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

Alwyn Young
London School of Economics
This draft: October 2023

Abstract

Analyzing the distributions of the pairs and wild bootstraps as those of permutation statistics provides a different approach to proving consistency, allowing the derivation of new results. For the case of independently but not-necessarily identically distributed (inid) data, this approach reveals moment conditions for consistency which cover more general regression models than earlier inid results and are often less demanding than previous results for independently and identically distributed data.

keywords: bootstrap consistency, permutation distribution

*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 ordinary least squares (OLS) covariance estimate to extend results concerning the consistency of the pairs and wild OLS bootstraps, 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 the realization of the external variable used by the wild bootstrap to transform residuals is equally likely. These equally likely permutations can be used to characterize the bootstrap distributions conditional on the data as normal given 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 OLS bootstrap. 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 coefficient distribution[1], and that of the Wald statistic based upon the (potentially incorrect) assumption of homoskedastic errors. Stute (1990) showed that for consistency of the coefficient distribution alone in the heteroskedastic iid model it is sufficient for both the squared regressors and the product of the squared regressors with the squared errors to have finite expectation. For similarly iid data, Mammen (1993) proved consistency of the wild OLS bootstrap coefficient and

[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).

homoskedasticity-based Wald test distributions with bounded expectations of the product of the fourth power of the regressors with the squared errors and an additional Lindeberg condition. This paper finds consistency of both the coefficient and heteroskedasticity robust Wald statistic distribution in a broader inid environment for both the pairs and wild bootstraps with finite expectations of only slightly more than second powers of the regressors and of the product of the second powers of the regressors with the second power of the errors. These are much less demanding assumptions than those used by Freedman and Mannon, requiring only slightly higher moments than used by Stute for the proof of only the pairs bootstrap coefficient distribution in a narrower iid environment. These results are useful because when data are drawn, for example, from distinct populations, geographic regions or time periods, the iid assumption is less likely to hold.[2] Moreover, the distribution of the homoskedasticity-based Wald test is not pivotal in a heteroskedastic iid or inid environment, as recognized by Freedman (1981) and Mammen (1993), whereas that of the heteroskedasticity robust Wald test is. Bootstrapped pivotal test statistics asymptotically provide higher order accuracy and faster convergence of rejection probabilities to nominal value (Singh 1981, Hall 1992).

For OLS models with inid data, the salient contribution is Liu (1988), who showed that the wild bootstrap provides consistent estimates of the second central moment of a linear combination of coefficients in an OLS regression model with bounded regressors provided the first and second moments of the wild bootstrap external variable are 0 and 1, respectively. Liu's result regarding the second central moment is easily extended to the case of the multivariate second central moments of coefficients for unbounded inid regressors without any additional restrictions on the moments of the external variable, as shown below. Our interest here, however, is in the full distribution of wild bootstrap coefficient and Wald statistic estimates, where our proof requires the existence of higher moments of the wild bootstrap external variable to ensure

[2] As examples: (i) Thornton (2008) used a randomized experiment to investigate the demand for and effects of learning HIV status across north, central and south Malawi, which differ systematically in their ethnicity and religion. (ii) Cai et al (2009) investigated saliency by randomly assigning restaurant arrivals in China to tables with different menu setups; not surprisingly, the total bill paid varies systematically with the time of day.

convergence of higher moments to the normal. As the external variable is selected by the practitioner, and not an exogenous characteristic of the data, these additional moment conditions pose no obstacle. The two point distribution proposed by Mammen (1993) and the Rademacher distribution, both often used in practical application (e.g. Davidson & Flachaire 2008), have moments of all order.

Liu's consideration of inid data has largely not been extended, as the OLS bootstrap literature since has focused on time series dependent data, where the absence of random sampling of independent observations raises different statistical issues and the use of different bootstrap methods (see the review in Hardle et al 2003). Djogbenou et al (2019), who prove consistency of the wild bootstrap t-statistic distribution for independently distributed cluster groupings of data, are a notable exception. With the moment assumptions used here plus the additional requirement of bounded slightly higher than fourth moments of the regressors, their proof allows for heterogeneity in the distribution of data across clusters. However, they limit that heterogeneity in requiring that the cross product of the regressors and the covariance matrix of coefficient estimates converge to matrices of constants, a condition that in other papers is typically motivated by an iid assumption.[3] The data generating process examined in this paper is more fully inid in that there is no restriction that such matrices converge to anything, and the moment assumptions are also less demanding. While the results in this paper are not revolutionary, the use of the permutation distribution allows a common proof of the consistency of the pairs and wild bootstrap distributions of both coefficients and Wald statistics in a fully inid framework with unbounded regressors and by and large less demanding moment conditions than used earlier.

The paper proceeds as follows: Section II reviews the OLS model, White's assumptions and results regarding OLS with inid data, and pairs and wild bootstrap methods for heteroskedastic data. Section III presents the foundational theorems regarding the asymptotic

[3]Canay et al (2021) who examine wild bootstrap consistency when the number of independent cluster groupings is fixed, similarly allow for heterogeneity across clusters while assuming convergence of the full sample cross-product and covariance matrices to matrices of constants and, additionally, convergence of the projection of regressors on each other within each cluster to a common matrix.

normality of permutation distributions that motivate the results. Section IV then combines these with White's (1980) result to derive sufficient conditions for pairs and wild OLS bootstrap consistency with inid data, concluding with remarks that more fully contrast the assumptions and results with those used in the papers cited above. The appendix and on-line appendix provide details of the proofs.

II. Framework and Notation

Our interest is in inference for the linear model $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$, $i = 1... N$, or in matrix form

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

where **y** represents the N x 1 matrix of observations on the dependent (outcome) variable, **X** the N x K matrix of observations of independent variables, $\boldsymbol{\beta}$ the K x 1 vector of unobserved parameters of interest **,** and $\boldsymbol{\varepsilon}$ the N x 1 matrix of unobserved disturbances. The ordinary least squares estimates $\hat{\boldsymbol{\beta}}_N$ of $\boldsymbol{\beta}$ minimize the sum of squared estimated residuals $\hat{\boldsymbol{\varepsilon}}_N'\hat{\boldsymbol{\varepsilon}}_N$, where $\hat{\boldsymbol{\varepsilon}}_N = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_N$, producing the estimates

$$(2)\ \hat{\boldsymbol{\beta}}_N = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

where for the purpose of describing limits below we use the subscript N to emphasize that the estimated coefficients and residuals are functions of N realized observations. If the disturbances ε_i are homoskedastic with common variance $\sigma_i^2 = \sigma^2$, one can use the homoskedastic variance estimate of $\hat{\boldsymbol{\beta}}_N$, $(\mathbf{X}'\mathbf{X})^{-1}\hat{\boldsymbol{\varepsilon}}_N'\hat{\boldsymbol{\varepsilon}}_N/(N-K)$, but we focus on more general inference using White's (1980) heteroskedasticity robust covariance estimate,

$$(3)\ \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}_N\}\{\hat{\boldsymbol{\varepsilon}}_N\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

where here and frequently later we use the notation $\{\mathbf{z}\}$ to denote a diagonal matrix with diagonal entries given by the vector **z**.

White (1980) provided conditions for valid inference in this model when the row vector of random variables associated with each observation i, $(\mathbf{x}_i', \varepsilon_i)$, are independently but not necessarily identically distributed (inid):

<u>Theorem I (following White 1980)</u>: If there exist strictly positive finite constants γ, Δ and η such that the following conditions hold

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

(Ib) (i) For all i $E(|x_{ij}x_{ik}|^{1+\gamma}) < \Delta$ for all j, k = 1...K; (ii) $\mathbf{M}_N = N^{-1}\sum_{i=1}^{N} E(\mathbf{x}_i\mathbf{x}_i')$ is non-singular with determinant($\mathbf{M}_N$) > η for all N sufficiently large,

(Ic) (i) For all i, $E(|\varepsilon_i^2 x_{ij}x_{ik}|^{1+\gamma}) < \Delta$ for all j, k = 1...K; (ii) $\mathbf{V}_N = N^{-1}\sum_{i=1}^{N} E(\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i')$ is non-singular with determinant($\mathbf{V}_N$) > η for all N sufficiently large,

then

$$\text{(i) } \hat{\boldsymbol{\beta}}_N \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \boldsymbol{\beta}, \quad \text{(ii) } \mathbf{V}_N^{-1/2}\mathbf{M}_N\sqrt{N}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}) \overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{n}_K,$$

(iii) $\mathbf{M}_N, \mathbf{V}_N$ and their inverses are uniformly bounded for all N sufficiently large,

$$\text{(iv) } N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N) - \mathbf{M}_N^{-1}\mathbf{V}_N\mathbf{M}_N^{-1} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK} \quad \& \quad \text{(v) } (\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta})'\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)^{-1}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}) \overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \chi_K^2,$$

where $\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ and $\overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ denote convergence almost surely and in distribution across ($\mathbf{X}$,$\boldsymbol{\varepsilon}$), respectively, $\mathbf{A}^{1/2}$ the "square root" of symmetric positive definite matrix $\mathbf{A}$,[4] $\mathbf{n}_K$ the K dimensional standard normal and χ_K^2 the central chi-squared with K degrees of freedom, and $\mathbf{0}_K$ and $\mathbf{0}_{KxK}$ vectors and matrices of zeros of the indicated dimensions.

White (1980) used (Ia) - (Ic) to prove (i), (ii) and parts of (iii) and added the assumption $E(|x_{ij}^2 x_{ik}x_{il}|^{1+\gamma}) < \Delta$ to prove (iv), (v) and other results. However, (Ia) - (Ic) suffice to prove (i) - (5), as shown in the appendix below.[5] White's covariance estimate often motivates inference with heteroskedasticity in an otherwise iid setting, such as when the variance of ε_i is a function of $\mathbf{x}_i$, but $(\mathbf{x}_i', \varepsilon_i)$ are otherwise iid draws from a fixed distribution. However, $\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)$ allows for asymptotically accurate inference in the much more general inid setting given above. Given the inid data $\mathbf{M}_N$, $\mathbf{V}_N$ and $N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N)$ do not necessarily converge to matrices of constants.

In this paper we examine two bootstrap techniques commonly used for OLS inference with heteroskedastic disturbances and prove the asymptotic consistency of their distributions for general inid data. Wu's (1986) external bootstrap, now commonly known as the wild bootstrap,

[4]With $\mathbf{E}$ equal to the matrix of eigenvectors and $\boldsymbol{\Lambda}$ the diagonal matrix of eigenvalues of $\mathbf{A}$, $\mathbf{A}^{1/2} = \mathbf{E}\boldsymbol{\Lambda}^{1/2}\mathbf{E}'$, where $\boldsymbol{\Lambda}^{1/2}$ is the diagonal matrix with entries equal to the square root of those of $\boldsymbol{\Lambda}$.

[5]White (1980) also assumed that $E(|\varepsilon_i^2|^{1+\gamma}) < \Delta$, but this is only used with (Ib) to prove $E(|x_{ij}\varepsilon_i|^{1+\gamma}) < \Delta$, which is actually implied by (Ic). The assumption $E(|\varepsilon_i^2|^{1+\gamma}) < \Delta$ is dropped in White (1984), c.f. exercise 5.12.

