On-Line Appendix for: "Consistency of the OLS Bootstrap for Independently but Not-Identically Distributed Data: A Permutation Perspective."

Alwyn Young, London School of Economics, October 2023

A. Multivariate Extension of Theorem II: pp. 1 - 11.

B. Proof of (L1p), (L2), (L3) and (9) for the Pairs Bootstrap: pp. 12 - 18.

C. Consistency of the Pairs Bootstrap with Sub-Sampling: pp. 19 - 27

A. Multivariate Extension of Theorem II

Following the presentation in the paper, let $\boldsymbol{\delta}' = (d_1, \ldots, d_N)$ denote a sequence of real numbers, $\mathbf{Z}' = (\mathbf{z}_1,...,\mathbf{z}_N)$ a sequence of K x 1 real vectors, and $\mathbf{O} = \mathbf{I}_{NxN} - \mathbf{1}_N\mathbf{1}'_N/N$ the centering matrix. We wish to show that across the row permutations **d** of **δ**

$$\text{(A1)}\quad \mathbf{v}(\mathbf{Z},\boldsymbol{\delta}) = \left(\frac{\mathbf{Z}'\mathbf{O}\mathbf{Z}}{N}\frac{\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-1/2}\frac{(\mathbf{Z}'\mathbf{O}\mathbf{d})}{\sqrt{N}}$$

is asymptotically distributed multivariate iid standard normal if for all N sufficiently large $\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}'$ is non-zero and the correlation matrix $\text{diag}(\mathbf{Z}'\mathbf{O}\mathbf{Z})^{-1/2}\,\mathbf{Z}'\mathbf{O}\mathbf{Z}\,\text{diag}(\mathbf{Z}'\mathbf{O}\mathbf{Z})^{-1/2}$, where $\text{diag}(\mathbf{Z}'\mathbf{O}\mathbf{Z})$ is the diagonal matrix with entries equal to the diagonal of $\mathbf{Z}'\mathbf{O}\mathbf{Z}$, is non-singular with determinant $> \Delta$ (a positive constant), while

$$\text{(A2)}\quad \lim_{N\to\infty}\frac{N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}[z_{ik}-m(z_{ik})]^{\tau}\sum_{i=1}^{N}[d_i-m(d_i)]^{\tau}}{\left(\sum_{i=1}^{N}[z_{ik}-m(z_{ik})]^2\right)^{\tau/2}\left(\sum_{i=1}^{N}[d_i-m(d_i)]^2\right)^{\tau/2}} = 0$$

for each element z_{ik} in the vector sequence $\mathbf{z}_i$. Hoeffding (1951) provides a proof for a broader, but univariate, permutation problem. The generalization to the multivariate case requires additional notation, 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

$$\text{(A3)}\ \ \tilde{\mathbf{Z}} = \mathbf{OZ}\left(\frac{\mathbf{Z'OZ}}{N}\right)^{-1/2} \ \ \&\ \ \tilde{\mathbf{d}} = \mathbf{Od}\left(\frac{\mathbf{d'Od}}{N}\right)^{-1/2},\ \text{so that}\ \mathbf{v}(\mathbf{Z},\boldsymbol{\delta}) = \frac{\tilde{\mathbf{Z}}'\tilde{\mathbf{d}}}{\sqrt{N}}.$$

For the k^{th} element of $\mathbf{v}$ we know that

$$\text{(A4)}\ \ v_k = \sum_{i=1}^{N}\frac{\tilde{z}_{ik}\tilde{d}_i}{N^{1/2}},\ \text{where}\sum_{i=1}^{N}\tilde{z}_{ik} = \sum_{i=1}^{N}\tilde{d}_i = 0\ \ \&\ \sum_{i=1}^{N}\tilde{z}_{ik}^2 = \sum_{i=1}^{N}\tilde{d}_i^2 = N\ \text{ for all } k,$$

$$\text{and}\ \sum_{i=1}^{N}\tilde{z}_{ik_1}\tilde{z}_{ik_2} = 0\ \ \text{for all } k_1 \neq k_2,\ \text{as}\ \tilde{\mathbf{Z}}'\mathbf{1}_N = \mathbf{0}_K,\ \tilde{\mathbf{d}}'\mathbf{1}_N = 0,\ \tilde{\mathbf{Z}}'\tilde{\mathbf{Z}} = N*\mathbf{I}_{NxN}\ \ \&\ \tilde{\mathbf{d}}'\tilde{\mathbf{d}} = N.$$

We shall show that all of the moments of the vector $\mathbf{v}$ 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 $\mathbf{d}$ is the row permutation of $\boldsymbol{\delta}$, $\tilde{\mathbf{d}} = \mathbf{Od}(\mathbf{d'Od}/N)^{-1/2}$ is simply the row permutation of $\tilde{\boldsymbol{\delta}} = \mathbf{O}\boldsymbol{\delta}(\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}/N)^{-1/2}$ and the sample moments of $\tilde{\mathbf{d}}$ are the same as those of $\tilde{\boldsymbol{\delta}}$. From the symmetry of the permutations, each element of $\tilde{\mathbf{d}}$ has the same distribution, with expectations across permutations $\mathbf{d}$ given by

$$\text{(A5)}\ \ E_{\mathbf{d}}(\tilde{d}_i) = \sum_{j=1}^{N}\frac{\tilde{d}_j}{N} = 0\ \ \&\ \ E_{\mathbf{d}}(\tilde{d}_i^2) = \sum_{j=1}^{N}\frac{\tilde{d}_j^2}{N} = 1,$$

while if $i_1 \neq i_2$ we have

$$\text{(A6)}\ \ E_{\mathbf{d}}(\tilde{d}_{i_1}\tilde{d}_{i_2}) = \sum_{j_1,j_2=1}^{N}\frac{\tilde{d}_{j_1}\tilde{d}_{j_2}}{N(N-1)} = \sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\frac{\tilde{d}_{j_1}\tilde{d}_{j_2}}{N(N-1)} - \sum_{j_1=1}^{N}\frac{\tilde{d}_{j_1}^2}{N(N-1)} = 0 - \frac{1}{N-1},$$

where we use the notation $j_1, j_2,...$ to denote summation across multiple indices, excluding ties between the indices. Using these, we compute the 1st and 2nd moments of the components of $\mathbf{v}$:

$$\text{(A7)}\quad E_{\mathbf{d}}(v_k)=\sum_{i=1}^{N}\frac{E_{\mathbf{d}}(\tilde{d}_i)\tilde{z}_{ik}}{N^{1/2}}=\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{d}_j\tilde{z}_{ik}}{N^{3/2}}=0$$

$$E_{\mathbf{d}}(v_k v_l)=\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\frac{E_{\mathbf{d}}(\tilde{d}_{i_1}\tilde{d}_{i_2})\tilde{z}_{i_1k}\tilde{z}_{i_2l}}{N}=\sum_{i_1=1}^{N}\frac{E_{\mathbf{d}}(\tilde{d}_{i_1}^2)\tilde{z}_{i_1k}\tilde{z}_{i_1l}}{N}+\sum_{i_1,i_2=1}^{N}\frac{E_{\mathbf{d}}(\tilde{d}_{i_1}\tilde{d}_{i_2})\tilde{z}_{i_1k}\tilde{z}_{i_2l}}{N}$$

$$=\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\frac{\tilde{d}_{j_1}^2\tilde{z}_{i_1k}\tilde{z}_{i_1l}}{N^2}+\sum_{i_1,i_2=1}^{N}\sum_{j_1,j_2=1}^{N}\frac{\tilde{d}_{j_1}\tilde{d}_{j_2}\tilde{z}_{i_1k}\tilde{z}_{i_2l}}{N^2(N-1)}$$

$$=\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\frac{\tilde{d}_{j_1}^2\tilde{z}_{i_1k}\tilde{z}_{i_1l}}{N^2}+\underbrace{\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\frac{\tilde{d}_{j_1}\tilde{d}_{j_2}\tilde{z}_{i_1k}\tilde{z}_{i_2l}}{N^2(N-1)}}_{=0}$$

$$-\underbrace{\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N}\frac{\tilde{d}_{j_1}^2\tilde{z}_{i_1k}\tilde{z}_{i_2l}}{N^2(N-1)}}_{=0}-\underbrace{\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\frac{\tilde{d}_{j_1}\tilde{d}_{j_2}\tilde{z}_{i_1k}\tilde{z}_{i_1l}}{N^2(N-1)}}_{=0}+\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\frac{\tilde{d}_{j_1}^2\tilde{z}_{i_1k}\tilde{z}_{i_1l}}{N^2(N-1)}$$

$$=1+\frac{1}{N-1}\ \text{(if } k=l\text{)}\ \text{ or } 0\ \text{ (otherwise).}$$

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, (A7) shows that the first moment of **v** is $\mathbf{0}_K$, while its second moments asymptotically equal the identity matrix, as desired. The next few pages focus on the higher moments.

Let $E_{\mathbf{d}}^{\tau}$ denote one of the τ^{th} moments of the joint distribution of **v** across the row permutations **d** of **δ**

$$\text{(A8)}\quad E_{\mathbf{d}}^{\tau}=E_{\mathbf{d}}\left[\prod_{p=1}^{\tau}v_{k(p)}\right]=E_{\mathbf{d}}\left[N^{-\tau/2}\sum_{i_1=1}^{N}...\sum_{i_\tau=1}^{N}\tilde{d}_{i_1}\tilde{z}_{i_1k(1)}...\tilde{d}_{i_\tau}\tilde{z}_{i_\tau k(p)}\right],$$

where the $k(p)$ indices may reference the same columns of $\tilde{\mathbf{Z}}$, i.e. $k(p)=k(q)$ for some $p\neq q$, so that the moment is across combinations of powers of the v_k. As can be seen from the second line of (A7) above, $E_{\mathbf{d}}^{\tau}$ needs to be separated into components based upon whether the i indices tie or not, which leads to elements of the form

$$\text{(A9)}\quad I(\tau,\{e_1\},\ldots,\{e_m\}) = E_{\mathbf{d}}\left[N^{-\tau/2}\sum_{i_1,\ldots,i_m=1}^{N}\tilde{d}_{i_1}^{e_1}\tilde{z}_{i_1}^{\{e_1\}}\ldots\tilde{d}_{i_m}^{e_m}\tilde{z}_{i_m}^{\{e_m\}}\right],\ \text{where}\ \sum_{i=1}^{m} e_i = \tau,\ e_i \geq 1\ \forall\ i$$

and $\sum_{i_1,\ldots,i_m}$ denotes the summation across m indices, excluding ties between the indices, the sets $\{e_1\}$, ... ,$\{e_m\}$ constitute a partition of the τ elements v_k of $\mathbf{v}$ used in $E_{\mathbf{d}}^{\tau}$, with the notation e_i without {} denoting the number of elements in $\{e_i\}$, and the $\tilde{d}^{e_i}$ and $\tilde{z}^{\{e_i\}}$ denoting the product of the elements within each set $\{e_i\}$. $\tilde{d}$ is raised to a cardinal number because there is only one column in $\tilde{\mathbf{d}}$, while we use the set notation $\{e_i\}$ on $\tilde{z}$ to denote the product of potentially distinct columns of $\tilde{\mathbf{Z}}$. The $\{e_i\}$ groupings tie elements together through their i and j indices. Thus, for example, we might have

$$\text{(A10)}\quad \{e_1\} = \{v_{k(1)}, v_{k(2)}\},\{e_2\} = \{v_{k(3)}\},\ldots,\{e_m\} = \{v_{k(\tau-1)}, v_{k(\tau)}\}$$

$$\tilde{d}_{i_1}^{e_1} = \tilde{d}_{i_1}^{2},\ \tilde{d}_{i_2}^{e_2} = \tilde{d}_{i_2},\ \ldots\ ,\tilde{d}_{i_m}^{e_m} = \tilde{d}_{i_m}^{2}$$

$$\text{and}\ \tilde{z}_{i_1}^{\{e_1\}} = \tilde{z}_{i_1k(1)}\tilde{z}_{i_1k(2)},\ \tilde{z}_{i_2}^{\{e_2\}} = \tilde{z}_{i_2k(3)},\ \ldots\ ,\tilde{z}_{i_m}^{\{e_m\}} = \tilde{z}_{i_mk(\tau-1)}\tilde{z}_{i_mk(\tau)}.$$

