Wild bootstrap test of overidentification with many instruments and heteroskedasticity

Wang, Wenjie

26 October 2022
Wild Bootstrap Test of Overidentification with Many Instruments and Heteroskedasticity

Wenjie Wang *

October 26, 2022

Abstract

This note studies the validity of bootstrapping the test of overidentifying restrictions under many/many weak instruments and heteroskedasticity. We propose a wild bootstrap procedure and establish this bootstrap consistently estimates the null limiting distributions of a jackknife overidentification test statistic under this asymptotic framework, no matter studentized or not. Monte Carlo simulations show that the wild bootstrap provides more reliable inference than asymptotic critical values. In particular, the studentized wild bootstrap test has the best finite sample performance in terms of both size and power.

JEL classification: C12, C15, C26.

Keywords: Wild Bootstrap, Overidentification Test, Many Instruments, Weak Instruments, Heteroskedasticity.

*Division of Economics, School of Social Sciences, Nanyang Technological University, HSS-04-65, 14 Nanyang Drive, 637332, Singapore. E-mail address: wang.wj@ntu.edu.sg. Wang acknowledges the financial support from NTU SUG Grant No.M4082262.SS0 and Singapore Ministry of Education Tier 1 grant RG53/20.
1 Introduction

Empirical applications of instrumental variables (IV) regressions often involve tests of overidentification. For the case with many/many weak instruments, Anatolyev and Gospodinov (2011) propose modifications of the J test of overidentifying restrictions so that the test can be robust to many instruments under conditional homoskedasticity. Chao, Hausman, Newey, Swanson, and Woutersen (2014) give a jackknife version of the overidentification test, which is asymptotically valid under heteroskedasticity.\textsuperscript{1} In addition, the literature on bootstrap methods for the IV model includes Davidson and MacKinnon (2008, 2010, 2015), Moreira, Porter, and Suarez (2009), Wang and Kaffo (2016), Kaffo and Wang (2017), Finlay and Magnusson (2019), among others. They find that for IV regressions, carefully designed bootstrap procedures typically provide finite sample improvement over asymptotic approximations. In this note, following Davidson and MacKinnon (2008, 2010, 2015), we propose a wild bootstrap procedure as an alternative method for implementing the overidentification test under many/many weak instruments and heteroskedasticity.

2 Setup

Following Chao et al. (2014), we consider a standard linear IV model given by

\begin{equation}
y = X \delta + \epsilon, \quad X = \Gamma + U,
\end{equation}

where \( y \) and \( X \) are, respectively, an \( n \times 1 \) vector of observations on the outcome variable and an \( n \times G \) matrix of observations on the endogenous regressors. \( \Gamma \) is the \( n \times G \) reduced form matrix, and \( \epsilon \) and \( U \) are, respectively, an \( n \times 1 \) vector and an \( n \times G \) matrix of disturbances. The estimation of \( \delta \) will be based on an \( n \times K \) matrix, \( Z \), of instrumental variable observations with \( rank(Z) = K \), and we treat \( Z \) as deterministic. Denote \( P = Z(Z'Z)^{-1}Z' \) and \( M = I_n - P \), where \( I_n \) is an identity matrix with dimension \( n \). We consider the case where \( G \), the dimension of \( \delta \), is small relative to \( n \), but we let \( K \to \infty \) as \( n \to \infty \) to model the effect of having many/many weak instruments. Also assume the other exogenous regressors have been partialled out from the model.

To define the jackknife overidentification statistic of Chao et al. (2014), let \( \hat{\epsilon}_i = y_i - X'_i \hat{\delta} \), where \( \hat{\delta} \) is certain IV estimator of \( \delta \), \( \hat{\epsilon} = (\hat{\epsilon}_1, ..., \hat{\epsilon}_n)' \), and \( \hat{\epsilon}(2) = (\hat{\epsilon}^2_1, ..., \hat{\epsilon}^2_n)' \). Let \( P_{ij} \) denote the \( ij \)-th element of \( P \), and let \( P(2) \) denote the \( n \)-dimensional square matrix with \( ij \)-th component equal to \( P^2_{ij} \). The test statistic takes the form

