Consistency of the OLS Bootstrap for Independently but Not-Identically Distributed Data: A Permutation Perspective

Alwyn Young
London School of Economics
July 2022

Abstract

Analyzing the distribution of the pairs and wild OLS bootstrap as that of a permutation statistic reveals moment conditions sufficient for consistency conditional on independently but not-necessarily identically distributed (inid) data which cover more general regression models than earlier inid results and are by and large less demanding than previous results for independently and identically distributed data

*I am grateful to Taisuke Otsu for helpful comments.

I. Introduction

This paper marries results on the asymptotic distribution of permutation statistics (Wald & Wolfowitz 1944, Noether 1949 and Hoeffding 1951) to White's (1980) proof of the consistency of the heteroskedasticity robust OLS covariance estimate to extend results concerning the consistency of the pairs and wild OLS bootstrap, which have mostly been derived for independently and identically distributed (iid) data, to general regression frameworks with independently but not-necessarily identically distributed (inid) data. Instead of considering the sampling distribution of the bootstraps, the usual approach, one can instead note that any permutation of the pairs bootstrap vector of sampling frequencies or wild bootstrap vector of iid residual transformations is equally likely. These equally likely permutations can be used to characterize the bootstrap distributions conditional on the data as normal given limiting restrictions on sample moments of the data. White's (1980) conditions for the asymptotic normality of OLS coefficients guarantee these restrictions almost surely, ensuring that the asymptotic distribution of pairs and wild bootstrapped coefficients and Wald statistics conditional on the data matches the unconditional distribution of the original OLS estimates.

This paper broadens earlier results on the consistency of the bootstrap. For OLS models with inid data, the salient contribution is Liu (1988), who showed that the wild bootstrap provides consistent estimates of the second moment of a linear combination of coefficients in an OLS regression model with bounded regressors provided the first and second moment of the iid bootstrap residual transformations are 0 and 1, respectively. We show below that if sufficiently high additional moments of the wild bootstrap residual transformations, regressors and errors exist, the wild bootstrap consistently estimates the multivariate distribution of coefficients and Wald statistics of potentially unbounded regressors. With the same moment restrictions on regressors and errors, the pairs bootstrap is shown to be consistent for inid data as well. For OLS models with iid data and potentially heteroskedastic errors, Freedman (1981) showed bounded fourth moments of both regressors and errors are sufficient for consistency of the pairs bootstrap.[1] Mammen (1993) proved consistency of the wild OLS bootstrap with iid data with bounded expectations of the product of the fourth moments of the regressors and second moment of the residuals and an additional Lindeberg condition. This paper finds that White's assumptions guarantee consistency in a broader inid environment for both the pairs and wild bootstraps with only slightly more than second moments of the errors and fourth moments of the regressors, extending iid consistency to an inid framework with by-and-large less demanding assumptions. These results are useful because when data is drawn, for example, from distinct populations, geographic regions or time periods, the iid assumption is less likely to hold.

The paper proceeds as follows: Following a short review of notation in section II, section III reviews foundational theorems regarding the asymptotic normality of permutation distributions. Section IV then combines these with White's (1980) result to derive sufficient conditions for pairs and wild OLS bootstrap consistency with inid data, while section V concludes with some remarks. The appendix and on-line appendix provide details of the proofs.

II. Framework and Notation

Our interest is in inference for the familiar OLS linear model

$$\text{(1)}\quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with $\mathbf{X}$ the N x K matrix of regressors, $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ the N x 1 matrices representing the dependent variable and residuals, and $\boldsymbol{\beta}$ the K x 1 vector of coefficients. We assume that the row vectors associated with each observation i, $(\mathbf{x}_i', \varepsilon_i)$, are independently but not necessarily identically distributed (inid). $\hat{\boldsymbol{\beta}}_N$ and $\hat{\boldsymbol{\varepsilon}}_N$ denote the OLS estimated coefficients and residuals, where for the purpose of describing limits below we use the subscript N to emphasize that these are functions of N realized observations. As the residuals are not necessarily homoskedastic, we use White's (1980) heteroskedasticity robust covariance estimate, $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, where the notation $\mathbf{D}_\mathbf{z}$ denotes a diagonal matrix with diagonal entries $\mathbf{z}$. Given the inid data

[1]With independent homoskedastic errors, the bootstrap resampling of estimated residuals (rather than the data itself) always yields consistent estimates of the coefficient distribution for a fixed number of OLS regressors (Bickel and Freedman 1983).

$N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)$ does not necessarily converge to a matrix of constants.

The pairs and wild bootstrap use different techniques to achieve inference for the linear model with inid data. The pairs bootstrap samples rows of dependent variables and regressors from the original data (**y**,**X**), and uses these to re-estimate the regression model and its covariance estimates. This process can be described as re-estimation using the data $\boldsymbol{\Delta}(\mathbf{y},\mathbf{X})$, where $\boldsymbol{\Delta}$ is an N x N matrix of 0s with a single 1 in each row.[2] Since $\boldsymbol{\Delta}\mathbf{y} = \boldsymbol{\Delta}\mathbf{X}\hat{\boldsymbol{\beta}}_N + \boldsymbol{\Delta}\hat{\boldsymbol{\varepsilon}}_N$, the estimated coefficients, residuals and covariance matrix for each pairs bootstrap sample are given by:

$$(2)\ \hat{\boldsymbol{\beta}}^p_{\boldsymbol{\delta}^p,N} = (\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}^p}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}^p}\mathbf{y} = \hat{\boldsymbol{\beta}}_N + (\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}^p}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}^p}\hat{\boldsymbol{\varepsilon}}_N$$

$$\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta}^p,N} = \boldsymbol{\Delta}\hat{\boldsymbol{\varepsilon}}_N + \boldsymbol{\Delta}\mathbf{X}(\hat{\boldsymbol{\beta}}_N - \hat{\boldsymbol{\beta}}^p_{\boldsymbol{\delta}^p,N}),\ \text{and}\ \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{\delta}^p,N}) = (\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}^p}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Delta}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta}^p,N}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta}^p,N}}\boldsymbol{\Delta}\mathbf{X}(\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}^p}\mathbf{X})^{-1},$$

where $\mathbf{D}_{\boldsymbol{\delta}^p} = \boldsymbol{\Delta}'\boldsymbol{\Delta}$ denotes the diagonal matrix whose i^{th} element δ_i^p represents the number of times the i^{th} row of the original data is sampled and we use subscripted $\boldsymbol{\delta}^p,N$ to distinguish the pairs bootstrap coefficient, variance and residual estimates associated with realized sampling frequencies $\boldsymbol{\delta}^p$ ($\hat{\boldsymbol{\beta}}^p_{\boldsymbol{\delta}^p,N}, \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}^p_{\boldsymbol{\delta}^p,N})$ & $\hat{\boldsymbol{\varepsilon}}^p_{\boldsymbol{\delta}^p,N}$) from those of the original sample ($\hat{\boldsymbol{\beta}}_N, \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)$ & $\hat{\boldsymbol{\varepsilon}}_N$).

Wu's (1986) bootstrap, commonly known as the wild bootstrap, holds the design matrix **X** constant and generates new realizations of the outcome vector **y** by multiplying the estimated residuals by a vector of independently and identically distributed random variables $\boldsymbol{\delta}^w$, so that $\mathbf{y}^w = \mathbf{X}\hat{\boldsymbol{\beta}}_N + \mathbf{D}_{\boldsymbol{\delta}^w}\hat{\boldsymbol{\varepsilon}}_N$ and

$$(3)\ \hat{\boldsymbol{\beta}}^w_{\boldsymbol{\delta}^w,N} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^w = \hat{\boldsymbol{\beta}}_N + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}^w}\hat{\boldsymbol{\varepsilon}}_N$$

$$\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta}^w,N} = \mathbf{D}_{\boldsymbol{\delta}^w}\hat{\boldsymbol{\varepsilon}}_N + \mathbf{X}(\hat{\boldsymbol{\beta}}_N - \hat{\boldsymbol{\beta}}^w_{\boldsymbol{\delta}^w,N}),\ \text{and}\ \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{\boldsymbol{\delta}^w,N}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta}^w,N}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta}^w,N}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

where we again use subscripted $\boldsymbol{\delta}^w,N$ to distinguish wild bootstrap estimates associated with a realized vector $\boldsymbol{\delta}^w$ from those of the original sample. We use the common notation **δ**, distinguished by superscripted p or w, for seemingly dissimilar objects in the pairs and wild bootstrap because these operate identically in the theorems and proofs below.