Since

$$\text{(A11)}\quad E_{\mathbf{d}}\left[\tilde{d}_{i_1}^{e_1}\ldots\tilde{d}_{i_m}^{e_m}\right] = \frac{N-m!}{N!}\sum_{j_1,\ldots,j_m}^{N}\tilde{d}_{j_1}^{e_1}\ldots\tilde{d}_{j_m}^{e_m}\ ,$$

we have

$$\text{(A12)}\quad I(\tau,\{e_1\},\ldots,\{e_m\}) = \underbrace{\frac{N-m!N^m}{N!}}_{\rightarrow 1}N^{-m-\frac{\tau}{2}}\sum_{i_1,\ldots,i_m}^{N}\sum_{j_1,\ldots,j_m}^{N}\tilde{d}_{j_1}^{e_1}\tilde{z}_{i_1}^{\{e_1\}}\ldots\tilde{d}_{j_m}^{e_m}\tilde{z}_{i_m}^{\{e_m\}}$$

$$\sim N^{-m-\frac{\tau}{2}}\sum_{i_1,\ldots,i_m}^{N}\sum_{j_1,\ldots,j_m}^{N}\tilde{d}_{j_1}^{e_1}\tilde{z}_{i_1}^{\{e_1\}}\ldots\tilde{d}_{j_m}^{e_m}\tilde{z}_{i_m}^{\{e_m\}},$$

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

$$\text{(A13)}\quad N^{-m-\frac{\tau}{2}}J(\tau,p,q,\{e_1\},\ldots,\{e_m\}) = N^{-m-\frac{\tau}{2}}\sum_{i_1=1}^{N}\ldots\sum_{i_p=1}^{N}\sum_{j_1=1}^{N}\ldots\sum_{j_q=1}^{N}\tilde{d}_{j_{d(1)}}^{e_1}\tilde{z}_{i_{c(1)}}^{\{e_1\}}\ldots\tilde{d}_{j_{d(m)}}^{e_m}\tilde{z}_{i_{c(m)}}^{\{e_m\}}$$

$$\text{with}\ 1 \leq p \leq m,\ 1 \leq q \leq m,\ 1 \leq c(g) \leq m,\ 1 \leq d(h) \leq m,\ (g,h = 1,\ldots,m)$$

and at least one c(g) and 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 the distinct $p \leq m$ and $q \leq m$ counters in the summations.

The third line of (A7) 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

$$\text{(A14)}\quad J(\tau,p,q,\{e_1\},...,\{e_m\}) = \prod_{u=1}^{s} J(\tau(u),p(u),q(u),\{e_{u1}\},...,\{e_{um(u)}\}),$$

where each $\{e_{ua}\}$ equals one of the original $\{e_i\}$, and the s sets $\{e_{u1}\},...,\{e_{um(u)}\}$ cover $\{e_1\},...,\{e_m\}$ in its entirety, with

$$\text{(A15)}\quad \sum_{i=1}^{m(u)} e_{ui} = \tau(u),\ \sum_{u=1}^{s}\tau(u) = \tau,\ \sum_{u=1}^{s} p(u) = p,\ \sum_{u=1}^{s} q(u) = q,\ \&\ \sum_{u=1}^{s} m(u) = m.$$

We assume that each J is subdivided into the greatest possible number of factors. In the fourth line of (A7) above, for example, we have:

$$\text{(A16)}\quad J(\tau=2,p=2,q=2,\{v_k\},\{v_l\}) = \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\tilde{d}_{j_1}\tilde{d}_{j_2}\tilde{z}_{i_1k}\tilde{z}_{i_2l} =$$

$$\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\tilde{d}_{j_1}\tilde{z}_{i_1k}\sum_{i_2=1}^{N}\sum_{j_2=1}^{N}\tilde{d}_{j_2}\tilde{z}_{i_2l} = J(\tau(1)=1,p(1)=1,q(1)=1,\{v_k\})J(\tau(2)=1,p(2)=1,q(2)=1,\{v_l\}),$$

while all three terms in the fifth line are indivisible because the i,j counters for the $\tilde{d}$ and $\tilde{z}$ elements connect at least one element of v_k to v_l. If $J(\tau(u),p(u),q(u),\{e_{u1}\},...,\{e_{um(u)}\})$ is indivisible, it is because the $2m(u)$ $c(u,g)$ and $d(u,h)$ subscript indices link across the $m(u)$ groups $\{e_{u1}\},...,\{e_{um(u)}\}$. To do so, there must be at least $m(u)$ - 1 equalities in these indices, i.e. at most $m(u)+1$ distinct values. At the same time, these indices cover every one of the numbers in 1...$p(u)$ and 1...$q(u)$, so we may conclude that

$$\text{(A17)}\quad p(u)+q(u) \le m(u)+1\,.$$

We note that if $(c(u,g),d(u,g)) = (c(u,h),d(u,h))$ for some $g \neq h$, we have more than the minimum $m(u)$ - 1 equalities necessary for indivisibility and (A17) holds with strict inequality. Summing across all s groups that make up $J(\tau,p,q,\{e_1\},...,\{e_m\})$,

$$\text{(A18)}\quad p+q \le m+s,$$

with strict inequality if $(c(u,g),d(u,g)) = (c(u,h),d(u,h))$ ever holds.

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

$$\text{(A19)}\quad \left|J(\tau(u),p(u),q(u),\{e_{u1}\},...,\{e_{um(u)}\})\right| \le \sum_{i_1=1}^{N}...\sum_{i_{p(u)}=1}^{N}\sum_{j_1=1}^{N}...\sum_{j_{q(u)}=1}^{N}\left|\tilde{d}_{j_{d(u,1)}}^{e_{u1}}\tilde{z}_{i_{c(u,1)}}^{\{e_{u1}\}}\right|...\left|\tilde{d}_{j_{d(u,m(u))}}^{e_{um(u)}}\tilde{z}_{i_{c(u,m(u))}}^{\{e_{um(u)}\}}\right|$$

$$\le \prod_{g=1}^{m(u)}\left(\sum_{i_1=1}^{N}...\sum_{i_{p(u)}=1}^{N}\sum_{j_1=1}^{N}...\sum_{j_{q(u)}=1}^{N}\left|\tilde{d}_{j_{d(u,g)}}^{e_{ug}}\tilde{z}_{i_{c(u,g)}}^{\{e_{ug}\}}\right|^{\tau(u)/e_{ug}}\right)^{e_{ug}/\tau(u)} = \prod_{g=1}^{m(u)}\left(N^{p(u)+q(u)-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tilde{d}_{j}^{e_{ug}}\tilde{z}_{i}^{\{e_{ug}\}}\right|^{\tau(u)/e_{ug}}\right)^{e_{ug}/\tau(u)},$$

where the reader is reminded that e_{ug} denotes the number of v_k in $\{e_{ug}\}$, with $\sum e_{ug} = \tau(u)$, allowing the application of Hölder's Inequality in the manner shown. We now decompose the set $\{e_{ug}\}$ into its constituent parts. Let $1...r$, $r \le \tau$, index the unique v_k variables across which the expectation $E_{\mathbf{d}}^{\tau}$ is taken, so that

$$\text{(A20)}\quad E_{\mathbf{d}}^{\tau} = E_{\mathbf{d}}\left[\prod_{p=1}^{\tau} v_{k(p)}\right] = E_{\mathbf{d}}\left[\prod_{a=1}^{r} v_{k(a)}^{f(a)}\right],$$

where in the first product different $k(p)$ may reference the same v_k, as earlier above, but in the second product each $k(a)$ references a unique v_k, with $f(a)$ denoting the power to which it is raised. Let $f(ug,1)...f(ug,r)$ similarly denote the power the unique v_k are raised to in the grouping $\{e_{ug}\}$. We then apply Hölder's Inequality once again

$$\text{(A21)}\ \prod_{g=1}^{m(u)}\left(N^{p(u)+q(u)-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tilde{d}_j^{e_{ug}}\tilde{z}_i^{\{e_{ug}\}}\right|^{\frac{\tau(u)}{e_{ug}}}\right)^{\frac{e_{ug}}{\tau(u)}}$$

$$=N^{p(u)+q(u)-2}\prod_{g=1}^{m(u)}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tilde{d}_j^{f(ug,1)}\ldots\tilde{d}_j^{f(ug,r)}\tilde{z}_{ik(1)}^{f(ug,1)}\ldots\tilde{z}_{ik(r)}^{f(ug,r)}\right|^{\frac{\tau(u)}{e_{ug}}}\right)^{\frac{e_{ug}}{\tau(u)}}\ \text{where}\ \sum_{g=1}^{m(u)}\sum_{a=1}^{r}f(ug,a)=\sum_{g=1}^{m(u)}e_{ug}=\tau(u),$$

$$\leq N^{p(u)+q(u)-2}\prod_{g=1}^{m(u)}\left(\prod_{a=1}^{r}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tilde{d}_j^{\tau(u)}\tilde{z}_{ik(a)}^{\tau(u)}\right|\right)^{\frac{f(ug,a)}{e_{ug}}}\right)^{\frac{e_{ug}}{\tau(u)}}=N^{p(u)+q(u)-2}\prod_{g=1}^{m(u)}\prod_{a=1}^{r}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tilde{d}_j^{\tau(u)}\tilde{z}_{ik(a)}^{\tau(u)}\right|\right)^{\frac{f(ug,a)}{\tau(u)}}$$

$$=N^{p(u)+q(u)-2}\prod_{a=1}^{r}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tilde{d}_j^{\tau(u)}\tilde{z}_{ik(a)}^{\tau(u)}\right|\right)^{\frac{f(u,a)}{\tau(u)}}\ \text{where}\ f(u,a)=\sum_{g=1}^{m(u)}f(ug,a)\ \&\ \sum_{a=1}^{r}f(u,a)=\tau(u),$$

$$=N^{p(u)+q(u)+\frac{\tau(u)}{2}-1}\prod_{a=1}^{r}\overline{M}(\tau(u),v_{k(a)})^{\frac{f(u,a)}{\tau(u)}}\ \text{where}\ \overline{M}(\tau(u),v_{k(a)})=N^{-\frac{\tau(u)}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tilde{d}_j^{\tau(u)}\tilde{z}_{ik(a)}^{\tau(u)}\right|.$$

Applying the bound to each element on the right hand side of (A14), we then have

$$\text{(A22)}\ N^{-m-\frac{\tau}{2}}\left|J(\tau,p,q,\{e_1\},\ldots,\{e_m\})\right|\leq N^{p+q-s-m}\prod_{u=1}^{s}\prod_{a=1}^{r}\overline{M}(\tau(u),v_{k(a)})^{f(u,a)/\tau(u)}.$$

Let us now assume (to be proven later) that (A2) implies that

$$\text{(A23)}\ N^{-\frac{\tau(u)}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{d}_j^{\tau(u)}\tilde{z}_{ik}^{\tau(u)}=o(1)\quad\text{for all } k \text{ and } \tau(u)=3,4,\ldots\ ,$$

then we see that if $\tau(u)$ is even and greater than 2, $\overline{M}(\tau(u),v_{k(a)})\rightarrow 0$. If $\tau(u)$ is odd and greater than 1, we can apply the Cauchy-Schwarz inequality

$$\text{(A24)}\ \overline{M}(2t+1,v_{k(a)})^2=\left(N^{-\frac{2t+1}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|v_{k(a)}\right|^{t}\left|v_{k(a)}\right|^{t+1}\right)^2$$

$$\leq\left(N^{-\frac{2t}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|v_{k(a)}\right|^{2t}\right)\left(N^{-\frac{2t+2}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|v_{k(a)}\right|^{2t+2}\right)$$

$$=\left(N^{-\frac{2t}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}v_{k(a)}^{2t}\right)\left(N^{-\frac{2t+2}{2}-1}\sum_{i=1}^{N}\sum_{j=1}^{N}v_{k(a)}^{2t+2}\right)=o(1)\ \text{for}\ t=1,2,\ldots$$

Finally, we have

$$\text{(A25)}\ \overline{M}(2,v_{k(a)})=N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\tilde{d}_j^2\tilde{z}_{ik(a)}^2\right|=N^{-2}\sum_{i=1}^{N}\tilde{d}_i^2\sum_{j=1}^{N}\tilde{z}_{jk(a)}^2=1.$$

Combining these results with (A22), and the fact that $p+q \le s+m$, we see that if $\tau(u) \ge 2$ for all u in 1...s and (a) $\tau(u) > 2$ for any u or (b) $\tau(u) = 2$ for all u and $p+q < s+m$, then $N^{-m-\tau/2}J(\tau...)$ asymptotically equals 0.

