On the Inconsistency of Nonparametric Bootstraps for the Subvector Anderson-Rubin Test

Wang, Wenjie

Nanyang Technological University

7 March 2020

On the Inconsistency of Nonparametric Bootstraps for the Subvector Anderson-Rubin Test

Wenjie Wang *

Nanyang Technological University

March 7, 2020

ABSTRACT

Bootstrap procedures based on instrumental variable (IV) estimates or t-statistics are generally invalid when the instruments are weak. The bootstrap may even fail when applied to identification-robust test statistics. For subvector inference based on the Anderson-Rubin (AR) statistic, Wang and Doko Tchatoka (2018) show that the residual bootstrap is inconsistent under weak IVs. In particular, the residual bootstrap depends on certain estimator of structural parameters to generate bootstrap pseudo-data, while the estimator is inconsistent under weak IVs. It is thus tempting to consider nonparametric bootstrap. In this note, under the assumptions of conditional homoskedasticity and one nuisance structural parameter, we investigate the bootstrap consistency for the subvector AR statistic based on the nonparametric i.i.d. bootstap and its recentered version proposed by Hall and Horowitz (1996). We find that both procedures are inconsistent under weak IVs: although able to mimic the weak-identification situation in the data, both procedures result in approximation errors, which leads to the discrepancy between the bootstrap world and the original sample. In particular, both bootstrap tests can be very conservative under weak IVs.

Key words: Nonparametric Bootstrap; Weak Identification; Weak Instrument; Subvector Inference; Anderson-Rubin Test.

JEL classification: C12; C13; C26.

* Division of Economics, School of Social Sciences, Nanyang Technological University. HSS-04-65, 14 Nanyang Drive, 637332, SINGAPORE. E-mail: wang.wj@ntu.edu.sg. This work was supported by NTU SUG Grant No.M4082262.SS0. We thank a referee for giving very insightful comments, which help to improve the paper substantially.

1. Introduction

Inference in the linear IV model with possibly weak instruments has received considerable attention. Recently, Young (2019) studies 1359 IV regressions in 31 papers published by the American Economic Association and finds that the IVs are often weak, and inference methods based-on normal approximations can be unreliable. Young (2019) advocates for the usage of bootstrap methods.

As pointed out by Andrews, Stock and Sun (2019, Sec.6), bootstrap procedures based on IV estimates or t-statistics are generally invalid when the instruments are weak. By contrast, appropriate bootstrap procedures based on identification-robust statistics may remain valid. For testing joint hypothesis in the homoskedastic case, Moreira, Porter and Suarez (2009) show the validity of residual bootstrap for score and Anderson-Rubin (AR) tests. However, for subvector inference based on the AR statistic, Wang and Doko Tchatoka (2018) show the inconsistency of the residual bootstrap. In particular, it depends on certain point estimator that is inconsistent under weak IVs.

It is thus tempting to consider nonparametric bootstrap, which is also the most widely used bootstrap method by empirical researchers. Under the assumptions of conditional homoskedasticity and one nuisance structural parameter, we investigate the bootstrap consistency for the subvector AR statistic based on the nonparametric i.i.d. bootstap (pairs bootstrap) and its recentered version proposed by Hall and Horowitz (1996). We show that both procedures are inconsistent under weak IVs: although able to mimic the weak-identification situation, both procedures result in approximation errors, leading to the discrepancy between the bootstrap and original sample. Asymptotic results show that both bootstrap tests can be very conservative under weak IVs, while the pairs bootstrap test can be very conservative even under strong IVs.

2. Setting and Preliminary Result

We consider the linear IV model

$$y = X\beta + W\gamma + \varepsilon, \tag{2.1}$$
$$(X : W) = Z(\Pi_x : \Pi_w) + (V_x : V_w), \tag{2.2}$$

where $y \in \mathbb{R}^n$ is dependent variable, $X \in \mathbb{R}^n$ and $W \in \mathbb{R}^n$ are endogenous explanatory variables, $Z \in \mathbb{R}^{n\times L}$ are instrumental variables, and $[\varepsilon : V_x : V_w] \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ are unobserved disturbances. $\beta \in \mathbb{R}$, $\gamma \in \mathbb{R}$, $\Pi_x \in \mathbb{R}^L$ and $\Pi_w \in \mathbb{R}^L$ are unknown parameters. We assume that L is fixed and $L \geq 2$.

We are interested in testing the subvector null hypothesis

$$H_0 : \beta = \beta_0 \ \text{ versus } \ H_1 : \beta \neq \beta_0, \tag{2.3}$$

where γ is a nuisance structural parameter in this context.

To introduce the test statistics, consider the problem of testing the joint hypothesis $H_0^*: \beta = \beta_0, \gamma = \gamma_0$, and define the AR test statistic as:

$$AR_n(\beta_0, \gamma_0) = \frac{(y - X\beta_0 - W\gamma_0)' P_Z (y - X\beta_0 - W\gamma_0)}{L\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0, \gamma_0)}, \tag{2.4}$$

where $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0, \gamma_0) = \frac{1}{n-L}(y - X\beta_0 - W\gamma_0)' M_Z (y - X\beta_0 - W\gamma_0)$, $P_Z = Z(Z'Z)^{-1}Z'$ and $M_Z = I_n - P_Z$. Then, the subvector AR statistic can be defined as:

$$AR_n(\beta_0, \tilde{\gamma}) = \min_{\gamma \in \mathbb{R}} AR_n(\beta_0, \gamma), \tag{2.5}$$

where $\tilde{\gamma} = \arg\min_{\gamma \in \mathbb{R}} AR_n(\beta_0, \gamma)$. It is well known that the solution of the minimization problem in (2.5) is given by the null-constrained LIML estimator of γ, i.e.,

$$\tilde{\gamma} = \left[W' \left(P_Z - \frac{\tilde{\kappa}}{n-L} M_Z \right) W \right]^{-1} W' \left(P_Z - \frac{\tilde{\kappa}}{n-L} M_Z \right) \tilde{y}(\beta_0), \tag{2.6}$$

where $\tilde{y}(\beta_0) = y - X\beta_0$, $\tilde{\kappa}$ is the smallest root of the characteristic polynomial

$$\left| \kappa \hat{\Omega}_W - (\tilde{y}(\beta_0) : W)' P_Z (\tilde{y}(\beta_0) : W) \right| = 0, \tag{2.7}$$

and $\hat{\Omega}_W = \frac{1}{n-L}(\tilde{y}(\beta_0) : W)' M_Z (\tilde{y}(\beta_0) : W)$. We make the following assumptions.