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 external random variables δ_i^w, so that $y_i^w = \mathbf{x}_i'\hat{\boldsymbol\beta}_N + \hat{\varepsilon}_i\delta_i^w$, or in matrix form $\mathbf{y}^w = \mathbf{X}\hat{\boldsymbol\beta}_N + \{\hat{\boldsymbol\varepsilon}_N\}\boldsymbol\delta^w$. Selecting $\hat{\boldsymbol\beta}_N^w$ so as to minimize the sum of squared residuals for this new data, $\hat{\boldsymbol\varepsilon}_N'^w\hat{\boldsymbol\varepsilon}_N^w$ = $(\mathbf{y}^w - \mathbf{X}\hat{\boldsymbol\beta}_N^w)'(\mathbf{y}^w - \mathbf{X}\hat{\boldsymbol\beta}_N^w)$, yields coefficient and covariance estimates

$$(4)\ \hat{\boldsymbol\beta}_N^w = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}^w = \hat{\boldsymbol\beta}_N + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\boldsymbol\delta^w$$

$$\text{and } \hat{\mathbf{V}}(\hat{\boldsymbol\beta}_N^w) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N^w\}\{\hat{\boldsymbol\varepsilon}_N^w\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

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

Repeated draws of the N dimensional iid vector $\boldsymbol\delta^w$ are made and the percentiles of the distribution of the deviation of the wild bootstrap coefficients from the mean of its data generating process, $\hat{\boldsymbol\beta}_N^w - \hat{\boldsymbol\beta}_N$, or Wald statistics for the same, $(\hat{\boldsymbol\beta}_N^w - \hat{\boldsymbol\beta}_N)'\hat{\mathbf{V}}(\hat{\boldsymbol\beta}_N^w)^{-1}(\hat{\boldsymbol\beta}_N^w - \hat{\boldsymbol\beta}_N)$, are used to evaluate the statistical significance of corresponding measures for tests of the null hypothesis $\boldsymbol\beta = \boldsymbol\beta_0$ for the original sample, i.e. $\hat{\boldsymbol\beta}_N - \boldsymbol\beta_0$ and $(\hat{\boldsymbol\beta}_N - \boldsymbol\beta_0)'\hat{\mathbf{V}}(\hat{\boldsymbol\beta}_N)^{-1}(\hat{\boldsymbol\beta}_N - \boldsymbol\beta_0)$. All permutations of any given outcome vector $\boldsymbol\delta^w$ are equally likely, a fact that plays a prominent role in the results of this paper.

The pairs bootstrap samples N observational pairs $(y_i,\mathbf{x}_i)$ from the rows of the original data $(\mathbf{y},\mathbf{X})$ with replacement, producing the data set $(\mathbf{y}^p, \mathbf{X}^p) = (\boldsymbol\Delta\mathbf{y}, \boldsymbol\Delta\mathbf{X})$, where $\boldsymbol\Delta$ is an N x N matrix of 0s with a single 1 in each row.[6] Selecting $\hat{\boldsymbol\beta}_N^p$ so as to minimize the sum of squared residuals for this new data, $\hat{\boldsymbol\varepsilon}_N'^p\hat{\boldsymbol\varepsilon}_N^p = (\mathbf{y}^p - \mathbf{X}^p\hat{\boldsymbol\beta}_N^p)'(\mathbf{y}^p - \mathbf{X}^p\hat{\boldsymbol\beta}_N^p)$, yields coefficient and covariance estimates

$$(5)\ \hat{\boldsymbol\beta}_N^p = (\mathbf{X}'\boldsymbol\Delta'\boldsymbol\Delta\mathbf{X})^{-1}\mathbf{X}'\boldsymbol\Delta'\boldsymbol\Delta\mathbf{y} = \hat{\boldsymbol\beta}_N + (\mathbf{X}'\boldsymbol\Delta'\boldsymbol\Delta\mathbf{X})^{-1}\mathbf{X}'\boldsymbol\Delta'\boldsymbol\Delta\hat{\boldsymbol\varepsilon}_N$$

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

where we use the fact that as $\mathbf{y} = \mathbf{X}\hat{\boldsymbol\beta}_N + \hat{\boldsymbol\varepsilon}_N$, so $\boldsymbol\Delta\mathbf{y} = \boldsymbol\Delta\mathbf{X}\hat{\boldsymbol\beta}_N + \boldsymbol\Delta\hat{\boldsymbol\varepsilon}_N$. Again, repeated samples are made and the resulting distribution of coefficients , $\hat{\boldsymbol\beta}_N^p - \hat{\boldsymbol\beta}_N$, and Wald statistics. $(\hat{\boldsymbol\beta}_N^p - \hat{\boldsymbol\beta}_N)'\hat{\mathbf{V}}(\hat{\boldsymbol\beta}_N^p)^{-1}(\hat{\boldsymbol\beta}_N^p - \hat{\boldsymbol\beta}_N)$, used to evaluate the statistical significance of corresponding measures for the original sample.

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

With some matrix algebra the equations for the pairs bootstrap can be transformed into a form that highlights their similarity with the wild bootstrap. If we use the notation $\mathbf{D}_{jk}$ to represent the jk^{th} element of matrix $\mathbf{D}$, $\mathbf{D}_{\bullet j}$ the j^{th} column, and $\mathbf{z}$ some vector, then using the fact that $\mathbf{\Delta}$ is a matrix of zeros with a single 1 in each row and the only potentially non-zero element of any $\{\mathbf{h}\}_{\bullet k}$ is h_k

$$(6)\ (\mathbf{\Delta}'\mathbf{\Delta})_{jk} = \mathbf{\Delta}'_{\bullet j}\mathbf{\Delta}_{\bullet k} = \sum_{i=1}^{N}\mathbf{\Delta}_{ij}\mathbf{\Delta}_{ik} = (\text{if } j = k)\sum_{i=1}^{N}\mathbf{\Delta}_{ij}^{2} = \sum_{i=1}^{N}\mathbf{\Delta}_{ij}$$

$$= (\text{if } j \neq k)\ 0$$

$$(\mathbf{\Delta}'\{\mathbf{\Delta z}\})_{jk} = \mathbf{\Delta}'_{\bullet j}\{\mathbf{\Delta z}\}_{\bullet k} = \mathbf{\Delta}_{kj}\sum_{i=1}^{N}\mathbf{\Delta}_{ki}z_i = \mathbf{\Delta}_{kj}^{2}z_j = \mathbf{\Delta}_{kj}z_j.$$

Consequently, if we define $\delta_j^p = \sum_{i=1}^{N}\mathbf{\Delta}_{ij}$, then $\mathbf{\Delta}'\mathbf{\Delta} = \{\boldsymbol{\delta}^p\}$ is a diagonal matrix with elements equal to the number of times each row is sampled, while $\mathbf{\Delta}'\{\mathbf{\Delta z}\} = \{\mathbf{z}\}\mathbf{\Delta}'$. Using this, we re-express (5) above as:

$$(7)\ \hat{\boldsymbol{\beta}}_N^p = \hat{\boldsymbol{\beta}}_N + (\mathbf{X}'\{\boldsymbol{\delta}^p\}\mathbf{X})^{-1}\mathbf{X}'\{\boldsymbol{\delta}^p\}\hat{\boldsymbol{\varepsilon}}_N = \hat{\boldsymbol{\beta}}_N + (\mathbf{X}'\{\boldsymbol{\delta}^p\}\mathbf{X})^{-1}\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}_N\}\boldsymbol{\delta}^p$$

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

As in the case of the wild bootstrap, conditional on the original data the estimated coefficients and covariance matrix are only a function of the realization of the Nx1 vector $\boldsymbol{\delta}^p$. All permutations of any given sampling frequency vector $\boldsymbol{\delta}^p$ are equally likely, a fact that plays a prominent role in the results of this paper. Consequently, we use the common notation $\boldsymbol{\delta}$, distinguished by superscripted p or w, for seemingly dissimilar objects because these operate identically in the theorems and proofs below.

Our interest is in deriving sufficient conditions for the conditional consistency of the bootstrap distributions in an inid framework. Specifically, we show that White's (1980) assumptions are sufficient to ensure that for the bootstrapped coefficient and heteroskedasticity robust covariance estimates, with b denoting p (pairs) or w (wild)

$$(8)\ \left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}_N\}\{\hat{\boldsymbol{\varepsilon}}_N\}\mathbf{X}}{N}\right)^{-\frac{1}{2}}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\sqrt{N}(\hat{\boldsymbol{\beta}}_N^b - \hat{\boldsymbol{\beta}}_N) \overset{d(\boldsymbol{\delta}^b)|as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{n}_K$$

$$\&\ \sqrt{N}(\hat{\boldsymbol{\beta}}_N^b - \hat{\boldsymbol{\beta}}_N)'[N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N^b)]^{-1}(\hat{\boldsymbol{\beta}}_N^b - \hat{\boldsymbol{\beta}}_N)\sqrt{N} \overset{d(\boldsymbol{\delta}^b)|as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \chi_K^2,$$

where $\overset{d(\boldsymbol{\delta})|as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ denotes convergence in distribution across $\boldsymbol{\delta}$ almost surely across realizations of $(\mathbf{X},\boldsymbol{\varepsilon})$. These results show that the asymptotic conditional distribution given the data $(\mathbf{X},\boldsymbol{\varepsilon})$ of the bootstrap equals the asymptotic distribution of the OLS estimates across $(\mathbf{X},\boldsymbol{\varepsilon})$, allowing for valid inference using the percentiles of bootstrapped coefficient estimates or Wald statistics.[7]

The key characteristic exploited in proofs below is that any of the row permutations of the vectors $\boldsymbol{\delta}$ are 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 permutation 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 ensure that these properties hold almost surely for $(\mathbf{X},\boldsymbol{\varepsilon})$, while the properties of the multinomial sampling frequencies $\boldsymbol{\delta}^p$ and moment assumptions on the iid elements of $\boldsymbol{\delta}^w$ ensure the requisite conditions on $\boldsymbol{\delta}$ also hold almost surely. Consequently, almost surely conditional on the data $(\mathbf{X},\boldsymbol{\varepsilon})$ the distributions of the bootstraps across the draws $\boldsymbol{\delta}$ that determine their coefficient estimates and Wald statistics converge to the distribution of their OLS counterparts for the original sample $(\mathbf{X},\boldsymbol{\varepsilon})$ across its data generating process.

III. Foundational Permutation Theorems

The proofs in this paper rely on a theorem first proven by Wald & Wolfowitz (1944) and later refined by Noether (1949) and Hoeffding (1951) concerning the asymptotic distribution of root-N times the correlation of a permuted sequence with another sequence:

Theorem II: Let $\mathbf{z}' = (z_1, \dots, z_N)$ and $\boldsymbol{\delta}' = (d_1, \dots, d_N)$ denote sequences of real numbers, not all equal, and $\mathbf{d}' = (d_1, \dots, d_N)$ any of the $N!$ equally likely permutations of $\boldsymbol{\delta}$. Then as $N \to \infty$ the distribution across the realizations of $\mathbf{d}$ of the random variable

$$\text{(IIa)}\quad v_N = \sum_{i=1}^{N} \frac{[z_i - m(z_i)][d_i - m(d_i)]}{s(z_i)s(d_i)N^{1/2}},$$

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

[7]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.

converges to that of the standard normal if for all integer $\tau > 2$

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

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 $\mathbf{Z}' = (\mathbf{z}_1,...,\mathbf{z}_N)$ is a sequence of K x 1 vectors and $\mathbf{O} = \mathbf{I}_{NxN} - \mathbf{1}_N\mathbf{1}'_N/N$ the centering matrix,[8] then

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

is asymptotically distributed multivariate iid standard normal if (IIb) 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 $\text{diag}(\mathbf{Z}'\mathbf{O}\mathbf{Z})^{-½}\mathbf{Z}'\mathbf{O}\mathbf{Z}\,\text{diag}(\mathbf{Z}'\mathbf{O}\mathbf{Z})^{-½}$, 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).

