KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.810

"Bootstrapping Anderson-Rubin Statistic and J Statistic in Linear IV Models with Many Instruments"

Wenjie Wang

February 2012

KYOTO UNIVERSITY

KYOTO, JAPAN

Bootstrapping Anderson-Rubin Statistic and J Statistic in Linear IV Models with Many Instruments

Wenjie Wang
Graduate School of Economics, Kyoto University

February, 2012

Abstract

A bootstrap method is proposed for the Anderson-Rubin test and the J test for overidentifying restrictions in linear instrumental variable models with many instruments. We show the bootstrap validity of these test statistics when the number of instruments increases at the same rate as the sample size. Moreover, since it has been shown in the literature to be valid when the number of instruments is small, the bootstrap technique is practically robust to the numerosity of the moment conditions. A small-scale Monte Carlo experiment shows that our procedure has outstanding small sample performance compared with some existing asymptotic procedures.

1 Introduction

The conventional asymptotic theory often provide a poor approximation of the finite sample distribution of instrumental variable estimators and test statistics. Examples are with weak instruments(e.g., Staiger and Stock(1997)[17], Stock and Wright(2000)[18]) or many instruments(e.g., Morimune(1983)[14], Bekker(1994)[4], Chao and Swanson(2005)[5], Andrews and Stock(2007)[3], Hansen et al.,(2008)[10], Anatolyev and Gospodinov(2011)[1]). And these problems have recently received considerable attention in the econometric literature.

However, despite the large literature on estimation in the presence of many (and possibly weak) instruments, the behavior of the tests for parameter and overidentifying restrictions has not been fully investigated. Andrews and Stock(2007)[3] derives the asymptotic distributions of some parameter and specification tests in models with moderately many instruments. But in their paper, the number of instruments grows much more slowly relative to the sample size. In contrast, Anatolyev and Gospodinov(2011)[1] argue that to obtain a good asymptotic approximation, one has to acknowledge the numerosity of instruments via a many instruments assumption of Bekker(1994)[4]. They also propose a modification of the Anderson-Rubin statistic and J statistic so that these tests can be robust to many instruments. However, the empirical size distortion of these modified tests tends to increase when the number of instruments becomes a nontrivial fraction of the sample size.

Instead of doing modification of the AR and J tests, we propose in this paper to bootstrap them directly and show the bootstrap validity of these test statistics under many instruments asymptotics of Bekker(1994)[4]. Furthermore, this bootstrap procedure is easy to implement in practice because it does not require an a priori choice of asymptotic framework, i.e., it is valid under both fixed and many instruments asymptotics. Monte Carlo simulations show that the bootstrap techniques provide a more reliable method to approximate the null distribution of the test statistics.

The remainder of the paper is organized as follows. Section 2 introduces the basic framework and the test statistics. The main results are established and discussed in Section 3. Section 4 presents Monte Carlo simulation results for the size properties of our bootstrap procedure in finite samples. Section 5 concludes. All proofs are relegated to the Appendix.

2 Model, Assumption and Statistics

We consider a standard linear instrumental variable regression given by

$$y_i = X_i'\beta + \epsilon_i$$

for $i = 1, ..., n$, where y_i is the scalar outcome variable and X_i is the $k \times 1$ vector of regressors that is possibly correlated with the unobservable error term ϵ_i. Let Z_i be a $l \times 1$ vector of instruments, which we treat as deterministic, where $k \leq l < n$. We also let $P_Z = Z(Z'Z)^{-1}Z'$ and $M_Z = I_n - P_Z$, where I_n is an identity matrix with dimension n. We further assume that

$$X_i = \pi' Z_i + v_i$$

where π is the $l \times k$ matrix of parameters whose value may depend on n as well as l. The model can be written in matrix form as

$$y = X\beta + \epsilon \quad (1)$$

$$X = Z\pi + v \quad (2)$$

where $y = (y_1, ..., y_n)'$ is $n \times 1$, $X = (x_1, ..., x_n)'$ is $n \times k$, $Z = (z_1, ..., z_n)'$ is $n \times l$, $\epsilon = (\epsilon_1, ..., \epsilon_n)'$ is $n \times 1$ and $v = (v_1, ..., v_n)'$ is $n \times k$. In this paper, we consider the case when k, the dimension of β, is small relative to n, but l is large and comparable to n.

The model and the data are assumed to satisfy the following conditions.

Assumption 1. The errors $\eta_i = (\epsilon_i, v_i')'$ are i.i.d. for $i = 1, ..., n$ with mean zero and positive definite variance matrix $\Sigma = \begin{pmatrix} \sigma_{\epsilon\epsilon} & \sigma_{v\epsilon}' \\ \sigma_{v\epsilon} & \Sigma_{vv} \end{pmatrix}$. ϵ_i and v_i have finite fourth moments.

Assumption 2. As $n \to \infty$, $\lambda_n = l/n \to \lambda$, where $0 < \lambda < 1$.

Assumption 2 adopts the many instruments asymptotic framework of Bekker(1994)[4] when the number of instruments is a nontrivial fraction of the sample size.

Following Anatolyev and Gospodinov(2011)[1], we also assume the following condition for the instruments.

Assumption 3. Under the asymptotics of Assumption 2, $n^{-1} \sum_{i=1}^{n} |z_i'(Z'Z)^{-1}z_i - \lambda| \to 0$.

As discussed in their paper, Assumption 3 requires that (almost) all diagonal elements of the projection matrix P_Z converge to λ (under standard or moderately many instruments asymptotics they converge to zero). And the validity of Assumption 3 follows from the literature on large dimensional covariance matrices(Silverstein(1995)[16]) in the case that Z are i.i.d. both across rows and columns, possibly after a rotating transformation, and have finite fourth moments. The i.i.d. requirement can be relaxed at the expense of existence of higher order moments.

Assumption 4. X_i has finite fourth moment.

Assumption 5. $\pi' Z' Z\pi/n \to Q$, where Q is a positive definite matrix.

Assumption 5 implies that the information accumulation by adding new instruments is limited and thus bounded even with $l \to \infty$. Note that this condition allows for moderately weak instruments though not as weak as the case considered by Chao and Swanson(2005)[5] or Hansen et al.(2008)[10].

We proceed to introduce the test statistics to be studied in this paper. In the literaure, statistics to test the hypothese $H_0 : \beta = \beta_0$ have been developped whose limiting distribution under H_0 does not depend on the value of π, see e.g. Anderson and Rubin(1949)[2], Kleibergen(2002)[11] and Moreira(2003)[12]. Therefore, these test statistics are able to provide correct asymptotic size no matter the instrument Z is strong or weak. However, these test statistics are not robust to many instruments. For example, the Anderson-Rubin(AR) test statistic takes the following form

$$AR = (n-l)\frac{\epsilon_0' P_Z \epsilon_0}{\epsilon_0' M_Z \epsilon_0} \tag{3}$$

where $\epsilon_0 = y - X\beta_0$ is a vector of null restricted error. Under $H_0 : \beta = \beta_0$ and conventional fixed instruments asymptotics(when l does not grow with the sample size), its limiting distribution is $\chi^2(l)$. However, the limiting distribution changes when the number of instruments increases with the sample size. Particularly, in the framework of moderately many instruments(when $l^3/n \to 0$ as $l, n \to \infty$), Andrews and Stock(2007)[3] show that

$$\sqrt{l}\left(\frac{AR}{l} - 1\right) \Rightarrow_{d_P} N(0,2)$$

Moreover, Anatolyev and Gospodinov(2011)[1] show that under the many instruments assumption of Bekker(1994)[4].

$$\sqrt{l}\left(\frac{AR}{l} - 1\right) \Rightarrow_{d_P} N(0, 2/(1-\lambda)) \tag{4}$$

and they proposed a modified test statistic based on equation (4). However, as can be found out in the results of their simulation and also in ours, this asymptotic approximation can become less reliable when l becomes an important fraction of the sample size.

We then turn to the J statistic. The standard J test statistic (Sargan, 1958) is defined as

$$J = \frac{\hat{\epsilon}' P_Z \hat{\epsilon}}{\hat{\sigma}_{\epsilon\epsilon}} \tag{5}$$

where $\hat{\epsilon} = y - X\hat{\beta}$ is the residual vector and $\hat{\sigma}_{\epsilon\epsilon} = \hat{\epsilon}'\hat{\epsilon}/n$. Let $\hat{\beta}_{2SLS} = (X' P_Z X)^{-1} X' P_Z y$ denote the two-stage least square (2SLS) estimator. It is well known that under the null of correct moment restriction $H_m : E[\epsilon_i Z_i] = 0$, the standard asymptotic theory gives $\hat{\beta}_{2SLS} - \beta = O_P(1/\sqrt{n})$. And for the J statistic based on $\hat{\beta}_{2SLS}$, we have

$$J \Rightarrow_{d_P} \chi^2(l-k) \tag{6}$$

as $n \to \infty$. When $l \to \infty$, however, $\hat{\beta}_{2SLS}$ is no longer a consistent estimator of β. The right hand side of (6) diverges and the asymptotic distribution of J is not well-defined.

