2}^2}{M}\sum_{d=1}^{C}\sum_{i\in d}\sum_{j\in d}I_{m_g=m_i}\leq\overline{N}^2\sum_{c=1}^{C}\sum_{g\in c}\sum_{h\in c}\frac{v_{g1}^2v_{h2}^2}{M}=\overline{N}^2m_c(v_{g1}^2,v_{h2}^2),$$

since $\sum_{j\in d}1\leq\overline{N}$ & $\sum_{d=1}^{C}\sum_{i\in d}I_{m_g=m_i}\leq\overline{N}$ given the upper bound on maximum grouping size.

So the expressions in (L7a) converge in mean square, and hence in probability to zero, as stated in the Lemma.[5]

Turning to (L7b), we first calculate the expectation:

$$\text{(C7.4)}\quad E_{\boldsymbol{\tau}}(t_{m1}t_{m2})=\omega(\boldsymbol{x}_{m1}\boldsymbol{x}_{m2})$$

$$E_{\boldsymbol{\tau}}(t_{m1}t_{n2\ (n\neq m)})=\sum_{m=1}^{M}\sum_{n=1}^{M}\frac{t_{m1}t_{n2}}{M(M-1)}-\sum_{m=1}^{M}\frac{t_{m1}t_{m2}}{M(M-1)}=\frac{M\omega(\boldsymbol{x}_{m1})\omega(\boldsymbol{x}_{m2})}{M-1}-\frac{\omega(\boldsymbol{x}_{m1}\boldsymbol{x}_{m2})}{M-1}=\omega(\boldsymbol{x}_{m1})\omega(\boldsymbol{x}_{m2})+o_{a.s.}(1)$$

$$E_{\boldsymbol{\tau}}(m_c(t_{i1}v_{i1},t_{j2}v_{j2}))=\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{E_{\boldsymbol{\tau}}(t_{m1}t_{m2})v_{i1}v_{j2}}{M}I_{m_i=m_j}+\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{E_{\boldsymbol{\tau}}(t_{m1}t_{n2\ (n\neq m)})v_{i1}v_{j2}}{M}I_{m_i\neq m_j}$$

$$=\omega(\boldsymbol{x}_{m1}\boldsymbol{x}_{m2})m_v(v_{i1},v_{j2})+\omega(\boldsymbol{x}_{m1})\omega(\boldsymbol{x}_{m2})[m_c(v_{i1},v_{j2})-m_v(v_{i1},v_{j2})]+o_{a.s.}(1),$$

[5]When comparing results in this section with the Lemma, keep in mind that in the proof t_{i1} may denote 1 or the product of 2 columns of **T**, but in the statement of the Lemma it only denotes one column.

where we use the notation $o_{a.s.}(1)$ to denote sequences which almost surely converge to zero and make use of the fact (Lemma B2c) that the $\omega()$ mean for products of up to 4 elements of $\boldsymbol{X}$ are known to be almost surely bounded, as are $m_c(v_{i1},v_{j2})$ and $m_v(v_{i1},v_{j2})$. Calculating the second moment is considerably more complicated.

We begin by calculating the expectation of the product of four arbitrary $t_{m1}... t_{m4}$, varying by whether indices do or don't match

$$\text{(C7.5)}\quad E_{\boldsymbol{T}}(t_{m1}t_{m2}t_{m3}t_{m4}) = \sum_{m=1}^{M}\frac{t_{m1}t_{m2}t_{m3}t_{m4}}{M} = \omega(x_{m1}x_{m2}x_{m3}x_{m4})$$

$$E_{\boldsymbol{T}}(t_{m1}t_{m2}t_{m3}\underset{(\neq m)}{t_{n4}}) = \sum_{m=1}^{M}\sum_{n=1}^{M}\frac{t_{m1}t_{m2}t_{m3}t_{n4}}{M(M-1)} - \sum_{m=1}^{M}\frac{t_{m1}t_{m2}t_{m3}t_{m4}}{M(M-1)} = \omega(x_{m1}x_{m2}x_{m3})\omega(x_{m4}) + o_{a.s.}(1)$$

$$E_{\boldsymbol{T}}(t_{m1}t_{m2}\underset{(\neq m)}{t_{n3}}\underset{(\neq m)}{t_{n4}}) = \sum_{m=1}^{M}\sum_{n=1}^{M}\frac{t_{m1}t_{m2}t_{n3}t_{n4}}{M(M-1)} - \sum_{m=1}^{M}\frac{t_{m1}t_{m2}t_{m3}t_{m4}}{M(M-1)} = \omega(x_{m1}x_{m2})\omega(x_{m3}x_{m4}) + o_{a.s.}(1)$$

$$E_{\boldsymbol{T}}(t_{m1}t_{m2}\underset{(\neq m)}{t_{n3}}\underset{(\neq m,n)}{t_{o4}}) = \frac{\left(\begin{array}{c}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{o=1}^{M}t_{m1}t_{m2}t_{n3}t_{o4} - \sum_{v=3}^{4}\sum_{m=1}^{M}\sum_{n=1}^{M}t_{nv}\prod_{k=1(\neq v)}^{4}t_{mk} \\ -\sum_{m=1}^{M}\sum_{n=1}^{M}t_{m1}t_{m2}t_{n3}t_{n4} + 2\sum_{m=1}^{M}\prod_{k=1}^{4}t_{mk}\end{array}\right)}{M(M-1)(M-2)} = \omega(x_{m1}x_{m2})\omega(x_{m3})\omega(x_{m4}) + o_{a.s.}(1)$$

$$E_{\boldsymbol{T}}(t_{m1}\underset{(\neq m)}{t_{n2}}\underset{(\neq m,n)}{t_{o3}}\underset{(\neq m,n,o)}{t_{p4}}) = \frac{\left(\begin{array}{c}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{o=1}^{M}\sum_{p=1}^{M}t_{m1}t_{n2}t_{o3}t_{p4} - \sum_{w=1}^{3}\sum_{v=w+1}^{4}\sum_{m=1}^{M}\sum_{n=1}^{M}\sum_{o=1}^{M}t_{nw}t_{ov}\prod_{k=1(\neq w,v)}^{4}t_{mk} + \\ 2\sum_{v=1}^{4}\sum_{m=1}^{M}\sum_{n=1}^{M}t_{nv}\prod_{k=1(\neq v)}^{4}t_{mk} + \sum_{v=2}^{4}\sum_{m=1}^{M}\sum_{n=1}^{M}t_{nv}\prod_{k=1(\neq v)}^{4}t_{mk} - 6\sum_{m=1}^{M}\prod_{k=1}^{4}t_{mk}\end{array}\right)}{M(M-1)(M-2)(M-3)} = \prod_{k=1}^{4}\omega(x_{mk}) + o_{a.s.}(1),$$

where $\prod_{k=1(\neq v)}^{4}t_{mk}$ denotes the product of across $k = 1$ through 4, excluding v (or w & v if $\neq w,v$), and where the $o_{a.s.}(1)$ terms incorporate all elements whose limit based upon Lemmas B2c and B2d are known to be 0, taking into account that each t_{mk} is at most the product of two columns of $\boldsymbol{X}$. (C7.5) retains terms which might be unbounded. For example, if each t_{mk} is the product of two columns, then $\omega(x_{m1}x_{m2}x_{m3}x_{m4})$ is the mean of the product of 8 columns, and bounds on this have not been established, but Lemma B2d tells us it will almost surely converge to 0 if divided by M^{2a}. Using (C7.5) we calculate the second moment:

$$\text{(C7.6)}\quad E_{\boldsymbol{\tau}}(m_c(t_{i1}v_{i1},t_{j2}v_{j2})^2)=E_{\boldsymbol{\tau}}\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{t_{g1}t_{h2}t_{i1}t_{j2}v_{g1}v_{h2}v_{i1}v_{j2}}{M^2}$$

$$=o_{a.s.}(1)+\omega(\boldsymbol{x}_{m1}^2\boldsymbol{x}_{m2}^2)\sum I_{m_g=m_h=m_i=m_j}+\omega(\boldsymbol{x}_{m1}^2)\omega(\boldsymbol{x}_{m2}^2)\sum(I_{m_g=m_i}I_{m_h=m_j}-I_{m_g=m_h=m_i=m_j})$$

$$+\omega(\boldsymbol{x}_{m1}\boldsymbol{x}_{m2})^2\sum(I_{m_g=m_h}I_{m_i=m_j}+I_{m_g=m_j}I_{m_h=m_i}-2I_{m_g=m_h=m_i=m_j})$$

$$+\omega(\boldsymbol{x}_{m1}^2\boldsymbol{x}_{m2})\omega(\boldsymbol{x}_{i2})\sum(I_{m_g=m_h=m_i}+I_{m_g=m_i=m_j}-2I_{m_g=m_h=m_i=m_j})$$

$$+\omega(\boldsymbol{x}_{m2}^2\boldsymbol{x}_{m1})\omega(\boldsymbol{x}_{i1})\sum(I_{m_g=m_h=m_j}+I_{m_h=m_i=m_j}-2I_{m_g=m_h=m_i=m_j})$$

$$+\begin{bmatrix}\omega(\boldsymbol{x}_{m1}\boldsymbol{x}_{m2})*\\ \omega(\boldsymbol{x}_{m1})\omega(\boldsymbol{x}_{m2})\end{bmatrix}\sum\begin{pmatrix}I_{m_g=m_h}+I_{m_i=m_j}+I_{m_g=m_j}+I_{m_h=m_i}-2I_{m_g=m_h}I_{m_i=m_j}-2I_{m_g=m_j}I_{m_h=m_i}\\+8I_{m_g=m_h=m_i=m_j}-2I_{m_g=m_h=m_i}-2I_{m_g=m_h=m_j}-2I_{m_g=m_i=m_j}-2I_{m_h=m_i=m_j}\end{pmatrix}$$

$$+\omega(\boldsymbol{x}_{m1}^2)\omega(\boldsymbol{x}_{m2})^2\sum(I_{m_g=m_i}-I_{m_g=m_i}I_{m_h=m_j}-I_{m_g=m_h=m_i}-I_{m_g=m_i=m_j}+2I_{m_g=m_h=m_i=m_j})$$

$$+\omega(\boldsymbol{x}_{m2}^2)\omega(\boldsymbol{x}_{m1})^2\sum(I_{m_h=m_j}-I_{m_g=m_i}I_{m_h=m_j}-I_{m_g=m_h=m_j}-I_{m_h=m_i=m_j}+2I_{m_g=m_h=m_i=m_j})$$

$$+\omega(\boldsymbol{x}_{m1})^2\omega(\boldsymbol{x}_{m2})^2\sum\begin{pmatrix}1+2I_{m_g=m_h=m_i}+2I_{m_g=m_i=m_j}+2I_{m_g=m_h=m_j}+2I_{m_h=m_i=m_j}+I_{m_g=m_h}I_{m_i=m_j}+I_{m_g=m_j}I_{m_h=m_i}\\+I_{m_g=m_i}I_{m_h=m_j}-6I_{m_g=m_h=m_i=m_j}-I_{m_g=m_h}-I_{m_i=m_j}-I_{m_g=m_j}-I_{m_h=m_i}-I_{m_g=m_i}-I_{m_h=m_j}\end{pmatrix},$$

$$\text{where }\sum I_{\mathrm{a}}(\text{or } I_{\mathrm{a}}I_{\mathrm{b}})=\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{g1}v_{h2}v_{i1}v_{j2}}{M^2}I_{\mathrm{a}}(\text{or } I_{\mathrm{a}}I_{\mathrm{b}}),$$

and where the $o_{a.s.}(1)$ emerges from the product of the $o_{a.s.}(1)$ terms in (C7.5) with the $\sum I$ terms in (C7.6) which, making frequent use of the Cauchy-Schwarz Inequality are all shown to be almost surely bounded:

$$\text{(C7.7)}\quad \sum I_{m_g=m_h}=\sum I_{m_i=m_j}=m_v(v_{g1},v_{h2})m_c(v_{g1},v_{h2}),\ \sum I_{m_g=m_h}I_{m_i=m_j}=m_v(v_{g1},v_{h2})^2,$$

$$\left|\sum I_{m_g=m_h=m_i=m_j}\right|\le\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{g1}^2v_{h2}^2}{M^2}I_{m_g=m_h=m_i=m_j}\le\frac{\overline{N}^2}{M}m_v(v_{g1}^2,v_{h2}^2)\xrightarrow{a.s.}0,$$

for $u=$"$m_g=m_h=m_i$", "$m_g=m_h=m_j$", "$m_g=m_i=m_j$" or "$m_h=m_i=m_j$":

$$\left|\sum I_u\right|\le\sqrt{\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{g1}^2v_{h2}^2}{M^2}I_u\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{i1}^2v_{j2}^2}{M^2}I_u}\le\sqrt{\frac{\overline{N}^4}{M^2}m_v(v_{g1}^2,v_{h2}^2)m_c(v_{g1}^2,v_{h2}^2)}\xrightarrow{a.s.}0,$$

for $a,b=m_i,m_j$ or $a,b=m_j,m_i$:

$$\left|\sum I_{m_g=a}I_{m_h=b}\right|\le\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{g1}^2v_{h2}^2}{M^2}I_{m_g=a}I_{m_h=b}\le\frac{\overline{N}^2}{M}m_c(v_{g1}^2,v_{h2}^2)\xrightarrow{a.s.}0,$$

for $a=g$ or h and $b=i$ or j

$$\left|\sum I_{m_a=m_b}\right|\le\sqrt{\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{g1}^2v_{h2}^2}{M^2}I_{m_a=m_b}\sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{g\in c}\sum_{h\in c}\sum_{i\in d}\sum_{j\in d}\frac{v_{i1}^2v_{j2}^2}{M^2}I_{m_a=m_b}}\le\frac{\overline{N}^2}{M}m_c(v_{g1}^2,v_{h2}^2)\xrightarrow{a.s.}0,$$

keeping in mind that v_{g1} & v_{g2} are always the product of elements of two columns of $(\mathbf{Z}_{+},\boldsymbol{\varepsilon})$, so by Lemma B2d we know that when divided by M all m_c and m_v terms involving the square of these converge to 0.

If each t_{ik} involves only $n_k = 1$ columns of $\mathbf{T}$, then all the ω in (C7.6) are known to be bounded, and hence vanish if multiplied by something that converges to 0, so using (C7.7) we can simplify to

$$\text{(C7.8)}\quad E_{\boldsymbol{\tau}}(m_c(t_{i1}v_{i1},t_{j2}v_{j2})^2)-\left(\omega(\boldsymbol{x}_{m1}\boldsymbol{x}_{m2})m_v(v_{i1},v_{j2})+\omega(\boldsymbol{x}_{m1})\omega(\boldsymbol{x}_{m2})[m_c(v_{i1},v_{j2})-m_v(v_{i1},v_{j2})]\right)^2 \xrightarrow{a.s.} 0$$

$$\Rightarrow E_{\boldsymbol{\tau}}(\left[m_c(t_{i1}v_{i1},t_{j2}v_{j2})-\omega(\boldsymbol{x}_{m1}\boldsymbol{x}_{m2})m_v(v_{i1},v_{j2})-\omega(\boldsymbol{x}_{m1})\omega(\boldsymbol{x}_{m2})[m_c(v_{i1},v_{j2})-m_v(v_{i1},v_{j2})]\right]^2)\xrightarrow{a.s.}0,$$

and see that the expression in (L7b) converges in mean square and hence in probability to 0. When $n_1 + n_2 > 2$, so that at least one of the t_{ik} involves more than one column of $\mathbf{T}$, the mean $\omega(\boldsymbol{x}_{m1}^2\boldsymbol{x}_{m2}^2)$ in (C7.6) involves the product of more than 4 columns of $\boldsymbol{\mathcal{X}}$, as will at least one of $\omega(\boldsymbol{x}_{m1}^2\boldsymbol{x}_{m2})$ and $\omega(\boldsymbol{x}_{m1}\boldsymbol{x}_{m2}^2)$, and hence we don't know if these means are bounded. However, from Lemma B2d we know that if multiplied by $M^{-a(2n_1+2n_2-4)}$ their limit is 0, as for $b1$ and $b2$ each equal to 1 or 2:[6]

$$\text{(C7.9)}\quad M^{-a(2n_1+2n_2-4)}\omega(\boldsymbol{x}_{m1}^{b1}\boldsymbol{x}_{m2}^{b2})=\overbrace{M^{-a(n_1+n_2-2)}}^{\to 0}\overbrace{M^{-a(\frac{2n_1+2n_2}{2}-2)}\omega(\boldsymbol{x}_{m1}^{b1}\boldsymbol{x}_{m2}^{b2})}^{\xrightarrow{a.s.}0\text{ (Lemma B2d)}}\xrightarrow{a.s.}0.$$

Hence, in this case we multiply $m_c(t_{i1}v_{i1},t_{j2}v_{j2})$ by $M^{-a(n_1+n_2-2)}$ and see that

$$\text{(C7.10)}\quad \underbrace{M^{-a(n_1+n_2-2)}E_{\boldsymbol{\tau}}m_c(t_{i1}v_{i1},t_{j2}v_{j2})}_{\text{by (C7.4)}}\xrightarrow{a.s.}0\ \ \&\ \ \underbrace{M^{-a(2n_1+2n_2-4)}E_{\boldsymbol{\tau}}(m_c(t_{i1}v_{i1},t_{j2}v_{j2})^2)}_{\text{by (C7.6), (C7.7) \& (C7.9)}}\xrightarrow{a.s.}0,$$

thereby ensuring convergence in mean square and probability to zero, verifying the claim in (L7c).

Lemma B8: From the proof of Lemma B2a we know that $\tilde{\boldsymbol{\mathcal{W}}}_{\boldsymbol{\varepsilon}}'\tilde{\boldsymbol{\mathcal{W}}}_{\boldsymbol{\varepsilon}}/M-\mathbf{W}_M\xrightarrow{a.s.}\mathbf{0}_{QxQ}$. When using the clustered covariance estimate the random vectors $\mathbf{w}_i\varepsilon_i$ and $\mathbf{w}_j\varepsilon_j$ are independent with expectation $\mathbf{0}_Q$ if in different clusters (U1b, U5c), so with intersection groupings v defined as the largest observational grouping such that all observations belong to at most one cluster grouping c and one treatment grouping m, with the number of such groupings $V \le N$:

$$\text{(C8.1)}\ \mathbf{W}_M=\sum_{m=1}^{M}\frac{E(\mathbf{W}_m'\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m'\mathbf{W}_m)}{M}=\sum_{m=1}^{M}\sum_{i\in m}\sum_{j\in m}\frac{E(\mathbf{w}_i\varepsilon_i\varepsilon_j\mathbf{w}_j')}{M}=$$

$$\sum_{m=1}^{M}\sum_{v\subseteq m}\sum_{i\in v}\sum_{j\in v}\frac{E(\mathbf{w}_i\varepsilon_i\varepsilon_j\mathbf{w}_j')}{M}=\sum_{m=1}^{M}E\left(\sum_{v\subseteq m}\frac{\mathbf{W}_v'\boldsymbol{\varepsilon}_v\boldsymbol{\varepsilon}_v'\mathbf{W}_v}{M}\right)=\sum_{v=1}^{V}E\left(\frac{\mathbf{W}_v'\boldsymbol{\varepsilon}_v\boldsymbol{\varepsilon}_v'\mathbf{W}_v}{M}\right).$$

The expectation of $|\,\mathbf{z}_{+vj}'\boldsymbol{\varepsilon}_v\mathbf{z}_{+vk}'\boldsymbol{\varepsilon}_v\,|^{1+\delta}$ is for all v, j and k bounded as

$$\text{(C8.2)}\ E(|\,\mathbf{z}_{+vj}'\boldsymbol{\varepsilon}_v\mathbf{z}_{+vk}'\boldsymbol{\varepsilon}_v\,|^{1+\delta})\le\overbrace{\sqrt{\prod_{n=j,k}E(|\,(\mathbf{z}_{+vn}'\boldsymbol{\varepsilon}_v)^2\,|^{1+\delta})}}^{\text{Hölder's Inequality}}\le\overbrace{\sqrt{\prod_{n=j,k}E(|\,\mathbf{z}_{+vn}'\mathbf{z}_{+vn}\boldsymbol{\varepsilon}_v'\boldsymbol{\varepsilon}_v\,|^{1+\delta})}}^{\text{Cauchy-Schwarz Inequality}}$$

$$\le\overbrace{\sqrt{\prod_{n=j,k}E(|\,\mathbf{z}_{+u_vn}'\mathbf{z}_{+u_vn}\boldsymbol{\varepsilon}_{u_v}'\boldsymbol{\varepsilon}_{u_v}\,|^{1+\delta})}}^{v\subseteq u}\overbrace{<\Delta}^{\text{U3a}},$$

where u_v denotes the union grouping to which intersection grouping v belongs. Since $\mathbf{W}$ is a part of $\mathbf{Z}_+$ (assumption A2), applying the LLNHDS Corollary and the fact that V/M is bounded between 1 and $\overline{N}$ (assumption A4), it then follows that

[6] In applying Lemma B2d here, as $b1$ and $b2$ are 1 or 2, I put $2n_1$ and $2n_2$ in the exponent on M to ensure that the lemma is satisfied (n_k denoting the number of columns of $\mathbf{T}$ in t_{ik} and hence the number of columns of $\boldsymbol{\mathcal{X}}$ in $\boldsymbol{x}_{mk}$).

$$\text{(C8.3)}\quad \frac{\mathscr{W}_{\varepsilon}'\mathscr{W}_{\varepsilon}}{M} - \mathbf{W}_M = \frac{V}{M}\left(\sum_{v=1}^{V}\frac{\mathbf{W}_v'\boldsymbol{\varepsilon}_v\boldsymbol{\varepsilon}_v'\mathbf{W}_v}{V} - \sum_{v=1}^{V}E\left(\frac{\mathbf{W}_v'\boldsymbol{\varepsilon}_v\boldsymbol{\varepsilon}_v'\mathbf{W}_v}{V}\right)\right) \xrightarrow{a.s.} \mathbf{0}_{QxQ},$$

thereby proving the lemma.

Table D1: Notation used in Appendices D & E (also reviewed as introduced in the appendices)

(1) Regression model: $\mathbf{y} = \mathbf{X_W}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ or $\mathbf{y} = \mathbf{Z}_+\boldsymbol{\gamma}_+ + \boldsymbol{\varepsilon}$ where $\mathbf{Z}_+ = (\mathbf{X_W},\mathbf{Z})$ and $\boldsymbol{\gamma}'_+ = (\boldsymbol{\beta}',\boldsymbol{\gamma}')$. Estimated parameters denoted by $\hat{}$. $\mathbf{X}$ is $NxPQ$, $\mathbf{Z}$ NxK and $\mathbf{Z}_+$ NxK_+.

(2) $\mathbf{A_B}$ and $\bullet$ denote the row by row Kronecker product, $\mathbf{A_B} = \mathbf{A} \bullet \mathbf{B}$. This appears in the form of NxP treatment variables $\mathbf{X}$ multiplied by NxQ interaction covariates $\mathbf{W}$ ($\mathbf{X_W}$) and the multiplication of $\mathbf{W}$ with errors $\boldsymbol{\varepsilon}$ ($\mathbf{W}_{\boldsymbol{\varepsilon}}$). $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product.

(3) $\mathbf{t_{W}}_k$ and $\mathbf{x_{W}}_l$ denote the k^{th} and l^{th} columns of $\mathbf{T_W}$ and $\mathbf{X_W}$, with the i^{th} elements of these vectors given by $t_{ip(k)}w_{iq(k)}$ and $x_{ip(l)}w_{iq(l)}$, where $p(j)$ and $q(j)$ denote the columns of $\mathbf{T}$ (or $\mathbf{X}$) and $\mathbf{W}$ associated with the j^{th} column of $\mathbf{T_W}$ (or $\mathbf{X_W}$).

(4) Sample of N observations divided into S strata with N_s observations in stratum s. Subscript s denotes the sub-matrix associated with stratum s, as in $\mathbf{X}_s$ and $\mathbf{W}_{\boldsymbol{\varepsilon}s}$. $\sum_{i \in s}$ denotes summation across observations i in stratum s.

(5) $\mathbf{T}$ denotes a stratified row permutation of $\mathbf{X}$.

(6) $\mathbf{y}_{\mathbf{T},\boldsymbol{\beta}_0} = \mathbf{y} + (\mathbf{T_W} - \mathbf{X_W})\boldsymbol{\beta}_0$ denotes the counterfactual value of $\mathbf{y}$ following stratified row permutation $\mathbf{T}$ of $\mathbf{X}$ under the null $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}$ the associated parameter estimates.

(7) $\mathbf{1}_N$ & $\mathbf{0}_N$ denote $Nx1$ vectors of 1s & 0s, $\mathbf{0}_{QxQ}$ a QxQ matrix of 0s & $\mathbf{I}_Q$ the QxQ identity matrix.

(8) Sample and stratum demeaned variables: $\widetilde{\mathbf{T}} = (\mathbf{I}_N - \mathbf{1}_N\mathbf{1}'_N/N)\mathbf{T}$ & $\widetilde{\mathbf{T}}_s = (\mathbf{I}_{N_s} - \mathbf{1}_{N_s}\mathbf{1}'_{N_s}/N_s)\mathbf{T}_s$.

(9) $m()$ and $m_s()$ denote full sample and stratum means, i.e.

$$m(x_{ip}) = \sum_{i=1}^{N} \frac{x_{ip}}{N},\ m_s(x_{ip}) = \sum_{i \in s} \frac{x_{ip}}{N_s},\ \&\ m(x_{ip}) = \sum_{s=1}^{S} \frac{N_s}{N} m_s(x_{ip}).$$

$\mathbf{m}(\mathbf{x}_i)$ & $\mathbf{m}(\mathbf{x}'_i)$ are column and row vector versions of these.

(10) $\xrightarrow{a.s.}$ & $\xrightarrow{p_{\mathbf{D}}}$ denote convergence almost surely and in probability across the probability law governing the data $\mathbf{D} = (\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})$. $\xrightarrow{p}$ & $\xrightarrow{d}$ denote convergence in probability and distribution across stratified permutations $\mathbf{T}$ of $\mathbf{X}$ in probability given the data $\mathbf{D}$. $E()$ denotes the expectation across the data $\mathbf{D}$.

(11) $\mathbf{n}_{PQ}$ denotes the multivariate iid standard normal, indicated by $\mathbf{n}_{PQ} \sim \mathrm{N}(\mathbf{0}_{PQ},\mathbf{I}_{PQ})$.

D. Stratification of Treatment

This appendix generalizes the results to allow for the stratified application of treatment. Specifically, it shows that White's assumptions justify the use of covariance estimates without adjustment for stratification (as is done universally in the published papers reviewed in Young 2019) and notes sufficient additional conditions for use of the distribution of coefficients and covariance estimates based upon within stratum permutations of treatment to yield asymptotically accurate inference. To minimize notational complexity, the analysis is done in the context of the framework presented in the paper, with treatment applied at the observation level and heteroskedasticity robust covariance estimates. Extension

to the grouped treatment/clustered covariances of Appendix B above is straightforward but uses much more notation.

The baseline regression model remains as in the paper

$$\text{(D.1)}\quad \mathbf{y} = \mathbf{X_W}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \text{ or } \mathbf{y} = \mathbf{Z}_+\boldsymbol{\gamma}_+ + \boldsymbol{\varepsilon},$$

where $\mathbf{Z}_+ = (\mathbf{X_W},\mathbf{Z})$ and $\boldsymbol{\gamma}'_+ = (\boldsymbol{\beta}',\boldsymbol{\gamma}')$. We now assume that the data are divided into a finite number S of strata, with the subscript s denoting the N_s rows of each matrix or vector associated with observations in stratum s, as in $\mathbf{Z}_s$. The notation $\mathbf{T}$ now denotes the *stratified* permutations of treatment $\mathbf{X}$, i.e. an outcome observed when permuting treatment within strata. We use the notation $\Sigma_{i\in s}$ to denote the sum across all observations i in stratum s, and $m()$ and $m_s()$ to denote the full sample and stratum means, as in:

$$\text{(D.2)}\quad m(x_{ip}) = \sum_{i=1}^{N}\frac{x_{ip}}{N} = \sum_{s=1}^{S}\frac{N_s}{N}\sum_{i\in s}\frac{x_{ip}}{N_s} = \sum_{s=1}^{S}\frac{N_s}{N}m_s(x_{ip}).$$

With $N = \sum_{s=1}^{S} N_s$ denoting the total number of observations in the regression, we assume that for all s $N_s/N > \Delta > 0$ for all N sufficiently large, so that $N \to \infty$ implies (and is of course implied by) $N_s \to \infty$ for all s.[7] All discussion below of limits is with respect to $N \to \infty$, and hence $N_s \to \infty$ in all strata.

In addition to the above, we assume:

(S1) White's assumptions W1 - W4 and the additional randomization specific assumptions A1 - A3 given in the paper hold for the entire sample. In addition, for all strata s $\Sigma_{i\in s} E(\mathbf{w}_i\mathbf{w}'_i\varepsilon_i^2)/N_s$ is non-singular with determinant $> \gamma > 0$ for all N_s sufficiently large.

(S2) While the expectation of non-treatment variables may differ systematically across strata, the asymptotic strata average first and second moments of treatment variables are almost surely identical, i.e.

$$\sum_{i\in s}\frac{E(\mathbf{x}_i)}{N_s} - \sum_{i\in t}\frac{E(\mathbf{x}_i)}{N_t} \xrightarrow{a.s.} \mathbf{0}_P \;\;\&\;\; \sum_{i\in s}\frac{E(\mathbf{x}_i\mathbf{x}'_i)}{N_s} - \sum_{i\in t}\frac{E(\mathbf{x}_i\mathbf{x}'_i)}{N_t} \xrightarrow{a.s.} \mathbf{0}_{PxP} \text{ for all } s, t = 1..S.$$

White's assumption W3b (given in the paper) ensured that the condition on the determinant in S1 holds for the average across the entire sample. The proofs later on require that it hold for individual strata as well. Assumption S2 corresponds to the A5 mentioned in the conclusion of the paper. It allows, but limits, heterogeneity of treatment across strata.

Since White's assumptions all hold, his result holds as well, and $\sqrt{N}(\hat{\boldsymbol{\gamma}}_+ - \boldsymbol{\gamma}_+)$ is asymptotically (across the data generating process for the data sequence $\mathbf{Z}_+,\boldsymbol{\varepsilon}$) normally distributed with mean $\mathbf{0}_{K_+}$ and positive definite covariance matrix $\mathbf{M}_N^{-1}\mathbf{V}_N\mathbf{M}_N^{-1}$ (as defined in W1-W4 in the paper), to which N times the heteroskedasticity robust covariance estimate calculated without consideration of the strata almost surely converges, $N\hat{\mathbf{V}}_h(\hat{\boldsymbol{\gamma}}_+) - \mathbf{M}_N^{-1}\mathbf{V}_N\mathbf{M}_N^{-1} \xrightarrow{a.s.} \mathbf{0}_{K_+xK_+}$ (see White 1980 or the more general proof for clustering in Appendix C above). As noted in the paper, White's assumptions in particular imply that average linear treatment effects do not vary systematically across strata, as a common parameter vector $\boldsymbol{\gamma}_+$ applies to all

[7] Given White's assumptions, strata for which this does not apply asymptotically almost surely have zero influence on coefficient and covariance estimates and can be ignored.

observations and $E(\mathbf{z}_{+i}\varepsilon_i) = \mathbf{0}_{K+}$ for all i. As in the non-stratified framework considered in the paper, White's assumption that the variables are independently distributed can be reconciled with dependence between observations within strata brought about by the application of a given distribution of treatment to the observations by appealing to O'Neill's (2009) de Finetti result that if the cumulative distribution function $F_{\mathbf{x}}$ of treatment within strata converges to a given distribution the exchangeable random variables asymptotically have an iid distribution.

Assumptions S1 and S2 ensure that the following hold:

Lemma D1: Across the probability distribution of the data generating process for $\mathbf{D} = (\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})$:

(a) $\mathbf{Z}'\mathbf{Z}/N$, $\mathbf{W}'\mathbf{W}/N$, $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N$ & $\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}/N$ are all almost surely strictly positive definite with determinant $> \nu > 0$ for all N sufficiently large, as are $\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s/N_s$ and $\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}/N_s$ for all strata s, while $\widetilde{\mathbf{W}}_s'\widetilde{\mathbf{W}}_s/N - \mathbf{W}_s'\mathbf{W}_s/N \xrightarrow{a.s.} \mathbf{0}_{QxQ}$ & $\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}/N_s - \mathbf{W}_{\boldsymbol{\varepsilon}s}'\mathbf{W}_{\boldsymbol{\varepsilon}s}/N_s \xrightarrow{a.s.} \mathbf{0}_{QxQ}$.

