Detecting Weak Identification by Bootstrap

Zhaoguo Zhan*

May 4, 2017

Abstract

This paper proposes bootstrap resampling as a diagnostic tool to detect weak instruments in the instrumental variable regression. When instruments are not weak, the bootstrap distribution of the standardized Two-Stage-Least-Squares estimator is close to the standard normal distribution. As a result, a substantial difference between these two distributions indicates the existence of weak instruments. A bootstrap-based test for evaluating the strength of instruments is thus developed. The size and power of the test are examined by simulation. For illustration, an empirical application is also discussed.

JEL Classification: C18; C26; C36

Keywords: Weak instruments; Weak identification; Bootstrap

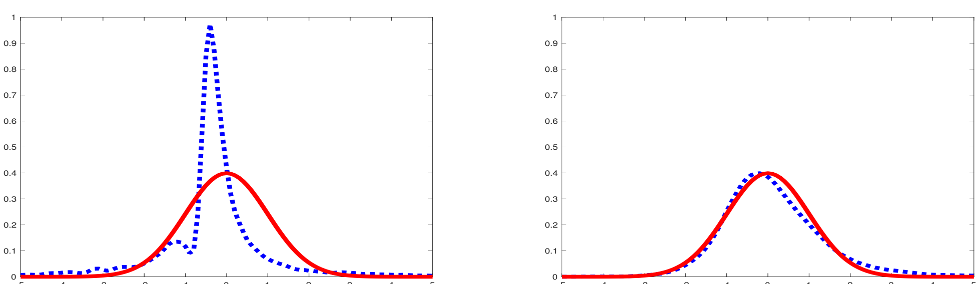
*Department of Economics, Finance and Quantitative Analysis, Coles College of Business, Kennesaw State University, GA 30144, USA. Email: zzhan@kennesaw.edu. Phone: (470)578-6537. Fax: (470)578-9022. Homepage: https://sites.google.com/site/zhaoguozhan.

1 Introduction

This paper proposes a simple method for detecting weak instruments in the linear Instrumental Variable (IV) regression model. The idea of the proposed method is conveyed by Figure 1. Panel (a) of Figure 1 plots the p.d.f. of a bootstrapped Two-Stage-Least-Squares (TSLS) estimator (after standardization) under a weak instrument, as well as the p.d.f. of the standard normal variate. The sharp difference of the two distributions in Panel (a) suggests that under the weak instrument, the bootstrap distribution of the TSLS estimator is substantially different from the normal distribution. By contrast, when a strong instrument is used to replace the weak instrument and draw Panel (b), the bootstrap distribution is close to the normal distribution. As indicated by Figure 1, the strength of instruments can be inferred by comparing the bootstrap distribution of the standardized TSLS estimator with the standard normal distribution.[1]

Figure 1: The comparison of the bootstrap distribution and the normal distribution

(· · · standardized bootstrap distribution — standard normal)



(a) Weak Instrument (b) Strong Instrument

Note: Panel (a) plots the bootstrap distribution (dotted) of a standardized TSLS estimator under a weak instrumental variable. This weak instrument is replaced by a strong one, in order to similarly draw the bootstrap distribution in Panel (b). In both Panel (a) and Panel (b), the p.d.f. of the standard normal variate (solid) is plotted for comparison. The data is from Card (1995).

There exists a sizable and growing literature on weak instruments and more generally, weak identification, which this paper builds upon. See, e.g., Stock, Wright and Yogo (2002)

[1]Figure 1 results from the empirical application in this paper, which is discussed in more detail later on.

for an early survey. It is now well known that when instruments are weak, the commonly used TSLS estimator is biased and its associated t/Wald-test suffers size distortion. To help evaluate the strength of instruments, Staiger and Stock (1997) and Stock and Yogo (2005) suggest the first stage F-test in the TSLS procedure. Since the TSLS is a standard toolkit for economists, the F-test, together with the F-statistic > 10 rule of thumb in Staiger and Stock (1997) for excluding weak instruments, is widely used in economic studies.[2] Motivated by the widely adopted F-test and its known limitations, this paper suggests an alternative bootstrap-based method for detecting weak instruments.[3]

Following the existing literature, the strength of instruments can be measured by the so-called concentration parameter. As a starting point, this paper shows that the concentration parameter appears in the leading term of the Edgeworth expansion of the TSLS estimator. This leading term reflects the difference between the distribution of the TSLS estimator and the normal distribution. To make this term small (e.g., less than 5%), the concentration parameter needs to exceed a certain threshold, which calls for the F-statistic much larger than 10 in the classical homoscedasticity and just-identification setup. The commonly used rule of thumb therefore does not imply that the distribution of the TSLS estimator is well approximated by the normal distribution.