\begin{align*}
\hat{T} &= \frac{\hat{\epsilon}'P\hat{\epsilon} - \sum_{i=1}^{n} P_{ii} \hat{\epsilon}_i^2}{\sqrt{\hat{V}}} + K = \frac{\sum_{i \neq j} \hat{\epsilon}_i P_{ij} \hat{\epsilon}_j}{\sqrt{\hat{V}}} + K, \\
\hat{V} &= \frac{\hat{\epsilon}(2)'P(2)\hat{\epsilon}(2) - \sum_{i=1}^{n} P^2_{ii} \hat{\epsilon}_i^4}{K} = \frac{\sum_{i \neq j} \hat{\epsilon}_i^2 P^2_{ij} \hat{\epsilon}_j^2}{K},
\end{align*}

(2)

where \( \sum_{i \neq j} \) denotes the double sum over all \( i \) not equal to \( j \).

\textsuperscript{1}For a comprehensive review of the related literature, see Anatolyev (2019) and the references therein.
For the choice of $\hat{\delta}$, Chao et al. (2014) consider the one proposed by Hausman, Newey, Woutersen, Chao, and Swanson (2012), referred to as HFUL, and show that the jackknife overidentification test with critical region $\hat{T} \geq q_{k-G}(1-\alpha)$, where $q_{r}(\tau)$ denotes the $\tau$-th quantile of the chi-squared distribution with $r$ degrees of freedom, has asymptotic rejection probability equal to $\alpha$ under many/many weak instruments and heteroskedasticity. For the wild bootstrap test, we also consider an unstudentized version of $\hat{T}$, namely,

$$\hat{T}_u = \sum_{i \neq j} \hat{\epsilon}_i P_{ij} \hat{\epsilon}_j.$$  

(3)

### 3 Wild bootstrap overidentificaiton tests

Our wild bootstrap procedure is as follows:

**Step 1:** The bootstrap error terms $\epsilon_i^*$ and $v_i^*$ are obtained by

$$\epsilon_i^* = \hat{\epsilon}_i w_i^*, \quad \text{and} \quad v_i^* = \hat{v}_i \omega_i^*, \quad i = 1, \ldots, n,$$

(4)

where $w_i^*$ is a random variable with mean zero, variance one, and independent from the data, while $\hat{v}_i$ is the residual from regressing $X_i$ on $(Z_i, \hat{\epsilon}_i)$, following the efficient bootstrap procedure proposed by Davidson and MacKinnon (2008, 2010, 2015).

**Step 2:** The bootstrap analogues of $X_i$ and $y_i$ are obtained by

$$X_i^* = Z_i' \hat{\Pi} + v_i^*, \quad y_i^* = X_i' \hat{\delta}^* + \epsilon_i^*, \quad i = 1, \ldots, n,$$

(5)

where $\hat{\Pi}$ is the obtained coefficient for $Z_i$ when regressing $X_i$ on $(Z_i, \hat{\epsilon}_i)$.

**Step 3:** For $i = 1, \ldots, n$, compute $\hat{\epsilon}_i^* = y_i^* - X_i' \hat{\delta}^*$, where the bootstrap analogue of HFUL $\hat{\delta}^*$ is computed using $(y^*, X^*, Z)$. Then, construct the bootstrap statistic

$$\hat{T}^* = \sum_{i \neq j} \hat{\epsilon}_i^* P_{ij} \hat{\epsilon}_j^* \sqrt{\hat{V}^*} + K,$$

where $\hat{V}^* = \sum_{i \neq j} \hat{\epsilon}_i^2 P_{ij} \hat{\epsilon}_j^2 / K$, and its unstudentized version

$$\hat{T}_u^* = \sum_{i \neq j} \hat{\epsilon}_i^* P_{ij} \hat{\epsilon}_j^*.$$  

(6)

(7)

**Step 4:** Repeat Steps 1-3 $B$ times, and compute the bootstrap P values as $\hat{p}_T^* = B^{-1} \sum_{b=1}^B I\{\hat{T}_b^* \geq \hat{T}\}$ and $\hat{p}_{T_u}^* = B^{-1} \sum_{b=1}^B I\{\hat{T}_{u,b}^* \geq \hat{T}_u\}$. We reject the null hypothesis of no misspecification if the bootstrap P value is smaller than $\alpha$.