Our interest is in deriving sufficient conditions for the conditional consistency of both

[2]The on-line appendix proves consistency for sub-sampling, with and without replacement, $M < N$ observations.

bootstraps in this inid framework. Specifically, we show that White's (1980) assumptions, which guarantee that for the OLS coefficient estimates and heteroskedasticity robust covariance estimate:

$$(4)\quad \left(\frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\sqrt{N}(\hat{\boldsymbol{\beta}}_N-\boldsymbol{\beta})\overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\mathbf{n}_K \quad \& \quad N(\hat{\boldsymbol{\beta}}_N-\boldsymbol{\beta})'[N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)]^{-1}(\hat{\boldsymbol{\beta}}_N-\boldsymbol{\beta})\overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\chi^2_K,$$

where $\overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ denotes convergence in distribution across the realizations of ($\mathbf{X}$,$\boldsymbol{\varepsilon}$), $\mathbf{n}_K$ the K-variate iid standard normal, χ^2_K the chi-squared with K degrees of freedom, and $\mathbf{A}^{½}$ the "square root" of symmetric positive definite matrix $\mathbf{A}$,[3] are also sufficient to ensure that for the bootstrapped coefficient and heteroskedasticity robust covariance estimates, with $b = p$ (pairs) or w (wild)

$$(5)\quad \left(\frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\sqrt{N}(\hat{\boldsymbol{\beta}}^b_{\boldsymbol{\delta}^b,N}-\hat{\boldsymbol{\beta}}_N)\overset{d(\boldsymbol{\delta}^b)|p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\mathbf{n}_K$$

$$\& \quad N(\hat{\boldsymbol{\beta}}^b_{\boldsymbol{\delta}^b,N}-\hat{\boldsymbol{\beta}}_N)'[N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}^b_{\boldsymbol{\delta}^b,N})]^{-1}(\hat{\boldsymbol{\beta}}^b_{\boldsymbol{\delta}^b,N}-\hat{\boldsymbol{\beta}}_N)\overset{d(\boldsymbol{\delta}^b)|p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\chi^2_K$$

where $\overset{d(\boldsymbol{\delta}^b)|p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ denotes convergence in distribution across $\boldsymbol{\delta}^b$ in probability across realizations of ($\mathbf{X}$,$\boldsymbol{\varepsilon}$). These results show that the conditional distribution given the data ($\mathbf{X}$,$\boldsymbol{\varepsilon}$) of the bootstrap equals the distribution of the OLS estimates across ($\mathbf{X}$,$\boldsymbol{\varepsilon}$), allowing for asymptotically valid inference using the percentiles of bootstrapped coefficient estimates or Wald statistics.[4]

The key characteristic exploited in proofs below is that any of the row permutations of a given vector $\boldsymbol{\delta}$ is equally likely. Consequently, the distribution of the bootstraps can be thought of as the distribution across permutations of $\boldsymbol{\delta}$ integrated across the ordered realizations of $\boldsymbol{\delta}$. Permutation theorems characterize this distribution as asymptotically normal with covariance matrix $N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)$ provided ($\mathbf{X}$,$\boldsymbol{\varepsilon}$) and $\boldsymbol{\delta}$ have certain moment properties. White's (1980) assumptions and the properties of $\boldsymbol{\delta}$ are sufficient to guarantee these properties hold in probability. Since White's assumptions also imply that $\sqrt{N}(\hat{\boldsymbol{\beta}}_N-\boldsymbol{\beta})$ is asymptotically normally distributed with variance $N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)$, this ensures consistency of the bootstraps.

[3]With $\mathbf{E}$ equal to the matrix of eigenvectors and $\boldsymbol{\Lambda}$ the diagonal matrix of eigenvalues of $\mathbf{A}$, $\mathbf{A}^{½} = \mathbf{E}\boldsymbol{\Lambda}^{½}\mathbf{E}'$, where $\boldsymbol{\Lambda}^{½}$ is the diagonal matrix with entries equal to the square root of those of $\boldsymbol{\Lambda}$.

[4]Although, as noted by Cavaliere and Georgiev (2020), even when conditional consistency does not hold valid inference using the bootstrap is still possible if the unconditional limit distribution of the sample test statistic equals the average of the random limit distribution of the bootstrap given the data.

III. Foundational Theorems

This paper makes use of a theorem first proven by Wald & Wolfowitz (1944) and later refined by Noether (1949) and Hoeffding (1951):

Theorem I: Let $\mathbf{z}' = (z_1, \ldots, z_N)$ and $\boldsymbol{\delta}' = (d_1, \ldots, d_N)$ denote sequences of real numbers, not all equal, and $\mathbf{d}' = (d_1, \ldots, d_N)$ any of the $N!$ equally likely permutations of $\boldsymbol{\delta}$. If for all integer $\tau > 2$

$$\text{(Ia)}\ \lim_{N\to\infty} \frac{N^{\frac{\tau}{2}-1}\sum_{i=1}^{N}[z_i - m(z_i)]^{\tau}\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^{\tau}}{\left(\sum_{i=1}^{N}[z_i - m(z_i)]^2\right)^{\tau/2}\left(\sum_{i=1}^{N}[\delta_i - m(\delta_i)]^2\right)^{\tau/2}} = 0,$$

then as $N \to \infty$ the distribution of the random variable

$$\text{(Ib)}\ v_N = \sum_{i=1}^{N}\frac{[z_i - m(z_i)][d_i - m(d_i)]}{s(z_i)s(d_i)N^{½}},$$

$$\text{where for } h = z \text{ or } d,\ m(h_i) = \sum_{i=1}^{N}\frac{h_i}{N}\ \ \&\ \ s(h_i)^2 = \sum_{i=1}^{N}\frac{[h_i - m(h_i)]^2}{N},$$

as calculated across the realizations of $\mathbf{d}$ converges to that of the standard normal. That is,

$$\text{(Ic)}\ v_N \xrightarrow{d(\mathbf{d})} n,$$

where $\xrightarrow{d(\mathbf{d})}$ denotes convergence in distribution across the permutations $\mathbf{d}$ of $\boldsymbol{\delta}$ and n is the standard normal.

The proof is based upon showing that the moments of v_N converge to those of the standard normal. A simple multivariate extension, proven in the on-line appendix, is that if $\mathbf{Z}' = (\mathbf{z}_1,\ldots,\mathbf{z}_N)$ is a sequence of K x 1 vectors and $\mathbf{O} = \mathbf{I} - \mathbf{1}\mathbf{1}'/N$ the centering matrix,[5] then

$$\text{(Id)}\ \mathbf{v}_N = \left(\frac{\mathbf{Z}'\mathbf{O}\mathbf{Z}}{N}\frac{\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-½}\frac{(\mathbf{Z}'\mathbf{O}\mathbf{d})}{\sqrt{N}}$$

is asymptotically distributed multivariate iid standard normal if (Ia) holds for each element in the vector sequence $\mathbf{z}_i$ and for all N sufficiently large $\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}'$ is non-zero and the correlation matrix $\mathbf{D}_{\mathbf{Z'OZ}}^{-½}\mathbf{Z}'\mathbf{O}\mathbf{Z}\mathbf{D}_{\mathbf{Z'OZ}}^{-½}$, where $\mathbf{D}_{\mathbf{Z'OZ}}$ is the diagonal matrix with diagonal entries equal to those of $\mathbf{Z}'\mathbf{O}\mathbf{Z}$, is non-singular with determinant > Δ (a positive constant).

[5]Where $\mathbf{I}$ denotes the N x N identity matrix & $\mathbf{1}$ an N x 1 vector of ones.

Theorem I is easily extended to a probabilistic environment by noting the following result due to Ghosh (1950) that allows us to translate the almost sure characteristics of an infinite number of moment conditions into an almost sure statement regarding a distribution:

Theorem II: If all the moments of the cumulative distribution function $F_N(x)$ converge almost surely (in probability) to those of $F(x)$ which possesses a density function and for which, with ν_{k+1} denoting the absolute moment of order k+1,

$$\text{(IIa)}\ \lim_{k\to\infty}\frac{\alpha^{k+2}\nu_{k+1}}{k+2!}=0\ \text{ for any given value of } \alpha,$$

then $F_N(x)$ converges almost surely (in probability) to $F(x)$.