In order to solve this problem, Anatolyev and Gospodinov(2011)[1] have proposed modified versions of the J testd statistic. The modified J tests are constructed such that under many instruments asymptotics, their asymptotic distributions under the null hypothesis are standard normal. Moreover, It turns out that the choice of $\hat{\beta}$ is not important for the asymptotic behavior of these modified test statistics as long as its rate of convergence is not slower than $\sqrt{n}$ under many instruments. More precisely, Anatolyev and Gospodinov(2001)[1] show that as long as the $\hat{\beta}$ satisfies $\sqrt{n}(\hat{\beta} - \beta) = O_P(1)$ under many instruments asymptotics, we can obtain

$$\sqrt{l}\left(\frac{J}{l} - 1\right) \Rightarrow_{d_P} N(0, 2(1-\lambda)) \tag{7}$$

However, similar to the AR test, the simulation results in Anatolyev and Gospodinov(2011)[1] show that the modified J test statistic can have serious size distortion when the number of instruments becomes an important fraction of the sample size(e.g., l/n=0.5). Note that similar ratios of number of moment conditions to sample size often arise in empirical applications such as linear asset pricing models of large portfolios and estimating structual macroeconomic models by matching impulse response functions. Therefore, we think it necessary to propose a many-instruments robust inference approach that can deliver more reliable finite sample performance.

3 Bootstrap Validity under Many Instruments Asymptotics

Instead of modifying the test statistics, we propose to bootstrap the AR and J statistics even when the number of instruments is large, and we show the bootstrap validity under many instruments asymptotics of Bekker(1994) [4] . In this paper, we shall consider the residual based i.i.d. bootstrap, and our bootstrap procedure for the Anderson-Rubin test statistic is carried out as follows:

Step 1: Given $\hat{\beta}$, consistent estimator of β, the residuals from the equation (1) are obtained as:

$$\hat{\epsilon} = y - X\hat{\beta}$$

As to the choice of $\hat{\beta}$, estimators that are consistent under both conventional fixed instruments asymptotics and many instruments asymptotics can be used. (eg., LIML estimator or Bias-corrected two stage least square estimator, see Newey(2004)[15] for more discussions.) Particularly, the Bias-corrected two stage least square estimator($\hat{\beta}_{B2SLS}$, say) takes the following form:

$$\hat{\beta}_{B2SLS} = \left(X^{'}P_Z X - \lambda_n X^{'}X\right)^{-1}\left(X^{'}P_Z y - \lambda_n X^{'}y\right) \tag{8}$$

where $\lambda_n = l/n$. We will use this Bias-corrected two stage least square estimator in both mathematical proof and numerical experiments below. But one can extend the results easily to the case of LIML estimator.

Step 2: The residuals are re-centered to yield $\tilde{\epsilon}$, then ϵ^* are drawn from the empirical distribution function of $\tilde{\epsilon}$.

Step 3: Our bootstrap version of Anderson-Rubin statistic takes the following form:

$$AR^* = (n-l)\frac{\epsilon^{*'}P_Z\epsilon^*}{\epsilon^{*'}M_Z\epsilon^*} \tag{9}$$

Step 4: Repeat Steps 1-3 B times, and obtain the empirical distribution of the B test statistics of AR^*. This empirical distribution is used to approximate the finite sample distribution of AR under H_0.

Step 5: Let AR^*_α be the α percentile of the bootstrap distribution from Step 4. We will reject the null hypothesis at significance level α if the observed $AR > AR^*_\alpha$.

Note that for the AR test, there is no need to generate the bootstrap resample ($\{X^*, y^*\}$, say). To see why, suppose we set

$$y^* \quad = \quad X^*\beta_0 + \epsilon^*$$

where X^* is generated from some resampling scheme for X, and we generate y^* under $H_0 : \beta = \beta_0$. But the resulting bootstrap version of the AR statistic will take exactly the same form as equation (9).

The following theorem shows the bootstrap validity of AR statistic under many instruments asymptotics.

THEOREM 1. Suppose Assumptions 1-5 holds. Then,

$$sup_{x\in R}\left|P^*\left(\sqrt{l}\left(\frac{AR^*}{l}-1\right)\leq x\right) - P\left(\sqrt{l}\left(\frac{AR}{l}-1\right)\leq x\right)\right| \to_p 0$$

where P^* denotes the probability measure induced by the i.i.d. bootstrap.

Proof. See the Appendix.

In the literature, several authors have considered improving the finite sample performance of the AR statistic and other identification robust statistics by using bootstrap technique(e.g., Davidson and Mackinnon(2008)[6], Moreira, Porter and Suarez(2009)[13]).

And they show that it is valid to bootstrap the AR test under the conventional fixed instruments asymptotics; Theorem 1 extends their result to many instruments case.

In practice, it is difficult to decide when we should use the fixed instruments asymptotics and when we should use the many instruments asymptotics. However, this decision is not necessary for the bootstrap approach since the actual procedure will be the same in both cases. For example, suppose we generate B=99 times bootstrap samples, and construct

$$\left\{\sqrt{l}\left(\frac{AR_1^*}{l}-1\right),...,\sqrt{l}\left(\frac{AR_{99}^*}{l}-1\right)\right\}$$

as their bootstrap validity has been shown in Theorem 1. Sorting all 100 statistics,

$$\left\{\sqrt{l}\left(\frac{AR}{l}-1\right),\sqrt{l}\left(\frac{AR_1^*}{l}-1\right),...,\sqrt{l}\left(\frac{AR_{99}^*}{l}-1\right)\right\}$$

for a 5 percent nominal level test, we shall reject $H_0:\beta=\beta_0$ if $\sqrt{l}\left(\frac{AR}{l}-1\right)$ is among the 5 largest ones. But this is equivalent to reject H_0 if AR is one of the 5 largest statistics in $\{AR, AR_1^*, ..., AR_{99}^*\}$, which is exactly the same the procedure as used in Davidson and MacKinnon(2008)[6] and Moreira, Porter and Suarez(2009)[13] for the fixed instruments case. Therefore, the bootstrap technique is robust to the numerosity of the instruments in the sense that it is valid for both few and many instruments.

The bootstrap procedure for the J test is a little bit more complicated than that of the AR test because we will use both equation (1) and equation (2) to generate the residuals.

Step 1: The residuals are obtained as:

$$\hat{\epsilon}=y-X\hat{\beta}$$
$$\hat{v}=X-Z\hat{\pi}$$

where $\hat{\pi}=(Z'Z)^{-1}Z'X$.

Step 2: The residuals are re-centered to yield $\{\tilde{\epsilon},\tilde{v}\}$, then $\{\epsilon^*,v^*\}$ are drawn from the empirical distribution function of $\{\tilde{\epsilon},\tilde{v}\}$.

Step 3: Next, we set

$$\begin{aligned} y^* &= X^*\hat{\beta}+\epsilon^* \\ X^* &= Z\hat{\pi}+v^* \end{aligned}$$

Step 4: Obtain the bootstrap residual $\hat{\epsilon}^*=y^*-X^*\hat{\beta}^*$, where $\hat{\beta}^*$ is the estimate of $\hat{\beta}$ using the bootstrap sample $\{X^*,y^*\}$. Then use the bootstrap residual $\hat{\epsilon}^*$ to construct the test statistic

$$J^*=\frac{\hat{\epsilon}^{*\prime}P_Z\hat{\epsilon}^*}{\hat{\sigma}_{\epsilon\epsilon}^*}$$

where $\hat{\sigma}_{\epsilon\epsilon}^*=\hat{\epsilon}^{*\prime}\hat{\epsilon}^*/n$.

Step 5: Repeat Steps 1-4 B times, and obtain the empirical distribution of the B test statistics of J^*. This empirical distribution is used to approximate the finite sample distribution of J under the null hypothesis.

Step 6: Let J^*_α be the α percentile of the bootstrap distribution from Step 5. We will reject the null hypothesis at significance level α if the observed $J > J^*_\alpha$.

And the bootstrap validity of J statistic under many instruments asymptotics is shown in Theorem 2.

THEOREM 2. Suppose Assumptions 1-5 holds. Then,

$$sup_{x\in R}\left|P^*\left(\sqrt{l}\left(\frac{J^*}{l}-1\right)\le x\right)-P\left(\sqrt{l}\left(\frac{J}{l}-1\right)\le x\right)\right|\to_p 0$$

where P^* denotes the probability measure induced by the i.i.d. bootstrap.

4 Monte Carlo Simulation

To evaluate the finite-sample performance of the proposed bootstrap procedure, we conduct some Monte Carlo experiments. The design of the experiment is similar to that considered by Hahn and Hausman(2002)[9] and Donald and Newey(2001) [7].