The following theorem states the asymptotic validity of the wild bootstrap. We assume the same regularity conditions as those in Chao et al. (2014), summarized by Assumption 1 in the Appendix.²

²The assumption rules out the case of weak identification, where the IV estimators becomes inconsistent. In this case, the bootstrap will also be inconsistent; e.g., see Wang and Doko Tchatoka (2018) and Wang (2020).
Theorem 3.1 Suppose that Assumption 1 holds. Then,
\[
\sup_{x \in \mathbb{R}} \left| P^* \left( \hat{T}^* \leq x \right) - P \left( \hat{T} \leq x \right) \right| \to^p 0, \quad \text{and} \quad \sup_{x \in \mathbb{R}} \left| P^* \left( \hat{T}_u^* \leq x \right) - P \left( \hat{T}_u \leq x \right) \right| \to^p 0,
\]
where \( P^* \) denotes the probability measure induced by the wild bootstrap procedure in (4)-(7).

Theorem 3.1 gives the validity of the wild bootstrap for both \( \hat{T} \) and \( \hat{T}_u \). In practice, the unstudentized wild bootstrap test is easier to compute given the simple formula of \( \hat{T}_u \). On the other hand, the studentized wild bootstrap test may achieve better size control as the test statistic is asymptotically pivotal under the current framework. In Section 4, we compare the finite sample performance of the two bootstrap tests in terms of both size and power.

4 Simulations

We conduct simulations by using the following data generating process:
\[
y_i = \delta X_i + \epsilon_i, \quad (8)
\]
\[
X_i = Z_i \pi + U_i, \quad (9)
\]
for \( i = 1, \ldots, n \), where \( U_i \sim N(0, 1) \), \( Z_i \sim N(0, I_K) \), \( \pi \equiv \frac{a}{\sqrt{K}} \iota_K \), and \( \iota_K \) is a \( K \)-vector of ones. Following Chao et al. (2014), we generate \( \epsilon_i \) as
\[
\epsilon_i = \rho U_i + \sqrt{\frac{1 - \rho^2}{\phi^2 + (0.86)^2}} (\phi v_{1i} + 0.86 v_{2i}), \quad (10)
\]
where \( v_{1i} \sim N(0, Z_{1i}^2) \), \( v_{2i} \sim N(0, (0.86)^2) \), \( \rho = 0.5 \), \( \phi = 0.2 \), and \( Z_{1i} \) is the first element in \( Z_i \). We set \( \delta = 1 \), \( B = 199 \), \( \alpha = 5\% \), and the number of Monte Carlo replications equals 5000. For \( w_i^* \) in the wild bootstrap, we use a Rademacher random variable with \( P(w_i^* = 1) = P(w_i^* = -1) = 1/2 \).

We compare the size and power of two asymptotic tests, namely, Hansen’s GMM J test (denoted as “asy.hansen.J”), Chao et al. (2014)’s jackknife J test (“asy.jack.J”), and the two wild bootstrap tests (“boot.unstud.jack.J” and “boot.stud.jack.J”). Throughout we use HFUL as \( \hat{\delta} \).

Figure 1 plots the size results as a function of \( \lambda \in \{0.1, 0.3, 0.5, 0.7, 0.9\} \), where \( \lambda = K/n \), for \( n = 300 \) and \( a \in \{1, 10, 50\} \). Both Hansen’s J test and Chao et al. (2014)’s J test tend to be rather conservative as \( \lambda \) increases, while the unstudentized bootstrap test has some slight over-rejections when \( a = 1 \). By contrast, the studentized bootstrap test has good size control.

Then, we investigate the power by generating the structural errors for (8) using \( \epsilon_i = \epsilon_i + \rho Z_1 U_i \). Figure 2 plots the power curves as a function of \( \rho Z \) for \( \lambda \in \{0.5, 0.9\} \), \( a \in \{1, 10, 50\} \), and \( n = 300 \). We observe that the studentized bootstrap test has an power improvement over the other tests, especially when the number of IVs is large relative to the sample size (\( \lambda = 0.9 \)).
5 Conclusion

We propose valid wild bootstrap tests for testing overidentifying restrictions under many/many weak instruments and heteroskedasticity. The studentized wild bootstrap test has excellent finite sample performance in terms of both size and power. We notice that Carrasco and Doukali (2022) recently proposed a regularized overidentification test, which is valid even when $K$ is larger than $n$. For future research agenda, it may be interesting to study the bootstrap validity for this regularized test.