Theorem II is easily extended to a probabilistic environment by noting the following result due to Ghosh (1950) that translates the almost sure or in probability characteristics of an infinite number of moment conditions into similar statements regarding a distribution:

> Theorem III: 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,
>
> $$\text{(IIIa)}\ \lim_{k\to\infty}\frac{\alpha^{k+2}\nu_{k+1}}{k+2!} = 0 \ \text{ for any given value of } \alpha,$$
>
> then $F_N(x)$ converges almost surely (in probability) to $F(x)$.

Condition (IIIa) is of course true for the normal distribution. Hoeffding (1952) generalized the result by showing that condition (IIIa) 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 III covers the multivariate case given in (IIc) 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. In light of

[8] Where $\mathbf{1}_N$ denotes an N vector of ones and $\mathbf{I}_{NxN}$ the NxN identity matrix.

Theorem III, in applying Theorem II below we use the notation $\overset{d(\mathbf{d})|as(\boldsymbol{\delta},\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ and $\overset{d(\mathbf{d})|as(\boldsymbol{\delta},\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$, i.e. almost surely across the realizations of ($\boldsymbol{\delta}$,$\mathbf{X}$,$\boldsymbol{\varepsilon}$) the distribution of $\mathbf{v}_N$ across permutations $\mathbf{d}$ of $\boldsymbol{\delta}$ converges to the multivariate standard normal. Theorems II and III are used to characterize the asymptotic distribution of $\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}_N\}\boldsymbol{\delta}/\sqrt{N}$, which appears in the expressions for the bootstrapped coefficient estimates in (4) and (7) above.

A less demanding form of Theorem II, 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 IV: 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 $\boldsymbol{\delta}$. Then as $N \rightarrow \infty$, across permutations $\mathbf{d}$ of $\boldsymbol{\delta}$

$$\text{(IVa)} \quad m(z_i d_i) - m(z_i)m(d_i) = \sum_{i=1}^{N} \frac{z_i d_i}{N} - \sum_{i=1}^{N} \frac{z_i}{N} \sum_{i=1}^{N} \frac{d_i}{N} \overset{p}{\rightarrow} 0,$$

if

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

Theorem IV is used in proofs to make statements regarding the convergence in probability of terms such as $\mathbf{Z}'\{\boldsymbol{\delta}\}\mathbf{Z}/N$, for some matrix $\mathbf{Z}$ made up of $\mathbf{X}$ and $\hat{\boldsymbol{\varepsilon}}_N$, which appear in (4) and (7) above. As satisfaction of (IVb) depends on the realized sample moments of ($\mathbf{X}$,$\boldsymbol{\varepsilon}$) and $\boldsymbol{\delta}$, we use the notation $\overset{p(\mathbf{d})|as(\boldsymbol{\delta},\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$, i.e. almost surely across the realizations of ($\boldsymbol{\delta}$,$\mathbf{X}$,$\boldsymbol{\varepsilon}$) $m(z_i d_i)$ converges in probability across the permutations $\mathbf{d}$ of $\boldsymbol{\delta}$ to $m(z_i)m(d_i)$.

IV. Bootstrap Consistency with INID Data

The following result is proven in the appendix further below:

Theorem V: Assume that for the wild bootstrap $E[\delta_i^w] = 0$, $E[(\delta_i^w)^2] = 1$ and $E[(\delta_i^w)^{2(1+\theta_1)}] < \Delta$ for some finite Δ and $\theta_1 > 1/\gamma$, with γ as given in Theorem I earlier. White's assumptions (Ia) - (Ic) given in Theorem I in combination with the properties of $\boldsymbol{\delta}$ are sufficient to ensure that across the permutations $\mathbf{d}$ of $\boldsymbol{\delta}^b$, for $b = p$ (pairs) or w (wild)

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

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

Bounded higher moments of δ_i^w are needed to ensure that conditions (IIb) and (IVb) in Theorems II and IV are satisfied.

Let $\boldsymbol{\delta}^*$ denote the ordered values of $\boldsymbol{\delta}$. Across permutations $\mathbf{d}$ of $\boldsymbol{\delta}^*$ (Va) and (Vb) hold. These permutations, integrated across the distribution of $\boldsymbol{\delta}^*$, characterize the entire distribution of $\boldsymbol{\delta}$. Adding the result[9]

$$\text{(9)}\ \frac{\boldsymbol{\delta}^{p\prime}\mathbf{O}\boldsymbol{\delta}^{p}}{N} \overset{p(\boldsymbol{\delta}^p)}{\rightarrow} 1 \ \text{ and } \ \frac{\boldsymbol{\delta}^{w\prime}\mathbf{O}\boldsymbol{\delta}^{w}}{N} \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 1,$$

implies that

$$\text{(10a)} \left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}_N\}\{\hat{\boldsymbol{\varepsilon}}_N\}\mathbf{X}}{N}\right)^{-½} \left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\sqrt{N}(\hat{\boldsymbol{\beta}}_N^b - \hat{\boldsymbol{\beta}}_N) \overset{d(\boldsymbol{\delta}^b)|as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{n}_K$$

$$\text{(10b)}\ N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N^b) - N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N) \overset{p(\boldsymbol{\delta}^b)|as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}\ ,$$

where the convergence in distribution in this case is across the bootstrap realizations of $\boldsymbol{\delta}^b$ that determine the bootstrap coefficient and covariance estimates, as in (4) and (7) earlier above. When combined with White's (1980) result in Theorem I regarding the asymptotic distribution of OLS coefficient estimates, this establishes that almost surely the conditional (on the data) distributions of the bootstrapped coefficients and Wald statistics converge to the unconditional distributions of their OLS regression counterparts.

Remark 1: assumptions on regressors and errors

For an OLS model with iid data and potentially heteroskedastic residuals, Mammen (1993) showed that for a fixed number of regressors the wild bootstrap distributions of linear combinations of the coefficients and Wald statistics based upon the homoskedastic covariance estimate are in probability consistent given $\sup_{\|\mathbf{c}\|=1} E[(\mathbf{c}'\mathbf{x}_i)^4(1+\varepsilon_i^2)] < \infty$ and the Lindeberg type

[9]For the wild bootstrap, (9) follows immediately from the assumptions on moments. The proof for the pairs bootstrap is lengthy and is given in the on-line appendix.

condition $E[(\mathbf{c}'\mathbf{x}_i)^2\varepsilon_i^2 I[(\mathbf{c}'\mathbf{x}_i)^2\varepsilon_i^2 \geq \gamma N] \rightarrow 0$ for every fixed $\gamma > 0$. For the same model, Freedman (1981) proved almost sure consistency of pairs bootstrap coefficients and homoskedastic-based Wald tests if the row vectors $(\mathbf{x}_i', y_i)$ are independently and identically distributed and $E[((\mathbf{x}_i', y_i)(\mathbf{x}_i', y_i)')^2] < \infty$. Stute (1990) tightened part of the result showing that almost sure convergence of the pairs bootstrap coefficients alone for iid data only requires $E(x_{ij}x_{ik})$ and $E(x_{ij}x_{ik}\varepsilon_i^2)$ to be finite. By adopting a permutation approach, this paper proves almost sure consistency of both coefficients and Wald statistics based upon the heteroskedasticity robust covariance estimate with inid data for both the pairs and wild bootstrap with the existence of only slightly higher moments than required by Stute (1990), i.e. $E|x_{ij}x_{ik}|^{1+\gamma} < \infty$ and $E|x_{ij}x_{ik}\varepsilon_i^2|^{1+\gamma} < \infty$ for some $\gamma > 0$. 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 and Stute allowed for sub-sampling $M < N$ observations . As shown in the on-line appendix, at the expense of complicating the proofs the permutation based pairs bootstrap consistency results can be extended to sub-sampling, with and without replacement, if $M/N \rightarrow 0$ and for some $\gamma^* > (1+\gamma)^{-1}$, M is such that $\lim_{\inf} M/N^{\gamma^*} > 0$.[10]

For inid data, Liu (1988) proved consistency in probability of the second central moment of the wild OLS bootstrap coefficient distribution with bounded regressors and finite second moments of ε_i. This paper proves almost sure consistency of the wild bootstrap distribution for unbounded regressors given the moment conditions described above. Djogbenou et al (2019) prove consistency in probability of the distribution of the wild bootstrap t-statistic for within cluster correlated but cross-cluster independent but not identically distributed data. In the case where clusters are composed of single observations, their assumptions on the existence of moments are those used in this paper plus the addition of the fourth moment restriction $E|x_{ij}^4|^{1+\gamma} < \infty$ for some $\gamma > 0$. They also impose asymptotic homogeneity of the data generating process in the form of assuming that $\mathbf{X}'\mathbf{X}/N$ converges to a matrix of constants, while for any

[10]The requirement that M not fall too rapidly relative to N is needed to ensure the existence and convergence of higher moments to the normal, as the proof of Theorem II is based upon the method of moments.

vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$ there exists a finite scalar $v_{\boldsymbol{\alpha}} > 0$ and non-random sequence $\mu_N \rightarrow \infty$ such that $\mu_N \boldsymbol{\alpha}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\boldsymbol{\varepsilon}\}\{\boldsymbol{\varepsilon}\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\alpha} \rightarrow v_{\boldsymbol{\alpha}}$. Thus, while papers usually use the iid assumption to motivate convergence of key matrices to matrices of constants, Djogbenou et al (2019) avoid the iid assumption but assume the data nevertheless converge to such matrices. This paper, using White's (1980) assumptions, requires no such convergence of the asymptotic regressor cross product and covariance matrix of coefficient estimates and as such covers more fundamentally inid data using more demanding moment assumptions. However, Djobgenou et al's analysis goes beyond this paper's in allowing for cluster correlated groupings of observations and dealing with issues concerning the asymptotic maximum size of any such grouping.

Remark 2: type of consistency proven

Aside from consistency of the coefficient distribution, Freedman (1981) and Mammen (1993) prove consistency of the Wald statistic based upon the covariance estimate with homoskedastic errors, while recognizing that with heteroskedastic errors its distribution is not pivotal. This paper focuses on the Wald statistic using the heteroskedasticity robust covariance estimate which is also asymptotically accurate with homoskedastic errors. This test statistic is asymptotically pivotal and hence provides higher order asymptotic bootstrap accuracy (Singh 1981, Hall 1992). Djogbenou et al (2019) prove consistency for t-statistics in a broader framework with clustered-robust covariance estimates, which allow for arbitrary correlations and heteroskedasticity within defined groups of observations and, when clusters are defined as single observations, encompass the heteroskedasticity-robust framework of this paper.