We now return to the equality in (A14), expressing $J(\tau...)$ as the product of s different $J(\tau(u)...)$. If $\tau(u) = 1$ we have $m(u) = p(u) = q(u) = 1$, and $J(\tau(u)...)$ is given by

$$\text{(A26)}\quad \sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\tilde{d}_{j_1}\tilde{z}_{i_1k(1)} = \sum_{i_1=1}^{N}\tilde{z}_{i_1k(1)}\sum_{j_1=1}^{N}\tilde{d}_{j_1} = 0,$$

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 $\tau(u) = 2$ for all u. This means that each $J(\tau(u), p(u), q(u), \{e_{u1}\},...,\{e_{um(u)}\})$ involves two elements, $v_{k(1)}$ and $v_{k(2)}$, divided into $m(u) = 1$ or 2 groups. If $m(u) = 2$, then $p(u) + q(u) \le 3$. If $p(u) + q(u) = 3$, then $J(\tau(u)...)$ is given by

$$\text{(A27)}\quad \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{N}\tilde{d}_{j_1}\tilde{d}_{j_1}\tilde{z}_{i_1k(1)}\tilde{z}_{i_2k(2)} \text{ or } \sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}\tilde{d}_{j_1}\tilde{d}_{j_2}\tilde{z}_{i_1k(1)}\tilde{z}_{i_1k(2)},$$

both of which are zero. If $p(u) + q(u) = 2$ for any u, 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 $J(\tau(u)...)$ we have $\tau(u) = 2$ and $m(u) = p(u) = q(u) = 1$ (as $p(u) \le m(u)$, $q(u) \le m(u)$), i.e. there is only one grouping of two v_ks (possibly the same), summed across one index for i and one for j, i.e.

$$\text{(A28)}\quad J(\tau(u) = 2, p(u) = 1, q(u) = 1, \{v_{k(1)}, v_{k(2)}\}) = \sum_{i_1=1}^{N}\sum_{j_1=1}^{N}\tilde{d}_{j_1}\tilde{d}_{j_1}\tilde{z}_{i_1k(1)}\tilde{z}_{i_1k(2)},$$

which equals N^2 if $k(1) = k(2)$ and 0 otherwise. Since $J(\tau...)$ is the product of $J(\tau(u)...)$, we then know that the only form of $N^{-m-\tau/2}J(\tau...)$ that is not identically or asymptotically zero is:

$$\text{(A29)}\quad N^{-m-\tau/2}J(\tau,p,q,\{v_{k(1)},v_{k(1)}\},...,\{v_{k(m)},v_{k(m)}\})\ \text{ with } m=p=q=\tau/2$$

$$=N^{-\tau}\sum_{i_1=1}^{N}\sum_{j_1=1}^{N}...\sum_{i_{\tau/2}=1}^{N}\sum_{j_{\tau/2}=1}^{N}\tilde{d}_{j_1}^2\tilde{z}_{i_1k(1)}^2...\tilde{d}_{j_{\tau/2}}^2\tilde{z}_{i_{\tau/2}k(\tau/2)}^2=N^{-\tau}N^{\tau}=1.$$

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 (A29). Hence, the only $I(\tau,...)$ that is not identically or asymptotically zero is that where τ is even and

$$\text{(A30)}\quad I(\tau,\{e_1\},...,\{e_m\})\sim N^{-m-\frac{\tau}{2}}\sum_{i_1,...,i_m}^{N}\sum_{j_1,...,j_m}^{N}\tilde{d}_{j_1}^{e_1}\tilde{z}_{i_1}^{\{e_1\}}...\tilde{d}_{j_m}^{e_m}\tilde{z}_{i_m}^{\{e_m\}}$$

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

$E_{\mathbf{d}}^{\tau}$ is made up of the sum of $I(\tau,...)$ which tie the τ v_k 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 v_k, so the only asymptotically non-zero $E_{\mathbf{d}}^{\tau}$ is that where the powers to which the r unique v_k 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 $\tau > 2$

$$\text{(A31)}\quad E_{\mathbf{d}}^{\tau}=E_{\mathbf{d}}\left[\prod_{a=1}^{r}v_{k(a)}^{f(a)}\right]\rightarrow\left[\prod_{a=1}^{r}(f(a)-1)!!\right]\ \text{(if all } f(a) \text{ even)},\ =0\ \text{(otherwise)},$$

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

All that remains is to show that (A2) implies (A23). Define

$$\text{(A32)}\quad \breve{z}_{ik}=\frac{z_{ik}-m(z_{ik})}{\left(\sum_{i=1}^{N}[z_{ik}-m(z_{ik})]^2\right)^{1/2}}\ \ \&\ \ \breve{d}_i=\frac{d_i-m(d_i)}{\left(\sum_{i=1}^{N}[d_i-m(d_i)]^2\right)^{1/2}},$$

so that (A2) may be re-expressed as

$$\text{(A33)}\quad \lim_{N\to\infty}N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\breve{z}_{ik}^{\tau}\sum_{j=1}^{N}\breve{d}_j^{\tau}=0\ \ \forall\ k\ \ \&\ \ \forall\ \tau=3,4,...\ .$$

If τ is even, we can equivalently say that

$$\text{(A2)}' \lim_{N\to\infty} N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\left|\breve{z}_{ik}^{\tau}\right|\sum_{j=1}^{N}\left|\breve{d}_{j}^{\tau}\right| = 0 \;\forall\; k.$$

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

$$\text{(A34)}\;\; N^{\frac{2\eta+1}{2}-1}\sum_{i=1}^{N}\left|\breve{z}_{ik}^{2\eta+1}\right|\sum_{i=1}^{N}\left|\breve{d}_{i}^{2\eta+1}\right| \le \left(N^{\frac{2\eta+2}{2}-1}\sum_{i=1}^{N}\left|\breve{z}_{ik}^{2\eta+2}\right|\sum_{i=1}^{N}\left|\breve{d}_{i}^{2\eta+2}\right|\right)^{½}\left(N^{\frac{2\eta}{2}-1}\sum_{i=1}^{N}\left|\breve{z}_{ik}^{2\eta}\right|\sum_{i=1}^{N}\left|\breve{d}_{i}^{2\eta}\right|\right)^{½},$$

so (A2)′ in fact applies for all τ = 3, 4, ...[1] We also note that

$$\text{(A35)}\;\; \tilde{\mathbf{d}} = \mathbf{Od}\left(\frac{\mathbf{d}'\mathbf{Od}}{N}\right)^{-½} = N^{½}\breve{\mathbf{d}} \;\;\&\;\; \tilde{\mathbf{Z}} = \mathbf{OZ}\left(\frac{\mathbf{Z}'\mathbf{OZ}}{N}\right)^{-½} = N^{½}\breve{\mathbf{Z}}\mathbf{W},\;\text{ where } \mathbf{W} = \text{diag}(\mathbf{Z}'\mathbf{OZ})^{½}(\mathbf{Z}'\mathbf{OZ})^{-½}.$$

The elements of **W** are asymptotically bounded as for all N sufficiently large the determinant of $\text{diag}(\mathbf{Z}'\mathbf{OZ})^{-½}\,\mathbf{Z}'\mathbf{OZ}\,\text{diag}(\mathbf{Z}'\mathbf{OZ})^{-½}$ is greater than some positive constant $\varDelta$, and so

$$\text{(A36)}\;\; \sum_{i=1}^{K}\sum_{j=1}^{K} w_{ij}^2 = trace(\mathbf{W}'\mathbf{W}) = trace(\text{diag}(\mathbf{Z}'\mathbf{OZ})^{½}\,\mathbf{Z}'\mathbf{OZ}^{-1}\,\text{diag}(\mathbf{Z}'\mathbf{OZ})^{½}) < K^{K+1}/\varDelta < \infty.$$

To see the last, note that by the properties of the Rayleigh quotient for any positive definite matrix **A**, $\mathbf{x}'\mathbf{x}\lambda_{\min} \le \mathbf{x}'\mathbf{Ax} \le \mathbf{x}'\mathbf{x}\lambda_{\max}$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of **A**. Consequently, $\mathbf{x}'\mathbf{Axx}'\mathbf{A}^{-1}\mathbf{x} \le (\mathbf{x}'\mathbf{x})^2\lambda_{\max}/\lambda_{\min}$, as the eigenvalues of $\mathbf{A}^{-1}$ are the inverse of those of **A**. Allowing **x** to equal a vector of zeros with a 1 in the i^{th} row, we see that the i^{th} diagonal element of $\mathbf{A}^{-1}$ is less than or equal to $\lambda_{\max}/\lambda_{\min}$ divided by the i^{th} diagonal element of **A**. For the K x K matrix $\text{diag}(\mathbf{Z}'\mathbf{OZ})^{-½}\,\mathbf{Z}'\mathbf{OZ}\,\text{diag}(\mathbf{Z}'\mathbf{OZ})^{-½}$ with determinant greater than $\varDelta$, all diagonal elements are 1, the largest eigenvalue is less than K (as the trace equals the sum of the eigenvalues), and the smallest eigenvalue must be greater than $\varDelta/K^{K-1}$ (as the determinant equals the product of the eigenvalues).

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

[1]When τ = 3 and η = 1, the second square root on the right-hand side of (A34) equals 1 while the first goes to 0; in all other cases both square roots on the right-hand side go to zero.

$$\text{(A37)}\quad \left| N^{-\frac{\tau}{2}-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{d}_j^{\tau} \tilde{z}_{ik}^{\tau} \right| = \left| N^{\frac{\tau}{2}-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \breve{d}_j^{\tau} \left(\sum_{e=1}^{K} \breve{z}_{ie} w_{ek} \right)^{\tau} \right|$$

$$= \left| N^{\frac{\tau}{2}-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \breve{d}_j^{\tau} \sum_{f(1)+\ldots+f(K)=\tau} \frac{\tau!}{f(1)!f(2)!\ldots f(K)!} \prod_{e=1}^{K} \breve{z}_{ie}^{f(e)} w_{ek}^{f(e)} \right|$$

$$\leq N^{\frac{\tau}{2}-1} \sum_{f(1)+\ldots+f(K)=\tau} \frac{\tau!}{f(1)!f(2)!\ldots f(K)!} \sum_{i=1}^{N} \sum_{j=1}^{N} \prod_{e=1}^{K} \left| \breve{d}_j^{f(e)} \breve{z}_{ie}^{f(e)} w_{ek}^{f(e)} \right|$$

$$\leq N^{\frac{\tau}{2}-1} \sum_{f(1)+\ldots+f(K)=\tau} \frac{\tau!}{f(1)!f(2)!\ldots f(K)!} \prod_{e=1}^{K} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \left| \breve{d}_j^{\tau} \breve{z}_{ie}^{\tau} w_{ek}^{\tau} \right| \right)^{f(e)/\tau}$$

$$\leq \sum_{f(1)+\ldots+f(K)=\tau} \frac{\tau!}{f(1)!f(2)!\ldots f(K)!} \prod_{e=1}^{K} \left(\left| w_{ek}^{\tau} \right| N^{\frac{\tau}{2}-1} \sum_{i=1}^{N} \left| \breve{z}_{ie}^{\tau} \right| \left| \sum_{j=1}^{N} \breve{d}_j^{\tau} \right| \right)^{f(e)/\tau} \to 0 \text{ [by (A2)}' \text{ above],}$$

where we use the notation $\sum_{f(1)+\ldots+f(K)=\tau}$ to denote the summation across all sets of K non-negative integers that sum to τ.