The simulation model is described by Eqs.(1) and (2). The n rows of $[\epsilon, v]$ are i.i.d. with mean zero, unit variance and correlation ρ. The correlation coefficient ρ represents the degree of endogeneity of X, and we consider in the simulation $\rho = 0, 0.2, 0.5$ and 0.8. We take the matrix of instruments, Z, to be distributed $N(0, I_l)$. And we consider different strengths of identification by introducing $R_f^2 = 0.01$ and 0.2 for the theoretical R^2 of the first-stage regression. $R_f^2 = 0.01$ reflects relatively weak instruments whereas $R_f^2 = 0.2$ reflects relatively strong instruments. We consider the sample size $n = 100$, and the number of instruments $l = 10, 30$ and 50. For all the results, the number of replication is 5000, and we generate $B = 399$ bootstrap resamples.

Table 1 and Table 2 report the empirical rejection frequency at 5 percent nominal level of the conventional(AR), the modified version(AR_{AG}) by Anatolyev and Gospodinov(2011)[1] and the bootstrap version($AR_{Bootstrap}$) of the AR test. Examining the results, we can find that the conventional AR test seriously over-reject when the number of instruments is large. The AR_{AG} test performs better than the AR test, but it tends to over-reject for $l = 50$. Our $AR_{Bootstrap}$ test has coverage very close to the nominal level for all values of ρ, l and R_f^2.

Table 3 and Table 4 report the empirical rejection frequency at 5 percent nominal level of the conventional(J), the modified version(J_{AG}) by Anatolyev and Gospodinov(2011)[1] and the bootstrap version($J_{Bootstrap}$) of the J test. Interestingly, different from the AR test, the conventional J test tends to seriously under-reject when the number of instruments is large. While the empirical size distortion of the J_{AG} test is much smaller than that of the J test, the overall performance of the $J_{Bootstrap}$ test turns out to be the best among the three.

5 Conclusions and Future Research

To summarize, we propose in this paper to bootstrap the standard AR and J tests of parameter and overidentifying restrictions in the presence of many instruments. The bootstrap validity is shown under the many instruments asymptotics of Bekker(1994)[4]. A small-scale Monte Carlo experiment shows that our bootstrap procedure has outstanding small sample performance compared with some existing asymptotic procedures. Furthermore, in practice, this bootstrap procedure can be implemented no matter the number of instruments is small or large. A currently undertaken extension is to show the bootstrap validity of other identification robust statistics(e.g., Kleibergen(2002)[11]'s K statistic and Moreira(2003)[12]'s CLR statistic) in the presence of many instruments. Another research topic that may be interesting is to consider non-i.i.d environments that accommodate heteroskedasticity and serial correlation in the error terms.

References

[1] Stanislav Anatolyev and Nikolay Gospodinov. Specification testing in models with many instruments. *Econometric Theory*, 27(02):427–441, April 2011.

[2] T. W. Anderson and Herman Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, 20(1):46–63, 1949.

[3] Donald W.K. Andrews and James H. Stock. Testing with many weak instruments. *Journal of Econometrics*, 138(1):24–46, May 2007.

[4] Paul A Bekker. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3):657–81, May 1994.

[5] John C. Chao and Norman R. Swanson. Consistent estimation with a large number of weak instruments. *Econometrica*, 73(5):1673–1692, 09 2005.

[6] Russell Davidson and James G. MacKinnon. Bootstrap inference in a linear equation estimated by instrumental variables. *Econometrics Journal*, 11(3):443–477, November 2008.

[7] Stephen G Donald and Whitney K Newey. Choosing the number of instruments. *Econometrica*, 69(5):1161–91, September 2001.

[8] Harry H. Kelejian and Ingmar R. Prucha. On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics*, 104(2):219–257, September 2001.

[9] Jinyong Hahn and Jerry Hausman. A new specification test for the validity of instrumental variables. *Econometrica*, 70(1):163–189, January 2002.

[10] Christian Hansen, Jerry Hausman, and Whitney Newey. Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, 26:398–422, 2008.

[11] Frank Kleibergen. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803, September 2002.

[12] Marcelo J. Moreira. A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048, 07 2003.

[13] Marcelo J. Moreira, Jack R. Porter, and Gustavo A. Suarez. Bootstrap validity for the score test when instruments may be weak. *Journal of Econometrics*, 149(1):52–64, April 2009.

[14] Kimio Morimune. Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica*, 51(3):821–41, May 1983.

[15] W.K. Newey. Many instruments asymptotics. *Manuscript, MIT*, 2004.

[16] J. W. Silverstein. Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. *Journal of Multivariate Analysis*, 55(2):331–339, November 1995.

[17] Douglas Staiger and James H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, May 1997.

[18] James H. Stock and Jonathan Wright. Gmm with weak identification. *Econometrica*, 68(5):1055–1096, September 2000.

6 APPENDIX

Throughout this Appendix, for any bootstrap statistic T^* we write $T^* \to_{P^*} 0$ in probability when $lim_{n\to\infty} P[P^*(|T^*| > \delta) > \delta] = 0$ for any $\delta > 0$, i.e. $P^*(|T^*| > \delta) = o_P(1)$. Also, we say that $T^* = O_{P^*}(n^\lambda)$ in probability if and only if $\forall\delta > 0$, There exists a $M_\delta < \infty$ such that $lim_{n\to\infty} P[P^*(|n^{-\lambda}T^*| > M_\delta) > \delta] = 0$, i.e. $\forall\delta > 0$, There exists a $M_\delta < \infty$ such that $P^*(|n^{-\lambda}T^*| > M_\delta) = o_P(1)$. Finally, we write $T^* \Rightarrow_{d_{P^*}} D$ in probability, for any distribution D, when weak convergence under the bootstrap probability measure occurs in a set with probability converging to one.

Following Anatolyev and Gospodinov(2011)[1], we also use throughout the proof that $0 \le z_i'(Z'Z)^{-1}z_i \le 1$ for each i, that

$$\frac{1}{n}\sum_{i=1}^{n} z_i'(Z'Z)^{-1}z_i = \frac{1}{n}Tr\left((Z'Z)^{-1}\sum_{i=1}^{n} z_i z_i'\right) = \frac{1}{n}Tr(I_l) = \lambda_n$$

and that

$$\frac{1}{n}\sum_{i=1}^{n}\left(z_i'(Z'Z)^{-1}z_i - \lambda_n\right)^2 \le \frac{1}{n}\sum_{i=1}^{n}|z_i'(Z'Z)^{-1}z_i - \lambda_n|$$

because $0 \le z_i'(Z'Z)^{-1}z_i \le 1$ and $u^2 \le |u|$ when $0 \le u \le 1$.

Lemma 1 Suppose $\hat{\beta} - \beta \to_P 0$, $E[\epsilon_i^4] < \infty$, $E \parallel X_i \parallel^4 < \infty$, then $E^*(\epsilon_i^{*4}) \equiv \tilde{\kappa}_4$ and $Var^*[\epsilon_i^*] \equiv \tilde{\sigma}_{\epsilon\epsilon}$ are bounded in probability.

Proof.

(a) Let us start with $\tilde{\kappa}_4$. Let $\bar{\epsilon} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Using Minkowski and Cauchy-Schwartz inequalities, we obtain

$$\begin{aligned}
\tilde{\kappa}_4 &= \frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_i^4 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\epsilon_i - \bar{\epsilon} - (X_i - \bar{X})'(\hat{\beta} - \beta)\right)^4 \\
&\le C_1\left\{\frac{1}{n}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^4 + \frac{1}{n}\sum_{i=1}^{n}|(X_i - \bar{X})'(\hat{\beta} - \beta)|^4\right\} \\
&\le C_2\left\{\frac{1}{n}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^4 + \parallel \hat{\beta} - \beta \parallel^4 \frac{1}{n}\sum_{i=1}^{n}\parallel X_i - \bar{X} \parallel^4\right\}
\end{aligned}$$

for large enough constants C_1 and C_2.

Using the Minkowski inequality again, we get

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\parallel X_i - \bar{X} \parallel^4 &\le C_1\left\{\frac{1}{n}\sum_{i=1}^{n}\parallel X_i \parallel^4 + \parallel \bar{X} \parallel^4\right\} \\
&\to_P C_1\left\{E \parallel X_i \parallel^4 + \parallel E[X_i] \parallel^4\right\}
\end{aligned}$$

using $\| \bar{X} \| \to_P \| E[X_i] \| \leq E \| X_i \| \leq (E \| X_i \|^4)^{1/4}$ by Jensen's inequality. Since $\hat{\beta} - \beta \to_P 0$ and $E[\epsilon_i^4] < \infty$, the term $\tilde{\kappa}_4$ is bounded in probability.