This paper proceeds to show the bootstrap provides a simple realistic check for weak instruments. The underlying reason is the bootstrap resample preserves the pattern of identification, so that if weak instruments exist in the given sample, they are also likely to exist

[2]At the time of writing, Staiger and Stock (1997) and Stock and Yogo (2005) have more than 10,000 Google scholar citations in total.

[3]The F-test is known to have its limitations. For instance, although the F-test aims to control the TSLS bias and the size distortion of t/Wald-test, the F-statistic itself does not directly reflect how severe the bias or the size distortion is, so it lacks an intuitive interpretation. In addition, the commonly used cutoff 10 for the F-statistic is derived under the homoscedasticity assumption; to account for potential heteroscedasticity, autocorrelation and clustering, Olea and Pflueger (2013) develop a modified F-test that calls for critical values larger than 10. Furthermore, if a one-sided t-test is of interest, the commonly used cutoff of the F-statistic is not sufficient for controlling size distortion. See, e.g., the numerical evidence provided by Feir, Lemieux and Marmer (2016) in fuzzy regression discontinuity that is nested by TSLS. Finally, when the classical IV model considered in Staiger and Stock (1997) is generalized to Hansen (1982)'s Generalized Method of Moments (GMM) framework, extensions of the F-test to similarly detect weak identification in various IV/GMM settings are in demand but often challenging (see, e.g., Wright 2003, Sanderson and Windmeijer 2016).

in the bootstrap resample. Consequently, the difference between the bootstrap distribution of the standardized TSLS estimator and the standard normal distribution conveys the useful information of identification strength. This difference is measured throughout the paper by the Kolmogorov-Smirnov distance, since it reflects the worst-case size distortion of the one-sided t-test using critical values from the normal distribution. By examining this difference, researchers can easily evaluate how strong or weak the instruments are. Unlike many existing methods for detecting weak instruments or weak identification (see, e.g., Hahn and Hausman 2002, Stock and Yogo 2005, Inoue and Rossi 2011), the proposed bootstrap approach has the unique feature of providing an intuitive graphical view, as shown in Figure 1, and its reported Kolmogorov-Smirnov statistic can be straightforwardly interpreted.

Associated with the Kolmogorov-Smirnov statistic, a confidence interval for the difference in distribution between the standardized TSLS estimator and the standard normal variate is constructed by iterative bootstrap. Based on this confidence interval, a test is developed in this paper: The null hypothesis of strong instruments is rejected at the α-level, if the adopted threshold for weak instruments (e.g., 5%) is exceeded by the lower bound of the one-sided $100(1-\alpha)\%$ confidence interval. Consistent with expectations, both the size and the power of the proposed test are found to perform well in Monte Carlo simulations.

Finally, it is worth noting that other than the t/Wald-test associated with TSLS, there exist several robust tests (see, e.g., Anderson and Rubin 1949, Stock and Wright 2000, Kleibergen 2002, 2005, Moreira 2003), which provide reliable inference regardless of the strength of instruments. Although these robust tests are naturally appealing, evaluating the strength of instruments is still of theoretical and practical interest (see, e.g., Olea and Pflueger 2013, Sanderson and Windmeijer 2016). The objective of this paper is therefore to propose the bootstrap for evaluating instruments. If the research objective is to conduct inference on the structural parameters in the TSLS model, then the aforementioned robust tests can be directly employed.[4]

[4] It is not recommended to use the bootstrap in this paper (or other similar diagnosis tools) as a pre-test, and proceed to the t/Wald-test after instruments pass the pre-test. See, e.g., Guggenberger (2010).

The rest of the paper is organized as follows. Section 2 describes the classical linear IV regression model and uses the Edgeworth expansion to illustrate the deviation of the standardized TSLS estimator from the standard normal variate. In Section 3, the bootstrap-based approach for detecting weak instruments is proposed. Section 4 contains the Monte Carlo analysis and an empirical application adopted from Card (1995). Section 5 concludes. The necessary proofs and additional numerical results are attached in the Appendix.