Figure 1: Size of asymptotic and bootstrap tests

Figure 2: Power of asymptotic and bootstrap tests
References


A Appendix

The following notations are used for the bootstrap asymptotics: for any bootstrap statistic $T^*$ we write $T^* \to P^*$ in probability if for any $\delta > 0$, $\epsilon > 0$, $\lim_{n \to \infty} P[|T^*| > \delta] > \epsilon = 0$, i.e., $P^*(|T^*| > \delta) = o_P(1)$. Also, we write $T^* = O_{P^*}(n^{\epsilon})$ in probability if and only if for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $\lim_{n \to \infty} P[|T^*| > M_\delta] > \delta = 0$, i.e., for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $P^*(|n^{-\epsilon}T^*| > M_\delta) = o_P(1)$. Finally, we write $T^* \to d^* T$ in probability if, conditional on the sample, $T^*$ weakly converges to $T$ under $P^*$, for all samples contained in a set with probability converging to one. Specifically, we write $T^* \to d^* T$ in probability if and only if $E^*(f(T^*)) \to E(f(T))$ in probability for any bounded and uniformly continuous function $f$. To be concise, we sometimes use the short version $T^* \to P^*$ 0 to say that $T^* \to P^*$ 0 in probability, and use $T^* = O_{P^*}(n^{\epsilon})$ for $T^* = O_{P^*}(n^{\epsilon})$ in probability.

Let $Z_i', \epsilon_i, U_i'$ and $\Gamma_i'$ denote the $i$-th row of $Z, \epsilon, U,$ and $\Gamma$, respectively. Below we give the regularity conditions needed for Theorem 3.1.

Assumption 1

(i) $Z$ includes among its columns a vector of ones, $\text{rank}(Z) = K$, and there is a constant $C$ such that $P_{ii} \leq C < 1$ ($i = 1, ..., n$), $K \to \infty$.

(ii) $\Gamma_i = S_n z_i / \sqrt{n}$ where $S_n = \hat{S} \text{diag}(\mu_1 n, ..., \mu_{Gn})$ and $\hat{S}$ is nonsingular. Also, for each $j$ either $\mu_{ijn} = \sqrt{n}$ or $\mu_{ijn} / \sqrt{n} \to 0$, $\mu_n = \min_{1 \leq j \leq G} \mu_{ijn} \to \infty$, and $\sqrt{K}/\mu_n^2 \to 0$. Also, there is $C > 0$ such that $\left\| \sum_{i=1}^n z_i z_i' / n \right\| \leq C$ and $\lambda_{\min}(\sum_{i=1}^n z_i z_i' / n) \geq 1/C$, for $n$ sufficiently large.

(iii) There is a constant $C$ such that $(\epsilon_1, U_1, ..., \epsilon_n, U_n)$ are independent, with $E[\epsilon_i] = 0$, $E[U_i] = 0$, $E[\epsilon_i^2] < C$, $E[||U_i||^2] \leq C$, $\text{Var}(\epsilon_i', U_i') = \text{diag}(\hat{\Omega}_i, 0)$, and $\lambda_{\min}\left(\sum_{i=1}^n \hat{\Omega}_i / n\right) \geq 1/C$.

(iv) There is $\pi_{Kn}$ such that $\sum_{i=1}^n \left\| z_i - \pi_{Kn} Z_i \right\|^2 / n \to 0$.

(v) There is a constant, $C > 0$, such that with probability one, $\sum_{i=1}^n ||z_i||^4 / n^2 \to 0$, $E[\epsilon_i^4] \leq C$ and $E[||U_i||^4] \leq C$.