(b) For $\tilde{\sigma}_{\epsilon\epsilon}$, note that by our bootstrap DGP, $E^*[\epsilon_i^*] = 0$. Therefore, to show $\tilde{\sigma}_{\epsilon\epsilon} = O_P(1)$, it suffices to show that $\frac{1}{n}\sum_{i=1}^n \tilde{\epsilon}_i^2 = O_P(1)$. But this follows when we apply the same arguments as for part (a). ■

Lemma 2 and Lemma 3 are for the bootstrap validity of J statistic. we shall first introduce some notations to be used in the Lemmas. By our i.i.d. bootstrap procedure, $E^*[v_i^* \epsilon_i^*] = \frac{1}{n}\sum_{i=1}^n \tilde{v}_i \tilde{\epsilon}_i \equiv \tilde{\sigma}_{v\epsilon}$ and $E^*[v_i^* v_i^{*'}] = \frac{1}{n}\sum_{i=1}^n \tilde{v}_i \tilde{v}_i' \equiv \tilde{\Sigma}_{vv}$ where $\tilde{\sigma}_{v\epsilon}$ is $k \times 1$ and $\tilde{\Sigma}_{vv}$ is $k \times k$.

Lemma 2 If Assumption 1-5 are satisfied, then the following statements are true as $n \to \infty$:

$$
\begin{aligned}
&(a) \quad v^{*'} P_Z \epsilon^*/l = \tilde{\sigma}_{v\epsilon} + O_{P^*}(1/\sqrt{l}), \text{in probability;}\\
&(b) \quad v^{*'} P_Z v^*/l = \tilde{\Sigma}_{vv} + O_{P^*}(1/\sqrt{l}), \text{in probability;}\\
&(c) \quad \epsilon^{*'} P_Z \epsilon^*/l = \tilde{\sigma}_{\epsilon\epsilon} + O_{P^*}(1/\sqrt{l}), \text{in probability;}\\
&(d) \quad \hat{\pi}' Z' v^*/n = O_{P^*}(1/\sqrt{n}), \text{in probability;}\\
&(e) \quad \hat{\pi}' Z' \epsilon^*/n = O_{P^*}(1/\sqrt{n}), \text{in probability.}
\end{aligned}
$$

Proof.

To prove part(a), note that it suffices to prove that $v^{*(g)'} P_Z \epsilon^*/l = \tilde{\sigma}_{v\epsilon}^{(g)} + O_{P^*}(1/\sqrt{l})$ as $n \to \infty$, where $v^{*(g)}$ denoted the gth column of v^*, so that $v^{*(g)'} P_Z \epsilon^*/l$ is the gth element of $v^{*'} P_Z \epsilon^*/l$, and where $\tilde{\sigma}_{v\epsilon}^{(g)}$ denotes the gth element of $\tilde{\sigma}_{v\epsilon}$, $g = 1, ..., k$.

From the bootstrap DGP, we can see that

$$
\begin{aligned}
E^*\left[\frac{v^{*(g)'} P_Z \epsilon^*}{l}\right] &= \frac{1}{l} E^*[Tr(v^{*(g)'} P_Z \epsilon^*)] = \frac{1}{l} Tr(P_Z E^*[\epsilon^* v^{*(g)'}])\\
&= \frac{\tilde{\sigma}_{v\epsilon}^{(g)}}{l} Tr(P_Z) = \tilde{\sigma}_{v\epsilon}^{(g)}
\end{aligned}
$$

because $E^*[\epsilon_i^* v_j^{*(g)}] = E^*[\epsilon_i^*] E^*[v_j^{*(g)}] = 0$ for $i \neq j$ by the property of i.i.d. bootstrap.

Furthermore, note that

$$
\begin{aligned}
& E^*\left[\frac{v^{*(g)\prime}P_Z\epsilon^*}{l}-\tilde{\sigma}_{v\epsilon}^{(g)}\right]^2 \\
= \; & \frac{1}{l^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n z_i'(Z'Z)^{-1}z_j z_k'(Z'Z)^{-1}z_l E^*[v_{ig}^*\epsilon_j^* v_{kg}^*\epsilon_l^*] \\
& -\frac{2\tilde{\sigma}_{v\epsilon}^{(g)}}{l}\sum_{i=1}^n\sum_{j=1}^n z_i'(Z'Z)^{-1}z_j E^*[v_{ig}^*\epsilon_j^*]+(\tilde{\sigma}_{v\epsilon}^{(g)})^2 \\
= \; & \frac{1}{l^2}E^*[v_{ig}^{*2}\epsilon_i^{*2}]\left[\sum_{i=1}^n(z_i'(Z'Z)^{-1}z_i)^2\right]+\frac{2}{l^2}\tilde{\Sigma}_{vv}^{(g,g)}\tilde{\sigma}_{\epsilon\epsilon}\left[\sum_{i=2}^n\sum_{j=1}^{i-1}(z_i'(Z'Z)^{-1}z_j)^2\right] \\
& +\left\{\frac{2}{l^2}(\tilde{\sigma}_{v\epsilon}^{(g)})^2\left[\sum_{i=2}^n\sum_{j=1}^{i-1}\left(z_i'(Z'Z)^{-1}z_i z_j'(Z'Z)^{-1}z_j+(z_i'(Z'Z)^{-1}z_j)^2\right)\right]-(\tilde{\sigma}_{v\epsilon}^{(g)})^2\right\} \\
\equiv \; & L_1+L_2+L_3
\end{aligned}
$$

The second equality follows from noting that $E^*[v_{ig}^*\epsilon_j^* v_{kg}^*\epsilon_l^*]$ equals zero except in the case where either $(i=j=k=l)$ or $(i=k, j=l)$ or $(i=j, k=l)$ or $(i=l, j=k)$ and from using $\sum_{i=1}^n z_i'(Z'Z)^{-1}z_i=l$.

Let us first focus on L_1. Note that

$$
\begin{aligned}
L_1 &= \frac{1}{l^2}\frac{1}{n}\sum_{i=1}^n\tilde{v}_{ig}^2\tilde{\epsilon}_i^2\left[\sum_{i=1}^n\left(z_i'(Z'Z)^{-1}z_i\right)^2\right] \\
&\leq \frac{1}{l^2}\left(\frac{1}{n}\sum_{i=1}^n\tilde{v}_{ig}^4\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n\tilde{\epsilon}_i^4\right)^{1/2}\left[\sum_{i=1}^n\left(z_i'(Z'Z)^{-1}z_i\right)^2\right] \\
&\leq \frac{1}{l}\left(\frac{1}{n}\sum_{i=1}^n\tilde{v}_{ig}^4\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n\tilde{\epsilon}_i^4\right)^{1/2}=O_P\left(\frac{1}{l}\right)
\end{aligned}
$$

where the first inequality follows from the Cauchy-Schwartz inequality, and the second inequality follows from using $\sum_{i=1}^n\left(z_i'(Z'Z)^{-1}z_i\right)^2\leq\sum_{i=1}^n z_i'(Z'Z)^{-1}z_i=l$. The last equality follows from using the same arguments as in Lemma 1.

Next, for L_2, we have

$$
\begin{aligned}
L_2 &\leq \frac{\tilde{\Sigma}_{vv}^{(g,g)}\tilde{\sigma}_{\epsilon\epsilon}}{l^2}\left[\sum_{i=1}^n\left(z_i'(Z'Z)^{-1}z_i\right)^2+2\sum_{i=2}^n\sum_{j=1}^{i-1}(z_i'(Z'Z)^{-1}z_j)^2\right] \\
&= \frac{\tilde{\Sigma}_{vv}^{(g,g)}\tilde{\sigma}_{\epsilon\epsilon}}{l}=O_P\left(\frac{1}{l}\right)
\end{aligned}
$$

because

$$
\sum_{i=1}^n\left(z_i'(Z'Z)^{-1}z_i\right)^2+2\sum_{i=2}^n\sum_{j=1}^{i-1}(z_i'(Z'Z)^{-1}z_j)^2=Tr(P_Z'P_Z)=Tr(P_Z)=l
$$

given that P_Z is symmetric and idempotent.

Finally, for L_3, we note that

$$\begin{aligned}
|L_3| &= \left| \frac{(\tilde{\sigma}_{v\epsilon}^{(g)})^2}{l^2} \left[(Tr(P_Z))^2 + Tr(P_Z' P_Z) - 2\sum_{i=1}^{n} (z_i'(Z'Z)^{-1} z_i)^2 \right] - (\tilde{\sigma}_{v\epsilon}^{(g)})^2 \right| \\
&= \left| \frac{(\tilde{\sigma}_{v\epsilon}^{(g)})^2}{l^2} \left(l^2 + l - 2\sum_{i=1}^{n} (z_i'(Z'Z)^{-1} z_i)^2 \right) - (\tilde{\sigma}_{v\epsilon}^{(g)})^2 \right| \\
&= \left| \frac{(\tilde{\sigma}_{v\epsilon}^{(g)})^2}{l^2} \left(l - 2\sum_{i=1}^{n} \left(z_i'(Z'Z)^{-1} z_i \right)^2 \right) \right| \\
&\leq \frac{(\tilde{\sigma}_{v\epsilon}^{(g)})^2}{l} + \frac{2(\tilde{\sigma}_{v\epsilon}^{(g)})^2 \sum_{i=1}^{n} z_i'(Z'Z)^{-1} z_i}{l^2} = O_P\left(\frac{1}{l}\right)
\end{aligned}$$

Therefore, $E^* \left[\frac{v^{*(g)'} P_Z \epsilon^*}{l} - \tilde{\sigma}_{v\epsilon}^{(g)} \right]^2 = O_P(1/l)$.