B. Proof of (L1p), (L2), (L3) and (9) for the Pairs Bootstrap

As we are only examining the pairs bootstrap, in this appendix we simplify notation and drop the superscript p on $\boldsymbol{\delta}$. (L1p), (L2) and (L3) appear in the Lemma in the paper's appendix and (9) in the text.

(L1p) and (9): We begin by deriving obvious results to familiarize the reader with the technique used in later, more challenging, steps. Define the random variable c_{ti} as a (0,1) indicator of whether observation i is chosen in bootstrap draw t. Obviously, c_{ti} and c_{tj} are interdependent, as only one of N observations is selected on any given draw, with, for γ and ζ each equal to any positive integer, $E(c_{ti}^{\gamma}) = N^{-1}$ and $E(c_{ti}^{\gamma}c_{tj}^{\zeta}) = 0$ if $i \neq j$, but c_{ti} and c_{sj} for $s \neq t$ are independent and identically distributed for all i and j with $E(c_{ti}^{\gamma}c_{sj}^{\zeta}) = N^{-2}$. Consequently:

$$\text{(B1)}\ m(\delta_i) = \frac{1}{N}\sum_{i=1}^{N}\delta_i = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{N}c_{ti} = 1, \quad E(m(\delta_i)) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{N}E(c_{ti}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{N}\frac{1}{N} = 1,$$

$$E(m(\delta_i)^2) = E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{t=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{N}c_{ti}c_{sj}\right] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{N}E(c_{ti}c_{tj}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s,t=1}^{N}E(c_{ti}c_{sj})$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{t=1}^{N}\frac{1}{N} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s,t=1}^{N}\frac{1}{N^2} = \frac{N^2}{N^3} + \frac{N^3(N-1)}{N^4} = 1,$$

$$\&\ \ E(m(\delta_i)^2) - E(m(\delta_i))^2 = 0 \ \ \text{[as expected]},$$

where we use subscripted s,t to denote the summation across values of the two indices, excluding ties between them. So, $m(\delta_i) = 1$ is a constant with zero variance (proving the first part of L1p in the Lemma).

Turning to $m(\delta_i^2)$

$$\text{(B2) } E(m(\delta_i^2)) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t_1=1}^{N}\sum_{t_2=1}^{N} E(c_{t_1 i}c_{t_2 i}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{N} E(c_{ti}^2) + \frac{1}{N}\sum_{i=1}^{N}\sum_{t_1,t_2=1}^{N} E(c_{t_1 i}c_{t_2 i}) = \frac{N^2}{N^2} + \frac{N^2(N-1)}{N^3} \to 2,$$

$$E(m(\delta_i^2)^2) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}\sum_{j=1}^{N}\sum_{s_1=1}^{N}\sum_{s_2=1}^{N} E(c_{t_1 i}c_{t_2 i}c_{s_1 j}c_{s_2 j})$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}\sum_{s_1=1}^{N}\sum_{s_2=1}^{N} E(c_{t_1 i}c_{t_2 i}c_{s_1 i}c_{s_2 i}) + \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}\sum_{s_1=1}^{N}\sum_{s_2=1}^{N} E(c_{t_1 i}c_{t_2 i}c_{s_1 j}c_{s_2 j})$$

$$= \frac{1}{N}\underbrace{\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}\sum_{s_1=1}^{N}\sum_{s_2=1}^{N} E(c_{t_1 i}c_{t_2 i}c_{s_1 i}c_{s_2 i})}_{a} + \frac{(N-1)}{N}\underbrace{\sum_{t_1=1}^{N}\sum_{t_2=1}^{N}\sum_{s_1=1}^{N}\sum_{s_2=1}^{N} E(c_{t_1 i}c_{t_2 i}c_{s_1 j}c_{s_2 j})}_{b},$$

where

$$\text{(B3) } a = \sum_{t=1}^{N} E(c_{ti}^4) + 4\sum_{t_1,t_2=1}^{N} E(c_{t_1 i}c_{t_2 i}^3) + 3\sum_{t_1,t_2=1}^{N} E(c_{t_1 i}^2 c_{t_2 i}^2) + 6\sum_{t_1,t_2,t_3=1}^{N} E(c_{t_1 i}^2 c_{t_2 i}c_{t_3 i}) + \sum_{t_1,t_2,t_3,t_4=1}^{N} E(c_{t_1 i}c_{t_2 i}c_{t_3 i}c_{t_4 i})$$

$$= \frac{N}{N} + \frac{7N(N-1)}{N^2} + \frac{6N!/(N-3)!}{N^3} + \frac{N!/(N-4)!}{N^4},$$

$$b = \sum_{t_1,s_1=1}^{N} E(c_{t_1 i}^2 c_{s_1 j}^2) + 2\sum_{t_1,t_2,s_1=1}^{N} E(c_{t_1 i}c_{t_2 i}c_{s_1 j}^2) + \sum_{t_1,t_2,s_1,s_2=1}^{N} E(c_{t_1 i}c_{t_2 i}c_{s_1 j}c_{s_2 j}) = \frac{N(N-1)}{N^2} + \frac{2N!/(N-3)!}{N^3} + \frac{N!/(N-4)!}{N^4}.$$

As shown, the expectation denoted by "*a*" is calculated by considering all ways in which the four indices might, based upon the equality of their values, be tied together in one, two, three or four groups, while the expectation of "*b*" is similarly calculated, but with the proviso that we can ignore cases where any *t* index equals an *s* index, as the expectation then is 0 (since $i \neq j$ in "*b*").

Having established the technique with these simple examples, we can consider the more general expectation for any integer power $\tau \geq 2$:

$$\text{(B4) } E(m(\delta_i^\tau)) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t_1=1}^{N}\ldots\sum_{t_\tau=1}^{N} E(c_{t_1 i}\ldots c_{t_\tau i}) = \sum_{t_1=1}^{N}\ldots\sum_{t_\tau=1}^{N} E(c_{t_1 i}\ldots c_{t_\tau i})$$

$$= \sum_{t=1}^{N} E(c_{ti}^\tau) + \sum_{a_1+a_2=\tau}\sum_{t_1,t_2=1}^{N} E(c_{t_1 i}^{a_1}c_{t_2 i}^{a_2}) + \sum_{a_1+a_2+a_3=\tau}\sum_{t_1,t_2,t_3=1}^{N} E(c_{t_1 i}^{a_1}c_{t_2 i}^{a_2}c_{t_3 i}^{a_3})\ldots\sum_{t_1,t_2,\ldots,t_\tau=1}^{N} E(c_{t_1 i}\ldots c_{t_\tau i})$$

$$= \frac{N}{N} + \sum_{a_1+a_2=\tau}\frac{N(N-1)}{N^2} + \sum_{a_1+a_2+a_3=\tau}\frac{N!/(N-3)!}{N^3} + \ldots. + \frac{N!/(N-\tau)!}{N^\tau}$$

$$= \sum_{j=1}^{\tau}\frac{N!/(N-j)!}{N^j} C_j^\tau = c_N(\tau),$$

where we use the notation $\sum_{a_1+\ldots+a_j=\tau}$ to denote the summation across all ways in which τ objects can be divided into j groups and C_j^τ to denote the number of ways this can be achieved (as all such objects have the same expectation), with $C_1^\tau = C_\tau^\tau = 1$. Similarly, we have

$$\text{(B5) } E(m(\delta_i^\tau)^2) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{t_1=1}^{N}\ldots\sum_{t_\tau=1}^{N}\sum_{j=1}^{N}\sum_{s_1=1}^{N}\ldots\sum_{s_\tau=1}^{N}E(c_{t_1 i}\ldots c_{t_\tau i}c_{s_1 j}\ldots c_{s_\tau j})$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{t_1=1}^{N}\ldots\sum_{t_\tau=1}^{N}\sum_{s_1=1}^{N}\ldots\sum_{s_\tau=1}^{N}E(c_{t_1 i}\ldots c_{t_\tau i}c_{s_1 i}\ldots c_{s_\tau i}) + \frac{1}{N^2}\sum_{i,j=1}^{N}\sum_{t_1=1}^{N}\ldots\sum_{t_\tau=1}^{N}\sum_{s_1=1}^{N}\ldots\sum_{s_\tau=1}^{N}E(c_{t_1 i}\ldots c_{t_\tau i}c_{s_1 j}\ldots c_{s_\tau j})$$

$$= \frac{1}{N}\underbrace{\sum_{t_1=1}^{N}\ldots\sum_{t_\tau=1}^{N}\sum_{s_1=1}^{N}\ldots\sum_{s_\tau=1}^{N}E(c_{t_1 i}\ldots c_{t_\tau i}c_{s_1 i}\ldots c_{s_\tau i})}_{d_N(\tau)} + \frac{N-1}{N}\underbrace{\sum_{t_1=1}^{N}\ldots\sum_{t_\tau=1}^{N}\sum_{s_1=1}^{N}\ldots\sum_{s_\tau=1}^{N}E(c_{t_1 i}\ldots c_{t_\tau i}c_{s_1 j}\ldots c_{s_\tau j})}_{e_N(\tau)},$$

where, keeping in mind that in "e_N" the expectation of any object with a tie between an s and t index is zero,

$$\text{(B6) } d_N(\tau) = \sum_{j=1}^{2\tau}\frac{N!/(N-j)!}{N^j}C_j^{2\tau},$$

$$e_N(\tau) = \sum_{s,t=1}^{N}E(c_{ti}^\tau)E(c_{sj}^\tau) + \sum_{a_1+a_2=\tau}\sum_{t_1,t_2,s=1}^{N}E(c_{t_1 i}^{a_1}c_{t_2 i}^{a_2})E(c_{sj}^\tau) + \ldots \sum_{t_1,t_2,\ldots,t_\tau,s=1}^{N}E(c_{t_1 i}\ldots c_{t_\tau i})E(c_{sj}^\tau)$$

$$+\sum_{b_1+b_2=\tau}\sum_{t,s_1,s_2=1}^{N}E(c_{ti}^\tau)E(c_{s_1 i}^{b_1}c_{s_2 i}^{b_2}) + \sum_{a_1+a_2=\tau}\sum_{b_1+b_2=\tau}\sum_{t_1,t_2,s_1,s_2=1}^{N}E(c_{t_1 i}^{a_1}c_{t_2 i}^{a_2})E(c_{s_1 i}^{b_1}c_{s_2 i}^{b_2}) + \ldots \sum_{b_1+b_2=\tau}\sum_{t_1,t_2,\ldots,t_\tau,s_1,s_2=1}^{N}E(c_{t_1 i}\ldots c_{t_\tau i})E(c_{s_1 i}^{b_1}c_{s_2 i}^{b_2})$$

$$+\ldots \quad = \sum_{j=1}^{\tau}\sum_{k=1}^{\tau}\frac{N!/(N-j-k)!}{N^{j+k}}C_j^\tau C_k^\tau.$$