(b) $\mathbf{Z}'\boldsymbol{\varepsilon}/N \xrightarrow{a.s.} \mathbf{0}_K$ & $\mathbf{X}_{\mathbf{W}}'\boldsymbol{\varepsilon}/N \xrightarrow{a.s.} \mathbf{0}_{PQ}$.

(c) The full sample and strata means of the product of the elements of one, two, three or four columns of $\mathbf{X}$ or the elements of one, two or four columns of $\mathbf{D}$ (no more than two of which are ε_i) are almost surely bounded, as are $(\mathbf{Z}'\mathbf{Z}/N)^{-1}$, $(\mathbf{W}'\mathbf{W}/N)^{-1}$, $(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N)^{-1}$, $(\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}/N)^{-1}$, $(\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s/N_s)^{-1}$ and $(\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}/N_s)^{-1}$.

(d) For p and q denoting columns of $\mathbf{X}$ & $\mathbf{W}$ used to make any column of $\mathbf{X_W}$:

$$\sqrt{N}\left(\sum_{s=1}^{S}\frac{N_s}{N}m_s(x_{ip})m_s(w_{iq}\mathbf{z}_i') - m(x_{ip})m(w_{iq}\mathbf{z}_i')\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{N} \xrightarrow{p_{\mathbf{D}}} 0.$$

(e) For p and q denoting columns of $\mathbf{X}$ & $\mathbf{W}$ used to make any column of $\mathbf{X_W}$:

$$\sqrt{N}\left(\sum_{s=1}^{S}\frac{N_s}{N}m_s(x_{ip})m_s(w_{iq}\varepsilon_i) - m(x_{ip})m(w_{iq}\varepsilon_i)\right) \xrightarrow{p_{\mathbf{D}}} 0.$$

In the paper, and generally in this on-line appendix, convergence almost surely is with reference to the probability distribution of the data while convergence in probability or distribution is with respect to the permutations $\mathbf{T}$ of $\mathbf{X}$. However, there are some instances in this appendix, such as in Lemma D1, where convergence in probability is with respect to the distribution of the data $\mathbf{D}$, which I note by using the notation $p_{\mathbf{D}}$. Because of this, result (R1) in the paper is now stated in terms of in probability, rather than almost surely, across the distribution of the data. Stronger assumptions ensure almost sure convergence, but convergence in probability is sufficient for the objective of this appendix. Lemmas D1d and D1e are used later on in proofs of the following Lemma:

Lemma D2: In probability, across the distribution of the data $\mathbf{D} = (\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})$, the following hold across the probability distribution generated by stratified permutations $\mathbf{T}$ of treatment $\mathbf{X}$:

(a) With no more than one of the d_{ij} denoting ε_i,

$$m(t_{ip}d_{ij}d_{ik}) - m(x_{ip})m(d_{ij}d_{ik}) \xrightarrow{p} 0,\ m(t_{ip}t_{iq}d_{ij}d_{ik}) - m(x_{ip}x_{iq})m(d_{ij}d_{ik}) \xrightarrow{p} 0.$$

(b) For n = 1, 2, 3, or 4 and not more than two of the d_{ij} denoting ε_i, for some a in (0,½)

$$m(N^{-a\max(n-2,0)}(\prod_{o=1}^{n} t_{ip(o)})\, d_{ij}d_{ik}d_{il}d_{im}) \;-\; m(N^{-a\max(n-2,0)}\prod_{o=1}^{n} x_{ip(o)})m(d_{ij}d_{ik}d_{il}d_{im}) \;\xrightarrow{p}\; 0.$$

(c) For p and q denoting columns of $\mathbf{X}$ & $\mathbf{W}$ used to make any column of $\mathbf{X}_\mathbf{W}$:

$$\sqrt{N}\,[\mathbf{m}(t_{ip}w_{iq}\mathbf{z}_i') - m(x_{ip})\mathbf{m}(w_{iq}\mathbf{z}_i')]\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{N} \xrightarrow{p} 0.$$

(d) $$\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}_\varepsilon'\widetilde{\mathbf{W}}_\varepsilon}{N}\right)^{-½}\frac{(\widetilde{\mathbf{T}}\bullet\widetilde{\mathbf{W}}_\varepsilon)'\mathbf{1}_N}{\sqrt{N}} \xrightarrow{d} \mathbf{n}_{PQ},\ \text{where}\ \mathbf{n}_{PQ}\sim \mathrm{N}(\mathbf{0}_{PQ},\mathbf{I}_{PQ}).$$

The proofs of the lemmas are given in Appendix E.

With Lemmas D1 and D2 in hand, one can follow the steps used in the paper's appendix and prove that in probability across the distribution of the data **D** for a finite $\mathbf{r}=\sqrt{N}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$, asymptotically $\hat{\mathbf{r}}=\sqrt{N}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}-\boldsymbol{\beta}_0)$ is, across stratified permutations **T** of **X**, distributed multivariate mean zero normal with a covariance matrix equal to N times the probability limit of the heteroskedasticity robust covariance estimate $\hat{\mathbf{V}}_r(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})$, result (R1) in the paper. Results (R2) - (R5) then follow as in the paper's appendix. For the sake of completeness, the two sections below prove (R1), although the steps are nearly identical to those used in the paper's appendix.

(a) Asymptotic Distribution of Coefficient Estimates

As in the text, counterfactual output is given by

$$\text{(D.3)}\quad \mathbf{y}_{\mathbf{T},\boldsymbol{\beta}_0}=\mathbf{y}+(\mathbf{T}\bullet\mathbf{W}-\mathbf{X}\bullet\mathbf{W})\boldsymbol{\beta}_0=\mathbf{X}_\mathbf{W}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\mathbf{Z}\boldsymbol{\gamma}+\boldsymbol{\varepsilon}+\mathbf{T}_\mathbf{W}\boldsymbol{\beta}_0,$$

the only difference being that the permutations **T** of **X** are stratified (i.e. treatment is permuted within strata). Consequently, as before we have

$$\text{(D.4)}\quad \sqrt{N}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}-\boldsymbol{\beta}_0)\;=\;\left(\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}}{N}\right)^{-1}\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{X}_\mathbf{W}}{N}\mathbf{r}+\left(\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}}{N}\right)^{-1}\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\boldsymbol{\varepsilon}}{\sqrt{N}}$$

where $\mathbf{r}=\sqrt{N}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)$. Let $\mathbf{t}_{\mathbf{W}k}$ and $\mathbf{x}_{\mathbf{W}l}$ denote the k^{th} and l^{th} columns of $\mathbf{T}_\mathbf{W}$ and $\mathbf{X}_\mathbf{W}$, with the i^{th} elements of these vectors given by $t_{ip(k)}w_{iq(k)}$ and $x_{ip(l)}w_{iq(l)}$, where $p(j)$ and $q(j)$ denote the columns of **T** (or **X**) and **W** associated with the j^{th} column of $\mathbf{T}_\mathbf{W}$ (or $\mathbf{X}_\mathbf{W}$). With this notation, we see that the kl^{th} element of $\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}/N$ can be expressed as

(D.5) $$\frac{\mathbf{t}'_{\mathbf{W}k}\mathbf{M}\mathbf{t}_{\mathbf{W}l}}{N} = \frac{\mathbf{t}'_{\mathbf{W}k}[\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{t}_{\mathbf{W}l}}{N} = m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}) - \mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\mathbf{m}(t_{ip(l)}w_{iq(l)}\mathbf{z}_i),$$

$$\text{so } \frac{\mathbf{t}'_{\mathbf{W}k}\mathbf{M}\mathbf{t}_{\mathbf{W}l}}{N} - [m(x_{ip(k)}x_{ip(l)}) - m(x_{ip(k)})m(x_{ip(l)})]m(w_{iq(k)}w_{iq(l)}) =$$

$$\underbrace{[m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}) - m(x_{ip(k)}x_{ip(l)})m(w_{iq(k)}w_{iq(l)})]}_{\xrightarrow{p}0 \text{ (Lemma D2a)}} + m(x_{ip(k)})m(x_{ip(l)})[m(w_{iq(k)}w_{iq(l)}) - \underbrace{\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\mathbf{m}(w_{iq(l)}\mathbf{z}_i)}_{\text{a.s.}=m(w_{iq(k)}w_{iq(l)}) \text{ for } N \text{ sufficiently large (Lemma D1a)}}]$$

$$-\underbrace{[\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) - m(x_{ip(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)]}_{\xrightarrow{p}\mathbf{0}'_K \text{ (Lemma D2a)}}\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}}_{\text{a.s. bounded (Lemma D1c)}}\underbrace{[\mathbf{m}(t_{ip(l)}w_{iq(l)}\mathbf{z}_i) - m(x_{ip(l)})\mathbf{m}(w_{iq(l)}\mathbf{z}_i)]}_{\xrightarrow{p}\mathbf{0}_K \text{ (Lemma D2a)}}$$

$$-\underbrace{m(x_{ip(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}}_{\text{a.s. bounded (Lemma D1c)}}\underbrace{[\mathbf{m}(t_{ip(l)}w_{iq(l)}\mathbf{z}_i) - m(x_{ip(l)})\mathbf{m}(w_{iq(l)}\mathbf{z}_i)]}_{\xrightarrow{p}\mathbf{0}_K \text{ (Lemma D2a)}}$$

$$-\underbrace{[\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) - m(x_{ip(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)]}_{\xrightarrow{p}\mathbf{0}'_K \text{ (Lemma D2a)}}\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}m(x_{ip(l)})\mathbf{m}(w_{iq(l)}\mathbf{z}_i)}_{\text{a.s. bounded (Lemma D1c)}} \xrightarrow{p} 0,$$

where $\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) = (m(t_{ip(k)}w_{iq(k)}z_{i1}), \ldots, m(t_{ip(k)}w_{iq(k)}z_{iK}))$, and in the third line we use the fact that as $\mathbf{m}(w_{iq(k)}, \mathbf{z}'_i) = \mathbf{w}'_{q(k)}\mathbf{Z}/N$, where $\mathbf{w}_{q(k)}$ is the $q(k)^{th}$ column of $\mathbf{W}$ which is included in the covariates $\mathbf{Z}$ (assumption A2), so for all N sufficiently large that $\mathbf{Z}'\mathbf{Z}/N$ is guaranteed to be invertible $\mathbf{w}'_{q(k)}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ is a row vector of zeros with a 1 in the column corresponding to the position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$. Similarly, the kl^{th} element of $\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}/N$ can be expressed as

(D.6) $$\frac{\mathbf{t}'_{\mathbf{W}k}\mathbf{M}\mathbf{x}_{\mathbf{W}l}}{N} = \frac{\mathbf{t}'_{\mathbf{W}k}[\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{x}_{\mathbf{W}l}}{N} = m(t_{ip(k)}w_{iq(k)}w_{iq(l)}x_{ip(l)}) - \mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\mathbf{x}_{\mathbf{W}l}}{N} =$$

$$\underbrace{[m(t_{ip(k)}w_{iq(k)}w_{iq(l)}x_{ip(l)}) - m(x_{ip(k)})m(w_{iq(k)}w_{iq(l)}x_{ip(l)})]}_{\xrightarrow{p}0 \text{ (Lemma D2a)}} - \underbrace{[\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) - m(x_{ip(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)]}_{\xrightarrow{p}\mathbf{0}'_K \text{ (Lemma D2a)}}\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\mathbf{x}_{\mathbf{W}l}}{N}}_{\text{a.s. bounded (Lemma D1c)}}$$

$$+ m(x_{ip(k)})m(w_{iq(k)}w_{iq(l)}x_{ip(l)}) - m(x_{ip(k)})\underbrace{\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\mathbf{x}_{\mathbf{W}l}}{N}}_{\text{a.s.}=m(w_{iq(k)}w_{iq(l)}x_{ip(l)}) \text{ for } N \text{ sufficiently large (Lemma D1a)}} \xrightarrow{p} 0,$$

where in the last line we again use the assumption that $\mathbf{W}$ is included in $\mathbf{Z}$.

Combining these results, we have:

(D.7) $$\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N} - \frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{N} \otimes \frac{\mathbf{W}'\mathbf{W}}{N} \xrightarrow{p} \mathbf{0}_{PQ\,x\,PQ} \;\;\&\;\; \frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}}{N} \xrightarrow{p} \mathbf{0}_{PQ\,x\,PQ}.$$

Finite values of $\mathbf{r} = \sqrt{N}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, multiplied by $\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}/N$, asymptotically have no influence in (D.4). As can be seen, assumption A2 that $\mathbf{W}$ is a part of $\mathbf{Z}$ (or the less plausible alternative that the mean of the x_{ip} are asymptotically zero) is key to this result.

The remaining part of (D.4) is the vector $\mathbf{T}'_{\mathbf{W}}\mathbf{M}\boldsymbol{\varepsilon}/\sqrt{N}$, the k^{th} term of which equals:

$$\text{(D.8)}\quad \frac{\mathbf{t}'_{\mathbf{W}k}\mathbf{M}\boldsymbol{\varepsilon}}{\sqrt{N}} = \sqrt{N}m(t_{ip(k)}w_{iq(k)}\varepsilon_i) - \sqrt{N}\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{N} = \underbrace{\sqrt{N}\begin{bmatrix} m(t_{ip(k)}w_{iq(k)}\varepsilon_i) - \\ m(x_{ip(k)})m(w_{iq(k)}\varepsilon_i)\end{bmatrix}}_{v_k}$$

$$\underbrace{-\sqrt{N}\begin{bmatrix}\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}'_i) - \\ m(x_{ip(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\end{bmatrix}\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{N}}_{\xrightarrow{a.s.}\mathbf{0}_K\text{ (LemmasD1b, D1c)}}}_{\xrightarrow{p}0\text{ (Lemma D2c)}} + \sqrt{N}\left[m(x_{ip(k)})m(w_{iq(k)}\varepsilon_i) - m(x_{ip(k)})\quad \underbrace{\mathbf{m}(w_{iq(k)}\mathbf{z}'_i)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{N}}_{\text{a.s.}=m(w_{iq(k)}\varepsilon_i)\text{ for }N\text{ sufficiently large (Lemma D1a)}}\right].$$

The only term that asymptotically is non-zero is v_k which, as $m(t_{ip(k)}) = m(x_{ip(k)})$, equals the k^{th} element of $(\widetilde{\mathbf{T}} \bullet \widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}})'\mathbf{1}_N / \sqrt{N}$. Applying Lemma D2d we then see that

$$\text{(D.9)}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{-½}\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\boldsymbol{\varepsilon}}{\sqrt{N}} \xrightarrow{d} \mathbf{n}_{PQ},\ \text{where}\ \mathbf{n}_{PQ} \sim \text{N}(\mathbf{0}_{PQ}, \mathbf{I}_{PQ}),$$

$$\text{so ... (D.10)}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{-½}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}'\mathbf{W}}{N}\right)\sqrt{N}(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0) - \boldsymbol{\beta}_0) =$$

$$\underbrace{\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{-½}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}'\mathbf{W}}{N}\right)}_{\text{almost surely bounded positive definite matrices (Lemmas D1a,D1c)}}\underbrace{\left(\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N}\right)^{-1}}_{\xrightarrow{p}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}'\mathbf{W}}{N}\right)^{-1}}\underbrace{\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{X}_{\mathbf{W}}}{N}}_{\xrightarrow{p}\mathbf{0}_{PQxPQ}}\underbrace{\mathbf{r}}_{\text{bounded}} +$$

$$\underbrace{\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{-½}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}'\mathbf{W}}{N}\right)\left(\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N}\right)^{-1}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{½}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}'_{\boldsymbol{\varepsilon}}\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{-½}\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\boldsymbol{\varepsilon}}{\sqrt{N}}}_{\xrightarrow{p}\mathbf{I}_{PQ}\qquad\qquad\qquad\qquad\qquad\qquad\xrightarrow{d}\mathbf{n}_{PQ}} \xrightarrow{d} \mathbf{n}_{PQ}.$$

(b) Probability Limit of the Heteroskedasticity Robust Covariance Estimate

For the heteroskedasticity robust covariance estimate we have

$$\text{(D.11)}\ N\mathbf{V}_r(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0)) = \left(\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N}\right)^{-1}\mathbf{A}\left(\frac{\mathbf{T}'_{\mathbf{W}}\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N}\right)^{-1},\ \text{where}\ \mathbf{A} = \frac{(\mathbf{M}\mathbf{T}_{\mathbf{W}} \bullet \hat{\boldsymbol{\varepsilon}}(\mathbf{T},\boldsymbol{\beta}_0))'(\mathbf{M}\mathbf{T}_{\mathbf{W}} \bullet \hat{\boldsymbol{\varepsilon}}(\mathbf{T},\boldsymbol{\beta}_0))}{N}$$

with kl^{th} term given by

$$\text{(D.12)}\ \mathbf{A}_{kl} = \frac{1}{N}\sum_{i=1}^{N}(t_{ip(k)}w_{iq(k)} - \sum_{a=1}^{K}z_{ia}\hat{\delta}_{ak})(t_{ip(l)}w_{iq(l)} - \sum_{b=1}^{K}z_{ib}\hat{\delta}_{bl})\hat{\varepsilon}_i(\mathbf{T},\boldsymbol{\beta}_0)^2,$$

$$\text{where}\ \hat{\varepsilon}_i(\mathbf{T},\boldsymbol{\beta}_0) = \varepsilon_i - \sum_{c=1}^{K}z_{ic}\hat{\eta}_c + \sum_{d=1}^{PQ}(x_{ip(d)}w_{iq(d)} - \sum_{e=1}^{K}z_{ie}\hat{\tau}_{ed})\frac{r_d}{\sqrt{N}} - \sum_{f=1}^{PQ}(t_{ip(f)}w_{iq(f)} - \sum_{g=1}^{K}z_{ig}\hat{\delta}_{gf})\frac{\hat{r}_f}{\sqrt{N}},$$

with $\mathbf{r} = \sqrt{N}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, $\hat{\mathbf{r}} = \sqrt{N}(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0) - \boldsymbol{\beta}_0)$. $\hat{\boldsymbol{\delta}}_k = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{t}_{\mathbf{W}k}$, $\hat{\boldsymbol{\tau}}_k = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{x}_{\mathbf{W}k}$ and $\hat{\boldsymbol{\eta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}$. From Lemmas D1b and D1c we have $\hat{\boldsymbol{\eta}} \xrightarrow{a.s.} \mathbf{0}_K$ and from D1c know that $\hat{\boldsymbol{\tau}}_k$ is almost surely bounded. The plim of $\hat{\boldsymbol{\delta}}_k$ across the distribution of $\mathbf{T}$ is bounded as

$$\text{(D.13)}\ \hat{\boldsymbol{\delta}}_k - \underbrace{m(x_{ip(k)})\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\mathbf{w}_{q(k)}}{N}}_{\text{a.s. bounded (Lemma D1c)}} = \underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}}_{\text{a.s. bounded (Lemma D1c)}}\underbrace{[\mathbf{m}(t_{ip(k)}w_{iq(k)}\mathbf{z}_i) - m(x_{ip(k)})\mathbf{m}(w_{iq(k)}\mathbf{z}_i)]}_{\xrightarrow{p}\mathbf{0}_K\text{ (Lemma D2a)}} \xrightarrow{p} 0.$$

As for all N sufficiently large $(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{w}_{q(k)}$ is a vector of zeros with a 1 in the row corresponding to the column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$ and $m(x_{ip(k)})$ is known to be bounded by Lemma D1c, $\operatorname{plim}\hat{\delta}_{ak}=0$ unless a is the column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$, in which case $\operatorname{plim}\hat{\delta}_{ak}-m(x_{ip(k)})=0$. The elements of $\mathbf{r}$ are finite and of $\hat{\mathbf{r}}$ are asymptotically normal with bounded variance, so when divided by a positive power of N have a plim of zero.

When the terms in (D.12) are multiplied out, most involve a product with an element of $\mathbf{r}/\sqrt{N}$, $\hat{\mathbf{r}}/\sqrt{N}$, or $\hat{\boldsymbol{\eta}}$ that has a plim of zero, 0 to 4 parameters $\hat{\tau}$ and $\hat{\delta}$ with bounded probability limits, and the mean of the product of the elements of 0 to 4 columns of $\mathbf{T}$ and the elements of 4 columns of $\mathbf{D}$ = $(\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})$ (no more than two of which are ε_i). From Lemma D1c we know that the sample means of the product of the elements of one through four columns of $\mathbf{X}$ or four columns of $\mathbf{D}$ are almost surely bounded. Consequently, in (D.12) every term that involves the product of an element of $\mathbf{r}/\sqrt{N}$, $\hat{\mathbf{r}}/\sqrt{N}$, or $\hat{\boldsymbol{\eta}}$ that has a plim of zero with the mean of the product of four columns of $\mathbf{D}$ with zero, one or two columns of $\mathbf{T}$ has, using Lemma D2a, a probability limit of zero. Every term in (D.12) that involves the product of n = three or four columns of $\mathbf{T}$ with four columns of $\mathbf{D}$ also includes at least n - 2 $\hat{\mathbf{r}}/\sqrt{N}$ terms which can be re-expressed as $(\hat{\mathbf{r}}/N^{½-a})(1/N^{a})$ for some a in (0,½). The $1/N^{a}$ parts satisfy Lemma D2b, while the $\hat{\mathbf{r}}/N^{½-a}$ parts converge in probability to 0. Thus, all such terms also have a plim of 0.

The above only leaves terms in (D.12) that involve the product of two or less columns of $\mathbf{T}$ and do not include an element of $\mathbf{r}/\sqrt{N}$, $\hat{\mathbf{r}}/\sqrt{N}$, or $\hat{\boldsymbol{\eta}}$, namely

$$\text{(D.14)}\ \sum_{i=1}^{N}\frac{t_{ip(k)}w_{iq(k)}t_{ip(l)}w_{iq(l)}\varepsilon_i^2}{N}-\sum_{a=1}^{K}\hat{\delta}_{ak}\sum_{i=1}^{N}\frac{t_{ip(l)}w_{iq(l)}z_{ia}\varepsilon_i^2}{N}-\sum_{b=1}^{K}\hat{\delta}_{bl}\sum_{i=1}^{N}\frac{t_{ip(k)}w_{iq(k)}z_{ib}\varepsilon_i^2}{N}+\sum_{a=1}^{K}\sum_{b=1}^{K}\hat{\delta}_{ak}\hat{\delta}_{bl}\sum_{i=1}^{N}\frac{z_{ia}z_{ib}\varepsilon_i^2}{N}$$

$$=m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}\varepsilon_i^2)-\sum_{a=1}^{K}\hat{\delta}_{ak}m(t_{ip(l)}w_{iq(l)}z_{ia}\varepsilon_i^2)-\sum_{b=1}^{K}\hat{\delta}_{bl}m(t_{ip(k)}w_{iq(k)}z_{ib}\varepsilon_i^2)+\sum_{a=1}^{K}\sum_{b=1}^{K}\hat{\delta}_{ak}\hat{\delta}_{bl}m(z_{ia}z_{ib}\varepsilon_i^2)$$

where $m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}\varepsilon_i^2)-m(x_{ip(k)}x_{ip(l)})m(w_{iq(k)}w_{iq(l)}\varepsilon_i^2)\underbrace{\xrightarrow{p}0}_{\text{Lemma 2a}}$ & $m(t_{ip(l)}w_{iq(l)}z_{ia}\varepsilon_i^2)-m(x_{ip(l)})m(w_{iq(l)}z_{ia}\varepsilon_i^2)\underbrace{\xrightarrow{p}0}_{\text{Lemma 2a}}$,

$$\text{so } \mathbf{A}_{kl}-[m(x_{ip(k)}x_{ip(l)})-m(x_{ip(k)})m(x_{ip(l)})]m(w_{iq(k)}w_{iq(l)}\varepsilon_i^2)\xrightarrow{p}0,$$

where we use the boundedness of means of products of up to four terms (Lemma D1c) and the fact noted above that $\operatorname{plim}\hat{\delta}_{ak}=0$ unless a is the column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$, in which case $\operatorname{plim}\hat{\delta}_{ak}=m(x_{ip_k})$ and $z_{ia}=w_{iq(k)}$. This allows us to state that

$$\text{(D.15)}\ \mathbf{A}-\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}_{\boldsymbol{\varepsilon}}'\mathbf{W}_{\boldsymbol{\varepsilon}}}{N}\xrightarrow{p}\mathbf{0}_{PQxPQ}$$

and consequently for the heteroskedasticity robust covariance estimate we have

$$\text{(D.16)}\ N\mathbf{V}_r(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})-\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}'\mathbf{W}}{N}\right)^{-1}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}_{\boldsymbol{\varepsilon}}'\mathbf{W}_{\boldsymbol{\varepsilon}}}{N}\right)\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}'\mathbf{W}}{N}\right)^{-1}\xrightarrow{p}\mathbf{0}_{PQxPQ},$$

which from (D.10) is seen to be the asymptotic covariance matrix of normally distributed $\sqrt{N}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}-\boldsymbol{\beta}_0)$.

E. Proofs of Lemmas used in Appendix D

This appendix references the Markov Corollary presented in the paper.

Lemma D1a-D1c: In the paper's appendix proofs for Lemma 1 in the paper the statements regarding the full sample in Lemma D1a - D1c were proven without reference to permutations, and simply using the assumptions regarding observation level moments and the determinants of sample average matrices of expectations given in assumptions W1-W4 and A1-A3. Consequently, all of these results continue to hold. Given the assumption in S1 on the determinant of strata level $\Sigma_{i\epsilon s} E(\mathbf{w}_i \mathbf{w}_i' \varepsilon_i^2)/N_s$ and on the moments of $\mathbf{x}_i$ in S2 along with assumption A1, the results for $\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon} s}' \widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon} s} / N_s$ and $\widetilde{\mathbf{X}}_s' \widetilde{\mathbf{X}}_s / N_s$ can be proven using the same techniques as was used for the full sample. The results regarding bounded sample means in D1c were proven in the paper's appendix using the Markov Corollary and the uniform bounds on the observation level moments of the **X** and **D** variables. As we assume that $N \to \infty$ implies $N_s \to \infty$, these results hold at the stratum level as well.

Lemma D1d: For the random vector $\boldsymbol{\mu}_N = \mathbf{Z}'\boldsymbol{\varepsilon}/\sqrt{N}$, we have:

$$\text{(E.1)}\ E\left(\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{\sqrt{N}}\right) = \sum_{i=1}^{N} \frac{E(\mathbf{z}_i \varepsilon_i)}{\sqrt{N}} \overbrace{=}^{\text{W1: } E(\mathbf{z}_{+i}\varepsilon_i)=\mathbf{0}_{K+}} \sum_{i=1}^{N} \frac{\mathbf{0}_K}{\sqrt{N}} = \mathbf{0}_K \ \text{ and } \ E\left(\frac{\mathbf{Z}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{Z}}{N}\right) = \overbrace{\sum_{i=1}^{N} \frac{E(\mathbf{z}_i \varepsilon_i^2 \mathbf{z}_i')}{N}}^{\text{W1: } E(\mathbf{z}_{+i}\varepsilon_i)=\mathbf{0}_{K+} \text{ \& observations independent}}.$$

From assumption W3 and Jensen's Inequality we have $E(|\varepsilon_i^2 z_{+ij}^2 z_{+ik}^2|) < \Delta^{1/(1+\delta)}$, so the covariance matrix of $\boldsymbol{\mu}_N$ is bounded. Consequently, by Chebyshev's Inequality the elements of μ_N are bounded in probability. The proof of Lemma 1c in the paper's appendix showed that the elements of $(\mathbf{Z}'\mathbf{Z}/N)^{-1}$ are almost surely bounded.

Assumption A3 combined with Jensen's Inequality tells us that for n = 1, 2, 3 or 4, $E(|x_{ip}^n|^{1+\theta^*}) < \Delta^{n/4}$. Using Hölder's Inequality, and with $p(1)$ ... $p(n)$ denoting arbitrary columns of **X**, these uniform bounds apply to the product of different column elements:

$$\text{(E.2)}\ E\left(\left|\prod_{o=1}^{n} x_{ip(o)}\right|^{1+\theta^*}\right) \le \sqrt[n]{\prod_{o=1}^{n} E\left(\left|x_{ip(o)}^n\right|^{1+\theta^*}\right)} < \Delta^{n/4}.$$

Consequently, applying the Markov Corollary at both the stratum and full sample level, and making use of Assumption S2 and the assumption that $N \to \infty$ implies $N_s \to \infty$,

$$\text{(E.3)}\ \overbrace{m_s(\prod_{o=1}^{n} x_{ip(o)}) - m_s(E(\prod_{o=1}^{n} x_{ip(o)})) \xrightarrow{a.s.} 0,\ m(\prod_{o=1}^{n} x_{ip(o)}) - m(E(\prod_{o=1}^{n} x_{ip(o)})) \xrightarrow{a.s.} 0,}^{\text{(E.2) and Markov Corollary}}$$

$$\text{\&}\ \overbrace{m_s(E(\prod_{o=1}^{n} x_{ip(o)})) - m(E(\prod_{o=1}^{n} x_{ip(o)})) \xrightarrow{a.s.} 0}^{\text{Assumption S2}} \ \text{ and so } \ m_s(\prod_{o=1}^{n} x_{ip(o)}) - m(\prod_{o=1}^{n} x_{ip(o)}) \xrightarrow{a.s.} 0.$$

We will make use of (E.3) further below as well. In the present context, we have

$$\text{(E.4)}\ \sum_{s=1}^{S} \frac{N_s}{N} m_s(x_{ip}) m_s(w_{iq}\mathbf{z}_i') - m(x_{ip}) m(w_{iq}\mathbf{z}_i') = \sum_{s=1}^{S} \frac{N_s}{N} \overbrace{[m_s(x_{ip}) - m(x_{ip})]}^{\xrightarrow{a.s.} 0} \overbrace{m_s(w_{iq}\mathbf{z}_i')}^{\text{a.s. bounded (Lemma 1c)}} \xrightarrow{a.s.} \mathbf{0}_K',$$

In sum,

$$\text{(E.5)}\quad \underbrace{\left(\sum_{s=1}^{S}\frac{N_s}{N}m_s(x_{ip})m_s(w_{iq}\mathbf{z}_i')-m(x_{ip})m(w_{iq}\mathbf{z}_i')\right)}_{\xrightarrow{a.s.}\mathbf{0}_K'}\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}}_{\text{a.s. bounded}}\underbrace{\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{\sqrt{N}}}_{\substack{\text{bounded in}\\ \text{probability}}},$$

and hence the entire expression converges in probability across the distribution of the data **D** to zero.