But, for any T^* such that $Var^*[T^*] = O_P(1/l)$, by the Tchebychev's inequality, we have for any $\delta > 0$ and any fixed $M_\delta > 0$,

$$P^*(|\sqrt{l}T^*| > M_\delta) \leq \frac{1}{M_\delta^2} Var^*(\sqrt{l}T^*) = \frac{1}{M_\delta^2} O_P(1),$$

Also, by the definition of $O_P(1)$, for δ, there exists a $M_\delta' < \infty$ such that

$$\lim_{n\to\infty} P(|O_P(1)| > M_\delta') = 0.$$

If we take $M_\delta = \sqrt{\frac{M_\delta'}{\delta}}$, i.e. $M_\delta^2 = \frac{M_\delta'}{\delta}$, then,

$$\begin{aligned}
\lim_{n\to\infty} P(|\frac{1}{M_\delta^2} O_P(1)| > \delta) &= \lim_{n\to\infty} P(\frac{\delta}{M_\delta'} |O_P(1)| > \delta) \\
&= \lim_{n\to\infty} P(|O_P(1)| > M_\delta') = 0.
\end{aligned}$$

These results show that $P^*(|\sqrt{l}T^*| > M_\delta) = o_P(1)$, i.e. $T^* = O_{P^*}(1/\sqrt{l})$ by the definitions at the beginning of the Appendix.

Therefore it follows that $v^{*(g)'} P_Z \epsilon^*/l - \tilde{\sigma}_{v\epsilon}^{(g)} = O_{P^*}(1/\sqrt{l})$, as required. This proves part (a). Parts (b) and (c) follow from proof similar to that of part (a).

The proof for parts (d) and (e) are similar, so we will only prove (d). To proceed, note that by the properties of Expectation and Trace operator,

$$\begin{aligned}
E^*\left[\| \frac{v^{*'} Z\hat{\pi}}{n} \|^2 \right] &= E^*\left[Tr\left(\frac{\hat{\pi}' Z' v^* v^{*'} Z\hat{\pi}}{n^2} \right) \right] \\
&= Tr(\tilde{\Sigma}_{vv}) E^* \left[\frac{Tr(\hat{\pi}' Z' Z\hat{\pi})}{n^2} \right] \\
&= \frac{1}{n} Tr(\tilde{\Sigma}_{vv}) Tr\left(\frac{\hat{\pi}' Z' Z\hat{\pi}}{n} \right)
\end{aligned}$$

Using $\hat{\pi} = (Z'Z)^{-1}Z'X$, we have

$$\frac{\hat{\pi}'Z'Z\hat{\pi}}{n} = \frac{\pi'Z'Z\pi}{n} + \frac{v'Z\pi}{n} + \frac{\pi'Z'v}{n} + \frac{v'P_Z v}{n} = Q + \lambda_n \Sigma_{vv} + o_P(1) = O_P(1)$$

Therefore

$$E^*\left[\| \frac{v^{*'}Z\hat{\pi}}{n} \|^2\right] = \frac{1}{n}O_P(1)O_P(1) = O_P\left(\frac{1}{n}\right)$$

because $\tilde{\Sigma}_{vv}$ is bounded in probability. It follows that $\hat{\pi}'Z'v^*/n = O_{P^*}(1/\sqrt{n})$. ■

We proceed to show the results for $\hat{\beta}^*_{B2SLS}$.

Lemma 3 Suppose that Assumptions 1-5 hold, then

$$\hat{\beta}^*_{B2SLS} - \hat{\beta}_{B2SLS} = O_{P^*}\left(\frac{1}{\sqrt{n}}\right)$$

Proof.

By the definition of Bias-corrected 2SLS estimator, we have

$$\hat{\beta}^*_{B2SLS} - \hat{\beta}_{B2SLS} = \left(\frac{X^{*'}P_Z X^*}{n} - \frac{\lambda_n X^{*'}X^*}{n}\right)^{-1}\left(\frac{X^{*'}P_Z\epsilon^*}{n} - \frac{\lambda_n X^{*'}\epsilon^*}{n}\right)$$

For the denominator, we have

$$\begin{aligned}
&\frac{X^{*'}P_Z X^*}{n} - \frac{\lambda_n X^{*'}X^*}{n} \\
= \;& \frac{\hat{\pi}'Z'Z\hat{\pi}}{n} + \frac{\hat{\pi}'Z'v^*}{n} + \frac{v^{*'}Z'\hat{\pi}}{n} + \frac{v^{*'}P_Z v^*}{n} - \lambda_n\left(\frac{\hat{\pi}'Z'Z\hat{\pi}}{n} + \frac{\hat{\pi}'Z'v^*}{n} + \frac{v^{*'}Z'\hat{\pi}}{n} + \frac{v^{*'}v^*}{n}\right) \\
= \;& (1-\lambda_n)\frac{\hat{\pi}'Z'Z\hat{\pi}}{n} + (1-\lambda_n)\left(\frac{\hat{\pi}'Z'v^*}{n} + \frac{v^{*'}Z'\hat{\pi}}{n}\right) + \frac{v^{*'}P_Z v^*}{n} - \lambda_n\frac{v^{*'}v^*}{n}
\end{aligned}$$

Note that $\hat{\pi}'Z'v^*/n = O_{P^*}(1/\sqrt{n})$ by Lemma 2. Also note that

$$\begin{aligned}
\frac{v^{*'}P_Z v^*}{n} &= \lambda_n\frac{v^{*'}P_Z v^*}{l} \\
&= \lambda_n\left\{\tilde{\Sigma}_{vv} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)\right\} \\
&= \lambda_n\tilde{\Sigma}_{vv} + O_{P^*}\left(\frac{\sqrt{l}}{n}\right)
\end{aligned}$$

where the second equality follows from Lemma 2. Analogously, we have

$$\lambda_n\frac{v^{*'}v^*}{n} = \lambda_n\tilde{\Sigma}_{vv} + O_{P^*}\left(\frac{1}{\sqrt{n}}\right)$$

Hence,

$$\begin{aligned}
\frac{X^{*'}P_Z X^*}{n} - \lambda_n\frac{X^{*'}X^*}{n} &= (1-\lambda_n)\frac{\hat{\pi}'Z'Z\hat{\pi}}{n} + \lambda_n\tilde{\Sigma}_{vv} - \lambda_n\tilde{\Sigma}_{vv} + o_{P^*}(1) \\
&= (1-\lambda_n)Q + (1-\lambda_n)\left(\frac{\hat{\pi}'Z'Z\hat{\pi}}{n} - Q\right) + o_{P^*}(1) \\
&= (1-\lambda_n)Q + (1-\lambda_n)\lambda_n\Sigma_{vv} + o_{P^*}(1) \geq C(Q + \lambda_n\Sigma_{vv})
\end{aligned}$$

for some $C > 0$, where the last inequality holds in the positive semidefinite sense with probability approaching 1 in probability. It then follows that

$$\left(\frac{X^{*'}P_Z X^*}{n} - \lambda_n \frac{X^{*'}X^*}{n}\right)^{-1} = O_{P^*}(1)$$

Similarly,

$$\begin{aligned}\frac{X^{*'}P_Z\epsilon^*}{n} - \lambda_n\frac{X^{*'}\epsilon^*}{n} &= (1-\lambda_n)\frac{\hat{\pi}'Z'\epsilon^*}{n} + \frac{v^{*'}P_Z\epsilon^*}{n} - \lambda_n\frac{v^{*'}\epsilon^*}{n} \\ &= O_{P^*}\left(\frac{1}{\sqrt{n}}\right) + \lambda_n\tilde{\sigma}_{v\epsilon} + O_{P^*}\left(\frac{\sqrt{l}}{n}\right) - \lambda_n\tilde{\sigma}_{v\epsilon} + O_{P^*}\left(\frac{1}{\sqrt{n}}\right) = O_{P^*}\left(\frac{1}{\sqrt{n}}\right)\end{aligned}$$

by Lemma 2. Therefore, $\hat{\beta}^*_{B2SLS} - \hat{\beta}_{B2SLS} = O_{P^*}(1)O_{P^*}\left(\frac{1}{\sqrt{n}}\right) = O_{P^*}\left(\frac{1}{\sqrt{n}}\right)$, as required. ■

