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 *

26 October, 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.[1] In addition, it is found in the literature that for IV regressions, carefully designed bootstrap procedures typically provide finite sample improvement over asymptotic approximations.[2] 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

$$y = X\delta + \epsilon, \quad X = \Gamma + U, \tag{1}$$

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. Γ is the $n \times G$ reduced form matrix, and ϵ and U are, respectively, an $n \times 1$ vector and an $n \times G$ matrix of disturbances. The estimation of δ 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 δ, 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 δ, $\hat{\epsilon} = (\hat{\epsilon}_1, ..., \hat{\epsilon}_n)'$, and $\hat{\epsilon}(2) = (\hat{\epsilon}_1^2, ..., \hat{\epsilon}_n^2)'$. 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_{ij}^2. The test statistic takes the form

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

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

[1]For a comprehensive review of the related literature, see Anatolyev (2019) and the references therein.

[2]See, e.g., Davidson and MacKinnon (2008, 2010, 2015), Moreira, Porter, and Suarez (2009), Wang and Liu (2015), Wang and Kaffo (2016), Kaffo and Wang (2017), Wang and Doko Tchatoka (2018), Finlay and Magnusson (2019), Wang (2020), Young (2020), Wang and Zhang (2021), among others.

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 τ-th quantile of the chi-squared distribution with r degrees of freedom, has asymptotic rejection probability equal to α under many/many weak instruments and heteroskedasticity. For the wild bootstrap test, we also consider an unstudentized version of $\hat{T}$, namely,

$$\hat{T}_u \quad = \quad \sum_{i \neq j} \hat{\epsilon}_i P_{ij} \hat{\epsilon}_j. \tag{3}$$

3 Wild bootstrap overidentificaiton tests

Our wild bootstrap procedure is as follows:

Step 1: The bootstrap error terms ϵ_i^* and v_i^* are obtained by

$$\epsilon_i^* = \hat{\epsilon}_i w_i^*, \text{ and } v_i^* = \tilde{v}_i \omega_i^*, \quad i = 1, ..., n, \tag{4}$$

where w_i^* is a random variable with mean zero, variance one, and independent from the data, while $\tilde{v}_i$ 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

$$X_i^* = Z_i' \tilde{\Pi} + v_i^*, \quad y_i^* = X_i' \hat{\delta} + \epsilon_i^*, \ i = 1, ..., n, \tag{5}$$

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

Step 3: For $i = 1, ..., n$, 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

$$\hat{T}^* \quad = \quad \frac{\sum_{i \neq j} \hat{\epsilon}_i^* P_{ij} \hat{\epsilon}_j^*}{\sqrt{\hat{V}^*}} + K, \text{ where } \hat{V}^* = \frac{\sum_{i \neq j} \hat{\epsilon}_i^{*2} P_{ij}^2 \hat{\epsilon}_j^{*2}}{K}, \tag{6}$$

and its unstudentized version

$$\hat{T}_u^* \quad = \quad \sum_{i \neq j} \hat{\epsilon}_i^* P_{ij} \hat{\epsilon}_j^*. \tag{7}$$

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 $\hat{p}_{Tu}^* = B^{-1} \sum_{b=1}^B I\{\hat{T}_{u,b}^* \geq \hat{T}_u\}$. We reject the null hypothesis of no misspecification if the bootstrap P value is smaller than α.

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.[3]

[3]The 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,*

$$\sup_{x\in R}\left|P^*\left(\hat{T}^*\leq x\right)-P\left(\hat{T}\leq x\right)\right|\to^p 0,\quad \textit{and}\quad \sup_{x\in R}\left|P^*\left(\hat{T}_u^*\leq x\right)-P\left(\hat{T}_u\leq x\right)\right|\to^p 0,$$

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 performace of the two bootstrap tests in terms of both size and power.