Lemma D1e:

We define the random variable τ_{pq} and calculate its expectation:

$$\text{(E.6)}\quad \tau_{pq}=\sqrt{N}\left(\sum_{s=1}^{S}\frac{N_s}{N}m_s(x_{ip})m_s(w_{iq}\varepsilon_i)-m(x_{ip})m(w_{iq}\varepsilon_i)\right)=\sum_{s=1}^{S}\frac{N_s}{\sqrt{N}}\sum_{i\in s}\sum_{j\in s}\frac{x_{ip}w_{jq}\varepsilon_j}{N_s^2}-\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{x_{ip}w_{jq}\varepsilon_j}{N^{3/2}}$$

$$E(\tau_{pq})=\sum_{s=1}^{S}\frac{N_s}{\sqrt{N}}\left(\sum_{i\in s}\frac{E(x_{ip}w_{iq}\varepsilon_i)}{N_s^2}+\sum_{i,j\in s}\frac{E(x_{ip})E(w_{jq}\varepsilon_j)}{N_s^2}\right)-\sum_{i=1}^{N}\frac{E(x_{ip}w_{iq}\varepsilon_i)}{N^{3/2}}-\sum_{i,j=1}^{N}\frac{E(x_{ip})E(w_{jq}\varepsilon_j)}{N^{3/2}}=0,$$

where, as elsewhere in the paper, we use subscripted *i*, *j* to denote the summation across both indices, excluding ties between them, and we recall assumption W1's statement that the observations are independent with $E(\mathbf{z}_{+i}\varepsilon_i)=\mathbf{0}_{K+}$, so $E(x_{ip}w_{iq}\varepsilon_i)$ and $E(w_{iq}\varepsilon_i)$ equal 0 as both $\mathbf{X_W}$ and $\mathbf{W}$ are part of $\mathbf{Z}_+$, and the fact that expectations of powers less than 4 of x_{ip} are bounded by A3 and Jensen's Inequality as for n = 1, 2, 3 or 4, $E(|x_{ip}^n|)<\Delta^{n/4(1+\theta^*)}$. The variance of τ_{pq} is given by:

$$\text{(E.7)}\quad E(\tau_{pq}^2)=E\left(\begin{array}{c}\displaystyle\sum_{s=1}^{S}\sum_{t=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in t}\sum_{j\in t}\frac{x_{gp}w_{hq}\varepsilon_h\,x_{ip}w_{jq}\varepsilon_j}{NN_sN_t}\\ \displaystyle+\sum_{g=1}^{N}\sum_{h=1}^{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{x_{gp}w_{hq}\varepsilon_h\,x_{ip}w_{jq}\varepsilon_j}{N^3}-2\sum_{g=1}^{N}\sum_{h=1}^{N}\sum_{s=1}^{S}\sum_{i\in s}\sum_{j\in s}\frac{x_{gp}w_{hq}\varepsilon_h\,x_{ip}w_{jq}\varepsilon_j}{N^2N_s}\end{array}\right)$$

$$=\sum_{s=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{NN_s^2}+\sum_{g=1}^{N}\sum_{h=1}^{N}\sum_{i=1}^{N}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^3}-2\sum_{g=1}^{N}\sum_{s=1}^{S}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}.$$

In the last line we simplify by making use of the fact that observations are independent and, as already noted, $E(x_{ip}w_{iq}\varepsilon_i)$ and $E(w_{iq}\varepsilon_i)$ equal 0 and expectations of first and second powers of x_{ip} are bounded, as is $E(x_{ip}^2w_{iq}\varepsilon_i)$, as by A3, W3a and Jensen's Inequality

$$\text{(E.8)}\quad |E(x_{ip}^2w_{iq}\varepsilon_i)|\le E(|x_{ip}^2w_{iq}\varepsilon_i|)\le\overbrace{\sqrt{E(x_{ip}^4)E(w_{iq}^2\varepsilon_i^2)}}^{\text{Holder's Inequality}}<\sqrt{\Delta^{1/(1+\theta^*)}\Delta^{1/(1+\delta)}}<\infty.$$

Consequently, all non-zero expectations in the first line of (E.7) must contain an ε_i^2, which rules out all cases where the *h* and *j* indices in the first line do not tie, leading to the expression in the second line.

In calculating the expectations in the second line of (E.7), it will once again be necessary to consider cases where the various indices do or don't tie. Since observational level expectations are bounded and cases that involve one or more ties are divided by powers of *N*, the latter vanish asymptotically, leaving only expectations in which there are no ties. Expectations that involve no ties will cancel. We can conclude that the variance of τ_{pq} converges to zero, and so it converges in mean square and hence in probability to the expectation given in (E.6), i.e. 0. As these statements are difficult to confirm without experience with such problems, the details are worked out for the reader below.

We begin by calculating the expectations in the second line of (E.7), paying careful attention to potential ties between indices:

$$\text{(E.9a)}\quad \sum_{g=1}^{N}\sum_{h=1}^{N}\sum_{i=1}^{N}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^3}=\left(\begin{array}{c}\displaystyle\sum_{i=1}^{N}\frac{E(x_{ip}^2w_{iq}^2\varepsilon_i^2)}{N^3}+2\sum_{h,i=1}^{N}\frac{E(x_{hp})E(x_{ip}w_{iq}^2\varepsilon_i^2)}{N^3}\\ \displaystyle+\sum_{h,i=1}^{N}\frac{E(x_{hp}^2)E(w_{iq}^2\varepsilon_i^2)}{N^3}+\sum_{g,h,i=1}^{N}\frac{E(x_{gp})E(x_{hp})E(w_{iq}^2\varepsilon_i^2)}{N^3}\end{array}\right)$$

$$\begin{aligned}&=\frac{m[E(x_{ip}^2w_{iq}^2\varepsilon_i^2)]}{N^2}+2\frac{m[E(x_{hp})]m[E(x_{ip}w_{iq}^2\varepsilon_i^2)]}{N}-2\frac{m[E(x_{ip})E(x_{ip}w_{iq}^2\varepsilon_i^2)]}{N^2}\\&+\frac{m[E(x_{hp}^2)]m[E(w_{iq}^2\varepsilon_i^2)]}{N}-\frac{m[E(x_{ip}^2)E(w_{iq}^2\varepsilon_i^2)]}{N^2}+m[E(x_{gp})]^2m[E(w_{iq}^2\varepsilon_i^2)]\\&-2\frac{m[E(x_{hp})]m[E(x_{gp})E(w_{gq}^2\varepsilon_g^2)]}{N}-\frac{m[E(x_{gp})^2]m[E(w_{iq}^2\varepsilon_i^2)]}{N}+2\frac{m[E(x_{ip})^2E(w_{iq}^2\varepsilon_i^2)]}{N^2}\\&=m[E(x_{gp})]^2m[E(w_{iq}^2\varepsilon_i^2)]+o(1).\end{aligned}$$

$$\text{(E.9b)}\quad \sum_{s=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{NN_s^2}=\sum_{s=1}^{S}\left(\begin{array}{c}\displaystyle\sum_{i\in s}\frac{E(x_{ip}^2w_{iq}^2\varepsilon_i^2)}{NN_s^2}+2\sum_{h,i\in s}\frac{E(x_{hp})E(x_{ip}w_{iq}^2\varepsilon_i^2)}{NN_s^2}\\ \displaystyle+\sum_{h,i\in s}\frac{E(x_{hp}^2)E(w_{iq}^2\varepsilon_i^2)}{NN_s^2}+\sum_{g,h,i\in s}\frac{E(x_{gp})E(x_{hp})E(w_{iq}^2\varepsilon_i^2)}{NN_s^2}\end{array}\right)$$

$$\begin{aligned}&=\sum_{s=1}^{S}\frac{\left(\begin{array}{c}m_s[E(x_{ip}^2w_{iq}^2\varepsilon_i^2)]+2N_sm_s[E(x_{hp})]m_s[E(x_{ip}w_{iq}^2\varepsilon_i^2)]\\-2m_s[E(x_{ip})E(x_{ip}w_{iq}^2\varepsilon_i^2)]+N_sm_s[E(x_{hp}^2)]m_s[E(w_{iq}^2\varepsilon_i^2)]\\-m_s[E(x_{ip}^2)E(w_{iq}^2\varepsilon_i^2)]+N_s^2m_s[E(x_{gp})]^2m_s[E(w_{iq}^2\varepsilon_i^2)]\\-2N_sm_s[E(x_{hp})]m_s[E(x_{gp})E(w_{gq}^2\varepsilon_g^2)]\\-N_sm_s[E(x_{gp})^2]m_s[E(w_{iq}^2\varepsilon_i^2)]+2m_s[E(x_{ip})^2E(w_{iq}^2\varepsilon_i^2)]\end{array}\right)}{NN_s}\\&=\sum_{s=1}^{S}\frac{N_s}{N}m_s[E(x_{gp})]^2m_s[E(w_{iq}^2\varepsilon_i^2)]+o(1).\end{aligned}$$

$$\text{(E.9c)}\quad \sum_{g=1}^{N}\sum_{s=1}^{S}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}=\sum_{s=1}^{S}\sum_{t=1}^{S}\sum_{g\in t}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}$$

$$=\sum_{s=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}+\sum_{s,t=1}^{S}\sum_{g\in t}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp})E(x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}$$

$$=\sum_{s=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}+\sum_{s,t=1}^{S}\sum_{g\in t}\sum_{h\in s}\frac{E(x_{gp})E(x_{hp}w_{hq}^2\varepsilon_h^2)}{N^2N_s}+\sum_{s,t=1}^{S}\sum_{g\in t}\sum_{h,i\in s}\frac{E(x_{gp})E(x_{hp})E(w_{iq}^2\varepsilon_i^2)}{N^2N_s}$$

$$=\sum_{s=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}+\sum_{s,t=1}^{S}\left(\begin{array}{c}\displaystyle\sum_{g\in t}\sum_{h\in s}\frac{E(x_{gp})E(x_{hp}w_{hq}^2\varepsilon_h^2)}{N^2N_s}+\sum_{g\in t}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp})E(x_{hp})E(w_{iq}^2\varepsilon_i^2)}{N^2N_s}\\ \displaystyle-\sum_{g\in t}\sum_{h\in s}\frac{E(x_{gp})E(x_{hp})E(w_{hq}^2\varepsilon_h^2)}{N^2N_s}\end{array}\right)$$

$$=\sum_{s=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}+\sum_{s,t=1}^{S}\frac{N_t}{N^2}\left(\begin{array}{c}m_t[E(x_{gp})]m_s[E(x_{hp}w_{hq}^2\varepsilon_h^2)]+N_sm_t[E(x_{gp})]m_s[E(x_{hp})]m_s[E(w_{iq}^2\varepsilon_i^2)]\\ -m_t[E(x_{gp})]m_s[E(x_{hp})E(w_{hq}^2\varepsilon_h^2)]\end{array}\right)$$

$$=\sum_{s=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}+\sum_{s=1}^{S}\sum_{t=1}^{S}\frac{N_t}{N^2}\left(\begin{array}{c}m_t[E(x_{gp})]m_s[E(x_{hp}w_{hq}^2\varepsilon_h^2)]\\ +N_sm_t[E(x_{gp})]m_s[E(x_{hp})]m_s[E(w_{iq}^2\varepsilon_i^2)]\\ -m_t[E(x_{gp})]m_s[E(x_{hp})E(w_{hq}^2\varepsilon_h^2)]\end{array}\right)$$

$$-\sum_{s=1}^{S}\frac{N_s}{N^2}\left(\begin{array}{c}m_s[E(x_{gp})]m_s[E(x_{hp}w_{hq}^2\varepsilon_h^2)]+N_sm_s[E(x_{gp})]^2m_s[E(w_{iq}^2\varepsilon_i^2)]\\ -m_s[E(x_{gp})]m_s[E(x_{hp})E(w_{hq}^2\varepsilon_h^2)]\end{array}\right)$$

$$=\sum_{s=1}^{S}\sum_{g\in s}\sum_{h\in s}\sum_{i\in s}\frac{E(x_{gp}x_{hp}w_{iq}^2\varepsilon_i^2)}{N^2N_s}+m[E(x_{gp})]\sum_{s=1}^{S}\frac{1}{N}\left(\begin{array}{c}m_s[E(x_{hp}w_{hq}^2\varepsilon_h^2)]+N_sm_s[E(x_{hp})]m_s[E(w_{iq}^2\varepsilon_i^2)]\\ -m_s[E(x_{hp})E(w_{hq}^2\varepsilon_h^2)]\end{array}\right)$$

$$-\sum_{s=1}^{S}\frac{N_s}{N^2}\left(\begin{array}{c}m_s[E(x_{gp})]m_s[E(x_{hp}w_{hq}^2\varepsilon_h^2)]+N_sm_s[E(x_{gp})]^2m_s[E(w_{iq}^2\varepsilon_i^2)]\\ -m_s[E(x_{gp})]m_s[E(x_{hp})E(w_{hq}^2\varepsilon_h^2)]\end{array}\right)$$

$$=\sum_{s=1}^{S}\frac{N_s^2}{N^2}m_s[E(x_{gp})]^2m_s[E(w_{iq}^2\varepsilon_i^2)]+m[E(x_{gp})]\sum_{s=1}^{S}\frac{N_s}{N}m_s[E(x_{hp})]m_s[E(w_{iq}^2\varepsilon_i^2)]$$

$$-\sum_{s=1}^{S}\frac{N_s^2}{N^2}m_s[E(x_{gp})]^2m_s[E(w_{iq}^2\varepsilon_i^2)]+o(1)$$

$$=m[E(x_{gp})]\sum_{s=1}^{S}\frac{N_s}{N}m_s[E(x_{hp})]m_s[E(w_{iq}^2\varepsilon_i^2)]+o(1)$$

where, (i) as elsewhere, subscripted *s*,*t* or *g*,*h*,*i* or *h*,*i* denote summations across the indices excluding ties between them; (ii) we simplify at the end of each expression using the fact that since all of the expectations are uniformly bounded those that are divided by any positive power of N or N_s are $o(1)$; and (iii) as the first term following the equal sign from the second line down in (E.9c) is identical (subject to a modification of the denominator) to (E.9b), we simply substitute using the results from that earlier calculation towards the end of (E.9c).

Combining the results of (E.9a) - (E.9c), we have:

$$(E.10)\quad E(\tau_{pq}^2) = m[E(x_{gp})]^2 m[E(w_{iq}^2\varepsilon_i^2)] + \sum_{s=1}^{S}\frac{N_s}{N} m_s[E(x_{gp})]^2 m_s[E(w_{iq}^2\varepsilon_i^2)]$$

$$- 2m[E(x_{gp})]\sum_{s=1}^{S}\frac{N_s}{N} m_s[E(x_{hp})]m_s[E(w_{iq}^2\varepsilon_i^2)] + o(1)$$

$$= \sum_{s=1}^{S}\frac{N_s}{N}\left(\underbrace{m_s[E(x_{gp})] - m[E(x_{gp})]}_{\xrightarrow{a.s.} 0 \text{ (assumption S2)}}\right)^2 \underbrace{m_s[E(w_{iq}^2\varepsilon_i^2)]}_{\text{a.s. bounded (W3)}} + o(1) \xrightarrow{a.s.} 0.$$

Consequently, τ_{pq} converges in mean square and in probability to zero, as claimed in the lemma.

Lemma D2a: The proof of Lemma 1d in the paper's appendix only involves the observation level moment conditions in W1 - W4 & A1 - A3. These lemmas now hold at the stratum level, with stratum means $m_s()$ and $N_s \to \infty$ taking the place of sample means $m()$ and $N \to \infty$. So:

$$(E.11)\quad m_s(t_{ip}d_{ij}d_{ik}) - m_s(x_{ip})m_s(d_{ij}d_{ik}) \xrightarrow{p} 0 \;\&\; m_s(t_{ip}t_{iq}d_{ij}d_{ik}) - m_s(x_{ip}x_{iq})m_s(d_{ij}d_{ik}) \xrightarrow{p} 0.$$

Consequently, using (E.3) earlier and the fact that N_s/N is bounded between 0 and 1, for n = 1 or 2

$$(E.12)\quad m((\prod_{o=1}^{n} t_{ip(o)})d_{ij}d_{ik}) - m(\prod_{o=1}^{n} x_{ip(o)})m(d_{ij}d_{ik}) =$$

$$\sum_{s=1}^{S}\frac{N_s}{N}\underbrace{\left[m_s((\prod_{o=1}^{n})t_{ip(o)}d_{ij}d_{ik}) - m_s(\prod_{o=1}^{n} x_{ip(o)})m_s(d_{ij}d_{ik})\right]}_{\xrightarrow{p} 0 \text{ (E.11)}} + \sum_{s=1}^{S}\frac{N_s}{N}\underbrace{\left[m_s(\prod_{o=1}^{n} x_{ip(o)}) - m(\prod_{o=1}^{n} x_{ip(o)})\right]}_{\xrightarrow{a.s.} 0 \text{ (E.3)}} \underbrace{m_s(d_{ij}d_{ik})}_{\text{a.s. bounded (Lemma D1c)}} \xrightarrow{p} 0$$

Lemma D2b: Again, the proof of Lemma 2 in the paper's appendix only involves observation level moment conditions, and hence now holds at the stratum level as well. As $N_s \to \infty$:

$$(E.13)\quad m_s(N_s^{-a\max(n-2,0)}(\prod_{o=1}^{n} t_{ip(o)})d_{ij}d_{ik}d_{il}d_{im}) - m_s(N_s^{-a\max(n-2,0)}\prod_{o=1}^{n} x_{ip(o)})m_s(d_{ij}d_{ik}d_{il}d_{im}) \xrightarrow{p} 0.$$

Using the fact that N_s/N is bounded between 0 and 1, and that sample and stratum means are almost surely bounded (Lemma D1c)

$$(E.14)\quad m(N^{-a\max(n-2,0)}(\prod_{o=1}^{n} t_{ip(o)})d_{ij}d_{ik}d_{il}d_{im}) - m(N^{-a\max(n-2,0)}\prod_{o=1}^{n} x_{ip(o)})m(d_{ij}d_{ik}d_{il}d_{im}) =$$

$$\sum_{s=1}^{S}\frac{N_s^{1+a\max(n-2,0)}}{N^{1+a\max(n-2,0)}}\underbrace{\left[m_s(N_s^{-a\max(n-2,0)}(\prod_{o=1}^{n} t_{ip(o)})d_{ij}d_{ik}d_{il}d_{im}) - m_s(N_s^{-a\max(n-2,0)}\prod_{o=1}^{n} x_{ip(o)})m_s(d_{ij}d_{ik}d_{il}d_{im})\right]}_{\xrightarrow{p} 0 \text{ (E.13)}}$$

$$+ \sum_{s=1}^{S}\frac{N_s^{1+a\max(n-2,0)}}{N^{1+a\max(n-2,0)}} N_s^{-a\max(n-2,0)}\underbrace{\left[m_s(\prod_{o=1}^{n} x_{ip(o)}) - m(\prod_{o=1}^{n} x_{ip(o)})\right]}_{\xrightarrow{a.s.} 0 \text{ (E.3)}} \underbrace{m_s(d_{ij}d_{ik}d_{il}d_{im})}_{\text{a.s. bounded (Lemma D1c)}} \xrightarrow{p} 0.$$

Lemma D2c: From Lemma 1d in the paper, which now holds within strata, if

$$(E.15)\quad \text{if } c_{N_s} \xrightarrow{a.s.} 0 \text{ then } \sqrt{N_s}[\,m_s(t_{ip}d_{ij}d_{ik}) - m_s(x_{ip})m_s(d_{ij}d_{ik})]c_{N_s} \xrightarrow{p} 0.$$

For any c_N, we can define $c_{Ns} = c_N$, based upon the value of N_s corresponding to N. Since $N \to \infty \Leftrightarrow N_s \to \infty$ for all $s = 1..S$, if $c_N \to 0$ as $N \to \infty$, we can also comfortably say $c_{Ns} \to 0$ as $N_s \to \infty$. The elements of $(\mathbf{Z}'\mathbf{Z}/N)^{-1}(\mathbf{Z}'\boldsymbol{\varepsilon}/N)$, which from Lemma D1b and D1c are known to converge almost surely to 0, are one such c_N. We now note that the expression in the Lemma can be decomposed into:

$$\text{(E.16)}\ \sqrt{N}[\mathbf{m}(t_{ip}w_{iq}\mathbf{z}_i') - m(x_{ip})\mathbf{m}(w_{iq}\mathbf{z}_i')]\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{N} =$$

$$\underbrace{\sum_{s=1}^{S}\sqrt{\frac{N_s}{N}}\left(\sqrt{N_s}[\,\mathbf{m}_s(t_{ip}w_{iq}\mathbf{z}') - m_s(x_{ip})\mathbf{m}_s(w_{iq}\mathbf{z}_i')]\underbrace{\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{N}}_{\text{elements are } c_{N_s}}\right)}_{\xrightarrow{p} 0 \text{ by (E.15)}}$$

$$+\underbrace{\sqrt{N}\left(\sum_{s=1}^{S}\frac{N_s}{N}m_s(x_{ip})\mathbf{m}_s(w_{iq}\mathbf{z}_i') - m(x_{ip})\mathbf{m}(w_{iq}\mathbf{z}_i')\right)\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{N}}_{\xrightarrow{p_{\mathbf{D}}} 0 \text{ by Lemma D1d}},$$

Consequently, in probability across the data sequences **D** the expression in D2c converges in probability (across the permutations **T** of **X**) to 0, as stated in the Lemma.

Lemma D2d: We use superscripted ˜ to denote either sample demeaned ($\widetilde{\mathbf{T}}$) or strata demeaned ($\widetilde{\mathbf{T}}_s$) variables, with the presence of the subscript s indicating the intent. Since the moment conditions in W1-W4 & A1-A3 apply to all observations, from the proof of Lemma 1e in the paper's appendix we know that across the stratified permutations $\mathbf{T}_s$ of $\mathbf{X}_s$, as $N_s \to \infty$

$$\text{(E.17)}\ \left(\frac{\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s}{N_s}\otimes\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}}{N_s}\right)^{-½}\frac{(\widetilde{\mathbf{T}}_s \bullet \widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s})'\mathbf{1}_{N_s}}{\sqrt{N_s}} \xrightarrow{d} \mathbf{n}_{PQs},\ \text{where}\ \mathbf{n}_{PQs} \sim \mathrm{N}(\mathbf{0}_{PQ}, \mathbf{I}_{PQ}),$$

provided the matrices $\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}/N_s$ and $\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s/N_s$, following their counterparts for the entire sample, are almost surely positive definite with determinant greater than some γ for all N_s sufficiently large (which is true by Lemma D1a). The $\mathbf{n}_{PQs}$ are clearly independent across strata, as the observations and permutations of each strata are independent of the others. The k^{th} element of $\mathbf{v} = (\widetilde{\mathbf{T}} \bullet \widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}})'\mathbf{1}_N/\sqrt{N}$ equals:

$$\text{(E.18)}\ v_k = \sum_{i=1}^{N}\frac{(t_{ip(k)} - m(t_{ip(k)}))(w_{iq(k)}\varepsilon_i - m(w_{iq(k)}\varepsilon_i))}{\sqrt{N}} = \sum_{s=1}^{S}\sqrt{\frac{N_s}{N}}\sum_{i\in s}\frac{t_{ip(k)}w_{iq(k)}\varepsilon_i}{\sqrt{N_s}} - \sqrt{N}m(x_{ip(k)})m(w_{iq(k)}\varepsilon_i) =$$

$$\sum_{s=1}^{S}\sqrt{\frac{N_s}{N}}\sum_{i\in s}\frac{[t_{ip(k)} - m_s(t_{ip(k)})][w_{iq(k)}\varepsilon_i - m_s(w_{iq(k)}\varepsilon_i)]}{\sqrt{N_s}} + \underbrace{\sqrt{N}\left(\sum_{s=1}^{S}\frac{N_s}{N}m_s(x_{ip(k)})m_s(w_{iq(k)}\varepsilon_i) - m(x_{ip(k)})m(w_{iq(k)}\varepsilon_i)\right)}_{\xrightarrow{p_{\mathbf{D}}} 0 \text{ by Lemma D1e}}.$$

In addition,

$$\text{(E.19)}\ \left(\sum_{s=1}^{S}\frac{N_s}{N}\frac{\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s}{N_s}\otimes\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}}{N_s}\right) - \left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right) = \sum_{s=1}^{S}\frac{N_s}{N}\underbrace{\left[\frac{\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s}{N_s} - \frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\right]}_{\xrightarrow{a.s.} \mathbf{0}_{PxP} \text{ (E.3)}}\otimes\underbrace{\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}}{N_s}}_{a.s.\ \text{bounded (Lemma D1c)}} \xrightarrow{a.s.} \mathbf{0}_{PQxPQ}.$$

Consequently, the covariance matrix of S independent $\mathbf{n}_{PQs}$ random variables each multiplied by $\sqrt{N_s/N}(\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s/N_s \otimes \widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}/N_s)^{½}$ converges almost surely to $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N \otimes \widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}/N$, and using (E.18)

$$\text{(E.20)}\quad \left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{-½}\frac{(\widetilde{\mathbf{T}}\bullet\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}})'\mathbf{1}_N}{\sqrt{N}} =$$

$$\underbrace{\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{-½}}_{\text{a.s. bounded (Lemma D1c)}}\sum_{s=1}^{S}\sqrt{\frac{N_s}{N}}\underbrace{\left(\frac{\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s}{N_s}\otimes\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}}{N_s}\right)^{½}}_{\text{a.s. bounded (Lemma D1c)}}\underbrace{\left(\frac{\widetilde{\mathbf{X}}_s'\widetilde{\mathbf{X}}_s}{N_s}\otimes\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s}}{N_s}\right)^{-½}\frac{(\widetilde{\mathbf{T}}_s\bullet\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}s})'\mathbf{1}_{N_s}}{\sqrt{N_s}}}_{\xrightarrow{d}\mathbf{n}_{PQs}}$$

$$+\underbrace{\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}'\widetilde{\mathbf{W}}_{\boldsymbol{\varepsilon}}}{N}\right)^{-½}}_{\text{a.s. bounded (Lemma D1c)}}\underbrace{\sum_{s=1}^{S}\sqrt{N}\frac{N_s}{N}[\mathbf{m}_s(\mathbf{x}_i')\bullet\mathbf{m}_s(\mathbf{w}_i'\varepsilon_i)]'-[\mathbf{m}(\mathbf{x}_i')\bullet\mathbf{m}(\mathbf{w}_i'\varepsilon_i)]'}_{\xrightarrow{p}\mathbf{0}_{PQ}\text{ (Lemma D1e)}}\xrightarrow{d}\mathbf{n}_{PQ},$$

as stated in the Lemma.

F. Papers used in Section V's Analysis of a Practical Sample & Alternative Figures Retaining Regressions and Papers Otherwise Dropped on the Basis of Growing Number of Strata or Unbalanced Stratification

Given below are the 39 papers whose OLS regressions are analyzed in Section V & in this appendix. The acronym at the beginning of each reference is the code used to identify the paper in the public use do-files. As noted in the text, I remove papers in the 53 paper Young (2019) sample without OLS regressions or where treatment is calculated from sample characteristics or applied using multiple cross-cutting criteria in a fashion that does not allow for counterfactual permutation. These papers (identified by the acronym used in Young 2019) are:

(i) No OLS regressions: CC1, CC2, CHKL, CILS, ER, FJP, LL, S, VDR.
(ii) Treatment does not allow counterfactual permutation: D, DR2, FG, GKN, MMW.

Some OLS regressions in ABHOT, DDK and DKR are removed on the basis of (ii) as well. In the figures in the text, I also remove papers where procedures are such that the number of strata grows with the sample size (AL & MMW2) and regressions with stratified treatment that is not, at least in principle, asymptotically balanced for asymptotically non-negligible strata (all of the regressions in GRS and KMP, some in BBLP, CGTTTV, DKR & FL). The results presented in Figures I-III & V in the paper are based upon the remaining regressions in 35 papers. Figures F1-F4 below include all regressions dropped for stratification issues and as can be seen are virtually identical to Figures I-III & V given in the paper. Figures F1 & F2 are based upon 3213 coefficients in 1066 regressions in 39 papers. Figure F3 is based upon 1730 coefficients in 467 regressions in 28 papers where the other-treatment-stratification permutation distribution for coefficients in multi-treatment equations is not degenerate. No regressions are dropped because of stratification issues in Figure IV of the paper. Figure F4 duplicates Figure V in the paper. With the addition of regressions dropped from Figure V for stratification reasons, the sample now consists of 2712 coefficients in 565 regressions with more than one treatment effect in 32 papers. This is the only case where there is any discernible difference with results given in the paper, showing more cases where low p-values and low leverage regressions have big increases in their p-values in the unconstrained max (panels a & c, compare with Figure V in the paper).

List of papers

(AFGH) Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470–49.

(AKL) Aker, Jenny C., Christopher Ksoll, and Travis J. Lybbert. 2012. "Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger." *American Economic Journal: Applied Economics* 4 (4): 94–120.

(ABHOT) Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken, and Julia Tobias. 2012. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." *American Economic Review* 102 (4): 1206–1240.

(ALO) Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics* 1 (1): 136–163.

(AL) Angrist, Joshua, and Victor Lavy. 2009. "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial." *American Economic Review* 99 (4): 1384–1414.

(A) Ashraf, Nava. 2009. "Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines." *American Economic Review* 99 (4): 1245–1277.

(ABS) Ashraf, Nava, James Berry, and Jesse M. Shapiro. 2010. "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia." *American Economic Review* 100 (5): 2383–2413.

(BBLP) Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L. Linden, and Francisco Perez-Calle. 2011. "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia." *American Economic Journal: Applied Economics* 3 (2): 167–195

(BM) Beaman, Lori and Jeremy Magruder. 2012. "Who Gets the Job Referral? Evidence from a Social Networks Experiment." *American Economic Review* 102 (7): 3574–3593.

(BL) Burde, Dana and Leigh L. Linden. 2013. "Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools." *American Economic Journal: Applied Economics* 5 (3): 27–40.

(CCF) Cai, Hongbin, Yuyu Chen, and Hanming Fang. 2009. "Observational Learning: Evidence from a Randomized Natural Field Experiment." *American Economic Review* 99 (3): 864–882.

(CMS) Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm. 2010. "Tournaments and Office Politics: Evidence from a Real Effort Experiment." *American Economic Review* 100 (1): 504–517.

(CL) Chen, Yan and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review* 99 (1): 431–457.

(CGTTTV) Cole, Shawn, Xavier Giné, Jeremy Tobacman, Petia Topalova, Robert Townsend, and James Vickery. 2013. "Barriers to Household Risk Management: Evidence from India." *American Economic Journal: Applied Economics* 5 (1): 104–135.

(DDK) Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (5): 1739–1774.

(DKR) Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101 (6): 2350–2390.

(DHR) Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–1278.

(DR) Dupas, Pascaline and Jonathan Robinson. 2013. "Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya." *American Economic Journal: Applied Economics* 5 (1): 163–192.

(EGN) Erkal, Nisvan, Lata Gangadharan, and Nikos Nikiforakis. 2011. "Relative Earnings and Giving in a Real-Effort Experiment." *American Economic Review* 101 (7): 3330–3348.

(FPPR) Field, Erica, Rohini Pande, John Papp, and Natalia Rigol. 2013. "Does the Classic Microfinance Model Discourage Entrepreneurship Among the Poor? Experimental Evidence from India." *American Economic Review* 103 (6): 2196–2226.

(FL) Fong, Christina M. and Erzo F. P. Luttmer. 2009. "What Determines Giving to Hurricane Katrina Victims? Experimental Evidence on Racial Group Loyalty." *American Economic Journal: Applied Economics* 1 (2): 64–87.

(GJKM) Giné, Xavier, Pamela Jakiela, Dean Karlan, and Jonathan Morduch. 2010. "Microfinance Games." *American Economic Journal: Applied Economics* 2 (3): 60-95.

(GKB) Gerber, Alan S., Dean Karlan, and Daniel Bergan. 2009. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1 (2): 35–52.

(GMR) Gertler, Paul J., Sebastian W. Martinez, and Marta Rubio-Codina. 2012. "Investing Cash Transfers to Raise Long-Term Living Standards." *American Economic Journal: Applied Economics* 4 (1): 164–192.

(GGY) Giné, Xavier, Jessica Goldberg, and Dean Yang. 2012. "Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi." *American Economic Review* 102 (6): 2923–2954.

(GRS) Galiani, Sebastian, Martín A. Rossi, and Ernesto Schargrodsky. 2011. "Conscription and Crime: Evidence from the Argentine Draft Lottery." *American Economic Journal: Applied Economics* 3 (2): 119–136.

(HS) Heffetz, Ori and Moses Shayo. 2009. "How Large Are Non-Budget-Constraint Effects of Prices on Demand?" *American Economic Journal: Applied Economics* 1 (4): 170–199.

(IZ) Ifcher, John and Homa Zarghamee. 2011. "Happiness and Time Preference: The Effect of Positive Affect in a Random-Assignment Experiment." *American Economic Review* 101 (7): 3109–3129.

(KL) Karlan, Dean and John A. List. 2007. "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment." *American Economic Review* 97 (5): 1774-1793.

(KMP) Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. 2012. "The Currency of Reciprocity: Gift Exchange in the Workplace." *American Economic Review* 102 (4): 1644–1662.

(KN) Kosfeld, Michael and Susanne Neckermann. 2011. "Getting More Work for Nothing? Symbolic Awards and Worker Performance." *American Economic Journal: Microeconomics* 3 (3): 86–99.

(LLLPR) Landry, Craig E., Andreas Lange, John A. List, Michael K. Price, and Nicholas G. Rupp. "Is a Donor in Hand Better than Two in the Bush? Evidence from a Natural Field Experiment." *American Economic Review* 100 (3): 958–983.

(LMW) Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber. 2012. "Sorting in Experiments with Application to Social Preferences." *American Economic Journal: Applied Economics* 4 (1): 136–163.

(MSV) Macours, Karen, Norbert Schady, and Renos Vakis. 2012. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." *American Economic Journal: Applied Economics* 4 (2): 247–273.

(MMW2) de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2013. "The Demand for, and Consequences of, Formalization among Informal Firms in Sri Lanka." *American Economic Journal: Applied Economics* 5 (2): 122–150.

