Detecting Weak Identification by Bootstrap*

Zhaoguo Zhan†

November 1, 2010

Abstract

This paper proposes to use bootstrap resampling as a diagnostic tool to detect the existence of weak identification in IV and GMM. When identification is strong, the bootstrap distribution of the IV/GMM estimator converges to a normal distribution. Consequently, the substantial difference between the bootstrap distribution and the normal distribution provides statistical evidence against the null of strong identification. I propose to extract the evidence by comparing the conventional t/Wald-based confidence interval with the bootstrap percentile confidence interval, and provide a quantitative threshold to distinguish between strong and weak identification. Unlike other existing tests, the proposed method has the unique feature of providing a graphic view of the identification strength, which is illustrated by an empirical example.

Keywords: Weak Identification, Bootstrap
JEL Classification: C12, C18, C26

*I am very thankful to Frank Kleibergen, Sophocles Mavroeidis, Blaise Melly and participants in lunch and seminar talks at Brown for their valuable comments. The Monte Carlo study in this paper was supported by Brown University through the use of the facilities of its Center for Computation and Visualization.

†Email: *Zhaoguo_Zhan@brown.edu*. Mail: Department of Economics, Box B, Brown University, Providence, RI 02912, USA. Web: *http://www.econ.brown.edu/students/zhaoguo_zhan*.

1 Introduction

The Instrumental Variable (IV) regression and Generalized Method of Moments (GMM) are becoming the standard toolkit for empirical economists, and it is now well known that both IV and GMM applications may suffer from the problem of weak identification. An example that has received sizable attention is the linear IV regression with weak instruments studied in Staiger and Stock (1997). When the strength of identification is weak, the finite sample distribution of IV/GMM estimators is poorly approximated by the normal distribution, which further induces the malfunction of conventional inference methods that rely on the property of asymptotic normality.[1] Two approaches co-exist to handle weak identification: the first approach is to use the robust methods in Anderson and Rubin (1949), Stock and Wright (2000), Kleibergen (2002)(2005), and Moreira (2003), which can produce confidence intervals/sets with the correct coverage, regardless of the strength of identification; the second approach that is popular in practice is to rule out weak identification by pretesting. Although identification-robust methods are available, excluding weak identification in IV/GMM applications has practical importance: if identification is not weak, then the rich set of conventional methods is applicable, making statistical inference and economic decisions much easier. For instance, other than its confidence interval, the point estimator of a parameter is usually preferred by policy makers, but it is not consistent or meaningful unless weak identification is excluded.

In this paper, I propose a method based on bootstrap resampling to detect whether weak identification exists in IV/GMM applications. This method has the unique feature of providing a graphic view of the strength of identification. In the econometric literature, there exists a group of tests on identification that this paper is in line with: for example, in the linear IV model, Hahn and Hausman (2002) and Stock and Yogo (2005) provide tests for the null of strong instruments and weak instruments respectively, and the first stage F test with the $F > 10$ rule of thumb proposed in Stock and Yogo (2005) is widely used; in Hansen (1982)'s GMM framework, which nests the linear IV model, the suggested tests include Wright (2002)(2003), Inoue and Rossi (2008), Bravo *et al.* (2009), etc.

The proposed method is illustrated by Figure 1. The exact finite sample distribution of IV/GMM estimators is generally unknown, but can be approximated either by the limiting normal distribution or by the bootstrap distribution. When the IV/GMM models are strongly identified, these two approximation methods are both valid, i.e.

[1] See the article by Stock *et al.* (2002) for a survey.

both the normal distribution and the bootstrap distribution are close to the exact distribution, and in practice, they can be used exchangeably. Consequently, strong identification implies that the bootstrap distribution is not far away from the normal distribution. When the bootstrap distribution is substantially different from the normal distribution, inference based on these two distributions might contradict; in this situation, it is inappropriate to consider the identification strength as strong. As a result, whether or not weak identification exists can be inferred by comparing the bootstrap distribution with the normal distribution.

Since its introduction by Efron (1979), the bootstrap has become a practical tool for statistical inference. The properties of the bootstrap are explained using the theory of the Edgeworth expansion in Hall (1992), and its econometric applications are illustrated in Horowitz (2001). As an alternative to the limiting distribution, the bootstrap approximates the distribution of a targeted statistic by resampling the data, and there is considerable evidence that it performs better than the first-order limiting distribution in finite samples.[2] However, the bootstrap does not always work well. When the IV/GMM models are not well identified, for instance, the bootstrap is known to be problematic when it is used to approximate the commonly used t/Wald statistic, as explained in Hall and Horowitz (1996). Nevertheless, the fact that the bootstrap fails still conveys useful information: as illustrated by Figure 1, the substantial difference between the bootstrap distribution and the normal distribution indicates that it is problematic to approximate the exact finite sample distribution of IV/GMM estimators by the normal distribution; in other words, the identification strength is weak.

Given the above introduction, it is tempting to apply normality tests, e.g. the Kolmogorov-Smirnov test, to examine whether the bootstrap distribution is normal in order to investigate the strength of identification. However, this route is not productive, when normality tends to be rejected as the number of bootstrap replications becomes large. This is due to the fact that, in general, the bootstrap distribution coincides with the normal distribution only at the limit where there are an infinite number of data points. In an IV/GMM application with a finite sample, the bootstrap distribution is not equivalent to the normal distribution, hence the null hypothesis for normality tests that the bootstrap distribution is identical to the normal distribution does not hold in empirical applications. Although normality tests do not work well, empirical researchers can still eyeball the graph of the bootstrap distribution to evaluate the strength of identification, and the test proposed in this paper does not

[2] See, for example, Horowitz (1994).

impose the equivalence of the bootstrap distribution and the normal distribution; only a substantial difference of these two distributions would induce the rejection of strong identification.

The rest of the paper is organized as follows: weak identification and the bootstrap strategy are illustrated in Section 2; in Section 3, the bootstrap-based method for determining whether weak identification exists is proposed; the linear IV regression model is used as an example in Section 4, with an application to Card (1995); in Section 5, Monte Carlo results are presented; Section 6 concludes the paper. Although the linear IV model is employed in this paper for expository purposes, the proposed method can be extended to the more general GMM framework.

Throughout the paper, the following notations are used: for an m by n matrix A, A_i (sometimes a_i is used when A is a vector) is its i^{th} row, and $P_A = A(A'A)^{-1}A'$, $M_A = I_m - P_A$, I_m is the m by m identity matrix; $vec(A)$ is the column vector containing the column by column vectorization of elements in A; for an object O, O^* is its bootstrap counterpart; $\Rightarrow$ stands for weak convergence; $\xrightarrow{p}$ stands for convergence in probability; $N(\mu, \sigma^2)$ is the normal distribution with mean μ, variance σ^2; $\otimes$ is the Kronecker product; $||.||$ is the Euclidean norm.

2 Motivation and Strategy

2.1 Weak Identification, Rank Tests

Let θ denote the population parameter of interest in an IV/GMM application, and $\hat{\theta}_n$ is the estimator for θ, e.g. the two stage least squares estimator. The subscript n for $\hat{\theta}_n$ refers to that $\hat{\theta}_n$ is computed by finite sample of n observations. Assuming that the rank condition and other regularity conditions are satisfied (cf., Wooldridge (2002)), the conventional first-order asymptotic theory yields that $\hat{\theta}_n$ is $\sqrt{n}-$consistent, and asymptotically normally distributed:

$$\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, \sigma^2), \text{ and there exists } \hat{\sigma} \xrightarrow{p} \sigma \tag{1}$$

As surveyed by Stock *et al.* (2002), the exact finite sample distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ can be substantially different from the normal distribution $N(0, \sigma^2)$, especially when the rank condition is weakly satisfied. The scenario that (1) does not provide a good approximation in finite sample applications is known as *weak identification*.

It is helpful to distinguish *weak identification* from *identification/noidentification*:

identification/noidentification refers to whether θ can be identified at the population level, hence it is unrelated to the sample size n; in contrast, *weak identification* targets that the first-order asymptotic theory may provide a poor approximation in finite samples. When θ is *identified* in the population, it is possible that the first-order approximation of (1) under a given sample[3] does not function well, hence *weak identification* may exist even though there is *identification*; the case that θ is *unidentified* is also nested by *weak identification*, since the first-order approximation of (1) also breaks down under *noidentification*. Loosely speaking, if the first-order approximation based on (1) is poor, then *weak identification* is concerned.

Since the conventional result of (1) breaks down under weak identification, it is important to investigate the identification strength in IV/GMM applications. A natural way is to examine the rank condition, as weak identification is more likely to happen when the rank condition is only weakly satisfied. There are rank tests available to serve this purpose. For example, in the linear IV model with a single endogenous regressor, the rank condition corresponds to that the correlation of the endogenous regressor and the instruments is non-zero, and Stock and Yogo (2005) propose the F test to examine this correlation: if the F statistic is greater than the tabled critical values, typically around 10 for small number of instruments, then the rank condition is considered strongly satisfied, and weak identification is excluded.

The approach of examining the rank condition, however, has some limitations when it is extended to the more general GMM framework: first of all, the rank statistic in the non-linear GMM may depend on the weakly identified parameters, hence could not be consistently estimated, see, e.g. Wright (2003); secondly, it is not clear how large the rank statistic needs to be in order to decide that identification in GMM is not weak, to the best of my knowledge. In the linear IV model, which can be seen as a special case of GMM, although the tabled critical values of the F test are available, they are derived under the i.i.d and homoscedasticity assumptions, hence it is not appropriate to apply them if heteroscedasticity or non-i.i.d. takes place.

Given the importance of detecting weak identification in IV/GMM applications, and the limitations of the rank test, a question naturally arises: is there a tool that does not have these limitations, and is universally applicable to both IV and GMM applications? This paper suggests that the bootstrap could be the tool.

[3]Even when the sample size is very large, weak identification could exist, e.g. in some regressions of Angrist and Krueger (1991), there are 329000 observations, but their instruments are weak.