4 Simulations

We conduct simulations by using the following data generating process:

$$y_i = \delta X_i + \epsilon_i, \tag{8}$$

$$X_i = Z_i'\pi + U_i, \tag{9}$$

for $i = 1, ..., n$, where $U_i \sim N(0,1)$, $Z_i \sim N(0, I_K)$, $\pi = \frac{a}{\sqrt{K}}\iota_K$, and ι_K is a K-vector of ones. Following Chao et al. (2014), we generate ϵ_i as

$$\epsilon_i = \rho U_i + \sqrt{\frac{1-\rho^2}{\phi^2 + (0.86)^4}}\left(\phi v_{1i} + 0.86 v_{2i}\right), \tag{10}$$

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 λ 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 $e_i = \epsilon_i + \rho_Z Z_{1i}$. Figure 2 plots the power curves as a function of ρ_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

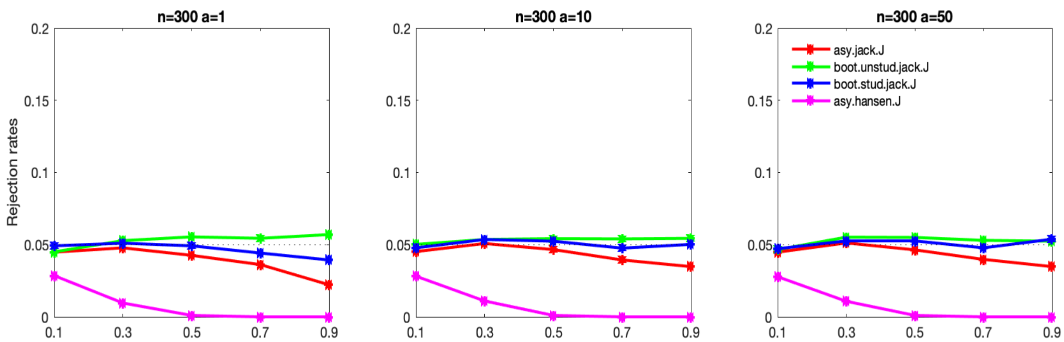
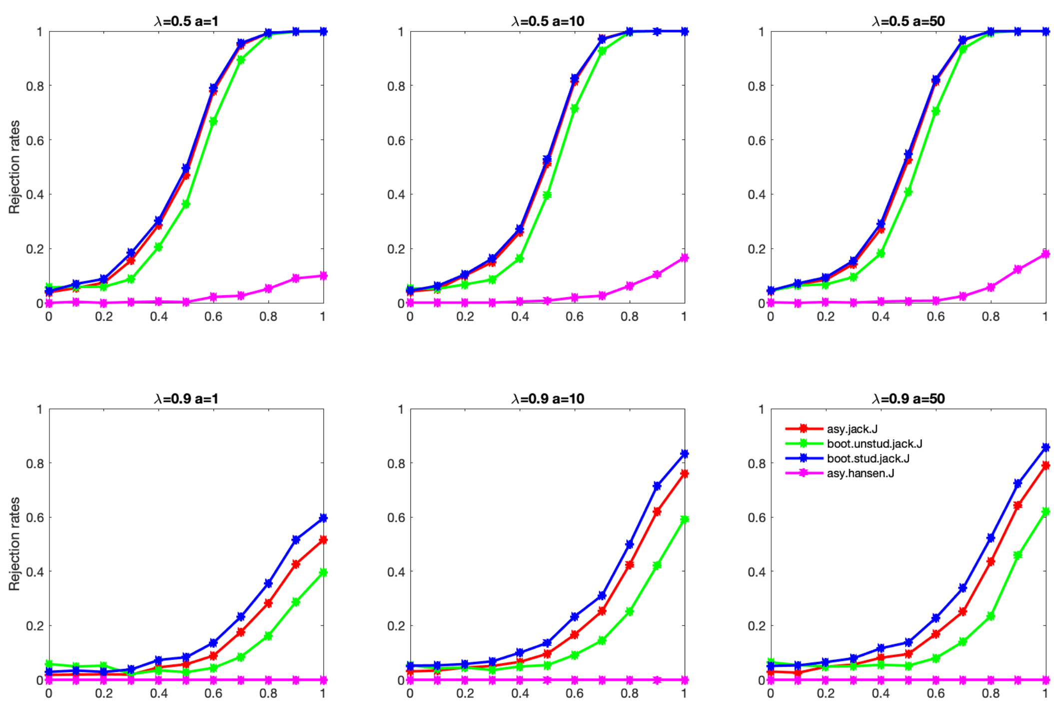