Liu (1988) proves consistency of the wild bootstrap second central moment with bounded regressors. Proving such consistency with the unbounded regressors of this paper is trivial. If we assume, as did Liu (1988), that $E[\boldsymbol{\delta}^w] = \mathbf{0}_N$ and $E[\boldsymbol{\delta}^w\boldsymbol{\delta}'^w] = \mathbf{I}_{NxN}$ (the identity matrix), then taking the expectation with respect to this variable for a given realization of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, we have

$$(11)\ E[\hat{\boldsymbol\beta}_N^w \mid \mathbf{X},\boldsymbol\varepsilon] = \hat{\boldsymbol\beta}_N + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}E[\boldsymbol\delta^w] = \hat{\boldsymbol\beta}_N$$

$$E[(\hat{\boldsymbol\beta}_N^w - E[\hat{\boldsymbol\beta}_N^w])(\hat{\boldsymbol\beta}_N^w - E[\hat{\boldsymbol\beta}_N^w])' \mid \mathbf{X},\boldsymbol\varepsilon] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}E[\boldsymbol\delta^w\boldsymbol\delta'^w]\{\hat{\boldsymbol\varepsilon}_N\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

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

Thus, for any sample size the variance of wild bootstrap coefficient estimates equals White's heteroskedasticity robust covariance estimate. Since under White's conditions $N\hat{\mathbf{V}}(\hat{\boldsymbol\beta}_N)$ is a consistent estimator of the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol\beta}_N - \boldsymbol\beta)$, it follows that for such general inid data the wild bootstrap coefficient variance is a consistent estimator as well, reproducing Liu's result in a more general framework.

A similar result for the pairs bootstrap is more problematic. The first two moments of the multinomial sampling frequencies ($\boldsymbol\delta^p$) for N draws with replacement from N observations are $E[\boldsymbol\delta^p] = \mathbf{1}_N$ (a vector of ones) and $E[\boldsymbol\delta^p\boldsymbol\delta'^p] = \mathbf{I}_{NxN} - N^{-1}\mathbf{1}_N\mathbf{1}'_N$. Examining the moments of the latter half of $\hat{\boldsymbol\beta}_N^p - \hat{\boldsymbol\beta}_N = (\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{X})^{-1}\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\boldsymbol\delta^p$, we see:

$$(12)\ E[\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\boldsymbol\delta^p \mid \mathbf{X},\boldsymbol\varepsilon] = \mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}E[\boldsymbol\delta^p] = \mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\mathbf{1}_N = \mathbf{X}'\hat{\boldsymbol\varepsilon}_N = \mathbf{0}_K,$$

$$\&\ E[(\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\boldsymbol\delta^p)(\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\boldsymbol\delta^p)' \mid \mathbf{X},\boldsymbol\varepsilon] = \mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}E[\boldsymbol\delta^p\boldsymbol\delta'^p]\{\hat{\boldsymbol\varepsilon}_N\}\mathbf{X}$$

$$= \mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\{\hat{\boldsymbol\varepsilon}_N\}\mathbf{X} - N^{-1}\mathbf{X}'\hat{\boldsymbol\varepsilon}_N\hat{\boldsymbol\varepsilon}'_N\mathbf{X} = \mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\{\hat{\boldsymbol\varepsilon}_N\}\mathbf{X} = (\mathbf{X}'\mathbf{X})\mathbf{V}(\hat{\boldsymbol\beta}_N)(\mathbf{X}'\mathbf{X}),$$

where we make use of the fact that $\mathbf{X}'\hat{\boldsymbol\varepsilon}_N = \mathbf{0}_K$ as the OLS estimates $\hat{\boldsymbol\beta}_N$ in (2) above minimize $\hat{\boldsymbol\varepsilon}'_N\hat{\boldsymbol\varepsilon}_N$. Were $\mathbf{X}'\{\hat{\boldsymbol\varepsilon}_N\}\boldsymbol\delta^p$ multiplied by $(\mathbf{X}'\mathbf{X})^{-1}$, this would prove consistency of the second central moment of pairs bootstrap coefficients, but unfortunately it is multiplied by $(\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{X})^{-1}$. However, it is easy to show that $(\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{X})^{-1}$ converges in probability to $(\mathbf{X}'\mathbf{X})^{-1}$ (see the appendix below). Using this fact, Tu and Shao (1995) prove consistency of the second central moment using the artifice of assuming that when the minimum eigenvalue of $(\mathbf{X}'\{\boldsymbol\delta^p\}\mathbf{X})^{-1}$ is less than ½ of the minimum eigenvalue of $(\mathbf{X}'\mathbf{X})^{-1}$, an event whose probability converges to zero, $\hat{\boldsymbol\beta}_N^p$ is set equal to $\hat{\boldsymbol\beta}_N$.

This paper, and most papers which prove consistency of bootstrap distributions, implicitly prove convergence in the sense of the Kolmogorov sup-norm since, by Polya's Theorem, if a distribution function F_N converges to F which is continuous, as is the normal distribution, then the convergence is uniform, i.e. $\lim_{N\to\infty}\sup_x | F_N(x) - F(x) | = 0$. The notable exception is

Freedman (1981), who proves convergence of the distribution of the pairs bootstrap in the sense of the Mallows (1972) metric, namely $d_k(F,G)^k = \inf_{\Gamma(\mathbf{x},\mathbf{y})} \|\mathbf{x}-\mathbf{y}\|^k$, where $\| \ \|$ denotes the Euclidean norm and $\Gamma(\mathbf{x},\mathbf{y})$ the collection of all possible joint distributions of the vectors $(\mathbf{x},\mathbf{y})$ whose marginal distributions are F and G, respectively. Freedman proves convergence in the d_1 metric of $\mathbf{X}'\{\boldsymbol{\delta}^p\}\mathbf{X}/N$ to $\mathbf{X}'\mathbf{X}/N$ and in the d_2 metric of $\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}_N\}\boldsymbol{\delta}^p/\sqrt{N}$ to $\mathbf{X}'\boldsymbol{\varepsilon}/\sqrt{N}$. Convergence in the d_k metric is equivalent to convergence in distribution plus convergence of the 1st through kth absolute moments (Bickel and Freedman 1981), so in this respect Freedman's results for the pairs bootstrap with iid data go beyond those presented in this and other papers. They do not, however, constitute a proof of convergence of the second moments of pairs bootstrap coefficients, as these involve the expectation of the product of $(\mathbf{X}'\{\boldsymbol{\delta}^p\}\mathbf{X})^{-1}$ with $\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}_N\}\boldsymbol{\delta}^p$, as was noted above.

It is well known that convergence in distribution does not imply convergence of moments, but the fact that the proof of Theorem II regarding the asymptotic permutation distribution of root-N correlation coefficients is based upon the method of moments (see Hoeffding 1951 and the on-line appendix of this paper) might lead to the erroneous conclusion that the results here imply consistency of all moments. They do not, most fundamentally because the proof of convergence in distribution of $\sqrt{N}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta})$ to the multivariate normal does not necessarily imply the existence of all higher moments for the original coefficients themselves. With regards to the bootstraps, in the appendix below Theorem II is used to prove that across the equally likely permutations $\mathbf{d}$ of a given $\boldsymbol{\delta}^b$, for $b = p$ (pairs) or w (wild)

$$(13)\quad \left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-\frac{1}{2}} \left(\frac{\boldsymbol{\delta}'^b\mathbf{O}\boldsymbol{\delta}^b}{N}\right)^{-\frac{1}{2}} \frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}}{\sqrt{N}} \overset{d(\mathbf{d})|as(\boldsymbol{\delta}^b,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{n}_K,$$

signifying, by the method of proof, that the moments across permutations $\mathbf{d}$ of $\boldsymbol{\delta}$ of the left hand side converge to those of the multivariate standard normal. Since this is true for all $\boldsymbol{\delta}$ such that $\boldsymbol{\delta}'^b\mathbf{O}\boldsymbol{\delta}^b > 0$, which almost surely holds (see (L2) in the appendix), we can equally say that across the distribution of $\boldsymbol{\delta}$ the moments of (13) converge to those of the multivariate standard normal. For the wild bootstrap $\sqrt{N}(\hat{\boldsymbol{\beta}}_N^w - \hat{\boldsymbol{\beta}}_N)$ consists of (13) multiplied by

$(\mathbf{X'X}/N)^{-1}(\boldsymbol{\delta}'^{b}\mathbf{O}\boldsymbol{\delta}^{b}/N)^{½}(\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N)^{½}$, and as $\boldsymbol{\delta}'^{w}\mathbf{O}\boldsymbol{\delta}^{w}/N \overset{as(\boldsymbol{\delta}^{w})}{\rightarrow} 1$ we can say that all the moments of $\sqrt{N}(\hat{\boldsymbol{\beta}}_{N}^{w}-\hat{\boldsymbol{\beta}}_{N})$ converge to those of the mulitvariate normal with covariance matrix $N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{N})$, although these need not be the asymptotic moments of $\sqrt{N}(\hat{\boldsymbol{\beta}}_{N}-\boldsymbol{\beta})$. In the case of the pairs bootstrap, $\sqrt{N}(\hat{\boldsymbol{\beta}}_{N}^{p}-\hat{\boldsymbol{\beta}}_{N})$ equals (13) multiplied by $(\mathbf{X}'\{\boldsymbol{\delta}^{p}\}\mathbf{X}/N)^{-1}(\boldsymbol{\delta}'^{p}\mathbf{O}\boldsymbol{\delta}^{p}/N)^{½}(\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N)^{½}$ and as both $\mathbf{X}'\{\boldsymbol{\delta}^{p}\}\mathbf{X}/N$ and $\boldsymbol{\delta}'^{p}\mathbf{O}\boldsymbol{\delta}^{p}/N$ are only shown to converge in probability, nothing can be said about the asymptotic moments of $\sqrt{N}(\hat{\boldsymbol{\beta}}_{N}^{p}-\hat{\boldsymbol{\beta}}_{N})$ without the use of an artifice such as that of Shao and Tu (1995) mentioned above.

Remark 3: assumptions on the wild bootstrap external variable

Liu (1988) proves consistency of the second central moment of the wild bootstrap coefficients assuming that the first and second moments of the wild bootstrap external variable δ_{i}^{w} are 0 and 1, respectively.[11] This paper extends the proof to consistency in distribution by additionally requiring that $E[(\delta_{i}^{w})^{2(1+\theta_{1})}]<\infty$ for $\theta_1 > 1/\gamma$ where $\gamma > 0$ is such that $E|x_{ij}x_{ik}|^{1+\gamma}<\infty$ and $E|x_{ij}x_{ik}\varepsilon_{i}^{2}|^{1+\gamma}<\infty$. As the proof of Theorem II is based on the method of moments, depending upon the existence of higher moments for the regressors higher moments on δ_{i}^{w} are needed to ensure that all moments of (13) above exist and converge to the normal. Proofs of the consistency of wild bootstrap distributions typically assume that the external variable δ_{i}^{w} comes from a particular distribution, such as the Rademacher, with moments of all order (e.g. Mammen 1993, Canay et al 2021). A notable exception is Djogbenou et al (2019), where the proof of convergence in distribution merely requires that $|\delta_{i}^{w}|^{2+\lambda}<\infty$ for some $\lambda > 0$. As that paper uses the central limit theorem rather than the method of moments, it can avail itself of tighter assumptions on δ_{i}^{w}. The wild bootstrap external variable, however, is under the control of the practitioner (i.e. not a characteristic of the given data) and at this time there appear to be no known advantages to using an external variable without higher moments.

[11] Liu (1988) also advocated selecting $E(\delta_{i}^{w3}) = 1$ so as to correct for skewness in the Edgeworth expansion. However, Monte Carlos find that forms of δ_{i}^{w} that make this assumption perform less well than those that do not (Davidson & Flachaire 2008, MacKinnon 2015)

References

Bickel, P.J., Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9, 1196-1217.