2.2 Test Strategy by Bootstrap

Let $\hat{\theta}_n^*$ denote the bootstrap estimator of θ, and $\hat{\theta}_n^*$ is the counterpart of $\hat{\theta}_n$. $\hat{\theta}_n$ is computed by the data sample of the IV/GMM application, while $\hat{\theta}_n^*$ is computed by resampling the data sample or the model estimated from the data. The bootstrap counterpart of $\sqrt{n}(\hat{\theta}_n - \theta)$ is $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$.

When identification is not weak, together with mild regularity conditions, the bootstrap distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ asymptotically coincides with the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ (cf., Horowitz (2001)):

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \Rightarrow N(0, \sigma^2) \tag{2}$$

There are now three objects: the exact distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$, the normal distribution $N(0, \hat{\sigma}^2)$, and the bootstrap distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$. By (1) and (2), both of which hold under mild conditions, these three distributions are asymptotically equivalent.

The proposed bootstrap test strategy for detecting weak identification is to compare the bootstrap distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ with the normal distribution $N(0, \hat{\sigma}^2)$, or equivalently, compare the standardized bootstrap distribution of $\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ with the standard normal distribution. Since the difference between these two distributions is negligible when identification is strong, a substantial difference is the evidence against strong identification.

I find that this bootstrap test strategy is appealing: (i) it provides a simple and intuitive way to detect weak identification, i.e. empirical researchers can draw the graph of the bootstrap distribution and compare it with the normal distribution to evaluate the strength of identification; (ii) it is universally applicable to IV/GMM applications, i.e. although the data of IV/GMM applications may not be i.i.d. and homoscedastic, there exist various bootstrap methods to construct the bootstrap distribution, e.g. the block bootstrap in Hall and Horowitz (1996) for time series data in GMM, the pair and wild bootstrap used in Davidson and MacKinnon (2010) for heteroscedasticity in IV. By the bootstrap test strategy, as long as it is feasible to construct the bootstrap distribution, detecting weak identification reduces to the comparison of two distributions: the bootstrap distribution, and the normal distribution.

Definition 1. *Conditional on the sample of the observed data, the standardized bootstrap estimator* $X = \frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ *follows a distribution with c.d.f.* $F(x)$*. From now on,* ***the bootstrap distribution*** *refers to* $F(x)$ *in this paper.*

2.3 Why Not Use Normality Tests

Normality tests appear to be the intuitive choice, since comparing the bootstrap distribution with the normal distribution is proposed to infer the strength of identification. However, applying normality tests here will almost always induce the rejection of normality, and lead to the conclusion of weak identification.

Take the classic Kolmogorov-Smirnov (KS) test for example. With B bootstrap replications, let $X_i = \frac{\hat{\theta}_n^{*i} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}, i = 1, ..., B$ denote the i.i.d. bootstrapped estimators after standardization. To test the hypothesis that $X_i, i = 1, ..., B$ are B points drawn from the standard normal distribution $\Phi(x)$, the KS statistic is the supremum distance between $\Phi(x)$ and the empirical c.d.f. $\hat{F}(x)$, scaled by the square root of the number of points. As $B \to \infty$, the KS statistic goes to infinity, instead of converging to the Kolmogorov distribution:

$$KS = \sqrt{B} \underset{x}{sup} |\hat{F}(x) - \Phi(x)| \to \infty$$

where $\hat{F}(x) = \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}(X_i \leq x) = \frac{1}{B} \sum_{i=1}^{B} \mathbf{1}(\frac{\hat{\theta}_n^{*i} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \leq x)$.

This is because the hypothesis $F(x) = \Phi(x)$ does not hold when $n < \infty$. In other words, in empirical applications where the sample size is finite, $F(x)$ differs from $\Phi(x)$, although the difference may not be substantial, e.g. by the Edgeworth expansion in Horowitz (2001), $F(x) - \Phi(x) = O(n^{-1/2})$. Consequently, even when the difference between $F(x)$ and $\Phi(x)$ is minor, as $B \to \infty$, $KS \to \infty$, i.e. the KS test tends to reject normality when the bootstrap replication gets large.

Said differently, the bootstrap distribution is *not identical to* the normal distribution, although it can be asymptotically equivalent to the normal distribution. The KS test is for testing the hypothesis that the bootstrap distribution is *identical to* the normal distribution. For the purpose of examining whether the bootstrap distribution is close to the normal distribution, it is inappropriate to apply this test.

2.4 Comparison of C.I.'s

Instead of verifying the equivalence of the bootstrap distribution and the normal distribution by normality tests, this paper provides a quantitative measure of the difference/distance between these two distributions: if the measure shows that the difference is substantial, then identification is considered as *weak*; on the contrary, if the measure shows that the difference is negligible, then identification is considered as *strong*.

The measure results from the comparison of two confidence intervals (C.I.): the conventional C.I. derived by inverting the t/Wald test, and the bootstrap percentile C.I.. Consider the practical task of constructing a confidence interval for θ, the parameter of interest. The $100(1-\alpha)\%$ C.I. of θ derived by inverting the t/Wald test is written as:

$$C_t \equiv (\hat{\theta}_n - z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta}_n + z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}) \tag{3}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $\Phi(x)$. For example, when $\alpha = 5\%$, $z_{1-\alpha/2} \approx 1.96$, and the 95% C.I. of θ is approximately $(\hat{\theta}_n - 1.96\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta}_n + 1.96\frac{\hat{\sigma}}{\sqrt{n}})$.

Alternatively, a C.I. can also be constructed by the bootstrap percentile method: order the bootstrapped estimators $\{\hat{\theta}_n^{*i}, i = 1, ..., B\}$, write the ordered sequence as $\{\hat{\theta}_n^{*(i)}, i = 1, ..., B\}$, where $\hat{\theta}_n^{*(i)}$ is the i^{th} smallest of $\{\hat{\theta}_n^{*i}, i = 1, ..., B\}$; define $\hat{\theta}^*_{n,\alpha/2}$ and $\hat{\theta}^*_{n,1-\alpha/2}$: $\hat{\theta}^*_{n,\alpha/2} \equiv \hat{\theta}_n^{*(\lceil B\alpha/2\rceil)}$, $\hat{\theta}^*_{n,1-\alpha/2} \equiv \hat{\theta}_n^{*(\lceil B(1-\alpha/2)\rceil)}$, where $\lceil x \rceil$ denotes the integer ceiling of x. The bootstrapped $100(1-\alpha)\%$ C.I. of θ is:

$$C_b \equiv (\hat{\theta}^*_{n,\alpha/2}, \hat{\theta}^*_{n,1-\alpha/2}) \tag{4}$$

The two intervals in (3) and (4) are asymptotically equivalent when identification is not weak. To see this, equalizing the boundaries of these two intervals yields:

$$\frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} = -z_{1-\alpha/2}, \quad \frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} = z_{1-\alpha/2}$$

These two equalities approximately hold if the distribution of $\hat{\theta}_n^*$ after standardization, i.e. subtracting the estimate and dividing by the standard error, is close to the standard normal distribution.

In practice, the above two methods of constructing the $100(1-\alpha)\%$ C.I. for θ are both commonly used. From a practical point of view, no matter which method empirical researchers use, the correspondent intervals need not be substantially different from each other. If these two intervals do substantially differ, then it is difficult to make reliable economic inference: for instance, one interval may include zero while the other one does not, hence decisions of whether θ is significantly different from zero based on the two different C.I.'s could contradict.[4] The difference between the two intervals in (3) and (4) boils down to the difference between the bootstrap distribution and the normal distribution. If these two intervals are substantially different, it indicates the

[4]This problem is encountered in the empirical example of Card (1995).

bootstrap distribution is substantially different from the normal distribution, hence the identification strength is weak.

The idea of comparing alternative C.I.'s to investigate the identification status comes from Wright (2002), where he compares the interval derived by inverting the robust tests with the conventional interval derived by inverting the t/Wald test. Different from Wright (2002), I use the bootstrap to construct a C.I. for comparison; in addition, Wright (2002) provides an identification test, while in this paper, I target weak identification instead of identification.

2.5 Quantitative Definition

Based on the comparison of the two intervals in (3) and (4), a measure of the difference between the bootstrap distribution and the normal distribution, as well as a quantitative definition of weak identification, is provided below.

Definition 2. *Define D as the measure of the difference between the bootstrap distribution $F(x)$ and the standard normal distribution $\Phi(x)$:*

$$D \equiv \frac{q_{1-\alpha/2} - q_{\alpha/2}}{2z_{1-\alpha/2}} - 1$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $\Phi(x)$, $q_{\alpha/2}, q_{1-\alpha/2}$ are assumed to be the two unique quantiles of the continuous c.d.f. $F(x)$, i.e. $F^{-1}(\alpha/2) = q_{\alpha/2}$, $F^{-1}(1-\alpha/2) = q_{1-\alpha/2}$.

D is the relative difference of the lengths between the $\alpha/2, 1-\alpha/2$ quantiles of the two distributions. $D = 0$ if $F(x) = \Phi(x)$. The deviation of D from 0 indicates the deviation of $F(x)$ from $\Phi(x)$, and hence suggests existence of weak identification.

Definition 3. *Suppose there is a cutoff $\gamma > 0$. The identification strength is considered as **weak** in this paper if $|D| > \gamma$, otherwise the identification strength is **strong**.*

It is important to set a non-zero γ: if $\gamma = 0$, then identification tends to be considered as *weak*, because $F(x)$ is not exactly normal except in the limit of $n \to \infty$. For the same reason, it is inappropriate to apply normality tests, which impose $\gamma = 0$ under their null hypothesis.

D can be estimated by the relative difference of (3) and (4):

$$\hat{D} = \frac{\hat{q}_{1-\alpha/2} - \hat{q}_{\alpha/2}}{2z_{1-\alpha/2}} - 1 = \frac{\frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} - \frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}}{2z_{1-\alpha/2}} - 1 = \frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}^*_{n,\alpha/2}}{2z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}} - 1$$

Empirical researchers may have a certain tolerance level for D, which is the threshold[5] γ. For example, a researcher may consider it acceptable if $|D| \leq \gamma = 0.25$, i.e. the relative difference of the two C.I.'s is less than a quarter. If $|D|$ goes above this tolerance level, then it is not unreasonable to determine that the identification strength is *weak*.