We note that

$$\text{(B7) } e_N(\tau) - c_N(\tau)^2 = \sum_{j=1}^{\tau}\sum_{k=1}^{\tau}\frac{N!}{N^{j+k}(N-j-k)!}C_j^\tau C_k^\tau - \sum_{j=1}^{\tau}\sum_{k=1}^{\tau}\frac{N!N!}{N^{j+k}(N-j)!(N-k)!}C_j^\tau C_k^\tau$$

$$= \sum_{j=1}^{\tau}\sum_{k=1}^{\tau}\frac{N!C_j^\tau C_k^\tau}{N^{j+k}(N-j-k)!}\left[1 - \frac{N(N-1)\ldots(N-j+1)}{(N-k)(N-k+1)\ldots(N-k-j+1)}\right] < 0.$$

Using the above, we see that $m(\delta_i^\tau)$ converges in mean square and hence in probability to the value given in (B4) as its variance is seen to be $O(N^1)$

$$\text{(B8) } E[m(\delta_i^\tau)^2] - E[m(\delta_i^\tau)]^2 = \frac{d_N(\tau)}{N} + \frac{(N-1)}{N}e_N(\tau) - c_N(\tau)^2 < \frac{d_N(\tau)}{N} < \frac{\sum_{j=1}^{2\tau}C_j^{2\tau}}{N}.$$

We also note that for all $\tau \geq 2$ (B4) converges to a finite constant

$$\text{(B9)} \sum_{j=1}^{\tau} \frac{N!/(N-j)!}{N^j} C_j^\tau \rightarrow \sum_{j=1}^{\tau} C_j^\tau .$$

For the case of $\tau = 2$, $E(m(\delta_i^2)) = 2 - N^{-1} \rightarrow 2$ while the variance is $O(N^{-1})$, so $m(\delta_i^2)$ converges in mean square and hence in probability to 2, as stated in the second part of (L1p). This also proves (9) in the text as well, as $m(\delta_i^2) - m(\delta_i)^2 \xrightarrow{p} 1$. For the almost sure result in the last part of (L1p), we note that from (B4) and (B9) above for all $\tau \geq 2$ and any $\eta > 0$

$$\text{(B10)} \; N^{-\eta} E(m(\delta_i^\tau)) = N^{-\eta} \sum_{j=1}^{\tau} \frac{N!/(N-j)!}{N^j} C_j^\tau \rightarrow 0 .$$

Since the variance of $m(\delta_i^\tau)$ is $O(N^{-1})$, the variance of $N^{-\eta} m(\delta_i^\tau)$ is $O(N^{-1-2\eta})$ and for any ε and sufficiently large N such that $\varepsilon > N^{-\eta} E(m(\delta_i^\tau))$

$$\text{(B11) Prob}\{N^{-\eta} m(\delta_i^\tau) > \varepsilon\} = \text{Prob}\left\{N^{-\eta} m(\delta_i^\tau) - N^{-\eta} E(m(\delta_i^\tau)) > \varepsilon - N^{-\eta} E(m(\delta_i^\tau))\right\}$$

$$\leq \text{Prob}\left\{\left|N^{-\eta} m(\delta_i^\tau) - N^{-\eta} E(m(\delta_i^\tau))\right| > \varepsilon - N^{-\eta} E(m(\delta_i^\tau))\right\} \leq \frac{O(N^{-1-2\eta})}{\left[\varepsilon - N^{-\eta} E(m(\delta_i^\tau))\right]^2} = O(N^{-1-2\eta}),$$

where the last inequality follows from Chebyshev's Inequality. Consequently,

$$\text{(B12)} \sum_{N=1}^{\infty} \text{Prob}\left\{N^{-\eta} m(\delta_i^\tau) > \varepsilon\right\} < \infty$$

and by the Borel-Cantelli Lemma it follows that almost surely $N^{-\eta} m(\delta_i^\tau)$ is greater than ε only a finite number of times. Hence, almost surely for every ε there exists an N' such that $N^{-\eta} m(\delta_i^\tau) < \varepsilon$ for all $N > N'$, i.e. $N^{-\eta} m(\delta_i^\tau) \xrightarrow{as} 0$. For $\tau = 2$, as $\theta > 0$ in the Lemma in the paper's appendix is > 0, this proves the last part of (L1p). We use the result for $\tau > 2$ further below.

(L2): We have already shown, in the proof above of (9), that $m(\delta_i^2) - m(\delta_i)^2 \xrightarrow{p} 1$, but we want to prove something stronger, i.e. that asymptotically it almost surely lies above some $\kappa > 0$. Let $\{\delta_i = 1\}$ denote the set of observations i that are sampled once and only once in the N iid draws from the sample of N observations and $\{\delta_i \neq 1\}$ the set of (remaining) observations that are sampled with some other frequency. As $m(\delta_i) = 1$, we have $(\delta_i - m(\delta_i))^2 = 0$ for $i \in \{\delta_i = 1\}$, while if $i \in \{\delta_i \neq 1\}$, then $(\delta_i - m(\delta_i))^2 = 1$ if $\delta_i = 0$ & $(\delta_i - m(\delta_i))^2 > 1$ otherwise. With k denoting the fraction of the N observations that are in $\{\delta_i = 1\}$, we have

$$\text{(B13)} \sum_{i=1}^{N} \frac{[\delta_i - m(\delta_i)]^2}{N} = \sum_{i \in \{\delta_i = 1\}} \frac{[1-1]^2}{N} + \sum_{i \in \{\delta_i \neq 1\}} \frac{[\delta_i - 1]^2}{N} \geq (1-k).$$

Let $P_N(X)$ denote the probability that in a single bootstrap sample of N observations X belong to the set $\{\delta_i = 1\}$ and $S_N(T)$ denote the sum of these probabilities from T to N. From the Borel-Cantelli Lemma, if for some $k < 1$

$$\text{(B14)} \sum_{N=1}^{\infty} S_N(\lceil kN \rceil) < \infty,$$

where $\lceil kN \rceil$ denotes the integer ceiling of kN, then almost surely $\sum_{i=1}^{N} [\delta_i - m(\delta_i)]^2 / N$ is bounded from below by 1- k, proving (L2).

The distribution of δ_i is the same as that of the distribution of N balls across N cells, i.e. the classical "occupancy problem." As shown in Geiringer (1938)[2]

$$\text{(B15)} \ P_N(X) = \sum_{Z=X}^{N} \frac{(-1)^{Z+X} N!N!(N-Z)^{N-Z}}{X!(Z-X)!(N-Z)!(N-Z)!N^N}.$$

So

$$\text{(B16)} \ S_N(T) = \sum_{X=T}^{N} P_N(X) = \sum_{Z=T}^{N} \frac{(-1)^Z N!N!(N-Z)^{N-Z}}{Z!(N-Z)!(N-Z)!N^N} \sum_{X=T}^{Z} \frac{(-1)^X Z!}{X!(Z-X)!}$$

$$= \sum_{Z=T}^{N} \frac{(-1)^Z N!N!(N-Z)^{N-Z}}{Z!(N-Z)!(N-Z)!N^N} \left(\sum_{X=0}^{Z} \frac{(-1)^X Z!}{X!(Z-X)!} - \sum_{X=0}^{T-1} \frac{(-1)^X Z!}{X!(Z-X)!} \right)$$

$$= \sum_{Z=T}^{N} \frac{(-1)^Z N!N!(N-Z)^{N-Z}}{Z!(N-Z)!(N-Z)!N^N} \left(0 - (-1)^{T-1} \frac{Z-1!}{T-1!Z-T!} \right)$$

$$= \sum_{Z=T}^{N} \underbrace{\frac{(-1)^{Z+T} N!N!(N-Z)^{N-Z}}{(T-1)!(Z-T)!(N-Z)!(N-Z)!ZN^N}}_{f(Z,T,N)},$$

where for subsequent use we name the terms in the last summation f. We note:

$$\text{(B17)} \ S_N(T) \leq \sum_{Z=T}^{N} |f(Z,T,N)| \leq (N-T) Max_{Z \in T \dots N} |f(Z,T,N)|,$$

[2]Geiringer, Hilda (1938). "On the Probability Theory of Arbitrarily Linked Events." *The Annals of Mathematical Statistics* 9 (4): 260-271.

where $Max_{Z \in T \ldots N}$ denotes the maximum across *Z* in integers *T* through *N*.

To find the maximum, we begin by noting that the proportional rate of change of *f* is given by:

$$\text{(B18) } h(Z,T,N) = \frac{|f(Z+1,T,N)|}{|f(Z,T,N)|} = \left(\frac{Z}{Z+1}\right)\left(\frac{N-Z}{Z+1-T}\right)\left(\frac{N-Z-1}{N-Z}\right)^{N-Z-1}.$$

If we divide numerator and denominator of each term in parentheses by *N*, set $Z = \tau N$ and $T = \lceil kN \rceil$, and use the fact that $\lceil kN \rceil / N \to k$ and $(1-N^{-1})^N \to e^{-1}$, we have

$$\text{(B19) } h(Z,T,N) = \left(\frac{\tau}{\tau + N^{-1}}\right)\left(\frac{1-\tau}{\tau + N^{-1} - \lceil kN \rceil / N}\right)\left(1 - \frac{1}{(1-\tau)N}\right)^{(1-\tau)N-1} \to (1)\left(\frac{1-\tau}{\tau - k}\right)\left(e^{-1}\right),$$

which is monotonically declining in τ and equals 1 when

$$\text{(B20) } \tau = \frac{1+ek}{1+e}.$$

For $T = \lceil kN \rceil$, asymptotically the maximum of $|f(Z,T,N)|$ is reached at $Z = \lceil \tau N \rceil$, $\lceil \tau N \rceil - 1$ or $\lceil \tau N \rceil + 1$. Hence asymptotically

$$\text{(B21) } (N-T) Max_{Z \in T \ldots N} |f(Z,T,N)| \le \frac{N(1-k)N!N!(N-\tau N)^{N-\tau N}}{(kN-1)!(\tau N - kN)!(N-\tau N)!(N-\tau N)!\tau N N^N}$$

$$= \frac{\frac{1-k}{\tau} N!N!\left(\frac{e(1-k)}{1+e}N\right)^{\frac{e(1-k)}{1+e}N}}{(kN-1)!\left(\frac{1-k}{1+e}N\right)!\left(\frac{e(1-k)}{1+e}N\right)!\left(\frac{e(1-k)}{1+e}N\right)!N^N},$$

where, since we can divide everything by *N*, we simplify notation by replacing $\lceil kN \rceil$ and $\lceil \tau N \rceil$ with limiting values *kN* and τN.

Applying Stirling's formula, the expression in (B21) is bounded from above by

$$\text{(B22)}\quad \frac{C_1 e^{-2N} N^{2N+1}\left[\frac{eN(1-k)}{1+e}\right]^{\frac{eN(1-k)}{1+e}} e^{\frac{2}{12N}} e^{-\frac{1}{12(kN-1)+1}-\frac{1+e}{12N(1-k)+1+e}-\frac{2(1+e)}{12eN(1-k)+1+e}}}{e^{-kN+1}(kN-1)^{kN-\frac{1}{2}} e^{-\frac{N(1-k)}{1+e}}\left[\frac{N(1-k)}{1+e}\right]^{\frac{N(1-k)}{1+e}+\frac{1}{2}} e^{-2\frac{eN(1-k)}{1+e}}\left[\frac{eN(1-k)}{1+e}\right]^{2\frac{eN(1-k)}{1+e}+1} N^N}$$

$$= C_2\left[\frac{1-k}{1+e}\right]^{-N(1-k)} e^{-N}k^{-kN}\underbrace{\left(1-\frac{1}{kN}\right)^{-kN+\frac{1}{2}}}_{\to e}\underbrace{e^{\frac{2}{12N}} e^{-\frac{1}{12(kN-1)+1}-\frac{1+e}{12N(1-k)+1+e}-\frac{2(1+e)}{12eN(1-k)+1+e}}}_{\to 1},$$

where C_1 and C_2 are finite non-zero constants and the $e^{1/2N}$ & $e^{1/(2N+1)}$ type-terms are the upper and lower Stirling's formula proportional error bounds. For $k = .9$,

$$\text{(B23)}\quad \left[\frac{1-k}{1+e}\right]^{-N(1-k)} e^{-N}k^{-kN} = \left[\frac{.1}{1+e}\right]^{-.1N} e^{-N}.9^{-.9N} < e^{-.5N}.$$

Consequently, $\sum_{N=1}^{\infty} S_N(\lceil .9N\rceil) < \infty$ and so $\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^2 / N$ is almost surely bounded from below by 1 - .9 = .1, proving (L2) for the pairs bootstrap.