Assumption 2.1 $\left\{ (\varepsilon_i, V_{xi}, V_{wi}, Z_i')' : 1 \leq i \leq n \right\}$ *are i.i.d. across i with distribution F.*

Assumption 2.2 *(i)* $E_F[U_i] = 0$ *and* $E_F[Z_i U_i'] = 0$ *where* $U_i = (\varepsilon_i, V_{xi}, V_{wi})'$; *and (ii)* $E_F[\|T_i\|^{2+\zeta}] \leq K < \infty$ *for some* $K \geq 0$, $\zeta > 0$ *and for all* $T_i \in \{Z_i\varepsilon_i, Z_i V_{w,i}, V_{w,i}\varepsilon_i, \varepsilon_i, V_{w,i}, Z_i\}$; $E_F[Z_i Z_i'] := Q_{ZZ}$, $E_F[U_i U_i'] := \Sigma_{UU}$, $E_F\left[vec(Z_i U_i')(vec(Z_i U_i'))'\right] = \Sigma_{UU} \otimes Q_{ZZ}$, *and for* $A \in \{Q_{ZZ}, \Sigma_{UU}\}$, $\lambda_{min}(A) \geq \varsigma$ *for some* $\varsigma > 0$. $E_F[\cdot]$ *denotes the expectation under F,* $\otimes$ *the Kronecker product, and* $\lambda_{min}(\cdot)$ *the smallest eigenvalue of a matrix.*

Assumption 2.3 *When the sample size n converges to infinity, we have*:

$$n^{1/2} Q_{ZZ}^{1/2} \Pi_{n,w} \sigma_{v_w v_w}^{-1/2} \to h_{ww} \in \mathbb{R}^L \text{ with } ||h_{ww}|| < \infty.$$

When Assumptions **2.1** - **2.2** hold, by Lyapunov-type CLTs we have

$$n^{-1/2} vec\left(Z'[\varepsilon : V_w]\right) \xrightarrow{d} vec\left(\psi_{Z\varepsilon}, \psi_{ZV_w}\right) \sim N(0, \Sigma \otimes Q_{ZZ}), \text{ with } \Sigma = \begin{bmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{v_w\varepsilon} \\ \sigma_{v_w\varepsilon} & \sigma_{v_w v_w} \end{bmatrix}.$$

Then define $\psi_{V_w} = Q_{ZZ}^{-1/2}\psi_{ZV_w}\sigma_{v_w v_w}^{-1/2}$, $\psi_\varepsilon = Q_{ZZ}^{-1/2}\psi_{Z\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1/2}$, where $vec(\psi_\varepsilon, \psi_{V_w}) \sim N(0, \Sigma_h \otimes I_L)$, with $\Sigma_h = \begin{pmatrix} 1 & h_{w\varepsilon} \\ h_{w\varepsilon} & 1 \end{pmatrix}$ and $h_{w\varepsilon} = \sigma_{v_w v_w}^{-1/2}\sigma_{v_w\varepsilon}\sigma_{\varepsilon\varepsilon}^{-1/2}$. Also define

$$\Delta_h \;=\; (\Psi_h'\Psi_h - \kappa_h)^{-1}(\Psi_h'\psi_\varepsilon - \kappa_h h_{w\varepsilon}),\;\; S_h = \psi_\varepsilon - \Psi_h\Delta_h, \tag{2.8}$$

where $\Psi_h = h_{ww} + \psi_{V_w}$, and κ_h is the smallest root of $\left|(\psi_\varepsilon : \Psi_h)'(\psi_\varepsilon : \Psi_h) - \kappa_h\Sigma_h\right| = 0$. The following theorem gives the null limiting distribution of $AR_n(\beta_0, \tilde{\gamma})$ under weak IVs, which is nonstandard and characterized by $h = (h_{ww}, h_{w\varepsilon})$. Note that $\|h_{ww}\|^2$ characterizes the identification strength for the nuisance parameter γ, and $h_{w\varepsilon}$ characterizes the degree of endogeneity.

Theorem 2.4 *(Theorem 3.2 of Wang and Doko Tchatoka (2018)) Suppose that Assumptions* **2.1**–**2.3** *are satisfied. If further H_0 holds, then we have:*

$$AR_n(\beta_0, \tilde{\gamma}) \;\;\to^d\;\; \xi_h = \frac{1}{L}\left\|\left(1 - 2h_{w\varepsilon}\Delta_h + \Delta_h^2\right)^{-1/2} S_h\right\|^2,$$

where Δ_h and S_h are defined in (2.8)*.*

3. Bootstrapping the Subvector Anderson-Rubin Test

In this section, we study nonparametric bootstrap procedures for the subvector AR test. The motivation is that to implement residual bootstrap, one has to use the null-restricted LIML estimator $\tilde{\gamma}$ to generated the bootstrap disturbances for the structural equation (2.1); e.g., see Section 4.1 of Wang and Doko Tchatoka (2018). However, $\tilde{\gamma}$ cannot consistently estimate γ under weak IVs, resulting in discrepancy between the bootstrap and original data.