To summarize, this paper proposes to use the difference D between the bootstrap distribution $F(x)$ and the standard normal distribution $\Phi(x)$ to evaluate the strength of identification. If $|D|$ is greater than a given threshold, identification is *weak*.

To my knowledge, Stock and Yogo (2005) is the first to provide a quantitative definition of weak instruments/identification. Like Stock and Yogo (2005), the quantitative Definition 3 of weak identification is also practically motivated: from the practical perspective, it is not appropriate to consider identification as strong if the bootstrap and the limiting normal distribution provide substantially different confidence intervals. Unlike Stock and Yogo (2005), Definition 3 is not directly related to the bias of the conventional IV/GMM estimator or the size of the t/Wald test, while Stock and Yogo (2005) define identification as weak if the relative bias of the conventional estimator or the size distortion of the t/Wald test exceeds a threshold. Although quantitatively different, the definition in this paper is qualitatively similar to the one in Stock and Yogo (2005): both definitions of weak identification indicate that the distributions of the conventional IV/GMM estimators are poorly approximated by the normal distribution, or said differently, the conventional approximation of (1) breaks down.

2.6 The Bootstrap Principle

Note that *weak identification* refers to the severe disparity between the exact distribution and the normal distribution, while its quantitative definition stated above rests on the disparity between the bootstrap distribution and the normal distribution. In essence, the bootstrap strategy for detecting weak identification is to use the difference between the bootstrap distribution and the normal distribution as a proxy for the the difference between the exact distribution and the normal distribution.

The so-called bootstrap principle (or the bootstrap analogy) in Hall (1992) can help clarify the proposed strategy. The bootstrap principle states that the mapping from the population to the sample (1^{st} mapping) is similar to the mapping from the sample, which is also the bootstrap population, to the bootstrap resample (2^{nd} mapping). By this principle, the identification strength in the 2^{nd} mapping is expected to be

[5] A choice of the threshold γ is suggested in the later part of Monte Carlo studies.

similar to the identification strength in the 1^{st} mapping. Consequently, the bootstrap strategy for detecting weak identification is to use the identification strength in the 2^{nd} mapping as the proxy for the identification strength in the 1^{st} mapping: the substantial disparity between the bootstrap distribution and the normal distribution corresponds to the weak identification strength in the 2^{nd} mapping, which further suggests the weak identification strength in the 1^{st} mapping, i.e. the disparity between the exact distribution and the normal distribution is also severe.

The advantage of the bootstrap strategy is clear: the bootstrap population is the given sample, hence the identification strength in the 2^{nd} mapping is known or recoverable. The only randomness in this mapping comes from the randomness of bootstrap resampling, and if the number of bootstrap replications is sufficiently large, this randomness is negligible. Once the identification strength in the 2^{nd} mapping is recovered, it is used to infer the identification strength in the 1^{st} mapping, since they are expected to be similar by the bootstrap principle.

3 Estimation and Test

With the quantitative definition as well as the advantage of the bootstrap test strategy, detecting weak identification becomes straightforward. By definition, $|D| > \gamma$ implies weak identification. D, the distance between the bootstrap distribution and the standard normal distribution, needs to be estimated.

3.1 Estimation

Draw B i.i.d. observations from $F(x)$ by bootstrap: $X_i = \frac{\hat{\theta}_n^{*i} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$, $i = 1, ..., B$. The bootstrap distribution $F(x)$ can be estimated by the empirical c.d.f. $\hat{F}(x)$ almost surely:

$$\hat{F}(x) = \frac{1}{B}\sum_{i=1}^{B}\mathbf{1}(X_i \leq x) = \frac{1}{B}\sum_{i=1}^{B}\mathbf{1}(\frac{\hat{\theta}_n^{*i} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \leq x) \overset{a.s.}{\to} F(x)$$

Consequently, D can be estimated almost surely, by the continuous mapping theorem:

$$\hat{D} = \frac{\hat{q}_{1-\alpha/2} - \hat{q}_{\alpha/2}}{2z_{1-\alpha/2}} - 1 \overset{a.s.}{\to} D$$

It would be ideal if B could be made infinity. For the given B realizations from $F(x)$, though B can be arbitrarily large, whether $|D|$ exceeds γ needs to be tested, and a test serving this purpose is presented next.

3.2 Test and Its Decision Rule

Assume the following conditions hold for an IV/GMM model with the conventional estimator $\hat{\theta}_n$, associated with standard error $\frac{\hat{\sigma}}{\sqrt{n}}$:

Assumption 1. *There exist $\{\hat{\theta}_n^{*i}, i = 1, ..., B\}$, the i.i.d. draws of the bootstrapped estimator $\hat{\theta}_n^*$; conditional on the sample, the standardized random variable $X = \frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ has a continuous density function $f(x)$ that is non-zero in a neighborhood of the two quantiles $q_{\alpha/2}, q_{1-\alpha/2}$, and can be consistently estimated by the non-parametric kernel estimation: $\hat{f}(x) \xrightarrow{p} f(x)$.*

Comments:

1. Under Assumption 1, the joint distribution of the two quantile estimators, namely, $\hat{q}_{\alpha/2} = \frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$, $\hat{q}_{1-\alpha/2} = \frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$, is asymptotically normal conditional on the sample, as $B \to \infty$ (see David and Nagaraja (2003)):

$$\sqrt{B} \begin{pmatrix} \frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} - q_{\alpha/2} \\ \frac{\hat{\theta}^*_{n,1-\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} - q_{1-\alpha/2} \end{pmatrix} \Rightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega \right)$$

   where

$$\Omega = \begin{pmatrix} \frac{(1-\alpha/2)\alpha/2}{f(q_{\alpha/2})^2} & \frac{(\alpha/2)^2}{f(q_{\alpha/2})f(q_{1-\alpha/2})} \\ \frac{(\alpha/2)^2}{f(q_{\alpha/2})f(q_{1-\alpha/2})} & \frac{(1-\alpha/2)\alpha/2}{f(q_{1-\alpha/2})^2} \end{pmatrix}$$

2. Silverman (1998) provides high level assumptions for the consistency of the non-parametric kernel density estimator, while $\hat{f}(x) \xrightarrow{p} f(x)$ is directly assumed here for simplicity. As $\frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \xrightarrow{p} q_{\alpha/2}$, $f(q_{\alpha/2})$ is consistently estimated by $\hat{f}(\frac{\hat{\theta}^*_{n,\alpha/2} - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}})$; similarly, $f(q_{1-\alpha/2})$ is consistently estimable. The covariance matrix Ω is thus consistently estimable: there exists $\hat{\Omega} \xrightarrow{p} \Omega$. The normal kernel and Silverman's rule of thumb for choosing the bandwidth are used in the empirical application and simulation studies of this paper.

Theorem. *Under Assumption 1, and conditional on the sample, the following result holds as $B \to \infty$:*

$$\sqrt{B}(\hat{D} - D) \quad \Rightarrow \quad N(0, \frac{\Omega_{11} + \Omega_{22} - 2\Omega_{12}}{4z^2_{1-\alpha/2}}) \tag{5}$$

where $\Omega_{i,j}$ is the element of Ω at row i, column j.

The quantitative definition of *strong* and *weak* identification implies the following decision rule: reject the null of *strong* identification when $|D| > \gamma$. There are two cases that would induce the rejection, $D > \gamma$ and $D < -\gamma$. Consequently, *strong* identification is rejected when $\hat{D}$ is significantly greater than γ (Case 1), or significantly less than $-\gamma$ (Case 2) in the test statistics below.

Case 1: *strong* identification is rejected at 5% if

$$b_1 = \sqrt{B}\frac{\hat{D} - \gamma}{[(\hat{\Omega}_{11} + \hat{\Omega}_{22} - 2\hat{\Omega}_{12})/(4z^2_{1-\alpha/2})]^{1/2}} > z_{95\%}$$

Case 2: *strong* identification is rejected at 5% if

$$b_2 = \sqrt{B}\frac{\hat{D} + \gamma}{[(\hat{\Omega}_{11} + \hat{\Omega}_{22} - 2\hat{\Omega}_{12})/(4z^2_{1-\alpha/2})]^{1/2}} < -z_{95\%}$$

Combining these two cases, reject *strong* identification at 5% if

$$|\hat{D}| > \gamma + z_{95\%}\sqrt{\frac{\hat{\Omega}_{11} + \hat{\Omega}_{22} - 2\hat{\Omega}_{12}}{4Bz^2_{1-\alpha/2}}}$$

As $B \to \infty$, the test above ends up with a rule of thumb: reject *strong* identification if $|\hat{D}| > \gamma$, hence this test can be substituted by the rule of thumb under sufficiently large B. From now on, this test is referred to as the b test, since it is based on the bootstrap.

4 IV and Bootstrap

In this section, a linear IV regression model is used as a specific example to further illustrate the bootstrap approach for detecting weak identification, with an application to Card (1995). Most of the analytical results are well known, for example, the convergence results in (1) and (2) under mild conditions, and proofs of the listed results are attached in the appendix. The main objective of this section is to show that the difference between the bootstrap distribution and the normal distribution is a suitable proxy for the difference between the exact distribution and the normal distribution, hence it is a reasonable indicator of the identification strength.