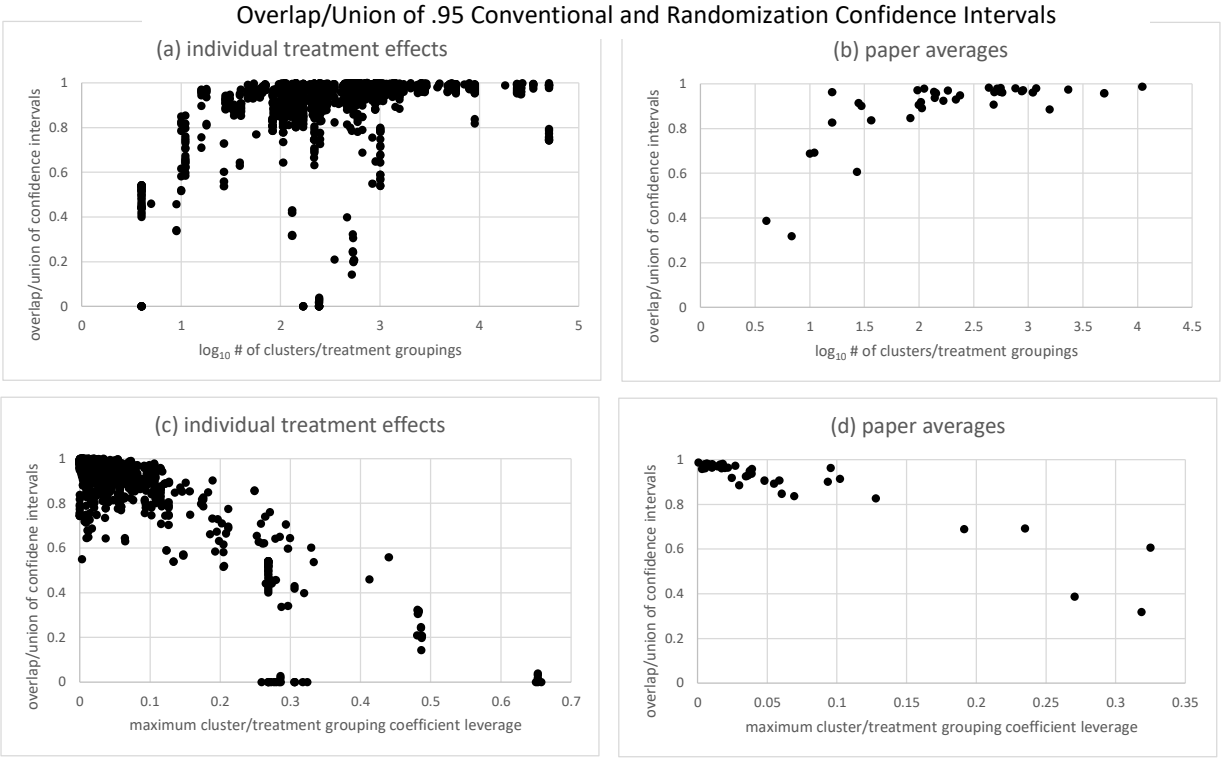
(OT) Oster, Emily and Rebecca Thornton. 2011. "Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation." *American Economic Journal: Applied Economics* 3 (1): 91–100.

(R) Robinson, Jonathan. 2012. "Limited Insurance within the Household: Evidence from a Field Experiment in Kenya." *American Economic Journal: Applied Economics* 4 (4): 140–164.

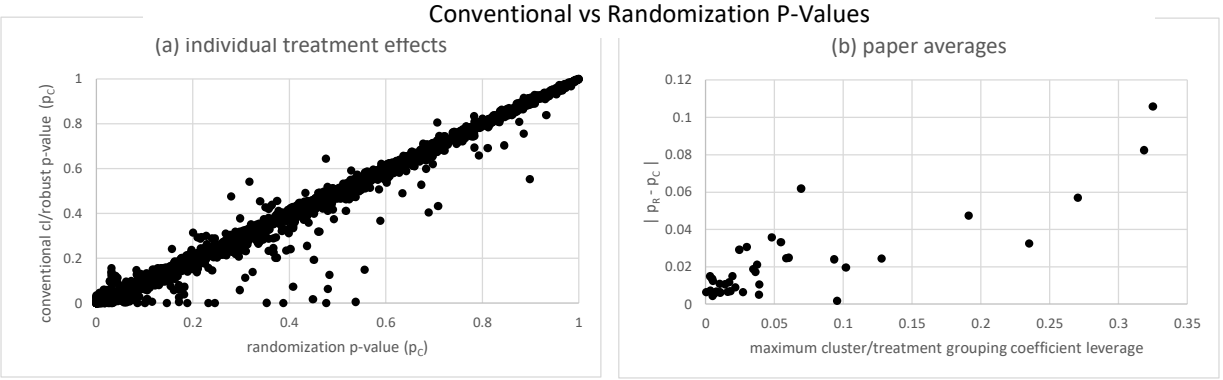
(T) Thornton, Rebecca L. 2008. "The Demand for, and Impact of, Learning HIV Status." *American Economic Review* 98 (5): 1829–1863.

(WDL) Wisdom, Jessica, Julie S. Downs, and George Loewenstein. 2010. "Promoting Healthy Choices: Information versus Convenience." *American Economic Journal: Applied Economics* 2 (2): 164–178.

Appendix Figure F1: Sensitivity Test for Figure I in Paper
Overlap/Union of .95 Conventional and Randomization Confidence Intervals

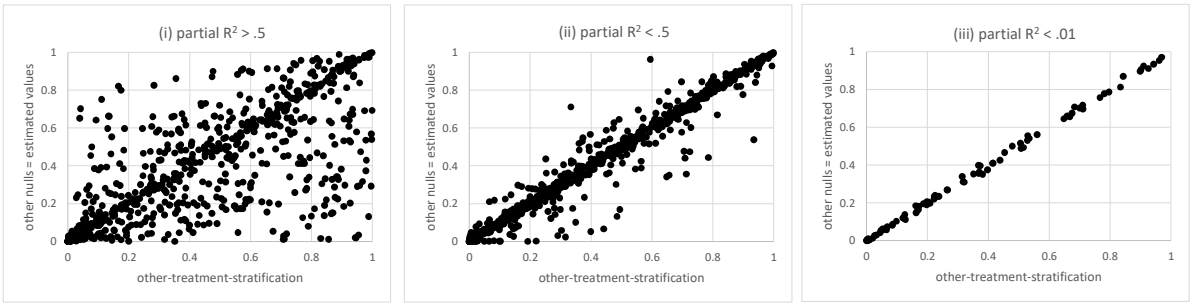


Appendix Figure F2: Sensitivity Test for Figure II in Paper
Conventional vs Randomization P-Values

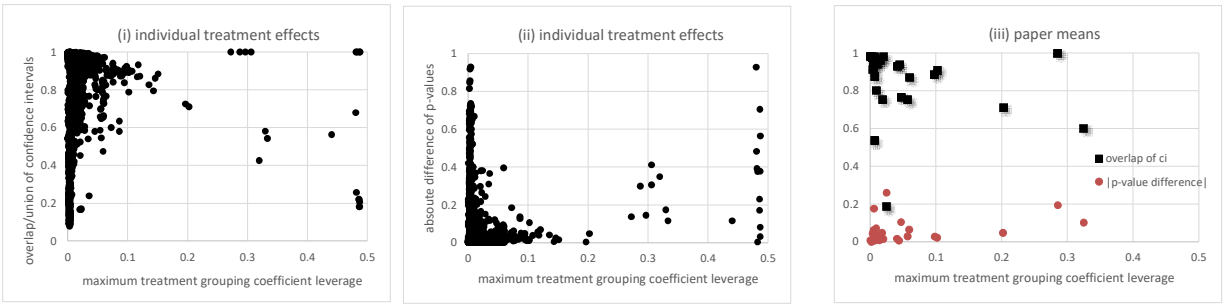


Appendix Figure F3: Sensitivity Test for Figure III in Paper
Randomization Inference using Other-Treatment-Stratification Compared with Other Methods

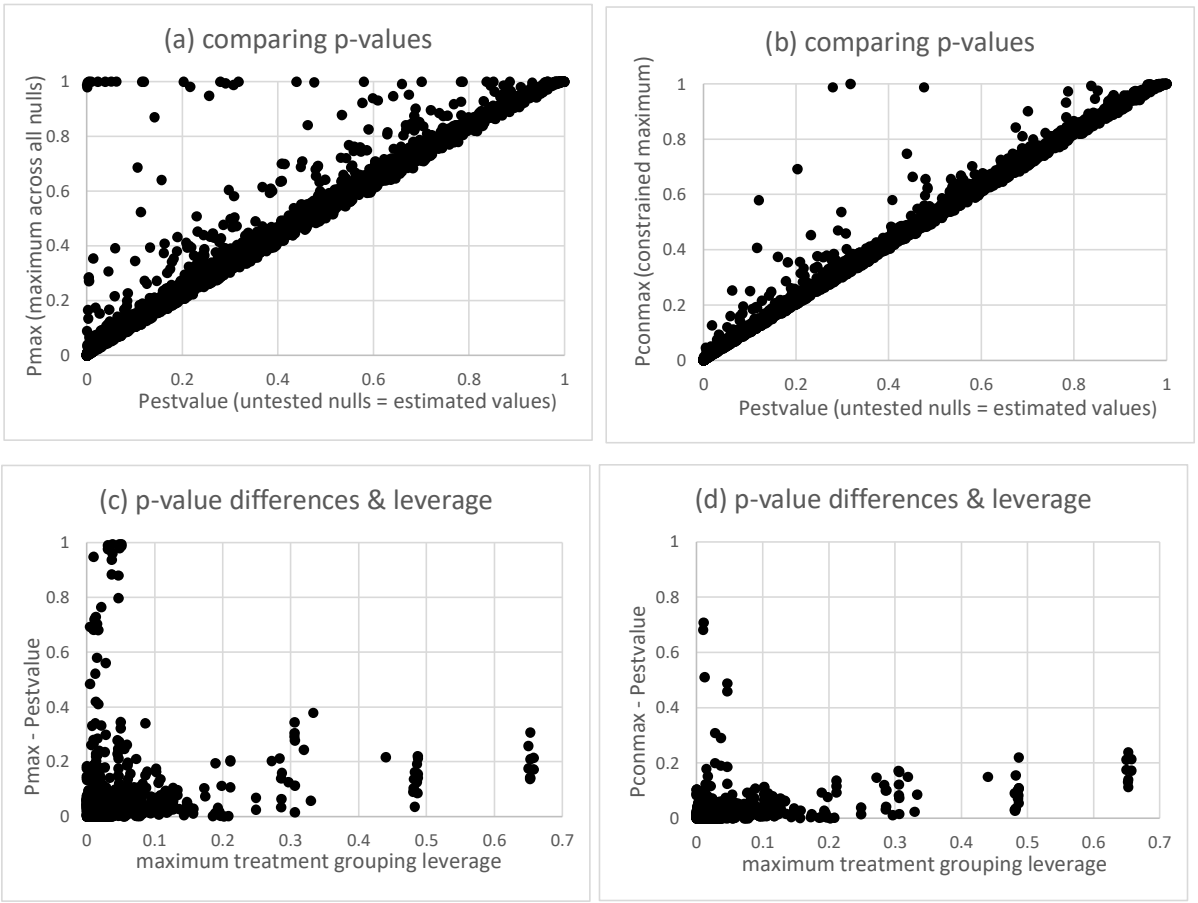
(a) Individual treatment effect p-values compared to those with null on untested coefficients = to estimated values by R^2 of regression of permuted treatment on other-treatment-stratification dummies



(b) Confidence intervals and p-values compared to those of conventional inference by maximum coefficient leverage of a single cluster/treatment grouping



Appendix Figure F4: Sensitivity Test for Figure V in the Paper
Maximum P-Value Across all Nulls for Untested Measures
Compared with Setting Untested Nulls Equal to Estimated Effects
(2712 individual treatment effects in 565 multi-treatment estimating equations)



G. Observations versus Leverage as Predictors of Differences in Conventional and Randomization Inference

Section V of the paper notes that while both the log number of treatment groupings and the maximum leverage share of a single treatment grouping are related to differences between conventional and randomization inference, the maximum leverage share appears to have greater predictive value (as measured by R^2) and is more robustly statistically significant. Table G1 below substantiates this claim with OLS regressions. The dependent variable is either the overlap divided by the union of the .95 confidence intervals or the absolute difference in the p-values produced by randomization and conventional clustered/robust inference, both measured at either the level of 2944 individual treatment effects or 35 paper averages. The independent variables are the maximum coefficient leverage share of a single treatment grouping, the $\log_{10}$ number of treatment groupings and a constant term or paper fixed effects. Standard errors in parentheses are either clustered at the paper level or, for paper averages, heteroskedasticity robust. Bootstrapped (at the paper level) p-values based upon the percentiles of t-statistics are also reported in brackets. As shown in the table, the R^2s associated with the use of maximum leverage as a regressor are very much higher than those found using the number of treatment groupings. While maximum leverage is almost always statistically significant at the .05 level, the number of treatment groupings is often statistically insignificant, especially when the bootstrap is used to evaluate significance or when entered jointly with maximum leverage in the same regression. Table G2 produces similar results using a 39 paper sample that includes the papers and regressions dropped because of stratification issues (see Appendix F above and discussion in Section V of the paper).

Table G1: Observations versus Leverage as Predictors of Differences between Randomization and Conventional Inference (2944 coefficients in 35 papers)

	(a) coefficient level no paper fixed effects s.e. clustered by paper			(b) coefficient level paper fixed effects s.e. clustered by paper			(c) paper averages heteroskedasticity robust s.e.		
	dependent variable: 1 - overlap divided by union of .95 confidence intervals								
max leverage	1.92 (.305) [.050]		1.90 (.236) [.023]	1.64 (.113) [.001]		1.66 (.117) [.002]	1.94 (.407) [.118]		2.06 (.582) [.147]
$\log_{10}$ # of treatment groupings		-.112 (.059) [.332]	-.005 (.026) [.854]		-.039 (.037) [.453]	.021 (.010) [.123]		-.175 (.055) [.104]	.018 (.043) [.714]
constant	.006 (.005) [.268]	.353 (.160) [.309]	.021 (.066) [.796]				-.005 (.015) [.752]	.512 (.142) [.085]	-.053 (.122) [.709]
R^2	.736	.176	.737	.868	.547	.870	.772	.457	.773
N	2944	2944	2944	2944	2944	2944	35	35	35
	dependent variable: absolute value of difference in conventional and randomization p-values								
max leverage	.377 (.038) [.000]		.404 (.034) [.000]	.408 (.027) [.000]		.415 (.030) [.001]	.260 (.054) [.058]		.309 (.055) [.010]
$\log_{10}$ # of treatment groupings		-.016 (.009) [.322]	.007 (.003) [.077]		-.006 (.008) [.580]	.009 (.004) [.041]		-.022 (.008) [.077]	.007 (.004) [.129]
constant	.005 (.001) [.001]	.057 (.023) [.237]	-.014 (.009) [.194]				.007 (.002) [.025]	.072 (.020) [.049]	-.012 (.012) [.325]
R^2	.446	.055	.454	.502	.187	.506	.707	.358	.722
N	2944	2944	2944	2944	2944	2944	35	35	35

Notes: clustered by 35 papers (panels a & b) or heteroskedasticity robust (panel c) standard errors in (); bootstrap p-values based upon 1999 bootstrap sampling (at paper level) draws from the distribution of t-statistics in [].

Table G2: Observations versus Leverage as Predictors of Differences between Randomization and Conventional Inference (3213 coefficients in 39 papers)

	(a) observation level no paper fixed effects s.e. clustered by paper			(b) coefficient level paper fixed effects s.e. clustered by paper			(c) paper averages heteroskedasticity robust s.e.		
	dependent variable: 1 - overlap divided by union of .95 confidence intervals								
max leverage	1.91 (.303) [.058]		1.90 (.239) [.025]	1.64 (.112) [.000]		1.65 (.117) [.001]	1.94 (.400) [.108]		2.04 (.548) [.135]
$\log_{10}$ # of treatment groupings		-.106 (.054) [.327]	-.004 (.022) [.866]		-.040 (.034) [.455]	.020 (.009) [.132]		-.157 (.050) [.098]	.014 (.033) [.724]
constant	.006 (.005) [.252]	.339 (.147) [.298]	.019 (.058) [.801]				-.003 (.014) [.821]	.473 (.130) [.079]	-.040 (.099) [.734]
R^2	.736	.176	.736	.867	.549	.868	.775	.435	.777
N	3213	3213	3213	3213	3213	3213	39	39	39
	dependent variable: absolute value of difference in conventional and randomization p-values								
max leverage	.377 (.038) [.001]		.402 (.035) [.000]	.407 (.028) [.000]		.414 (.030) [.000]	.262 (.053) [.038]		.285 (.060) [.007]
$\log_{10}$ # of treatment groupings		-.016 (.008) [.285]	.006 (.003) [.076]		-.007 (.008) [.543]	.008 (.003) [.061]		-.021 (.007) [.048]	.003 (.005) [.488]
constant	.005 (.001) [.001]	.056 (.021) [.213]	-.012 (.009) [.212]				.007 (.002) [.009]	.071 (.018) [.024]	-.001 (.014) [.908]
R^2	.444	.059	.451	.503	.194	.506	.680	.362	.683
N	3213	3213	3213	3213	3213	3213	39	39	39

Notes: clustered by 39 papers (panels a & b) or heteroskedasticity robust (panel c) standard errors in (); bootstrap p-values based upon 1999 bootstrap sampling (at paper level) draws from the distribution of t-statistics in [].

H. Determinants of Difference Between "Max across $\beta_{0\sim j}$" and "$\beta_{0\sim j} = \hat{\beta}_{0\sim j}$" Randomization Inference P-Values

Table H1 below uses regression analysis to examine the determinants of differences between the randomization inference p-value for individual coefficients in multi-treatment equations found by maximizing across all nulls for untested coefficients versus that found by setting the null for untested coefficients equal to estimated values. Panel (a) looks at the unconstrained maximum, whereas panel (b) looks at the constrained maximum wherein the null for untested coefficients is restricted to have a p-value greater than 10^{-10} using the conventional Wald test. The sample is the same as the practical sample examined in the paper and appendices above, except that it is restricted to equations with more than one treatment measure and hence consists of 2462 individual coefficient p-values found in 28 papers. The independent variables are the maximum coefficient leverage share of a single treatment grouping, the $\log_{10}$ number of treatment groupings, the $\log_{10}$ number minus 1 of treatment measures (so that an equation with two treatment measures, the lowest possible, has a $\log_{10}$ value of 0), the "$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$" p-value, its square and a constant term or paper fixed effects. Standard errors in parentheses are either clustered at the paper level or, for paper averages, heteroskedasticity robust. Bootstrapped (at the paper level) p-values based upon the percentiles of t-statistics are also reported in brackets.

As shown in the table, the maximum leverage and number of treatment measures are consistently significant, while the number of treatment groupings is never significant. The "$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$" p-value and its square are statistically significant when the regression is run using individual coefficients as the dependent variable, but completely insignificant when paper averages are used and much of the tail variation of these variables is eliminated. The amount a large p-value can be increased is obviously limited, and the coefficients on the p-values (when significant) indicate this, while also suggesting that small p-values are inherently more robust to the search across $\boldsymbol{\beta}_{0\sim j}$. When the sample is expanded to include regressions dropped on the basis of having asymptotically unbalanced treattment across strata and hence not fitting into the framework studied in this paper and the on-line appendix (Table H2), the regression fit is somewhat worse (with lower R^2s) and the number of treatment measures becomes insignificant in some specifications, but the patterns are similar.

Table H1: Determinants of Differences Between Maximum P-value across $\boldsymbol{\beta}_{0\sim j}$ and Setting $\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$

	(a) coefficient level no paper fixed effects s.e. clustered by paper				(b) coefficient level paper fixed effects s.e. clustered by paper				(c) paper averages heteroskedasticity robust s.e.			
	(a) dependent variable: maximum p-value across $\boldsymbol{\beta}_{0\sim j}$ minus p-value setting $\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$											
maximum leverage	.252 (.022) [.000]			.254 (.020) [.000]	.228 (.023) [.004]			.215 (.029) [.006]	.405 (.092) [.021]			.380 (.085) [.025]
$\log_{10}$ # of treatment groups		-.010 (.009) [.299]				-.022 (.016) [.323]				-.021 (.019) [.325]		
$\log_{10}$ # - 1 of treatment measures			.033 (.010) [.024]	.034 (.008) [.005]			.058 (.018) [.040]	.052 (.017) [.064]			.075 (.019) [.012]	.064 (.014) [.011]
$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ p-value	.211 (.023) [.001]	.220 (.024) [.001]	.193 (.026) [.001]	.185 (.025) [.001]	.204 (.025) [.001]	.212 (.025) [.001]	.203 (.024) [.001]	.194 (.023) [.001]	-.240 (.403) [.578]	.025 (.375) [.933]	.398 (.334) [.255]	-.078 (.332) [.832]
$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ p-value squared	-.214 (.024) [.001]	-.224 (.025) [.001]	-.205 (.026) [.001]	-.197 (.025) [.001]	-.208 (.025) [.001]	-.217 (.026) [.001]	-.209 (.024) [.001]	-.201 (.024) [.001]	.400 (.529) [.488]	.018 (.463) [.969]	-.692 (.444) [.164]	.073 (.452) [.883]
constant	.010 (.003) [.013]	.041 (.024) [.120]	-.001 (.006) [.875]	-.007 (.005) [.238]					.022 (.028) [.448]	.084 (.050) [.173]	.020 (.029) [.505]	.010 (.019) [.620]
R^2	.156	.102	.144	.205	.280	.248	.264	.300	.462	.097	.230	.607
N	2462	2462	2462	2462	2462	2462	2462	2462	28	28	28	28
	(a) dependent variable: constrained maximum p-value across $\boldsymbol{\beta}_{0\sim j}$ minus p-value setting $\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$											
maximum leverage	.182 (.020) [.000]			.183 (.024) [.001]	.190 (.025) [.012]			.184 (.026) [.019]	.198 (.046) [.027]			.184 (.038) [.017]
$\log_{10}$ # of treatment groups		-.005 (.004) [.175]				-.016 (.008) [.272]				-.010 (.009) [.341]		
$\log_{10}$ # - 1 of treatment measures			.023 (.003) [.000]	.023 (.002) [.000]			.029 (.010) [.297]	.024 (.010) [.352]			.041 (.009) [.036]	.036 (.007) [.015]
$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ p-value	.113 (.016) [.003]	.119 (.016) [.000]	.101 (.015) [.005]	.095 (.015) [.020]	.109 (.019) [.031]	.116 (.020) [.011]	.112 (.018) [.015]	.105 (.017) [.040]	-.196 (.134) [.220]	-.058 (.143) [.690]	.126 (.120) [.300]	-.105 (.084) [.241]
$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ p-value squared	-.103 (.015) [.002]	-.110 (.015) [.001]	-.097 (.014) [.005]	-.091 (.015) [.022]	-.102 (.018) [.041]	-.109 (.019) [.015]	-.105 (.018) [.015]	-.099 (.017) [.041]	.299 (.164) [.131]	.099 (.169) [.530]	-.254 (.167) [.150]	.117 (.099) [.260]
constant	.002 (.001) [.161]	.020 (.011) [.071]	-.005 (.002) [.072]	-.010 (.002) [.031]					.015 (.013) [.301]	.044 (.027) [.220]	.013 (.013) [.327]	.008 (.008) [.326]
R^2	.253	.137	.217	.341	.367	.272	.281	.383	.529	.083	.313	.753
N	2462	2462	2462	2462	2462	2462	2462	2462	28	28	28	28

Notes: clustered by 28 papers (panels a & b) or heteroskedasticity robust (panel c) standard errors in (); bootstrap p-values based upon 1999 bootstrap sampling (at paper level) draws from the distribution of t-statistics in [].

Table H2: Determinants of Differences Between Maximum P-value across $\boldsymbol{\beta}_{0\sim j}$ and Setting $\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ (including regressions dropped on the basis of unbalanced treatment across strata)

	(a) coefficient level no paper fixed effects s.e. clustered by paper				(b) coefficient level paper fixed effects s.e. clustered by paper				(c) paper averages heteroskedasticity robust s.e.			
	dependent variable: maximum p-value across $\boldsymbol{\beta}_{0\sim j}$ - p-value with $\boldsymbol{\beta}_{0\sim j}$ = estimated values											
maximum leverage	.278			.284	.235			.217	.402			.371
	(.033)			(.032)	(.023)			(.029)	(.080)			(.075)
	[.000]			[.000]	[.002]			[.005]	[.004]			[.008]
$\log_{10}$ # of treatment groups		-.021				-.022				-.029		
		(.014)				(.015)				(.015)		
		[.197]				[.282]				[.081]		
$\log_{10}$ # - 1 of treatment measures			.041	.042			.076	.071			.142	.136
			(.018)	(.017)			(.028)	(.028)			(.072)	(.073)
			[.089]	[.074]			[.116]	[.116]			[.321]	[.375]
$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ p-value	.167	.178	.145	.137	.188	.195	.184	.176	-.838	-.643	.014	-.395
	(.050)	(.048)	(.058)	(.058)	(.029)	(.028)	(.031)	(.031)	(.758)	(.810)	(.515)	(.486)
	[.082]	[.043]	[.129]	[.144]	[.001]	[.000]	[.000]	[.001]	[.414]	[.524]	[.982]	[.490]
$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ p-value squared	-.178	-.189	-.166	-.158	-.196	-.203	-.193	-.187	1.07	.809	-.341	.282
	(.042)	(.041)	(.047)	(.048)	(.026)	(.026)	(.027)	(.027)	(.898)	(.973)	(.546)	(.514)
	[.018]	[.004]	[.026]	[.042]	[.000]	[.000]	[.000]	[.000]	[.363]	[.502]	[.547]	[.596]
constant	.021	.082	.005	-.001	.019	.082	-.027	-.027	.093	.174	.054	.049
	(.012)	(.047)	(.009)	(.008)	(.005)	(.039)	(.017)	(.016)	(.080)	(.107)	(.053)	(.049)
	[.327]	[.188]	[.611]	[.871]	[.030]	[.195]	[.247]	[.230]	[.425]	[.292]	[.413]	[.467]
R^2	.068	.045	.070	.107	.358	.342	.359	.376	.148	.087	.246	.351
N	2712	2712	2712	2712	2712	2712	2712	2712	32	32	32	32
	dependent variable: maximum p-value across bounded $\boldsymbol{\beta}_{0\sim j}$ - p-value with $\boldsymbol{\beta}_{0\sim j}$ = estimated values											
maximum leverage	.185			.188	.190			.183	.195			.185
	(.021)			(.025)	(.024)			(.026)	(.041)			(.033)
	[.000]			[.002]	[.010]			[.014]	[.018]			[.010]
$\log_{10}$ # of treatment groups		-.007				-.015				-.011		
		(.004)				(.008)				(.006)		
		[.082]				[.270]				[.119]		
$\log_{10}$ # - 1 of treatment measures			.023	.023			.031	.026			.047	.044
			(.004)	(.003)			(.010)	(.010)			(.012)	(.012)
			[.000]	[.000]			[.152]	[.162]			[.039]	[.066]
$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ p-value	.113	.119	.102	.096	.115	.121	.117	.111	-.272	-.150	.076	-.128
	(.015)	(.014)	(.014)	(.014)	(.018)	(.019)	(.017)	(.017)	(.155)	(.168)	(.129)	(.093)
	[.001]	[.000]	[.002]	[.009]	[.009]	[.001]	[.001]	[.010]	[.140]	[.401]	[.576]	[.209]
$\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$ p-value squared	-.103	-.109	-.098	-.092	-.107	-.114	-.109	-.104	.382	.208	-.187	.123
	(.014)	(.013)	(.013)	(.013)	(.018)	(.018)	(.017)	(.016)	(.183)	(.197)	(.164)	(.103)
	[.001]	[.000]	[.001]	[.008]	[.010]	[.001]	[.001]	[.011]	[.091]	[.326]	[.265]	[.237]
constant	.003	.026	-.006	-.010	.003	.048	-.014	-.014	.025	.056	.014	.011
	(.001)	(.012)	(.002)	(.002)	(.003)	(.020)	(.009)	(.008)	(.016)	(.027)	(.013)	(.009)
	[.108]	[.044]	[.031]	[.009]	[.221]	[.268]	[.223]	[.192]	[.193]	[.112]	[.302]	[.286]
R^2	.211	.122	.184	.287	.343	.270	.280	.359	.394	.119	.335	.680
N	2712	2712	2712	2712	2712	2712	2712	2712	32	32	32	32

Notes: clustered by 32 papers (panels a & b) or heteroskedasticity robust (panel c) standard errors in (); bootstrap p-values based upon 1999 bootstrap sampling (at paper level) draws from the distribution of t-statistics in [].

I. Practical Results using Other Treatment & Covariate Stratification

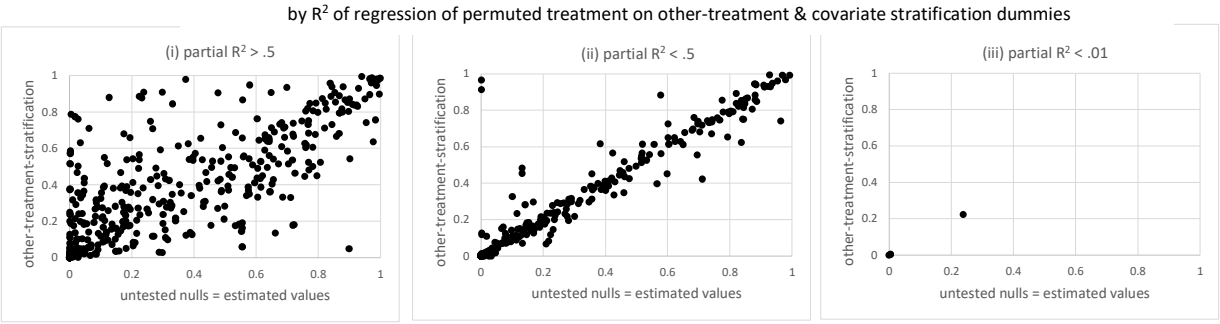
As noted in the paper, D'Haultfoeuille & Tuvaandorj's (2022) method of subset testing involves stratified permutation by all other covariates, but in the paper I only stratify treatment permutation by other treatment variables, as randomly applied treatment is independent of non-treatment covariates. In this appendix I implement their method in full, stratifying by non-permuted treatment values crossed with covariate values crossed with stratification measures (when these exist, to ensure that the resulting permutations are a valid subset of the potential outcomes of the original experimental procedure). The resulting distributions are non-degenerate for only 720 of the 2712 estimated treatment effects residing in equations with more than one treatment measure.[8] Figure I1a below graphs the p-values found setting the null equal to estimated effects against those found using other-treatment and covariate stratification. As in the paper, I divide results by the partial R^2 of the regression of the permuted treatment measure on the stratification dummies.[9] Also as in the paper, differences are greatest in the 419 cases where the R^2 is greater than .5, much less in the 301 cases where the R^2 is less than .5, and largely non-existent in the 4 cases where the R^2 is less than .01. Figure I1b graphs the overlap/union of the confidence intervals and the absolute value of the difference in p-values of other-treatment & covariates stratified permutation and conventional clustered/robust inference against the maximum coefficient leverage share of a single cluster/treatment grouping. Again, as is the case in the figures in the paper, large differences appear even when maximum leverage is near zero, so that the influence of individual observations is minimal and conventional inference is more likely (given amenable error moments) to have its desirable asymptotic properties. While the patterns in Figure I1a & I1b mimic those shown in Figure III in the paper & Figure F3 above, they appear somewhat more extreme, with for example more frequent and larger differences at negligible values of leverage (panel iiib), as the average partial R^2 is .57, as opposed to the .43 and .45 found with other-treatment-stratification alone in the samples of Figure III in the paper and Figure F3 above, respectively.

[8] Section V in the paper notes this as 667 because throughout the discussion there I drop regressions where the number of base treatment strata grew with the number of observations or treatment was unbalanced across strata (see Appendix F above and discussion in the paper). However, as D'Haultfoeuille & Tuvaandorj's theory does not exclude such cases, I include them here (as well as in Appendix F).