Bickel, P.J., Freedman, D.A. (1983). Bootstrapping regression models with many parameters. In *A festschrift for Erich L. Lehmann in honor of his sixty-fifth birthday* (Bickel, P. J., Doksum, K.A., Hodges, J.L., eds.). Wadsworth, Belmont, CA.

Cai, H., Chen, Y., Fang, H. (2009). Observational learning: evidence from a randomized natural field experiment. *American Economic Review*, 99, 864–882.

Canay, I.A., Santos, A., Shaikh, A.M. (2021). The wild bootstrap with a "small" number of "large" clusters. *Review of Economics and Statistics*, 103, 346-363.

Cavaliere, G., Georgiev, I. (2020). Inference under random limit bootstrap measures. *Econometrica* 88, 2547-2574.

Davidson, R., Flachaire, E. (2008). The wild bootstrap, tamed at last. *Journal of Econometrics*, 146, 162-169.

Djogbenou, A.A., MacKinnon, J.G., Nielsen, M.O. (2019). Asymptotic theory and wild bootstrap inference with clustered errors. *Journal of Econometrics*, 212, 393-412.

Freedman, D.A. (1981). Bootstrapping regression models. *The Annals of Statistics*, 9, 1218-1228.

Galambos, J. (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.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hardle, W., Horowitz, J., Kreiss, J-P. (2003). Bootstrap methods for time series. *International Statistical Review*, 71, 435-459.

Hoeffding, W. (1951). A combinatorial central limit theorem. *The Annals of Mathematical Statistics*, 22, 558-566.

Hoeffding, W. (1952). The large sample power of tests based upon permutations of observations. *The Annals of Mathematical Statistics*, 23, 169-192.

Liu, R.Y. (1988). Bootstrap procedures under some non-i.i.d. models. *The Annals of Statistics*, 16, 1696-1708.

MacKinnon, J.G. (2015). Wild bootstrap cluster confidence intervals. *L'Actualité Économique*, 9, 11-33.

Mallows, C.L. (1972). A note on asymptotic joint normality. *The Annals of Mathematical Statistics*, 43, 508-515.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21, 255-285.

Noether, G.E. (1949). On a theorem by Wald and Wolfowitz. *The Annals of Mathematical Statistics*, 20, 455-458.

Shao, J., Dongsheng, T. (1995). *The Jackknife and Bootstrap*. New York, Springer-Verlag.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, 9, 1187-1195.

Stute, W. (1990). Bootstrap of the linear correlation model. *Statistics: A Journal of Theoretical and Applied Statistics*, 21, 433-436.

Thornton, R.L. (2008). The demand for, and impact of, learning HIV status. *American Economic Review*, 98, 1829–1863

Wald, A., Wolfowitz, J. (1944). Statistical tests based on permutations of the observations. *The Annals of Mathematical Statistics*, 15, 358-372.

White, H. (1980). A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 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*, 14, 1261-1295.

Appendix

A. Proof of Theorem I

The following corollaries to Markov's Law of Large Numbers and the Continuous Mapping Theorem given in 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

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

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 $b_N(\omega) - c_N \overset{as}{\rightarrow} 0$, and for all N sufficiently large, c_N is interior to C, uniformly in N. Then $g(b_N(\omega)) - g(c_N) \overset{as}{\rightarrow} 0$.

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. Prob($x_j < x$)), and u_N be a nondecreasing sequence of real numbers such that for all j Prob($x_j < \sup_N u_N$) = 1. Then

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

Turning to Theorem I, as noted earlier White (1980) showed that (Ia) - (Ic) are sufficient for (i) $\hat{\boldsymbol{\beta}}_N \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \boldsymbol{\beta}$ and (ii) $\mathbf{V}_N^{-\frac{1}{2}}\mathbf{M}_N\sqrt{N}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}) \overset{d(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{n}_K$. For (iii), from (Ib), (Ic) and Jensen's Inequality we have $E(|x_{ij}x_{ik}|) < \Delta^{1/(1+\gamma)}$ and $E(|\varepsilon_i^2 x_{ij}x_{ik}|) < \Delta^{1/(1+\gamma)}$, so $\mathbf{M}_N$ and $\mathbf{V}_N$ are uniformly bounded and, with determinants > $\eta > 0$, invertible for all N sufficiently large. As the sum of the eigenvalues of a matrix equal its trace and the product its determinant, their maximum eigenvalues are less than $K\Delta^{1/(1+\gamma)}$ and their minimum eigenvalues greater than $\eta/(K\Delta^{1/(1+\gamma)})^{K-1}$ for all N sufficiently large. The minimum and maximum eigenvalues of their inverses are the inverses of these. Consequently, for all N sufficiently large the determinants of their inverses are greater than $(K\Delta^{1/(1+\gamma)})^{-K} > 0$ and, by the spectral decomposition of a real symmetric matrix, the absolute value of their elements bounded by $(K\Delta^{1/(1+\gamma)})^{K-1}/\eta$.[12] This establishes (iii) in Theorem I.

$E(|x_{ij}x_{ik}|^{1+\gamma}) < \Delta$ in Theorem (Ib) and the Markov Corollary ensure that $\mathbf{X}'\mathbf{X}/N - \mathbf{M}_N \overset{as(\mathbf{X})}{\rightarrow} \mathbf{0}_{KxK}$ and, by the Continuous Mapping Theorem Corollary, that $\mathbf{X}'\mathbf{X}/N$ is invertible for all N sufficiently large with $(\mathbf{X}'\mathbf{X}/N)^{-1} - \mathbf{M}_N^{-1} \overset{as(\mathbf{X})}{\rightarrow} \mathbf{0}_{KxK}$. The jk^{th} element of $\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N$ equals:

$$\text{(A1)} \sum_{i=1}^{N} x_{ij}x_{ik}\hat{\varepsilon}_i^2/N = \sum_{i=1}^{N} x_{ij}x_{ik}(\varepsilon_i + \sum_{p=1}^{K}(\beta_p - \hat{\beta}_{pN})x_{ip})^2/N$$

$$= \sum_{i=1}^{N}\frac{x_{ij}x_{ik}\varepsilon_i^2}{N} + 2\sum_{p=1}^{K}\frac{(\beta_p - \hat{\beta}_{pN})}{N^{(\theta-1)/2}}\sum_{i=1}^{N}\frac{x_{ij}x_{ik}x_{ip}\varepsilon_i}{N^{1+(1-\theta)/2}} + \sum_{p=1}^{K}\sum_{q=1}^{K}\frac{(\beta_p - \hat{\beta}_{pN})}{N^{(\theta-1)/2}}\frac{(\beta_q - \hat{\beta}_{qN})}{N^{(\theta-1)/2}}\sum_{i=1}^{N}\frac{x_{ij}x_{ik}x_{ip}x_{iq}}{N^{2-\theta}}.$$

[12]Let $\mathbf{E}$, $\boldsymbol{\lambda}$ and $\lambda_{\max}$ denote the eigenvectors, eigenvalues and maximum eigenvalue of symmetric positive definite matrix $\mathbf{A}$, a_{ij} the ij^{th} element and $\boldsymbol{\alpha}_i$ a vector of 0s with a 1 in the i^{th} position. By the Cauchy-Schwarz Inequality and properties of the Rayleigh quotient, $a_{ij}^2 = (\boldsymbol{\alpha}_i'\mathbf{E}\{\boldsymbol{\lambda}\}\mathbf{E}'\boldsymbol{\alpha}_j)^2 \le (\boldsymbol{\alpha}_i'\mathbf{E}\{\boldsymbol{\lambda}\}\mathbf{E}'\boldsymbol{\alpha}_i)(\boldsymbol{\alpha}_j'\mathbf{E}\{\boldsymbol{\lambda}\}\mathbf{E}'\boldsymbol{\alpha}_j) \le \lambda_{\max}^2$.

The Markov Corollary and $E(|\varepsilon_i^2 x_{ij}x_{ik}|^{1+\gamma}) < \Delta$ in Theorem (Ic) ensure that the first summation almost surely converges to $\sum_{i=1}^N E(\varepsilon_i^2 x_{ij}x_{ik})/N$, which is the ij^{th} element of $\mathbf{V}_N$. From the Cauchy-Schwarz Inequality we have

$$\text{(A2)} \left|\sum_{i=1}^N \frac{x_{ij}x_{ik}x_{ip}\varepsilon_i}{N^{1+(1-\theta)/2}}\right| \leq \sqrt{\sum_{i=1}^N \frac{x_{ij}^2 x_{ik}^2}{N^{2-\theta}} \sum_{i=1}^N \frac{x_{ip}^2\varepsilon_i^2}{N}} \leq \sqrt{\max_{i\leq N}\frac{x_{ij}^2}{N^{1-\theta}} \sum_{i=1}^N \frac{x_{ik}^2}{N} \sum_{i=1}^N \frac{x_{ip}^2\varepsilon_i^2}{N}}$$

$$\& \quad \left|\sum_{i=1}^N \frac{x_{ij}x_{ik}x_{ip}x_{iq}}{N^{2-\theta}}\right| \leq \sqrt{\sum_{i=1}^N \frac{x_{ij}^2 x_{ik}^2}{N^{2-\theta}} \sum_{i=1}^N \frac{x_{ip}^2 x_{iq}^2}{N^{2-\theta}}} \leq \sqrt{\max_{i\leq N}\frac{x_{ij}^2}{N^{1-\theta}} \max_{i\leq N}\frac{x_{ip}^2}{N^{1-\theta}} \sum_{i=1}^N \frac{x_{ik}^2}{N} \sum_{i=1}^N \frac{x_{iq}^2}{N}}.$$

Using Markov's Inequality and $E(|x_{ij}^2|^{1+\gamma}) < \Delta$ in Theorem (Ib), we can state that for any $\delta >$ $1/(1+\gamma)$ but < 1

$$\text{(A3)} \sum_{N=1}^{\infty} \text{Prob}(x_{Nj}^2 \geq N^\delta) \leq \sum_{N=1}^{\infty} \frac{E(|x_{Nj}^2|^{1+\gamma})}{N^{\delta(1+\gamma)}} < \sum_{N=1}^{\infty} \frac{\Delta}{N^{\delta(1+\gamma)}} < \infty.$$

So, by the Borel-Cantelli Corollary, $\max_{i\leq N} x_{ij}^2$ is asymptotically almost surely less than N^δ and hence $\max_{i\leq N} x_{ij}^2/N^{1-\theta}$ almost surely converges to zero for $1 > 1-\theta > 1/(1+\gamma)$, i.e. $0 < \theta < \gamma/(1+\gamma)$. Together with the fact that $\sum_{i=1}^N x_{ij}^2/N$ almost surely converges to the bounded j^{th} diagonal term of $\mathbf{M}_N$, this establishes that both left hand side terms in (A2) almost surely converge to 0. Theorem I (i), (ii) and (iii) show that $\sqrt{N}(\beta_p - \hat{\beta}_{pN})$ is asymptotically normally distributed with mean zero and bounded variance less than some $\sigma^2 > 0$. Hence, asymptotically the probability $|\sqrt{N}(\beta_p - \hat{\beta}_{pN})| > N^\delta$ for any δ such that $\theta > \delta > 0$ can be bounded by

$$\text{(A4)} \ \frac{2}{\sigma\sqrt{2\pi}} \int_{N^\delta}^{\infty} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx < \frac{2}{\sigma\sqrt{2\pi}} \int_{N^\delta}^{\infty} \frac{x}{N^\delta} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \frac{2\sigma}{\sqrt{2\pi}} \frac{1}{N^\delta} \exp\left(\frac{-N^{2\delta}}{2\sigma^2}\right),$$

which is less than $N^{-1-\delta}$ for all N sufficiently large. So

$$\text{(A5)} \sum_{N=1}^{\infty} \text{Prob}(|\sqrt{N}(\beta_p - \hat{\beta}_{pN})| \geq N^\delta) < \infty$$

and by the Borel-Cantelli Lemma $\sqrt{N}(\beta_p - \hat{\beta}_{pN})/N^\theta \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0$. Putting all of the above together, we see that $\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N - \mathbf{V}_N \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}$ and $N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N) - \mathbf{M}_N^{-1}\mathbf{V}_N\mathbf{M}_N^{-1} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}$, establishing (iv) in Theorem I, while (v) follows from (i) - (iv).