4.1 Model Setup

$$\begin{cases} Y = X\theta + U \\ X = Z\Pi + V \end{cases}$$

$Y = (Y_1, ..., Y_n)', X = (X_1, ..., X_n)'$ are $n \times 1$ vectors of endogenous observations, and $Z = (Z_1, ..., Z_n)'$ is the $n \times k$ matrix of instruments, $k \geq 1$. $U = (u_1, ..., u_n)', V = (v_1, ..., v_n)'$, where the error term $(u_i, v_i)', i = 1, ..., n$, is assumed to have mean zero, and to be i.i.d., homoscedastic with covariance matrix $\Sigma = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix}$. The parameter of interest is θ, and $\hat{\theta}_n$ is the IV estimator of θ:

$$\hat{\theta}_n = (X'P_zX)^{-1}X'P_zY$$

It is central to derive the distribution of $\hat{\theta}_n$ for statistical inference. The exact finite sample distribution of $\hat{\theta}_n$, however, is unknown without making further distributional assumptions. Instead, two alternative methods are often used to approximate the exact distribution in econometric applications: the limiting normal distribution, and the bootstrap distribution.

4.2 First-Order Asymptotics

Under the conventional asymptotic theory where the $k \times 1$ vector Π is modeled as non-zero and fixed, the IV estimator $\hat{\theta}_n$ is asymptotically normally distributed as the sample size n gets large. In contrast, to explore the distribution of $\hat{\theta}_n$ when the instruments are only weakly related to the endogenous variable, Staiger and Stock (1997) develop weak instrument asymptotics, i.e. Π is modeled as local to zero.

Assumption 2. *(a)* $\Pi = \Pi_0 \neq 0$, *and* Π_0 *is fixed; (a')* $\Pi = \Pi_n = \frac{C}{\sqrt{n}}$, *and* C *is fixed.*

The asymptotics under Assumption 2(a) are called *Strong Instrument Asymptotics*, and the asymptotics under Assumption 2(a') are called *Weak Instrument Asymptotics*. (a)(a') are two alternative rank conditions, and in (a') the rank condition is only weakly satisfied. The following results and notations are used: $Z'Z/n \xrightarrow{p} Q_{zz} \equiv E(Z_i'Z_i)$, $(\frac{Z'U}{\sqrt{n}}, \frac{Z'V}{\sqrt{n}}) \Rightarrow (\Psi_{zu}, \Psi_{zv})$, and $(\Psi_{zu}, \Psi_{zv})'$ is distributed $N(0, \Sigma \otimes Q_{zz})$. The validity of these results follows from law of large numbers and a central limit theorem, after assuming the existence of second moments. By the similar derivation as in Staiger and Stock (1997), the following two well-known results hold.

Under *Strong Instrument Asymptotics*:

$$\sqrt{n}(\hat{\theta}_n - \theta) \;\; \Rightarrow \;\; (\Pi_0' Q_{zz} \Pi_0)^{-1} \Pi_0' \Psi_{zu} \sim N(0, (\Pi_0' Q_{zz} \Pi_0)^{-1} \sigma_u^2) \tag{6}$$

Under *Weak Instrument Asymptotics*:

$$\hat{\theta}_n - \theta \;\; \Rightarrow \;\; [(Q_{zz} C + \Psi_{zv})' Q_{zz}^{-1} (Q_{zz} C + \Psi_{zv})]^{-1} (Q_{zz} C + \Psi_{zv})' Q_{zz}^{-1} \Psi_{zu} \tag{7}$$

If $k = 1$, i.e. the model is exactly identified, then (7) reduces to:

$$\hat{\theta}_n - \theta \;\; \Rightarrow \;\; (Q_{zz} C + \Psi_{zv})^{-1} \Psi_{zu}$$

The conventional result of (6) indicates that when instruments are strong, the IV estimator $\hat{\theta}_n$ is both consistent and asymptotically normally distributed. In contrast, the result of (7) indicates that the IV estimator $\hat{\theta}_n$ is neither consistent nor asymptotically normally distributed, if the rank condition is weak. As the magnitude of Π increases, however, the distribution of $\hat{\theta}_n$ in (7) gets closer to the normal distribution in (6). In the extreme case that $C = \sqrt{n}\Pi_0$, (6) and (7) coincide. Π, the vector of nuisance parameters, is thus the driving force of the linear IV regression model: it determines whether θ can be consistently estimated, and whether the estimator $\hat{\theta}_n$ can be well approximated by the normal distribution.

As a function of Π, the concentration parameter μ^2 is a unit-less measure of the identification strength in the studies of weak instruments:

$$\mu^2 = \frac{\Pi' Z' Z \Pi}{\sigma_v^2}$$

The greater μ^2, the stronger the identification of the parameter θ, and the distribution of $\hat{\theta}_n$ gets closer to the normal distribution, as shown in Rothenberg (1984). In addition, Stock and Yogo (2005) suggest that there is a threshold of the concentration parameter for the set of weak instruments, i.e. instruments as well as identification are considered weak if μ^2/k is under the threshold. The first stage F test is suggested in Stock and Yogo (2005) to check whether the threshold is exceeded: if the F statistic is greater than the tabled critical values, typically around 10 for small k, then instruments as well as identification are not weak:

$$F = \frac{\hat{\Pi}_n' Z' Z \hat{\Pi}_n / k}{\hat{\sigma}_v^2}$$

where $\hat{\Pi}_n = (Z'Z)^{-1}Z'X$, $\hat{\sigma}_v^2 = (X - Z\hat{\Pi}_n)'(X - Z\hat{\Pi}_n)/(n-k)$.

4.3 Residual Bootstrap

As an alternative to the limiting normal distribution, the bootstrap provides another way of approximating the distribution of $\hat{\theta}_n$. For the linear IV regression model under homoscedasticity, the residual bootstrap is a commonly used bootstrap method. See, for example, Moreira *et al.* (2009). This bootstrap procedure or algorithm is described as follows.

1. $\hat{U}, \hat{V}$ are the residuals induced by $\hat{\theta}_n, \hat{\Pi}_n$ in the linear IV model:

$$\hat{U} = Y - X\hat{\theta}_n, \ \hat{V} = X - Z\hat{\Pi}_n$$

2. Re-center $\hat{U}, \hat{V}$ to yield $\tilde{U}, \tilde{V}$, by pre-multiplying a constant matrix M_e, where $M_e = I_n - P_e$, and e is the n by 1 vector of ones:

$$\tilde{U} = M_e\hat{U}, \ \tilde{V} = M_e\hat{V}$$

3. Sampling the rows of $(\tilde{U}, \tilde{V})$ and Z independently n times with replacement, and let (U^*, V^*) and Z^* denote the outcome. The dependent variables (X^*, Y^*) are generated by:

$$\begin{cases} Y^* = X^*\hat{\theta}_n + U^* \\ X^* = Z^*\hat{\Pi}_n + V^* \end{cases}$$

4. As the counterpart of the IV estimator $\hat{\theta}_n$, the bootstrapped IV estimator $\hat{\theta}_n^*$ is computed by the bootstrap resample (X^*, Y^*, Z^*):

$$\hat{\theta}_n^* \quad = \quad (X^{*'}P_{z^*}X^*)^{-1}X^{*'}P_{z^*}Y^*$$

5. Re-do Steps 2-4 B times, and $\{\hat{\theta}_n^{*i}, i = 1, ..., B\}$ are B i.i.d. estimators.

The bootstrap data generation process (D.G.P.) above aims to mimic the D.G.P. of the linear IV regression model: when instruments are strong, the equation $\hat{\Pi}_n = \Pi + O_p(n^{-1/2})$ indicates that $\hat{\Pi}_n$ is not substantially different from Π; in addition, the variance of the bootstrap error term (u_i^*, v_i^*) converges to Σ, the variance of (u_i, v_i).

Consequently, it is natural to expect that the mimicking process works well under strong instruments, and the distributions of $\hat{\theta}_n^*$ and $\hat{\theta}_n$ are alike. This conjecture on the bootstrapped estimator $\hat{\theta}_n^*$ is confirmed by the result below.

Under *Strong Instrument Asymptotics*:

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \quad \Rightarrow \quad (\Pi_0' Q_{zz} \Pi_0)^{-1} \Pi_0' \Psi_{zu} \tag{8}$$

The result of (8) motivates the usage of the bootstrap as a tool to detect the identification strength: under strong identification, the distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ is asymptotically identical to the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$, and the asymptotic distribution is normal; if the distribution of $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$ is found to be substantially different from normal, then it indicates that identification is weak.

The bootstrap mimicking process also helps explain why the bootstrap becomes problematic when instruments are weak: first of all, if Π is local to zero, then the connection $\hat{\Pi}_n = \Pi + O_p(n^{-1/2})$ implies that the difference between $\hat{\Pi}_n$ and Π becomes substantial, hence the identification strength in the bootstrap resample is substantially different from the identification strength in the sample; secondly, when $\hat{\theta}_n$ does not consistently estimate θ under weak instruments, the residual $\hat{U}$ does not converge to the error term U, since $\hat{U} = U - X(\hat{\theta}_n - \theta)$. Both of these two facts contribute to that the bootstrap D.G.P. does not mimic the D.G.P. of the linear IV model well under weak instruments, and consequently, the distributions of $\hat{\theta}_n^*$ and $\hat{\theta}_n$ are different.

4.4 Comparison of μ^2, μ^{2*}

To compare the identification strength in the bootstrap resample (X^*, Y^*, Z^*) with the identification strength in the sample (X, Y, Z), consider the concentration parameter μ^{2*}, the bootstrap counterpart of μ^2:

$$\mu^{2*} = \frac{\hat{\Pi}_n' Z^{*\prime} Z^* \hat{\Pi}_n}{\sigma_v^{2*}}, \text{where } \sigma_v^{2*} = \frac{V^{*\prime} V^*}{n}$$

Under *Strong Instrument Asymptotics*:

$$\mu^2 \to \infty, \text{and } \mu^{2*} \to \infty \tag{9}$$

Under *Weak Instrument Asymptotics*:

$$\mu^2 \xrightarrow{p} \frac{C'Q_{zz}C}{\sigma_v^2}, \text{and } \mu^{2*} \Rightarrow \frac{C'Q_{zz}C + 2C'\Psi_{zv} + \Psi_{zv}'Q_{zz}^{-1}\Psi_{zv}}{\sigma_v^2} \tag{10}$$

The result of (9) states that when identification is strong, μ^2 and μ^{2*} go to infinity, implying that the distributions of $\hat{\theta}_n$ and $\hat{\theta}_n^*$ are both asymptotically normal, as stated in (6) and (8). On the contrary, the result of (10) states that when identification is weak, μ^{2*} and μ^2 do not go to infinity, and are asymptotically different, which implies that the distributions of $\hat{\theta}_n$ and $\hat{\theta}_n^*$ are both asymptotically non-normal, and their asymptotical distributions are not identical. The asymptotic difference in μ^2 and μ^{2*} is $\frac{2C'\Psi_{zv}+\Psi_{zv}'Q_{zz}^{-1}\Psi_{zv}}{\sigma_v^2}$: under homoscedasticity, it has mean k.