2 Instrumental Variable, Edgeworth Expansion

2.1 Linear IV Model

Following Stock and Yogo (2005), this paper focuses on the linear IV regression model under homoscedasticity, with n i.i.d. observations:

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

where $U = (U_1, ..., U_n)'$, $V = (V_1, ..., V_n)'$ are $n \times 1$ vectors of structural and reduced form errors, respectively; $(U_i, V_i)', i = 1, ..., n$, is assumed to have mean zero and the covariance matrix $\Sigma = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$; $Z = (Z_1, ..., Z_n)'$ is the $n \times k$ matrix of instruments with $\mathbb{E}(Z_i U_i) = 0$ and $\mathbb{E}(Z_i V_i) = 0$; $Y = (Y_1, ..., Y_n)'$ and $X = (X_1, ..., X_n)'$ are $n \times 1$ vectors of endogenous observations. The structural parameter of interest is θ, while Π is the $k \times 1$ vector of nuisance parameters.

In addition, $(\frac{Z'U}{\sqrt{n}}, \frac{Z'V}{\sqrt{n}}) \xrightarrow{d} (\Psi_{zu}, \Psi_{zv})$ by the central limit theorem and $Z'Z/n \xrightarrow{p} Q_{zz}$ by the law of large numbers, provided the moments exist, where $(\Psi'_{zu}, \Psi'_{zv})' \sim N(\mathbf{0}, \Sigma \otimes Q_{zz})$, $Q_{zz} = \mathbb{E}(Z_i Z_i')$ is assumed to be nonsingular.

Given the sample (X, Y, Z), the commonly used TSLS estimator of θ, denoted by $\hat{\theta}_n$, is written as:

$$\hat{\theta}_n = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}Z'Y \tag{1}$$

Assumption 1 (Strong Instrument Asymptotics)**.** $\Pi = \Pi_0 \neq 0$*, and* Π_0 *is fixed.*

Under Assumption 1 and the setup of the model, the classical result is that $\hat{\theta}_n$ is asymptotically normally distributed, according to the conventional first-order asymptotics:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} (\Pi' Q_{zz} \Pi)^{-1} \Pi' \Psi_{zu} \tag{2}$$

The asymptotic result above can be rewritten to standardize $\hat{\theta}_n$:

$$\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} \xrightarrow{d} N(0,1) \tag{3}$$

where $\sigma = \sqrt{\sigma^2}$, $\sigma^2 = (\Pi' Q_{zz} \Pi)^{-1} \sigma_u^2$ and $N(0,1)$ stands for the standard normal variate, whose c.d.f. and p.d.f. are denoted by $\Phi(\cdot)$ and $\phi(\cdot)$, respectively.

2.2 Weak Instruments

The sizable literature on weak instruments highlights that in finite samples, the difference in distribution between the TSLS estimator and the normal variate could be substantial. See Nelson and Startz (1990), Bound, Jaeger and Baker (1995), etc. The substantial difference further induces the TSLS bias as well as the size distortion of the conventional t/Wald-test using critical values from the normal distribution. This issue is often referred to as the weak instrument problem, or more generally, the weak identification problem.

To my knowledge, Stock and Yogo (2005) are the first to suggest that weak instruments can be quantitatively defined based on either the TSLS bias or the size distortion of the t/Wald test. For instance, if the size distortion of the t/Wald-test associated with TSLS exceeds 5% (e.g., the rejection rate under the null exceeds 10% at the 5% level), then instruments are deemed weak in Stock and Yogo (2005).

This paper adopts the 5% threshold and considers the quantitative definition of weak instruments as follows.

Definition 1. *The identification strength of IV applications is deemed weak, if the maximum difference between the cumulative distribution function of the standardized TSLS estimator and that of the standard normal variate exceeds* 5%*; otherwise the identification is deemed strong. In other words, instruments are deemed weak iff the Kolmogorov-Smirnov distance below exceeds* 5%*:*

$$KS = \sup_{-\infty<c<\infty} \left| P(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} \leq c) - \Phi(c) \right| \tag{4}$$

Since $\sqrt{n}\frac{\hat{\theta}_n-\theta}{\sigma}$ is the non-studentized t-statistic, Definition 1 can be motivated as follows. Consider a one-sided non-studentized t-test using critical values from the standard normal distribution. If the size distortion of this test at any level exceeds 5%, then instruments are deemed weak. Consequently, Definition 1 can also be viewed as an extension of the definition in Stock and Yogo (2005): Definition 1 considers the size distortion at any nominal level, while the definition of weak instruments based on size distortion in Stock and Yogo (2005) relies on a specific preset significance level (e.g., the commonly used 5% or 10%).

2.3 Edgeworth Expansion

In order to describe the finite sample distribution of the TSLS estimator, Theorem 1 employs the Edgeworth expansion, which results from Hall (1992).