We write P^* to denote the probability measure induced by a bootstrap procedure conditional on the data, and write E^* and Var^* to denote the expected value and variance with respect to P^*. Following Gonçalves and White (2004), for any bootstrap statistic T^* we write $T^* \to^{P^*} 0$ in probability if for any $\delta > 0$, $\varepsilon > 0$, $lim_{n\to\infty}P[P^*(|T^*| > \delta) > \varepsilon] = 0$, i.e., $P^*(|T^*| > \delta) = o_P(1)$. 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.

The nonparametric i.i.d. bootstrap procedure (pairs bootstrap) is implemented by sampling $\left(y_1^*, X_1^*, W_1^*, Z_1^{*'}\right)$, ..., $\left(y_n^*, X_n^*, W_n^*, Z_n^{*'}\right)$ randomly with replacement from the sample. Then, the bootstrap statistic can be defined as:

$$AR_{n,p}^*(\beta_0, \tilde{\gamma}^*) = \frac{(y^* - X^*\beta_0 - W^*\tilde{\gamma}^*)' P_{Z^*}(y^* - X^*\beta_0 - W^*\tilde{\gamma}^*)}{L\hat{\sigma}_{\varepsilon\varepsilon}^*(\beta_0, \tilde{\gamma}^*)}, \tag{3.1}$$

where $\tilde{\gamma}^*$ and $\hat{\sigma}^*_{\varepsilon\varepsilon}(\beta_0, \tilde{\gamma}^*)$ are the analogues of $\tilde{\gamma}$ and $\hat{\sigma}_{\varepsilon\varepsilon}(\beta_0, \tilde{\gamma})$ computed using bootstrap samples.

To understand the bootstrap failure under weak IVs, we note that

$$n^{-1/2}Z^{*'}(y^* - X^*\beta_0 - W^*\tilde{\gamma}^*) = n^{-1/2}Z^{*'}\varepsilon^* + n^{-1/2}Z^{*'}W^*(\gamma - \tilde{\gamma}^*) \tag{3.2}$$

where $\varepsilon^* = y^* - X^*\beta_0 - W^*\gamma$. Furthermore, $n^{-1/2}Z^{*'}\varepsilon^* = n^{-1/2}\left(Z^{*'}\varepsilon^* - Z'\varepsilon\right) + n^{-1/2}Z'\varepsilon$, $n^{-1/2}Z^{*'}W^* = \left(n^{-1}Z^{*'}Z^*\right)n^{1/2}\Pi_{n,w} + n^{-1/2}\left(Z^{*'}V_w^* - Z'V_w\right) + n^{-1/2}Z'V_w$, and a bootstrap CLT can be applied to $n^{-1/2}\left(Z^{*'}\varepsilon^* - Z'\varepsilon\right)$ and $n^{-1/2}\left(Z^{*'}V_w^* - Z'V_w\right)$. Therefore, the following (conditional) convergence in distribution holds:

$$\begin{aligned} & vec\left((n^{-1}Z^{*'}Z^*)^{-1/2}n^{-1/2}Z^{*'}\varepsilon^*\sigma_{\varepsilon\varepsilon}^{*-1/2},\ (n^{-1}Z^{*'}Z^*)^{-1/2}n^{-1/2}Z^{*'}W^*\sigma_{v_wv_w}^{*-1/2}\right) \\ \to^{d^*} \quad & vec\left(\psi_\varepsilon + \psi_\varepsilon^B, \Psi_h + \psi_{V_w}^B\right) \text{ in probability,} \end{aligned} \tag{3.3}$$

where $vec\left(\psi_\varepsilon^B, \psi_{V_w}^B\right) \sim N(0, \Sigma_h \otimes I_L)$, $\Psi_h = h_{ww} + \psi_{V_w}$, $\sigma^*_{\varepsilon\varepsilon} = E^*\left(\varepsilon_i^{*2}\right)$ and $\sigma^*_{v_wv_w} = E^*\left(V_{wi}^{*2}\right)$.

ψ_ε^B and $\psi_{V_w}^B$ are the bootstrap counterparts of ψ_ε and ψ_{V_w}, and correctly replicates the randomness in the original data. However, the nonparametric bootstrap is inconsistent under weak IVs. In particular, the original identification strength for γ is characterized by $\|h_{ww}\|^2$, while conditional on the data, the corresponding identification strength in the bootstrap world is characterized by $\|\Psi_h\|^2 = \|h_{ww} + \psi_{V_w}\|^2$. Therefore, although able to mimic the weak-identification situation ($\|\Psi_h\|^2$ is finite with probability approaching one when $\|h_{ww}\|^2$ is finite), the bootstrap generates approximation errors ψ_ε and ψ_{V_w}, whose values will depend on the specific realization of the sample. Theorem **3.1** presents the null limiting distribution of $AR^*_{n,p}(\beta_0, \tilde{\gamma}^*)$ under weak IVs.

Theorem 3.1 *Suppose that Assumptions* **2.1**–**2.3** *are satisfied. If further* H_0 *holds, then we have:*

$$AR^*_{n,p}(\beta_0, \tilde{\gamma}^*) \quad \to^{d^*} \quad \xi^B_{h,p} = \frac{1}{L}\left\|\left(1 - 2h_{w\varepsilon}\Delta_h^B + (\Delta_h^B)^2\right)^{-1/2} S^B_{h,p}\right\|^2,$$

in probability, where $\Delta_h^B = \left\{\left(\Psi_h + \psi_{v_w}^B\right)'\left(\Psi_h + \psi_{v_w}^B\right) - \kappa_h^B\right\}^{-1}\left\{\left(\Psi_h + \psi_{v_w}^B\right)'\left(\psi_\varepsilon + \psi_\varepsilon^B\right) - \kappa_h^B h_{w\varepsilon}\right\}$, κ_h^B *is the smallest root of* $\left|\kappa_h\Sigma_h - \left(\psi_\varepsilon + \psi_\varepsilon^B : \Psi_h + \psi_{v_w}^B\right)'\left(\psi_\varepsilon + \psi_\varepsilon^B : \Psi_h + \psi_{v_w}^B\right)\right| = 0$, *and* $S^B_{h,p} = (\psi_\varepsilon + \psi_\varepsilon^B) - (\Psi_h + \psi_{v_w}^B)\Delta_h^B$.