Table 1 reports the difference between μ^2 and μ^{2*} when $k = 1$ by Monte Carlo studies: the relative difference is found to be substantial when μ^2 is small, and negligible when μ^2 is large; overall, μ^{2*} is greater than μ^2. Another interpretation of (10) is that, loosely speaking, the F statistic of the bootstrap resample is above the F statistic of the sample by 1, because of the relation $E(F) \approx \mu^2/k + 1$ in Stock *et al.* (2002).

The comparison of μ^{2*} and μ^2 indicates a useful result in the linear IV model where μ^2/k and μ^{2*}/k are measures of the identification strength: on average, the identification strength in the bootstrap resample is similar to, and slightly stronger (plus 1) than the identification strength in the sample; consequently, if the identification strength in the bootstrap resample is weak, then the identification strength in the sample must also be weak. In this sense, the proposed bootstrap strategy for detecting weak identification is conservative.

4.5 Edgeworth Expansion

The Edgeworth expansion provides another look at the bootstrap strategy for detecting weak identification. For the purpose of this paper, it suffices to consider the two-term expansion of the standardized estimator $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$, where $\sigma^2 = (\Pi_0'Q_{zz}\Pi_0)^{-1}\sigma_u^2$.

Define $R_i = (Z_iX_i, Z_iY_i, vec(Z_i'Z_i)')'$, $\mu = E(V_i) = (Q_{zz}\Pi, Q_{zz}\Pi\beta, vec(Q_{zz})')'$. Rewrite $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ in the form of $\sqrt{n}A(\bar{R})$, where $\bar{R} = \frac{1}{n}\sum_{i=1}^n R_i$, and $A(\mu) = 0$. The following result is the application of the smooth function model and Theorem 2.2 in Hall (1992). A similar result is available in Moreira *et al.* (2009).

Theorem. *Under Assumption 2(a), and assume two conditions: (i)* $E(||R_i||^3) < \infty$, *(ii)* $lim\ sup_{||t||\to\infty}|Eexp(it'R_i)| < 1$, $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ *and its bootstrap counterpart admit two-term*

Edgeworth expansions uniformly in x:

$$P(\frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} \leq x) = \Phi(x) + n^{-1/2}p(x)\phi(x) + o(n^{-1/2}) \quad (11)$$

$$P(\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \leq x) = \Phi(x) + n^{-1/2}p^*(x)\phi(x) + o(n^{-1/2}) \quad (12)$$

where $p(x)$ is a polynomial of degree 2, with coefficients depending on θ, Π, and moments of R_i up to order 3, $p^*(x)$ is the bootstrap counterpart of $p(x)$ with coefficients depending on $\hat{\theta}_n, \hat{\Pi}_n$, and moments of R_i^*.

Figure 2 is drawn based on the results of the Edgeworth expansion. The distance between the exact distribution of $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ and the standard normal distribution has order $O(n^{-1/2})$, which is the same as the order of the distance between the bootstrap distribution of $\frac{\hat{\theta}_n^*-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ and the standard normal distribution. Compared with the normal distribution, the bootstrap distribution has the well known property of asymptotic refinement: it is closer to the the exact distribution, as the distance has order $O(n^{-1})$, which results from $p^*(x) - p(x) = O(n^{-1/2})$: $\hat{\theta}_n, \hat{\Pi}_n$, and moments of R_i^* approach θ, Π, and moments of R_i at rate $n^{-1/2}$.

As discussed above, the bootstrap strategy for detecting weak identification is to use the distance between the bootstrap distribution and the normal as a proxy of the distance between the exact distribution and the normal distribution. This strategy is supported by the Edgeworth expansion in two ways: firstly, these two distances have the same order $O(n^{-1/2})$; secondly, the price of using the proxy is low, i.e. the proxy error is $O(n^{-1})$, because $[P(\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}} \leq x) - \Phi(x)] - [P(\frac{\hat{\theta}_n^*-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}} \leq x) - \Phi(x)] = O(n^{-1})$.

To summarize, the bootstrap distribution is a good proxy for the exact distribution for the purpose of this paper. Under the null hypothesis of strong identification, the disparity between the exact distribution and the normal distribution is reflected by the disparity between the bootstrap distribution and the normal distribution. If the bootstrap distribution is found to be substantially different from normal, then it is appropriate to conclude that the exact distribution is substantially different from normal as well, hence the identification strength is weak.

4.6 Heteroscedasticity and Pair Bootstrap

So far, the discussion is restricted to the linear IV model under the homoscedasticity assumption. However, it is well understood that this assumption is unlikely to hold in practice: for example, $E(u_i^2|Z_i)$ may not be constant, but depend on Z_i. As a result, it is common that empirical researchers need to take the existence of heteroscedasticity into consideration.

Once the homoscedasticity assumption is loosened, the validity of the popular F test for detecting weak instruments/identification is under doubt: the way in which Stock and Yogo (2005) derive the critical values of the F test crucially depends on the homoscedasticity assumption. Consequently, it is not clear whether these critical values can still be used when heteroscedasticity instead of homoscedasticity takes place.

The validity of the bootstrap test approach, on the contrary, stays unaffected, when the pair bootstrap replaces the residual bootstrap. Freedman (1984) shows that the pair bootstrap of the IV estimator remains valid under heteroscedasticity: the IV estimator and its bootstrap counterpart asymptotically have the same normal distribution under strong instruments. The idea of the pair bootstrap is to directly resample the data. In the special case of $k = 1$, the pair bootstrap is to draw the bootstrap resample (X^*, Y^*, Z^*) from the empirical distribution of (X, Y, Z). Compared with the residual bootstrap above, the pair bootstrap is non-parametric, and it preserves the possible heteroscedastic relations in the IV model, as proved in Freedman (1984). Consequently, if heteroscedasticity is concerned, the bootstrap test procedure involves two minor modifications: (i) use the pair bootstrap to compute $\hat{\theta}_n^*$; (ii) use a heteroscedasticity-robust estimator for $\hat{\sigma}$.

4.7 An Empirical Example: Card (1995)

In this section, an empirical application is investigated to illustrate the bootstrap approach for detecting weak identification. The same data as in Card (1995) is used. By employing the IV approach, Card (1995) answers the following question: what is the return to education? Or specifically, how much more can an individual earn if he/she completes an extra year of schooling?

The dataset is ultimately taken from the National Longitudinal Survey of Young Men between 1966-1981 with 3010 observations, and there are two variables in the dataset that measure college proximity: $nearc2$ and $nearc4$, both are dummy variables, and are 1 if there is a 2-year, 4-year college in the local area respectively. See Card (1995) for the detailed description of the data. To identify the return to education, Card (1995) considers a structural wage equation as follows:

$$lwage = \alpha + \theta edu + W'\beta + u$$

where $lwage$ is the log of wage, edu is the years of schooling, the covariate vector W contains the control variables, and u is the error term. Among the set of parameters (α, θ, β'), θ measuring the return to education is of interest.

In the basic specification, Card (1995) uses five control variables: experience, the square of experience, the dummy for race, the dummy for living in the south, and the dummy for living in the standard metropolitan statistical area (SMSA). To bypass the issue that experience is also endogenous, Davidson and MacKinnon (2010) replace experience, the square of experience with age, the square of age. Following Davidson and MacKinnon (2010), I used age, square of age, and the three dummy variables as control variables, hence edu is the only endogenous regressor. In addition, an extra instrument $nearc2 \wedge nearc4$ is constructed: $nearc2 \wedge nearc4$ is a dummy variable, and is 1 if there are both a 2-year and a 4-year college in the local area. Unlike Davidson and MacKinnon (2010), I use the three instruments, $nearc2$, $nearc2 \wedge nearc4$, $nearc4$, one by one as the single instrument for edu to better illustrate the approach discussed in this paper, while Davidson and MacKinnon (2010) simultaneously use more than one instrument.

The identification strength under the three potential IV's is examined by the first stage F test in Stock and Yogo (2005): if $nearc2$ is used as the IV, $F \approx 0.54$; if $nearc2 \wedge nearc4$ is used as the IV, $F \approx 6.98$; if $nearc4$ is used as the IV, $F \approx 10.22$. According to the rule of thumb $F > 10$ suggested in Stock and Yogo (2005), these F statistics suggest that $nearc4$ is a strong IV, $nearc2$, $nearc2 \wedge nearc4$ are not. Table 2 reports the point estimate and 95% confidence interval of θ under each of these instruments. If the point estimate of the return to education is of interest, 0.0936 derived by $nearc4$ is more reliable, compared with the other point estimates. Based on these empirical results, an additional year of education increases the wage by about 9.36%; however, the possibility that this effect is zero can not be rejected at 95%.

As proposed in this paper, the bootstrap can help evaluate the identification strength. To allow for heteroscedasticity, the pair bootstrap is employed, and the IV estimator of θ is computed $B = 9999$ times, using the three instruments one by one: specifically, the bootstrap resample is directly drawn from the sample with replacement, and the size of the resample is equal to the number of observations; the bootstrapped IV estimator is computed by each bootstrap resample, and this process is replicated B times. To make it comparable to the standard normal variate, the bootstrapped estimator is standardized by subtracting the IV estimate and dividing by the standard error of the IV estimator, where the standard error is computed in the way of White (1980). The p.d.f and quantiles of the bootstrapped IV estimator after standardization are plotted against the standard normal variate in Figure 3. Figure 3 shows that, the bootstrap distribution is closer to the normal distribution, when the instrument is stronger. The figure of the bootstrap distribution and the Q-Q plot[6] hence are useful tools to help detect weak identification, since they provide empirical researchers a graphic evaluation of the identification strength.