Condition (IIa) is of course true for the normal distribution. Hoeffding (1952) generalized the result by showing that condition (IIa) is not even needed for convergence in probability at all points of continuity of any $F(x)$ that is uniquely determined by its moments. By virtue of the Cramér-Wold device, Theorem II covers the multivariate case given in (Id) above, as for all $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda}'\boldsymbol{\lambda}=1$, all moments of $\boldsymbol{\lambda}'\mathbf{v}_N$ converge to those of the standard normal. We apply Theorem II in Theorem I using the notation $\xrightarrow{d(\mathbf{d})|p(\boldsymbol{\delta},\mathbf{z})}$, i.e. in probability across the realizations of $\boldsymbol{\delta}$ and $\mathbf{z}$ the distribution of $\mathbf{v}_N$ across permutations $\mathbf{d}$ converges to the multivariate standard normal. Theorems I and II are used below to characterize the asymptotic distribution of $\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}}\hat{\boldsymbol{\varepsilon}}/\sqrt{N}$ in the formula for the bootstrap coefficient estimates $\hat{\boldsymbol{\beta}}_{\boldsymbol{\delta},N}$ in (2) and (3) above.

A less demanding form of Theorem I, proven in the appendix below, provides a weaker condition under which the mean of products converges in probability across permutations to the product of means:

Theorem III: Let $\mathbf{z}' = (z_1, \ldots, z_N)$ and $\boldsymbol{\delta}' = (d_1, \ldots, d_N)$ denote sequences of real numbers, possibly all equal, and $\mathbf{d}' = (d_1, \ldots, d_N)$ any of the N! equally likely permutations of $\boldsymbol{\delta}$. If

$$\text{(IIIa)}\ \lim_{N\to\infty}\frac{\sum_{i=1}^{N}\frac{[z_i-m(z_i)]^2}{N}\sum_{i=1}^{N}\frac{[\delta_i-m(\delta_i)]^2}{N}}{N}=0.$$

then as $N\to\infty$

$$\text{(IIIb)}\ m(z_id_i)-m(z_i)m(d_i)=\sum_{i=1}^{N}\frac{z_id_i}{N}-\sum_{i=1}^{N}\frac{z_i}{N}\sum_{i=1}^{N}\frac{d_i}{N}\xrightarrow{p(\mathbf{d})}0,$$

where $\xrightarrow{p(\mathbf{d})}$ denotes convergence in probability across permutations $\mathbf{d}$ of $\boldsymbol{\delta}$.

Theorem III is used in proofs to make statements regarding the convergence in probability of terms such as $\mathbf{X}'\mathbf{D}_{\boldsymbol{\delta}}\mathbf{X}/N$ in the pairs bootstrap coefficient and $\mathbf{X}'\boldsymbol{\Delta}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta},N}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta},N}}\boldsymbol{\Delta}\mathbf{X}/N$ and $\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta},N}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta},N}}\mathbf{X}/N$ in the pairs and wild bootstrap variance estimates in (2) and (3) above. As (IIIa), based as it is on the moments of **z** and **δ**, will hold in probability, we use the notation $\overset{p(\mathbf{d})|p(\boldsymbol{\delta},\mathbf{z})}{\rightarrow}$, i.e. in probability across the realizations of **δ** and **z**, $m(z_i d_i)$ converges in probability across the permutations **d** of **δ** to $m(z_i)m(d_i)$.

IV. Sufficient Conditions for Bootstrap Consistency with INID Data

The consistency of the bootstrap in this paper follows from two theorems, proven in the appendix below. The first establishes sufficient conditions for the distribution of the bootstrap across permutations **d** of **δ** to converge to the normal:

Theorem IV: Let τ denote any integer greater than 2 and assume that for the wild bootstrap $E[\delta_i^w]=0$, $E[(\delta_i^w)^2]=1$ and $E[(\delta_i^w)^\tau]<\infty$ for all τ. Moreover, let k denote any of the 1...K regressors and assume that the following limiting conditions on the regressors and OLS estimated residuals hold

$$\text{(IVa)}\quad \frac{\mathbf{X}'\mathbf{X}}{N}-\mathbf{V}_{N_1}\overset{p(\mathbf{X})}{\rightarrow}\mathbf{0}_{KxK}\quad \&\quad \frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}}{N}-\mathbf{V}_{N_2}\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\mathbf{0}_{KxK}$$

$$\text{where } \mathbf{V}_{N_i}\ (i=1,2) \text{ is bounded and non-singular with}$$
$$\text{determinant}(\mathbf{V}_{N_i})>\eta_i>0 \text{ for all } N \text{ sufficiently large;}$$

$$\text{(IVb)}\quad \forall\ k\ \&\ \tau:\ \frac{\sum_{i=1}^{N}x_{ik}^{\tau}\hat{\varepsilon}_i^{\tau}}{\left(\sum_{i=1}^{N}x_{ik}^{2}\hat{\varepsilon}_i^{2}\right)^{\tau/2}}\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}0;$$

$$\text{(IVc)}\ \forall\ k:\ \sum_{i=1}^{N}\frac{x_{ik}^4}{N^2}\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}0;\quad \text{(IVd)}\ \forall\ k:\ \sum_{i=1}^{N}\frac{x_{ik}^8}{N^4}\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}0;$$

where $\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ denotes convergence (element by element in the case of matrices) in probability across the distribution of (**X**,**ε**) and $\mathbf{0}_{KxK}$ a KxK matrix of zeros. Then for $b=p$ (pairs) or w (wild) and **d** denoting the permutation of $\boldsymbol{\delta}^b$

$$\text{(IVe)}\left(\frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\left(\frac{\boldsymbol{\delta}^{b\prime}\mathbf{O}\boldsymbol{\delta}^{b}}{N}\right)^{-½}\sqrt{N}(\hat{\boldsymbol{\beta}}^{b}_{\mathbf{d},N}-\hat{\boldsymbol{\beta}}_N)\overset{d(\mathbf{d})|p(\boldsymbol{\delta}^b,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\mathbf{n}_K$$

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

where as described earlier above the notation $\overset{d(\mathbf{d})|p(\boldsymbol{\delta},\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ and $\overset{p(\mathbf{d})|p(\boldsymbol{\delta},\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ denote convergence in distribution and probability across permutations **d** of **δ** in probability across the realizations of (**δ**,**X**,**ε**).

The second theorem notes that the conditions given in White (1980) sufficient to ensure that OLS coefficient estimates are asymptotically distributed multivariate normal with the heteroskedasticity robust covariance estimate also ensure that (IVa) - (IVd) hold almost surely and hence in probability as well:

Theorem V: Assume as in White (1980) that the following conditions hold

(Va) $(\mathbf{x}'_i,\varepsilon_i)$ is a sequence of independent but not necessarily identically distributed random vectors such that $E(\mathbf{x}_i\varepsilon_i)=\mathbf{0}_{Kx1}$.

(Vb) (i) There exist positive finite constants γ_1 and Δ_1 such that for all i, $E(|\varepsilon_i^2|^{1+\gamma_1})<\Delta_1$ and $E(|x_{ij}x_{ik}|^{1+\gamma_1})<\Delta_1$ for all j, k = 1...K; (ii) $\mathbf{V}_{N_1}=N^{-1}\sum_{i=1}^{N}E(\mathbf{x}_i\mathbf{x}'_i)$ is non-singular with determinant($\mathbf{V}_{N_1}$) > η_1 > 0 for all N sufficiently large.

(Vc) (i) There exist positive finite constants γ_2 and Δ_2 such that for all i, $E(|\varepsilon_i^2x_{ij}x_{ik}|^{1+\gamma_2})<\Delta_2$ for all j, k = 1...K; (ii) $\mathbf{V}_{N_2}=N^{-1}\sum_{i=1}^{N}E(\varepsilon_i^2\mathbf{x}_i\mathbf{x}'_i)$ is non-singular with determinant($\mathbf{V}_{N_2}$) > η_2 > 0 for all N sufficiently large.