Now we consider the nonparametric bootstrap procedure proposed by Hall and Horowitz (1996), which recenters the moment conditions in the bootstrap world. This procedure leads to

$$AR^*_{n,r}(\beta_0, \tilde{\gamma}^*) = \frac{(\tilde{\varepsilon}^*(\beta_0, \tilde{\gamma}^*)'Z^* - \tilde{\varepsilon}(\beta_0, \tilde{\gamma})'Z)\,(Z^{*'}Z^*)^{-1}(Z^{*'}\tilde{\varepsilon}^*(\beta_0, \tilde{\gamma}^*) - Z'\tilde{\varepsilon}(\beta_0, \tilde{\gamma}))}{L\hat{\sigma}^*_{\varepsilon\varepsilon}(\beta_0, \tilde{\gamma}^*)}, \tag{3.4}$$

where $\tilde{\varepsilon}^*(\beta_0, \tilde{\gamma}^*) = y^* - X^*\beta_0 - W^*\tilde{\gamma}^*$, and $\tilde{\varepsilon}(\beta_0, \tilde{\gamma}) = y - X\beta_0 - W\tilde{\gamma}$. We note that

$$\begin{aligned} & n^{-1/2}\left(Z^{*\prime}\tilde{\varepsilon}^*(\beta_0,\tilde{\gamma}^*) - Z'\tilde{\varepsilon}(\beta_0,\tilde{\gamma})\right) \\ = \; & n^{-1/2}\left(Z^{*\prime}\varepsilon^* - Z'\varepsilon\right) + n^{-1/2}Z^{*\prime}W^*(\gamma - \tilde{\gamma}^*) + n^{-1/2}\left(Z'\varepsilon - Z'\tilde{\varepsilon}(\beta_0,\tilde{\gamma})\right), \end{aligned} \tag{3.5}$$

and the last term does not vanish under weak identification. This is very different from the strong-identification case in which the recentering bootstrap is shown by Hall and Horowitz (1996) to achieve asymptotic refinement for various tests. Theorem **3.2** characterizes the null limiting distribution of $AR^*_{n,r}(\beta_0, \tilde{\gamma}^*)$ under weak IVs and shows that the recentering bootstrap is inconsistent.

Theorem 3.2 *Suppose that Assumptions* **2.1**–**2.3** *are satisfied. If further* H_0 *holds, then we have:*

$$AR^*_{n,r}(\beta_0, \tilde{\gamma}^*) \quad \to^{d^*} \quad \xi^B_{h,r} = \frac{1}{L}\left\|\left(1 - 2h_{w\varepsilon}\Delta^B_h + (\Delta^B_h)^2\right)^{-1/2} S^B_{h,r}\right\|^2,$$

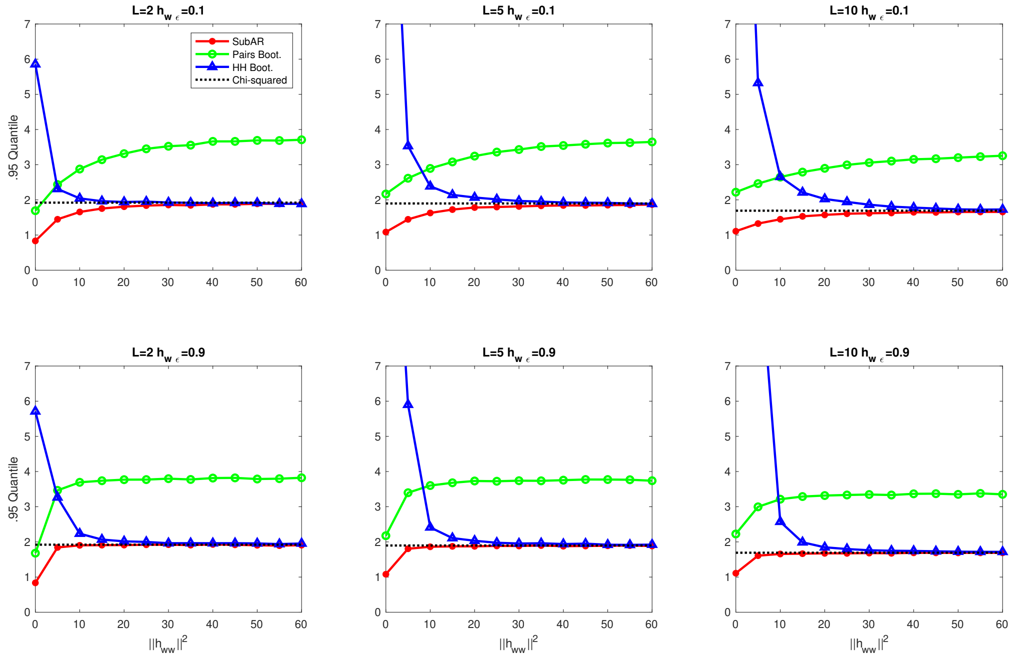
in probability, where Δ^B_h *is defined in Theorem 3.1, and* $S^B_{h,r} = (\Psi_h\Delta_h + \psi^B_\varepsilon) - (\Psi_h + \psi^B_{v_w})\Delta^B_h$.