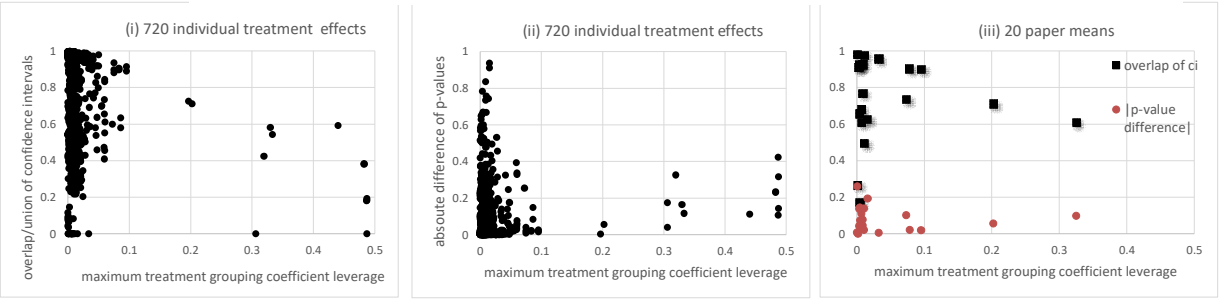
[9] "Partial" because I take the regression of treatment on the original stratification dummies (when the experiment was stratified) as the reduced model.

Figure I1: Randomization Inference using Other-Treatment & Covariate Stratification Compared with Other Methods

(a) 720 individual treatment effect p-values compared to those with null on untested coefficients = to estimated values by R^2 of regression of permuted treatment on other-treatment & covariate stratification dummies



(b) Confidence intervals and p-values compared to those of conventional inference by maximum coefficient leverage of a single cluster/treatment grouping



J. Formulae and Methods used in *Randcmdci* to Calculate Randomization Confidence Intervals

This appendix presents the formulae and methods used by *randcmdci* to calculate randomization confidence intervals and p-values for individual treatment effects. *Randcmdci* asks the user to indicate the base treatment variables that are permutable across observations or groups of observations, possibly within strata alone (all as specified by the user). The programme then executes any calculations given by the user to generate the regressors associated with treatment variables. The permutable treatment variables are not necessarily the treatment regressors themselves, as the calculations may transform the treatment values. The regression model is $\mathbf{y} = \mathbf{X\beta} + \mathbf{Z\gamma} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ are the K_X treatment regressors and $\mathbf{Z}$ the K_Z covariates. We depart from the notation in the paper and use $\mathbf{X}$ rather than $\mathbf{X_W}$ to indicate treatment regressors, as it is up to the user how treatment regressors are generated from base treatment variables (i.e. more may be done than simply interacting them with covariates). As long as base treatment variables are permutable across possibly stratified and grouped observations, as indicated by the user, and the calculation procedures given to transform these values into treatment regressors follow those used in the experiment, the resulting treatment regressors represent potential outcomes of the experiment and the randomization p-values are exact for sharp nulls. Asymptotic accuracy for heterogeneous treatment effects, however, is only guaranteed for the framework presented in the paper, where the only post permutation calculation used to generate a treatment regressor is one in which the permuted treatment variable is (possibly) multiplied by a covariate.[10] We use $\mathbf{T}$ to denote the treatment regressors generated by a permutation of the base treatment variables and the execution of any calculations indicated by the user. Again, to emphasize, our notation departs from that in the paper, as $\mathbf{T}$ here is not merely a permutation of treatment, but the resulting treatment following any post-permutation calculations, including multiplication with covariates ($\mathbf{T_W}$ in the notation of the paper) which fits the asymptotic theorems in the paper, but also allowing for other calculations which do not while still providing finite sample exact tests of sharp nulls.

The baseline coefficient estimates for treatment regressors are the vector $\hat{\boldsymbol{\beta}}$ with individual components $\hat{\beta}_j$ and clustered or heteroskedasticity robust covariance estimates $V(\hat{\beta}_j)$. Following each permutation of underlying treatment, counterfactual outcomes $\mathbf{y}_{\mathbf{T},\boldsymbol{\beta}_0} = \mathbf{y} + (\mathbf{T_W} - \mathbf{X_W})\boldsymbol{\beta}_0$ for the sharp null $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ (with j^{th} element β_{0j}) are calculated. The vector of coefficient estimates associated with these are $\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}$ with individual components $\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_{0j}}$ and associated clustered or heteroskedasticity robust covariance estimates $V(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_{0j}})$. Our first objective is to calculate which nulls $\boldsymbol{\beta}_0$ are consistent with randomization p-values greater than level α, i.e. a 1- α confidence interval. We do this by comparing the percentiles of Wald statistics for individual coefficients, i.e.

[10] Thus, availing ourselves of the example of Duflo, Dupas & Kremer's (2011) random assignment of students to class sections discussed in the paper, the assigned section is permutable across students, but the average quality of assigned peers (the treatment regressor) is not. Consequently, the regressor does not fit the framework discussed in the paper. Nevertheless, one can calculate confidence intervals for sharp nulls by permuting assignment, recalculating the quality of peers for each such assignment, and using it as the treatment regressor.

$$\text{(J.1)}\quad \frac{(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j}-\beta_{0j})^2}{V(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j})} \text{ to } \frac{(\hat{\beta}_j-\beta_{0j})^2}{V(\hat{\beta}_j)}.$$

Randcmdci begins by calculating coefficient estimates and residuals associated with a realized permutation of treatment **T** under the null that $\boldsymbol{\beta}_0 = \mathbf{0}$, that is

$$\text{(J.2)}\quad \hat{\boldsymbol{\beta}}_{\mathbf{T},\mathbf{0}} = (\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{T})^{-1}\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{y}_{\mathbf{T},\mathbf{0}} = (\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{T})^{-1}\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{y}$$
$$\hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\mathbf{0}} = \mathbf{M}_{\mathbf{Z}}(\mathbf{y}-\mathbf{T}\hat{\boldsymbol{\beta}}_{\mathbf{T},\mathbf{0}}), \text{ where } \mathbf{M}_{\mathbf{Z}} = \mathbf{I}_N - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})\mathbf{Z}'.$$

The coefficient estimates and residuals for any null $\boldsymbol{\beta}_0$ can then be calculated as:

$$\text{(J.3)}\quad \hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} = (\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{T})^{-1}\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{y}_{\mathbf{T},\boldsymbol{\beta}_0} = \boldsymbol{\beta}_0 + \hat{\boldsymbol{\beta}}_{\mathbf{T},\mathbf{0}} - \mathbf{A}\boldsymbol{\beta}_0 \quad [\text{where } \mathbf{A} = (\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{T})^{-1}(\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{X})\,]$$
$$\&\quad \hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\boldsymbol{\beta}_0} = \mathbf{M}_{\mathbf{Z}}(\mathbf{y}_{\mathbf{T},\boldsymbol{\beta}_0} - \mathbf{T}\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}) = \hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\mathbf{0}} - (\breve{\mathbf{X}} - \breve{\mathbf{T}}\mathbf{A})\boldsymbol{\beta}_0,$$

where we use $\breve{}$ to denote residuals from the projection on **Z**, as in $\breve{\mathbf{X}} = \mathbf{M}_{\mathbf{Z}}\mathbf{X}$. Let $\sum_{i\in c}$ denote summation across the observations i in cluster c and C the total number of clusters. When the covariance estimate is not clustered and merely heteroskedasticity robust, $C = N$ and each "cluster" contains one observation. Let $\mathbf{a}_k$ and a_{jk} denote the k^{th} column and jk^{th} element of **A**, $\breve{\breve{\mathbf{X}}} = \mathbf{M}_{\mathbf{Z}}\mathbf{X}(\mathbf{X}'\mathbf{M}_{\mathbf{Z}}\mathbf{X})^{-1}$ & $\breve{\breve{\mathbf{T}}} = \mathbf{M}_{\mathbf{Z}}\mathbf{T}(\mathbf{T}'\mathbf{M}_{\mathbf{Z}}\mathbf{T})^{-1}$ and let $\breve{\mathbf{t}}_i'$ denote the i^{th} row of $\breve{\mathbf{T}}$, $\breve{x}_{ij}$ the ij^{th} element of $\breve{\mathbf{X}}$, $\breve{\breve{t}}_{ij}$ the ij^{th} element of $\breve{\breve{\mathbf{T}}}$, & $\hat{\varepsilon}_{\mathbf{T},\mathbf{0}i}$ the i^{th} element of $\hat{\boldsymbol{\varepsilon}}_{\mathbf{T},\mathbf{0}}$. Further, define the K_X x 1 vector **d** and K_X x K_X matrix **S** with elements

$$\text{(J.4)}\quad d_k = \sum_{c=1}^{C}\left(\sum_{i\in c}\breve{\breve{t}}_{ij}\hat{\varepsilon}_{i\mathbf{T},\mathbf{0}}\right)\left(\sum_{i\in c}\breve{\breve{t}}_{ij}(\breve{x}_{ik}-\breve{\mathbf{t}}_i'\mathbf{a}_k)\right) \quad \&\quad S_{kl} = \sum_{c=1}^{C}\left(\sum_{i\in c}\breve{\breve{t}}_{ij}(\breve{x}_{ik}-\breve{\mathbf{t}}_i'\mathbf{a}_k)\right)\left(\sum_{i\in c}\breve{\breve{t}}_{ij}(\breve{x}_{il}-\breve{\mathbf{t}}_i'\mathbf{a}_l)\right).$$

With this notation in mind, the coefficient and clustered/robust variance estimate associated with the j^{th} treatment measure with regressors **T** and null $\boldsymbol{\beta}_0$ are given by:

$$\text{(J.5)}\quad \hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j} - \beta_{0j} = \hat{\beta}_{\mathbf{T},\mathbf{0}j} - \mathbf{a}_{j\sim j}'\boldsymbol{\beta}_{0\sim j} - a_{jj}\beta_{0j}$$
$$(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j} - \beta_{0j})^2 = (\hat{\beta}_{\mathbf{T},\mathbf{0}j} - a_{jj}\beta_{0j})^2 - 2(\hat{\beta}_{\mathbf{T},\mathbf{0}j} - a_{jj}\beta_{0j})\mathbf{a}_{j\sim j}'\boldsymbol{\beta}_{0\sim j} + \boldsymbol{\beta}_{0\sim j}'\mathbf{a}_{j\sim j}\mathbf{a}_{j\sim j}'\boldsymbol{\beta}_{0\sim j}$$
$$\&\ V(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j}) = \sum_{c=1}^{C}\left(\sum_{i\in c}\breve{\breve{t}}_{ij}[\hat{\varepsilon}_{i\mathbf{T},\mathbf{0}} - \sum_k(\breve{x}_{ik}-\breve{\mathbf{t}}_i'\mathbf{a}_k)\beta_{0k}]\right)^2 = \sum_{c=1}^{C}\left(\sum_{i\in c}\breve{\breve{t}}_{ij}\hat{\varepsilon}_{i\mathbf{T},\mathbf{0}}\right)^2 - 2\mathbf{d}'\boldsymbol{\beta}_0 + \boldsymbol{\beta}_0'\mathbf{S}\boldsymbol{\beta}_0$$
$$= \underbrace{V(\hat{\beta}_{\mathbf{T},\mathbf{0}j}) - 2\mathbf{d}_{\sim j}'\boldsymbol{\beta}_{0\sim j} + \boldsymbol{\beta}_{0\sim j}'\mathbf{S}_{\sim j\sim j}\boldsymbol{\beta}_{0\sim j}}_{c_3} + 2\beta_{0j}\underbrace{\left[-d_j + \mathbf{S}_{j\sim j}\boldsymbol{\beta}_{0\sim j}\right]}_{c_2} + \beta_{0j}^2\underbrace{\mathbf{S}_{jj}}_{c_1},$$

where subscripted ~j denotes excluding the j^{th} element, as in $\mathbf{a}_{j\sim j}$ is the j^{th} column of **A** excluding its j^{th} element and $\mathbf{S}_{\sim j\sim j}$ is **S** excluding its j^{th} row and column.[11] The letters c_1, c_2 and c_3 are used below to indicate the expressions given above.

The first step to calculating the randomization confidence interval is to calculate the roots implied by equality of the Wald statistics in (J.1) above. This equality defines a 4th order polynomial in β_{0j}:

[11] I follow Stata's convention and multiply the variance estimate by an adjustment for the finite sample bias in the case of normal iid errors based upon the number of observations, clusters and regressors, but as this appears on both sides of (I.1), I omit it in the equations above to minimize clutter.

$$\text{(J.6)}\quad f(\beta_{0j}) = \frac{(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_{0j}} - \beta_{0j})^2}{V(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_{0j}})} - \frac{(\hat{\beta}_j - \beta_{0j})^2}{V(\hat{\beta}_j)} = 0 \Rightarrow$$

$$(\hat{\beta}_{\mathbf{T},\mathbf{0}j} - \mathbf{a}'_{j\sim j}\boldsymbol{\beta}_{0\sim j} - a_{jj}\beta_{0j})^2 V(\hat{\beta}_j) - (\hat{\beta}_j^2 - 2\hat{\beta}_j\beta_{0j} + \beta_{0j}^2)(c_3 + 2c_2\beta_{0j} + c_1\beta_{0j}^2) = 0 \Rightarrow$$

$$\underbrace{-c_1}_{a}\beta_{0j}^4 + \underbrace{[2\hat{\beta}_j c_1 - 2c_2]}_{b}\beta_{0j}^3 + \underbrace{[a_{jj}^2 V(\hat{\beta}_j) + 4c_2\hat{\beta}_j - c_1\hat{\beta}_j^2 - c_3]}_{c}\beta_{0j}^2$$

$$+ \underbrace{[-2V(\hat{\beta}_j)a_{jj}(\hat{\beta}_{\mathbf{T},\mathbf{0}j} - \mathbf{a}'_{j\sim j}\boldsymbol{\beta}_{0\sim j}) + 2\hat{\beta}_j c_3 - 2c_2\hat{\beta}_j^2]}_{d}\beta_{0j} + \underbrace{[(\hat{\beta}_{\mathbf{T},\mathbf{0}j} - \mathbf{a}'_{j\sim j}\boldsymbol{\beta}_{0\sim j})^2 V(\hat{\beta}_j) - c_3\hat{\beta}_j^2]}_{e} = 0.$$

(J.6) allows for up to 4 real roots, which *randcmdci* calculates along with the derivative of the expression at the values of the real roots. When no real roots exist, *randcmdci* notes the value of *e*, which then tells whether the Wald statistic for post-permutation regressor outcome **T** is greater or less than that for **X** for all real β_{0j}. When *a*, *b*, *c*, *d* & *e* all equal to 0, which arises for example when **T** equals **X** or -**X**, the Wald statistics are identical for all β_{0j} (a "universal tie"). *randcmdci* checks for this case as well. In equations with more than one treatment measure, all of the above depends upon the nulls for the elements of $\boldsymbol{\beta}_{0\sim j}$. Based on the results in the paper, for the baseline calculation of confidence intervals *randcmdci* sets these equal to the estimated values $\hat{\boldsymbol{\beta}}_{0\sim j}$ although, as covered below, in calculating p-values alternate values are considered as well. Of course, where there is only one treatment measure $\sim j$ is the null set and all terms involving $\sim j$ in (J.5) & (J.6) are set equal to 0 (i.e. don't exist).

Randcmdci then calculates for each treatment regressor outcome **T** whether the lim as $\beta_{0j} \rightarrow -\infty$ of $f(\beta_{0j})$ is > , < or = to 0. Where real roots in (J.6) exist, this is determined by taking the sign of the derivative at the smallest root.[12] Where real roots do not exist, this is determined, as already noted above, by the sign of *e* or (in the case of a universal tie) by the fact that *a*, *b*, *c*, *d* & *e* all equal 0. The number of draws where $f(\beta_{0j})$ is found to be greater than or equal to 0 at this limit can be termed G[-∞] & E[-∞]. *Randcmdci* then orders all the real roots calculated in *D* draws of **T** along the real line, which we might denote as $r_1 < r_2 < r_3$ .... Moving along the real line indexed by *r*, using the value of the derivative at each r_i, the value of G[*r*] & E[*r*] is determined. A single draw of *U* distributed uniformly on (0,1) is used to calculate the p-value at each point on the real line as equal to (G[*r*]+*U**(E[*r*]+1))/(*D*+1),[13] and these p-values are used to calculate the .9, .95 & .99 confidence intervals. The program notes when the confidence interval is non-convex, in which case it alerts the user to this fact and reports the convex cover of the non-convex set. Temp variables and matrices used in the program's code follow the notation above, e.g. *a*, *b*, *c*, *d*, *e*, c_1, c_2, c_3, **d**, **S**, etc, except that subscript *k* rather than *j* is used to denote the coefficient of interest.

[12]The reader may note that c_1 above is ≥ 0. In the usual case, with $c_1 > 0$, *a* in (I.6) is < 0 and the derivative associated with the smallest real root (if it exists) is always positive, i.e. the draw **T** cannot contribute to G[-∞]. However, cases might arise where $c_1 = 0$ and the derivative on the smallest real root (if it exists) is negative, so the draw **T** contributes to G[-∞], so *randcmdci* checks for this possibility.

[13]Recall from (2.4) in the paper that the original treatment draw **X** is treated as a tie with itself, hence the +1.

A sidebar: Because of machine precision, a, b, c, d & e are often not exactly equal to 0, even when **T** does generate a universal tie with **X** (as when **T** equals **X** or -**X**). Consequently, it is necessary to use a non-zero cutoff as an indicator of 0. Ideally, this cutoff should not be sensitive to units of measure, i.e. a scaling of variables, so *randcmdci* uses a normalization to adjust for units of measure. Let k_y, k_j & $k_{\sim j}$ be scalars that multiply **y**, $\mathbf{x}_j$ ($\mathbf{t}_j$) and any $\mathbf{x}_{\sim j}$ ($\mathbf{t}_{\sim j}$). The following is the fashion in which the measures in (J.6) scale with these:

$$\text{(J.7)}\quad c_1 : k_y^0, k_j^0, k_{\sim j}^0,\ c_2 : k_y^1, k_j^{-1}, k_{\sim j}^0,\ c_3 : k_y^2, k_j^{-2}, k_{\sim j}^0,\ a_{jj} : k_y^0, k_j^0, k_{\sim j}^0,\ \mathbf{a}_{j\sim j} : k_y^0, k_j^{-1}, k_{\sim j}^1$$
$$\hat{\beta}_{\mathbf{T},\mathbf{0}j} : k_y^1, k_j^{-1}, k_{\sim j}^0,\ \hat{\beta}_j : k_y^1, k_j^{-1}, k_{\sim j}^0,\ \boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{\sim j} : k_y^1, k_j^0, k_{\sim j}^{-1},\ V(\hat{\beta}_j) : k_y^2, k_j^{-2}, k_{\sim j}^0$$
$$a : k_y^0, k_j^0, k_{\sim j}^0,\ b : k_y^1, k_j^{-1}, k_{\sim j}^0,\ c : k_y^2, k_j^{-2}, k_{\sim j}^0,\ d : k_y^3, k_j^{-3}, k_{\sim j}^0,\ b : k_y^4, k_j^{-4}, k_{\sim j}^0$$

Consequently, *randcmdci* uses the following indicator which is unaffected by scale:

$$\text{(J.8)}\quad dif = |a| + \frac{|b|}{V(\hat{\beta}_j)^{1/2}} + \frac{|c|}{V(\hat{\beta}_j)} + \frac{|d|}{V(\hat{\beta}_j)^{3/2}} + \frac{|e|}{V(\hat{\beta}_j)^2}.$$

In the more than 6 million realizations of **T** across 1999 permutations each of thousands of treatment measures in my practical sample, there is a gap in the distribution of *dif*. 26826 realizations of *dif* are less than $3\text{x}10^{-10}$ (and these realizations can be confirmed through examination of **T** and **X** to generate universal ties), and the remaining 6 million+ are greater than $2\text{x}10^{-3}$. *Randcmdci* uses a value of $dif < 10^{-9}$ to identify universal ties.

Continuing, to compute the p-value of the null of zero effects for an individual treatment effect when setting the null for untested measures equal to estimated values ($\beta_{0j} = 0$, $\boldsymbol{\beta}_{0\sim j} = \hat{\boldsymbol{\beta}}_{0\sim j}$), *randcmdci* calculates the number of instances G & E where:

$$\text{(J.9)}\quad G[0] : I\left(\left|\frac{\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j}}{V(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j})^{1/2}}\right| > \left|\frac{\hat{\beta}_j}{V(\hat{\beta}_j)^{1/2}}\right| + 10^{-9}\right),\ G[0] + E[0] : I\left(\left|\frac{\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j}}{V(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_0 j})^{1/2}}\right| > \left|\frac{\hat{\beta}_j}{V(\hat{\beta}_j)^{1/2}}\right| - 10^{-9}\right)$$

where I is an indicator for the event occurring. The p-value is then given by $(G[0]+U*(E[0]+1))/(D+1)$, using the same U used to calculate the randomization confidence interval above. As can be seen in (J.9), as an allowance for machine precision an absolute difference in the absolute value of the t-statistics of less than 10^{-9} is considered a tie (contributing to E). Again, in more than 6 million permutations of treatment across my practical sample, there is a gap in the distribution of the difference of the absolute value of t-statistics, as in (J.9). In 27701 instances it is less than $4\text{x}10^{-12}$ in absolute value,[14] and in the remaining 6 million+ instances it is greater than $4\text{x}10^{-7}$ in absolute value.

Randcmdci also allows the user to call for the calculation of the maximum p-value for the test of zero effects ($\beta_{0j} = 0$) for an individual treatment effect across all possible nulls $\boldsymbol{\beta}_{0\sim j}$, which ensures control of the null rejection probability below nominal level in the case of sharp nulls (see the discussion in the

[14]These include the universal ties identified above plus 875 additional ties for the specific null $\beta_{0j} = 0$ where the absolute value of the difference in absolute t-stats is less than $6\text{x}10^{-13}$.

paper). Setting $\beta_{0j} = 0$ and substituting using the definitions in (J.5), we can write:

$$\text{(J.10)}\quad (\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_{0j}} - \beta_{0j})^2 V(\hat{\beta}_j) - (\hat{\beta}_j - \beta_{0j})^2 V(\hat{\beta}_{\mathbf{T},\boldsymbol{\beta}_{0j}}) =$$

$$(\hat{\beta}_{\mathbf{T},\mathbf{0}j} - \mathbf{a}'_{j\sim j}\boldsymbol{\beta}_{0\sim j})^2 V(\hat{\beta}_j) - \hat{\beta}_j^2 (V(\hat{\beta}_{\mathbf{T},\mathbf{0}j}) - 2\mathbf{d}'_{\sim j}\boldsymbol{\beta}_{0\sim j} + \boldsymbol{\beta}'_{0\sim j}\mathbf{S}_{\sim j\sim j}\boldsymbol{\beta}_{0\sim j}) =$$

$$\boldsymbol{\beta}'_{0\sim j}\underbrace{[V(\hat{\beta}_j)\mathbf{a}_{j\sim j}\mathbf{a}'_{j\sim j} - \hat{\beta}_j^2\mathbf{S}_{\sim j\sim j}]}_{\mathbf{A}_0}\boldsymbol{\beta}_{0\sim j} + \underbrace{[2\hat{\beta}_j^2\mathbf{d}_{\sim j} - 2\hat{\beta}_{\mathbf{T},\mathbf{0}j}V(\hat{\beta}_j)\mathbf{a}_{j\sim j}]'}_{\mathbf{b}_0}\boldsymbol{\beta}_{0\sim j} + \underbrace{[\hat{\beta}_{\mathbf{T},\mathbf{0}j}^2 V(\hat{\beta}_j) - V(\hat{\beta}_{\mathbf{T},\mathbf{0}j})\hat{\beta}_j^2]}_{c_0} = g(\boldsymbol{\beta}_{0\sim j})$$

$g(\boldsymbol{\beta}_{0\sim j}) = 0$ defines a quadratic equation in $\boldsymbol{\beta}_{0\sim j}$. As was the case above, there are cases where all elements of $\mathbf{A}_0$, $\mathbf{b}_0$ and c_0 are zero, there is a universal tie[15] and $g(\boldsymbol{\beta}_{0\sim j})$ is identically zero. As before, because of machine precision it is necessary to construct a non-zero cutoff to distinguish such cases, and this cutoff should not be sensitive to scaling of variables. Following the definitions used earlier above, note that

$$\text{(J.11)}\quad \mathbf{a}_{j\sim j} : k_y^0, k_j^{-1}, k_{\sim j}^1,\ \mathbf{S}_{\sim j\sim j} : k_y^0, k_j^{-2}, k_{\sim j}^2,\ \mathbf{d}_{\sim j} : k_y^1, k_j^{-2}, k_{\sim j}^1$$

$$\hat{\beta}_{\mathbf{T},\mathbf{0}j} : k_y^1, k_j^{-1}, k_{\sim j}^0,\ \hat{\beta}_j : k_y^1, k_j^{-1}, k_{\sim j}^0,\ V(\hat{\beta}_j) : k_y^2, k_j^{-2}, k_{\sim j}^0,\ V(\hat{\beta}_{\sim j}) : k_y^2, k_j^0, k_{\sim j}^{-2}$$

$$\mathbf{A}_0 : k_y^2, k_j^{-4}, k_{\sim j}^2,\ \mathbf{b}_0 : k_y^3, k_j^{-4}, k_{\sim j}^1,\ c_0 : k_y^4, k_j^{-4}, k_{\sim j}^0$$

and construct the following indicator which is unaffected by scale:

$$\text{(J.12)}\quad dif_2 = \frac{|c_0|}{V(\hat{\beta}_j)^2} + \sum_{k\in\sim j}\frac{V(\hat{\beta}_k)^{1/2}|b_{0k}|}{V(\hat{\beta}_j)^2} + \sum_{k\in\sim j}\sum_{l\in\sim j}\frac{V(\hat{\beta}_k)^{1/2}V(\hat{\beta}_l)^{1/2}|A_{0kl}|}{V(\hat{\beta}_j)^2},$$

where b_{0k} denotes the k^{th} element of $\mathbf{b}_0$, A_{0kl} the k^{th} x l^{th} element of $\mathbf{A}_0$, and $k \in \sim j$ summation across all k in $1...K_x$ excluding j. As before there is a gap in the distribution of dif_2. In more than 5 million permutations of treatment in multi-treatment equations across my practical sample, it is less than $2\text{x}10^{-12}$ in 6327 instances and greater than $2\text{x}10^{-5}$ in the remaining 5 million+. *Randcmdci* uses a value of $dif_2 < 10^{-9}$ to identify universal ties in these computations.

Returning to (J.10), when there are only two treatment measures, i.e. $\boldsymbol{\beta}_{0\sim j}$ is the scalar $\beta_{0\sim j}$, we can solve for the roots for each $\mathbf{T}$ in D draws and line these up along the real line. In some instances there are no roots, in some instances the quadratic and/or linear term is zero, and there is also the possibility of universal ties where all terms are zero, as identified by dif_2. *Randcmdci* follows the same procedure used in the case of confidence intervals for β_{0j} above, calculating the number of cases where $g(\beta_{0\sim j})$ is greater than 0 or equal to 0 (the universal ties) as $\beta_{0\sim j} \rightarrow$-∞ and then, with the location of real roots r of $g(\beta_{0\sim j})$ and the sign of the derivative of g at those roots for each treatment outcome $\mathbf{T}$ in hand, moving along the real line indexed by r and keeping track of G[r] & E[r]. With the p-value at each point given by $(G[r]+U^*(E[r]+1))/(D+1)$, using the same U as was used to calculate the confidence interval for β_{0j} above, *randcmdci* calculates the maximum p-value across all values r. This is the maximum p-value for the test of $\beta_{0j} = 0$ across all possible nulls for $\beta_{0\sim j}$.

[15]The universal tie here is different than in the case considered in (I.6), (I.8) & (I.9), as that concerned a universal tie across all β_{0j} given $\boldsymbol{\beta}_{0\sim j}$ = estimated values whereas here we are examining a universal tie across all $\boldsymbol{\beta}_{0\sim j}$ given $\beta_{0j} = 0$.

For the case of three treatment measures, where $\boldsymbol{\beta}_{0\sim j}$ is a 2 x 1 vector, *randcmdci* transforms $\boldsymbol{\beta}_{0\sim j}$ into polar coordinates, giving:

$$\text{(J.13)}\quad g(\boldsymbol{\beta}_{0\sim j}) \Rightarrow g(\theta, r) = \mathbf{x}'\mathbf{A}_0\mathbf{x}\,r^2 \;+\; \mathbf{b}_0'\mathbf{x}\,r \;+\; c_0 = 0, \quad \text{where } \mathbf{x}' = [\sin(\theta), \cos(\theta)],\ \theta \in [0, 2\pi).$$

For a given value of θ this is a quadratic equation in *r*, and the procedure described above can be used to solve for the maximum p-value across all positive and negative *r* given θ, which we may call $g^*(\theta)$. *Randcmdci* then performs a line search across θ dividing [0,π] into # (as given by the user option *maxlevel*(#)) evenly spaced points and takes the maximum p-value across these. Because the maximum across *r* for each θ allows for both positive and negative values of *r*, this search automatically considers the maximum p-value along the opposite ray where θ lies in [π,2π].

For the case of K_x = four or more treatment measures, $\boldsymbol{\beta}_{0\sim j}$ is a K_x - 1 vector and we transform into n-dimensional spherical coordinates, where with $\boldsymbol{\theta}$ a *k* (= K_x - 2) x 1 vector:

$$\text{(J.14)}\quad g(\boldsymbol{\beta}_{0\sim j}) \Rightarrow g(\boldsymbol{\theta}, r) = \mathbf{x}'\mathbf{A}_0\mathbf{x}\,r^2 \;+\; \mathbf{b}_0'\mathbf{x}\,r \;+\; c_0 = 0,$$

$$\text{where } x_1 = \cos(\theta_1),\ x_2 = \sin(\theta_1)\cos(\theta_2),\ x_3 = \sin(\theta_1)\sin(\theta_2)\cos(\theta_3),\ \ldots.,\ x_{k+1} = \prod_{l=1}^{k}\sin(\theta_l),$$

$$\text{with } \theta_l \in [0, \pi] \text{ for } l = 1 \ldots k\text{-}1 \;\;\&\;\; \theta_k \in [0, 2\pi).$$

For a given value of $\boldsymbol{\theta}$ this is a quadratic equation in *r*, and the procedure described above can be used to solve for the maximum p-value across *r* given that $\boldsymbol{\theta}$, which we may call $g^*(\boldsymbol{\theta})$. *Randcmdci* then performs # (as given by *maxlevel*) iterations of the Nelder-Mead (1965) search procedure. For each iteration, *k*+1 draws from the uniform distribution across $[0,\pi]^k$ are used to find *k*+1 initial $g^*(\boldsymbol{\theta})$ values and then the Nelder-Mead simplex method is executed until the $g^*(\boldsymbol{\theta})$ for the *k*+1 vectors in the simplex are identical. *Randcmdci* takes the maximum across the # independent Nelder-Mead searches and reports that as the maximum p-value. On each round of simplex optimization I draw $\mathbf{u} = (u_1, u_2, u_3)$ from the 3 dimensional iid uniform distribution on (0,1) and use these to randomly set the Nelder-Mead reflection coefficient α = u_1, expansion coefficient γ = 1+ u_2, and contraction coefficient β = u_3.

