Economics Letters 157 (2017) 107–111

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

On bootstrap validity for specification testing with many weak instruments

Maximilien Kaffo [a], Wenjie Wang [b,*]

[a] International Monetary Fund, 700 19th St NW, Washington, DC 20431, USA
[b] Graduate School of Social Sciences, Hiroshima University, 1-2-1 Kagamiyama, Higashihiroshima, 7398525, Japan

HIGHLIGHTS

- We show bootstrap validity for the J test and Anderson–Rubin test under many/many weak instruments.
- The bootstrap does not require an a priori choice of asymptotic framework.
- Monte Carlo simulation shows that the bootstrap has a good finite-sample performance under many/many weak instruments.

ARTICLE INFO

Article history:
Received 4 May 2017
Received in revised form 2 June 2017
Accepted 3 June 2017
Available online 13 June 2017

JEL classification:
C12
C15
C26

Keywords:
Bootstrap
J test
Anderson–Rubin test
Many instruments
Weak instruments

ABSTRACT

This paper studies the asymptotic validity of bootstrapping the J test of over-identifying restrictions and the Anderson–Rubin (AR) test under many/many weak instrument sequences. We show that the (residual-based) bootstrap consistently estimates the limiting distributions of interest under these asymptotic frameworks. Interestingly, such bootstrap validity holds even if the bootstrap cannot mimic well certain important properties in the model. In addition, the studied bootstrap procedures are easy to implement in practice because they do not require an a priori choice between the conventional asymptotics and the many/many weak instrument asymptotics. Monte Carlo simulation shows that the bootstrap techniques provide a more reliable method to approximate the null distribution of the J and AR test statistics under many/many weak instruments.

1. Introduction

The conventional asymptotic theory often provides a poor approximation to the finite-sample distribution of instrumental variable (IV) estimators and test statistics, especially with weak instruments (Staiger and Stock, 1997) or many instruments (Bekker, 1994; Chao and Swanson, 2005). Despite the large literature on estimation with many/many weak instruments, the literature on corresponding tests remains relatively sparse. Anatolyev and Gospodinov (2011, henceforth AG) propose modifications of the J test of overidentifying restrictions and the Anderson–Rubin (AR) test so that these tests can be robust to many instruments.

In this paper, we study the bootstrap as an alternative inference method for the J and AR tests under many/many weak instruments, and show the bootstrap validity. Interestingly, such validity holds even if the bootstrap cannot mimic well certain important properties of the IV model. Simulations show that the bootstrap provides a more reliable method for the J and AR tests.

2. Model, assumptions and test statistics

We consider a standard linear IV regression given by

$$y = X\beta + \epsilon, \tag{1}$$

$$X = Z\Pi + V, \tag{2}$$

where y and X are an $n\times 1$ vector and an $n\times k$ matrix of observations on the endogenous variables, respectively, and Z is an $n \times l$ matrix of observation on the instruments, which we treat as deterministic. Alternatively, the results that follow can be interpreted as being conditional on Z. ϵ and V are an $n \times 1$ vector and an $n \times k$ matrix of random disturbances, respectively. Denote $P_Z = Z(Z'Z)^{-1}Z'$ and

* Corresponding author. Fax: +81 82 424 7212.
E-mail addresses: mkaffomelou@imf.org (M. Kaffo), wenjiew@hiroshima-u.ac.jp (W. Wang).

http://dx.doi.org/10.1016/j.econlet.2017.06.004

$M_Z = I_n - P_Z$, where I_n is an identity matrix with dimension n. The model and data are assumed to satisfy the following conditions.

Assumption 1. The errors $\eta_i = (\epsilon_i, V_i')'$ are i.i.d. for $i = 1, \ldots, n$ 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 $n \to \infty$, $\lambda_n = l/n \to \lambda$ for some constant λ satisfying $0 < \lambda < 1$. There exists a non-decreasing sequence of positive real numbers r_n such that, as $n \to \infty$, $r_n/n \to \kappa$ for some constant κ, with $0 \le \kappa < \infty$, and such that $\Pi'Z'Z\Pi/r_n \to Q$, where Q is a positive definite matrix. In addition, $\sqrt{l}/r_n \to 0$ as $n \to \infty$.

Assumption 2 allows l to be a nontrivial fraction of the sample size. Moreover, r_n can be interpreted as the rate at which the concentration parameter, $\Sigma_{vv}^{-1/2}\Pi'Z'Z\Pi\Sigma_{vv}^{-1/2}$, grows as n increases. Thus, one can characterize the quality of instruments by the order of magnitude of r_n. In particular, the case $r_n = n$ corresponds to the many (strong) instrument asymptotics in Bekker (1994), and the case $r_n < n$ corresponds to the many weak instrument asymptotics in Chao and Swanson (2005).

Assumption 3. As $n \to \infty$, $n^{-1}\sum_{i=1}^{n}|z_i(Z'Z)^{-1}z_i - \lambda| \to 0$.

Assumption 3 requires that the diagonal elements of P_Z do not exhibit variation asymptotically; i.e., are asymptotically nonstochastic. Anatolyev and Yaskov (2017) have systematically studied this assumption, and provided specific examples where it holds or fails. In particular, they show that situations with indicator instruments of equal group sizes, with independent instruments (including Gaussian), with instruments drawn from a log-concave distribution, with instruments distributed according to Gaussian copula, or with instruments following a factor model belong to the asymptotically nonstochastic case. On the other hand, situations with indicator instruments of unequal group sizes, or with dummy instruments (both stand-alone and those that interact with other instruments) lead to nontrivial asymptotic variation in diagonal elements, and Assumption 3 does not hold in these cases.