Theorem 1. *Under Assumption 1, if the following two conditions for the random vector* $R_i = (X_i Z_i', Y_i Z_i', vec(Z_i Z_i')')'$ *hold: i.* $\mathbb{E}(||R_i||^3) < \infty$*, ii.* $lim\ sup_{||t||\to\infty} |\mathbb{E}exp(it'R_i)| < 1$*, then* $\sqrt{n}\frac{\hat{\theta}_n-\theta}{\sigma}$ *admits the two-term Edgeworth expansion uniformly in c,* $-\infty < c < \infty$*:*

$$P(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} \leq c) = \Phi(c) + n^{-1/2}p(c)\phi(c) + O(n^{-1}) \tag{5}$$

where $p(c)$ *is a polynomial of degree* 2*, whose coefficients depend on the moments of* R_i *up* *to order* 3.[5]

Proof. A straightforward application of Theorem 2.2 in Hall (1992) to TSLS. □

Based on the Edgeworth expansion above, the leading term that reflects the deviation of $\sqrt{n}\frac{\hat{\theta}_n-\theta}{\sigma}$ from normality is $n^{-1/2}p(c)\phi(c)$, which shrinks to zero as n goes to infinity. The corollary below provides the expression of $p(c)$ in the empirically relevant just-identified case.

Corollary 1. *Under the conditions of Theorem 1, if Z_i is independent of (U_i, V_i) and the model is just-identified, i.e., $k = 1$, then $p(c) = \rho c^2/\sqrt{\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}}$ and the two-term Edgeworth expansion of $\sqrt{n}\frac{\hat{\theta}_n-\theta}{\sigma}$ reads:*

$$P(\sqrt{n}\frac{\hat{\theta}_n-\theta}{\sigma} \leq c) = \Phi(c) + \frac{\rho c^2 \phi(c)}{\sqrt{\mu^2}} + O(n^{-1}), \textit{with } \mu^2 \equiv n\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}. \tag{6}$$

Proof. See Appendix A. □

Corollary 1 shows that the departure of the TSLS estimator from normality in the just-identified model is crucially affected by two parameters: ρ, the correlation coefficient that measures the degree of endogeneity; μ^2, the concentration parameter that measures the strength of instruments.

2.4 Insufficient Cutoff of F-statistic

The result in Corollary 1 helps illustrate the insufficiency of the F-test, together with its commonly used cutoff, for controlling the deviation in Definition 1, as shown in this subsection.

For the linear IV model described above, the first stage F-statistic reads:

$$F = \frac{\widehat{\Pi}_n' Z' Z \widehat{\Pi}_n}{k\hat{\sigma}_v^2} \tag{7}$$

[5] A similar result is provided in Moreira, Porter and Suarez (2009), where the necessity of the two technical conditions i and ii is explained: the first condition is the imposed minimum moment assumption, while the second is Cramér's condition discussed in Bhattacharya and Ghosh (1978).

where $\widehat{\Pi}_n = (Z'Z)^{-1}Z'X$, $\hat{\sigma}_v^2$ is the sample variance of the residual $\widehat{V} = X - Z\widehat{\Pi}_n$. Thus this F-statistic is an estimator of the concentration parameter μ^2 divided by k. According to the rule of thumb in Staiger and Stock (1997), if the F-statistic is larger than 10, then instruments in Z are not deemed weak.[6]

To illustrate the insufficiency of the rule of thumb, let's omit the $O(n^{-1})$ term of the Edgeworth expansion as if instruments are not weak, so the leading term $\frac{\rho c^2 \phi(c)}{\sqrt{\mu^2}}$ mainly determines size distortion. By combining Corollary 1 and Definition 1, when weak instruments are excluded in the just-identified model, it approximately implies that:

$$\left| \frac{\rho c^2 \phi(c)}{\sqrt{\mu^2}} \right| \leq 5\% \quad \Longleftrightarrow \quad \mu^2 \geq 400 \rho^2 c^4 \phi^2(c) \tag{8}$$

At the point $c = \pm\sqrt{2}$, $400\rho^2 c^4 \phi^2(c)$ achieves its maximum. When the degree of endogeneity is severe, i.e., $|\rho| \approx 1$, the maximum is above 34. Consequently, the concentration parameter μ^2, loosely speaking, needs to exceed 34 to avoid the 5% distortion.

The argument above quickly suggests that the commonly used cutoff of the F-statistic is not quantitatively sufficient. In the empirically relevant just-identified case with $k = 1$, the F-statistic is just an estimator of the concentration parameter μ^2. In order to statistically establish that μ^2 is above 34, the F-statistic needs to be larger than it. To put it differently, if the F-statistic is only slightly above 10, then the chance that the distribution of the TSLS estimator differs substantially from the normal distribution is still large.