To better understand the bootstrap statistics, we apply Theorems **2.4**, **3.1**, and **3.2**, and plot the 95% quantiles of ξ_h, $\xi^B_{h,p}$ and $\xi^B_{h,r}$ in Figure 1 with $||h_{ww}||^2 \in \{0, 10, \ldots, 60\}$, $L \in \{2, 5, 10\}$, and $h_{w\varepsilon} \in \{0.1, 0.9\}$ (by 100,000 simulation replications). The corresponding χ^2_{L-1} critical values (divided by L) are also plotted. The quantiles of $\xi^B_{h,p}$ turn out to be always higher than those of ξ_h, suggesting that the pairs bootstrap tests can be very conservative no matter the IVs are strong or weak. Indeed, we note that under H_0, $AR_n(\beta_0, \tilde{\gamma})$ is equivalent to a version of the J statistic, while Giurcanu and Presnell (2018, Theorem 2 9(e)) show that in the standard strong-identification case, instead of having a central chi-squared limiting distribution, the pairs bootstrap analogue of the J statistic has a non-central chi-squared limiting distribution. This bootstrap is therefore inconsistent even under strong IVs. By contrast, the quantiles of $\xi^B_{h,r}$ converge to those of χ^2_{L-1}/L when $||h_{ww}||^2$ become large. This is in line with Hall and Horowitz (1996), which shows the consistency of the recentering bootstrap for J tests under strong identification. However, we note that due to the inclusion of $\tilde{\gamma}$ in (3.4) when recentering the bootstrap moment conditions, the quantiles of $\xi^B_{h,r}$ can be much higher than those of ξ_h under weak IVs (as $\tilde{\gamma}$ becomes inconsistent), suggesting that the recentering bootstrap tests can be very conservative in this case.

Now we study the asymptotic size of the two bootstrap tests. Following Guggenberger, Kleibergen, Mavroeidis and Chen (2012) and Guggenberger, Kleibergen and Mavroeidis (2019), we first define the parameter space under the null hypothesis in (2.3):

$$\Theta = \left\{\theta = (\gamma, \Pi_x, \Pi_w, F): \; \gamma \in \mathbb{R}, \; \Pi_x \in \mathbb{R}^L, \; \Pi_w \in \mathbb{R}^L \text{ and } F \text{ such that Assumptions } \mathbf{2.1\text{–}2.2} \text{ hold}\right\}. \tag{3.6}$$

Figure 1. 95% quantiles of ξ_h, $\xi^B_{h,p}$ and $\xi^B_{h,r}$



Note: The results are based on 100,000 simulation replications.

Then the asymptotic size of the bootstrap tests is defined as:

$$AsySz\left[\hat{c}_{n,j}(1-\alpha)\right] := \limsup_{n\to\infty} \sup_{\theta\in\Theta} P_\theta\left[AR_n(\beta_0,\tilde{\gamma}_j) > \hat{c}_{n,j}(1-\alpha)\right], \tag{3.7}$$

where P_θ denotes probability of an event when the null data generating process is pinned down by $\theta \in \Theta$, and $\hat{c}_{n,j}(1-\alpha)$ denotes the $(1-\alpha)$-th quantile of the distribution of $AR^*_{n,j}(\beta_0,\tilde{\gamma}^*)$ for $j \in \{p,r\}$. The next theorem gives an explicit formula of the asymptotic size.

Theorem 3.3 *For $j \in \{p,r\}$, $AsySz[\hat{c}_{n,j}(1-\alpha)]$ equals $sup_{h\in H} P[\xi_h > \tilde{c}_{h,j}(1-\alpha)]$, where $\tilde{c}_{h,j}(1-\alpha)$ is the $(1-\alpha)$-th quantile of $\xi^B_{h,j}$ and H is defined in* (A.5)*.*

Table 1 reports the asymptotic sizes of the bootstrap tests for $\alpha = 0.05$ and $L \in \{2,...,11\}$, which are based on Theorem **3.3** and 100,000 simulation replications. The asymptotic sizes of the pairs bootstrap tests ("Pairs boot.") are much smaller than 0.05. By contrast, the recentering bootstrap tests ("HH boot.") achieves correct asymptotic size (up to simulation error), since it does consistently estimate the distribution of interest under strong identification so that its asymptotic null rejection probability equals 0.05 in this case. However, according to Figure 1, the recentering bootstrap can be very conservative under weak identification. In sum, we could not recommend either bootstrap method as there exist methods that both have correct asymptotic size and are less conservative such as the conditional subvector AR test proposed by Guggenberger et al. (2019).

Table 1. $AsySz\left[\hat{c}_{n,j}(1-\alpha)\right]$ for nominal size $\alpha = 0.05$.

L	Pairs Boot.	HH Boot.	L	Pairs Boot.	HH Boot.
2	0.0097	0.051	7	0.00061	0.049
3	0.0045	0.052	8	0.00048	0.050
4	0.0026	0.051	9	0.00030	0.049
5	0.0016	0.050	10	0.00021	0.049
6	0.0010	0.051	11	0.00017	0.050

Note: The results are based on 100,000 simulation replications.

4. Monte Carlo Simulation

We examine the finite sample performance of bootstrap tests by a small-scale Monte Carlo experiment. The disturbances are i.i.d. normal with mean zero, unit variance, and $h_{w\varepsilon} \in \{0.1, 0.5, 0.9\}$. Z_i's are distributed i.i.d. $N(0, I_L)$ with $L \in \{2, 10\}$. The IV strength is set at $||h_{ww}||^2 \in \{0, 4, 16, 64\}$. The experiment is executed with $n = 200$, $5,000$ Monte Carlo replications, and 299 replications of bootstrap samples. The nominal level is 5%, and Table 2 compares the pairs bootstrap, the recentering bootstrap, and the residual bootstrap in Moreira et al. (2009) and Wang and Doko Tchatoka (2018) ("Resid. Boot."). The pairs bootstrap does not reject, while the rejection frequencies of the recentering bootstrap increase when $||h_{ww}||^2$ or $h_{w\varepsilon}$ increases. However, the recentering bootstrap is also very conservative under weak IVs. These findings are in line with the asymptotic results. The residual bootstrap has the best performance, although it is also conservative when $||h_{ww}||^2$ is small (as it is also inconsistent for the subvector AR test under weak IVs).