Figure 2: Power of asymptotic and bootstrap tests



References

Anatolyev, S. (2019): "Many instruments and/or regressors: A friendly guide," *Journal of Economic Surveys*, 33(2), 689–726.

Anatolyev, S., and N. Gospodinov (2011): "Specification Testing In Models With Many Instruments," *Econometric Theory*, 27(02), 427–441.

Carrasco, M., and M. Doukali (2022): "Testing overidentifying restrictions with many instruments and heteroscedasticity using regularised jackknife IV," *The Econometrics Journal*, 25(1), 71–97.

Chao, J. C., J. A. Hausman, W. K. Newey, N. R. Swanson, and T. Woutersen (2014): "Testing overidentifying restrictions with many instruments and heteroskedasticity," *Journal of Econometrics*, 178, 15–21.

Davidson, R., and J. G. MacKinnon (2008): "Bootstrap inference in a linear equation estimated by instrumental variables," *Econometrics Journal*, 11(3), 443–477.

——— (2010): "Wild Bootstrap Tests for IV Regression," *Journal of Business & Economic Statistics*, 28(1), 128–144.

——— (2015): "Bootstrap tests for overidentification in linear regression models," *Econometrics*, 3(4), 825–863.

Finlay, K., and L. M. Magnusson (2019): "Two applications of wild bootstrap methods to improve inference in cluster-IV models," *Journal of Applied Econometrics*, 34(6), 911–933.

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012): "Instrumental variable estimation with heteroskedasticity and many instruments," *Quantitative Economics*, 3(2), 211–255.

Kaffo, M., and W. Wang (2017): "On bootstrap validity for specification testing with many weak instruments," *Economics Letters*, 157, 107–111.

Moreira, M. J., J. R. Porter, and G. A. Suarez (2009): "Bootstrap validity for the score test when instruments may be weak," *Journal of Econometrics*, 149(1), 52–64.

Wang, W. (2020): "On the inconsistency of nonparametric bootstraps for the subvector Anderson–Rubin test," *Economics Letters*, p. 109157.

Wang, W., and F. Doko Tchatoka (2018): "On bootstrap inconsistency and Bonferroni-based size-correction for the subset Anderson–Rubin test under conditional homoskedasticity," *Journal of Econometrics*, 207(1), 188–211.

Wang, W., and M. Kaffo (2016): "Bootstrap inference for instrumental variable models with many weak instruments," *Journal of Econometrics*, 192(1), 231–268.

Wang, W., and Q. Liu (2015): "Bootstrap-based selection for instrumental variables model," *Economics Bulletin*, 35(3), 1886–1896.

Wang, W., and Y. Zhang (2021): "Wild bootstrap for instrumental variables regressions with weak and few clusters," *arXiv preprint arXiv:2108.13707*.

Young, A. (2020): "Consistency without inference: Instrumental variables in practical application," Discussion paper, London School of Economics.

A Appendix

The following notations are used for the bootstrap asymptotics: for any bootstrap statistic T^* we write $T^* \to^{P^*} 0$ in probability if for any $\delta > 0$, $\epsilon > 0$, $lim_{n\to\infty} P\left[P^*\left(|T^*| > \delta\right) > \epsilon\right] = 0$, i.e., $P^*\left(|T^*| > \delta\right) = o_P(1)$. Also, we write $T^* = O_{P^*}\left(n^{\varphi}\right)$ in probability if and only if for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $lim_{n\to\infty} P\left[P^*\left(|n^{-\varphi}T^*| > M_\delta\right) > \delta\right] = 0$, i.e., for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $P^*\left(|n^{-\varphi}T^*| > M_\delta\right) = o_P(1)$. 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 $E^*\left(f(T^*)\right) \to E\left(f(T)\right)$ 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 $T^* = O_{P^*}\left(n^{\varphi}\right)$ for $T^* = O_{P^*}\left(n^{\varphi}\right)$ in probability.