B. Proof of Theorem IV

If either the z_i or δ_i are all identical ($z_i = z$ or $\delta_i = \delta$), Theorem IV 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 **d** of **δ**:

$$\text{(B1)}\quad E_{\mathbf{d}}(d_i) = \sum_{i=1}^{N} \frac{\delta_i}{N} = m(\delta_i),\ E_{\mathbf{d}}(d_i^2) = \sum_{i=1}^{N} \frac{\delta_i^2}{N} = m(\delta_i^2)$$

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

We then calculate the mean and variance of $m(z_i d_i)$ - $m(z_i)m(d_i)$ across the realizations of **d**:

$$\text{(B2)}\quad E_{\mathbf{d}}(m(z_i d_i) - m(z_i)m(\delta_i)) = \sum_{i=1}^{N} \frac{z_i E_{\mathbf{d}}(d_i)}{N} - m(z_i)m(\delta_i) = 0,$$

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

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

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

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

where subscripted i,j denotes the summation across the two indices excluding ties between them. The last line shows that if (IVb) 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 IV.

C. Proof of Theorem V

We begin by noting the following Lemma, proven in Appendix D further below

Lemma: Let $\overset{as(\boldsymbol{\delta})}{\rightarrow}$ or $\overset{p(\boldsymbol{\delta})}{\rightarrow}$ denote convergence almost surely or in probability across the distribution of **δ**, τ any integer greater than 2, $b = p$ (pairs) or w (wild), $\gamma > 0$ be as given in Theorem I, $\theta_1 > 0$ as in Theorem V, and η_1 some constant > 0 . For all θ such that $\gamma/(1+\gamma)$ > $\theta > 0$ (pairs) or $\gamma/(1+\gamma) > \theta > 1/(1+\theta_1)$ (wild):

$$\text{(L1w) } m(\delta_i^w) \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 0,\ m((\delta_i^w)^2) \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 1 \ \&\ N^{-\theta} m((\delta_i^w)^4) \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 0;$$

$$\text{(L1p) } m(\delta_i^p) = 1,\ m((\delta_i^p)^2) \overset{p(\boldsymbol{\delta}^p)}{\rightarrow} 2,\ \&\ N^{-\theta} m((\delta_i^p)^2) \overset{as(\boldsymbol{\delta}^p)}{\rightarrow} 0;$$

(L2) for some constant $\kappa > 0$ almost surely for all N sufficiently large $\sum_{i=1}^{N} \frac{[\delta_i^b - m(\delta_i^b)]^2}{N} > \kappa$;

$$\text{(L3) } \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)} \sum_{i=1}^{N} [\delta_i^b - m(\delta_i^b)]^\tau}{\left(\sum_{i=1}^{N} [\delta_i^b - m(\delta_i^b)]^2\right)^{\tau/2}} \overset{as(\boldsymbol{\delta}^b)}{\rightarrow} 0;$$

(L4) almost surely for all N sufficiently large $\frac{\mathbf{X}'\mathbf{X}}{N}$, $\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}$,

and their inverses are bounded and positive definite with determinant $> \eta_1 > 0$;

$$\text{(L5) } \forall\ k \ \&\ \tau: \frac{N^{\theta\left(\frac{\tau}{2}-1\right)} \sum_{i=1}^{N} x_{ik}^\tau \hat{\varepsilon}_i^\tau}{\left(\sum_{i=1}^{N} x_{ik}^2 \hat{\varepsilon}_i^2\right)^{\tau/2}} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0;$$

$$\text{(L6) } \forall\ j,k: \sum_{i=1}^{N} \frac{x_{ij}^2 x_{ik}^2}{N^{2-\theta}} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0;\quad \text{(L7) } \forall\ j,k: \sum_{i=1}^{N} \frac{x_{ij}^4 x_{ik}^4}{N^{4-3\theta}} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

For a permutation $\mathbf{d}$ of $\boldsymbol{\delta}^w$ or $\boldsymbol{\delta}^p$, the coefficient estimates of the pairs and wild bootstrap are, following (4) and (7) in the text, given by $\sqrt{N}(\hat{\boldsymbol{\beta}}_N^p - \hat{\boldsymbol{\beta}}_N) = \mathbf{C}^{-1}\mathbf{a}$ and $\sqrt{N}(\hat{\boldsymbol{\beta}}_N^w - \hat{\boldsymbol{\beta}}_N) =$ $(\mathbf{X}'\mathbf{X}/N)^{-1}\mathbf{a}$, where $\mathbf{C} = \mathbf{X}'\{\mathbf{d}\}\mathbf{X}/N$, $\mathbf{a} = \mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}/\sqrt{N}$ and we simplify notation here and later by dropping the subscript N on $\hat{\boldsymbol{\varepsilon}}$. Regarding the jk^{th} element of $\mathbf{C}$, given by $\sum_{i=1}^{N} x_{ij}x_{ik}d_i/N$, we can apply Theorem IV with $z_i = x_{ij}x_{ik}$. Condition IVb in this case requires that:

$$\text{(C1) } \left[\frac{m(x_{ij}^2 x_{ik}^2) - m(x_{ij}x_{ik})^2}{N^{1-\theta}}\right]\left[\frac{m((\delta_i^p)^2) - m(\delta_i^p)^2}{N^\theta}\right] \overset{as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

which is guaranteed by (L1p), (L4) and (L6) above. So,

$$\text{(C2) } \underbrace{\frac{\mathbf{X}'\{\mathbf{d}\}\mathbf{X}}{N}}_{\mathbf{C}} - \frac{\mathbf{X}'\mathbf{X}}{N} \underbrace{m(\delta_i^p)}_{=1} \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}.$$

By the corollary to the Continuous Mapping Theorem given above, $(\mathbf{X}'\{\mathbf{d}\}\mathbf{X}/N)^{-1}$ converges in probability to bounded positive definite $(\mathbf{X}'\mathbf{X}/N)^{-1}$ (as in L4).

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 II in the text with $z_{ik} = x_{ik}\hat{\varepsilon}_i$, or $\mathbf{Z} = \{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}$. Since $\mathbf{1}'_N \{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X} = \hat{\boldsymbol{\varepsilon}}'\mathbf{X} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{X} = \mathbf{y}'\mathbf{X} - \mathbf{y}'\mathbf{X} = \mathbf{0}'_K$, the mean of z_{ik} is zero and so we have $\mathbf{Z}'\mathbf{O}\mathbf{Z} = \mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}$ and $\mathbf{Z}'\mathbf{O}\mathbf{d} = \mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}$. From (L2) we know that almost surely $\mathbf{d}'\mathbf{O}\mathbf{d} = \boldsymbol{\delta}'^{b}\mathbf{O}\boldsymbol{\delta}^{b}$ is non-zero, while (L4) ensures 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 $\delta > 0$.[13] Hence, following Theorems II and III, the distribution across **d** of

$$\text{(C3)}\left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-½}\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}}{\sqrt{N}}$$

converges almost surely (across $\boldsymbol{\delta}^b$,**X**,**ε**) to that of the iid multivariate standard normal as by (L3) and (L5) for all integer τ greater than 2

$$\text{(C4)}\quad \frac{N^{\frac{\tau}{2}-1}\sum_{i=1}^{N} x_{ik}^{\tau}\hat{\varepsilon}_i^{\tau}}{\left(\sum_{i=1}^{N} x_{ik}^{2}\hat{\varepsilon}_i^{2}\right)^{\tau/2}}\frac{\sum_{i=1}^{N}[\delta_i^b - m(\delta_i^b)]^{\tau}}{\left(\sum_{i=1}^{N}[\delta_i^b - m(\delta_i^b)]^{2}\right)^{\tau/2}} = \frac{N^{\theta\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N} x_{ik}^{\tau}\hat{\varepsilon}_i^{\tau}}{\left(\sum_{i=1}^{N} x_{ik}^{2}\hat{\varepsilon}_i^{2}\right)^{\tau/2}}\frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}[\delta_i^b - m(\delta_i^b)]^{\tau}}{\left(\sum_{i=1}^{N}[\delta_i^b - m(\delta_i^b)]^{2}\right)^{\tau/2}} \xrightarrow{as(\boldsymbol{\delta}^b,\mathbf{X},\boldsymbol{\varepsilon})} 0.$$

Using (L4) and the fact that $\boldsymbol{\delta}'^{b}\mathbf{O}\boldsymbol{\delta}^{b}/N = \mathbf{d}'\mathbf{O}\mathbf{d}/N$ is a scalar, we then have:

$$\text{(C5)}\quad \left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\left(\frac{\boldsymbol{\delta}^{p}{}'\mathbf{O}\boldsymbol{\delta}^{p}}{N}\right)^{-½}\sqrt{N}(\hat{\boldsymbol{\beta}}_N^{p} - \hat{\boldsymbol{\beta}}_N) =$$

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

$$\left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)\left(\frac{\boldsymbol{\delta}^{w}{}'\mathbf{O}\boldsymbol{\delta}^{w}}{N}\right)^{-½}\sqrt{N}(\hat{\boldsymbol{\beta}}_N^{w} - \hat{\boldsymbol{\beta}}_N) = \left(\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N}\right)^{-½}\left(\frac{\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-½}\frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{d}}{\sqrt{N}} \xrightarrow{d(\mathbf{d})|as(\boldsymbol{\delta}^{w},\mathbf{X},\boldsymbol{\varepsilon})} \mathbf{n}_K,$$

thereby establishing the claim in (Va).