Randcmdci also allows the user to ask for a bounded search. If the *boundcoef*(#) option is chosen, $(\boldsymbol{\beta}_{0\sim j} - \hat{\boldsymbol{\beta}}_{0\sim j})'(\boldsymbol{\beta}_{0\sim j} - \hat{\boldsymbol{\beta}}_{0\sim j})$ must be less than $\#^2$. Using the change of variables

$$\text{(J.15)}\quad g(\boldsymbol{\beta}_{0\sim j}) = \boldsymbol{\beta}_{0\sim j}'\mathbf{A}_0\boldsymbol{\beta}_{0\sim j} \;+\; \mathbf{b}_0'\boldsymbol{\beta}_{0\sim j} \;+\; c_0 = \tilde{\boldsymbol{\beta}}_{0\sim j}'\tilde{\mathbf{A}}_0\tilde{\boldsymbol{\beta}}_{0\sim j} \;+\; \tilde{\mathbf{b}}_0'\tilde{\boldsymbol{\beta}}_{0\sim j} \;+\; \tilde{c}_0 = g(\tilde{\boldsymbol{\beta}}_{0\sim j})$$

$$\text{where } \tilde{\boldsymbol{\beta}}_{0\sim j} = \boldsymbol{\beta}_{0\sim j} - \hat{\boldsymbol{\beta}}_{0\sim j},\ \tilde{\mathbf{A}}_0 = \mathbf{A}_0,\ \tilde{\mathbf{b}}_0 = \mathbf{b}_0 + 2\mathbf{A}_0\hat{\boldsymbol{\beta}}_{0\sim j},\ \&\ \tilde{c}_0 = c_0 + \mathbf{b}_0'\hat{\boldsymbol{\beta}}_{0\sim j} + \hat{\boldsymbol{\beta}}_{0\sim j}'\mathbf{A}_0\hat{\boldsymbol{\beta}}_{0\sim j}.$$

If the user specifies the *boundwald*(#) option, $(\boldsymbol{\beta}_{0\sim j} - \hat{\boldsymbol{\beta}}_{\sim j})'V(\hat{\boldsymbol{\beta}}_{\sim j})^{-1}(\boldsymbol{\beta}_{0\sim j} - \hat{\boldsymbol{\beta}}_{\sim j})$ must be less than $\#^2$. As $V(\hat{\boldsymbol{\beta}}_{\sim j})$ is strictly positive definite, we have:

$$\text{(J.16)}\quad V(\hat{\boldsymbol{\beta}}_{\sim j})^{-1} = \mathbf{V}^{-½}\mathbf{V}^{-½}, \text{ where } \mathbf{V}^{-½} = \mathbf{E}\boldsymbol{\Lambda}^{-½}\mathbf{E}' \;\&\; \mathbf{V}^{½} = \mathbf{E}\boldsymbol{\Lambda}^{½}\mathbf{E}'$$

and where $\mathbf{E}$ are the eigenvectors of $V(\hat{\boldsymbol{\beta}}_{\sim j})$, and $\boldsymbol{\Lambda}^{½}$ & $\boldsymbol{\Lambda}^{-½}$ denote diagonal matrices whose elements are the square root & inverse of the square root, respectively, of the eigenvalues of $V(\hat{\boldsymbol{\beta}}_{\sim j})$. Using the change of variables:

$$\text{(J.17)}\quad g(\boldsymbol{\beta}_{0\sim j}) = \boldsymbol{\beta}'_{0\sim j}\mathbf{A}_0\boldsymbol{\beta}_{0\sim j} + \mathbf{b}'_0\boldsymbol{\beta}_{0\sim j} + c_0 = \widetilde{\boldsymbol{\beta}}'_{0\sim j}\widetilde{\mathbf{A}}_0\widetilde{\boldsymbol{\beta}}_{0\sim j} + \widetilde{\mathbf{b}}'_0\widetilde{\boldsymbol{\beta}}_{0\sim j} + \widetilde{c}_0 = g(\widetilde{\boldsymbol{\beta}}_{0\sim j})$$

where $\widetilde{\boldsymbol{\beta}}_{0\sim j} = \mathbf{V}^{-½}(\boldsymbol{\beta}_{0\sim j} - \hat{\boldsymbol{\beta}}_{\sim j})$, $\widetilde{\mathbf{A}}_0 = \mathbf{V}^{½}\mathbf{A}_0\mathbf{V}^{½}$, $\widetilde{\mathbf{b}}_0 = \mathbf{V}^{½}\mathbf{b}_0 + 2\mathbf{V}^{½}\mathbf{A}_0\hat{\boldsymbol{\beta}}_{\sim j}$, & $\widetilde{c}_0 = c_0 + \mathbf{b}'_0\hat{\boldsymbol{\beta}}_{\sim j} + \hat{\boldsymbol{\beta}}'_{\sim j}\mathbf{A}_0\hat{\boldsymbol{\beta}}_{\sim j}$.

For both *boundcoef* and *boundwald*, the search for a maximum p-value is now restricted to values of *r*, as in the previous paragraphs, $\leq \#$. For $K_x = 2$ the calculation of a maximum across the unbounded or bounded space for *r* can proceed concurrently. The same can be done with the line search across θ in $[0,\pi]$ when $K_x = 3$, as for each pre-determined θ a maximum across bounded and unbounded values of *r* can be calculated. However, when $K_x \geq 4$, the values of $\boldsymbol{\theta}$ chosen by the Nelder-Mead algorithm at each step depend upon the values of the maximand $g^*(\boldsymbol{\theta})$ in the simplex, which is different when *r* is bounded. Consequently, for $K_x \geq 4$ separate searches are conducted, with and without bounds on *r*.

Finally, it should be noted that if requested by users *randcmdci* will also provide p-values for tests of specific (joint) nulls $\boldsymbol{\beta}_0$. This is done by calculating the number of instances *G* & *E* where:

$$\text{(J.18)}\quad G[\boldsymbol{\beta}_0] : I\begin{pmatrix}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0)'V(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})^{-1}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0) \\ > (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'V(\hat{\boldsymbol{\beta}})^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + 10^{-9}\end{pmatrix},\ G[\boldsymbol{\beta}_0] + E[\boldsymbol{\beta}_0] : I\begin{pmatrix}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0)'V(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})^{-1}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0} - \boldsymbol{\beta}_0) \\ > (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'V(\hat{\boldsymbol{\beta}})^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 10^{-9}\end{pmatrix}.$$

where as before *I* is an indicator for the event occurring. With a random draw *U* from the uniform distribution on (0,1), the p-value is given by $(G[\boldsymbol{\beta}_0]+U*(E[\boldsymbol{\beta}_0]+1))/(D+1)$.

K. Convergence in Distribution for any $\boldsymbol{\beta}_0$ (notation follows that used in the paper & its appendices)

In this appendix we prove a version of (R1) that establishes convergence in distribution for any fixed $\boldsymbol{\beta}_0$, not merely for drifting sequences $\boldsymbol{\beta}_0$ that lie in a root-N neighborhood of the true parameter values $\boldsymbol{\beta}$. (R1) is modified to read:

(RR1) Given White's (1980) assumptions W1 - W4 and the additional assumptions A1 - A3, as in the paper and modified below, for any $\boldsymbol{\beta}_0$ the Wald statistic $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ based on the heteroskedasticity robust covariance estimate is asymptotically distributed chi-squared with PQ degrees of freedom

$$\tau(\mathbf{T},\boldsymbol{\beta}_0) \overset{d(\mathbf{T})|a.s.(\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})}{\rightarrow} \chi^2_{PQ},$$

where $\overset{d(\mathbf{T})|a.s.(\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})}{\rightarrow}$ denotes convergence as $N \rightarrow \infty$ in distribution across the permutations $\mathbf{T}$ of $\mathbf{X}$ almost surely given the realization of the data $(\mathbf{X_W},\mathbf{Z},\boldsymbol{\varepsilon})$.

(RR1) does not subsume (R1) given in the paper, as no claim is made that $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ converges in probability to $\tau(\mathbf{T},\boldsymbol{\beta})$. The fixed value of $\boldsymbol{\beta}_0$ asymptotically affects the variance of the counterfactual coefficient estimates, but the covariance estimate properly adjusts for this producing a chi-squared distribution. The individual realizations of $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ depend upon $\boldsymbol{\beta}_0$ and do not necessarily equal $\tau(\mathbf{T},\boldsymbol{\beta})$.

(RR1) requires changing assumptions (A1) and (A3) as given in the paper to read:

(AA1) There exists a finite positive constant γ such that (a) $\mathbf{G}_N = \sum_{i=1}^N E(\mathbf{x}_i\mathbf{x}_i')/N$ - $\sum_{i=1}^N E(\mathbf{x}_i)/N \sum_{i=1}^N E(\mathbf{x}_i')/N$ is non-singular for all N sufficiently large with determinant $\mathbf{G}_N > \gamma > 0$; (b) with $\boldsymbol{\Phi}_N = (\sum_{i=1}^N E(\mathbf{z}_i\mathbf{z}_i')/N)^{-1} \sum_{i=1}^N E(\mathbf{z}_i\mathbf{x}_{\mathbf{W}i}')/N$, $\breve{\mathbf{x}}_{\mathbf{W}i}' = \mathbf{x}_{\mathbf{W}i}' - \mathbf{z}_i'\boldsymbol{\Phi}_N$ & $\omega_i = \boldsymbol{\alpha}'\mathbf{w}_i\mathbf{w}_i'\boldsymbol{\alpha}$, for all $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\boldsymbol{\alpha}=1$ & for all N sufficiently large

$$\sum_{i=1}^N \frac{E(\omega_i\varepsilon_i^2)}{N} - \sum_{i=1}^N \frac{E(\omega_i\varepsilon_i\breve{\mathbf{x}}_{\mathbf{W}i}')}{N}\mathbf{K}^{-1}\sum_{i=1}^N \frac{E(\omega_i\varepsilon_i\breve{\mathbf{x}}_{\mathbf{W}i})}{N} > \gamma > 0,$$

$$\text{where } \mathbf{K} = \sum_{i=1}^N \frac{E(\omega_i\breve{\mathbf{x}}_{\mathbf{W}i}\breve{\mathbf{x}}_{\mathbf{W}i}')}{N} \text{ with determinant } \mathbf{K} > \gamma > 0.$$

(AA3) There exist positive finite constants θ, θ^* and $\Delta < \infty$, with $\theta(1+2\theta^*) > 1$, such that for all $i, j = 1...K_+$, $p = 1...P$, and $q = 1...Q$, $E(|w_{iq}^2\varepsilon_i^2|^{1+\theta}) < \Delta$, $E(|w_{iq}^2 z_{+ij}^2|^{1+\theta}) < \Delta$ and $E(|x_{ip}^4|^{1+\theta^*}) < \Delta$.

Relative to (A1) in the paper, the addition of AA1b requires that there is enough independent variation in the errors that their residual variation in a regression on the treatment variables weighted by any combination of the interaction covariates does not asymptotically go to zero, (i.e., that the R^2 in the weighted regression of the outcome $\mathbf{y}$ on $\mathbf{X_W}$, net of the effects of covariates $\mathbf{Z}$, does not go 1), and ensures that the covariance matrix of $\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0)$ based on counterfactual outcomes remains non-singular for any null $\boldsymbol{\beta}_0$. $E(|w_{iq}^2 z_{+ij}^2|^{1+\theta}) < \Delta$ in (AA3) is an added moment condition that, depending upon the value of θ^*, may be more demanding than the $E(|z_{+ij}^4|^{1+\delta}) < \Delta$ given in White's assumption (W4). Otherwise, the framework and notation is as given in the paper and its appendices.

Following each permutation of treatment, the dependent variable is adjusted in accordance with the null and the realization **T** of treatment

$$\text{(K.1)}\quad \mathbf{y}(\mathbf{T},\boldsymbol{\beta}_0) = \mathbf{y} + (\mathbf{T}\bullet\mathbf{W} - \mathbf{X}\bullet\mathbf{W})\boldsymbol{\beta}_0 = \mathbf{X}_\mathbf{W}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} + \mathbf{T}_\mathbf{W}\boldsymbol{\beta}_0 .$$

With $\mathbf{M} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ denoting the residual maker with respect to **Z**, the estimated coefficients and residuals associated with **T** and $\boldsymbol{\beta}_0$ are

$$\text{(K.2)}\quad \hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0) = (\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W})^{-1}\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{y}(\mathbf{T},\boldsymbol{\beta}_0) = (\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W})^{-1}\mathbf{T}_\mathbf{W}'\mathbf{M}\boldsymbol{\xi} + \boldsymbol{\beta}_0 \text{ where } \boldsymbol{\xi} = \mathbf{M}\mathbf{X}_\mathbf{W}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \mathbf{M}\boldsymbol{\varepsilon},$$

$$\text{(K.3)}\quad \hat{\boldsymbol{\varepsilon}}(\mathbf{T},\boldsymbol{\beta}_0) = \mathbf{M}\mathbf{y}(\mathbf{T},\boldsymbol{\beta}_0) - \mathbf{M}\mathbf{T}_\mathbf{W}\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0) = \boldsymbol{\xi} - \mathbf{M}\mathbf{T}_\mathbf{W}(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0) - \boldsymbol{\beta}_0),$$

and where we use the fact that $\mathbf{M}\mathbf{Z} = \mathbf{0}_{NxK}$ and define $\boldsymbol{\xi}$ which will be used repeatedly below.

All Lemmas proven in the paper continue to hold, as the moment conditions have, if anything, been strengthened, and will be referenced below as Lemma 1, Lemma 2, etc. The following additional Lemma, proven below, will also be useful:

Lemma K1: White's assumptions W1 - W4 and the additional A1 - A3 as modified above ensure that

(a) The means of the products of four columns of $\mathbf{E} = (\mathbf{Z}_+,\boldsymbol{\xi})$, no more than two of which are $\boldsymbol{\xi}$, are almost surely bounded.

(b) With $\mathbf{W}_\xi = \mathbf{W}\bullet\boldsymbol{\xi}$, almost surely for all N sufficiently large $\tilde{\mathbf{W}}_\xi'\tilde{\mathbf{W}}_\xi / N = \mathbf{W}_\xi'\mathbf{W}_\xi / N$ & $\mathbf{W}_\xi'\mathbf{W}_\xi / N$ is bounded and strictly positive definite with determinant $> \gamma > 0$, while $(\mathbf{W}_\xi'\mathbf{W}_\xi / N)^{-1}$ is bounded.

(c) x_{ip} & $w_{iq}\xi_i$ almost surely satisfy condition Ib of Theorem I for all column pairs p of **X** and q of $\mathbf{W}_\xi$, while $\tilde{\mathbf{X}}'\tilde{\mathbf{X}} / N$ & $\tilde{\mathbf{W}}_\xi'\tilde{\mathbf{W}}_\xi / N$ are bounded with determinant $> \gamma > 0$ for all N sufficiently large, so that across the row permutations **T** of **X** we have

$$\left(\frac{\tilde{\mathbf{X}}'\tilde{\mathbf{X}}}{N} \otimes \frac{\tilde{\mathbf{W}}_\xi'\tilde{\mathbf{W}}_\xi}{N}\right)^{-½} \frac{(\tilde{\mathbf{T}}\bullet\tilde{\mathbf{W}}_\xi)'\mathbf{1}_N}{\sqrt{N}} \xrightarrow{d} \mathbf{n}_{PQ}, \text{ where } \mathbf{n}_{PQ} \sim \mathrm{N}(\mathbf{0}_{PQ}, \mathbf{I}_{PQ}).$$

(d) For some a in (0,½) condition IIIb of Theorem III almost surely holds for the mean of the product of the elements of n = 1, 2, 3 or 4 of the columns of **T** divided by $N^{a\max(n-2,0)}$ with the elements of four columns of $\mathbf{E} = (\mathbf{Z}_+,\boldsymbol{\xi})$, no more than two of which are $\boldsymbol{\xi}$, so that across permutations **T** of **X**

$$m(N^{-a\max(n-2,0)}(\prod_{o=1}^{n} t_{ip(o)})e_{ij}e_{ik}e_{il}e_{im}) - m(N^{-a\max(n-2,0)}\prod_{o=1}^{n} x_{ip(o)})m(e_{ij}e_{ik}e_{il}e_{im}) \xrightarrow{p} 0.$$

As elsewhere in this paper, almost sure limits are with respect to the data sequence $(\mathbf{X}_\mathbf{W},\mathbf{Z},\boldsymbol{\varepsilon})$, while probability limits and limiting distributions are with respect to the probability distribution generated by the $N!$ equally likely row permutations **T** of **X**.

(a) Asymptotic Distribution of Coefficient Estimates

Multiplying (K.2) by $\sqrt{N}$, we have

$$\text{(K.4)}\quad \sqrt{N}(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0) - \boldsymbol{\beta}_0) = \left(\frac{\mathbf{T}_\mathbf{W}'\mathbf{M}\mathbf{T}_\mathbf{W}}{N}\right)^{-1}\frac{\mathbf{T}_\mathbf{W}'\boldsymbol{\xi}}{\sqrt{N}},$$

From (B.5) in the paper we know that:

$$\text{(K.5)}\quad \frac{\mathbf{T}_{\mathbf{W}}'\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N} - \frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\mathbf{W}'\mathbf{W}}{N} \xrightarrow{p} \mathbf{0}_{PQ\,x\,PQ}.$$

The k^{th} term of $\mathbf{T}_{\mathbf{W}}'\boldsymbol{\xi}/\sqrt{N}$ equals:

$$\text{(K.6)}\quad \frac{\mathbf{t}_{\mathbf{W}k}'\boldsymbol{\xi}}{\sqrt{N}} = \sum_{i=1}^{N} \frac{[t_{ip(k)} - m(t_{ip(k)})][w_{iq(k)}\xi_i - m(w_{iq(k)}\xi_i)]}{\sqrt{N}} + \sqrt{N}\, m(x_{ip(k)})m(w_{iq(k)}\xi_i).$$

However,

$$\text{(K.7)}\quad m(w_{iq(k)}\xi_i) = \frac{\mathbf{w}_{q(k)}'[\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})\mathbf{Z}'][\mathbf{X}_{\mathbf{W}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \boldsymbol{\xi}]}{N} = 0 \text{ almost surely for } N \text{ sufficiently large}$$

$$\text{as } \frac{\mathbf{w}_{q(k)}'\mathbf{Z}}{N}\left(\frac{\mathbf{Z}'\mathbf{Z}}{N}\right)^{-1}\mathbf{Z}' = \mathbf{w}_{q(k)}' \text{ almost surely for } N \text{ sufficiently large (Lemma 1a),}$$

where because of A2 for all N sufficiently large that $\mathbf{Z}'\mathbf{Z}/N$ is guaranteed to be invertible $\mathbf{w}_{q(k)}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ is a row vector of zeros with a 1 in the column corresponding to the position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$. Consequently, the second term in (K.6) is zero for sufficiently large N and applying Lemma K1c we have

$$\text{(K.8)}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\widetilde{\mathbf{W}}_{\xi}'\widetilde{\mathbf{W}}_{\xi}}{N}\right)^{-½} \frac{\mathbf{T}_{\mathbf{W}}'\boldsymbol{\xi}}{\sqrt{N}} \xrightarrow{d} \mathbf{n}_{PQ}, \text{ where } \mathbf{n}_{PQ} \sim \mathrm{N}(\mathbf{0}_{PQ}, \mathbf{I}_{PQ}),$$

so that

$$\text{(K.9)}\quad \overbrace{\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\widetilde{\mathbf{W}}_{\xi}'\widetilde{\mathbf{W}}_{\xi}}{N}\right)^{-½}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\mathbf{W}'\mathbf{W}}{N}\right)}^{\text{almost surely bounded positive definite matrices (Lemmas 1a-1c)}} \sqrt{N}(\hat{\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\beta}_0) - \boldsymbol{\beta}_0) =$$

$$\underbrace{\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\widetilde{\mathbf{W}}_{\xi}'\widetilde{\mathbf{W}}_{\xi}}{N}\right)^{-½}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\mathbf{W}'\mathbf{W}}{N}\right)\overbrace{\left(\frac{\mathbf{T}_{\mathbf{W}}'\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N}\right)^{-1}}^{\xrightarrow{p}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\mathbf{W}'\mathbf{W}}{N}\right)^{-1}}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\widetilde{\mathbf{W}}_{\xi}'\widetilde{\mathbf{W}}_{\xi}}{N}\right)^{½}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N} \otimes \frac{\widetilde{\mathbf{W}}_{\xi}'\widetilde{\mathbf{W}}_{\xi}}{N}\right)^{-½}\frac{\mathbf{T}_{\mathbf{W}}'\boldsymbol{\xi}}{\sqrt{N}}}_{\xrightarrow{p}\mathbf{I}_{PQ} \qquad\qquad \xrightarrow{d}\mathbf{n}_{PQ}} \xrightarrow{d} \mathbf{n}_{PQ}.$$

(b) Probability Limit of the Heteroskedasticity Robust Covariance Estimate

For the heteroskedasticity robust covariance estimate we have

$$\text{(K.10)}\ N\mathbf{V}_r(\hat{\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\beta}_0)) = \left(\frac{\mathbf{T}_{\mathbf{W}}'\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N}\right)^{-1}\mathbf{A}\left(\frac{\mathbf{T}_{\mathbf{W}}'\mathbf{M}\mathbf{T}_{\mathbf{W}}}{N}\right)^{-1}, \text{ where } \mathbf{A} = \frac{(\mathbf{M}\mathbf{T}_{\mathbf{W}} \bullet \hat{\boldsymbol{\varepsilon}}(\mathbf{T}, \boldsymbol{\beta}_0))'(\mathbf{M}\mathbf{T}_{\mathbf{W}} \bullet \hat{\boldsymbol{\varepsilon}}(\mathbf{T}, \boldsymbol{\beta}_0))}{N}.$$

Using the formula for $\hat{\boldsymbol{\varepsilon}}(\mathbf{T}, \boldsymbol{\beta}_0)$ from (K.3) earlier, the kl^{th} term of $\mathbf{A}$ is given by

$$\text{(K.11)}\quad \mathbf{A}_{kl} = \frac{1}{N}\sum_{i=1}^{N}(t_{ip(k)}w_{iq(k)} - \sum_{a=1}^{K} z_{ia}\hat{\delta}_{ak})(t_{ip(l)}w_{iq(l)} - \sum_{b=1}^{K} z_{ib}\hat{\delta}_{bl})\left[\xi_i - \sum_{c=1}^{PQ}(t_{ip(c)}w_{iq(c)} - \sum_{d=1}^{K} z_{id}\hat{\delta}_{dc})\frac{\hat{r}_c}{\sqrt{N}}\right]^2$$

with $\hat{\mathbf{r}} = \sqrt{N}(\hat{\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\beta}_0) - \boldsymbol{\beta}_0)$ & $\hat{\boldsymbol{\delta}}_k = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{t}_{\mathbf{W}k}$. From (B.13) in the paper the plim of $\hat{\boldsymbol{\delta}}_k$ is known to equal 0 unless a is the column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$, in which case plim $\hat{\delta}_{ak} - m(x_{ip(k)}) = 0$. From (K.9), the elements of $\hat{\mathbf{r}}$ are asymptotically multivariate normal with bounded variance, so when divided by any positive power of N have a probability limit of zero.

When (K.11) is multiplied out, all terms multiplied by an element of $\hat{\mathbf{r}}$ involve the mean of the product of the elements of 0 to 4 columns of **T** and the elements of 4 columns of **E** = ($\mathbf{Z}_{+}$,$\boldsymbol{\xi}$), no more than two of which are $\boldsymbol{\xi}$. From Lemma 1c and K1a we know that the sample means of the product of the elements of one through four columns of **X** or four columns of **E** are almost surely bounded. Consequently, using Lemma K1d, in (K.11) every term that involves the product of an element of $\hat{\mathbf{r}}/\sqrt{N}$ that has a plim of zero with the mean of the product of four columns of **E** with zero, one or two columns of **T** (and also possibly with an element of bounded $\hat{\boldsymbol{\delta}}_k$) has a probability limit of zero. Every term in (K.11) that involves the product of n = 3 or 4 columns of **T** with four columns of **E** also includes at least n - 2 $\hat{\mathbf{r}}/\sqrt{N}$ terms which can be re-expressed as $(\hat{\mathbf{r}}/N^{½-a})(1/N^{a})$ for some a in (0,½). The $1/N^{a}$ part can be used to satisfy Lemma K1d, while from (K.9) the $\hat{\mathbf{r}}/N^{½-a}$ part converges in probability to 0. Thus, all such terms also have a plim of 0.

The above only leaves terms in (K.11) that do not include an element of $\hat{\mathbf{r}}/\sqrt{N}$ namely

$$\text{(K.12)}\ \sum_{i=1}^{N}\frac{t_{ip(k)}w_{iq(k)}t_{ip(l)}w_{iq(l)}\xi_i^2}{N}-\sum_{a=1}^{K}\hat{\delta}_{ak}\sum_{i=1}^{N}\frac{t_{ip(l)}w_{iq(l)}z_{ia}\xi_i^2}{N}-\sum_{b=1}^{K}\hat{\delta}_{bl}\sum_{i=1}^{N}\frac{t_{ip(k)}w_{iq(k)}z_{ib}\xi_i^2}{N}+\sum_{a=1}^{K}\sum_{b=1}^{K}\hat{\delta}_{ak}\hat{\delta}_{bl}\sum_{i=1}^{N}\frac{z_{ia}z_{ib}\xi_i^2}{N}$$

$$=m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}\xi_i^2)-\sum_{a=1}^{K}\hat{\delta}_{ak}m(t_{ip(l)}w_{iq(l)}z_{ia}\xi_i^2)-\sum_{b=1}^{K}\hat{\delta}_{bl}m(t_{ip(k)}w_{iq(k)}z_{ib}\xi_i^2)+\sum_{a=1}^{K}\sum_{b=1}^{K}\hat{\delta}_{ak}\hat{\delta}_{bl}m(z_{ia}z_{ib}\xi_i^2)$$

$$\text{where}\ \ m(t_{ip(k)}t_{ip(l)}w_{iq(k)}w_{iq(l)}\xi_i^2)-m(x_{ip(k)}x_{ip(l)})m(w_{iq(k)}w_{iq(l)}\xi_i^2)\ \underbrace{\xrightarrow{p}0}_{\text{Lemma K1d}}$$

$$\&\ \ m(t_{ip(l)}w_{iq(l)}z_{ia}\xi_i^2)-m(x_{ip(l)})m(w_{iq(l)}z_{ia}\xi_i^2)\ \underbrace{\xrightarrow{p}0}_{\text{Lemma K1d}}\ ,$$

$$\text{so}\ \mathbf{A}_{kl}-[m(x_{ip(k)}x_{ip(l)})-m(x_{ip(k)})m(x_{ip(l)})]m(w_{iq(k)}w_{iq(l)}\xi_i^2)\xrightarrow{p}0,$$

where we recall the boundedness of means of products of up to four terms (Lemma 1c & K1a) and the fact noted above that plim $\hat{\delta}_{ak}$ = 0 unless a is the column position of $\mathbf{w}_{q(k)}$ in **Z**, in which case plim $\hat{\delta}_{ak}=m(x_{ip(k)})$ and $z_{ia}=w_{iq(k)}$. This allows us to state that

$$\text{(K.13)}\ \ \mathbf{A}-\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}_{\xi}'\mathbf{W}_{\xi}}{N}\xrightarrow{p}\mathbf{0}_{PQxPQ}$$

and consequently for the heteroskedasticity robust covariance estimate we have

$$\text{(K.14)}\ N\mathbf{V}_r(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0))-\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}'\mathbf{W}}{N}\right)^{-1}\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}_{\xi}'\mathbf{W}_{\xi}}{N}\right)\left(\frac{\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}}{N}\otimes\frac{\mathbf{W}'\mathbf{W}}{N}\right)^{-1}\xrightarrow{p}\mathbf{0}_{PQxPQ}\ ,$$

which from (K.9) and Lemma K1b is seen to be the asymptotic covariance matrix of normally distributed $\sqrt{N}(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0)-\boldsymbol{\beta}_0)$. This establishes that the distribution of the Wald statistic $\tau(\mathbf{T},\boldsymbol{\beta}_0)$ across permutations **T** converges to that of the chi-squared with PQ degrees of freedom.

(c) Proof of Lemma K

Lemma K1a: Regarding the means of products of four columns of **E** = ($\mathbf{Z}_{+}$,$\boldsymbol{\xi}$), we note that

$$\text{(K.15)}\quad \xi_i = \varepsilon_i - \sum_{a=1}^{K} z_{ia}\hat{\tau}_a + \sum_{b=1}^{PQ}\left(x_{ip(b)}w_{iq(b)} - \sum_{c=1}^{K} z_{ic}\hat{\mathbf{\Phi}}_{cb}\right)(\beta_b - \beta_{0b})$$

where $\hat{\boldsymbol{\tau}} = (\mathbf{Z}'\mathbf{Z}/N)^{-1}(\mathbf{Z}'\boldsymbol{\varepsilon}/N)$ & $\hat{\mathbf{\Phi}} = (\mathbf{Z}'\mathbf{Z}/N)^{-1}(\mathbf{Z}'\mathbf{X}_{\mathbf{W}}/N)$. From Lemma 1a-1c in the paper we know that the elements of $\hat{\mathbf{\Phi}}$ are almost surely bounded, those of $\hat{\boldsymbol{\tau}}$ almost surely converge to 0, and the mean of the product of four columns of $\mathbf{D} = (\mathbf{X}_{\mathbf{W}},\mathbf{Z},\boldsymbol{\varepsilon})$, no more than two of which are $\boldsymbol{\varepsilon}$, is bounded. β_b - β_{0b} is a constant. The mean of the product of four columns of $(\mathbf{Z}_+,\boldsymbol{\xi})$ (no more than two of which are $\boldsymbol{\xi}$) is made up of the sum of the means of products of four columns of $(\mathbf{X}_{\mathbf{W}},\mathbf{Z},\boldsymbol{\varepsilon})$ (no more than two of which are $\boldsymbol{\varepsilon}$) times terms from $\hat{\boldsymbol{\tau}}$, $\hat{\mathbf{\Phi}}$ and $\boldsymbol{\beta}$ - $\boldsymbol{\beta}_0$, and hence, by the results just noted, is almost surely bounded.

Lemma K1b: (K.7) showed that $m(w_{iq}\xi_i)$ equals zero for all N sufficiently large, which establishes $\widetilde{\mathbf{W}}_{\xi}'\widetilde{\mathbf{W}}_{\xi}/N = \mathbf{W}_{\xi}'\mathbf{W}_{\xi}/N$ for such N. From Lemma K1a we see that the elements of $\mathbf{W}_{\xi}'\mathbf{W}_{\xi}/N$, formed of the means of the product of four columns of $(\mathbf{Z}_+,\boldsymbol{\xi})$, two of which are $\boldsymbol{\xi}$, are almost surely bounded. With regards to the determinant, by the properties of the Rayleigh quotient we know that if $\boldsymbol{\alpha}'\mathbf{W}_{\xi}'\mathbf{W}_{\xi}\boldsymbol{\alpha}/N > \gamma > 0$ for all $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$, then $\mathbf{W}_{\xi}'\mathbf{W}_{\xi}/N$ is positive definite with determinant greater than $\gamma^{Q} > 0$. From the above:

$$\text{(K.16)}\quad \frac{\boldsymbol{\alpha}'\mathbf{W}_{\xi}'\mathbf{W}_{\xi}\boldsymbol{\alpha}}{N} = N^{-1}\sum_{i=1}^{N}\omega_i(\varepsilon_i - \mathbf{z}_i'\hat{\boldsymbol{\tau}} + (\mathbf{x}_{\mathbf{W}i}' - \mathbf{z}_i'\hat{\mathbf{\Phi}})(\boldsymbol{\beta} - \boldsymbol{\beta}_0))^2 \text{ where } \omega_i = \boldsymbol{\alpha}'\mathbf{w}_i\mathbf{w}_i'\boldsymbol{\alpha},$$

$$\hat{\boldsymbol{\tau}} \overset{a.s.}{\rightarrow} \left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{z}_i')}{N}\right)^{-1}\left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\varepsilon_i)}{N}\right) = \mathbf{0}_K \quad \& \quad \hat{\mathbf{\Phi}} - \mathbf{\Phi} \overset{a.s.}{\rightarrow} \mathbf{0}_{KxPQ}, \text{ where } \mathbf{\Phi} = \left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{z}_i')}{N}\right)^{-1}\left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{x}_{\mathbf{W}i}')}{N}\right),$$

and the almost sure limits follow from the Markov Corollary given in the paper and use of the moment conditions in (W1)-(W3): $E(|z_{+ij}z_{+ik}|^{1+\delta}) < \Delta$, $E(\mathbf{z}_{+i}\varepsilon_i) = \mathbf{0}_{K+}$, & (using Jensen's Inequality) $E(|\varepsilon_i z_{+ik}|^{1+\delta}) \leq E(|\varepsilon_i^2 z_{+ik}^2|^{1+\delta})^{1/2} < \Delta^{1/2}$. By (W2), the matrix inverse in $\mathbf{\Phi}$ is known to exist, as it is a sub-matrix of $\mathbf{M}_N$, and the elements of $\mathbf{\Phi}$ are also known to be bounded.[16] $\boldsymbol{\alpha}$ is a vector of finite constants. When multiplied out, the remaining components of (K.16) are seen to be the means of four columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$, no more than two of which are $\boldsymbol{\varepsilon}$. Using (W3), (W4) and Hölder's Inequality

$$\text{(K.17)}\quad E(|z_{+ij}z_{+ik}\varepsilon_i^2|^{1+\delta}) \leq \sqrt{\prod_{a=j,k}E(|z_{+ia}^2\varepsilon_i^2|^{1+\delta})} < \Delta, \quad E(|z_{+ij}z_{+ik}z_{+il}z_{+im}|^{1+\delta}) \leq \sqrt[4]{\prod_{a=j,k,l,m}E(|z_{+ia}^4|^{1+\delta})} < \Delta,$$

$$\& \quad E(|z_{+ij}z_{+ik}z_{+il}\varepsilon_i|^{1+\delta}) \leq \sqrt[4]{E(|z_{+ij}^4|^{1+\delta})E(|z_{+ik}^4|^{1+\delta})}\sqrt{E(|z_{+il}^2\varepsilon_i^2|^{1+\delta})} < \Delta,$$

so by the Markov Corollary these means converge almost surely to the mean of their expectations and

$$\text{(K.18)}\quad \frac{\boldsymbol{\alpha}'\mathbf{W}_{\xi}'\mathbf{W}_{\xi}\boldsymbol{\alpha}}{N} - h_N \overset{a.s.}{\rightarrow} 0, \text{ where } h_N = \sum_{i=1}^{N}\frac{E(\omega_i\varepsilon_i^2)}{N} + 2(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\sum_{i=1}^{N}\frac{E(\omega_i\varepsilon_i\breve{\mathbf{x}}_{\mathbf{W}i})}{N} + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\sum_{i=1}^{N}\frac{E(\omega_i\breve{\mathbf{x}}_{\mathbf{W}i}\breve{\mathbf{x}}_{\mathbf{W}i}')}{N}(\boldsymbol{\beta} - \boldsymbol{\beta}_0).$$

where $\breve{\mathbf{x}}_{\mathbf{W}i}' = \mathbf{x}_{\mathbf{W}i}' - \mathbf{z}_i'\mathbf{\Phi}$. Minimizing the right hand side with respect to $\boldsymbol{\beta}_0$ using the fact that the mean of $E(\omega_i\breve{\mathbf{x}}_{\mathbf{W}i}\breve{\mathbf{x}}_{\mathbf{W}i}')$ is positive definite and invertible for all N sufficiently large (assumption AA1b above),

[16] As the trace of $\mathbf{M}_N$ is bounded from above by $K_+\Delta$ and its determinant from below by γ, its smallest eigenvalue is greater than $\lambda = \gamma/(K_+\,\Delta^{1/(1+\delta)})\text{^}(K_+\text{-}1)$ and the largest eigenvalue of the inverse in $\mathbf{\Phi}$ is bounded by λ^{-1}.

$$\text{(K.19)}\quad \boldsymbol{\beta}-\boldsymbol{\beta}_0=-\left(\sum_{i=1}^{N}\frac{E(\omega_i\breve{\mathbf{x}}_{\mathbf{W}i}\breve{\mathbf{x}}'_{\mathbf{W}i})}{N}\right)^{-1}\sum_{i=1}^{N}\frac{E(\omega_i\varepsilon_i\breve{\mathbf{x}}_{\mathbf{W}i})}{N},$$

$$\text{so}\quad h_N\geq\sum_{i=1}^{N}\frac{E(\omega_i\varepsilon_i^2)}{N}-\sum_{i=1}^{N}\frac{E(\omega_i\varepsilon_i\breve{\mathbf{x}}'_{\mathbf{W}i})}{N}\left(\sum_{i=1}^{N}\frac{E(\omega_i\breve{\mathbf{x}}_{\mathbf{W}i}\breve{\mathbf{x}}'_{\mathbf{W}i})}{N}\right)^{-1}\sum_{i=1}^{N}\frac{E(\omega_i\varepsilon_i\breve{\mathbf{x}}_{\mathbf{W}i})}{N}>\gamma>0,$$

where we use (AA1b) in the last as well. This establishes that the minimum eigenvalue of $\mathbf{W}'_\xi\mathbf{W}_\xi/N$ is almost surely $>\gamma>0$ for all N sufficiently large, so its determinant is greater than some $\gamma^Q>0$ and, as the eigenvalues of its inverse are bounded from above by the inverse of its smallest eigenvalue, that $(\mathbf{W}'_\xi\mathbf{W}_\xi/N)^{-1}$ is also almost surely bounded.