(vi) $\mu_n S_n^{-1} \to S_0$ and either (I) $K/\mu^2_n \to \kappa$ for finite $\kappa$ or (II) $K/\mu^2_n \to \infty$. Also, each of the following exists: $H_P = \lim_{n \to \infty} \sum_{i=1}^n (1 - P_{ii}) z_i z_i' / n$, $\Sigma_P = \lim_{n \to \infty} \sum_{i=1}^n (1 - P_{ii})^2 z_i z_i' \sigma_i^2 / n$, $\Psi = \lim_{n \to \infty} \sum_{i \neq j} p_{ij}^2 \left( \sigma_i^2 E[\tilde{U}_j U_i'] + E[\tilde{U}_j \epsilon_i] E[\epsilon_j U_i'] \right) / K$, where $\sigma_i^2 = E[\epsilon_i^2], \gamma_n = \sum_{i=1}^n E[U_i \epsilon_i] / \sum_{i=1}^n \sigma_i^2$, and $\tilde{U} = U - \epsilon \gamma_n$ having $i$-th row $U_{i}'$.

Proof of Theorem 3.1. We focus on the proof for the studentized version of the bootstrap test.
The proof of the unstudentized version is very similar, thus omitted. First, note that

\[
\frac{\sum_{i \neq j} \epsilon_i^* P_{ij} \epsilon_j^*}{\sqrt{K}} = \sum_{i \neq j} \left( \epsilon_i^* - X_i^*(\delta^* - \hat{\delta}) \right)' P_{ij} \left( \epsilon_j^* - X_j^*(\hat{\delta} - \hat{\delta}) \right) / \sqrt{K}
\]

\[
= \frac{\sum_{i \neq j} \epsilon_i^* P_{ij} \epsilon_j^*}{\sqrt{K}} + (\delta^* - \hat{\delta})' S_n \left( S_n^{-1} \sum_{i \neq j} X_i^* P_{ij} X_j^* S_n^{-1} \right) S_n' (\hat{\delta} - \hat{\delta}) / \sqrt{K}
\]

\[
+ 2(\delta^* - \hat{\delta})' S_n \left( S_n^{-1} \sum_{i \neq j} X_i^* P_{ij} \epsilon_j^* \right) / \sqrt{K}. \quad (A.11)
\]

Second, by using similar arguments as those in Wang and Kaffo (2016) and Theorem 2 of Hausman et al. (2012), we can show that for both Case (I) \((K/\mu_n \rightarrow \kappa < \infty)\) and Case (II) \((K/\mu_n \rightarrow \infty)\), 

\[
S_n' (\hat{\delta} - \hat{\delta}) = O_p(1). \quad \text{More specifically, let } \hat{\alpha}^*(\delta) = \sum_{i \neq j} \epsilon_i^*(\delta) P_{ij} \epsilon_j^*(\delta) / \epsilon^*(\delta)' \epsilon^*(\delta), \text{ where } \epsilon_i^*(\delta) = y_i^* - X_i^* \delta, \text{ and}
\]

\[
\hat{D}^*(\delta) = - \left[ \frac{\epsilon^*(\delta)' \epsilon^*(\delta)}{2} \right] \frac{\partial}{\partial \delta} \left[ \sum_{i \neq j} \epsilon_i^*(\delta) P_{ij} \epsilon_j^*(\delta) / \epsilon^*(\delta)' \epsilon^*(\delta) \right]
\]

\[
= \sum_{i \neq j} X_i^* P_{ij} \epsilon_j^*(\delta) - \epsilon^*(\delta)' \epsilon^*(\delta) \hat{\alpha}^*(\delta) \hat{\gamma}^*(\delta), \quad (A.12)
\]

where \(\hat{\gamma}^*(\delta) = X^*(\delta)' / \epsilon^*(\delta)' \epsilon^*(\delta)\). Note that \(S_n' (\hat{\delta} - \hat{\delta}) = \left( S_n^{-1} (\partial \hat{D}^*(\hat{\delta}) / \partial \delta) S_n^{-1} \right)^{-1} S_n^{-1} \hat{D}^*(\hat{\delta})\),

where \(\hat{\delta}\) lies on the line joining \(\delta^*\) and \(\hat{\delta}\). Also note that by Markov inequality and the current wild bootstrap procedure,

\[
\epsilon^* / n = \sum_{i=1}^n E^*[\epsilon_i^2] / n + O_p(1 / \sqrt{n}) = \sum_{i=1}^n \epsilon_i^2 / n + O_p(1 / \sqrt{n}) = O_p(1),
\]

\[
X^* \epsilon^* / n = \sum_{i=1}^n E^*[X_i^* \epsilon_i^*] / n + O_p(1 / \sqrt{n}) = \sum_{i=1}^n \tilde{v}_i \epsilon_i / n + O_p(1 / \sqrt{n}) = O_p(1). \quad (A.13)
\]