(L3): We have:

$$\text{(B24)}\quad \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^{\tau}}{\left(\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^2\right)^{\tau/2}} = \frac{N^{-\theta\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}\sum_{k=0}^{\tau}\frac{\tau!}{k!\tau-k!}\frac{\delta_i^k(-m(\delta_i))^{\tau-k}}{N}}{\left(\sum_{i=1}^{N}\frac{[\delta_i - m(\delta_i)]^2}{N}\right)^{\tau/2}}$$

$$= \frac{N^{-\theta\left(\frac{\tau}{2}-1\right)}\sum_{k=0}^{\tau}\sum_{i=1}^{N}(-1)^{\tau-k}\frac{\tau!}{k!\tau-k!}\frac{\delta_i^k}{N}}{\left(\sum_{i=1}^{N}\frac{[\delta_i - m(\delta_i)]^2}{N}\right)^{\tau/2}} = \frac{\sum_{k=0}^{\tau}(-1)^{\tau-k}\frac{\tau!}{k!\tau-k!}m(\delta_i^k)N^{-\theta\left(\frac{\tau}{2}-1\right)}}{\left(\sum_{i=1}^{N}\frac{[\delta_i - m(\delta_i)]^2}{N}\right)^{\tau/2}}.$$

From the results in (L1) and (L2) above, we know that the denominator of the last is almost surely bounded away from zero by $\kappa > 0$, while the numerator almost surely converges to zero. Hence, the entire expression almost surely converges to 0, proving (L3) for the pairs bootstrap.

C. Consistency of the Pairs Bootstrap with Sub-Sampling

This appendix proves consistency of the pairs bootstrap with sub-sampling $M < N$ observations, with and without replacement. As noted in the paper, we assume:

(C1a) for some $\gamma^* > 1/(1+\gamma)$ there exists a $c > 0$ such that $\lim_{\inf} M/N^{\gamma^*} > c$.

(C1b) $M/N \rightarrow 0$.

(C1a) is needed to ensure that sufficiently high moments of the bootstrap distribution exist, as the proof of the permutation distribution uses the method of moments.

We modify the notation, so that $\mathbf{\Delta}$ is now an M x N matrix of 0s with a single 1 in each row. Otherwise, the notation is as before, with the bootstrap sample given by $\mathbf{\Delta}(\mathbf{y},\mathbf{X})$ and the associated estimated coefficients, residuals and covariance matrix:

$$\text{(C2)}\; \hat{\boldsymbol\beta}_N^p(\boldsymbol\delta^p) = (\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{X})^{-1}\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{y} = \hat{\boldsymbol\beta}_N + (\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{X})^{-1}\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\boldsymbol\delta^p$$

$$\hat{\mathbf{V}}(\hat{\boldsymbol\beta}_N^p) = (\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{X})^{-1}\mathbf{X}'\{\mathbf{X}(\hat{\boldsymbol\beta}_N - \hat{\boldsymbol\beta}_N^p) + \hat{\boldsymbol\varepsilon}_N\}\{\boldsymbol\delta^p\}\{\mathbf{X}(\hat{\boldsymbol\beta}_N - \hat{\boldsymbol\beta}_N^p) + \hat{\boldsymbol\varepsilon}_N\}\mathbf{X}(\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{X}).$$

To simplify notation, below we drop the superscript p on $\boldsymbol\delta$ and subscript N on $\hat{\boldsymbol\varepsilon}$. We wish to show that across permutations $\mathbf{d}$ of $\boldsymbol\delta$ a version of Theorem V holds, namely

$$\text{(C3a)}\; \left(\frac{\mathbf{X}'\{\hat{\boldsymbol\varepsilon}\}\{\hat{\boldsymbol\varepsilon}\}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\left(\frac{\boldsymbol\delta'\mathbf{O}\boldsymbol\delta}{M}\right)^{-½}\sqrt{M}(\hat{\boldsymbol\beta}_N^p - \hat{\boldsymbol\beta}_N) \overset{d(\mathbf{d})|p(\boldsymbol\delta)as(\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow} \mathbf{n}_K$$

$$\text{(C3b)}\; M\hat{\mathbf{V}}(\hat{\boldsymbol\beta}_N^p) - N\hat{\mathbf{V}}(\hat{\boldsymbol\beta}_N) \overset{p(\mathbf{d})|p(\boldsymbol\delta)as(\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow} \mathbf{0}_{KxK},$$

where $\overset{d(\mathbf{d})|p(\boldsymbol\delta)as(\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow}$ and $\overset{p(\mathbf{d})|p(\boldsymbol\delta)as(\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow}$ signify convergence in distribution and in probability across the permutations $\mathbf{d}$ of $\boldsymbol\delta$ in probability across the distribution of $\boldsymbol\delta$ and almost surely across the distribution of ($\mathbf{X}$,$\boldsymbol\varepsilon$). Along with the result that $\boldsymbol\delta'\mathbf{O}\boldsymbol\delta / M \overset{p(\boldsymbol\delta)}{\rightarrow} 1$, these results ensure that almost surely conditional on each realization of the data ($\mathbf{X}$,$\boldsymbol\varepsilon$) the bootstrapped Wald statistics and (multiplied by $\sqrt{M/N}$) coefficient estimates have the same distribution as the unconditional (i.e. across the data generating process) original sample Wald and coefficient estimates. In the case of sub-sampling without replacement, the $p(\boldsymbol{\delta})$ here and everywhere later can be removed as the results hold for all realizations of $\boldsymbol{\delta}$.

As shown in the appendix in the paper, White's (1980) assumptions are sufficient to ensure that (L4) - (L7) in that appendix's Lemma hold for all θ in $(0, \gamma/(1+\gamma))$. With γ^* as in (C1a), we apply (L4) - (L7) below assuming that θ lies in $(1-\gamma^*, \gamma/(1+\gamma))$, so that $1 - \theta - \gamma^* < 0$. In this case, following assumption (C1a)

$$\text{(C4)}\quad \frac{N}{MN^{\theta}} = \frac{N^{1-\theta-\gamma^*}}{M/N^{\gamma^*}} = O(N^{1-\theta-\gamma^*}) \to 0.$$

The following Lemma, proven at the end of this appendix, provides sub-sampling counterparts of (L1p), (L2) and (L3) in the paper's appendix:

<u>Lemma C5</u>: Let $\overset{p(\boldsymbol{\delta})}{\to}$ denote convergence in probability across the distribution of $\boldsymbol{\delta}$, τ any integer greater than 2, and θ a constant in $(1-\gamma^*, \gamma/(1+\gamma))$. Then:

$$\text{(C5a)}\ m\left(\frac{N}{M}\delta_i\right) = 1;\quad \text{(C5b)}\ m\left(\frac{N}{M}\delta_i^2\right) \overset{p(\boldsymbol{\delta})}{\to} 1;\quad \text{(C5c)}\ N^{-\theta} m\left(\frac{N^2}{M^2}\delta_i^2\right) \overset{p(\boldsymbol{\delta})}{\to} 0;$$

$$\text{(C5d)}\ \text{ for some constant } \kappa > 0 \text{ for all } N \text{ sufficiently large } \sum_{i=1}^{N}\frac{[\delta_i - m(\delta_i)]^2}{M} > \kappa;$$

$$\text{(C5e)}\ \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^{\tau}}{\left(\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^2\right)^{\tau/2}} \overset{p(\boldsymbol{\delta})}{\to} 0.$$

(C5a), (C5b) and (C1b) together imply the requirement $\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}/M \overset{p(\boldsymbol{\delta})}{\to} 1$ noted above, as

$$\text{(C6)}\quad \frac{\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}}{M} = \sum_{i=1}^{N}\frac{[\delta_i^2 - m(\delta_i)]^2}{M} = \sum_{i=1}^{N}\frac{\delta_i^2}{M} - \frac{N}{M}m(\delta_i)^2 = m\left(\frac{N}{M}\delta_i^2\right) - \frac{M}{N} \overset{p(\boldsymbol{\delta})}{\to} 1.$$

Starting with the bootstrap coefficient estimates, we have:

$$\text{(C7)}\quad \sqrt{M}(\hat{\boldsymbol{\beta}}_N^p - \hat{\boldsymbol{\beta}}_N) = \mathbf{C}^{-1}\mathbf{a},\ \text{ where }\ \mathbf{C} = \frac{\mathbf{X}'\{\mathbf{d}\}\mathbf{X}(N/M)}{N}\ \text{ and }\ \mathbf{a} = \frac{\sqrt{(N/M)}\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}}{\sqrt{N}}.$$

Regarding the jk^{th} element of **C**, given by $\sum_{i=1}^{N} x_{ij}x_{ik}(N/M)d_i/N$, we can apply Theorem IV in the paper with $z_i = x_{ij}x_{ik}$ and "d_i" = $(N/M)d_i$. As $m(x_{ij}x_{ik})$ is almost surely bounded and $m(x_{ij}^2 x_{ik}^2)/N^{1-\theta} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0$ by (L4) and (L6) in the paper, we see that condition (IVb) is satisfied

$$\text{(C8)}\ \underbrace{\frac{m(x_{ij}^2 x_{ik}^2) - m(x_{ij}x_{ik})^2}{N^{1-\theta}}}_{\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (L4 \& L6)}} \left(\underbrace{\frac{1}{N^{\theta}} m\left(\frac{N^2}{M^2}\delta_i^2\right)}_{\overset{p(\boldsymbol{\delta})}{\rightarrow} 0 \text{ (C5c)}} - \underbrace{\frac{1}{N^{\theta}} m\left(\frac{N}{M}\delta_i\right)^2}_{\rightarrow 0 \text{ (C5a)}} \right) \overset{p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

and so

$$\text{(C9)}\ \underbrace{\frac{\mathbf{X}'\{\mathbf{d}\}\mathbf{X}(N/M)}{N}}_{\mathbf{C}} - \frac{\mathbf{X}'\mathbf{X}}{N} \underbrace{m\left(\frac{N}{M}\delta_i\right)}_{=1 \text{ (C5a)}} \overset{p(\mathbf{d})|p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}.$$

By the Corollary to the Continuous Mapping Theorem and (L4) given in the paper's appendix, it follows that $(\mathbf{X}'\{\mathbf{d}\}\mathbf{X}/M)^{-1}$ converges in probability across permutations **d** to bounded $(\mathbf{X}'\mathbf{X}/N)^{-1}$.