(Vd) There exist positive finite constants γ_3 and Δ_3 such that for all i, $E(|x_{ij}^2x_{ik}x_{il}|^{1+\gamma_3})<\Delta_3$ for all j, k, l = 1...K or, equivalently, $E(|x_{ij}^4|^{1+\gamma_3})<\Delta_3$ for all j = 1...K.

Then (IVa) - (IVd) hold almost surely and, as in White (1980),

$$\left(\frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\sqrt{N}(\hat{\boldsymbol{\beta}}_N-\hat{\boldsymbol{\beta}})\overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}\mathbf{n}_K\,,$$

where $\overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ denotes convergence in distribution across (**X**,**ε**).

In combination, Theorems IV and V establish that White's conditions are sufficient to ensure conditional consistency of the bootstraps. Let $\boldsymbol{\delta}^*$ denote the ordered values of **δ**. In probability, across permutations **d** of $\boldsymbol{\delta}^*$ (IVe) and (IVf) hold. These permutations, integrated across the distribution of $\boldsymbol{\delta}^*$, characterize the entire distribution of $\boldsymbol{\delta}$. Adding the result[6]

$$(6)\ \frac{\boldsymbol{\delta}^{b\prime}\mathbf{O}\boldsymbol{\delta}^b}{N} \overset{p(\boldsymbol{\delta}^b)}{\rightarrow} 1 \text{ [for } b = p \text{ or } w],$$

and noting that $(\mathbf{X}'\mathbf{X}/N)(\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}/N)^{-1}(\mathbf{X}'\mathbf{X}/N) = (N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N))^{-1}$, this implies that

$$(7)\left(\frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}}{N}\right)^{-1/2}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\sqrt{N}(\hat{\boldsymbol{\beta}}^b_{\boldsymbol{\delta}^b,N} - \hat{\boldsymbol{\beta}}_N) \overset{d(\boldsymbol{\delta}^b)|p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{n}_K$$

$$\& \ N(\hat{\boldsymbol{\beta}}^b_{\boldsymbol{\delta}^b,N} - \hat{\boldsymbol{\beta}}_N)'[N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}^b_{\boldsymbol{\delta}^b,N})]^{-1}(\hat{\boldsymbol{\beta}}^b_{\boldsymbol{\delta}^b,N} - \hat{\boldsymbol{\beta}}_N) \overset{d(\boldsymbol{\delta}^b)|p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \chi^2_K.$$

When combined with White's (1980) result regarding the asymptotic distribution of OLS coefficient estimates, this gives (4) & (5) earlier, i.e. that the conditional distributions of the bootstrapped coefficients and Wald statistics converge to the unconditional distributions of their OLS regression counterparts.

V. Remarks

For an OLS model with iid data and potentially heteroskedastic residuals, Mammen's (1993) results show that the wild bootstrap is in probability consistent for a linear combination of a fixed number of regressors given $\sup_{\|\mathbf{c}\|=1} E[(\mathbf{c}'\mathbf{x}_i)^4(1+\varepsilon_i^2)] < \infty$ and a Lindeberg condition. For the same model, Freedman (1981) proved consistency of the pairs bootstrap in probability if the row vectors $(\mathbf{x}_i', y_i)$ are independently and identically distributed and $E[((\mathbf{x}_i', y_i)(\mathbf{x}_i', y_i)')^2] < \infty$. By adopting a permutation approach, Theorems IV and V extend consistency to inid data at the cost of slightly higher than fourth moments for the x_{ik}, while requiring only slightly greater than second moments of the ε_i, which are by and large weaker conditions. It should be noted, however, that Mammen's result was part of a broader framework that allowed for a growing number of regressors, while Freedman allowed for sub-sampling $M < N$ observations. As shown in the on-line appendix, at the expense of complicating the proofs the permutation based consistency results can be extended to sub-sampling, with and without replacement, if for some γ^* > $\max((1+\gamma_2)^{-1},(1+\gamma_3)^{-1})$, $M(N)$ is such that $\lim_{\inf} M/N^{\gamma^*} > 0$.

[6]For the wild bootstrap, (6) follows immediately from the assumptions on moments in Theorem IV. The proof for the pairs bootstrap is somewhat lengthy and given in the on-line appendix.

Liu (1988) proved the second central moment of the wild OLS bootstrap coefficient distribution with inid data and bounded regressors converges to the asymptotic variance of the OLS coefficients provided $E[\delta_i^w]=0$ and $E[(\delta_i^w)^2]=1$. With the additional assumption that $E[(\delta_i^w)^\tau]<\infty$ for all integer $\tau > 2$, the theorems above extend these results to convergence of the conditional distribution of bootstrapped coefficient and Wald statistics to the unconditional distribution of the OLS regression counterparts with unbounded inid data. As shown in the on-line appendix, the moment conditions on δ_i^w for the permutation based proof can be weakened to only requiring that $E[\delta_i^w]=0$, $E[(\delta_i^w)^2]=1$ and $E[(\delta_i^w)^{2(1+\theta_1)}]<\Delta$ for some finite Δ and $\theta_1 \geq \max(1,1/\gamma_2+\kappa,1/\gamma_3+\kappa)$, for some $\kappa > 0$ and γ_2 & γ_3 as given in Theorem V. There are no apparent advantages, however, from drawing δ_i^w from a distribution without bounded higher moments.

Bibliography

Bickel, P.J. and D.A. Freedman (1983). "Bootstrapping Regression Models with Many Parameters." In A Festschrift for Erich L. Lehmann in Honor of his Sixty-fifth Birthday (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.). Wadsworth, Belmont, CA.

Cavaliere, Giuseppe and Iliyan Georgiev (2020). "Inference under Random Limit Bootstrap Measures." *Econometrica* 88 (6): 2547-2574.

Freedman, D.A. (1981). "Bootstrapping Regression Models". *The Annals of Statistics* 9 (6): 1218-1228.

Galambos, Janos (1987). The Asymptotic Theory of Extreme Order Statistics, 2nd edition. Malambar Florida: Robert E. Krieger Publishing Co.

Ghosh, M.N. (1950). "Convergence of Random Distribution Functions." *Bulletin of the Calcutta Mathematical Society* 42: 217-226.

Hoeffding, Wassily (1951). "A Combinatorial Central Limit Theorem." *The Annals of Mathematical Statistics*, Vol. 22, No. 4 (Dec. 1951): 558-566.

Hoeffding, Wassily (1952). "The Large Sample Power of Tests Based upon Permutations of Observations." *The Annals of Mathematical Statistics* 23 (2): 169-192.

Liu, Regina Y (1988). "Bootstrap Procedures under some Non-I.I.D. Models". *The Annals of Statistics*, Vol. 16, No. 4: 1696-1708.

Mammen, Enno (1993). "Bootstrap and Wild Bootstrap for High Dimensional Linear Models." *The Annals of Statistics* Vol. 21, No. 1: 255-285.

Noether, Gottfried E (1949). "On a Theorem by Wald and Wolfowitz". *The Annals of Mathematical Statistics*, Vol. 20, No. 3 (Sep. 1949): 455-458.

Wald, Abraham. and Jacob Wolfowitz (1944). "Statistical Tests Based on Permutations of the Observations." *The Annals of Mathematical Statistics*, Vol. 15, No. 4 (Dec. 1944): 358-372.

White, Halbert (1980). "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* Vol. 48, No. 4 (May 1980): 817-838.

White, Halbert (1984). Asymptotic Theory for Econometricians. London: Academic Press.

Wu, C. F. J. (1986). "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis." *The Annals of Statistics* Vol. 14, No. 4: 1261-1295.