Let Z_i', ϵ_i, U_i' and Γ_i' denote the i-th row of Z, ϵ, U, and Γ, respectively. Below we give the regularity conditions needed for Theorem 3.1.

Assumption 1

(i) Z includes among its columns a vector of ones, $rank(Z) = K$, 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 $S_n = \tilde{S} diag(\mu_{1n}, ..., \mu_{Gn})$ and $\tilde{S}$ is nonsingular. Also, for each j either $\mu_{jn} = \sqrt{n}$ or $\mu_{jn}/\sqrt{n} \to 0$, $\mu_n = \min_{1\leq j\leq G} \mu_{jn} \to \infty$, 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 $\lambda_{min}\left(\sum_{i=1}^n z_i z_i'/n\right) \geq 1/C$, for n sufficiently large.

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

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

(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) $K/\mu_n^2 \to \kappa$ for finite κ or (II) $K/\mu_n^2 \to \infty$. Also, each of the following exists: $H_P = \lim_{n\to\infty} \sum_{i=1}^n (1 - P_{ii}) z_i z_i'/n$, $\Sigma_P = \lim_{n\to\infty} \sum_{i=1}^n (1 - P_{ii})^2 z_i z_i' \sigma_i^2/n$, $\Psi = \lim_{n\to\infty} \sum_{i\neq j} P_{ij}^2 \left(\sigma_i^2 E[\tilde{U}_j \tilde{U}_j'] + E[\tilde{U}_i \epsilon_i] E[\epsilon_j \tilde{U}_j']\right)/K$, 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 $\tilde{U}_i'$.

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

$$
\begin{aligned}
\frac{\sum_{i\neq j}\hat{\epsilon}_i^* P_{ij}\hat{\epsilon}_j^*}{\sqrt{K}} &= \sum_{i\neq j}\left(\epsilon_i^* - X_i^{*'}(\hat{\delta}^*-\hat{\delta})\right)' P_{ij}\left(\epsilon_j^* - X_j^{*'}(\hat{\delta}^*-\hat{\delta})\right)/\sqrt{K} \\
&= \frac{\sum_{i\neq j}\epsilon_i^* P_{ij}\epsilon_j^*}{\sqrt{K}} + (\hat{\delta}^*-\hat{\delta})' S_n\left(S_n^{-1}\sum_{i\neq j} X_i^* P_{ij} X_j^{*'} S_n^{-1'}\right) S_n'(\hat{\delta}^*-\hat{\delta})/\sqrt{K} \\
&\quad +2(\hat{\delta}^*-\hat{\delta})' S_n\left(S_n^{-1}\sum_{i\neq j} X_i^* P_{ij}\epsilon_j^*\right)/\sqrt{K}. \qquad (A.11)
\end{aligned}
$$

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) ($K/\mu_n^2 \to \kappa < \infty$) and Case (II) ($K/\mu_n^2 \to \infty$), $S_n'(\hat{\delta}^*-\hat{\delta}) = O_{P^*}(1)$. More specifically, let $\tilde{\alpha}^*(\delta) = \sum_{i\neq j}\epsilon_i^*(\delta)P_{ij}\epsilon_j^*(\delta)/\epsilon^*(\delta)'\epsilon^*(\delta)$, where $\epsilon_i^*(\delta) = y_i^* - X_i^{*'}\delta$, and