Table 2 reports the 95% C.I. of the return to education by the bootstrap percentile method, in addition, the identification-robust conditional likelihood ratio (CLR)[7] test by Moreira (2003) is also applied to construct a C.I. for comparison. Instead of the F test, the proposed b test is applied to determine whether the strength of identification is strong or weak, with the tentative threshold $\gamma = 0.25$.

$nearc2$: the bootstrap C.I. of the return to education is found to be $(-4.3279, 4.8868)$, while the C.I. by t/Wald is $(-0.8188, 1.8346)$; the relative difference is significantly larger than the threshold with $\hat{D} \approx 2.46 > \gamma$, $b_1 \approx 12.09 > z_{95\%}$, hence the null of strong identification under $nearc2$ is rejected.

$nearc2 \wedge nearc4$: the bootstrap C.I. of the return to education is found to be $(0.0019, 0.4664)$, while the C.I. by t/Wald is $(-0.0099, 0.2692)$; the relative difference is significantly larger than the threshold with $\hat{D} \approx 0.66 > \gamma$, $b_1 \approx 6.36 > z_{95\%}$, hence the null of strong identification under $nearc2 \wedge nearc4$ is rejected.

$nearc4$: the bootstrap C.I. of the return to education is found to be $(0.0034, 0.2579)$, while the C.I. by t/Wald is $(-0.0027, 0.1899)$; the relative difference is significantly larger than the threshold with $\hat{D} \approx 0.32 > \gamma$, $b_1 \approx 2.71 > z_{95\%}$, hence the null of strong identification under $nearc4$ is rejected.

To conclude, both the F test and the bootstrap based b test can detect the weak instrument $nearc2$, $nearc2 \wedge nearc4$ and support the view that $nearc4$ is the strongest among the three potential instruments. The difference is, the b test considers $nearc4$ as weak, while the F test treats $nearc4$ as strong (but just across the threshold). It is surprising that, although $F > 10$ holds under $nearc4$, the relative difference of the C.I. by t/Wald and the bootstrap C.I. is as large as 0.32. It thus indicates that the $F > 10$ rule is not strict enough, i.e. the disparity between the two C.I.'s is still severe, although $F > 10$ holds in this example.

The b and F tests are proposed on different grounds, i.e. the b test is based on the comparison of the length of confidence intervals, while the F test is based on the threshold of the rank condition, hence it is not surprising that these two tests can imply

[6]Compared to the p.d.f., the Q-Q plot is known to be a better way of comparing distributions.

[7]Note: The IV model under consideration is just identified, hence CLR is equivalent to the AR test in Anderson and Rubin (1949) and K test in Kleibergen (2002).

different outcomes. In contrast with the F test, the b test has its advantages: firstly of all, it provides a graphic view of the identification strength, i.e. the Q-Q plot in Figure 3 shows that the bootstrap distribution is not very close to the normal distribution, hence the identification strength appears weak; secondly, unlike the F test, the b test does not rely on the restrictive homoscedasticity assumption, and has the potential of being extended to the generalized GMM framework.

5 Simulation

This section presents Monte Carlo results to evaluate the power of the proposed b test. The disparity between the two intervals, (3) and (4), is reported. The threshold γ is calibrated to the $F > 10$ rule.

5.1 Settings

The linear IV model described in Section 4.1 is employed in the D.G.P., with the following choice of parameters: $\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$, where $\rho \in \{0.99, 0.50, 0.01\}$ to introduce high, moderate and low degrees of endogeneity, respectively; $\theta = 0$; $z_i \sim NID(0,1)$, $i = 1, ..., n$, and $n = 1000$.

A sequence of μ^2, $\mu^2 \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 40, 60, 80\}$, is chosen by assigning different values to Π. For each μ^2, the data of (X, Y, Z) is generated by the linear IV model with the parameters specified above, and the bootstrapped IV estimator $\hat{\theta}_n^*$ is computed $B = 9999$ times by the residual bootstrap. The number of replications equals 1000.

5.2 Results

The results of the Monte Carlo studies are reported in Table 3.

C_0 **vs.** C_t: C_0 denotes the interval derived by taking the 2.5%, 97.5% quantiles of $\hat{\theta}_n$. C_t is the 95% C.I. derived by inverting the t test. Table 3 reports the relative difference in lengths of these two intervals: the median absolute difference in the lengths of C_0, C_t weighted by the length of C_t is reported. The departure of C_t from C_0 is more severe for larger ρ, and as expected, the departure of C_t from C_0 shrinks as μ^2 increases. It is found that the relative difference can be as high as 0.62 when $\mu^2 = 10$. This is a bit surprising since the identification strength under $\mu^2 = 10$ is generally not considered as very weak: as reported in the table, there is about half chance that the F test will consider the identification under $\mu^2 = 10$ as *strong*. In other words, although the F test and its decision rule may treat the identification strength under $\mu^2 = 10$ as *strong*, the departure of C_t from C_0 can still be severe.

C_b **vs.** C_t: the 95% C.I. C_b by the bootstrap is compared with C_t. Table 3 reports the median absolute value of the relative difference in the lengths of C_b, C_t. As expected, the difference shrinks as μ^2 increases. In particular, when $\mu^2 > 10$, the relative difference does not exceed 0.50 in absolute value; when $\mu^2 \geq 20$, the relative difference does not exceed 0.25 in absolute value; when $\mu^2 \geq 80$, the difference is negligible.

b **test**: the proposed b test is applied to test the null of strong identification, and the percentage of rejecting strong identification is reported. Two tentative cutoffs, $\gamma = 0.25, 0.50$ are considered. Under the stricter rule of $\gamma = 0.25$, the b test rejects the null more often.

F **test**: the F test and its decision rule in Stock and Yogo (2005) are applied to provide a benchmark for the b test. Table 3 reports the percentage of concluding weak identification by the F test (the frequency of the F statistic is less than its critical value) for different ρ's. As discussed above, the F test examines whether μ^2/k exceeds the cutoff, hence it does not depend on ρ, the degree of endogeneity.

By comparing the performance of the b test with the F test, $\gamma = 0.25$ appears to be a reasonable choice. It corresponds to μ^2 around 10 with endogeneity close to zero. With this threshold, the frequency of concluding weak identification by the b test is comparable to the F test, although b rejects strong identification slightly more often. This paper hence suggests $\gamma = 0.25$ as the quantitative threshold for distinguishing strong and weak identification, and this simple rule is not off the mark for practical reasons.

6 Conclusion

This paper suggests that the bootstrap is a useful tool for detecting weak identification in IV/GMM applications. The distinguishing feature of the bootstrap based approach is that it provides a graphic view of the identification strength. By eyeballing the graph of the bootstrap distribution, and comparing it with the normal distribution through the Q-Q plot, empirical researchers can evaluate whether or not weak identification exists. The underlying reason is simple: strong identification implies the bootstrap distribution is close to, and asymptotically identical to the normal distribution.

A quantitative threshold for distinguishing *strong* and *weak* identification is suggested based on the comparison of two C.I.'s for the parameter of interest: the C.I. by inverting the t/Wald test and the bootstrap percentile C.I.. The difference of these two C.I.'s boils down to the difference between the bootstrap distribution and the normal distribution, and exceeding the threshold implies that the relative difference of the two C.I.'s is at least as large as a quarter. For practical purposes, the identification strength is considered as *weak* in this paper once this threshold is exceeded. Monte Carlo experiments show that this threshold is comparable to and slightly stricter than the $F > 10$ rule of thumb in Stock and Yogo (2005). Even in the i.i.d. and homoscedasticity setting, $F > 10$ is found to be not strict enough, when it comes to the comparison of C.I.'s: the relative difference of the two commonly used C.I.'s named above can be very large, even when $F > 10$ holds; an application to Card (1995) also makes the same point.

References

Anderson, T. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, pp. 46–63.

Angrist, J. and Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, pp. 979–1014.

Bravo, F., Escanciano, J. and Otsu, T. (2009). Testing for Identification in GMM under Conditional Moment Restrictions, working Paper.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling, Aspects of labour market behaviour: essays in honour of John Vanderkamp. *ed. L.N. Christofides, E.K. Grant, and R. Swidinsky.*

David, H. and Nagaraja, H. (2003). *Order statistics*. Wiley-Interscience.

Davidson, R. and MacKinnon, J. (2010). Wild bootstrap tests for IV regression. *Journal of Business and Economic Statistics*, **28** (1), 128–144.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, pp. 1–26.

Freedman, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *The Annals of Statistics*, **12** (3), 827–842.

Hahn, J. and Hausman, J. (2002). A new specification test for the validity of instrumental variables. *Econometrica*, **70** (1), 163–189.

Hall, P. (1992). *The bootstrap and Edgeworth expansion*. Springer Verlag.

— and Horowitz, J. (1996). Bootstrap critical values for tests based on generalized method of moments estimators. *Econometrica*, **64** (4), 891–916.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pp. 1029–1054.

Horowitz, J. (1994). Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, **61** (2), 395–411.

— (2001). The Bootstrap. Handbook of Econometrics, Vol. 5.

Inoue, A. and Rossi, B. (2008). Testing for weak identification in possibly nonlinear models, working paper.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, pp. 1781–1803.

— (2005). Testing parameters in GMM without assuming that they are identified. *Econometrica*, pp. 1103–1123.

Moreira, M. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, pp. 1027–1048.

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

Rothenberg, T. (1984). Approximating the distributions of econometric estimators and test statistics. *Handbook of econometrics*, **2**, 881–935.

Silverman, B. (1998). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, pp. 557–586.

Stock, J. and Wright, J. (2000). GMM with weak identification. *Econometrica*, pp. 1055–1096.