Table 2. Null rejection frequencies (%) for $H_0 : \beta = \beta_0$ at $\alpha = 5\%$

		$L = 2$			$L = 10$		
$h_{w\varepsilon}$	$\|\|h_{ww}\|\|^2$	Pairs Boot.	HH Boot.	Resid. Boot.	Pairs Boot.	HH Boot.	Resid. Boot.
0.1	0	0	0	0.66	0	0	0.28
0.1	4	0	0.04	2.70	0	0	1.16
0.1	16	0	0.94	4.82	0	0	3.26
0.1	64	0	4.20	4.84	0	0.9	4.92
0.5	0	0	0	0.86	0	0	0.26
0.5	4	0	0.08	3.28	0	0	1.36
0.5	16	0	1.06	4.70	0	0	4.22
0.5	64	0	4.08	4.68	0	1.48	4.74
0.9	0	0	0	0.72	0	0	0.38
0.9	4	0	0.16	4.44	0	0	3.48
0.9	16	0	2.06	4.68	0	0.1	4.68
0.9	64	0	4.26	4.94	0	1.86	4.70

5. Conclusions

We show the inconsistency of two nonparametric bootstraps under weak IVs for the subvector AR test. Both methods can be very conservative under weak IVs and the pairs bootstrap can be very conservative even under strong IVs. We note that in the homoskedastic case, Guggenberger et al. (2012) provides appropriate chi-squared critical value, and Guggenberger et al. (2019) proposes a data-dependent critical value to further improve power. Kleibergen (2019) provides a subvector conditional likelihood ratio test. Wang and Doko Tchatoka (2018) proposes a Bonferroni-based size-correction method. For heteroskedastic data, Andrews (2017) proposes a two-step Bonferroni method that applies to nonlinear models.

References

Andrews, D. W., 2017. Identification-robust subvector inference. Technical report, Cowles Foundation Discussion Paper 2105.

Andrews, D. W., Guggenberger, P., 2010. Asymptotic size and a problem with subsampling and with the *m* out of *n* bootstrap. Econometric Theory 26(2), 426–468.

Andrews, I., Stock, J. H., Sun, L., 2019. Weak instruments in instrumental variables regression: Theory and practice. Annual Review of Economics 11, 727–753.

Giurcanu, M., Presnell, B., 2018. Bootstrap inference for misspecified moment condition models. Annals of the Institute of Statistical Mathematics 70(3), 605–630.

Gonçalves, S., White, H., 2004. Maximum likelihood and the bootstrap for nonlinear dynamic models. Journal of Econometrics 119(1), 199–219.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., 2019. A more powerful subvector anderson rubin test in linear instrumental variables regression. Quantitative Economics 10(2), 487–526.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., Chen, L., 2012. On the asymptotic sizes of subset anderson–rubin and lagrange multiplier tests in linear instrumental variables regression. Econometrica 80(6), 2649–2666.

Hall, P., Horowitz, J. L., 1996. Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica 64(4), 891–916.

Kleibergen, F., 2019. Efficient size correct subset inference in homoskedastic linear instrumental variables regression. Technical report, University of Amsterdam.

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

Wang, W. , Doko Tchatoka, F. , 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.

Young, A., 2019. Consistency without inference: Instrumental variables in practical application. Technical report, London School of Economics.

A. Appendix

The Appendix contains the proofs of the theoretical results in the paper.

PROOF OF THEOREM **3.1**

First, we note that the following decompositions hold:

$$\begin{aligned}
&\left(Z^{*\prime}Z^*\right)^{-1/2}Z^{*\prime}W^*\sigma_{v_wv_w}^{*-1/2}\\
=\;&\left(n^{-1}Z^{*\prime}Z^*\right)^{1/2}n^{1/2}\Pi_w\sigma_{v_wv_w}^{*-1/2}+\left(n^{-1}Z^{*\prime}Z^*\right)^{-1/2}n^{-1/2}\left(Z^{*\prime}V_w^*-Z'V_w\right)\sigma_{v_wv_w}^{*-1/2}\\
&+\left(n^{-1}Z^{*\prime}Z^*\right)^{-1/2}n^{-1/2}Z'V_w\sigma_{v_wv_w}^{*-1/2};\\
&\left(Z^{*\prime}Z^*\right)^{-1/2}Z^{*\prime}\varepsilon^*\sigma_{\varepsilon\varepsilon}^{*-1/2}\\
=\;&\left(n^{-1}Z^{*\prime}Z^*\right)^{-1/2}n^{-1/2}\left(Z^{*\prime}\varepsilon^*-Z'\varepsilon\right)\sigma_{\varepsilon\varepsilon}^{*-1/2}+\left(n^{-1}Z^{*\prime}Z^*\right)^{-1/2}n^{-1/2}Z'\varepsilon\sigma_{\varepsilon\varepsilon}^{*-1/2}.
\end{aligned}$$