$$
\begin{aligned}
\hat{D}^*(\delta) &= -\left[\frac{\epsilon^*(\delta)'\epsilon^*(\delta)}{2}\right]\frac{\partial}{\partial\delta}\left[\frac{\sum_{i\neq j}\epsilon_i^*(\delta)P_{ij}\epsilon_j^*(\delta)}{\epsilon^*(\delta)'\epsilon^*(\delta)}\right] \\
&= \sum_{i\neq j} X_i^* P_{ij}\epsilon_j^*(\delta) - \epsilon^*(\delta)'\epsilon^*(\delta)\tilde{\alpha}^*(\delta)\tilde{\gamma}^*(\delta), \qquad (A.12)
\end{aligned}
$$

where $\tilde{\gamma}^*(\delta) = X^{*'}\epsilon^*(\delta)/\epsilon^*(\delta)'\epsilon^*(\delta)$. Note that $S_n'(\hat{\delta}^*-\hat{\delta}) = \left(S_n^{-1}(\partial\hat{D}^*(\bar{\delta}^*)/\partial\delta)S_n^{-1'}\right)^{-1} S_n^{-1}\hat{D}^*(\hat{\delta})$, where $\bar{\delta}^*$ lies on the line joining $\hat{\delta}^*$ and $\hat{\delta}$. Also note that by Markov inequality and the current wild bootstrap procedure,

$$
\begin{aligned}
\epsilon^{*'}\epsilon^*/n &= \sum_{i=1}^n E^*[\epsilon_i^{*2}]/n + O_{P^*}\left(1/\sqrt{n}\right) = \sum_{i=1}^n \hat{\epsilon}_i^2/n + O_{P^*}\left(1/\sqrt{n}\right) = O_{P^*}(1), \\
X^{*'}\epsilon^*/n &= \sum_{i=1}^n E^*[X_i^*\epsilon_i^*]/n + O_{P^*}(1/\sqrt{n}) = \sum_{i=1}^n \tilde{v}_i\hat{\epsilon}_i/n + O_{P^*}\left(1/\sqrt{n}\right) = O_{P^*}(1). \qquad (A.13)
\end{aligned}
$$

Also, let $\tilde{\alpha}^*$ and $\tilde{\gamma}^*$ denote $\tilde{\alpha}^*(\hat{\delta})$ and $\tilde{\gamma}^*(\hat{\delta})$, 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

$$
\begin{aligned}
S_n^{-1}\hat{D}^*(\hat{\delta}) &= S_n^{-1}\left(X^{*'}P\epsilon^* - \sum_{i=1}^n P_{ii}X_i^*\epsilon_i^* + \epsilon^{*'}\epsilon^*\tilde{\alpha}^*\tilde{\gamma}^*\right) \\
&= S_n^{-1}\sum_{i\neq j} X_i^* P_{ij}\epsilon_j^* + o_{P^*}(1) \\
&= S_n^{-1}\sum_{i\neq j}\tilde{\Upsilon}_i P_{ij}\epsilon_j^* + S_n^{-1}\sum_{i\neq j} v_i^* P_{ij}\epsilon_j^* + o_{P^*}(1) = O_{P^*}(1), \qquad (A.14)
\end{aligned}
$$

where $\tilde{\Upsilon}_i = Z_i'\tilde{\Pi}$, 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

$$-S_n^{-1}(\partial\hat{D}^*(\bar{\delta}^*)/\partial\delta)S_n^{-1'} = S_n^{-1}\sum_{i\neq j} X_i^* P_{ij} X_j^{*'} S_n^{-1'} + o_{P^*}(1) = O_{P^*}(1), \tag{A.15}$$

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),

$$\frac{\sum_{i\neq j}\hat{\epsilon}_i^* P_{ij}\hat{\epsilon}_j^*}{\sqrt{K}} = \frac{\sum_{i\neq j}\epsilon_i^* P_{ij}\epsilon_j^*}{\sqrt{K}} + o_{P^*}(1), \tag{A.16}$$

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 $V_n^* = \sum_{i\neq j}\sigma_i^{*2}P_{ij}^2\sigma_j^{*2}/K$, 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 $E^*\left[\sum_{i\neq j}(\epsilon_i^* P_{ij}\epsilon_j^*)^2\right] = KV_n^*$. It follows by Lemma A2 of Chao et al. (2012) that

$$\frac{\sum_{i\neq j}\epsilon_i^* P_{ij}\epsilon_j^*}{\sqrt{KV_n^*}} \to^{d^*} N(0,1) \text{ in probability.} \tag{A.17}$$