Also, let \(\tilde{\alpha}^*\) and \(\tilde{\gamma}^*\) denote \(\hat{\alpha}^*(\hat{\delta})\) and \(\hat{\gamma}^*(\hat{\delta})\), respectively. By (A.13), \(\tilde{\gamma}^* = O_p(1)\). Then, we obtain that \(\tilde{\alpha}^* = o_p(\mu_n^2 / n)\) under the same arguments as in the proof for Lemma A5 of Hausman et al. (2012), and we have

\[
S_n^{-1} \hat{D}^*(\hat{\delta}) = S_n^{-1} \left( X^* P \epsilon^* - \sum_{i=1}^n P_{ii} X_i^* \epsilon_i^* + \epsilon^* \epsilon^* \tilde{\alpha}^* \tilde{\gamma}^* \right)
\]

\[
= S_n^{-1} \sum_{i \neq j} X_i^* P_{ij} \epsilon_j^* + o_p(1)
\]

\[
= S_n^{-1} \sum_{i \neq j} \tilde{Y}_i P_{ij} \epsilon_j^* + S_n^{-1} \sum_{i \neq j} v_i^* P_{ij} \epsilon_j^* + o_p(1) = O_p(1), \quad (A.14)
\]

where \(\tilde{Y}_i = Z_i^2 \tilde{I}\), by using the fact that \(E^*[\epsilon_i^*] = \hat{\epsilon}_i E^*[\omega_i^*] = 0\), \(E^*[v_i^*] = \tilde{v}_i E^*[v_i^*] = 0\), and by Markov inequality. Similarly, by following the arguments in the proof of Lemma A7 of Hausman et al. (2012),
we obtain
\[-S_n^{-1}(\partial \hat{D}^*(\hat{\delta}))/\partial \delta)S_n^{-1}' = S_n^{-1} \sum_{i \neq j} X_i^* P_{ij} X_j^* S_n^{-1}' + o_{P^*}(1) = O_{P^*}(1), \]
\[\text{(A.15)}\]
Therefore, \(S_n' (\hat{\delta}^* - \hat{\delta}) = O_{P^*}(1)\), given (A.14) and (A.15).