Note that $E^*\left[n^{-1}Z^{*\prime}Z^*\right]=n^{-1}Z'Z$ and $n^{-1}Z^{*\prime}Z^*-n^{-1}Z'Z\to^{P^*}0$ in probability, by the Law of Large Numbers. Moreover, $n^{-1}Z'Z\xrightarrow{p}Q_{ZZ}$ which is positive definite, therefore we obtain $\left(n^{-1}Z^{*\prime}Z^*\right)^{-1}\to^{P^*}Q_{ZZ}^{-1}$ in probability. Then, by using the similar arguments as in the proof of Theorem 4.3 in Wang and Doko Tchatoka (2018), we obtain conditional convergence in distribution under weak IVs:

$$W^{*\prime}P_{Z^*}W^*\sigma_{v_wv_w}^{*-1}\to^{d^*}\left(\Psi_h+\psi_{v_w}^B\right)'\left(\Psi_h+\psi_{v_w}^B\right), \tag{A.1}$$

in probability, where $\Psi_h=h_{ww}+\psi_{v_w}$. Similarly, we have

$$W^{*\prime}P_{Z^*}\varepsilon^*(\sigma_{\varepsilon\varepsilon}^*\sigma_{v_wv_w}^*)^{-1/2}\to^{d^*}\left(\Psi_h+\psi_{v_w}^B\right)'\left(\psi_\varepsilon+\psi_\varepsilon^B\right), \tag{A.2}$$

in probability.

Second, note that $\tilde{\kappa}^*$ is the smallest root of $\left|\kappa\hat{\Omega}_w^*-(\tilde{y}^*(\beta_0):W^*)'P_{Z^*}(\tilde{y}^*(\beta_0):W^*)\right|=0$, where $\hat{\Omega}_w^*=\frac{1}{n-L}(\tilde{y}^*(\beta_0):W^*)'M_{Z^*}(\tilde{y}^*(\beta_0):W^*)$. And this is equivalent to

$$\left|\kappa\hat{\Sigma}^*-(\varepsilon^*:Z^*\Pi_w+V_w^*)'P_{Z^*}(\varepsilon^*:Z^*\Pi_w+V_w^*)\right|=0, \tag{A.3}$$

where $\hat{\Sigma}^*=\frac{1}{n-L}(\varepsilon^*:W^*)'M_{Z^*}(\varepsilon^*:W^*)$. Then, by combining eqs (A.1)-(A.3), we obtain:

$$\begin{aligned}
\tilde{\gamma}^*-\gamma\;&\to^{d^*}\;\sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{v_wv_w}^{-1/2}\left\{\left(\Psi_h+\psi_{v_w}^B\right)'\left(\Psi_h+\psi_{v_w}^B\right)-\kappa_h^B\right\}^{-1}\left\{\left(\Psi_h+\psi_{v_w}^B\right)'\left(\psi_\varepsilon+\psi_\varepsilon^B\right)-\kappa_h^Bh_{w\varepsilon}\right\}\\
&=\;\sigma_{\varepsilon\varepsilon}^{1/2}\sigma_{v_wv_w}^{-1/2}\Delta_h^B\text{ in probability,}
\end{aligned}$$

where κ_h^B is the smallest root of $\left|\kappa_h \Sigma_h - \left(\psi_\varepsilon + \psi_\varepsilon^B : \Psi_h + \psi_{v_w}^B\right)' \left(\psi_\varepsilon + \psi_\varepsilon^B : \Psi_h + \psi_{v_w}^B\right)\right| = 0$.

For the denominator of the subvector AR statistic, we have the following decomposition:

$$
\begin{aligned}
& \frac{1}{n-L}\left(\tilde{y}^*(\beta_0) - W^*\tilde{\gamma}^*\right)' M_{Z^*}\left(\tilde{y}^*(\beta_0) - W^*\tilde{\gamma}^*\right) \\
= \; & (n-L)^{-1}\varepsilon^{*'}\varepsilon^* - 2(n-L)^{-1}\varepsilon^{*'}M_{Z^*}W^*(\tilde{\gamma}^* - \gamma) + (n-L)^{-1}W^{*'}M_{Z^*}W^*(\tilde{\gamma}^* - \gamma)^2.
\end{aligned}
$$

Then, by using similar arguments as those for $\tilde{\gamma}^*$, we have

$$
\frac{1}{n-L}\left(\tilde{y}^*(\beta_0) - W^*\tilde{\gamma}^*\right)' M_{Z^*}\left(\tilde{y}^*(\beta_0) - W^*\tilde{\gamma}^*\right) \to^{d^*} \sigma_{\varepsilon\varepsilon}\left(1 - 2h_{w\varepsilon}\Delta_h^B + (\Delta_h^B)^2\right), \qquad \text{(A.4)}
$$

in probability. For the numerator of the subvector AR statistic, we note that

$$
\begin{aligned}
& \left(Z^{*'}Z^*\right)^{-1/2} Z^{*'}\left(\tilde{y}^*(\beta_0) - W^*\tilde{\gamma}^*\right) \\
= \; & \sigma_{\varepsilon\varepsilon}^{*1/2}\Big\{(n^{-1}Z^{*'}Z^*)^{-1/2} n^{-1/2} Z^{*'}\varepsilon^*\sigma_{\varepsilon\varepsilon}^{*-1/2} \\
& + \left[\left(n^{-1}Z^{*'}Z^*\right)^{-1/2} n^{-1/2} Z^{*'}W^*\sigma_{v_w v_w}^{*-1/2}\right]\sigma_{\varepsilon\varepsilon}^{*-1/2}\sigma_{v_w v_w}^{*1/2}(\gamma - \tilde{\gamma}^*)\Big\}.
\end{aligned}
$$