Regarding the wild bootstrap heteroskedasticity robust covariance estimates, we have

---

[13]By (L4) $\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N$ is bounded with determinant > η_1 > 0. Let u denote the upper bound on the diagonal elements. By the trace property of eigenvalues, we know that the largest eigenvalue is less than Ku, and hence the smallest must be greater than $\eta_1/(Ku)^{K-1}$. The smallest eigenvalue of diag($\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N$)$^{-½}$ is greater than $u^{-½}$. 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 diag($\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N$)$^{-½}$($\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N$) diag($\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N$)$^{-½}$ is greater than $\eta_1/K^{K-1}u^K$, and hence the determinant greater than $(\eta_1/K^{K-1}u^K)^K$.

$$\text{(C6) } N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N^w) = \left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1} \mathbf{A} \left(\frac{\mathbf{X}'\mathbf{X}}{N}\right)^{-1}, \quad \text{where } \mathbf{A} = \frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}_N^w\}\{\hat{\boldsymbol{\varepsilon}}_N^w\}\mathbf{X}}{N}.$$

Using the formula for $\hat{\boldsymbol{\varepsilon}}_N^w$ in (4), the jk^{th} element of $\mathbf{A}$ is given by:

$$\text{(C7) } \sum_{i=1}^{N} \frac{x_{ij}x_{ik}\left(d_i\hat{\varepsilon}_i - \sum_{p=1}^{K} \frac{x_{ip}}{N^{½(1\text{-}\theta)}} \sqrt{\frac{\boldsymbol{\delta}^{w}{}'\mathbf{O}\boldsymbol{\delta}^w}{N^{1+\theta}}}\hat{\eta}_p\right)^2}{N}, \quad \left[\text{where } \hat{\boldsymbol{\eta}} = \left(\frac{\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-½} \sqrt{N}(\hat{\boldsymbol{\beta}}_N^w - \hat{\boldsymbol{\beta}}_N)\right]$$

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

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

$$\text{(C8) } \left[\frac{m(x_{ij}^2x_{ik}^2\hat{\varepsilon}_i^4) - m(x_{ij}x_{ik}\hat{\varepsilon}_i^2)^2}{N^{1-\theta}}\right]\left[\frac{m(\delta_i^{w4}) - m(\delta_i^{w2})^2}{N^{\theta}}\right] \overset{as(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

From (L1w) and (L4), we know that $[m(\delta_i^{w4}) - m(\delta_i^{w2})^2]/N^{\theta}$ and $m(x_{ij}x_{ik}\hat{\varepsilon}_i^2)^2/N^{1-\theta} \overset{as(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0$. Applying the Cauchy-Schwarz Inequality (here, and frequently below)

$$\text{(C9) } \frac{m(x_{ij}^2x_{ik}^2\hat{\varepsilon}_i^4)}{N^{1-\theta}} = \sum_{i=1}^{N}\frac{x_{ij}^2x_{ik}^2\hat{\varepsilon}_i^4}{N^{2-\theta}} \leq \sqrt{\left(\sum_{i=1}^{N}\frac{x_{ij}^4\hat{\varepsilon}_i^4}{N^{2-\theta}}\right)\left(\sum_{i=1}^{N}\frac{x_{ik}^4\hat{\varepsilon}_i^4}{N^{2-\theta}}\right)} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

where the last is guaranteed by (L4) and (L5) as

$$\text{(C10) } \sum_{i=1}^{N}\frac{x_{ij}^4\hat{\varepsilon}_i^4}{N^{2-\theta}} = \frac{\overbrace{N^{\theta}\sum_{i=1}^{N}(x_{ij}\hat{\varepsilon}_i)^4}^{\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0 \text{ (L5 with } \tau=4)}}{\left(\sum_{i=1}^{N}x_{ij}^2\hat{\varepsilon}_i^2\right)^{4/2}}\overbrace{\left(\sum_{i=1}^{N}\frac{x_{ij}^2\hat{\varepsilon}_i^2}{N}\right)^2}^{as \text{ bounded (L4)}} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

So, by Theorem IV

$$\text{(C11) "}a\text{": } m(x_{ij}x_{ik}\hat{\varepsilon}_i^2d_i^2) - m(x_{ij}x_{ik}\hat{\varepsilon}_i^2)m(\delta_i^{w2}) \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0, \text{ where } m(\delta_i^{w2}) \underbrace{\overset{as(\boldsymbol{\delta}^w)}{\rightarrow}}_{\text{by (L1w)}} 1.$$

For "b", we apply Theorem IV with $z_i = x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i / N^{½(1\text{-}\theta)}$, so condition (IVb) requires that

$$\text{(C12) } \frac{m(x_{ij}^2x_{ik}^2x_{ip}^2\hat{\varepsilon}_i^2/N^{1\text{-}\theta}) - m(x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i/N^{½(1\text{-}\theta)})^2}{N^{1-\theta}}\left[\frac{m(\delta_i^{w2}) - m(\delta_i^{w})^2}{N^{\theta}}\right] \overset{as(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

Using (L1w) and

$$\text{(C13)} \left| m\left(\frac{x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i}{N^{½(1-\theta)}}\right)\right| \leq \sum_{i=1}^{N}\frac{\left|x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i\right|}{N^{1+½(1-\theta)}} \leq \underbrace{\sqrt{\sum_{i=1}^{N}\frac{x_{ij}^2x_{ik}^2}{N^{2-\theta}}}}_{as(\mathbf{X},\boldsymbol{\varepsilon})\to 0\,(L6)}\underbrace{\sqrt{\sum_{i=1}^{N}\frac{x_{ip}^2\hat{\varepsilon}_i^2}{N}}}_{as\text{ bounded (L4)}} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0,$$

$$\text{(C14)} \frac{m(x_{ij}^2x_{ik}^2x_{ip}^2\hat{\varepsilon}_i^2/N^{1-\theta})}{N^{1-\theta}} = \sum_{i=1}^{N}\frac{x_{ij}^2x_{ik}^2x_{ip}^2\hat{\varepsilon}_i^2}{N^{3-2\theta}} \leq \underbrace{\sqrt{\sum_{i=1}^{N}\frac{x_{ij}^4x_{ik}^4}{N^{4-3\theta}}}}_{as(\mathbf{X},\boldsymbol{\varepsilon})\to 0\,(L7)}\underbrace{\sqrt{\sum_{i=1}^{N}\frac{x_{ip}^4\hat{\varepsilon}_i^4}{N^{2-\theta}}}}_{as(\mathbf{X},\boldsymbol{\varepsilon})\to 0\,(C10)} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0,$$

we see that condition (IVb) is met and by Theorem IV we then have

$$\text{(C15) "}b\text{": } m\left(\frac{x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i d_i}{N^{½(1-\theta)}}\right) - \underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i}{N^{½(1-\theta)}}\right)}_{as(\mathbf{X},\boldsymbol{\varepsilon})\to 0\ (C13)}\ \underbrace{m(\delta_i^w)}_{as(\boldsymbol{\delta}^w)\to 0\,(L1w)} \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0.$$

For "c", we note that

$$\text{(C16)} \left| m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}}{N^{1-\theta}}\right)\right| \leq \sqrt{\sum_{i=1}^{N}\frac{x_{ij}^2x_{ik}^2}{N^{2-\theta}}\sum_{i=1}^{N}\frac{x_{ip}^2x_{iq}^2}{N^{2-\theta}}}\ \underbrace{\overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\to} 0}_{\text{by (L6)}}.$$

From the above, we see that the $\hat{\eta}_p$ in (C7) are multiplied by $\sqrt{\boldsymbol{\delta}^{w\prime}\mathbf{O}\boldsymbol{\delta}^w/N^{1+\theta}}$ which from (L1w) converges almost surely (across $\boldsymbol{\delta}^w$) to 0, "c" terms which almost surely (across $\mathbf{X},\boldsymbol{\varepsilon}$) converge to 0, and "$b$" terms which also almost surely (across $\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon}$) converge in probability across permutations **d** to zero. As the $\hat{\eta}_p$, from (L4) and (C5) almost surely (across $\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon}$) converge in distribution across permutations **d** of $\boldsymbol{\delta}^w$ to normal variables with bounded variance, it follows that when so multiplied they converge in probability across permutations **d** to zero. This leaves only the "a" term, and consequently using (C11) we see that

$$\text{(C17)}\quad \mathbf{A} - \frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N} \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\to} \mathbf{0}_{KxK} \quad \text{and hence} \quad N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N^w) - N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N) \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^w,\mathbf{X},\boldsymbol{\varepsilon})}{\to} \mathbf{0}_{KxK},$$

which establishes (Vb) for the wild bootstrap.

For the pairs bootstrap heteroskedasticity robust covariance estimates, from (7) we have

$$\text{(C18)}\ N\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_N^p) = \left(\frac{\mathbf{X}'\{\mathbf{d}\}\mathbf{X}}{N}\right)^{-1}\mathbf{B}\left(\frac{\mathbf{X}'\{\mathbf{d}\}\mathbf{X}}{N}\right)^{-1},$$

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

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

$$\text{(C19)} \sum_{i=1}^{N} \frac{x_{ij}x_{ik}d_i\left(\hat{\varepsilon}_i - \sum_{p=1}^{K}\frac{x_{ip}}{N^{½(1-\theta)}}\sqrt{\frac{\boldsymbol{\delta}^{p}{}'\mathbf{O}\boldsymbol{\delta}^{p}}{N^{1+\theta}}}\hat{\eta}_p\right)^2}{N}, \left[\text{where } \hat{\boldsymbol{\eta}} = \left(\frac{\mathbf{d}'\mathbf{O}\mathbf{d}}{N}\right)^{-½}\sqrt{N}(\hat{\boldsymbol{\beta}}_N^p - \hat{\boldsymbol{\beta}}_N)\right]$$

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

For "d", we apply Theorem IV with $z_i = x_{ij}x_{ik}\hat{\varepsilon}_i^2$ and, as by (L1p), (L4) and (C9) $[m(\delta_i^{p2}) - m(\delta_i^p)^2]/N^{\theta}$, $m(x_{ij}x_{ik}\hat{\varepsilon}_i^2)^2/N^{1-\theta}$ and $m(x_{ij}^2x_{ik}^2\hat{\varepsilon}_i^4)/N^{1-\theta}$ all $\overset{as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ 0, condition (IVb) is met so

$$\text{(C20) "}d\text{": } m(x_{ij}x_{ik}\hat{\varepsilon}_i^2 d_i) - m(x_{ij}x_{ik}\hat{\varepsilon}_i^2)\underbrace{m(\delta_i^p)}_{=1\,(\text{L1p})} \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

For "e", we apply Theorem IV with $z_i = x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i/N^{½(1-\theta)}$ and, as by (L1p), (C13) and (C14) $[m(\delta_i^{p2}) - m(\delta_i^p)^2]/N^{\theta}$, $m(x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i/N^{½(1-\theta)})^2/N^{1-\theta}$ and $m(x_{ij}^2x_{ik}^2x_{ip}^2\hat{\varepsilon}_i^2/N^{1-\theta})/N^{1-\theta}$ all $\overset{as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ 0, condition (IVb) is met so

$$\text{(C21) "}e\text{": } m\left(\frac{x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i d_i}{N^{½(1-\theta)}}\right) - \underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}\hat{\varepsilon}_i}{N^{½(1-\theta)}}\right)}_{\text{by (C13)} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0}\underbrace{m(\delta_i^p)}_{=1\,(\text{L1p})} \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

For "f" we apply Theorem IV with $z_i = x_{ij}x_{ik}x_{ip}x_{iq}/N^{1-\theta}$ and see condition (IVb) holds as by (L1p) and (C16) $[m(\delta_i^{p2}) - m(\delta_i^p)^2]/N^{\theta}$ and $m(x_{ij}x_{ik}x_{ip}x_{iq}/N^{1-\theta})^2/N^{1-\theta}$ $\overset{as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow}$ 0, while by (L7)

$$\text{(C22)} \quad \frac{m(x_{ij}^2x_{ik}^2x_{ip}^2x_{iq}^2/N^{2(1-\theta)})}{N^{1-\theta}} \le \sqrt{\sum_{i=1}^{N}\frac{x_{ij}^4x_{ik}^4}{N^{4-3\theta}}\sum_{i=1}^{N}\frac{x_{ip}^4x_{iq}^4}{N^{4-3\theta}}} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0,$$

so

$$\text{(C23) "}f\text{": } m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}d_i}{N^{1-\theta}}\right) - \underbrace{m\left(\frac{x_{ij}x_{ik}x_{ip}x_{iq}}{N^{1-\theta}}\right)}_{\text{by (C16)} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0}\underbrace{m(\delta_i^p)}_{=1\,(\text{L1p})} \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0.$$

Similar to the case of the wild bootstrap, the $\hat{\eta}_p$ in (C19), which from (L4) and (C5) almost surely (across $\boldsymbol{\delta}^p$,**X**,**ε**) converge in distribution across permutations **d** of $\boldsymbol{\delta}^p$ to normal variables with bounded variance, are multiplied by $\sqrt{\boldsymbol{\delta}^{p}{}'\mathbf{O}\boldsymbol{\delta}^{p}/N^{1+\theta}}$ which from (L1p) converges almost surely (across $\boldsymbol{\delta}^p$) to 0 and "e" and "f" terms which almost surely (across $\boldsymbol{\delta}^p$,**X**,**ε**) converge in probability across permutations **d** to zero, and hence when so multiplied converge in probability across permutations **d** to zero. This leaves only the "d" term and so, using (C2) earlier

$$\text{(C24)}\quad \mathbf{B} - \frac{\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}}{N} \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK},\ \frac{\mathbf{X}'\{\mathbf{d}\}\mathbf{X}}{N} - \frac{\mathbf{X}'\mathbf{X}}{N} \overset{p(\mathbf{d})|as(\boldsymbol{\delta}^p,\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}$$

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

which establishes (Vb) for the pairs bootstrap.