Appendix

A. Proof of Theorem III

If either the z_i or δ_i are all identical ($z_i = z$ or $\delta_i = \delta$), Theorem III follows immediately. Assuming this is not the case, we first use the symmetry and equal likelihood of permutations to calculate the expectation of d_i and products of d_i across the row permutations $\mathbf{d}$ of $\boldsymbol{\delta}$:

$$\text{(A1)}\quad E_{\mathbf{d}}(d_i) = \sum_{i=1}^{N} \frac{\delta_i}{N} = m(\delta_i),\; E_{\mathbf{d}}(d_i^2) = \sum_{i=1}^{N} \frac{\delta_i^2}{N} = m(\delta_i^2)$$

$$\& \; E_{\mathbf{d}}(d_i d_{j\neq i}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\delta_i \delta_j}{N(N-1)} - \sum_{i=1}^{N} \frac{\delta_i^2}{N(N-1)} = \frac{m(\delta_i)^2 N}{N-1} - \frac{m(\delta_i^2)}{N-1}.$$

We then calculate the mean and variance of $m(z_i d_i)$ - $m(z_i)m(d_i)$ across the realizations of $\mathbf{d}$:

$$\text{(A2)}\quad E_{\mathbf{d}}(m(z_i d_i) - m(z_i)m(\delta_i)) = \sum_{i=1}^{N} \frac{z_i E_{\mathbf{d}}(d_i)}{N} - m(z_i)m(\delta_i) = 0,$$

$$E_{\mathbf{d}}((m(z_i d_i) - m(z_i)m(\delta_i))^2) = \sum_{i,j=1}^{N} \frac{z_i z_j E_{\mathbf{d}}(d_i d_j)}{N^2} + \sum_{i=1}^{N} \frac{z_i^2 E_{\mathbf{d}}(d_i^2)}{N^2} - m(z_i)^2 m(\delta_i)^2$$

$$= \left(\frac{m(\delta_i)^2 N}{N-1} - \frac{m(\delta_i^2)}{N-1} \right)\left(\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{z_i z_j}{N^2} - \sum_{i=1}^{N} \frac{z_i^2}{N^2} \right) + m(\delta_i^2)\sum_{i=1}^{N} \frac{z_i^2}{N^2} - m(z_i)^2 m(\delta_i)^2$$

$$= \left(\frac{m(\delta_i)^2 N}{N-1} - \frac{m(\delta_i^2)}{N-1} \right)\left(m(z_i)^2 - \frac{m(z_i^2)}{N} \right) + m(\delta_i^2)\frac{m(z_i^2)}{N} - m(z_i)^2 m(\delta_i)^2$$

$$= \frac{[m(z_i^2) - m(z_i)^2][m(\delta_i^2) - m(\delta_i)^2]}{N-1}$$

where subscripted i,j denotes the summation across the two indices excluding ties between them. The last line shows that if (IIIa) holds, then across the permutations **d** of **δ** $m(z_i d_i)$ - $m(z_i)m(d_i)$ converges in mean square and hence in probability to 0, as stated in Theorem III.

B. Proof of Theorem IV

We begin by noting the following Lemma.

Lemma: Let $\overset{p(\boldsymbol{\delta}^b)}{\rightarrow}$ denote convergence in probability across the distribution of $\boldsymbol{\delta}^b$ [$b = p$ or w], τ any integer greater than 2, and $\mu_{b,}^3\ \mu_{b,}^4$ ... sequences of constants < ∞. Then:

$$\text{(L1w) } m(\delta_i^w) \overset{p(\boldsymbol{\delta}^w)}{\rightarrow} 0,\ m((\delta_i^w)^2) \overset{p(\boldsymbol{\delta}^w)}{\rightarrow} 1;\ \ \text{(L1p) } m(\delta_i^p) = 1.$$

$$\text{(L2b)} \sum_{i=1}^{N} \frac{[\delta_i^b - m(\delta_i^b)]^2}{N} \overset{p(\boldsymbol{\delta}^b)}{\rightarrow} 1;\ \ \&\ \ \text{(L3b) } m((\delta_i^b)^\tau) \overset{p(\boldsymbol{\delta}^b)}{\rightarrow} \mu_b^\tau.$$

The results for $\boldsymbol{\delta}^w$ follow from the Strong Law of Large Numbers and the bounds on the moments of iid δ_i^w in Theorem IV. The proof for $\boldsymbol{\delta}^p$ is more involved and is given in the on-line appendix. We also make use below of a corollary to the Continuous Mapping Theorem given by White (1984) that extends the result to sequences that do not necessarily converge to a constant:

Corollary to Continuous Mapping Theorem: Let g: $\mathbb{R}^k \rightarrow \mathbb{R}^l$ be continuous on a compact set $C \subset \mathbb{R}^k$. Suppose that $b_N(\omega)$ and c_N are kx1 vectors such that $b_N(\omega)$ - $c_N \overset{p}{\rightarrow} 0$, and for all N sufficiently large, c_N is interior to C, uniformly in N. Then $g(b_N(\omega))$ - $g(c_N) \overset{p}{\rightarrow} 0$.

This corollary allows us to follow White (1984) in taking (IVa) in the text as indicating that in probability $\mathbf{X}'\mathbf{X}/N$ & $\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_N}\mathbf{X}/N$ are asymptotically invertible with determinants strictly greater than some $\eta > 0$.

For a permutation **d** of $\boldsymbol{\delta}^w$ or $\boldsymbol{\delta}^p$, the coefficient estimates of the pairs and wild bootstrap are, following (2) and (3) in the text, given by $\sqrt{N}(\hat{\boldsymbol{\beta}}_{\mathbf{d},N}^p - \hat{\boldsymbol{\beta}}_N) = \mathbf{C}^{-1}\mathbf{a}$ and $\sqrt{N}(\hat{\boldsymbol{\beta}}_{\mathbf{d},N}^w - \hat{\boldsymbol{\beta}}_N) =$ $(\mathbf{X}'\mathbf{X}/N)^{-1}\mathbf{a}$, where $\mathbf{C} = \mathbf{X}'\mathbf{D}_\mathbf{d}\mathbf{X}/N$, $\mathbf{a} = \mathbf{X}'\mathbf{D}_\mathbf{d}\hat{\boldsymbol{\varepsilon}}/\sqrt{N}$ and we simplify notation here and later by dropping the subscript N on $\hat{\boldsymbol{\varepsilon}}$. Regarding the jk^{th} element of **C**, given by $\sum_{i=1}^{N} x_{ij}x_{ik}d_i/N$, we can apply Theorem III with $z_i = x_{ij}x_{ik}$. Condition IIIa in this case requires that:

$$\text{(B1)} \frac{m(x_{ij}^2 x_{ik}^2) - m(x_{ij}x_{ik})^2}{N}\left[m((\delta_i^p)^2) - m(\delta_i^p)^2\right] \overset{p(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

Noting that $m(x_{ij}x_{ik})$ is in probability bounded from (IVa), given (L2b) above we see that (B1) is satisfied if (using the Cauchy-Schwarz Inequality here, and frequently later as well)

$$\text{(B2)} \sum_{i=1}^{N} \frac{x_{ij}^2 x_{ik}^2}{N^2} \leq \sqrt{\sum_{i=1}^{N} \frac{x_{ij}^4}{N^2} \sum_{i=1}^{N} \frac{x_{ik}^4}{N^2}} \overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

which is guaranteed by assumption (IVc) in Theorem IV. So,

$$\text{(B3)} \underbrace{\frac{\mathbf{X}'\mathbf{D}_{\mathbf{d}}\mathbf{X}}{N}}_{\mathbf{C}} - \frac{\mathbf{X}'\mathbf{X}}{N} \underbrace{m(\delta_i^p)}_{=1} \overset{p(\mathbf{d})|p(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}.$$

Noting that the k^{th} element of **a** equals $\sum_{i=1}^{N} x_{ik}\hat{\varepsilon}_i d_i / \sqrt{N}$, we apply the multivariate extension of Theorem I in the text with $z_{ik} = x_{ik}\hat{\varepsilon}_i$, or $\mathbf{Z} = \mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}$. As the mean of z_{ik} is zero, we have $\mathbf{Z}'\mathbf{O}\mathbf{Z} = \mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}$ and $\mathbf{Z}'\mathbf{O}\mathbf{d} = \mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{d} = \mathbf{X}'\mathbf{D}_{\mathbf{d}}\hat{\boldsymbol{\varepsilon}}$. From (L2b) we know that in probability $\mathbf{d}'\mathbf{O}\mathbf{d} = \boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}$ is non-zero, while assumption (IVa) and the corollary to the Continuous Mapping Theorem ensure that $\mathbf{D}^{-½}_{\mathbf{X}'\mathbf{D}_{\hat{\varepsilon}}\mathbf{D}_{\hat{\varepsilon}}\mathbf{X}}\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}\mathbf{D}^{-½}_{\mathbf{X}'\mathbf{D}_{\hat{\varepsilon}}\mathbf{D}_{\hat{\varepsilon}}\mathbf{X}}$ is in probability non-singular with determinant greater than some $\Delta > 0$.[7] Hence, following Theorems I and II, the distribution across **d** of

$$\text{(B4)} \left(\frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}}{N}\right)^{-½} \left(\frac{\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-½} \frac{\mathbf{X}'\mathbf{D}_{\mathbf{d}}\hat{\boldsymbol{\varepsilon}}}{\sqrt{N}}$$