The widely used J statistic of overidentifying restrictions can be defined as

$$J = \frac{\epsilon(\hat{\beta})'P_Z\epsilon(\hat{\beta})}{\hat{\sigma}_{\epsilon\epsilon}(\hat{\beta})} \tag{3}$$

where $\epsilon(\hat{\beta}) = y - X\hat{\beta}$, $\hat{\sigma}_{\epsilon\epsilon}(\hat{\beta}) = \epsilon(\hat{\beta})'\epsilon(\hat{\beta})/n$, and $\hat{\beta}$ is a consistent IV estimator under many/many weak instrument sequences; e.g., the limited information maximum likelihood (LIML) estimator or the bias-corrected TSLS estimator. The second test statistic we study is the Anderson–Rubin (AR) statistic for $H_0 : \beta = \beta_0$, which takes the following form

$$AR = \frac{\epsilon(\beta_0)'P_Z\epsilon(\beta_0)}{\frac{1}{n-l}\epsilon(\beta_0)'M_Z\epsilon(\beta_0)} \tag{4}$$

where $\epsilon(\beta_0) = y - X\beta_0$.

Under the standard fixed l asymptotics, J is distributed as $\chi^2(l-k)$ and AR is distributed as $\chi^2(l)$ under the null; we reject when $J > q_\alpha^{\chi^2(l-k)}$ and $AR > q_\alpha^{\chi^2(l)}$. However, they are no longer valid when the number of instruments becomes large. To solve this problem, AG propose corrected tests that are valid under Bekker (1994)'s many (strong) instrument asymptotics. Specifically, the corrected J test rejects when

$$J > q^{\chi^2(l-k)}_{\Phi(\sqrt{1-\lambda}\Phi^{-1}(\alpha))} \tag{5}$$

and the corrected AR test rejects when

$$J > q^{\chi^2(l-k)}_{\Phi(\Phi^{-1}(\alpha)/\sqrt{1-\lambda})} \tag{6}$$

where $\Phi(x)$ is the standard normal cumulative distribution function. Interestingly, the limiting distributions derived in their paper also hold under many weak instrument asymptotics.

Corollary 2.1. *Suppose that Assumptions 1–3 holds. Then,*

$$\sqrt{l}\left(\frac{J}{l} - 1\right) \to^d N(0, 2(1-\lambda))$$

and under $H_0 : \beta = \beta_0$,

$$\sqrt{l}\left(\frac{AR}{l} - 1\right) \to^d N(0, 2/(1-\lambda)).$$

Therefore, the corrected tests in (5)–(6) remain valid even under many weak instruments. However, we find in simulations that the asymptotic approximation may become less correct when the instruments are relatively weak.

3. Bootstrap J and AR tests with many/many weak instruments

We study the bootstrap as an alternative method, and show the bootstrap validity under many/many weak instruments. We consider two residual-based procedures: the standard bootstrap and the efficient bootstrap in Davidson and MacKinnon (2008), which are carried out as follows:

Step 1: The residuals are obtained as:

$$\epsilon(\hat{\beta}) = y - X\hat{\beta}$$
$$\widehat{V} = X - Z\widehat{\Pi}$$

where $\widehat{\Pi} = (Z'Z)^{-1}Z'X$ for the standard bootstrap and $\widehat{\Pi} = (Z'Z)^{-1}Z'\left(X - \epsilon(\hat{\beta})\frac{\epsilon(\hat{\beta})'M_ZX}{\epsilon(\hat{\beta})'M_Z\epsilon(\hat{\beta})}\right)$ for the efficient bootstrap.

Step 2: $\left(\epsilon(\hat{\beta}), \widehat{V}\right)$ are re-centered to yield $\left(\tilde{\epsilon}, \widetilde{V}\right)$. Then, (ϵ^*, V^*) are drawn from the empirical distribution function of $\left(\tilde{\epsilon}, \widetilde{V}\right)$.

Step 3: We set

$$y^* = X^*\hat{\beta} + \epsilon^*$$
$$X^* = Z\widehat{\Pi} + V^*$$

Step 4: Obtain $\epsilon^*(\hat{\beta}^*) = y^* - X^*\hat{\beta}^*$, where $\hat{\beta}^*$ is computed using the bootstrap data. Then, we construct the bootstrap test statistic

$$J^* = \frac{\epsilon^*(\hat{\beta}^*)'P_Z\epsilon^*(\hat{\beta}^*)}{\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)} \tag{7}$$

where $\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*) = \epsilon^*(\hat{\beta}^*)'\epsilon^*(\hat{\beta}^*)/n$. For the bootstrap AR test, we let

$$AR^* = \frac{\epsilon^{*\prime}P_Z\epsilon^*}{\frac{1}{n-l}\epsilon^{*\prime}M_Z\epsilon^*} \tag{8}$$

where ϵ^* is generated in Step 2. Note that for the AR test, one may also generate the bootstrap d.g.p. under the null $H_0 : \beta = \beta_0$ by replacing $\hat{\beta}$ with β_0 in Step 1.

Step 5: Repeat Steps 1–4 B times, and compute the bootstrap P values $\widehat{p}_J^* = B^{-1}\sum_{j=1}^{B} I(J_j^* > J)$ and $\widehat{p}_{AR}^* = B^{-1}\sum_{j=1}^{B} I(AR_j^* > AR)$, where $I(\cdot)$ is the indicator function. We reject the null hypothesis if the bootstrap P value is smaller than α.

The following result states the bootstrap validity for the tests under many/many weak instruments.

Theorem 3.1. *Suppose that Assumptions 1–3 holds. Then,*

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

and under $H_0 : \beta = \beta_0$,

$$sup_{x\in R}\left|P^*\left(AR^* \le x\right) - P\left(AR \le x\right)\right| \to_p 0$$

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