PROOF OF THEOREM 1

To proceed, let us consider the order of

$$\frac{J_0^*}{l} \equiv \frac{\epsilon^{*'}P_Z\epsilon^*}{l\tilde{\sigma}_{\epsilon\epsilon}}$$

$$\begin{aligned}E^*\left[\frac{J_0^*}{l} - 1\right] &= \frac{1}{\tilde{\sigma}_{\epsilon\epsilon}}E^*\left[Tr(\epsilon^{*'}Z(Z'Z)^{-1}Z'\epsilon^*)\right] - 1 \\ &= \frac{1}{\tilde{\sigma}_{\epsilon\epsilon}}Tr((Z'Z)^{-1}Z'E^*[\epsilon^*\epsilon^{*'}]Z) - 1\end{aligned}$$

where E^* denotes the expectation under P^*. Note that by the definition of i.i.d bootstrap,

$$E^*[\epsilon_i^*\epsilon_j^*] = E^*[\epsilon_i^*]E^*[\epsilon_j^*] = 0$$

for $i \neq j$. Therefore,

$$\begin{aligned}E^*\left[\frac{J_0^*}{l} - 1\right] &= \frac{1}{\tilde{\sigma}_{\epsilon\epsilon}}Tr((Z'Z)^{-1}Z'(\tilde{\sigma}_{\epsilon\epsilon}I_n)Z) - 1 \\ &= \frac{1}{l}Tr(I_l) - 1 = 0\end{aligned}$$

and

$$\begin{aligned}\frac{J_0^*}{l} - 1 &= \frac{1}{l}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i'(Z'Z)^{-1}z_j\frac{\epsilon_i^*\epsilon_j^*}{\tilde{\sigma}_{\epsilon\epsilon}} - 1 \\ &= \frac{1}{l}\sum_{i=1}^{n} z_i'(Z'Z)^{-1}z_i\left(\frac{\epsilon_i^{*2}}{\tilde{\sigma}_{\epsilon\epsilon,}} - 1\right) + \frac{1}{l}\sum_{i\neq j} z_i'(Z'Z)^{-1}z_j\frac{\epsilon_i^*\epsilon_j^*}{\tilde{\sigma}_{\epsilon\epsilon}} \\ &\equiv A_1^* + A_2^*\end{aligned}$$

A_1^* and A_2^* are uncorrelated under P^*.

Then, we proceed to calculate the variance of A_1^* and A_2^* under P^*,

$$\begin{aligned}
Var^*\left[A_1^*\right] &= Var^*\left[\frac{1}{l}\sum_{i=1}^{n} z_i'(Z'Z)^{-1}z_i\left(\frac{\epsilon_i^{*2}}{\tilde{\sigma}_{\epsilon\epsilon}}-1\right)\right] \\
&= \frac{1}{l^2}\sum_{i=1}^{n}\left(z_i'(Z'Z)^{-1}z_i\right)^2 Var^*\left[\frac{\epsilon_i^{*2}}{\tilde{\sigma}_{\epsilon\epsilon}}\right] \\
&= \frac{1}{l^2}\sum_{i=1}^{n}\left(z_i'(Z'Z)^{-1}z_i\right)^2\left(\frac{\tilde{\kappa}_4}{(\tilde{\sigma}_{\epsilon\epsilon})^2}-1\right) \\
&\leq \frac{n}{l^2}\left(\frac{\tilde{\kappa}_4}{(\tilde{\sigma}_{\epsilon\epsilon})^2}-1\right) = O_P\left(\frac{1}{l}\right)
\end{aligned}$$

using $0 \leq z_i(Z'Z)^{-1}z_i \leq 1$ for each i and the results in Lemma 1.

$$\begin{aligned}
Var^*[A_2^*] &= Var^*\left[\frac{1}{l}\sum_{i\neq j} z_i'(Z'Z)^{-1}z_j\frac{\epsilon_i^*\epsilon_j^*}{\tilde{\sigma}_{\epsilon\epsilon}}\right] \\
&= \frac{1}{l^2}E^*\left[\sum_{i\neq j}\sum_{k\neq l} z_i'(Z'Z)^{-1}z_j z_k'(Z'Z)^{-1}z_l\frac{\epsilon_i^*\epsilon_j^*}{\tilde{\sigma}_{\epsilon\epsilon}}\frac{\epsilon_k^*\epsilon_l^*}{\tilde{\sigma}_{\epsilon\epsilon}}\right] \\
&= \frac{2}{l^2}\sum_{i\neq j}(z_i'(Z'Z)^{-1}z_j)^2 \\
&= \frac{2}{l^2}\sum_{i=1}^{n} z_i'(Z'Z)^{-1}\left(\sum_{j=1,j\neq i}^{n} z_j z_j'\right)(Z'Z)^{-1}z_i \\
&= \frac{2}{l^2}\sum_{i=1}^{n}\left(z_i'(Z'Z)^{-1}z_i-(z_i'(Z'Z)^{-1}z_i)^2\right) \leq \frac{2}{l^2}n = O\left(\frac{1}{l}\right)
\end{aligned}$$

Thus, the variance of A_1^* and A_2^* under P^* is of order $O_P\left(1/l\right)$, therefore

$$\frac{J_0^*}{l}-1 = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$$

Then, let us consider the order of $\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}}-1$.

$$E^*\left[\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}}-1\right] = \frac{1}{n\tilde{\sigma}_{\epsilon\epsilon}}E^*\left[\sum_{i=1}^{n}\epsilon_i^{*2}\right]-1 = 0$$

$$\begin{aligned}
Var^*\left[\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}}-1\right] &= \frac{1}{n^2(\tilde{\sigma}_{\epsilon\epsilon})^2}Var^*\left[\sum_{i=1}^{n}\epsilon_i^{*2}\right] \\
&= \frac{1}{n(\tilde{\sigma}_{\epsilon\epsilon})^2}(E^*[\epsilon_i^{*4}]-(E^*[\epsilon_i^{*2}])^2) \\
&= \frac{1}{n(\tilde{\sigma}_{\epsilon\epsilon})^2}(\tilde{\kappa}_4-(\tilde{\sigma}_{\epsilon\epsilon})^2) = O_P\left(\frac{1}{n}\right) = O_P\left(\frac{1}{l}\right)
\end{aligned}$$

Therefore

$$\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1 = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$$

Then, we rewrite AR^* in the following form:

$$\frac{AR^*}{l} = (1-\lambda_n)\left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - \lambda_n \frac{J_0^*}{l}\right)^{-1} \frac{J_0^*}{l} \tag{10}$$

Note that

$$\begin{aligned}
\left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - \lambda_n \frac{J_0^*}{l}\right)^{-1} &= \left[\left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) - \lambda_n\left(\frac{J_0^*}{l} - 1\right) + (1-\lambda_n)\right]^{-1} \\
&= (1-\lambda_n)^{-1} - (1-\lambda_n)^{-2}\left[\left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) - \lambda_n\left(\frac{J_0^*}{l} - 1\right)\right] + O_{P^*}\left(\frac{1}{l}\right)
\end{aligned}$$

because $\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1$ and $\frac{J_0^*}{l} - 1$ are of order $O_{P^*}\left(1/\sqrt{l}\right)$. Putting this into equation(11), we obtain

$$\begin{aligned}
\frac{AR^*}{l} &= (1-\lambda_n)\left[(1-\lambda_n)^{-1} - (1-\lambda_n)^{-2}\left[\left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) - \lambda_n\left(\frac{J_0^*}{l} - 1\right)\right]\right]\frac{J_0^*}{l} + O_{P^*}\left(\frac{1}{l}\right) \\
&= \left(\frac{J_0^*}{l} - 1\right) + 1 - (1-\lambda_n)^{-1}\left[\left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) - \lambda_n\left(\frac{J_0^*}{l} - 1\right)\right] + O_{P^*}\left(\frac{1}{l}\right)
\end{aligned}$$

from which we get

$$(1-\lambda_n)\left(\frac{AR^*}{l} - 1\right) = \left(\frac{J_0^*}{l} - 1\right) - \left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) + O_{P^*}\left(\frac{1}{l}\right)$$

Thus, up to an $o_{P^*}(1)$ remainder,

$$\begin{aligned}
(1-\lambda_n)\sqrt{l}\left(\frac{AR^*}{l} - 1\right) &= \sqrt{l}\left\{\left(\frac{J_0^*}{l} - 1\right) - \left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right)\right\} \\
&= \frac{1}{\sqrt{l}}\sum_{i=1}^{n}(z_i'(Z'Z)^{-1}z_i - \lambda_n)\left(\frac{\epsilon_i^{*2}}{\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) + \frac{1}{\sqrt{l}}\sum_{i\neq j} z_i'(Z'Z)^{-1}z_j\frac{\epsilon_i^*\epsilon_j^*}{\tilde{\sigma}_{\epsilon\epsilon}} \\
&\equiv B_1^* + B_2^*
\end{aligned}$$

Exactly as before, we compute the variance of B_1^* under P^*, which yields:

$$\begin{aligned}
Var^*[B_1^*] &= Var^*\left[\frac{1}{\sqrt{l}}\sum_{i=1}^{n}(z_i'(Z'Z)^{-1}z_i - \lambda_n)\left(\frac{\epsilon_i^{*2}}{\tilde{\sigma}_{\epsilon\epsilon}} - 1\right)\right] \\
&= \frac{1}{l}\sum_{i=1}^{n}(z_i'(Z'Z)^{-1}z_i - \lambda_n)^2\left(\frac{\tilde{\kappa}_4}{(\tilde{\sigma}_{\epsilon\epsilon})^2} - 1\right) \\
&\leq \frac{1}{\lambda_n}\frac{1}{n}\sum_{i=1}^{n}\left|z_i'(Z'Z)^{-1}z_i - \lambda_n\right|\left(\frac{\tilde{\kappa}_4}{(\tilde{\sigma}_{\epsilon\epsilon})^2} - 1\right) \to_P 0
\end{aligned}$$

using Assumption 3. Therefore, $B_1^* = o_{P^*}(1)$.