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

$$\text{(B5)} \frac{\sum_{i=1}^{N} x_{ik}^{\tau}\hat{\varepsilon}_i^{\tau}}{\left(\sum_{i=1}^{N} x_{ik}^2\hat{\varepsilon}_i^2\right)^{\tau/2}} \frac{\sum_{i=1}^{N} \frac{[\delta_i - m(\delta_i)]^{\tau}}{N}}{\left(\sum_{i=1}^{N} \frac{[\delta_i - m(\delta_i)]^2}{N}\right)^{\tau/2}} \overset{p(\boldsymbol{\delta},\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

which given the Lemma is satisfied if (IVb) in Theorem IV holds.

Using the preceding and the fact that $\boldsymbol{\delta}'\mathbf{O}\boldsymbol{\delta}/N = \mathbf{d}'\mathbf{O}\mathbf{d}/N$ is a scalar, we have:

---

[7] By (IVa) $\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}/N$ is bounded with determinant $> \eta > 0$. Let κ denote the upper bound on the diagonal elements. By the trace property of eigenvalues, we know that the largest eigenvalue is less than $K\kappa$, and hence the smallest must be greater than $\eta/(K\kappa)^{K-1}$. The smallest eigenvalue of $\mathbf{D}^{-½}_{\mathbf{X}'\mathbf{D}_{\hat{\varepsilon}}\mathbf{D}_{\hat{\varepsilon}}\mathbf{X}/N}$ is greater than $\kappa^{-½}$. As the smallest eigenvalue of **AB** is greater than or equal to the product of their smallest eigenvalues, we have that the smallest eigenvalue of $\mathbf{D}^{-½}_{\mathbf{X}'\mathbf{D}_{\hat{\varepsilon}}\mathbf{D}_{\hat{\varepsilon}}\mathbf{X}}\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}\mathbf{D}^{-½}_{\mathbf{X}'\mathbf{D}_{\hat{\varepsilon}}\mathbf{D}_{\hat{\varepsilon}}\mathbf{X}}$ is greater than $\eta/K^{K-1}\kappa^K$, and hence the determinant greater than $(\eta/K^{K-1}\kappa^K)^K$.

$$\text{(B6)}\quad \left(\frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\left(\frac{\boldsymbol{\delta}^{p}{}'\mathbf{O}\boldsymbol{\delta}^{p}}{N}\right)^{-½}\sqrt{N}(\hat{\boldsymbol{\beta}}^{p}_{\mathbf{d},N}-\hat{\boldsymbol{\beta}}_{N}) =$$

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

$$\left(\frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\left(\frac{\boldsymbol{\delta}^{w}{}'\mathbf{O}\boldsymbol{\delta}^{w}}{N}\right)^{-½}\sqrt{N}(\hat{\boldsymbol{\beta}}^{w}_{\mathbf{d},N}-\hat{\boldsymbol{\beta}}_{N}) =$$

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

thereby establishing the claim in (IVe).

Regarding the wild bootstrap heteroskedasticity robust covariance estimates, we have

$$\text{(B7)}\ N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}^{w}_{\mathbf{d},N}) = \left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}\mathbf{A}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1},\quad \text{where } \mathbf{A} = \frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\mathbf{d},N}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}_{\mathbf{d},N}}\mathbf{X}}{N}.$$

Using the formula for $\hat{\boldsymbol{\varepsilon}}_{\boldsymbol{\delta}^{w},N}$ in (3), the jk^{th} element of $\mathbf{A}$ is given by:

$$\text{(B8)}\ \sum_{i=1}^{N}\frac{x_{ij}x_{ik}(d_i\hat{\varepsilon}_i-\sum_{p=1}^{K}\frac{x_{ip}}{\sqrt{N}}\gamma_N^{½}\hat{\eta}_p)^2}{N},\ \left[\text{where } \hat{\boldsymbol{\eta}} = \gamma_N^{-½}\sqrt{N}(\hat{\boldsymbol{\beta}}^{w}_{\mathbf{d},N}-\hat{\boldsymbol{\beta}}_{N})\ \&\ \gamma_N = \frac{\boldsymbol{\delta}^{w}{}'\mathbf{O}\boldsymbol{\delta}^{w}}{N}\right]$$

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

For "a", we note that d_i^2 is the permutation of δ_i^{w2} and apply Theorem III with $z_i = x_{ij}x_{ik}\hat{\varepsilon}_i^2$. Condition (IIIa) requires that:

$$\text{(B9)}\ \frac{m(x_{ij}^2x_{ik}^2\hat{\varepsilon}_i^4)-m(x_{ij}x_{ik}\hat{\varepsilon}_i^2)^2}{N}\left[m(\delta_i^{w4})-m(\delta_i^{w2})^2\right]\overset{p(\boldsymbol{\delta}^{w},\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}0.$$

From the Lemma and (IVa) in Theorem IV, we know that [$m(\delta_i^{w4})-m(\delta_i^{w2})^2$] and $m(x_{ij}x_{ik}\hat{\varepsilon}_i^2)^2$ are in probability bounded. Hence, (B9) is met if

$$\text{(B10)}\ \frac{m(x_{ij}^2x_{ik}^2\hat{\varepsilon}_i^4)}{N} = \sum_{i=1}^{N}\frac{x_{ij}^2x_{ik}^2\hat{\varepsilon}_i^4}{N^2} \le \sqrt{\left(\sum_{i=1}^{N}\frac{x_{ij}^4\hat{\varepsilon}_i^4}{N^2}\right)\left(\sum_{i=1}^{N}\frac{x_{ik}^4\hat{\varepsilon}_i^4}{N^2}\right)}\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}0,$$

which is guaranteed by (IVa) and (IVb) in Theorem IV as

$$\text{(B11)} \sum_{i=1}^{N} \frac{x_{ij}^4 \hat{\varepsilon}_i^4}{N^2} = \overbrace{\frac{\sum_{i=1}^{N} (x_{ij}\hat{\varepsilon}_i)^4}{\left(\sum_{i=1}^{N} x_{ij}^2 \hat{\varepsilon}_i^2\right)^{4/2}}}^{\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (IVb with } \tau=4)} \overbrace{\left(\sum_{i=1}^{N} \frac{x_{ij}^2 \hat{\varepsilon}_i^2}{N}\right)^2}^{\text{in probability bounded (IVa)}} \overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

So, by Theorem III

$$\text{(B12) "}a\text{": } m(x_{ij}x_{ik}\hat{\varepsilon}_i^2 d_i^2) - m(x_{ij}x_{ik}\hat{\varepsilon}_i^2) m(\delta_i^{w2}) \overset{p(\mathbf{d})|p(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0, \text{ where } m(\delta_i^{w2}) \underbrace{\overset{p(\boldsymbol{\delta}^w)}{\rightarrow}}_{\text{by (L1w)}} 1.$$