D. Proof of Lemma in Appendix B

(L1), (L2), (L3): We prove these for the wild bootstrap, placing the more involved proofs for the pairs in the on-line appendix. From the assumptions $E[\delta_i^w] = 0$ & $E[(\delta_i^w)^2] = 1$ (Theorem V) and the Strong Law of Large Numbers we know that $m(\delta_i^w) \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 0$ and $m((\delta_i^w)^2) \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 1$. Markov's Inequality, $E[(\delta_i^w)^{2(1+\theta_1)}] < \Delta$ (Theorem V) and $\theta > 1/(1+\theta_1)$ (Lemma) imply there exists a v in $(1/(1+\theta_1),\theta)$ such that

$$\text{(D1)}\quad \sum_{N=1}^{\infty} \text{Prob}(\delta_N^2 \geq N^v) \leq \sum_{N=1}^{\infty} \frac{E(|\delta_N^2|^{1+\theta_1})}{N^{v(1+\theta_1)}} < \sum_{N=1}^{\infty} \frac{\Delta}{N^{v(1+\theta_1)}} < \infty,$$

and thus by the Borel-Cantelli Corollary given above $\max_{i \leq N} \delta_i^{w2}/N^\theta \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 0$ and so

$$\text{(D2)}\quad \frac{m(\delta_i^{w4})}{N^\theta} = \sum_{i=1}^{N} \frac{\delta_i^{w4}}{N^{1+\theta}} \leq \max_{i \leq N} \frac{\delta_i^{w2}}{N^\theta} m(\delta_i^{w2}) \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 0.$$

This establishes (L1w). As $\boldsymbol{\delta}'^w\mathbf{O}\boldsymbol{\delta}^w/N \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 1$, for all N sufficiently large $\boldsymbol{\delta}'^w\mathbf{O}\boldsymbol{\delta}^w/N$ is almost surely greater than some κ such that $1 > \kappa > 0$, as stated in (L2). Regarding (L3), we have

$$\text{(D3)}\quad \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\left|\sum_{i=1}^{N}[\delta_i^w - m(\delta_i^w)]^\tau\right|}{\left(\sum_{i=1}^{N}[\delta_i^w - m(\delta_i^w)]^2\right)^{\tau/2}} \leq \frac{N^{(1-\theta)\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^{N}\left|\delta_i^w - m(\delta_i^w)\right|^\tau}{\left(\sum_{i=1}^{N}[\delta_i^w - m(\delta_i^w)]^2\right)^{\tau/2}}$$

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

From the above, we know the denominator of the last almost surely converges to 1, while as for the numerator, using (D2)

$$\text{(D4)} \max_{i\leq N}\frac{[\delta_i^w - m(\delta_i^w)]^2}{N^\theta} \leq \max_{i\leq N}\frac{\delta_i^{w2}}{N^\theta} + 2\left|\frac{m(\delta_i^w)}{N^{½\theta}}\right|\max_{i\leq N}\left(\frac{\delta_i^{w2}}{N^\theta}\right)^{½} + \frac{m(\delta_i^w)^2}{N^\theta} \overset{as(\boldsymbol{\delta}^w)}{\rightarrow} 0.$$

Consequently, (D3) almost surely converges to 0 for $\theta > 1/(1+\theta_1)$, proving (L3).

(L4): In the proof of Theorem I we saw that $\mathbf{X}'\mathbf{X}/N - \mathbf{M}_N \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}$ and $\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N - \mathbf{V}_N \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}$, where the determinants of $\mathbf{M}_N$ and $\mathbf{V}_N$ are > η > 0 for all N sufficiently large and the absolute values of their elements are uniformly bounded by $\varDelta^{1/(1+\gamma)}$. By the Continuous Mapping Theorem Corollary given above $(\mathbf{X}'\mathbf{X}/N)^{-1} - \mathbf{M}_N^{-1} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}$ and $(\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N)^{-1} - \mathbf{V}_N^{-1} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} \mathbf{0}_{KxK}$, where for all N sufficiently large the determinants of $\mathbf{M}_N^{-1}$ and $\mathbf{V}_N^{-1}$ are greater than $(K\varDelta^{1/(1+\gamma)})^{-K} > 0$ and the absolute values of their elements bounded by $(K\varDelta^{1/(1+\gamma)})^{K-1}/\eta$. It follows that almost surely for all N sufficiently large $\mathbf{X}'\mathbf{X}/N$, $\mathbf{X}'\{\hat{\boldsymbol{\varepsilon}}\}\{\hat{\boldsymbol{\varepsilon}}\}\mathbf{X}/N$ and their inverses have the same properties.

(L5), (L6) & L(7): Following the same logic used in (D3), we note that:

$$\text{(D5a)} \quad \frac{N^{\theta\left(\frac{\tau}{2}-1\right)}\left|\sum_{i=1}^N x_{ik}^\tau \hat{\varepsilon}_i^\tau\right|}{\left(\sum_{i=1}^N x_{ik}^2\hat{\varepsilon}_i^2\right)^{\tau/2}} \leq \frac{N^{\theta\left(\frac{\tau}{2}-1\right)}\sum_{i=1}^N \left|x_{ik}^\tau \hat{\varepsilon}_i^\tau\right|}{\left(\sum_{i=1}^N x_{ik}^2\hat{\varepsilon}_i^2\right)^{\tau/2}} \leq \left(\frac{\max_{i\leq N} x_{ik}^2\hat{\varepsilon}_i^2/N^{1-\theta}}{\sum_{i=1}^N x_{ik}^2\hat{\varepsilon}_i^2/N}\right)^{\frac{\tau}{2}-1}$$

$$\text{(D5b)} \quad \sum_{i=1}^N \frac{x_{ij}^2 x_{ik}^2}{N^{2-\theta}} \leq \frac{\max_{i\leq N} x_{ik}^2}{N^{1-\theta}}\sum_{i=1}^N\frac{x_{ij}^2}{N} \quad \& \quad \sum_{i=1}^N \frac{x_{ij}^4 x_{ik}^4}{N^{4-3\theta}} \leq \left(\frac{\max_{i\leq N} x_{ik}^2}{N^{1-\theta}}\right)^2 \frac{\max_{i\leq N} x_{ij}^2}{N^{1-\theta}}\sum_{i=1}^N\frac{x_{ij}^2}{N}.$$

So, to prove (L5) - (L7) it is sufficient to show that the right hand sides of the inequalities above converge to zero. In Appendix A we already showed that almost surely $\sum_{i=1}^N x_{ik}^2/N$ is bounded and $\max_{i\leq N} x_{ik}^2/N^{1-\theta}$ converges to 0, which establishes this for (D5b).

Turning to (D5a), as shown in Appendix A $\sum_{i=1}^N x_{ik}^2\hat{\varepsilon}_i^2/N$ almost surely converges to the diagonal element of $\mathbf{V}_N$ in Theorem (Ic), whose smallest eigenvalue is greater than $\eta/(K\varDelta^{1/(1+\gamma)})^{K-1}$ for all N sufficiently large. From the Schur-Horn Theorem, we know that the smallest diagonal element of $\mathbf{V}_N$ 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 (D5a) is almost surely greater than $\eta / (K\Delta^{1/(1+\gamma)})^{K-1} > 0$ for all N sufficiently large. Regarding the max term in the numerator, we have

$$\text{(D6)}\quad x_{ik}^2 \hat{\varepsilon}_i^2 / N^{1-\theta} = x_{ik}^2 (\varepsilon_i + \sum_{j=1}^{K} (\beta_j - \hat{\beta}_{jN}) x_{ij})^2 / N^{1-\theta}$$

$$\leq \frac{x_{ik}^2 \varepsilon_i^2}{N^{1-\theta}} + 2\sum_{j=1}^{K} \left| \frac{\beta_j - \hat{\beta}_{jN}}{N^{½(\theta-1)}} \right| \sqrt{\frac{x_{ij}^2 \varepsilon_i^2}{N^{1-\theta}} \frac{x_{ik}^4}{N^{2-2\theta}}} + \sum_{j=1}^{K} \sum_{l=1}^{K} \left| \frac{\beta_j - \hat{\beta}_{jN}}{N^{½(\theta-1)}} \right| \left| \frac{\beta_l - \hat{\beta}_{lN}}{N^{½(\theta-1)}} \right| \sqrt[4]{\frac{x_{ij}^4}{N^{2-2\theta}} \frac{x_{il}^4}{N^{2-2\theta}}} \sqrt{\frac{x_{ik}^4}{N^{2-2\theta}}}.$$

So, as it was shown in Appendix A that $(\beta_j - \hat{\beta}_{jN}) / N^{½(\theta-1)} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0$ & $\max_{i \leq N} x_{ik}^2 / N^{1-\theta} \overset{as(\mathbf{X},\boldsymbol{\varepsilon})}{\rightarrow} 0$, to prove that $\max_{i \leq N} x_{ik}^2 \hat{\varepsilon}_i^2 / N^{1-\theta}$ converges almost surely to zero it is sufficient to show that $\max_{i \leq N} x_{ik}^2 \varepsilon_i^2 / N^{1-\theta}$ converges almost surely to zero. However, $E(|\varepsilon_i^2 x_{ij} x_{ik}|^{1+\gamma}) < \Delta$ in Theorem (Ic), by the same argument used in (A3) above, ensures that this is the case for $0 < \theta < \gamma/(1+\gamma)$. In sum, White's assumptions ensure that (D5a) - (D5b) converge to 0 for all θ in $(0,\gamma/(1+\gamma))$, proving (L5) - (L7). As $\theta_1 > 1/\gamma$ in Theorem V, the condition $\theta > 1/(1+\theta_1)$ for the wild bootstrap in the Lemma and the proof of (L3) above can also be met without contradiction.