In view of the above, a question arises: Other than the commonly used F-statistic, can we develop another method for detecting weak instruments? Ideally, this method should provide the information on the deviation of the standardized TSLS estimator from the standard normal variate. The rest of the paper suggests that bootstrap resampling could be such a method.

[6] As explained in Stock and Yogo (2005), this rule of thumb is motivated by the definition of weak instruments based on the TSLS bias, rather than size distortion of the t/Wald-test. Consequently, F-statistic > 10 does not imply that the size distortion of the t/Wald-test is controlled. See their Table 5.1 and Table 5.2.

3 Bootstrap

Since its introduction by Efron (1979), the bootstrap has become a commonly used tool of economists. As an alternative to the limiting distribution by asymptotics, the bootstrap approximates the distribution of a targeted statistic by resampling the data. In addition, there is considerable evidence that the bootstrap performs better than the first-order limiting distribution in finite samples. See, e.g., Horowitz (2001).

However, the bootstrap does not always perform well. For instance, when instruments are weak, the bootstrap is known to provide a poor approximation to the exact finite sample distribution of the TSLS estimator. Nevertheless, the fact that the bootstrap fails still conveys useful information for the purpose of this paper. Specifically, when the bootstrap fails, the bootstrap distribution of the standardized TSLS estimator differs substantially from the standard normal distribution, which signals the existence of weak instruments, as detailed in this section.

3.1 Residual Bootstrap

To detect whether there exist weak instruments, the Kolmogorov-Smirnov distance between the exact distribution of the standardized TSLS estimator and the standard normal distribution is of interest. Since the exact distribution is unknown, this paper suggests replacing it by its bootstrap counterpart.

For this purpose, the residual bootstrap for the linear IV model is employed, which is described as follows.

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

$$\begin{cases} \widehat{U} = Y - X\hat{\theta}_n \\ \widehat{V} = X - Z\widehat{\Pi}_n \end{cases}$$

Step 2: Re-center $\widehat{U}, \widehat{V}$ to get $\tilde{U}, \tilde{V}$, so that they have mean zero and are orthogonal to

Z.

Step 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, following the naming convention of the bootstrap.

Step 4: The dependent variables (X^*, Y^*) are generated by:

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

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

$$\hat{\theta}_n^* = \left[X^{*'} Z^* (Z^{*'} Z^*)^{-1} Z^{*'} X^* \right]^{-1} X^{*'} Z^* (Z^{*'} Z^*)^{-1} Z^{*'} Y^* \tag{9}$$

Under Assumption 1, $\hat{\theta}_n$ is asymptotically normally distributed and admits an Edgeworth expansion in Theorem 1. For the same reason, the bootstrapped $\hat{\theta}_n^*$ also admits an Edgeworth expansion, as stated in Theorem 2.

Theorem 2. *Under the conditions in Theorem 1, the bootstrap version of the standardized TSLS estimator admits the two-term Edgeworth expansion uniformly in c,* $-\infty < c < \infty$*:*

$$P(\sqrt{n} \frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}} \leq c | \mathcal{X}_n) = \Phi(c) + n^{-1/2} p^*(c) \phi(c) + O(n^{-1}) \tag{10}$$

where $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$, $\hat{\sigma}^2 = (\widehat{\Pi}_n' \frac{Z'Z}{n} \widehat{\Pi}_n)^{-1} \frac{\tilde{U}'\tilde{U}}{n}$, $\mathcal{X}_n = (X, Y, Z)$ *denotes the sample of observations,* $p^*(c)$ *is the bootstrap counterpart of* $p(c)$.

Proof. Similar to Theorem 1, Theorem 2 results from Theorem 2.2 in Hall (1992). □

3.2 KS^* as the Proxy for KS

With the help of the bootstrap, the unknown Kolmogorov-Smirnov distance KS is proxied by its bootstrapped counterpart denoted by KS^*:

$$KS^* = \sup_{-\infty<c<\infty} \left| P(\sqrt{n}\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}} \le c|\mathcal{X}_n) - \Phi(c) \right| \tag{11}$$

where $P(\sqrt{n}\frac{\hat{\theta}_n^*-\hat{\theta}_n}{\hat{\sigma}} \le c|\mathcal{X}_n)$ can be derived by bootstrapping as follows. Re-do the residual bootstrap procedure B times to compute $\{\hat{\theta}_n^{*i}, i = 1, ..., B\}$, where $\hat{\theta}_n^{*i}$ denotes the i^{th} TSLS estimator by bootstrapping. By the strong law of large numbers, $\frac{1}{B}\sum_{i=1}^{B} \mathbf{1}(\sqrt{n}\frac{\hat{\theta}_n^{*i}-\hat{\theta}_n}{\hat{\sigma}} \le c|\mathcal{X}_n) \overset{a.s}{\to} P(\sqrt{n}\frac{\hat{\theta}_n^*-\hat{\theta}_n}{\hat{\sigma}} \le c|\mathcal{X}_n)$, i.e., $P(\sqrt{n}\frac{\hat{\theta}_n^*-\hat{\theta}_n}{\hat{\sigma}} \le c|\mathcal{X}_n)$ is effectively given, by making B sufficiently large. Consequently, KS^* is known, if the sample (X, Y, Z) denoted by $\mathcal{X}_n$ is provided.