For "b", we apply Theorem III with $z_i = x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i / \sqrt{N}$ and note that

$$\text{(B13)} \left| m\left(\frac{x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i}{\sqrt{N}}\right) \right| \le \sum_{i=1}^{N} \frac{\left|x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i\right|}{N^{3/2}} \le \sqrt[4]{\underbrace{\sum_{i=1}^{N} \frac{x_{ij}^4}{N^2} \sum_{i=1}^{N} \frac{x_{ik}^4}{N^2}}_{\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (IVc)}}} \sqrt{\underbrace{\sum_{i=1}^{N} \frac{x_{ip}^2 \hat{\varepsilon}_i^2}{N}}_{\text{in probability bounded (IVa)}}} \overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

so again using the Lemma we see that condition (IIIa) in this case is met if

$$\text{(B14)} \frac{m(x_{ij}^2 x_{ik}^2 x_{ip}^2 \hat{\varepsilon}_i^2 / N)}{N} = \sum_{i=1}^{N} \frac{x_{ij}^2 x_{ik}^2 x_{ip}^2 \hat{\varepsilon}_i^2}{N^3} \le \sqrt[4]{\underbrace{\sum_{i=1}^{N} \frac{x_{ij}^8}{N^4} \sum_{i=1}^{N} \frac{x_{ik}^8}{N^4}}_{\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (IVd)}}} \sqrt{\underbrace{\sum_{i=1}^{N} \frac{x_{ip}^4 \hat{\varepsilon}_i^4}{N^2}}_{\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (B11)}}} \overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

which as noted is guaranteed by (IVd) and the result in (B11). By Theorem III we then have

$$\text{(B15) "}b\text{": } m\left(\frac{x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i d_i}{\sqrt{N}}\right) - \underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i}{\sqrt{N}}\right)}_{\text{where } \overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ by (B13)}} \underbrace{m(\delta_i^w)}_{\overset{p(\boldsymbol{\delta}^w)}{\rightarrow} 0 \text{ by (L1w)}} \overset{p(\mathbf{d})|p(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

For "c", we note that

$$\text{(B16) } m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}}{N}\right) \le \sqrt[4]{\sum_{i=1}^{N} \frac{x_{ij}^4}{N^2} \sum_{i=1}^{N} \frac{x_{ik}^4}{N^2} \sum_{i=1}^{N} \frac{x_{ip}^4}{N^2} \sum_{i=1}^{N} \frac{x_{iq}^4}{N^2}} \underbrace{\overset{p(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0}_{\text{by (IVc)}}.$$

From the above, we see that in probability the $\hat{\eta}_p$ in (B8), which from (IVa) and (B6) converge to normal variables with bounded finite variance, are multiplied by γ_N which from (L2b) converges to 1 and by "b" and "c" terms which from (B15) and (B16) converge to zero, and hence when so multiplied converge in probability to zero. Consequently, using (B12) we see that

$$\text{(B17)} \quad \mathbf{A} - \frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}}{N} \overset{p(\mathbf{d})|p(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ and hence } N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{\mathbf{d},N}^w) - N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N) \overset{p(\mathbf{d})|p(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

which establishes (IVf) for the wild bootstrap.

Regarding the pairs bootstrap heteroskedasticity robust covariance estimates, we have

$$\text{(B18)}\ N\hat{\mathbf{V}}(\hat{\boldsymbol\beta}^p_{\mathbf{d},N}) = \left(\frac{\mathbf{X}'\mathbf{D}_\mathbf{d}\mathbf{X}}{N}\right)^{-1}\mathbf{B}\left(\frac{\mathbf{X}'\mathbf{D}_\mathbf{d}\mathbf{X}}{N}\right)^{-1},\ \text{where } \mathbf{B} = \frac{\mathbf{X}'\boldsymbol\Delta'\mathbf{D}_{\hat{\boldsymbol\varepsilon}_{\mathbf{d},N}}\mathbf{D}_{\hat{\boldsymbol\varepsilon}_{\mathbf{d},N}}\boldsymbol\Delta\mathbf{X}}{N}$$

and where it should be kept in mind that since $\mathbf{D}_{\boldsymbol\delta^p} = \boldsymbol\Delta'\boldsymbol\Delta$, the permutation $\mathbf{d}$ of $\boldsymbol\delta^p$ is associated with a conforming permutation of the columns of $\boldsymbol\Delta$. Using the formula for $\hat{\boldsymbol\varepsilon}_{\boldsymbol\delta^p,N}$ in (2), the jk^{th} element of $\mathbf{B}$ is given by

$$\text{(B19)}\ \sum_{i=1}^{N}\frac{x_{ij}x_{ik}d_i(\hat\varepsilon_i - \sum_{p=1}^{K}\frac{x_{ip}}{\sqrt{N}}\gamma_N^{½}\hat\eta_p)^2}{N},\left[\text{where } \hat{\boldsymbol\eta} = \gamma_N^{-½}\sqrt{N}(\hat{\boldsymbol\beta}^p_{\mathbf{d},N} - \hat{\boldsymbol\beta}_N)\ \&\ \gamma_N = \frac{\boldsymbol\delta^{p}{}'\mathbf{O}\boldsymbol\delta^p}{N}\right]$$

$$= \underbrace{m(x_{ij}x_{ik}\hat\varepsilon_i^2 d_i)}_{d} - 2\sum_{p=1}^{K}\gamma_N^{½}\hat\eta_p \underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}\hat\varepsilon_i d_i}{\sqrt{N}}\right)}_{e} + \sum_{p=1}^{K}\sum_{q=1}^{K}\gamma_N\hat\eta_p\hat\eta_q \underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}d_i}{N}\right)}_{f}.$$

For "*d*", we apply Theorem III with $z_i = x_{ij}x_{ik}\hat\varepsilon_i^2$ and using the Lemma, (B10) and (IVa) see that condition IIIa is met, so

$$\text{(B20)}\ "d":\ m(x_{ij}x_{ik}\hat\varepsilon_i^2 d_i) - m(x_{ij}x_{ik}\hat\varepsilon_i^2)\underbrace{m(\delta_i^p)}_{=1\,(\text{L1p})} \overset{p(\mathbf{d})|p(\boldsymbol\delta^p,\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow} 0.$$

For "*e*", we apply Theorem III with $z_i = x_{ij}x_{ik}x_{ip}\hat\varepsilon_i/\sqrt{N}$ and using the Lemma, (B13) and (B14) see that condition IIIa is met and

$$\text{(B21)}\ "e":\ m\left(\frac{x_{ij}x_{ik}x_{ip}\hat\varepsilon_i d_i}{\sqrt{N}}\right) - \underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}\hat\varepsilon_i}{\sqrt{N}}\right)}_{\text{which by (B13)} \overset{p(\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow} 0}\underbrace{m(\delta_i^p)}_{=1\,(\text{L1p})} \overset{p(\mathbf{d})|p(\boldsymbol\delta^p,\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow} 0.$$

For "*f*" we apply Theorem III with $z_i = x_{ij}x_{ik}x_{ip}x_{iq}/N$ and see that condition IIIa holds by the Lemma, (B16) and the fact that

$$\text{(B22)}\ \frac{m(x_{ij}^2x_{ik}^2x_{ip}^2x_{iq}^2/N^2)}{N} \le \sqrt[4]{\sum_{i=1}^{N}\frac{x_{ij}^8}{N^4}\sum_{i=1}^{N}\frac{x_{ik}^8}{N^4}\sum_{i=1}^{N}\frac{x_{ip}^8}{N^4}\sum_{i=1}^{N}\frac{x_{iq}^8}{N^4}}\underbrace{\overset{p(\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow} 0}_{\text{by (IVd)}},$$

so

$$\text{(B23)}\ "f":\ m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}d_i}{N}\right) - \underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}}{N}\right)}_{\text{which by (B16)} \overset{p(\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow} 0}\underbrace{m(\delta_i^p)}_{=1\,(\text{L1p})} \overset{p(\mathbf{d})|p(\boldsymbol\delta^p,\mathbf{X},\boldsymbol\varepsilon)}{\rightarrow} 0.$$

Consequently, similar to the case of the wild bootstrap, in (B19) in probability the $\hat{\eta}_p$, which are random normal variables with bounded variances, are all multiplied by γ_N which by (L2b) converge to 1 and "*e*" and "*f*" other terms which by (B21) and (B23) converge to zero, and hence when so multiplied in probability converge to zero as well. Using (B3), we then have

$$\text{(B24)} \quad \mathbf{B} - \frac{\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}}{N} \overset{p(\mathbf{d})|p(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}, \quad \frac{\mathbf{X}'\mathbf{D}_{\mathbf{d}}\mathbf{X}}{N} - \frac{\mathbf{X}'\mathbf{X}}{N} \overset{p(\mathbf{d})|p(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}$$

$$\text{and hence } N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}^p_{\mathbf{d},N}) - N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N) \overset{p(\mathbf{d})|p(\boldsymbol{\delta}^b,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK},$$

which establishes (IVf) for the pairs bootstrap.