For B_2^*, we check the conditions for the Central Limit Theorem by Kelejian and Prucha(2001, Thm.1)[8]. Let $\xi_{i,n}^* \equiv \epsilon_i^*/\sqrt{\tilde{\sigma}_{\epsilon\epsilon}}$. It is easy to see that $E^*[\xi_{i,n}^*] = 0$, and $\xi_{1,n}^*, ..., \xi_{n,n}^*$ are independent (conditional on the data). Therefore, Assumption 1 of this CLT is satisfied. Assumption 2 of this CLT is satisfied for $a_{ij,n} \equiv \frac{1}{\sqrt{n}} z_i'(Z'Z)^{-1} z_j$, as has been shown in Anatolyev and Gospodinov(2011)[1](Page 439). Finally, for Assumption 3, $sup_{1\leq i\leq n, n\geq 1} E^*[|\xi_{i,n}^*|^{2+\delta}]$ has to be bounded in probability for some $\delta > 0$, but this has been shown in Lemma 1.

Then, we proceed to calculate the variance of B_2^* under P^*:

$$\begin{aligned} Var^*[B_2^*] &= Var^*\left[\frac{1}{\sqrt{l}}\sum_{i\neq j} z_i'(Z'Z)^{-1}z_j \frac{\epsilon_i^*\epsilon_j^*}{\tilde{\sigma}_{\epsilon\epsilon}}\right] \\ &= \frac{2}{l}\sum_{i=1}^{n}\left(z_i'(Z'Z)^{-1}z_i - (z_i'(Z'Z)^{-1}z_i)^2\right) \\ &= \frac{2}{\lambda_n}\left((1-2\lambda_n)\frac{1}{n}\sum_{i=1}^{n} z_i'(Z'Z)^{-1}z_i + \lambda_n^2 - \frac{1}{n}\sum_{i=1}^{n}(z_i'(Z'Z)^{-1}z_i - \lambda_n)^2\right) \to 2(1-\lambda) \end{aligned}$$

using Assumption 3.

Therefore, we obtain that

$$(1-\lambda_n)\sqrt{l}\left(\frac{AR^*}{l} - 1\right) \Rightarrow_{d_{P^*}} N(0, 2(1-\lambda))$$

in probability, and $\Rightarrow_{d_{P^*}}$ denotes weak convergence under the bootstrap probability measure.

The result follows by Polya's Theorem, given that the normal distribution is everywhere continuous. ■

PROOF OF THEOREM 2

For J^*, first note that by the results in Lemma 1 and Lemma 2,

$$\begin{aligned} \frac{\hat{\epsilon}^{*\prime} P_Z \hat{\epsilon}^*}{l\tilde{\sigma}_{\epsilon\epsilon}} &= \frac{\left(\epsilon^* - X^*(\hat{\beta}_{B2SLS}^* - \hat{\beta}_{B2SLS})\right)' P_Z \left(\epsilon^* - X^*(\hat{\beta}_{B2SLS}^* - \hat{\beta}_{B2SLS})\right)}{l\tilde{\sigma}_{\epsilon\epsilon}} \\ &= \frac{J_0^*}{l} - 2(\hat{\beta}_{B2SLS}^* - \hat{\beta}_{B2SLS})' \frac{X^{*\prime} P_Z \epsilon^*}{l\tilde{\sigma}_{\epsilon\epsilon}} + \frac{(\hat{\beta}_{B2SLS}^* - \hat{\beta}_{B2SLS})' X^{*\prime} P_Z X^* (\hat{\beta}_{B2SLS}^* - \hat{\beta}_{B2SLS})}{l\tilde{\sigma}_{\epsilon\epsilon}} \end{aligned}$$

For the third term, we can see from the results in Lemma 2 and Lemma 3 that

$$\begin{aligned} \frac{X^{*\prime} P_Z X^*}{l\tilde{\sigma}_{\epsilon\epsilon}} &= \frac{1}{\lambda_n}\frac{\left(\hat{\pi}'Z'Z\hat{\pi} + \hat{\pi}'Z'v^* + v^{*\prime}Z\hat{\pi} + v^{*\prime}P_Z v^*\right)}{n\tilde{\sigma}_{\epsilon\epsilon}} \\ &= \frac{1}{\lambda_n\tilde{\sigma}_{\epsilon\epsilon}}\left(Q + \lambda_n\Sigma_{vv} + o_P(1) + O_{P^*}\left(\frac{1}{\sqrt{n}}\right) + \lambda_n\tilde{\Sigma}_{vv} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)\right) = O_{P^*}(1) \end{aligned}$$

Therefore,

$$\frac{\hat{\epsilon}^{*\prime} P_Z \hat{\epsilon}^*}{l\tilde{\sigma}_{\epsilon\epsilon}} = \frac{J_0^*}{l} - 2(\hat{\beta}^*_{B2SLS} - \hat{\beta}_{B2SLS})' \frac{X^{*\prime} P_Z \epsilon^*}{l\tilde{\sigma}_{\epsilon\epsilon}} + O_{P^*}\left(\frac{1}{l}\right) \quad (11)$$

using the result that $\hat{\beta}^*_{B2LS} - \hat{\beta}_{B2SLS} = O_{P^*}\left(\frac{1}{\sqrt{n}}\right) = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$. Analogously,

$$\begin{aligned} \frac{\hat{\sigma}^*_{\epsilon\epsilon}}{\tilde{\sigma}_{\epsilon\epsilon}} - 1 &= \frac{\left(\epsilon^* - X^*(\hat{\beta}^*_{B2SLS} - \hat{\beta}_{B2SLS})\right)' \left(\epsilon^* - X^*(\hat{\beta}^*_{B2SLS} - \hat{\beta}_{B2SLS})\right)}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1 \quad (12) \\ &= \left(\frac{\epsilon^{*\prime}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) - 2(\hat{\beta}^*_{B2SLS} - \hat{\beta}_{B2SLS})'\lambda_n \frac{X^{*\prime}\epsilon^*}{l\tilde{\sigma}_{\epsilon\epsilon}} + O_{P^*}\left(\frac{1}{l}\right) \quad (13) \end{aligned}$$

by the results in Lemma 2 and Lemma 3.

Next, we rewrite J^* in the following form

$$\begin{aligned} \frac{J^*}{l} - 1 &= \left(\frac{\hat{\epsilon}^{*\prime} P_Z \hat{\epsilon}^*}{l\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) \frac{\tilde{\sigma}_{\epsilon\epsilon}}{\hat{\sigma}^*_{\epsilon\epsilon}} + \left(\frac{\tilde{\sigma}_{\epsilon\epsilon}}{\hat{\sigma}^*_{\epsilon\epsilon}} - 1\right) \quad (14) \\ &= \left(\frac{\hat{\epsilon}^{*\prime} P_Z \hat{\epsilon}^*}{l\tilde{\sigma}_{\epsilon\epsilon}} - 1\right)\left(1 + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)\right) + \left(\frac{\tilde{\sigma}_{\epsilon\epsilon}}{\hat{\sigma}^*_{\epsilon\epsilon}} - 1\right) \quad (15) \end{aligned}$$

Note that

$$\frac{\tilde{\sigma}_{\epsilon\epsilon}}{\hat{\sigma}^*_{\epsilon\epsilon}} = \left(1 + \left(\frac{\hat{\sigma}^*_{\epsilon\epsilon}}{\tilde{\sigma}_{\epsilon\epsilon}} - 1\right)\right)^{-1} = 1 - \left(\frac{\hat{\sigma}^*_{\epsilon\epsilon}}{\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) + o_{P^*}\left(\frac{1}{\sqrt{l}}\right) \quad (16)$$