For KS^* in (11) to be a good proxy for KS in (4), the bootstrap distribution needs to well approximate the exact distribution of the standardized TSLS estimator. In fact, under strong instruments, the bootstrap distribution does approximate the exact distribution well. Consequently, KS^* and KS lie close to each other, as stated in Theorem 3.

Theorem 3. *Under the conditions in Theorem 1,* $KS = O_p(n^{-1/2})$*,* $KS^* = O_p(n^{-1/2})$*,* $KS^* - KS = O_p(n^{-1})$*, where the* $O_p(\cdot)$ *orders are sharp rates.*

Proof. See Appendix B. □

Theorem 3 shows that under strong instruments, both KS and KS^* shrink to zero as the sample size n gets large. In addition, the bootstrap-based KS^* is the super consistent proxy for KS, since their difference is of the magnitude $O_p(n^{-1})$.

3.3 Iterative Bootstrap

To account for the sampling variation of KS^*, this paper takes one step further. Just as the bootstrap-based KS^* serves as the proxy for KS, by the same principle, KS^* can be approximated by its bootstrap counterpart denoted by KS^{**}. This amounts to the double-bootstrap procedure (see, e.g., Beran 1988, Horowitz 2001) for constructing a bootstrap confidence interval of KS.

1. First bootstrap: compute KS^* in (11).

2. Second bootstrap: construct the bootstrap confidence interval of KS, as follows.

   (a) Take a bootstrap resample from $\mathcal{X}_n$, call it $\mathcal{X}_n^1$.

   (b) Repeat first bootstrap by replacing $\mathcal{X}_n$ with $\mathcal{X}_n^1$ to get

$$KS^{**1} = \sup_{-\infty<c<\infty} \left| P(\sqrt{n}\frac{\hat{\theta}_n^{**} - \hat{\theta}_n^{*1}}{\hat{\sigma}^{*1}} \le c|\mathcal{X}_n^1) - \Phi(c) \right| \qquad (12)$$

   where $\hat{\theta}_n^{**}$ is the bootstrapped TSLS estimator computed by the resample taken from $\mathcal{X}_n^1$; $\hat{\theta}_n^{*1}, \hat{\sigma}^{*1}$ defined in $\mathcal{X}_n^1$ are the counterparts of $\hat{\theta}_n, \hat{\sigma}$ defined in $\mathcal{X}_n$.

   (c) Repeatedly take $\mathcal{X}_n^2, ..., \mathcal{X}_n^B$ from $\mathcal{X}_n$ and similarly compute $KS^{**2}, ..., KS^{**B}$ as for KS^{**1} in (12).

   (d) The *two-sided* $100(1-\alpha)\%$ confidence interval of KS results from taking lower and upper $\alpha/2$ quantiles of $KS^{**1}, ..., KS^{**B}$, where $0 < \alpha < 1$. Similarly, for the *one-sided* $100(1-\alpha)\%$ confidence interval of KS, its lower bound results from taking the lower α quantile of $KS^{**1}, ..., KS^{**B}$.

If KS^* approximates KS well, as suggested by Theorem 3, then by the same logic, KS^{**1}, ..., KS^{**B} also mimic KS^* well. Consequently, the resulting bootstrap confidence interval is expected to have good coverage rates, as will be shown by simulation later on.

3.4 Test by Bootstrap

With the pieces above, a bootstrap-based test is developed. Consider

$$H_0 : KS \le 5\% \quad against \quad H_1 : KS > 5\% \qquad (13)$$

where, according to Definition 1, the null hypothesis H_0 states that instruments are strong, while the alternative H_1 is that instruments are weak.

Let the level of the *one-sided* test be denoted by α. The decision rule based on the bootstrap is thus to reject H_0, if the lower α quantile of $KS^{**1}, ..., KS^{**B}$ exceeds 5%. In other words, the null hypothesis of strong instruments is rejected at the α-level, if the 5% threshold for weak instruments is below the left bound of the lower *one-sided* $100(1-\alpha)\%$ bootstrap confidence interval of KS.