Noting that the k^{th} element of **a** equals $\sum_{i=1}^{N} x_{ik}\hat{\varepsilon}_i d_i \sqrt{N/M}/\sqrt{N}$, we apply the multivariate extension of Theorem II in the text with $z_{ik} = x_{ik}\hat{\varepsilon}_i$, or $\mathbf{Z} = \{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}$ and "d_i" = $d_i\sqrt{N/M}$. As was shown in the paper's appendix, the mean of z_{ik} is zero and we have $\mathbf{Z}'\mathbf{O}\mathbf{Z} = \mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}$ and $\mathbf{Z}'\mathbf{O}\mathbf{d} = \mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}$, while it was also shown that $\text{diag}(\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X})^{-½}(\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X})\text{diag}(\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X})^{-½}$ is almost surely non-singular with determinant greater than some $\Delta > 0$ for all N sufficiently large. From (C5d) we know that $\mathbf{d}'\mathbf{O}\mathbf{d} = \boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}$ is greater than zero for all N sufficiently large (i.e. the δ_i are not all equal). Hence, following Theorems II and III in the paper, the distribution across **d** of

$$\text{(C10)} \left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-½} \left(\frac{(N/M)\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-½} \frac{\sqrt{(N/M)}\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}}{\sqrt{N}}$$

converges in probability across **δ** and almost surely across (**X**,**ε**) to that of the iid multivariate standard normal provided that for all integer τ greater than 2

$$\text{(C11)}\quad \frac{N^{\theta\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N} x_{ik}^{\tau}\hat{\varepsilon}_i^{\tau}}{\left(\sum_{i=1}^{N} x_{ik}^2\hat{\varepsilon}_i^2\right)^{\tau/2}} \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}\left[\sqrt{\frac{N}{M}}\delta_i - m(\sqrt{\frac{N}{M}}\delta_i)\right]^{\tau}}{\left(\sum_{i=1}^{N}\left[\sqrt{\frac{N}{M}}\delta_i - m(\sqrt{\frac{N}{M}}\delta_i)\right]^2\right)^{\tau/2}}$$

$$= \frac{N^{\theta\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N} x_{ik}^{\tau}\hat{\varepsilon}_i^{\tau}}{\left(\sum_{i=1}^{N} x_{ik}^2\hat{\varepsilon}_i^2\right)^{\tau/2}} \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}\left[\delta_i - m(\delta_i)\right]^{\tau}}{\left(\sum_{i=1}^{N}\left[\delta_i - m(\delta_i)\right]^2\right)^{\tau/2}} \overset{p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

which given (L5) in the paper and (C5e) above is satisfied. Putting the preceding together, we have:

$$\text{(C12)}\quad \left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\left(\frac{\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}}{M}\right)^{-½}\sqrt{M}\,(\hat{\boldsymbol{\beta}}_N^p - \hat{\boldsymbol{\beta}}_N) =$$

$$\underbrace{\left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\left(\frac{\frac{N}{M}\mathbf{X}'\{\mathbf{d}\}\mathbf{X}}{N}\right)^{-1}\left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{½}}_{\overset{p(\mathbf{d})|p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\mathbf{I}_{KxK}}\underbrace{\left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-½}\left(\frac{\frac{N}{M}\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-½}\frac{\sqrt{\frac{N}{M}}\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}}{\sqrt{N}}}_{\overset{d(\mathbf{d})|p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\mathbf{n}_K} \overset{d(\mathbf{d})|p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{n}_K,$$

where all of the matrices and inverses involving **X** are by (L4) in the paper known to be almost surely bounded and positive definite with determinant > η > 0 for all N sufficiently large. This establishes (C3a).

Regarding the sub-sampling heteroskedasticity robust covariance estimates, similar to the paper's appendix we have

$$\text{(C13)}\ M\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N^p) = \left(\frac{\mathbf{X}'\{\mathbf{d}\}\mathbf{X}(N/M)}{N}\right)^{-1}\mathbf{B}\left(\frac{\mathbf{X}'\{\mathbf{d}\}\mathbf{X}(N/M)}{N}\right)^{-1},$$

$$\text{where}\ \ \mathbf{B} = \frac{\frac{N}{M}\mathbf{X}'\{\mathbf{X}(\hat{\boldsymbol{\beta}}_N - \hat{\boldsymbol{\beta}}_N^p) + \hat{\boldsymbol{\varepsilon}}\}\{\boldsymbol{\delta}^p\}\{\mathbf{X}(\hat{\boldsymbol{\beta}}_N - \hat{\boldsymbol{\beta}}_N^p) + \hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}.$$

The jk^{th} element of **B** is given by

$$\text{(C14)} \sum_{i=1}^{N} \frac{\frac{N}{M} x_{ij} x_{ik} d_i \left(\hat{\varepsilon}_i - \sum_{p=1}^{K} \frac{x_{ip}}{M^{½}} \sqrt{\frac{\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}}{M}} \hat{\eta}_p \right)^2}{N}, \left[\text{where } \hat{\boldsymbol{\eta}} = \left(\frac{\mathbf{d}'\mathbf{O}\mathbf{d}}{M} \right)^{-½} \sqrt{M} (\hat{\boldsymbol{\beta}}_N^p - \hat{\boldsymbol{\beta}}_N) \right]$$

$$= \underbrace{m\left(x_{ij} x_{ik} \hat{\varepsilon}_i^2 d_i \frac{N}{M} \right)}_{a} - 2 \sum_{p=1}^{K} \sqrt{\frac{\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}}{M}} \hat{\eta}_p \underbrace{m\left(\frac{x_{ij} x_{ik} x_{ip} \hat{\varepsilon}_i d_i \frac{N}{M}}{M^{½}} \right)}_{b} + \sum_{p=1}^{K} \sum_{q=1}^{K} \frac{\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}}{M} \hat{\eta}_p \hat{\eta}_q \underbrace{m\left(\frac{x_{ij} x_{ik} x_{ip} x_{iq} d_i \frac{N}{M}}{M} \right)}_{c}.$$

For "a", we apply Theorem IV with $z_i = x_{ij} x_{ik} \hat{\varepsilon}_i^2$ and "d_i" = $d_i(N/M)$. As $m(x_{ij} x_{ik} \hat{\varepsilon}_i^2)$ is almost surely bounded and $m(x_{ij}^2 x_{ik}^2 \hat{\varepsilon}_i^4)/N^{1-\theta} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0$ by (L4) & (C9) in the paper, condition (IVb) is met

$$\text{(C15)} \underbrace{\frac{m(x_{ij}^2 x_{ik}^2 \hat{\varepsilon}_i^4) - m(x_{ij} x_{ik} \hat{\varepsilon}_i^2)^2}{N^{1-\theta}}}_{\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (L4 \& C9 in paper)}} \left(\underbrace{N^{-\theta} m\left(\frac{N^2}{M^2} \delta_i^2 \right)}_{\overset{p(\boldsymbol{\delta})}{\rightarrow} 0 \text{(C5c)}} - \underbrace{N^{-\theta} m\left(\frac{N}{M} \delta_i \right)^2}_{\rightarrow 0 \text{(C5a)}} \right) \overset{p(\boldsymbol{\delta}) as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

and so

$$\text{(C16) "}a\text{": } m(x_{ij} x_{ik} \hat{\varepsilon}_i^2 d_i (N/M)) - m(x_{ij} x_{ik} \hat{\varepsilon}_i^2) \underbrace{m\left(\frac{N}{M} \delta_i \right)}_{=1 \text{ (C5a)}} \overset{p(\mathbf{d})|p(\boldsymbol{\delta}) as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

For "b", we apply Theorem IV with $z_i = x_{ij} x_{ik} x_{ip} \hat{\varepsilon}_i / M^{½}$ and "d_i" = $d_i(N/M)$. By (C13) and (C14) in the paper, $m(x_{ij} x_{ik} x_{ip} \hat{\varepsilon}_i / N^{½(1-\theta)}) \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0$ & $m(x_{ij}^2 x_{ik}^2 x_{ip}^2 \hat{\varepsilon}_i^2 / N^{1-\theta}) / N^{1-\theta} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0$, and so condition (IVb) is met

$$\text{(C17)} \underbrace{\frac{N^{1-\theta}}{M}}_{\rightarrow 0 \text{ (C4)}} \underbrace{\frac{m\left(\frac{x_{ij}^2 x_{ik}^2 x_{ip}^2 \hat{\varepsilon}_i^2}{N^{1-\theta}} \right) - m\left(\frac{x_{ij} x_{ik} x_{ip} \hat{\varepsilon}_i}{N^{½(1-\theta)}} \right)^2}{N^{1-\theta}}}_{\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (C13 \& C14 in paper)}} \left(\underbrace{N^{-\theta} m\left(\frac{N^2}{M^2} \delta_i^2 \right)}_{\overset{p(\boldsymbol{\delta})}{\rightarrow} 0 \text{(C5c)}} - \underbrace{N^{-\theta} m\left(\frac{N}{M} \delta_i \right)^2}_{\rightarrow 0 \text{(C5a)}} \right) \overset{p(\boldsymbol{\delta}) as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

and so

$$\text{(C18) "}b\text{": } m\left(\frac{x_{ij} x_{ik} x_{ip} \hat{\varepsilon}_i d_i \frac{N}{M}}{M^{½}} \right) - \underbrace{\frac{N^{½(1-\theta)}}{M^{½}}}_{\rightarrow 0 \text{ (C4)}} \underbrace{m\left(\frac{x_{ij} x_{ik} x_{ip} \hat{\varepsilon}_i}{N^{½(1-\theta)}} \right)}_{\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (C13 in paper)}} \underbrace{m\left(\frac{N}{M} \delta_i \right)}_{=1 \text{(C5a)}} \overset{p(\mathbf{d})|p(\boldsymbol{\delta}) as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

For "c", we apply Theorem IV with $z_i = x_{ij}x_{ik}x_{ip}x_{iq}/M$ and "d_i" = $d_i(N/M)$. By (C16) and (C22) in the paper $m(x_{ij}x_{ik}x_{ip}x_{iq}/N^{1-\theta}) \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0$ & $m(x_{ij}^2x_{ik}^2x_{ip}^2x_{iq}^2/N^{2-2\theta})/N^{1-\theta} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0$, and so condition (IVb) is met

$$\text{(C19)}\ \underbrace{\frac{N^{2-2\theta}}{M^2}}_{\to 0\ (\text{C4})}\underbrace{\frac{m\left(\frac{x_{ij}^2x_{ik}^2x_{ip}^2x_{iq}^2}{N^{2-2\theta}}\right)-m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}}{N^{1-\theta}}\right)^2}{N^{1-\theta}}}_{\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0\ (\text{C16 \& C22 in paper})}\left(\underbrace{N^{-\theta}m\left(\frac{N^2}{M^2}\delta_i^2\right)}_{\overset{p(\boldsymbol{\delta})}{\to} 0\,(\text{C5c})}-\underbrace{N^{-\theta}m\left(\frac{N}{M}\delta_i\right)^2}_{\to 0\,(\text{C5a})}\right)\overset{p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0$$

and

$$\text{(C20) "}c\text{": } m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}d_i\frac{N}{M}}{M}\right)-\underbrace{\frac{N^{1-\theta}}{M}}_{\to 0\ (\text{C4})}\underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}}{N^{1-\theta}}\right)}_{\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0\ (\text{C16 in paper})}\underbrace{m\left(\frac{N}{M}\delta_i\right)}_{=1\,(\text{C5a})}\overset{p(\mathbf{d})|as(\boldsymbol{\delta},\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0.$$

In sum, in (C14) the $\hat{\eta}_p$ are multiplied by $\sqrt{\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}/M}$ which converges in probability (across **δ**) to 1 and "b" and "c" terms which in probability across **δ** and almost surely across (**X**,**ε**) across permutations **d** of **δ** converge in probability to zero. As across the permutations **d** of **δ** the $\hat{\eta}_p$ themselves converge to normal random variables with almost surely (across **X**,**ε**) bounded variance, when so multiplied the product converges in probability across permutations **d** to zero. This leaves only the "a" term which, using (C9), establishes (C3b):

$$\text{(C21)}\quad \mathbf{B}-\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\overset{p(\mathbf{d})|p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\to}\mathbf{0}_{KxK}\ \ \&\ \ M\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N^p)-N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)\overset{p(\mathbf{d})|p(\boldsymbol{\delta})as(\mathbf{X},\boldsymbol{\varepsilon})}{\to}\mathbf{0}_{KxK}.$$