—, — and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, **20** (4), 518–529.

— and Yogo, M. (2005). Testing for weak instruments in linear IV regression, Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. *ed. DW Andrews and JH Stock*, pp. 80–108.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, **48** (4), 817–838.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. MIT press.

Wright, J. (2002). Testing the Null of Identification in GMM. *International Finance Discussion Papers*, **732**.

— (2003). Detecting lack of identification in GMM. *Econometric Theory*, **19** (02), 322–330.

Figure 1: Three Related Distributions

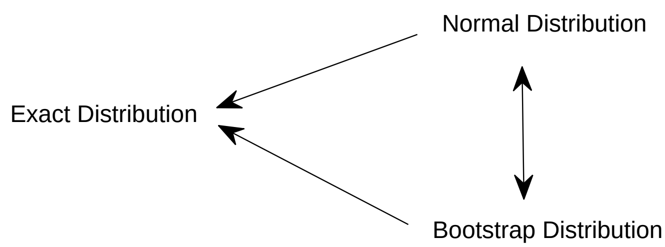
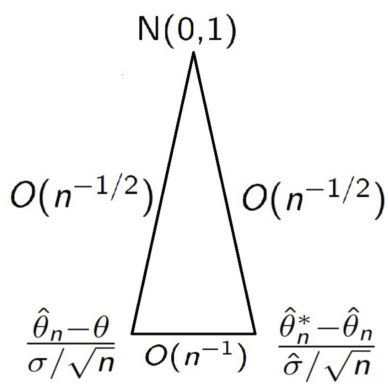


Figure 2: Expressions and Distances



Notes: The normal distribution and the bootstrap distribution are two approximations to the exact distribution of IV/GMM estimators. If the exact distribution of $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ is well approximated by the normal distribution $N(0,1)$, then the bootstrap distribution of $\frac{\hat{\theta}^*_n-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ is well approximated by $N(0,1)$ as well, because of the same magnitude $O(n^{-1/2})$ of the approximation error. A substantial difference between $\frac{\hat{\theta}^*_n-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}$ and $N(0,1)$ indicates the difference between $\frac{\hat{\theta}_n-\theta}{\sigma/\sqrt{n}}$ and $N(0,1)$ is also substantial, hence provides the evidence of weak identification.

Table 1: A Monte Carlo study of μ^2 and μ^{2*}

	μ^2												
	2	4	6	8	10	12	14	16	18	20	40	60	80
μ^{2*}	3.1	5.1	7.1	9.1	11.2	13.2	15.2	17.2	19.2	21.2	41.4	61.5	81.6

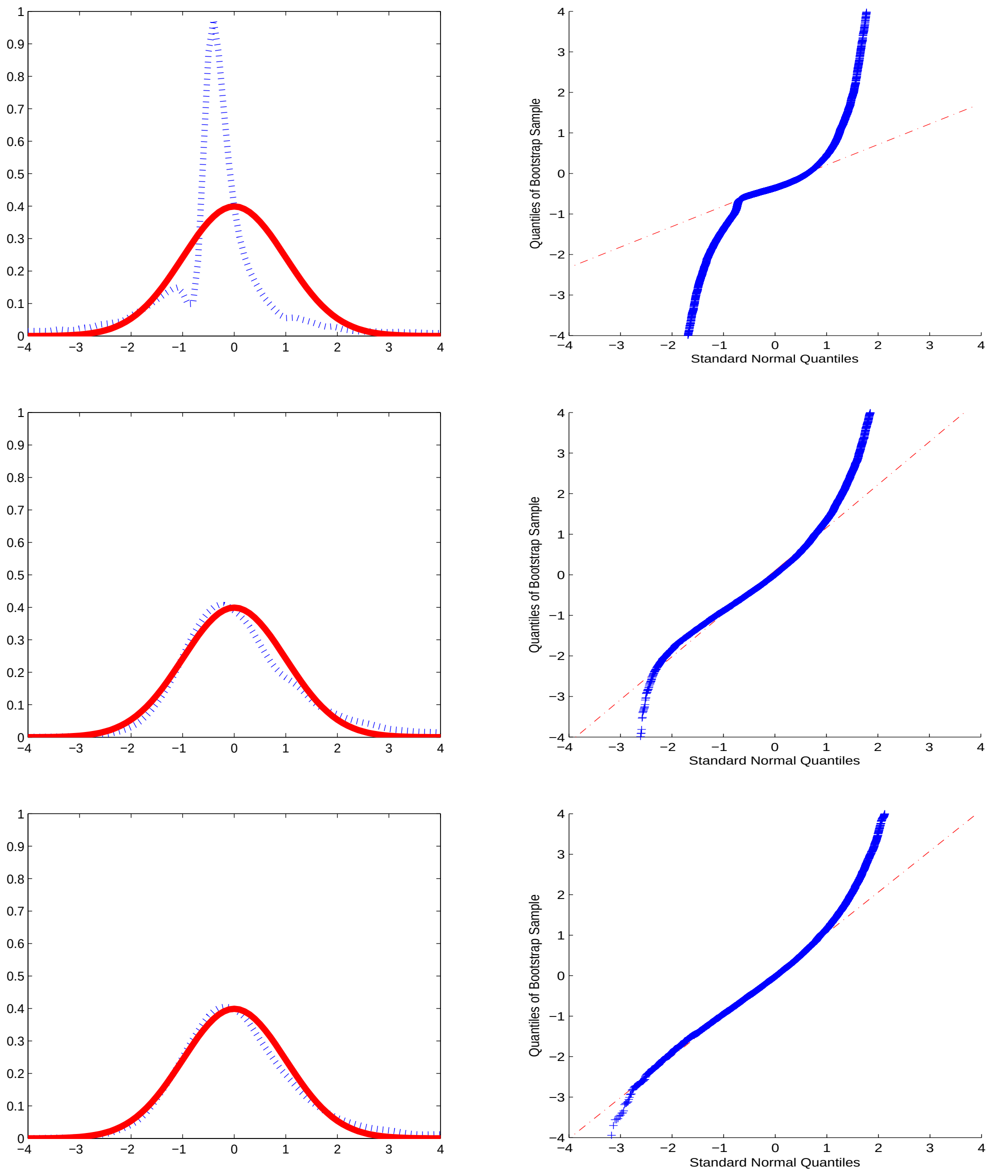
Notes: This table compares the concentration parameter μ^2 with μ^{2*}, the bootstrap counterpart of μ^2, by a Monte Carlo study. For each μ^2, the data of X, Z, V are generated by $x_i = z_i\Pi + v_i$, where: (i) x_i, z_i, v_i are the i^{th} elements of X, Z, V; (ii) $z_i \sim NID(0,1)$, $v_i \sim NID(0,1)$, $i = 1, ..., 1000$; (iii) Π is determined by the value of μ^2. The reported μ^{2*} is the sample average of 1000 replications; in each replication, the residual bootstrap is conducted 1000 times.

Table 2: Return to Education

IV	$nearc2$	$nearc2 \wedge nearc4$	$nearc4$
F statistic	0.54	6.98	10.22
$\hat{\theta}_n$	0.5079	0.1297	0.0936
95% C.I. by t/Wald	(-0.8188, 1.8346)	(-0.0099, 0.2692)	(-0.0027, 0.1899)
by bootstrap	(-4.3279, 4.8868)	(0.0019, 0.4664)	(0.0034, 0.2579)
by CLR	$(-\infty, -0.1750] \cup [0.0867, +\infty)$	$[0.0133, 0.5253]$	[0.0009, 0.2550]

Notes: This table presents the estimate $\hat{\theta}_n$ and confidence interval for return to education using the data of Card (1995). The first stage F statistic is reported for the three instrumental variables, $nearc2$, $nearc2 \wedge nearc4$, $nearc4$, which are used one by one for the endogenous years of schooling. The included control variables are age, square of age, and three dummy variables for race, living in the south, living in the SMSA.

Figure 3: The bootstrap distribution, p.d.f and Q-Q plot



Notes: The p.d.f. and Q-Q plot of the bootstrap distribution are presented, under three instrumental variables in the application of Card (1995). 1^{st} **row:** *nearc*2 as IV; 2^{nd} **row:** $nearc2 \wedge nearc4$ as IV; 3^{rd} **row:** *nearc*4 as IV; **Left:** p.d.f of the bootstrapped IV estimator after standardization (dotted) against the standard normal (solid); **Right:** the Q-Q plot. 9999 bootstrap replications are conducted.

Table 3: The performance of b, F for detecting weak identification

	μ^2												
	2	4	6	8	10	12	14	16	18	20	40	60	80
Comparison of C.I.													
C_0 vs. C_t:													
$\rho = 0.99$	5.24	1.64	1.37	0.80	0.62	0.60	0.50	0.41	0.38	0.37	0.22	0.18	0.14
0.50	2.62	1.10	0.83	0.62	0.41	0.38	0.33	0.29	0.29	0.26	0.15	0.13	0.11
0.01	1.90	0.95	0.53	0.39	0.32	0.27	0.23	0.21	0.21	0.19	0.12	0.10	0.08
C_b vs. C_t:													
$\rho = 0.99$	1.87	1.61	1.10	0.86	0.61	0.45	0.37	0.33	0.27	0.23	0.11	0.07	0.05
0.50	1.56	1.06	0.72	0.47	0.37	0.29	0.22	0.19	0.16	0.14	0.07	0.05	0.03
0.01	1.45	0.97	0.61	0.40	0.29	0.22	0.18	0.15	0.13	0.12	0.05	0.04	0.03
Rejection freq. of b													
$\gamma = 0.50$:													
$\rho = 0.99$	90.1	84.3	75.9	65.4	53.9	40.3	30.4	22.9	14.4	10.1	0.1	0.0	0.0
0.50	83.3	69.8	57.7	43.5	34.0	24.3	16.9	13.6	9.0	5.7	0.0	0.0	0.0
0.01	81.4	68.2	52.3	38.3	27.4	18.2	12.3	7.7	4.7	3.0	0.0	0.0	0.0
$\gamma = 0.25$:													
$\rho = 0.99$	96.4	96.0	94.4	89.7	81.7	73.3	63.1	56.6	44.7	35.5	1.6	0.0	0.0
0.50	93.6	87.5	82.6	67.6	60.3	49.5	37.9	29.0	23.8	17.9	0.6	0.0	0.0
0.01	92.2	87.1	77.2	62.2	50.2	39.2	29.5	22.6	15.7	11.5	0.2	0.0	0.0
Rejection freq. of F													
$\rho = 0.99$	94.1	82.7	70.5	53.0	43.3	32.1	22.6	16.1	12.5	7.9	0.0	0.0	0.0
0.50	93.9	82.1	66.3	56.9	42.9	32.7	22.4	15.7	12.5	8.2	0.0	0.0	0.0
0.01	93.4	83.7	69.4	54.9	45.7	33.3	22.3	15.2	11.6	7.4	0.0	0.0	0.0