The performance of the proposed test therefore relies on the coverage property of the bootstrap confidence interval. Since Theorem 3 establishes the feasibility of the bootstrap and thus the bootstrap confidence interval, the bootstrap-based test is expected to have the right size. To examine the performance of the test, as well as the coverage property of the bootstrap confidence interval, this paper presents the simulation evidence in the latter Section 4.

3.5 Weak Instrument Asymptotics

The bootstrap-based test described above is expected to have power, since the bootstrap distribution of the standardized TSLS estimator under weak instruments differs substantially from the standard normal distribution. To explain why this difference occurs, this paper adopts Assumption 2.

Assumption 2 (Weak Instrument Asymptotics). $\Pi = \frac{\Pi_0}{n^\delta}$*, where* Π_0 *is fixed,* $0 \leq \delta < \infty$*.*

When $\delta = 0$, Assumption 2 reduces to the conventional strong instrument asymptotics in Assumption 1. When $\delta = \frac{1}{2}$, Assumption 2 corresponds to the influential local-to-zero weak instrument asymptotics in Staiger and Stock (1997). When $\delta = \infty$, there is no identification. In addition, when $0 < \delta < \frac{1}{2}$, the identification strength is considered by Antoine and Renault (2009) as *nearly weak*; when $\frac{1}{2} < \delta < \infty$, the identification strength is treated as *near non-identified* in Hahn and Kuersteiner (2002). Overall, Assumption 2 corresponds to the asymptotically vanishing identification strength, which is weaker than in Assumption 1.

The discussion below focuses on the concentration parameter to show the bootstrap distribution under weak instruments differs from the normal distribution. This is because the weak instrument literature has established that the concentration parameter needs to accumulate to infinity, for the TSLS estimator to approach normality. See, e.g., Rothenberg (1984), Stock, Wright and Yogo (2002).

Under Assumption 2, the concentration parameter μ^2 now reads:

$$\mu^2 = n^{1-2\delta} \frac{\Pi_0' Q_{zz} \Pi_0}{\sigma_v^2} \tag{14}$$

which results from plugging $\Pi = \frac{\Pi_0}{n^\delta}$ to the definition of μ^2 in Corollary 1.

Similarly, the bootstrap counterpart of μ^2, denoted by μ^{*2}, is provided by:

$$\mu^{*2} = \frac{[\frac{\Pi_0}{n^\delta} + (Z'Z)^{-1}Z'V]' Z' Z [\frac{\Pi_0}{n^\delta} + (Z'Z)^{-1}Z'V]}{\tilde{V}'\tilde{V}/n} \tag{15}$$

which results from replacing Π, Q_{zz}, σ_v^2 with the bootstrap counterparts $\widehat{\Pi}_n$, $Z'Z/n$, $\tilde{V}'\tilde{V}/n$.

Theorem 4 explores the large sample difference between μ^2 and μ^{*2}.

Theorem 4. *Under Assumption 2 and the setup of the model, as $n \to \infty$:*

1. *If $0 \leq \delta < \frac{1}{2}$, then $\mu^2 = O_p(n^{1-2\delta}) \to \infty$; $\mu^{*2} = O_p(n^{1-2\delta}) \to \infty$.*
2. *If $\delta = \frac{1}{2}$, then $\mu^2 = \frac{\Pi_0' Q_{zz} \Pi_0}{\sigma_v^2}$; $\mu^{*2} \xrightarrow{d} \frac{\Pi_0' Q_{zz}\Pi_0 + 2\Pi_0'\Psi_{zv} + \Psi_{zv}' Q_{zz}^{-1}\Psi_{zv}}{\sigma_v^2}$; $\mu^{*2} - \mu^2 \xrightarrow{d} \frac{2\Pi_0'\Psi_{zv} + \Psi_{zv}' Q_{zz}^{-1}\Psi_{zv}}{\sigma_v^2}$.*
3. *If $\frac{1}{2} < \delta < \infty$, then $\mu^2 = o_p(1)$; $\mu^{*2} \xrightarrow{d} \chi_k^2$; $\mu^{*2} - \mu^2 \xrightarrow{d} \chi_k^2$.*

Proof. See Appendix C. □

When instruments are strong or nearly weak in Case 1 with $0 \leq \delta < \frac{1}{2}$, Theorem 4 shows that both μ^{*2} and μ^2 accumulate to infinity. By contrast, when instruments are weak or almost useless in Case 2 and 3 with $\frac{1}{2} \leq \delta < \infty$, neither μ^{*2} nor μ^2 goes to infinity; in addition, the bootstrap does not accurately preserve the identification strength in the sample, since μ^{*2} asymptotically differs from μ^2. The non-negligible difference between μ^{*2}

and μ^2 helps explain why the bootstrap fails to approximate the exact distribution of the TSLS estimator under weak instruments.