Putting the results in eqs (11), (13), (15) and (16) together, we obtain

$$\frac{J^*}{l} - 1 = \left(\frac{J_0^*}{l} - 1\right) - \left(\frac{\epsilon^{*\prime}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) - \frac{2}{\lambda_n\tilde{\sigma}_{\epsilon\epsilon}}(\hat{\beta}^*_{B2SLS} - \hat{\beta}_{B2SLS})' \frac{X^{*\prime}(P_Z - \lambda_n I)\epsilon^*}{n} + o_{P^*}\left(\frac{1}{\sqrt{l}}\right)$$

Now, let us consider the order of the third term

$$\frac{X^{*\prime}(P_Z - \lambda_n I)\epsilon^*}{\tilde{\sigma}_{\epsilon\epsilon} n} = \frac{\hat{\pi}' Z'(P_Z - \lambda_n I)\epsilon^*}{\tilde{\sigma}_{\epsilon\epsilon} n} + \frac{v^{*\prime}(P_Z - \lambda_n I)\epsilon^*}{\tilde{\sigma}_{\epsilon\epsilon} n}$$

$$\begin{aligned} E^*\left[\frac{\hat{\pi}' Z'(P_Z - \lambda_n I)\epsilon^*}{\tilde{\sigma}_{\epsilon\epsilon} n}\right] &= \frac{\hat{\pi}' Z'(P_Z - \lambda_n I)E^*\left[\epsilon^*\right]}{\tilde{\sigma}_{\epsilon\epsilon} n} = 0 \\ Var^*\left[\frac{\hat{\pi}' Z'(P_Z - \lambda I)\epsilon^*}{\tilde{\sigma}_{\epsilon\epsilon} n}\right] &= (1 - \lambda_n)^2 \frac{\hat{\pi}' Z' Z \hat{\pi}}{n^2} = O_P\left(\frac{1}{n}\right) \end{aligned}$$

by Assumption 5. Therefore, $\hat{\pi}' Z'(P_Z - \lambda I)\epsilon^*/(\tilde{\sigma}_{\epsilon\epsilon} n)$ is $O_{P^*}(\frac{1}{\sqrt{n}})$. For the second part, note that

$$\begin{aligned} E^*\left[\frac{v^{*\prime}(P_Z - \lambda_n I)\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}}\right] &= E^*\left[\frac{v^{*\prime} P_Z \epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}}\right] - \lambda_n E^*\left[\frac{v^{*\prime}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}}\right] \\ &= \frac{Tr\left(P_Z E^*[\epsilon^{*\prime} v^*]\right)}{n} - \lambda_n E^*\left[\frac{v^{*\prime}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}}\right] \\ &= \lambda_n \frac{\tilde{\sigma}_{v\epsilon}}{\tilde{\sigma}_{\epsilon\epsilon}} - \lambda_n \frac{\tilde{\sigma}_{v\epsilon}}{\tilde{\sigma}_{\epsilon\epsilon}} = 0 \end{aligned}$$

And along the lines of Newey(2004, proof of Lemma 1), one can see that its variance under P^* is $O_P(1/n)$, which leads to the conclusion that $X^{*'}(P_Z - \lambda_n I)\epsilon^*/(\tilde{\sigma}_{\epsilon\epsilon} n) = O_{P^*}(1/\sqrt{n})$.

Thus up to an $o_{P^*}(1)$ remainder,

$$\begin{aligned}
\sqrt{l}\left(\frac{J^*}{l} - 1\right) &= \sqrt{l}\left\{\left(\frac{J_0^*}{l} - 1\right) - \left(\frac{\epsilon^{*'}\epsilon^*}{n\tilde{\sigma}_{\epsilon\epsilon}} - 1\right)\right\} \\
&= \frac{1}{\sqrt{l}}\sum_{i=1}^{n}(z_i'(Z'Z)^{-1}z_i - \lambda_n)\left(\frac{\epsilon_i^{*2}}{\tilde{\sigma}_{\epsilon\epsilon}} - 1\right) + \frac{1}{\sqrt{l}}\sum_{i\neq j} z_i'(Z'Z)^{-1}z_j\frac{\epsilon_i^*\epsilon_j^*}{\tilde{\sigma}_{\epsilon\epsilon}} \\
&\equiv B_1^* + B_2^*
\end{aligned}$$

and using the same arguments as in the case of AR^*, we obtain that

$$\sqrt{l}\left(\frac{J^*}{l} - 1\right) \Rightarrow_{d_{P^*}} N(0, 2(1-\lambda))$$

in probability.

Finally, the result follows by Polya's Theorem, given that the normal distribution is everywhere continuous. ■

Table 1. Empirical rejection frequency at 0.05 nominal level of the AR tests, $R_f = 0.01$.

ρ=0	l=10	l=30	l=50
AR	0.0648	0.0912	0.143
AR_{AG}	0.0564	0.0606	0.0672
$AR_{Bootstrap}$	0.0492	0.0492	0.0462
ρ=0.2	l=10	l=30	l=50
AR	0.0644	0.0952	0.149
AR_{AG}	0.0552	0.0636	0.0736
$AR_{Bootstrap}$	0.0462	0.0488	0.0522
ρ=0.5	l=10	l=30	l=50
AR	0.0666	0.0956	0.1496
AR_{AG}	0.057	0.061	0.0734
$AR_{Bootstrap}$	0.0512	0.0488	0.0532
ρ=0.8	l=10	l=30	l=50
AR	0.0706	0.0984	0.148
AR_{AG}	0.0606	0.068	0.0698
$AR_{Bootstrap}$	0.0542	0.0544	0.053

Table 2. Empirical rejection frequency at 0.05 nominal level of the AR tests, $R_f = 0.2$.

ρ=0	l=10	l=30	l=50
AR	0.0664	0.0988	0.1408
AR_{AG}	0.0582	0.0646	0.0698
$AR_{Bootstrap}$	0.0532	0.0534	0.0486
ρ=0.2	l=10	l=30	l=50
AR	0.0654	0.1038	0.1478
AR_{AG}	0.0566	0.0652	0.07
$AR_{Bootstrap}$	0.0504	0.05	0.0464
ρ=0.5	l=10	l=30	l=50
AR	0.0666	0.0978	0.1532
AR_{AG}	0.0568	0.063	0.072
$AR_{Bootstrap}$	0.0516	0.0514	0.0514
ρ=0.8	l=10	l=30	l=50
AR	0.0636	0.1012	0.145
AR_{AG}	0.053	0.0654	0.0736
$AR_{Bootstrap}$	0.0476	0.054	0.05

Note. AR, AR_{AG} and $AR_{Bootstrap}$ denote the conventional AR test , the modified AR test proposed in Anatolyev and Gospodinov(2011), and the Bootstrapping AR test proposed in this paper, respectively.

Table 3. Empirical rejection frequency at 0.05 nominal level of the J tests, $R_f = 0.01$.

ρ=0	l=10	l=30	l=50
J	0.0206	0.0078	0.0022
J_{AG}	0.0264	0.0208	0.0214
$J_{Bootstrap}$	0.037	0.0262	0.0262
ρ=0.2	l=10	l=30	l=50
J	0.0236	0.0094	0.003
J_{AG}	0.0298	0.0238	0.0218
$J_{Bootstrap}$	0.0396	0.0288	0.0246
ρ=0.5	l=10	l=30	l=50
J	0.0232	0.0086	0.0016
J_{AG}	0.03	0.0234	0.023
$J_{Bootstrap}$	0.0418	0.0282	0.0284
ρ=0.8	l=10	l=30	l=50
J	0.039	0.0128	0.0032
J_{AG}	0.047	0.0318	0.0258
$J_{Bootstrap}$	0.0632	0.0398	0.0326

Table 4. Empirical rejection frequency at 0.05 nominal level of the J tests, $R_f = 0.2$.

ρ=0	l=10	l=30	l=50
J	0.0452	0.0138	0.0016
J_{AG}	0.0398	0.0318	0.031
$J_{Bootstrap}$	0.0452	0.0378	0.0342
ρ=0.2	l=10	l=30	l=50
J	0.0302	0.011	0.0028
J_{AG}	0.0366	0.0324	0.025
$J_{Bootstrap}$	0.0408	0.0354	0.0288
ρ=0.5	l=10	l=30	l=50
J	0.0448	0.0204	0.00302
J_{AG}	0.054	0.0458	0.0356
$J_{Bootstrap}$	0.0564	0.0478	0.0388
ρ=0.8	l=10	l=30	l=50
J	0.051	0.043	0.0178
J_{AG}	0.0598	0.0764	0.0744
$J_{Bootstrap}$	0.056	0.0714	0.0794

Note. J, J_{AG} and $J_{Bootstrap}$ denote the conventional J test , the modified J test proposed in Anatolyev and Gospodinov(2011), and the Bootstrapping J test proposed in this paper, respectively.