Given the previous results, it is clear that

$$
\left(Z^{*'}Z^*\right)^{-1/2} Z^{*'}\left(\tilde{y}^*(\beta_0) - W^*\tilde{\gamma}^*\right) \to^{d^*} \sigma_{\varepsilon\varepsilon}^{1/2}\left\{(\psi_\varepsilon + \psi_\varepsilon^B) - (\Psi_h + \psi_{v_w}^B)\Delta_h^B\right\},
$$

in probability. The desired result follows. □

PROOF OF THEOREM **3.2**

For the recentering bootstrap, we note that

$$
\begin{aligned}
& \left(Z^{*'}Z^*\right)^{-1/2}\left(Z^{*'}\tilde{\varepsilon}^*(\beta_0, \tilde{\gamma}^*) - Z'\tilde{\varepsilon}(\beta_0, \tilde{\gamma})\right) \\
= \; & \sigma_{\varepsilon\varepsilon}^{*1/2}\Big\{(n^{-1}Z^{*'}Z^*)^{-1/2} n^{-1/2}\left[(Z^{*'}\varepsilon^* - Z'\varepsilon) + (Z'\varepsilon - Z'\tilde{\varepsilon}(\beta_0, \tilde{\gamma}))\right]\sigma_{\varepsilon\varepsilon}^{*-1/2} \\
& + \left[\left(n^{-1}Z^{*'}Z^*\right)^{-1/2} n^{-1/2} Z^{*'}W^*\sigma_{v_w v_w}^{*-1/2}\right]\sigma_{\varepsilon\varepsilon}^{*-1/2}\sigma_{v_w v_w}^{*1/2}(\gamma - \tilde{\gamma}^*)\Big\}.
\end{aligned}
$$

In addition, $n^{-1/2}\left(Z'\varepsilon - Z'\tilde{\varepsilon}(\beta_0, \tilde{\gamma})\right) = n^{-1/2}Z'W(\tilde{\gamma} - \gamma)$. Then, it is clear that

$$
\left(Z^{*'}Z^*\right)^{-1/2}\left(Z^{*'}\tilde{\varepsilon}^*(\beta_0, \tilde{\gamma}^*) - Z'\tilde{\varepsilon}(\beta_0, \tilde{\gamma})\right) \to^{d^*} \sigma_{\varepsilon\varepsilon}^{1/2}\left\{(\Psi_h\Delta_h + \psi_\varepsilon^B) - (\Psi_h + \psi_{v_w}^B)\Delta_h^B\right\},
$$

in probability. The desired result follows. □

Define the localization parameter space:

$$H = \Big\{ h = (h_{ww}, h_{w\varepsilon}) : \exists \{\theta_n = (\gamma_n, \Pi_{n,x}, \Pi_{n,w}, F_n) \in \Theta : n \geq 1\} \text{ such that}$$
$$n^{1/2} Q_{n,ZZ}^{1/2} \Pi_{n,w} \sigma_{n,v_w v_w}^{-1/2} \to h_{ww} \in [-\infty, +\infty]^L \text{ and } \sigma_{n,v_w v_w}^{-1/2} \sigma_{n,v_w \varepsilon} \sigma_{n,\varepsilon\varepsilon}^{-1/2} \to h_{w\varepsilon} \in [-1, 1] \Big\},$$
$$\text{where } Q_{n,ZZ} = E_{F_n}\left(Z_i Z_i'\right), \sigma_{n,v_w\varepsilon} = E_{F_n}\left(V_{w,i}\varepsilon_i\right), \sigma_{n,v_w v_w} = E_{F_n}\left(V_{w,i}^2\right), \text{ and } \sigma_{n,\varepsilon\varepsilon} = E_{F_n}\left(\varepsilon_i^2\right). \tag{A.5}$$

PROOF OF THEOREM **3.3**

We follow Andrews and Guggenberger (2010) [e.g., the proof of Theorem 1] and note that there exists a "worst case sequence" $\theta_n = (\gamma_n, \Pi_{x,n}, \Pi_{w,n}, F_n) \in \Theta$ such that:

$$\begin{aligned} & AsySz[\hat{c}_{n,j}(1-\alpha)] \\ = & \limsup_{n\to\infty} \sup_{\theta \in \Theta} P_\theta[AR_n(\beta_0, \tilde{\gamma}) > \hat{c}_{n,j}(1-\alpha)] \\ = & \limsup_{n\to\infty} P_{\theta_n}[AR_n(\beta_0, \tilde{\gamma}) > \hat{c}_{n,j}(1-\alpha)] \\ = & \lim_{n\to\infty} P_{\theta_{m_n}}[AR_{m_n}(\beta_0, \tilde{\gamma}) > \hat{c}_{m_n,j}(1-\alpha)], \end{aligned} \tag{A.6}$$

where the first equality in (A.6) holds by the definition of asymptotic size and the second equality holds by the choice of the sequence $\{\theta_n : n \geq 1\}$. And $\{m_n : n \geq 1\}$ is a subsequence of $\{n : n \geq 1\}$; such a subsequence always exists. Furthermore, there exists a subsequence $\{\omega_n : n \geq 1\}$ of $\{m_n : n \geq 1\}$ such that:

$$\begin{aligned} & \lim_{n\to\infty} P_{\theta_{m_n}}[AR_{m_n}(\beta_0, \tilde{\gamma}) > \hat{c}_{m_n,j}(1-\alpha)] \\ = & \lim_{n\to\infty} P_{\theta_{\omega_n,h}}[AR_{\omega_n}(\beta_0, \tilde{\gamma}) > \hat{c}_{\omega_n,j}(1-\alpha)] \end{aligned} \tag{A.7}$$

for some $h \in H$. But, for any $h \in H$, any subsequence $\{\omega_n : n \geq 1\}$ of $\{n : n \geq 1\}$, and any sequence $\{\theta_{\omega_n,h} : n \geq 1\}$, we have $\left(AR_{\omega_n}(\beta_0, \tilde{\gamma}), \hat{c}_{\omega_n,j}(1-\alpha)\right) \xrightarrow{d} \left(\xi_h, \tilde{c}_{h,j}(1-\alpha)\right)$ jointly. It follows that $AsySz[\hat{c}_{n,j}(1-\alpha)] = sup_{h\in H} P[\xi_h > \tilde{c}_{h,j}(1-\alpha)]$.

□