Journal of Econometrics 192 (2016) 231–268

Contents lists available at ScienceDirect

Journal of Econometrics

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

Bootstrap inference for instrumental variable models with many weak instruments

Wenjie Wang [a,1], Maximilien Kaffo [b,*]

[a] Graduate School of Social Sciences, Hiroshima University, 1-2-1 Kagamiyama, Higashihiroshima, 7398525, Japan

[b] International Monetary Fund, 700 19th St NW, Washington, DC 20431, USA

ARTICLE INFO

Article history:
Received 9 June 2014
Received in revised form
3 April 2015
Accepted 25 December 2015
Available online 6 January 2016

JEL classification:
C15
C26

Keywords:
Bootstrap
Many instruments
Weak instruments
LIML
FULL
Corrected standard error

ABSTRACT

This study's main contribution is to theoretically analyze the application of bootstrap methods to instrumental variable models when the available instruments may be weak and the number of instruments goes to infinity with the sample size. We demonstrate that a standard residual-based bootstrap procedure cannot consistently estimate the distribution of the limited information maximum likelihood estimator or Fuller (1977) estimator under many/many weak instrument sequence. The primary reason is that the standard procedure fails to capture the instrument strength in the sample adequately. In addition, we consider the restricted efficient (RE) bootstrap of Davidson and MacKinnon (2008, 2010, 2014) that generates bootstrap data under the null (restricted) and uses an efficient estimator of the coefficient of the reduced-form equation (efficient). We find that the RE bootstrap is also invalid; however, it effectively mimics more key features in the limiting distributions of interest, and thus, is less distorted in finite samples than the standard bootstrap procedure. Finally, we propose modified bootstrap procedures that provide a valid distributional approximation for the two estimators with many/many weak instruments. A Monte Carlo experiment shows that hypothesis testing based on the asymptotic normal approximation can have severe size distortions in finite samples. Instead, our modified bootstrap procedures greatly reduce these distortions.

1. Introduction

Empirical applications of instrumental variable (IV) estimation often produce imprecise results. It is now well understood that standard first-order asymptotic theory breaks down when the instruments are weakly correlated with the endogenous regressors. In this case, commonly used IV estimators such as two-stage least square (TSLS) and limited information maximum likelihood (LIML) estimators can lose consistency; cf., Dufour (1997) and Staiger and Stock (1997), among others. However, as demonstrated by Chao and Swanson (2005), having many instruments in such a weakly identified situation can help to improve estimation accuracy. Indeed, using a large number of instruments can enhance the growth of the so-called concentration parameter even if each individual instrument is only weakly correlated with the endogenous explanatory variables. Chao and Swanson (2005) show that for well-centered IV estimators such as LIML, consistency can be established even when instrument weakness is such that the rate of growth of the concentration parameter is much slower than the sample size n. In addition, Hansen et al. (2008) reveal in an application from Angrist and Krueger (1991) that using 180 instruments, rather than 3, substantially improves estimator accuracy.

Moreover, for implementing inferences in the context of many/many weak instruments, Hansen et al. (2008) provide corrected standard errors (CSE). The CSE are an extension of those of Bekker (1994) to the case of non-Gaussian disturbances and are correct under a variety of asymptotic frameworks, including the many weak instrument sequence of Chao and Swanson (2005) and Stock and Yogo (2005), as well as the many instrument sequence of Kunitomo (1980), Morimune (1983) and Bekker (1994). Recently, the CSE are extended further by Chao et al. (2012) and Hausman et al. (2012) to the heteroscedastic case and by Newey and Windmeijer (2009) to continuously updating generalized method of moments (CUE) and other generalized empirical likelihood (GEL) estimators.

* Corresponding author.
E-mail addresses: wenjiew@hiroshima-u.ac.jp (W. Wang), mkaffomelou@imf.org (M. Kaffo).
[1] Tel.: +81 82 424 7262; fax: +81 82 424 7212.

http://dx.doi.org/10.1016/j.jeconom.2015.12.016
0304-4076/

However, our simulation evidence shows that hypothesis testing or confidence intervals (CIs) based on the CSE can be distorted severely in finite samples, especially in the case of strong endogeneity. This provides motivation for the use of the bootstrap instead of the asymptotic normal approximation to improve the quality of inference. Furthermore, the CSE have a rather tedious form, and thus, can be difficult to implement in practice; this also motivates the use of bootstrap methods. In particular, the bootstrap would help to avoid computing the tedious form of the CSE if bootstrap variance estimators or percentile-type bootstrap methods are valid under many/many weak instruments.

The existing literature on bootstrapping IV models turns out to be rather limited. Moreira et al. (2009) provide theoretical proof that guarantees the bootstrap validity of Kleibergen (2002)'s score statistic even under Staiger and Stock (1997)'s weak instrument asymptotics, in which the coefficients of the instruments are specified to be in an $n^{-1/2}$ shrinking neighborhood of zero and the number of instruments is kept fixed. Davidson and MacKinnon (2008, 2010, 2014) study various bootstrap methods (pairs bootstrap and residual-based bootstrap) of hypothesis testing and constructing confidence sets for the IV model. Their extensive simulation results show that the bootstrap approaches typically perform very well relative to the normal approximation, including the case in which instruments are quite weak. However, all these studies focus on the case in which the number of instruments is kept small relative to the sample size.