Lemma K1c: Lemmas 1a & 1c in the paper and K1b above already established that $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N$ & $\widetilde{\mathbf{W}}'_\xi\widetilde{\mathbf{W}}_\xi/N$ are almost surely bounded with determinant $>\gamma>0$ for all N sufficiently large, so all that remains is condition Ib. Define $w_{iq\xi}=w_{iq}\xi_i$ and, as elsewhere, let superscripted ~ denote sample demeaned values. Our objective is to prove that for all integer $\tau>2$ and all p and q

$$\text{(K.20)}\quad N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\widetilde{x}_{ip}^{\tau}\sum_{i=1}^{N}\widetilde{w}_{iq\xi}^{\tau}\Bigg/\left(\sum_{i=1}^{N}\widetilde{x}_{ip}^{2}\right)^{\tau/2}\left(\sum_{i=1}^{N}\widetilde{w}_{iq\xi}^{2}\right)^{\tau/2}\xrightarrow{a.s.}0.$$

We begin by noting that:

$$\text{(K.21)}\quad \left|\frac{N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\widetilde{x}_{ip}^{\tau}\sum_{i=1}^{N}\widetilde{w}_{iq\xi}^{\tau}}{\left(\sum_{i=1}^{N}\widetilde{x}_{ip}^{2}\sum_{i=1}^{N}\widetilde{w}_{iq\xi}^{2}\right)^{\tau/2}}\right|\leq\left|\frac{N^{\frac{\tau}{2}-1}\left(\underset{i\leq N}{\text{Max}}\,\widetilde{x}_{ip}^{2}\,\underset{i\leq N}{\text{Max}}\,\widetilde{w}_{iq\xi}^{2}\right)^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\widetilde{x}_{ip}^{2}\sum_{i=1}^{N}\widetilde{w}_{iq\xi}^{2}}{\left(\sum_{i=1}^{N}\widetilde{x}_{ip}^{2}\sum_{i=1}^{N}\widetilde{w}_{iq\xi}^{2}\right)^{\tau/2}}\right|=\left|\frac{\frac{\underset{i\leq N}{\text{Max}}\,\widetilde{x}_{ip}^{2}\,\underset{i\leq N}{\text{Max}}\,\widetilde{w}_{iq\xi}^{2}}{N}}{\sum_{i=1}^{N}\frac{\widetilde{x}_{ip}^{2}}{N}\sum_{i=1}^{N}\frac{\widetilde{w}_{iq\xi}^{2}}{N}}\right|^{\frac{\tau}{2}-1}.$$

As noted in the paper's appendix, for a K x K matrix with determinant $>\gamma>0$ and non-negative diagonal elements bounded from above by Δ', the smallest eigenvalue is bounded from below by $\lambda(K)=\gamma/(K\Delta')\text{^}(K-1)$. By the Schur-Horn Theorem, the smallest diagonal element of a real symmetric matrix is greater than or equal to its smallest eigenvalue. Consequently, given the properties already established for $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}/N$ & $\widetilde{\mathbf{W}}'_\xi\widetilde{\mathbf{W}}_\xi/N$ we know the smallest diagonal elements of these matrices are almost surely greater than $\lambda(P)$ and $\lambda(Q)$, establishing that the denominator of (K.21) is almost surely bounded away from zero.

Turning to the numerator, since for any sequence d_i

$$\text{(K.22)}\ \underset{i\leq N}{\text{Max}}\,\widetilde{d}_i^{\,2}\leq\underset{i\leq N}{\text{Max}}\,d_i^2+2\sqrt{\underset{i\leq N}{\text{Max}}(d_i^2)}\left|m(d_i)\right|+m(d_i)^2,$$

while almost surely $m(w_{iq\xi})=m(w_{iq}\xi_i)=0$ for N sufficiently large (K.7 earlier) and $m(x_{ip})$ is bounded (Lemma 1c), to prove (Ib) all that remains is to show that $\text{Max}_{i\leq N}x_{ip}^2\text{Max}_{i\leq N}w_{iq\xi}^2/N\xrightarrow{a.s.}0$. Using (K.15) and recalling that $x_{ip(b)}w_{ip(b)}$ is an element of $\mathbf{z}_{+i}$, we see that $w_{iq\xi}^2$ is made up of the sum of the product of terms $h_i=w_{iq}^2\varepsilon_i^2$, $w_{iq}^2\varepsilon_i z_{+ij}$, or $w_{iq}^2 z_{+ij}z_{+ik}$ times almost surely bounded elements of $\hat{\boldsymbol{\tau}}$, $\hat{\boldsymbol{\Phi}}$ and $\boldsymbol{\beta}$ - $\boldsymbol{\beta}_0$. From (AA3) above and Hölder's Inequality we have $E(|w_{iq}^2z_{+ij}z_{+ik}|^{1+\theta})<E(|w_{iq}^2z_{+ij}^2|^{1+\theta})^{½}E(|w_{iq}^2z_{+ik}^2|^{1+\theta})^{½}<\Delta$, $E(|w_{iq}^2\varepsilon_i^2|^{1+\theta})<\Delta$, $E(|w_{iq}^2\varepsilon_i z_{+ij}|^{1+\theta})<E(|w_{iq}^2\varepsilon_i^2|^{1+\theta})^{½}E(|w_{iq}^2z_{+ij}^2|^{1+\theta})^{½}<\Delta$, & $E(|x_i^4|^{1+\theta^*})<\Delta$ with $\theta(1+2\theta^*)>1$. Applying Markov's Inequality

$$\text{(K.23)}\ \sum_{N=1}^{\infty}\Pr\{x_{Np}^2 \geq N^a\} = \sum_{N=1}^{\infty}\Pr\{x_{Np}^4 \geq N^{2a}\} \leq \sum_{N=1}^{\infty}\frac{\Delta}{N^{2a(1+\theta^*)}} < \infty \ \text{ if } \ 2a(1+\theta^*) > 1$$

$$\& \ \sum_{N=1}^{\infty}\Pr\{h_N \geq N^b\} \leq \sum_{N=1}^{\infty}\frac{\Delta}{N^{b(1+\theta)}} < \infty \ \text{ if } \ b(1+\theta) > 1.$$

Both conditions can be met with $a > 0$, $b > 0$ and $a + b < 1$ if $\theta(1+2\theta^*) > 1$ as

$$\text{(K.24)}\ 1 > a+b > \frac{1}{2(1+\theta^*)} + \frac{1}{1+\theta} = 1 - \frac{\theta(1+2\theta^*)-1}{2(1+\theta^*)(1+\theta)}$$

poses no contradiction. Applying the Borel-Cantelli Corollary in the paper, we see that $\text{Max}_{i\leq N}x_{ip}^2\text{Max}_{i\leq N}w_{iq\xi}^2 / N^{a+b}$ is almost surely bounded by 1, so $\text{Max}_{i\leq N}x_{ip}^2\text{Max}_{i\leq N}w_{iq\xi}^2 / N \xrightarrow{a.s.} 0.$

Lemma K1d: The sample means of 1 through 4 columns of **X** and any 4 columns of **E** = (**Z**$_+$,**ξ**) (no more than two of which are **ξ**) are almost surely bounded (Lemma 1c and K1a), so to establish condition IIIb it suffices that there exists an a in (0,½) such that the following are almost surely bounded

$$\text{(K.25)}\ m\left(\frac{x_{ij}^2x_{ik}^2x_{il}^2}{N^{2a}}\right) \leq \max_{i\leq N}\frac{x_{ij}^2}{N^{2a}}m(x_{ik}^2x_{il}^2),\ m\left(\frac{x_{ij}^2x_{ik}^2x_{il}^2x_{im}^2}{N^{4a}}\right) \leq \max_{i\leq N}\frac{x_{ij}^2}{N^{2a}}\max_{i\leq N}\frac{x_{im}^2}{N^{2a}}m(x_{ik}^2x_{il}^2)$$

$$\& \ \sum_{i=1}^{N}\frac{e_{ij}^2e_{ik}^2e_{il}^2e_{im}^2}{N^2} \leq \max_{i\leq N}\frac{e_{ij}^2e_{ik}^2}{N}m(e_{il}^2e_{im}^2)$$

where we select indices so that when two elements are ξ_i, e_{ij} represents one and e_{il} the other. The proof of Lemma 2 in the paper's appendix already established the condition for the products of x_{ij} above. Using (K.15) we know that when $e_{ij} = \xi_i$, $e_{ij}^2e_{ik}^2$ is made up of the sum of the product of terms $h_i = z_{+ij}^2\varepsilon_i^2$, $z_{ij}^2\varepsilon_iz_{+ik}$, or $z_{+ij}^2z_{+ik}z_{+il}$ times almost surely bounded elements of $\hat{\boldsymbol{\tau}}$, $\hat{\boldsymbol{\Phi}}$ and $\boldsymbol{\beta}$ - $\boldsymbol{\beta}_0$, and otherwise we can say that $h_i = e_{ij}^2e_{ik}^2 = z_{+ij}^2z_{+ik}^2$. From (W3), (W4) & Hölder's Inequality we have $E(|\, z_{+ij}^2\varepsilon_i^2\,|^{1+\delta}) < \Delta$, $E(|\, z_{+ij}^2z_{+ik}z_{+il}\,|^{1+\delta}) < \Delta$, $E(|\, z_{+ij}^2\varepsilon_iz_{+ik}\,|^{1+\delta}) < E(|\, z_{+ij}^2\varepsilon_i^2\,|^{1+\delta})^{½}E(|\, z_{+ij}^2z_{+ik}^2\,|^{1+\delta})^{½} < \Delta$, & $E(|\, z_{ij}^2z_{ik}^2\,|^{1+\delta}) < \Delta$. Applying Markov's Inequality

$$\text{(K.26)}\ \sum_{N=1}^{\infty}\Pr\{h_N \geq N^a\} \leq \sum_{N=1}^{\infty}\frac{\Delta}{N^{a(1+\delta)}} < \infty \ \text{ if } \ a(1+\delta) > 1.$$

As $\delta > 0$, we know that there exists an $a < 1$ such that (K.26) holds, so by the Borel-Cantelli Corollary for any of the h_i described above $\max_{i\leq N} h_i / N^a$ is almost surely bounded by 1, and hence $\max_{i\leq N} e_{ij}^2e_{ik}^2 / N \xrightarrow{a.s.} 0$, establishing IIIb.

L. Convergence in Distribution for any $\boldsymbol{\beta}_0$ (Grouped Treatment) (notation follows Appendix B)

In this appendix we prove (RR1), as given in Appendix K above, for grouped treatment. The framework and notation follows Appendix B above, but we change assumptions (A1), (A3) & (U5) given therein to read:

(AA1) There exists a finite positive constant γ such that (a) $\mathbf{G}_M = \sum_{m=1}^{M} E(\boldsymbol{\varkappa}_m \boldsymbol{\varkappa}_m')/M - \sum_{m=1}^{M} E(\boldsymbol{\varkappa}_m)/M \sum_{m=1}^{M} E(\boldsymbol{\varkappa}_m')/M$ is non-singular for all M sufficiently large with determinant $\mathbf{G}_M > \gamma > 0$; (b) with $\boldsymbol{\Phi}_N = (\sum_{i=1}^{N} E(\mathbf{z}_i \mathbf{z}_i')/N)^{-1} \sum_{i=1}^{N} E(\mathbf{z}_i \mathbf{x}_{\mathbf{W}i}')/N$, $\breve{\mathbf{X}}_{\mathbf{W}m}' = \mathbf{X}_{\mathbf{W}m} - \mathbf{Z}_m \boldsymbol{\Phi}$ & $\boldsymbol{\omega}_m = \mathbf{W}_m \boldsymbol{\alpha}$, for all $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$ & for all M sufficiently large

$$\sum_{m=1}^{M} \frac{E(\boldsymbol{\omega}_m' \boldsymbol{\varepsilon}_m \boldsymbol{\omega}_m' \boldsymbol{\varepsilon}_m)}{M} - \sum_{m=1}^{M} \frac{E(\boldsymbol{\omega}_m' \boldsymbol{\varepsilon}_m \boldsymbol{\omega}_m' \breve{\mathbf{X}}_{\mathbf{W}m})}{M} \mathbf{K}^{-1} \sum_{m=1}^{M} \frac{E(\breve{\mathbf{X}}_{\mathbf{W}m}' \boldsymbol{\omega}_m \boldsymbol{\omega}_m' \boldsymbol{\varepsilon}_m)}{M} > \gamma > 0,$$

$$\text{where } \mathbf{K} = \sum_{m=1}^{M} \frac{E(\breve{\mathbf{X}}_{\mathbf{W}m}' \boldsymbol{\omega}_m \boldsymbol{\omega}_m' \breve{\mathbf{X}}_{\mathbf{W}m})}{M} \text{ with determinant } \mathbf{K} > \gamma > 0.$$

(AA3) There exist finite positive constants θ, θ^*, γ and Δ, with $\theta(1+2\theta^*) > 1$, such that (a) for all $m, j = 1...K_+$, $q = 1...Q$ and $p = 1...P$, $E(|\mathbf{w}_{mq}' \boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}_m' \mathbf{w}_{mq}|^{1+\theta}) < \Delta$, $E(|\mathbf{w}_{mq}' \mathbf{z}_{+mj} \boldsymbol{\varepsilon}_m' \mathbf{z}_{+mj}|^{1+\theta}) < \Delta$ and $E(|\boldsymbol{\varkappa}_{mp}^4|^{1+\theta^*}) < \Delta$; (b) $\mathbf{W}_M = M^{-1} \sum_{m=1}^{M} E(\mathbf{W}_m' \boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}_m' \mathbf{W}_m)$ is non-singular for all M sufficiently large, with determinant $\mathbf{W}_M > \gamma > 0$.

(UU5) Clustering is done at the treatment grouping level or above, i.e. treatment groups are contained within clusters, and, following assumption (U5c) of Appendix B above, errors are independently distributed across clusters with $E(\mathbf{z}_{+c_1 j}' \boldsymbol{\varepsilon}_{c_1} \mathbf{z}_{+c_2 k}' \boldsymbol{\varepsilon}_{c_2}) = 0$ for all $j, k = 1...K_+$ if cluster $c_1 \neq c_2$.

(AA1) and (AA3) are the grouped treatment versions of the extensions of (AA1) and (AA3) discussed in appendix **K**. The assumption $E(\mathbf{z}_{+c_1 j}' \boldsymbol{\varepsilon}_{c_1} \mathbf{z}_{+c_2 k}' \boldsymbol{\varepsilon}_{c_2}) = 0$ from Appendix B merely ensures that the conventional Wald statistic is asymptotically distributed chi-squared. Clustering must take place at the treatment level or above because a fixed deviation $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}$ contributes, via the original regressors $\mathbf{X}_\mathbf{W}$, to the error term of counterfactual outcomes $y(\mathbf{T}, \boldsymbol{\beta}_0)$. When treatment groups cut across clusters this introduces a correlation across cluster groups, resulting in a divergence between the variance of $\hat{\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\beta}_0)$ and that calculated using the clustered variance estimate, as shown analytically at the end of this appendix.

Turning to the proof of (RR1), the estimated coefficients and residuals associated with the null $\boldsymbol{\beta}_0$ and row permutations $\boldsymbol{\tau}$ of $\boldsymbol{\varkappa}$, producing observation level treatment measures **T**, are

$$\text{(L.1)} \quad \hat{\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\beta}_0) = (\mathbf{T}_\mathbf{W}' \mathbf{M} \mathbf{T}_\mathbf{W})^{-1} \mathbf{T}_\mathbf{W}' \mathbf{M} \mathbf{y}(\mathbf{T}, \boldsymbol{\beta}_0) = (\mathbf{T}_\mathbf{W}' \mathbf{M} \mathbf{T}_\mathbf{W})^{-1} \mathbf{T}_\mathbf{W}' \mathbf{M} \boldsymbol{\xi} + \boldsymbol{\beta}_0 \text{ where } \boldsymbol{\xi} = \mathbf{M} \mathbf{X}_\mathbf{W} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \mathbf{M} \boldsymbol{\varepsilon},$$

$$\text{(L.2)} \quad \hat{\boldsymbol{\varepsilon}}(\mathbf{T}, \boldsymbol{\beta}_0) = \mathbf{M} \mathbf{y}(\mathbf{T}, \boldsymbol{\beta}_0) - \mathbf{M} \mathbf{T}_\mathbf{W} \hat{\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\beta}_0) = \boldsymbol{\xi} - \mathbf{M} \mathbf{T}_\mathbf{W} (\hat{\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\beta}_0) - \boldsymbol{\beta}_0).$$

All Lemmas given and proven in Appendices B and C above continue to hold, as the moment conditions have, if anything, been strengthened, and will be referenced below as Lemma B2, Lemma B3, etc. The following additional Lemma, proven below, will also be useful:

Lemma L1: Define $\boldsymbol{\mathcal{W}}_\xi$ as the M x Q matrix whose mq^{th} element $\mathcal{w}_{mq\xi}$ is the sum of the observational elements corresponding to the m^{th} treatment group in the N x Q matrix $\mathbf{W}_\xi$ (i.e. $\mathcal{w}_{mq\xi} = \sum_{i\epsilon m} w_{iq} \xi_i$). If assumptions U1 - U4 and A1 - A4 as modified above hold, then

(a) With $e_{i1}e_{i2}$ and $e_{i3}e_{i4}$ each denoting the product of the elements of two columns of $\mathbf{E} = (\mathbf{Z}_+, \boldsymbol{\xi})$ (with at most one in each case being $\boldsymbol{\xi}$), $|m(e_{i1}e_{i2})|$, $|m(e_{i1}e_{i2}e_{i3}e_{i4})|$, $|m_c(e_{i1}e_{i2}, e_{i3}e_{i4})|$, $|m_m(e_{i1}e_{i2}, e_{i3}e_{i4})|$ & $|m_v(e_{i1}e_{i2}, e_{i3}e_{i4})|$ are all almost surely bounded.

(b) Almost surely for all M sufficiently large $\tilde{\boldsymbol{\mathcal{w}}}_\xi' \tilde{\boldsymbol{\mathcal{w}}}_\xi / M = \boldsymbol{\mathcal{w}}_\xi' \boldsymbol{\mathcal{w}}_\xi / M$ & $\boldsymbol{\mathcal{w}}_\xi' \boldsymbol{\mathcal{w}}_\xi / M$ is bounded and strictly positive definite with determinant $> \gamma > 0$, while $(\boldsymbol{\mathcal{w}}_\xi' \boldsymbol{\mathcal{w}}_\xi / M)^{-1}$ is bounded.

(c) $\mathcal{x}_{mp}$ & $\mathcal{w}_{mq\xi}$ almost surely satisfy condition Ib of Theorem I for all column pairs of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{W}}_\xi$, while $\tilde{\boldsymbol{\mathcal{x}}}' \tilde{\boldsymbol{\mathcal{x}}} / M$ & $\tilde{\boldsymbol{\mathcal{w}}}_\xi' \tilde{\boldsymbol{\mathcal{w}}}_\xi / M$ are almost surely bounded with determinant $> \gamma > 0$ for all M sufficiently large, so that across the row permutations $\boldsymbol{\mathcal{T}}$ of $\boldsymbol{\mathcal{X}}$ we have

$$\left(\frac{\tilde{\boldsymbol{\mathcal{x}}}' \tilde{\boldsymbol{\mathcal{x}}}}{M} \otimes \frac{\tilde{\boldsymbol{\mathcal{w}}}_\xi' \tilde{\boldsymbol{\mathcal{w}}}_\xi}{M} \right)^{-1/2} \frac{(\tilde{\boldsymbol{\mathcal{T}}} \bullet \tilde{\boldsymbol{\mathcal{w}}}_\xi)' \mathbf{1}_{PQ}}{\sqrt{M}} \xrightarrow{d} \mathbf{n}_{PQ}, \text{ where } \mathbf{n}_{PQ} \sim \mathrm{N}(\mathbf{0}_{PQ}, \mathbf{I}_{PQ}).$$

(d) With $t_{i1} \ldots t_{i4}$ denoting columns of $\mathbf{T}$ and u_{i1} and u_{j2} each the product of the elements of two columns of $\mathbf{E} = (\mathbf{Z}_+, \boldsymbol{\xi})$, no more than one of which is $\boldsymbol{\xi}$, for some a in (0,½)

(d1) $m_c(t_{i1}u_{i1}, u_{j2}) - \omega(\mathcal{x}_{m1}) m_c(u_{i1}, u_{j2}) \xrightarrow{p} 0$ & $m_c(t_{i1}t_{i2}u_{i1}, u_{j2}) - \omega(\mathcal{x}_{m1}\mathcal{x}_{m2}) m_c(u_{i1}, u_{j2}) \xrightarrow{p} 0$

(d2) $m_c(t_{i1}u_{i1}, t_{j2}u_{j2}) - \left([\omega(\mathcal{x}_{m1}\mathcal{x}_{m2}) - \omega(\mathcal{x}_{m1})\omega(\mathcal{x}_{m2})] m_v(u_{i1}, u_{j2}) + \omega(\mathcal{x}_{m1})\omega(\mathcal{x}_{m2}) m_c(u_{i1}, u_{j2}) \right) \xrightarrow{p} 0,$

(d3) $M^{-a} m_c(t_{i1}t_{i2}u_{i1}, t_{j3}u_{j2}) \xrightarrow{p} 0$ & $M^{-2a} m_c(t_{i1}t_{i2}u_{i1}, t_{j3}t_{j4}u_{j2}) \xrightarrow{p} 0.$

As elsewhere in this paper, almost sure limits are with respect to the data sequence $(\mathbf{X_W}, \mathbf{Z}, \boldsymbol{\varepsilon})$, while probability limits and limiting distributions are with respect to the probability distribution generated by the $M!$ equally likely row permutations $\boldsymbol{\mathcal{T}}$ of $\boldsymbol{\mathcal{X}}$.

(a) Asymptotic Distribution of Coefficient Estimates

Multiplying (L.2) by $\sqrt{M}$, we have

$$\text{(L.3)} \quad \sqrt{M}\,(\hat{\boldsymbol{\beta}}(\mathbf{T}, \boldsymbol{\beta}_0) - \boldsymbol{\beta}_0) = \left(\frac{\mathbf{T}_\mathbf{W}' \mathbf{M} \mathbf{T}_\mathbf{W}}{M} \right)^{-1} \frac{\mathbf{T}_\mathbf{W}' \boldsymbol{\xi}}{\sqrt{M}},$$

From (B.9) in Appendix B above we know that:

$$\text{(L.4)} \quad \frac{\mathbf{T}_\mathbf{W}' \mathbf{M} \mathbf{T}_\mathbf{W}}{M} - \frac{\tilde{\boldsymbol{\mathcal{x}}}' \tilde{\boldsymbol{\mathcal{x}}}}{M} \otimes \frac{\mathbf{W}'\mathbf{W}}{M} \xrightarrow{p} \mathbf{0}_{PQxPQ}.$$

The remaining part of (L.3) is the vector $\mathbf{T}_\mathbf{W}' \boldsymbol{\xi} / \sqrt{M}$, the k^{th} term of which equals:

$$\text{(L.5)} \quad \frac{\mathbf{t}_{\mathbf{W}k}' \boldsymbol{\xi}}{\sqrt{M}} = \sum_{i=1}^{N} \frac{t_{ip(k)} w_{iq(k)} \xi_i}{\sqrt{M}} = \sum_{i=1}^{M} \frac{t_{mp(k)} \sum_{i\epsilon m} w_{iq(k)} \xi_i}{\sqrt{M}} = \sum_{i=1}^{M} \frac{t_{mp(k)} \mathcal{w}_{mq(k)\xi}}{\sqrt{M}}$$

$$= \sum_{m=1}^{M} \frac{[t_{mp(k)} - \omega(t_{mp(k)})][\mathcal{w}_{mq(k)\xi} - \omega(\mathcal{w}_{mq(k)\xi})]}{\sqrt{M}} + \sqrt{M}\,\omega(t_{mp(k)})\omega(\mathcal{w}_{mq(k)\xi}).$$

However,

$$\text{(L.6)}\quad \omega(w_{mq(k)\xi}) = m(w_{iq(k)}\xi_i) = \frac{\mathbf{w}'_{q(k)}[\mathbf{I}_N - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})\mathbf{Z}'][\mathbf{X}_\mathbf{W}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\boldsymbol{\varepsilon}]}{M} = 0 \text{ a.s. for } M \text{ sufficiently large}$$

$$\text{as } \frac{\mathbf{w}'_{q(k)}\mathbf{Z}}{M}\left(\frac{\mathbf{Z}'\mathbf{Z}}{M}\right)^{-1}\mathbf{Z}' = \mathbf{w}'_{q(k)} \text{ almost surely for } M \text{ sufficiently large (Lemma B2a).}$$

where, as elsewhere, we make use of the fact that $\mathbf{w}_{q(k)}$ is an element of $\mathbf{Z}$. Consequently, for sufficiently large M the second term in (L.5) is zero and $\mathbf{T}'_\mathbf{W}\boldsymbol{\xi}/\sqrt{M} = (\tilde{\boldsymbol{\mathcal{T}}} \bullet \tilde{\boldsymbol{\mathcal{W}}}_\xi)'\mathbf{1}_{PQ}/\sqrt{M}$. Applying Lemma L1c:

$$\text{(L.7)}\left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M}\otimes\frac{\tilde{\boldsymbol{\mathcal{W}}}'_\xi\tilde{\boldsymbol{\mathcal{W}}}_\xi}{M}\right)^{-½}\frac{\mathbf{T}'_\mathbf{W}\boldsymbol{\xi}}{\sqrt{M}}\xrightarrow{d}\mathbf{n}_{PQ},\ \text{where } \mathbf{n}_{PQ}\sim \mathrm{N}(\mathbf{0}_{PQ},\mathbf{I}_{PQ}).$$

Combining the preceding results, we see that

$$\overbrace{\left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M}\otimes\frac{\tilde{\boldsymbol{\mathcal{W}}}'_\xi\tilde{\boldsymbol{\mathcal{W}}}_\xi}{M}\right)^{-½}\left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)}^{\text{almost surely bounded and positive definite (Lemmas B2a.2c, L1b)}}\sqrt{M}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}-\boldsymbol{\beta}_0) =$$

$$\underbrace{\left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M}\otimes\frac{\tilde{\boldsymbol{\mathcal{W}}}'_\xi\tilde{\boldsymbol{\mathcal{W}}}_\xi}{M}\right)^{-½}\left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)\left(\frac{\mathbf{T}'_\mathbf{W}\mathbf{M}\mathbf{T}_\mathbf{W}}{M}\right)^{-1}\left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M}\otimes\frac{\tilde{\boldsymbol{\mathcal{W}}}'_\xi\tilde{\boldsymbol{\mathcal{W}}}_\xi}{M}\right)^{½}}_{\xrightarrow{p}\mathbf{I}_{PQ}}\underbrace{\left(\frac{\tilde{\boldsymbol{\mathcal{X}}}'\tilde{\boldsymbol{\mathcal{X}}}}{M}\otimes\frac{\tilde{\boldsymbol{\mathcal{W}}}'_\xi\tilde{\boldsymbol{\mathcal{W}}}_\xi}{M}\right)^{-½}\frac{\mathbf{T}'_\mathbf{W}\boldsymbol{\xi}}{\sqrt{M}}}_{\xrightarrow{d}\mathbf{n}_{PQ}}\xrightarrow{d}\mathbf{n}_{PQ}.$$

(b) Probability Limit of the Clustered Covariance Estimate

For the clustered covariance estimate we again have the sandwich formula,

$$\text{(L.9)}\ M\mathbf{V}_{cl}(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0)) = \left(\frac{\mathbf{T}'_\mathbf{W}\mathbf{M}\mathbf{T}_\mathbf{W}}{M}\right)^{-1}\mathbf{A}\left(\frac{\mathbf{T}'_\mathbf{W}\mathbf{M}\mathbf{T}_\mathbf{W}}{M}\right)^{-1}$$

but with the kl^{th} element of $\mathbf{A}$ this time given by

$$\text{(L.10)}\ \mathbf{A}_{kl} = \frac{1}{M}\sum_{c=1}^{C}\left(\sum_{i\in c}(t_{ip(k)}w_{iq(k)}-\sum_{a=1}^{K}z_{ia}\hat{\delta}_{ak})\left[\xi_i-\sum_{c=1}^{PQ}(t_{ip(c)}w_{iq(c)}-\sum_{d=1}^{K}z_{id}\hat{\delta}_{dc})\frac{\hat{r}_c}{\sqrt{M}}\right]\right)*$$
$$\left(\sum_{j\in c}(t_{jp(l)}w_{jq(l)}-\sum_{b=1}^{K}z_{jb}\hat{\delta}_{bl})\left[\xi_j-\sum_{e=1}^{PQ}(t_{jp(e)}w_{jq(e)}-\sum_{f=1}^{K}z_{jf}\hat{\delta}_{fe})\frac{\hat{r}_e}{\sqrt{M}}\right]\right),$$

with $\hat{\mathbf{r}} = \sqrt{M}(\hat{\boldsymbol{\beta}}(\mathbf{T},\boldsymbol{\beta}_0)-\boldsymbol{\beta}_0)$ & $\hat{\boldsymbol{\delta}}_k = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{t}_{\mathbf{W}k}$. From (B.20) in Appendix B above the plim of $\hat{\boldsymbol{\delta}}_k$ is known to equal 0 unless a is the column position of $\mathbf{w}_{q(k)}$ in $\mathbf{Z}$, in which case plim $\hat{\delta}_{ak}-\omega(x_{mp(k)})=0$. From (L.8) the elements of $\hat{\mathbf{r}}$ are asymptotically multivariate normal with bounded variance, so when divided by a positive power of M have a probability limit of zero.