C. Proof of Theorem V

We prove that White's assumptions imply that (IVa) - (IVd) hold almost surely. The following corollary to Markov's Law of Large Numbers, given by White (1984), will be useful:

Corollary to Markov's Law: Let z_i be a sequence of independent random variables such that $E(|z_i|^{1+\gamma}) < \Delta < \infty$ for some $\gamma > 0$ and all i. Then

$$m(z_i) - m(E(z_i)) \overset{a.s.}{\rightarrow} 0.$$

as will the following Borel-Cantelli type corollary by Galambos (1987):

Borel-Cantelli Corollary: Let $x_1, x_2, \ldots$ be an infinite sequence of random variables, $F_j(x)$ be the cumulative distribution function of x_j (i.e. $\Pr\{x_j < x\}$), and u_N be a nondecreasing sequence of real numbers such that for all j Prob($x_j < \sup_N u_N$) = 1. Then

$$\sum_{j=1}^{\infty}[1 - F_j(u_j)] < \infty \Rightarrow \text{Prob}(\underset{i \le N}{\text{Max}}\, x_i \ge u_N \text{ infinitely often}) = 0.$$

From the corollary to Markov's Law, and (Va) and (Vb) we see that $\mathbf{X}'\mathbf{X}/N$ converges almost surely to $V_{N_1} = N^{-1}\sum_{i=1}^{N} E(\mathbf{x}_i\mathbf{x}_i')$ which by assumption is bounded non-singular with determinant > η_1 > 0 for all N sufficiently large. Similarly, White (1980) showed that his conditions are sufficient to ensure that $\mathbf{X}'\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{D}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{X}/N$ converges almost surely to $V_{N_2} = N^{-1}\sum_{i=1}^{N} E(\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i')$ in (Vc), which is bounded non-singular with determinant > η_2 > 0 for all N sufficiently large. This establishes (IVa).

Next, we note that:

$$\text{(C1a)}\quad \frac{\left|\sum_{i=1}^{N} x_{ik}^{\tau}\hat{\varepsilon}_i^{\tau}\right|}{\left(\sum_{i=1}^{N} x_{ik}^{2}\hat{\varepsilon}_i^{2}\right)^{\tau/2}} \leq \frac{\sum_{i=1}^{N}\left|x_{ik}^{\tau}\hat{\varepsilon}_i^{\tau}\right|}{\left(\sum_{i=1}^{N} x_{ik}^{2}\hat{\varepsilon}_i^{2}\right)^{\tau/2}} \leq \frac{(\operatorname*{Max}_i x_{ik}^{2}\hat{\varepsilon}_i^{2})^{\frac{\tau}{2}-1}\sum_{i=1}^{N} x_{ik}^{2}\hat{\varepsilon}_i^{2}}{\left(\sum_{i=1}^{N} x_{ik}^{2}\hat{\varepsilon}_i^{2}\right)^{\tau/2}} = \left(\frac{\operatorname*{Max}_i x_{ik}^{2}\hat{\varepsilon}_i^{2}/N}{\sum_{i=1}^{N} x_{ik}^{2}\hat{\varepsilon}_i^{2}/N}\right)^{\frac{\tau}{2}-1}$$

$$\text{(C1b)}\quad \sum_{i=1}^{N}\frac{x_{ik}^{4}}{N^{2}} \leq \frac{\operatorname*{Max}_i x_{ik}^{2}}{N}\sum_{i=1}^{N}\frac{x_{ik}^{2}}{N} \quad \& \quad \sum_{i=1}^{N}\frac{x_{ik}^{8}}{N^{4}} \leq \left(\frac{\operatorname*{Max}_i x_{ik}^{2}}{N}\right)^{3}\sum_{i=1}^{N}\frac{x_{ik}^{2}}{N}.$$

So, to prove (IVb) - (IVd) it is sufficient to show that the right hand sides of the inequalities above converge to zero.

Focusing on (C1b) and (IVc) - (IVd), we have already noted that $\sum_{i=1}^{N} x_{ik}^2/N$ converges almost surely to $\sum_{i=1}^{N} E(x_{ik}^2)/N$, which is bounded from above. Using Markov's inequality and the assumption that $E(|x_{ij}^2|^{1+\gamma_1}) < \Delta_1$ in (Vb), we can state that for some $\gamma > 1/(1+\gamma_1)$ but < 1

$$\text{(C2)}\quad \sum_{i=1}^{\infty} P(x_{ik}^2 \geq N^{\gamma}) \leq \sum_{i=1}^{\infty}\frac{E(|x_{ik}^2|^{1+\gamma_1})}{N^{\gamma(1+\gamma_1)}} < \sum_{i=1}^{\infty}\frac{\Delta_1}{N^{\gamma(1+\gamma_1)}} < \infty.$$

By the Borel-Cantelli corollary, $\text{Max}_i\, x_{ik}^2$ is asymptotically almost surely less than N^{γ} and hence $\text{Max}_i\, x_{ik}^2/N$ almost surely converges to zero, which in conjunction with the almost sure convergence of $\sum_{i=1}^{N} x_{ik}^2/N$ to a bounded value establishes (C1b) and hence (IVc) & (IVd).

Turning to (C1a) and (IVb), as already noted White (1980) showed that his assumptions ensure that $\mathbf{X}'\mathbf{D}_{\hat{\varepsilon}}\mathbf{D}_{\hat{\varepsilon}}\mathbf{X}/N$ converges almost surely to $V_{N_2} = N^{-1}\sum_{i=1}^{N} E(\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i')$, so $\sum_{i=1}^{N} x_{ik}^2\hat{\varepsilon}_i^2/N$ almost surely converges to the bounded diagonal element of V_{N_2} given by $\sum_{i=1}^{N} E(x_{ik}^2\varepsilon_i^2)/N$. From (Vc) and Hölder's inequality we know that $E(|x_{ik}^2\varepsilon_i^2|) < \Delta_2^{1/(1+\gamma_2)} < \infty$. As the sum of the eigenvalues of a matrix equal its trace and the product its determinant, we know that the largest eigenvalue of V_{N_2} in (Vc) is less than $K\Delta_2^{1/(1+\gamma_2)}$, while, since determinant $V_{N_2} > \eta_2$, the smallest eigenvalue must be greater than $\eta_2/K^{K-1}\Delta_2^{(K-1)/(1+\gamma_2)}$. From the Schur-Horn Theorem, we know that the smallest diagonal element of V_{N_2} is greater than or equal to its smallest eigenvalue, and hence the term $\sum_{i=1}^{N} x_{ik}^2\hat{\varepsilon}_i^2/N$ in the denominator of (C1a) is almost surely bounded from below away from zero.

Regarding the max term in (C1a), we have

$$\text{(C3)}\quad x_{ik}^2\hat{\varepsilon}_i^2 = x_{ik}^2(\varepsilon_i + \sum_{j=1}^{K}(\beta_j - \hat{\beta}_j)x_{ij})^2$$

$$\leq x_{ik}^2\varepsilon_i^2 + 2\sum_{j=1}^{K}\left|\beta_j - \hat{\beta}_j\right|\sqrt{x_{ij}^2\varepsilon_i^2 x_{ik}^4} + \sum_{j=1}^{K}\sum_{l=1}^{K}\left|\beta_j - \hat{\beta}_j\right|\left|\beta_l - \hat{\beta}_l\right|\sqrt[4]{x_{ij}^4 x_{il}^4}\sqrt{x_{ik}^4}.$$

Hence, given that White (1980) showed that his conditions ensure that $\hat{\boldsymbol{\beta}}_N \overset{a.s.(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \boldsymbol{\beta}$, to prove that $\text{Max}_i x_{ik}^2\hat{\varepsilon}_i^2 / N$ converges almost surely to zero it is sufficient to show that $\text{Max}_i x_{ik}^2\varepsilon_i^2 / N$ and $\text{Max}_i x_{ik}^4 / N$ converge almost surely to zero for all *k*. However, the assumptions $E(|\,\varepsilon_i^2 x_{ij}x_{ik}\,|^{1+\gamma_2}) < \Delta_2$ and $E(|\,x_{ij}^4\,|^{1+\gamma_3}) < \Delta_3$ in (Vc) and (Vd), by the same argument used in (C2) above, ensure that this is the case, establishing (IVb).