In this paper, we analyze bootstrap-based inference methods under many/many weak instruments. Based on the excellent results for the cases with a small number of instruments, one may expect the bootstrap also to perform well when the number of instruments becomes large. Surprisingly, we find that the bootstrap typically fails to mimic the limiting distributions of IV estimators in this context. We first consider a standard residual-based bootstrap method, in which the residual of the structural-form equation is obtained by using the LIML or Fuller (1977, FULL) estimator and the residual of the reduced-form equation is obtained by using the least squares estimator. We show analytically that this procedure cannot estimate the limiting distribution of LIML or FULL consistently. In particular, when the number of instruments is of the same order of magnitude as the rate of growth of the concentration parameter, the bootstrap analogue correctly replicates the convergence rate of the estimator, but the bootstrap limiting distribution has an asymptotic variance different from the original one. Furthermore, when the number of instruments grows faster than the concentration parameter, the convergence rate of the bootstrap analogue becomes even faster than that of LIML or FULL.

The primary reason of this bootstrap failure is that the standard procedure generates in the bootstrap sample "pseudo" instrument strength, which has at least the same order of magnitude as the original instrument strength. In addition, in the case with a large number of instruments, the bootstrap d.g.p. cannot mimic well important features of the disturbances in the IV model. Because of these inconsistencies, commonly used bootstrap-based inference approaches such as bootstrap variance estimator or percentile type bootstrap methods will be invalid in the case of many/many weak instruments. Similar results can be shown for other IV estimators such as the TSLS estimator, the bias-corrected TSLS estimator (Nagar, 1959; Rothenberg, 1984), and various jackknife IV estimators (Phillips and Hale, 1977; Angrist et al., 1999; Chao et al., 2012; Hausman et al., 2012; Bekker and Crudu, 2015).

We then consider the restricted efficient (RE) bootstrap procedure of Davidson and MacKinnon (2008, 2010, 2014), which generates bootstrap data under the null hypothesis (restricted) and uses efficient estimates of the reduced-form equation (efficient). These studies demonstrate that the RE bootstrap performs very well relative to the standard procedure. Here, we show that in the current context, the RE bootstrap also cannot estimate the limiting distribution of LIML or FULL consistently. However, we find that it is typically more robust to the instrument weakness than the standard bootstrap, and hence, exhibits relatively less distortion in finite samples.

Finally, we propose modifications to the RE bootstrap and justify that our modified bootstrap procedures provide a valid distributional approximation for LIML or FULL under many/many weak instrument sequences. In particular, we modify the RE bootstrap procedure by accurately rescaling the residuals and by introducing alternative reduced-form estimators, which allows the bootstrap to mimic well the instrument strength in the sample. Furthermore, we show analytically that all the bootstrap procedures analyzed in this study are asymptotically valid under percentile-t type methods. A Monte Carlo experiment demonstrates that the CSE-based normal approximation can have severe size distortions when the concentration parameter is small and/or when the degree of endogeneity is high. Our modified procedures can largely remove these distortions. In particular, one of our modified bootstrap performs best among all the procedures in terms of size control, while our second procedure is relatively balanced between size and power.

To the best of our knowledge, this study is the first to theoretically analyze the bootstrap validity under many/many weak instruments, and we obtain interesting implications of the properties of bootstrap methods that can be overlooked under conventional asymptotics. Indeed, the asymptotic approach taken here forces the distributional approximations to be more sensitive to the number and strength of available instruments and our findings highlight a fragility of bootstrap-based approximations with respect to these key features. In particular, conditions much more restrictive than those for the CSE-based normal approximation are necessary for existing bootstrap methods to estimate the limiting distributions of IV estimators consistently under many/many weak instruments. Furthermore, our results include modified, valid bootstrap procedures for the IV models, which effectively mimics the important features in the limiting distribution of interest.

The remainder of the paper is organized as follows. Section 2 introduces the model and provides a summary of the asymptotic theory for the estimators of interest and the CSE. Section 3 analyzes various residual-based bootstrap procedures and documents the inconsistency of the standard and RE bootstraps under many/many weak instrument sequences. Furthermore, we show that our modified bootstrap procedures provide a valid distributional approximation for LIML or FULL in this context. Section 4 contains the Monte Carlo results, and Section 5 concludes. All proofs are relegated to the Appendix.

2. The model, assumptions and asymptotic theory

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. ϵ and V are an $n \times 1$ vector and an $n \times k$ matrix of random disturbances, respectively. We denote $P_Z = Z(Z'Z)^{-1}Z'$ and $M_Z = I_n - P_Z$, where I_n is an identity matrix with dimension n. Throughout this study, we consider the case in which k, the dimension of β, is small relative to n, but we let $l \to \infty$ as $n \to \infty$ in order to model the effect of having many/many weak instruments. In addition, we assume that the included exogenous