Proof of (C5) Lemma

We prove the Lemma used above. (C5a) holds automatically, as the mean of δ_i will always equal *M/N*. For the case of sampling without replacement, δ_i is 1 for *M* observations and 0 for *N-M*. For this case, all elements of (C5) are easily proven:

$$\text{(C22a)}\ m\left(\frac{N}{M}\delta_i\right)=\frac{N}{M}\sum_{i=1}^{N}\frac{\delta_i}{N}=\frac{N}{M}\frac{M}{N}=1\,;\ \ \text{(C22b)}\ m\left(\frac{N}{M}\delta_i^2\right)=\frac{N}{M}\frac{M}{N}=1\,;$$

$$\text{(C22c)}\ N^{-\theta}m\left(\frac{N^2}{M^2}\delta_i^2\right)=N^{-\theta}\frac{N^2}{M^2}\frac{M}{N}=\underbrace{\frac{N}{MN^{\theta}}\to 0}_{\text{(C4)}};$$

$$\text{(C22d)}\ \sum_{i=1}^{N}\frac{[\delta_i-m(\delta_i)]^2}{M}=\frac{M}{M}\left(1-\frac{M}{N}\right)^2+\frac{(N-M)}{M}\left(\frac{M}{N}\right)^2=1-\frac{M}{N}\to 1>\kappa>0;$$

$$\text{(C22e)}\ \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}[\delta_i-m(\delta_i)]^{\tau}}{\left(\sum_{i=1}^{N}[\delta_i-m(\delta_i)]^2\right)^{\tau/2}}=\frac{\left(\frac{N}{MN^{\theta}}\right)^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\sum_{v=0}^{\tau}\frac{\tau!}{v!(\tau-v)!}\frac{\delta_i^{v}}{M}(-m(\delta_i))^{\tau-v}}{\left(\sum_{i=1}^{N}\frac{[\delta_i-m(\delta_i)]^2}{M}\right)^{\tau/2}}$$

$$=\frac{\left(\frac{N}{MN^{\theta}}\right)^{\frac{\tau}{2}-1}\left(\frac{N}{M}\left(-\frac{M}{N}\right)^{\tau}+\sum_{v=1}^{\tau}\frac{\tau!}{v!(\tau-v)!}\left(-\frac{M}{N}\right)^{\tau-v}\right)}{\left(1-\frac{M}{N}\right)^{\tau/2}}=\underbrace{O(N^{(1-\theta-\gamma^*)(\frac{\tau}{2}-1)})\to 0}_{\text{(C4)}}.$$

For sampling with replacement, we define the random variable c_{ti} as in Appendix B above and following the notation and methods used there see that:

$$\text{(C23)}\ E\left(m\left(\frac{N}{M}\delta_i\right)\right)=\frac{1}{M}\sum_{i=1}^{N}\sum_{t=1}^{M}E(c_{ti})=\frac{1}{M}\sum_{i=1}^{N}\sum_{t=1}^{M}\frac{1}{N}=1,$$

$$E\left(m\left(\frac{N}{M}\delta_i\right)^2\right)=E\left[\frac{1}{M^2}\sum_{i=1}^{N}\sum_{t=1}^{M}\sum_{j=1}^{N}\sum_{s=1}^{M}c_{ti}c_{sj}\right]=\frac{1}{M^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{M}E(c_{ti}c_{tj})+\frac{1}{M^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s,t=1}^{M}E(c_{ti}c_{sj})$$

$$=\frac{1}{M^2}\sum_{i=1}^{N}\sum_{t=1}^{M}\frac{1}{N}+\frac{1}{M^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s,t=1}^{M}\frac{1}{N^2}=\frac{1}{M}+\frac{M(M-1)}{M^2}=1,$$

$$\&\ \ E\left(m(\delta_i)^2\right)-E\left(m(\delta_i)\right)^2=0\ \text{[as expected]},$$

So, $m((N/M)\delta_i)=1$ is a constant with zero variance (proving C5a).

Continuing with the notation and techniques of Appendix B, we calculate the expectation of $m((N/M)\delta_i^{\tau})$ for integer $\tau\geq 2$.

$$\text{(C24) } E\left(m\left(\frac{N}{M}\delta_i^\tau\right)\right) = \frac{1}{M}\sum_{i=1}^{N}\sum_{t_1=1}^{M}...\sum_{t_\tau=1}^{M}E(c_{t_1 i}...c_{t_\tau i}) = \frac{N}{M}\sum_{t_1=1}^{M}...\sum_{t_\tau=1}^{M}E(c_{t_1 i}...c_{t_\tau i}) = \frac{N}{M}c_N(\tau),$$

$$\text{with } c_N(\tau) = \sum_{j=1}^{\tau}\frac{M!/(M-j)!}{N^j}C_j^\tau,$$

and the expectation of $m((N/M)\delta_i^\tau)^2$

$$\text{(C25) } E\left(m\left(\frac{N}{M}\delta_i^\tau\right)^2\right) = \frac{1}{M^2}\sum_{i=1}^{N}\sum_{t_1=1}^{M}...\sum_{t_\tau=1}^{M}\sum_{j=1}^{N}\sum_{s_1=1}^{M}...\sum_{s_\tau=1}^{M}E(c_{t_1 i}...c_{t_\tau i}c_{s_1 j}...c_{s_\tau j})$$

$$= \frac{1}{M^2}\sum_{i=1}^{N}\sum_{t_1=1}^{M}...\sum_{t_\tau=1}^{M}\sum_{s_1=1}^{M}...\sum_{s_\tau=1}^{M}E(c_{t_1 i}...c_{t_\tau i}c_{s_1 i}...c_{s_\tau i}) + \frac{1}{M^2}\sum_{i,j=1}^{N}\sum_{t_1=1}^{M}...\sum_{t_\tau=1}^{M}\sum_{s_1=1}^{M}...\sum_{s_\tau=1}^{M}E(c_{t_1 i}...c_{t_\tau i}c_{s_1 j}...c_{s_\tau j})$$

$$= \frac{N}{M^2}\underbrace{\sum_{t_1=1}^{M}...\sum_{t_\tau=1}^{M}\sum_{s_1=1}^{M}...\sum_{s_\tau=1}^{M}E(c_{t_1 i}...c_{t_\tau i}c_{s_1 i}...c_{s_\tau i})}_{d_N(\tau)} + \frac{N(N-1)}{M^2}\underbrace{\sum_{t_1=1}^{M}...\sum_{t_\tau=1}^{M}\sum_{s_1=1}^{M}...\sum_{s_\tau=1}^{M}E(c_{t_1 i}...c_{t_\tau i}c_{s_1 j}...c_{s_\tau j})}_{e_N(\tau)},$$

$$\text{with } d_N(\tau) = \sum_{j=1}^{2\tau}\frac{M!/(M-j)!}{N^j}C_j^{2\tau} \text{ and } e_N(\tau) = \sum_{j=1}^{\tau}\sum_{k=1}^{\tau}\frac{M!/(M-j-k)!}{N^{j+k}}C_j^\tau C_k^\tau.$$

We note that

$$\text{(C26) } e_N(\tau) - c_N(\tau)^2 = \sum_{j=1}^{\tau}\sum_{k=1}^{\tau}\frac{M!}{N^{j+k}(M-j-k)!}C_j^\tau C_k^\tau - \sum_{j=1}^{\tau}\sum_{k=1}^{\tau}\frac{M!M!}{N^{j+k}(M-j)!(M-k)!}C_j^\tau C_k^\tau$$

$$= \sum_{j=1}^{\tau}\sum_{k=1}^{\tau}\frac{M!C_j^\tau C_k^\tau}{N^{j+k}(M-j-k)!}\left[1 - \frac{M(M-1)...(M-j+1)}{(M-k)(M-k+1)...(M-k-j+1)}\right] < 0.$$

From this we see that $m((N/M)\delta_i^\tau)$ converges in mean square and hence in probability to the finite value given in (C24) as its variance is $O(M^{-1})$:

$$\text{(C27) } E\left[m(\delta_i^\tau) - E\left[m(\delta_i^\tau)\right]\right]^2 = E\left[m(\delta_i^\tau)^2\right] - E\left[m(\delta_i^\tau)\right]^2$$

$$= \frac{N}{M^2}d_N(\tau) + \frac{N(N-1)}{M^2}e_N(\tau) - \frac{N^2}{M^2}c_N(\tau)^2 < \frac{N}{M^2}d_N(\tau) = \frac{N}{M^2}\sum_{j=1}^{2\tau}\frac{M!/(M-j)!}{N^j}C_j^{2\tau} = O(M^{-1}) \to 0,$$

where the limit follows from the fact that assumption (C1a), $\lim_{\inf} M/N^{\gamma^*} > c > 0$, implies that M goes to infinity as $N \to \infty$. For (C5b) we have

$$\text{(C28) } E\left(\frac{N}{M}m(\delta_i^2)\right) = \frac{N}{M}c_N(2) = \frac{N}{M}\sum_{j=1}^{2}\frac{M!/(M-j)!}{N^j}C_j^2 = \frac{N}{M}\left(\frac{M}{N} + \frac{M(M-1)}{N^2}\right) \to 1.$$

Thus, $m((N/M)\delta_i^2)$ converges in mean square and hence in probability to 1, as stated in (C5b).

For (C5c), we have:

$$\text{(C29) } E\left(m\left(\frac{N^{2-\theta}}{M^2}\delta_i^2\right)\right) = \underbrace{\frac{N^{1-\theta}}{M}}_{\to 0 \text{ (C4)}} + N^{-\theta} - \frac{N^{-\theta}}{M} \to 0,$$

while the variance of $m((N^{2-\theta}/M^2)\delta_i^2)$ is $O(N^{2\text{-}2\theta}/M^3) \to 0$ (by C4). Consequently, $m((N^{2-\theta}/M^2)\delta_i^2)$ converges in mean square and hence in probability to 0, as stated in (C5c).

For (C5d), the lowest value is achieved when no observation is sampled more than once, so

$$\text{(C30) } \sum_{i=1}^{N}\frac{[\delta_i - m(\delta_i)]^2}{M} = \frac{M}{M}\left(1-\frac{M}{N}\right)^2 + \frac{N-M}{M}\left(\frac{M}{N}\right)^2 = 1-\frac{M}{N} \to 1 > \kappa > 0,$$

proving (C5d). Finally, with regards to (C5e) we have

$$\text{(C31) } \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^\tau}{\left(\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^2\right)^{\tau/2}} = \frac{\left(\frac{N^{1-\theta}}{M}\right)^{\frac{\tau}{2}-1}\sum_{i=1}^{N}\sum_{k=0}^{\tau}\frac{\delta_i^k m(\delta_i)^{\tau-k}}{M}}{\left(\sum_{i=1}^{N}\frac{[\delta_i - m(\delta_i)]^2}{M}\right)^{\tau/2}}$$

$$= \frac{\left(\frac{N^{1-\theta}}{M}\right)^{\frac{\tau}{2}-1}\sum_{k=0}^{\tau}\frac{\tau!}{k!\tau-k!}m\left(\frac{N}{M}\delta_i^k\right)\left(\frac{M}{N}\right)^{\tau-k}}{\left(\sum_{i=1}^{N}\frac{[\delta_i - m(\delta_i)]^2}{M}\right)^{\tau/2}} \xrightarrow{p(\boldsymbol{\delta})} 0,$$

as from (C30) we know the denominator is bounded away from zero, while $m((N/M)\delta_i^k)$ converges in probability to the finite number in (C24), and by (C1b) and (C4) $M/N \to 0$ and $N^{1\text{-}\theta}/M \to 0$.