When (L.10) is multiplied out, all terms multiplied by an element of $\hat{\mathbf{r}}/\sqrt{M}$ involve the m_c means of the product of the elements of 0 to 4 columns of $\mathbf{T}$ and u_{i1} and u_{j2} each the product of the elements of two columns of $\mathbf{E} = (\mathbf{Z}_+,\boldsymbol{\xi})$, no more than one of which in each case is $\boldsymbol{\xi}$, as in $m_c(u_{i1},u_{j2})$, $m_c(t_{i1}u_{i1},u_{j2})$, $m_c(t_{i1}u_{i1},t_{j2}u_{j2})$, $m_c(t_{i1}t_{i2}u_{i1},u_{j2})$, $m_c(t_{i1}t_{i2}u_{i1},t_{j3}u_{j2})$ and $m_c(t_{i1}t_{i2}u_{i1},t_{j3}t_{j4}u_{j2})$, where t_{i1} .. t_{i4} represent columns of $\mathbf{T}$. From Lemma B2c in Appendix B above and Lemma L1a we know that $|\omega(\prod_{k=1}^{n}x_{mk})|$ for n = 1..4 and $m_c(u_{i1},u_{j2})$ are almost surely bounded. Consequently, using Lemma L1d, in (L.10) every term that

involves the product of an element of $\hat{\mathbf{r}}/\sqrt{M}$ that has a plim of zero with the mean of the product of four columns of **E** with zero, one or two columns of **T** (and also possibly with an element of bounded $\hat{\boldsymbol{\delta}}_k$) has a probability limit of zero. Every term in (L.10) that involves the product of n = 3 or 4 columns of **T** with four columns of **E** also includes at least n - 2 $\hat{\mathbf{r}}/\sqrt{M}$ terms which can be re-expressed as $(\hat{\mathbf{r}}/M^{½-a})(1/M^{a})$ for some a in (0,½). The $1/M^{a}$ part can be used to satisfy Lemma L1d, while from (L.8) the $\hat{\mathbf{r}}/M^{½-a}$ part converges in probability to 0. Thus, all such terms also have a plim of 0.

This only leaves terms that do not include $\hat{\mathbf{r}}/\sqrt{M}$, namely

$$\text{(L.11)}\ \underbrace{\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{t_{ip(k)}w_{iq(k)}\xi_i t_{jp(l)}w_{jq(l)}\xi_j}{M}}_{m_c(t_{ip(k)}w_{iq(k)}\xi_i,t_{jp(l)}w_{jq(l)}\xi_j)}-\sum_{b=1}^{K}\hat{\delta}_{bl}\underbrace{\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{t_{ip(k)}w_{iq(k)}\xi_i z_{jb}\xi_j}{M}}_{m_c(t_{ip(k)}w_{iq(k)}\xi_i,z_{jb}\xi_j)}$$
$$-\sum_{a=1}^{K}\hat{\delta}_{ak}\underbrace{\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{z_{ia}\xi_i t_{jp(l)}w_{jq(l)}\xi_j}{M}}_{m_c(t_{ip(l)}w_{iq(l)}\xi_i,z_{ja}\xi_j)}+\sum_{a=1}^{K}\sum_{b=1}^{K}\hat{\delta}_{ak}\hat{\delta}_{bl}\underbrace{\sum_{c=1}^{C}\sum_{i\in c}\sum_{j\in c}\frac{z_{ia}\xi_i z_{jb}\xi_j}{M}}_{m_c(z_{ia}\xi_i,z_{jb}\xi_j)},$$

so using Lemma L1d we see that

$$\text{(L.12)}\ \mathbf{A}_{kl}-[\omega(\boldsymbol{\varkappa}_{mp(k)}\boldsymbol{\varkappa}_{mp(l)})-\omega(\boldsymbol{\varkappa}_{mp(k)})\omega(\boldsymbol{\varkappa}_{mp(l)})]m_v(w_{iq(k)}\xi_i,w_{jq(l)}\xi_j)\xrightarrow{p}0,$$

where we once again use the fact that plim $\hat{\delta}_{ak}$ is only non-zero in the column position of $\mathbf{w}_{q(k)}$ in **Z**, when plim $\hat{\delta}_{ak}=\omega(\boldsymbol{\varkappa}_{mp(k)})$ and $z_{ia}=w_{iq(k)}$. Consequently, for the clustered robust covariance estimate

$$\text{(L.13)}\ \mathbf{A}-\frac{\tilde{\boldsymbol{\varkappa}}'\tilde{\boldsymbol{\varkappa}}}{M}\otimes\frac{\tilde{\boldsymbol{\omega}}_\xi'\tilde{\boldsymbol{\omega}}_\xi}{M}\xrightarrow{p}\mathbf{0}_{PQxPQ}\quad\left[\text{where}\ \frac{\tilde{\boldsymbol{\omega}}_\xi'\tilde{\boldsymbol{\omega}}_\xi}{M}=\sum_{v=1}^{V}\sum_{i\in v}\sum_{j\in v}\frac{\mathbf{w}_i\xi_i\xi_j\mathbf{w}_j'}{M}=\sum_{v=1}^{V}\frac{\mathbf{W}_v'\boldsymbol{\xi}_v\boldsymbol{\xi}_v'\mathbf{W}_v}{M}\right]$$
$$\Rightarrow M\mathbf{V}_{cl}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})-\left(\frac{\tilde{\boldsymbol{\varkappa}}'\tilde{\boldsymbol{\varkappa}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)^{-1}\left(\frac{\tilde{\boldsymbol{\varkappa}}'\tilde{\boldsymbol{\varkappa}}}{M}\otimes\frac{\tilde{\boldsymbol{\omega}}_\xi'\tilde{\boldsymbol{\omega}}_\xi}{M}\right)\left(\frac{\tilde{\boldsymbol{\varkappa}}'\tilde{\boldsymbol{\varkappa}}}{M}\otimes\frac{\mathbf{W}'\mathbf{W}}{M}\right)^{-1}\xrightarrow{p}\mathbf{0}_{PQxPQ}.$$

We now note the following lemma:

Lemma L2: If U5 holds, then $\tilde{\boldsymbol{\omega}}_\xi'\tilde{\boldsymbol{\omega}}_\xi/M-\boldsymbol{\omega}_\xi'\boldsymbol{\omega}_\xi/M\xrightarrow{a.s.}\mathbf{0}_{QxQ}$.

From Lemma L2, (L.8) & (L.13) we see that when the errors are clustered, $M\hat{\mathbf{V}}_{cl}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0})$ converges in probability to the asymptotic covariance matrix of asymptotically normally distributed $\sqrt{M}(\hat{\boldsymbol{\beta}}_{\mathbf{T},\boldsymbol{\beta}_0}-\boldsymbol{\beta}_0)$, so the Wald statistic is asymptotically distributed chi-squared with PQ degrees of freedom.

(c) Proof of Lemmas

Lemma L1a: We note that

$$\text{(L.14)}\ \xi_i=\varepsilon_i-\sum_{a=1}^{K}z_{ia}\hat{\tau}_a+\sum_{b=1}^{PQ}\left(x_{ip(b)}w_{iq(b)}-\sum_{c=1}^{K}z_{ic}\hat{\boldsymbol{\Phi}}_{cb}\right)(\beta_b-\beta_{0b})$$

where $\hat{\boldsymbol{\tau}}=(\mathbf{Z}'\mathbf{Z}/M)^{-1}(\mathbf{Z}'\boldsymbol{\varepsilon}/M)$, $\hat{\boldsymbol{\Phi}}=(\mathbf{Z}'\mathbf{Z}/M)^{-1}(\mathbf{Z}'\mathbf{X}_{\mathbf{W}}/M)$ and β_b - β_{0b} is a constant. From Lemmas B2a-B2c in Appendix B above we know that the elements of $\hat{\boldsymbol{\Phi}}$ are almost surely bounded and those of $\hat{\boldsymbol{\tau}}$ almost surely converge to 0, while Lemma B2c also showed that for $d_{i1}d_{i2}$ and $d_{i3}d_{i4}$ each the product of the elements of two columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ (with at most one in each being $\boldsymbol{\varepsilon}$), $|\omega(\prod_{k=1}^{n}\boldsymbol{\varkappa}_{mk})|$, $|m(d_{i1}d_{i2})|$, $|m(d_{i1}d_{i2}d_{i3}d_{i4})|$, $|m_c(d_{i1}d_{i2},d_{i3}d_{i4})|$, $|m_m(d_{i1}d_{i2},d_{i3}d_{i4})|$ & $|m_v(d_{i1}d_{i2},d_{i3}d_{i4})|$ are all almost surely bounded. From

(L.14) we see that the m, m_c, m_m and m_v means of the product of $e_{i1}e_{i2}$ and $e_{i3}e_{i4}$, each denoting two columns of $(\mathbf{Z}_+,\boldsymbol{\xi})$ (with no more than one in each case being $\boldsymbol{\xi}$), are made up of corresponding means of products of four columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ (no more than two of which are $\boldsymbol{\varepsilon}$) times almost surely bounded terms $\hat{\boldsymbol{\tau}}$, $\hat{\boldsymbol{\Phi}}$ and $\boldsymbol{\beta}$ - $\boldsymbol{\beta}_0$, and hence, by the results just noted, are almost surely bounded.

Lemma L1b: (L.6) showed that $\omega(\mathcal{W}_{mq(k)\xi}) = m(w_{iq(k)}\xi_i) = 0$ for all M sufficiently large, which establishes $\tilde{\mathcal{W}}_{\xi}'\tilde{\mathcal{W}}_{\xi}/M = \mathcal{W}_{\xi}'\mathcal{W}_{\xi}/M$ for such M. From Lemma L1a we see that the elements of $\mathcal{W}_{\xi}'\mathcal{W}_{\xi}/M$, formed of the means $m_m(e_{i1}e_{i2}.e_{i3}e_{i4})$ of columns of $\mathbf{E} = (\mathbf{Z}_+,\boldsymbol{\xi})$, are almost surely bounded. With regards to the determinant, by the properties of the Rayleigh quotient we know that if $\boldsymbol{\alpha}'\mathcal{W}_{\xi}'\mathcal{W}_{\xi}\boldsymbol{\alpha}/M > \gamma > 0$ for all $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$, then $\mathcal{W}_{\xi}'\mathcal{W}_{\xi}/M$ is positive definite with determinant greater than $\gamma^Q > 0$. From the above, with $\omega_i = \boldsymbol{\alpha}'\mathbf{w}_i$:

$$\text{(L.15)}\quad \frac{\boldsymbol{\alpha}'\mathcal{W}_{\xi}'\mathcal{W}_{\xi}\boldsymbol{\alpha}}{M} = \sum_{m=1}^{M}\frac{\sum_{i\in m}\omega_i(\varepsilon_i - \mathbf{z}_i'\hat{\boldsymbol{\pi}} + (\mathbf{x}_{\mathbf{W}i}' - \mathbf{z}_i'\hat{\boldsymbol{\Phi}})(\boldsymbol{\beta}-\boldsymbol{\beta}_0))\sum_{j\in m}\omega_j(\varepsilon_j - \mathbf{z}_j'\hat{\boldsymbol{\pi}} + (\mathbf{x}_{\mathbf{W}j}' - \mathbf{z}_j'\hat{\boldsymbol{\Phi}})(\boldsymbol{\beta}-\boldsymbol{\beta}_0))}{M}$$

$$\text{where } \hat{\boldsymbol{\pi}} \xrightarrow{a.s.} \left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{z}_i')}{M}\right)^{-1}\left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\varepsilon_i)}{M}\right) = \mathbf{0}_K \;\;\&\;\; \hat{\boldsymbol{\Phi}} - \boldsymbol{\Phi} \xrightarrow{a.s.} \mathbf{0}_{KxPQ}, \text{ with } \boldsymbol{\Phi} = \left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{z}_i')}{M}\right)^{-1}\left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{x}_{\mathbf{W}i}')}{M}\right),$$

and the almost sure limits following from the Markov Corollary and, using assumption U2a, $E(|\,\mathbf{z}_{+uj}'\mathbf{z}_{+uk}\,|^{1+\delta}) < \Delta$ & $E(|\,\boldsymbol{\varepsilon}_u'\boldsymbol{\varepsilon}_u\,|^{1+\delta}) < \Delta$ which imply, from Hölder's Inequality, that $E(|\,\mathbf{z}_{+uj}'\boldsymbol{\varepsilon}_u\,|^{1+\delta}) < E(|\,\mathbf{z}_{+uj}'\mathbf{z}_{+uj}\,|^{1+\delta})^{½}E(|\,\boldsymbol{\varepsilon}_u'\boldsymbol{\varepsilon}_u\,|^{1+\delta})^{½} < \Delta$. When multiplied out, (L.15) is seen to consist of the sum of terms consisting of elements of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ - $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\pi}}$ & $\hat{\boldsymbol{\Phi}}$ times means of $\mathbf{d}_{m1}'\mathbf{d}_{m2}\mathbf{d}_{m3}'\mathbf{d}_{m4}$, with each $\mathbf{d}$ representing one of the columns of $(\mathbf{Z}_+,\boldsymbol{\varepsilon})$ with no more than one in each pair being $\boldsymbol{\varepsilon}$, and

$$\text{(L.16)}\quad E(|\,\mathbf{d}_{m1}'\mathbf{d}_{m2}\mathbf{d}_{m3}'\mathbf{d}_{m4}\,|^{1+\delta}) \overbrace{\leq \sqrt{E((\mathbf{d}_{m1}'\mathbf{d}_{m2})^{2(1+\delta)})E((\mathbf{d}_{m3}'\mathbf{d}_{m4})^{2(1+\delta)})}}^{\text{Hölder's Inequality}}$$

$$\overbrace{\leq \sqrt{E((\mathbf{d}_{m1}'\mathbf{d}_{m1}\mathbf{d}_{m2}'\mathbf{d}_{m2})^{1+\delta})E((\mathbf{d}_{m3}'\mathbf{d}_{m3}\mathbf{d}_{m4}'\mathbf{d}_{m4})^{1+\delta})}}^{\text{Cauchy-Schwarz Inequality}} \overbrace{\leq \sqrt{E((\mathbf{d}_{u1}'\mathbf{d}_{u1}\mathbf{d}_{u2}'\mathbf{d}_{u2})^{1+\delta})E((\mathbf{d}_{u3}'\mathbf{d}_{u3}\mathbf{d}_{u4}'\mathbf{d}_{u4})^{1+\delta})}}^{m\subset u} \overbrace{< \Delta}^{\text{U3 \& U4}},$$

so by the LLNHDS Corollary in Appendix C these means converge almost surely to the mean of their expectations and

$$\text{(L.17)}\quad \frac{\boldsymbol{\alpha}'\mathcal{W}_{\xi}'\mathcal{W}_{\xi}\boldsymbol{\alpha}}{M} - h_m \xrightarrow{a.s.} 0, \text{ where}$$

$$h_m = \sum_{m=1}^{M}\frac{E(\boldsymbol{\omega}_m'\boldsymbol{\varepsilon}_m\boldsymbol{\omega}_m'\boldsymbol{\varepsilon}_m)}{M} + 2(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\sum_{m=1}^{M}\frac{E(\breve{\mathbf{X}}_{\mathbf{W}m}'\boldsymbol{\omega}_m\boldsymbol{\omega}_m'\boldsymbol{\varepsilon}_m)}{M} + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\sum_{m=1}^{M}\frac{E(\breve{\mathbf{X}}_{\mathbf{W}m}'\boldsymbol{\omega}_m\boldsymbol{\omega}_m'\breve{\mathbf{X}}_{\mathbf{W}m})}{M}(\boldsymbol{\beta}-\boldsymbol{\beta}_0),$$

and $\breve{\mathbf{X}}_{\mathbf{W}m}' = \mathbf{X}_{\mathbf{W}m} - \mathbf{Z}_m\boldsymbol{\Phi}$. Minimizing h_m with respect to $\boldsymbol{\beta}_0$ using the fact that the mean of $E(\breve{\mathbf{X}}_{\mathbf{W}m}'\boldsymbol{\omega}_m\boldsymbol{\omega}_m'\breve{\mathbf{X}}_{\mathbf{W}m})$ is positive definite and invertible for all M sufficiently large (assumption AA1b),

$$\text{(L.18)}\quad \boldsymbol{\beta}-\boldsymbol{\beta}_0 = -\left(\sum_{m=1}^{M}\frac{E(\breve{\mathbf{X}}_{\mathbf{W}m}'\boldsymbol{\omega}_m\boldsymbol{\omega}_m'\breve{\mathbf{X}}_{\mathbf{W}m})}{M}\right)^{-1}\sum_{m=1}^{M}\frac{E(\breve{\mathbf{X}}_{\mathbf{W}m}'\boldsymbol{\omega}_m\boldsymbol{\omega}_m'\boldsymbol{\varepsilon}_m)}{M}, \text{ so}$$

$$h_m \geq \sum_{m=1}^{M}\frac{E(\boldsymbol{\omega}_m'\boldsymbol{\varepsilon}_m\boldsymbol{\omega}_m'\boldsymbol{\varepsilon}_m)}{M} - \sum_{m=1}^{M}\frac{E(\boldsymbol{\omega}_m'\boldsymbol{\varepsilon}_m\boldsymbol{\omega}_m'\breve{\mathbf{X}}_{\mathbf{W}m})}{M}\left(\sum_{m=1}^{M}\frac{E(\breve{\mathbf{X}}_{\mathbf{W}m}'\boldsymbol{\omega}_m\boldsymbol{\omega}_m'\breve{\mathbf{X}}_{\mathbf{W}m})}{M}\right)^{-1}\sum_{m=1}^{M}\frac{E(\breve{\mathbf{X}}_{\mathbf{W}m}'\boldsymbol{\omega}_m\boldsymbol{\omega}_m'\boldsymbol{\varepsilon}_m)}{M} > \gamma > 0,$$

which establishes that the minimum eigenvalue of $\mathbf{w}_\xi'\mathbf{w}_\xi/M$ is almost surely $> \gamma > 0$ for all M sufficiently large, so its determinant is greater than some $\gamma^Q > 0$ and, as the eigenvalues of its inverse are bounded from above by the inverse of its smallest eigenvalue, that $(\mathbf{w}_\xi'\mathbf{w}_\xi/M)^{-1}$ is also almost surely bounded.

Lemma L1c: Lemmas B2a, B2c from Appendix B & L1b above already established that $\tilde{\mathbf{x}}'\tilde{\mathbf{x}}/M$ & $\tilde{\mathbf{w}}_\xi'\tilde{\mathbf{w}}_\xi/M$ are almost surely bounded with determinant $> \gamma > 0$ for all M sufficiently large, so all that remains is condition Ib. Our objective is to prove that for all integer $\tau > 2$ and all p and q

$$\text{(L.19)}\quad M^{\frac{\tau}{2}-1}\sum_{m=1}^{M}\tilde{x}_{mp}^{\tau}\sum_{m=1}^{M}\tilde{w}_{mq\xi}^{\tau}\Bigg/\left(\sum_{m=1}^{M}\tilde{x}_{mp}^{2}\right)^{\tau/2}\left(\sum_{m=1}^{M}\tilde{w}_{mq\xi}^{2}\right)^{\tau/2}\xrightarrow{a.s.}0,$$

where $w_{mq\xi}=\sum_{i\in m}w_{iq}\xi_i$. We begin by noting that:

$$\text{(L.20)}\quad \left|\frac{M^{\frac{\tau}{2}-1}\sum_{m=1}^{M}\tilde{x}_{mp}^{\tau}\sum_{m=1}^{M}\tilde{w}_{mq\xi}^{\tau}}{\left(\sum_{m=1}^{M}\tilde{x}_{mp}^{2}\sum_{m=1}^{M}\tilde{w}_{mq\xi}^{2}\right)^{\tau/2}}\right|\le\left|\frac{M^{\frac{\tau}{2}-1}\left(\underset{m\le M}{\text{Max}}\,\tilde{x}_{mp}^{2}\,\underset{m\le M}{\text{Max}}\,\tilde{w}_{mq\xi}^{2}\right)^{\frac{\tau}{2}-1}\sum_{m=1}^{M}\tilde{x}_{mp}^{2}\sum_{m=1}^{M}\tilde{w}_{mq\xi}^{2}}{\left(\sum_{m=1}^{M}\tilde{x}_{mp}^{2}\sum_{m=1}^{M}\tilde{w}_{mq\xi}^{2}\right)^{\tau/2}}\right|=\left|\frac{\dfrac{\underset{m\le M}{\text{Max}}\,\tilde{x}_{mp}^{2}\,\underset{m\le M}{\text{Max}}\,\tilde{w}_{mq\xi}^{2}}{M}}{\sum_{m=1}^{M}\dfrac{\tilde{x}_{mp}^{2}}{M}\sum_{m=1}^{M}\dfrac{\tilde{w}_{mq\xi}^{2}}{M}}\right|^{\frac{\tau}{2}-1}.$$

As noted earlier, for a K x K matrix with determinant $> \gamma > 0$ and non-negative diagonal elements bounded from above by Δ', the smallest eigenvalue is bounded from below by $\lambda(K) = \gamma/(K\Delta')\text{^}(K\text{-}1)$. By the Schur-Horn Theorem, the smallest diagonal element of a real symmetric matrix is greater than or equal to its smallest eigenvalue. Consequently, given the properties already established for $\tilde{\mathbf{x}}'\tilde{\mathbf{x}}/M$ & $\tilde{\mathbf{w}}_\xi'\tilde{\mathbf{w}}_\xi/M$ we know the smallest diagonal elements of these matrices are almost surely greater than $\lambda(P)$ and $\lambda(Q)$, establishing that the denominator of (L.20) is almost surely bounded away from zero.

Turning to the numerator, since for any sequence d_m

$$\text{(L.21)}\quad \underset{m\le M}{\text{Max}}\,\tilde{d}_m^{2}\le\underset{m\le M}{\text{Max}}\,d_m^{2}+2\sqrt{\underset{m\le M}{\text{Max}}(d_m^{2})}\left|\omega(d_m)\right|+\omega(d_m)^{2}$$

and almost surely $\omega(w_{iq\xi}) = m(w_{iq\xi}) = 0$ for all M sufficiently large (L.6 earlier) and $\omega(x_{mp})$ is bounded (Lemma B2c), to prove (Ib) all that remains is to show that $\text{Max}_{m\le M}x_{mp}^2\text{Max}_{m\le M}w_{mq\xi}^2/M\xrightarrow{a.s.}0$. Using (L.14) and recalling that $x_{ip(b)}w_{ip(b)}$ is an element of $\mathbf{z}_{+i}$, we see that $w_{mq\xi}^2$ is made up of the sum of the product of terms $h_m = (\mathbf{w}_{mq}'\boldsymbol{\varepsilon}_m)^2$, $\mathbf{w}_{mq}'\boldsymbol{\varepsilon}_m\mathbf{w}_{mq}'\mathbf{z}_{+mj}$, or $\mathbf{w}_{mq}'\mathbf{z}_{+mj}\mathbf{w}_{mq}'\mathbf{z}_{+mk}$ times almost surely bounded elements of $\hat{\boldsymbol{\tau}}$, $\hat{\boldsymbol{\Phi}}$ and $\boldsymbol{\beta}$ - $\boldsymbol{\beta}_0$. From (AA3) above and Hölder's Inequality we have $E((\mathbf{w}_{mq}'\boldsymbol{\varepsilon}_m)^{2(1+\theta)})<\Delta$, $E(|\mathbf{w}_{mq}'\boldsymbol{\varepsilon}_m\mathbf{w}_{mq}'\mathbf{z}_{+mj}|^{1+\theta})<E((\mathbf{w}_{mq}'\boldsymbol{\varepsilon}_m)^{2(1+\theta)})^{½}E((\mathbf{w}_{mq}'\mathbf{z}_{+mj})^{2(1+\theta)})^{½}<\Delta$, $E(|\mathbf{w}_{mq}'\mathbf{z}_{+mj}\mathbf{w}_{mq}'\mathbf{z}_{+mk}|^{1+\theta})<E((\mathbf{w}_{mq}'\mathbf{z}_{+mj})^{2(1+\theta)})^{½}E((\mathbf{w}_{mq}'\mathbf{z}_{+mk})^{2(1+\theta)})^{½},<\Delta$ & $E(|x_{mp}^4|^{1+\theta^*})<\Delta$ with $\theta(1+2\theta^*)>1$. Consequently, applying Markov's Inequality

$$\text{(L.22)}\quad \sum_{M=1}^{\infty}P(x_{Mp}^2\ge M^a)=\sum_{M=1}^{\infty}P(x_{Mp}^4\ge M^{2a})\le\sum_{M=1}^{\infty}\frac{\Delta}{M^{2a(1+\theta^*)}}<\infty\ \text{ if }\ 2a(1+\theta^*)>1$$

$$\&\ \sum_{M=1}^{\infty}P(h_M\ge M^b)\le\sum_{M=1}^{\infty}\frac{\Delta}{M^{b(1+\theta)}}<\infty\ \text{ if }\ b(1+\theta)>1.$$

Both conditions can be met with $a > 0$, $b > 0$ and $a + b < 1$ if $\theta(1+2\theta^*) > 1$ as

$$\text{(L.23)}\quad 1 > a + b > \frac{1}{2(1+\theta^*)} + \frac{1}{1+\theta} = 1 - \frac{\theta(1+2\theta^*)-1}{2(1+\theta^*)(1+\theta)}$$

poses no contradiction. From the Borel-Cantelli Lemma Corollary, we see that $\text{Max}_{m\le M}\boldsymbol{x}_{mp}^2\text{Max}_{m\le M}\boldsymbol{w}_{mq\xi}^2 / M^{a+b}$ is almost surely bounded by 1, so $\text{Max}_{m\le M}\boldsymbol{x}_{mp}^2\text{Max}_{m\le M}\boldsymbol{w}_{mq\xi}^2 / M \xrightarrow{a.s.} 0$.

Lemma L1d: Lemma B7 in Appendix B above proved analogous results for the means of products of columns of **T** and v_{i1} and v_{j2} each denoting the product of the elements of two columns of ($\mathbf{Z}_+$,**ε**). From (L.14) above, we see that all terms in Lemma L1d can be expressed as the sum of means of the type seen in Lemma B7, possibly times elements of $\hat{\boldsymbol{\tau}}$, $\hat{\boldsymbol{\Phi}}$ and $\boldsymbol{\beta} - \boldsymbol{\beta}_0$ which are almost surely bounded. Applying Lemma B7 to each of these terms and summing up we get the results of Lemma L1d.

Lemma L2: Above we saw that $\tilde{\boldsymbol{w}}_\xi'\tilde{\boldsymbol{w}}_\xi / M = \boldsymbol{w}_\xi'\boldsymbol{w}_\xi / M$ for M sufficiently large. Based upon the moment results in (L.16) and the LLNHDS of Appendix C, as in the proof of Lemma L1b we have that

$$\text{(L.24)}\quad \frac{\boldsymbol{w}_\xi'\boldsymbol{w}_\xi}{M} - \mathbf{H}_m \xrightarrow{a.s.} \mathbf{0}_{Q\times Q}, \text{ where}$$

$$\mathbf{H}_m = \sum_{m=1}^{M}\frac{E(\mathbf{W}_m'\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m'\mathbf{W}_m)}{M} + 2\sum_{m=1}^{M}\frac{E(\mathbf{W}_m'\breve{\mathbf{X}}_{\mathbf{W}m}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\boldsymbol{\varepsilon}_m'\mathbf{W}_m)}{M} + \sum_{m=1}^{M}\frac{E(\mathbf{W}_m'\breve{\mathbf{X}}_{\mathbf{W}m}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\breve{\mathbf{X}}_{\mathbf{W}m}'\mathbf{W}_m)}{M}.$$

Since the intersection groupings v are a subset of the union grouping u, a similar appeal to (L.16) and the LLNHDS establishes that:

$$\text{(L.25)}\quad \frac{\boldsymbol{w}_\xi'\boldsymbol{w}_\xi}{M} - \mathbf{H}_v \xrightarrow{a.s.} \mathbf{0}_{Q\times Q}, \text{ where}$$

$$\mathbf{H}_v = \sum_{m=1}^{M}\frac{E(\mathbf{W}_v'\boldsymbol{\varepsilon}_v\boldsymbol{\varepsilon}_v'\mathbf{W}_v)}{M} + 2\sum_{m=1}^{M}\frac{E(\mathbf{W}_v'\breve{\mathbf{X}}_{\mathbf{W}v}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\boldsymbol{\varepsilon}_v'\mathbf{W}_v)}{M} + \sum_{m=1}^{M}\frac{E(\mathbf{W}_v'\breve{\mathbf{X}}_{\mathbf{W}v}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\breve{\mathbf{X}}_{\mathbf{W}v}'\mathbf{W}_v)}{M}.$$

When clustering at or above the treatment grouping level, the intersection groupings v equal the treatment groupings m, so $\mathbf{H}_m = \mathbf{H}_v$, completing the proof.

If cluster groupings are across or below treatment levels, the intersection groupings v are subsets of the treatment groupings m, which means that for (L.24) and (L.25) to be equal the sum of expectations of products of observations from the same treatment groups that are in different clusters in (L.24) must equal zero. This is unlikely to be the case. As treatment is generally iid and independent of the other regressors **Z**, we have:

$$\text{(L.26)}\quad \boldsymbol{\Phi} = \left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{z}_i')}{N}\right)^{-1}\left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i(\mathbf{x}_i \bullet \mathbf{w}_i)')}{N}\right) = \left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{z}_i')}{N}\right)^{-1} E(\mathbf{x}') \otimes \left(\sum_{i=1}^{N}\frac{E(\mathbf{z}_i\mathbf{w}_i')}{N}\right) = E(\mathbf{x}') \otimes \boldsymbol{\Psi},$$

where, as **W** is part of **Z**, **Ψ** is a matrix of 0s with a diagonal matrix of 1s along the diagonal position of **W** in **Z**. Consequently, $\mathbf{X}_\mathbf{W} - \mathbf{Z}\boldsymbol{\Phi} = [\mathbf{X} - \mathbf{1}_N E(\mathbf{x}')]\bullet\mathbf{W}$. This means that for all observations within treatment groups that cut across cluster groupings the diagonal elements of the last summation in $\mathbf{H}_m$ involve the non-zero variance of the treatment measures. Unless the expectation of the product of terms from **W** in such cases is zero, which is unlikely to true if treatment is iid and independent of the regressors, these expectations will not be 0 and so $\mathbf{H}_m$ will not equal $\mathbf{H}_v$. In contrast, as was seen in Appendix B above,

when $\boldsymbol{\beta}_0$ is root-N local to $\boldsymbol{\beta}$, the last two summations in (L.24) and (L.25) can be eliminated and if $E(\mathbf{z}'_{+c_1 j}\boldsymbol{\varepsilon}_{c_1}\mathbf{z}'_{+c_2 k}\boldsymbol{\varepsilon}_{c_2}) = 0$ for $c_1 \neq c_2$, $\mathbf{H}_m$ equals $\mathbf{H}_v$. If the ε_i are mean zero and independent across clusters, this condition can be met although, in the case of treatment groups that cut across clusters, it would require that there are no treatment related heterogeneous effects included in the residuals. This merely follows the analysis in Appendix B, which shows that if the conditions on the errors are such that the homoskedastic, heteroskedastic or clustered (at any level) covariance estimate allow for asymptotically accurate conventional inference, then randomization inference using Wald statistics based upon the same covariance estimate is equally asymptotically accurate. In contrast, when $\boldsymbol{\beta}_0$ is no longer root-N local to $\boldsymbol{\beta}$, as in this appendix, if clustering takes place below or across treatment groupings randomization inference based upon Wald statistics is likely to be inaccurate, even when conventional inference is not.

M. References in On-Line Appendix not included in Paper's Bibliography

Fang, Yuguang, Kenneth A. Loparo & Xiangbo Feng (1994). "Inequalities for the Trace of a Matrix Product." *IEEE Transactions on Automatic Control* 39 (12): 2489-2490.

Hoadley, Bruce (1971). "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case." *The Annals of Mathematical Statistics* 42 (6): 1977-1991.

White, Halbert (1980a). "Nonlinear Regression on Cross-Section Data." *Econometrica* 48 (3): 721-746.