Now we show $\hat{V}_n^* - V_n^* \to^{P^*} 0$. By $\hat{\delta}^* \to^{P^*} \hat{\delta}$, we obtain that w.p.a.1, $\left\|\hat{\delta}^* - \hat{\delta}\right\|^2 \le \left\|\hat{\delta}^* - \hat{\delta}\right\|$ and

$$|\hat{\epsilon}_i^{*2} - \epsilon_i^{*2}| \le 2||X_i||\left\|\hat{\delta}^* - \hat{\delta}\right\| + ||X_i||^2\left\|\hat{\delta}^* - \hat{\delta}\right\|^2 \le d_i\left\|\hat{\delta}^* - \hat{\delta}\right\|, \tag{A.18}$$

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 $E\left[\sum_{i\neq j} P_{ij}^2 d_i d_j\right]/K \le C\sum_{i\neq j} P_{ij}^2/K \le C$, 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,

$$E^*\left[\sum_{i\neq j} P_{ij}^2\epsilon_i^{*2}d_j\right]/K \le C\sum_{i\neq j} P_{ij}^2/K \le C. \tag{A.19}$$

Therefore, for $\hat{V}_n^* = \sum_{i\neq j} P_{ij}^2\hat{\epsilon}_i^{*2}\hat{\epsilon}_j^{*2}/K$ and $\tilde{V}_n^* = \sum_{i\neq j} P_{ij}^2\epsilon_i^{*2}\epsilon_j^{*2}/K$ we have

$$\begin{aligned}\left|\hat{V}_n^* - \tilde{V}_n^*\right| &\le \sum_{i\neq j} P_{ij}^2\left|\hat{\epsilon}_i^{*2}\hat{\epsilon}_j^{*2} - \epsilon_i^{*2}\epsilon_j^{*2}\right|/K \\ &\le \left\|\hat{\delta}^* - \hat{\delta}\right\|^2\sum_{i\neq j} P_{ij}^2 d_i d_j/K + 2\left\|\hat{\delta}^* - \hat{\delta}\right\|\sum_{i\neq j} P_{ij}^2\epsilon_i^{*2}d_j/K \to^{P^*} 0.\end{aligned} \tag{A.20}$$

In addition, note that by the choice of w_i^* for the wild bootstrap procedure, $V_n^* = \sum_{i\neq j} P_{ij}^2\hat{\epsilon}_i^2\hat{\epsilon}_j^2/K = \sum_{i\neq j} P_{ij}^2\epsilon_i^{*2}\epsilon_j^{*2}/K = \tilde{V}_n^*$. Therefore, $\hat{V}_n^* - V_n^* \to^{P^*} 0$.

By the Slutzky Theorem and (A.17),

$$\frac{\sum_{i\neq j}\hat{\epsilon}_i^* P_{ij}\hat{\epsilon}_j^*}{\sqrt{K\hat{V}_n^*}} = \frac{\sum_{i\neq j}\epsilon_i^* P_{ij}\epsilon_j^*}{\sqrt{K\hat{V}_n^*}} + \frac{o_{P^*}(1)}{\sqrt{\hat{V}_n^*}} = \sqrt{\frac{V_n^*}{\hat{V}_n^*}}\frac{\sum_{i\neq j}\epsilon_i^* P_{ij}\epsilon_j^*}{\sqrt{KV_n^*}} + o_{P^*}(1) \to^{d^*} N(0,1), \tag{A.21}$$

in probability. In addition, $\sum_{i\neq j}\hat{\epsilon}_i P_{ij}\hat{\epsilon}_j/\sqrt{K\hat{V}_n} \to^d N(0,1)$ by Theorem 1 of Chao et al. (2014). The result of bootstrap validity follows by Polya's Theorem. ■