Then, we have for both Case (I) and Case (II),
\[\frac{\sum_{i \neq j} \hat{\epsilon}_i^* P_{ij} \hat{\epsilon}_j^*}{\sqrt{K}} = \frac{\sum_{i \neq j} \epsilon_i^* P_{ij} \epsilon_j^*}{\sqrt{K}} + o_{P^*}(1), \]
\[\text{(A.16)}\]
by using (A.11), \(S_n' (\hat{\delta}^* - \hat{\delta}) = O_{P^*}(1)\), \(S_n^{-1} \sum_{i \neq j} X_i^* P_{ij} X_j^* S_n^{-1}' = O_{P^*}(1)\), and \(S_n^{-1} \sum_{i \neq j} X_i^* P_{ij} \epsilon_j^* = O_{P^*}(1)\). In addition, let \(V_n^* = \sum_{i \neq j} \sigma_i^2 P_{ij}^2 \epsilon_j^2 / K\), where \(\sigma_i^2 \equiv E^*[\epsilon_i^2]\). We note that \(E^*[\epsilon_i^4] = \hat{\epsilon}_i^4\) is bounded in probability by Assumption 1(v), and \(E^* \left[ \sum_{i \neq j} (\epsilon_i^* P_{ij} \epsilon_j^*)^2 \right] = K V_n^*\). It follows by Lemma A2 of Chao et al. (2012) that
\[\frac{\sum_{i \neq j} \epsilon_i^* P_{ij} \epsilon_j^*}{\sqrt{K V_n^*}} \to d^* N(0, 1) \text{ in probability.} \]
\[\text{(A.17)}\]
Now we show \(\tilde{V}_n^* - V_n^* \to_{P^*} 0\). By \(\hat{\delta}^* \to_{P^*} \hat{\delta}\), we obtain that w.p.a.1, \(\|\hat{\delta}^* - \hat{\delta}\|^2 \leq \|\delta^* - \delta\|^2\)
\[\|\hat{\epsilon}_i^2 - \epsilon_i^2\| \leq 2 \|X_i\| \|\hat{\delta}^* - \hat{\delta}\| + \|X_i\| \|\delta^* - \delta\|^2 \leq d_i \|\delta^* - \delta\|, \]
\[\text{(A.18)}\]
for \(d_i = 3(1 + \|X_i\|^2)\). Also by \(\sum_{i=1}^n \sum_{j=1}^n P_{ij}^2 = \sum_{i=1}^n P_{ii} = K\), we have \(E \left[ \sum_{i \neq j} P_{ij}^2 d_i d_j \right] / K \leq C \sum_{i \neq j} P_{ij}^2 / K \leq C\), so that by Markov inequality \(\sum_{i \neq j} P_{ij}^2 d_i d_j / K = O_{P^*}(1)\), which implies that \(\sum_{i \neq j} P_{ij}^2 d_i d_j / K = O_{P^*}(1)\). Similarly, since \(\hat{\epsilon}_i^2\) is bounded in probability, we have w.p.a.1,
\[E^* \left[ \sum_{i \neq j} P_{ij}^2 \epsilon_i^2 \epsilon_j^2 d_j / K \right] / K \leq C \sum_{i \neq j} P_{ij}^2 / K \leq C. \]
\[\text{(A.19)}\]
Therefore, for \(V_n^* = \sum_{i \neq j} P_{ij}^2 \epsilon_i^2 \epsilon_j^2 / K\) and \(\tilde{V}_n^* = \sum_{i \neq j} P_{ij}^2 \epsilon_i^2 \epsilon_j^2 / K\) we have
\[\left| \tilde{V}_n^* - V_n^* \right| \leq \sum_{i \neq j} P_{ij}^2 \left( \epsilon_i^2 \epsilon_j^2 - \epsilon_i^2 \epsilon_j^2 \right) / K \]
\[\leq \|\hat{\delta}^* - \hat{\delta}\|^2 \sum_{i \neq j} P_{ij}^2 d_i d_j / K + 2 \|\hat{\delta}^* - \hat{\delta}\| \sum_{i \neq j} P_{ij}^2 \epsilon_i^2 \epsilon_j^2 d_j / K \to_{P^*} 0. \]
\[\text{(A.20)}\]
In addition, note that by the choice of \(u_i^*\) for the wild bootstrap procedure, \(V_n^* = \sum_{i \neq j} P_{ij}^2 \epsilon_i^2 \epsilon_j^2 / K = \sum_{i \neq j} P_{ij}^2 \epsilon_i^2 \epsilon_j^2 / K = \tilde{V}_n^*\). Therefore, \(\tilde{V}_n^* - V_n^* \to_{P^*} 0\).

By the Slutsky Theorem and (A.17),
\[\frac{\sum_{i \neq j} \hat{\epsilon}_i^* P_{ij} \hat{\epsilon}_j^*}{\sqrt{K V_n^*}} = \frac{\sum_{i \neq j} \epsilon_i^* P_{ij} \epsilon_j^*}{\sqrt{K V_n^*}} + o_{P^*}(1) \to \sqrt{\frac{V_n^*}{\tilde{V}_n^*}} \binom{\tilde{V}_n^*}{\sqrt{K V_n^*}} + o_{P^*}(1) \to d^* N(0, 1), \]
\[\text{(A.21)}\]
in probability. In addition, \(\sum_{i \neq j} \hat{\epsilon}_i P_{ij} \hat{\epsilon}_j / \sqrt{K V_n^*} \to_{d^*} N(0, 1)\) by Theorem 1 of Chao et al. (2014). The result of bootstrap validity follows by Polya’s Theorem. □