Notes: This table presents the Monte Carlo results of comparing the bootstrap based b test proposed in this paper with the F test of Stock and Yogo (2005). The frequencies of concluding weak identification by the b test is reported for each μ^2, and the frequencies of concluding weak identification by the F test is also reported. μ^2 is the concentration parameter, and the greater μ^2 is, the stronger the strength of identification; ρ is the correlation coefficient, and the greater ρ is, the stronger the endogeneity; C_0, C_t and C_b are the C.I. derived by taking the 2.5%, 97.5% quantiles of $\hat{\theta}_n$, the 95% C.I. by inverting t, the 95% C.I. by the bootstrap respectively, and their relative difference in length (the median absolute value) is reported; γ is the threshold, identification is considered as *weak* by the b test if the relative difference between C_t and C_b exceeds γ.

Appendix

Proof. (5)

By the joint normal distribution of quantile estimators, the difference of two quantile estimators is also normally distributed:

$$\sqrt{B}\left[\left(\frac{\hat{\theta}^*_{n,1-\alpha/2}-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}-\frac{\hat{\theta}^*_{n,\alpha/2}-\hat{\theta}_n}{\hat{\sigma}/\sqrt{n}}\right)-(q_{1-\alpha/2}-q_{\alpha/2})\right]\Rightarrow N(0,\Omega_{11}+\Omega_{22}-2\Omega_{12})$$

Rewrite the LHS:

$$\sqrt{B}\left[\left(\frac{\hat{\theta}^*_{n,1-\alpha/2}-\hat{\theta}^*_{n,\alpha/2}}{2z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}}-1\right)-\left(\frac{q_{1-\alpha/2}-q_{\alpha/2}}{2z_{1-\alpha/2}}-1\right)\right]\cdot 2z_{1-\alpha/2}\Rightarrow N(0,\Omega_{11}+\Omega_{22}-2\Omega_{12})$$

The result follows after the substitution of $\hat{D}=\frac{\hat{\theta}^*_{n,1-\alpha/2}-\hat{\theta}^*_{n,\alpha/2}}{2z_{1-\alpha/2}\hat{\sigma}/\sqrt{n}}-1,\ D=\frac{q_{1-\alpha/2}-q_{\alpha/2}}{2z_{1-\alpha/2}}-1.$ □

Proof. (6)

$$\begin{aligned}
\frac{X'P_zX}{n} &= \frac{X'Z(Z'Z)^{-1}Z'X}{n}\\
&= X'Z(Z'Z)^{-1}\frac{Z'Z}{n}(Z'Z)^{-1}Z'X\\
&\xrightarrow{p} \Pi_0'Q_{zz}\Pi_0\\
\frac{X'P_zU}{\sqrt{n}} &= X'Z(Z'Z)^{-1}\frac{Z'U}{\sqrt{n}}\\
&\Rightarrow \Pi_0'\Psi_{zu}\\
\sqrt{n}(\hat{\theta}_n-\theta) &= (\frac{X'P_zX}{n})^{-1}\frac{X'P_zU}{\sqrt{n}}\\
&\Rightarrow (\Pi_0'Q_{zz}\Pi_0)^{-1}\Pi_0'\Psi_{zu}
\end{aligned}$$

□

Proof. (7)

$$\begin{aligned}
X'P_zX &= \frac{X'Z}{\sqrt{n}}(\frac{Z'Z}{n})^{-1}\frac{Z'X}{\sqrt{n}}\\
&= (\frac{Z'Z}{n}C+\frac{Z'V}{\sqrt{n}})'(\frac{Z'Z}{n})^{-1}(\frac{Z'Z}{n}C+\frac{Z'V}{\sqrt{n}})\\
&\Rightarrow (Q_{zz}C+\Psi_{zv})'Q_{zz}^{-1}(Q_{zz}C+\Psi_{zv})
\end{aligned}$$

$$\begin{aligned} X'P_zU &= \frac{X'Z}{\sqrt{n}}(\frac{Z'Z}{n})^{-1}\frac{Z'U}{\sqrt{n}} \\ &\Rightarrow (Q_{zz}C+\Psi_{zv})'Q_{zz}^{-1}\Psi_{zu} \\ \hat{\theta}_n-\theta &= (X'P_zX)^{-1}X'P_zU \\ &\Rightarrow [(Q_{zz}C+\Psi_{zv})'Q_{zz}^{-1}(Q_{zz}C+\Psi_{zv})]^{-1}(Q_{zz}C+\Psi_{zv})'Q_{zz}^{-1}\Psi_{zu} \end{aligned}$$

If exactly identified:

$$\begin{aligned} \hat{\theta}_n-\theta &= (Z'X)^{-1}Z'U \\ &\Rightarrow (Q_{zz}C+\Psi_{zv})^{-1}\Psi_{zu} \end{aligned}$$

□

Proof. (8)

$$\begin{aligned} \hat{\Pi}_n^* &\equiv (Z^{*'}Z^*)^{-1}Z^{*'}X^* \\ &= \hat{\Pi}_n+(\frac{Z^{*'}Z^*}{n})^{-1}\frac{Z^{*'}V^*}{n} \\ &= \Pi_0+(\frac{Z'Z}{n})^{-1}\frac{Z'V}{n}+(\frac{Z^{*'}Z^*}{n})^{-1}\frac{Z^{*'}V^*}{n} \\ &\xrightarrow{p} \Pi_0 \\ \frac{Z^{*'}U^*}{\sqrt{n}} &\Rightarrow \Psi_{zu} \text{ by Lyapunov's Central Limit Theorem:} \end{aligned}$$

Let $m_i = Z_i^{*'}u_i^*$, rewrite $\frac{Z^{*'}U^*}{\sqrt{n}} = \frac{\sum_{i=1}^n Z_i^{*'}u_i^*}{\sqrt{n}} = \frac{\sum_{i=1}^n m_i}{\sqrt{n}}$. By construction, $m_i's$ are independent with mean $\mu_i = 0$, variance $\sigma_i^2 = \frac{Z'Z}{n}\frac{\tilde{U}'\tilde{U}}{n}$, and $\sigma_i^2 \xrightarrow{p} Q_{zz}\sigma_u^2$ under *Strong Instrument Asymptotics*. To verify the Lyapunov's condition: for $1 > \delta > 0$, the expected values $\mathbb{E}[|m_i|^{2+\delta}] < \infty$, and $\lim_{n\to\infty}\frac{1}{(\sum_{i=1}^n\sigma_i^2)^{\frac{2+\delta}{2}}}\sum_{i=1}^n\mathbb{E}[|m_i-\mu_i|^{2+\delta}] = \lim_{n\to\infty}\frac{O_p(n)}{O_p(n+\frac{n\delta}{2})} = 0$.

Combining the two results above:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^*-\hat{\theta}_n) &= \left[X^{*'}Z^*(Z^{*'}Z^*)^{-1}\frac{Z^{*'}Z^*}{n}(Z^{*'}Z^*)^{-1}Z^{*'}X^*\right]^{-1}X^{*'}Z^*(Z^{*'}Z^*)^{-1}\frac{Z^{*'}U^*}{\sqrt{n}} \\ &= \left[\hat{\Pi}_n^{*'}\left(\frac{Z^{*'}Z^*}{n}\right)\hat{\Pi}_n^*\right]^{-1}\hat{\Pi}_n^*\frac{Z^{*'}U^*}{\sqrt{n}} \\ &\Rightarrow (\Pi_0'Q_{zz}\Pi_0)^{-1}\Pi_0'\Psi_{zu} \end{aligned}$$

□

Proof. (9)

$$\begin{aligned} \mu^2 &= \frac{\Pi' Z' Z \Pi}{\sigma_v^2} \\ &= O_p(n) \\ &\to \infty \end{aligned}$$

$$\hat{\Pi}_n \xrightarrow{p} \Pi,\ \sigma_v^{2*} \xrightarrow{p} \sigma_v^2,\ \frac{Z^{*'} Z^*}{n} \xrightarrow{p} Q_{zz}$$

$$\begin{aligned} \mu^{2*} &= \frac{\hat{\Pi}_n' Z^{*'} Z^* \hat{\Pi}_n}{\sigma_v^{2*}} \\ &= O_p(n) \\ &\to \infty \end{aligned}$$

□

Proof. (10)

$$\begin{aligned} \mu^2 &= \frac{C' \frac{Z'Z}{n} C}{\sigma_v^2} \\ &\xrightarrow{p} \frac{C' Q_{zz} C}{\sigma_v^2} \\ \mu^{2*} &= \frac{[C + (\frac{Z'Z}{n})^{-1} \frac{Z'V}{\sqrt{n}}]' \frac{Z^{*'} Z^*}{n} [C + (\frac{Z'Z}{n})^{-1} \frac{Z'V}{\sqrt{n}}]}{\sigma_v^{2*}} \\ &\Rightarrow \frac{C' Q_{zz} C + 2C' \Psi_{zv} + \Psi_{zv}' Q_{zz}^{-1} \Psi_{zv}}{\sigma_v^2} \end{aligned}$$

□