Although the bootstrap does not accurately preserve the identification strength, Theorem 4 shows that it does preserve the pattern of identification. That is, under strong instruments, both μ^{*2} and μ^2 accumulate to infinity, so the bootstrap distribution asymptotically coincides with the normal distribution; while under weak instruments, neither μ^{*2} nor μ^2 accumulates to infinity, so the bootstrap distribution differs from the normal distribution. This nice pattern therefore allows us to detect weak instruments by bootstrap.

The mean of the asymptotic difference between μ^{*2} and μ^2 is equal to k when $\frac{1}{2} \leq \delta < \infty$ in Theorem 4. Consequently, the identification strength of the bootstrap resample measured by μ^{*2} appears similar to but stronger than its counterpart μ^2, for small k's that are commonly encountered in practice. For the purpose of detecting weak instruments, this property is desirable. It implies that the use of the bootstrapped μ^{*2} is unlikely to exaggerate the weak instrument problem. The proposed bootstrap approach for detecting weak instruments is therefore conservative from this perspective.

3.6 Many Instruments

The discussion so far assumes that k is fixed, which corresponds to the situation where the number of instruments k is small relative to the sample size n. Another related strand of research concerns that the number of instruments is large. See, e.g., Bekker (1994), Chao and Swanson (2005), where k is allowed to grow as n increases. Although the focus of this paper is not on the scenario of many instruments, Assumption 3 is made in order to shed light on the performance of the bootstrap under large k's.

Assumption 3 (Many Instruments)**.** *As $n \to \infty$, $k/n \to \gamma$, and γ, $\Pi' Q_{zz} \Pi / \sigma_v^2$ are fixed.*

Assumption 3 is adopted from Bekker (1994), who shows that the TSLS estimator $\hat{\theta}_n$ under many instruments is inconsistent. Similarly, the bootstrapped TSLS estimator $\hat{\theta}_n^*$ is also inconsistent in Theorem 5.

Theorem 5. *Under Assumption 3 and the setup of the model, both* $\hat{\theta}_n$ *and* $\hat{\theta}_n^*$ *are inconsistent, i.e.,* $plim(\hat{\theta}_n) \neq \theta$, $plim(\hat{\theta}_n^*)|\mathcal{X}_n \neq \hat{\theta}_n$, *as* $n \to \infty$.

Proof. The condition for the TSLS consistency in Chao and Swanson (2005) is that the concentration parameter μ^2 (similarly, μ^{*2}) needs to grow at a faster rate than that of k. This rate condition is not satisfied for $\hat{\theta}_n$ (similarly, $\hat{\theta}_n^*$) under Assumption 3. □

Bekker (1994) also shows that the TSLS estimator under many instruments is asymptotically normal around its probability limit. However, the inconsistency result in Theorem 5 effectively explains why these instruments can be deemed weak in this paper and why the bootstrap has the power for detecting such instruments.

The subtle point is that the distribution of the *standardized* TSLS estimator, not the estimator itself, is concerned in this paper. More importantly, the TSLS estimator is considered to be standardized by using its estimand and standard error from conventional asymptotics (see, e.g., the expressions of KS and KS^* in (4) and (11)), rather than its probability limit and standard error from many instruments asymptotics. Consequently, even if the TSLS estimator itself is asymptotically normal under many instruments, once (improperly) standardized in the conventional manner, the resulting standardized version differs from the standard normal variate. The definition of weak instruments therefore also applies to the scenario of many instruments. In addition, whether the joint strength of many instruments is deemed weak in this paper relies on the validity of conventional asymptotics, including but not limited to the asymptotic normality of the TSLS estimator.

To put it differently, the joint strength of many instruments can be deemed weak in this paper, on the grounds that conventional asymptotics no longer function well when k is large. Theorem 5 shows the bootstrapped TSLS estimator under Assumption 3 is inconsistent, which further implies that after its conventional standardization, the resulting bootstrap distribution differs substantially from the standard normal distribution.[7] There-

[7] The TSLS standard error also matters for standardization, though it suffices here to focus on inconsistency for illustrating the point.

fore, the bootstrap remains to have power in the scenario of many instruments, for which the simulation evidence is presented in Appendix D.[8]

4 Monte Carlo Analysis and Application

This section presents the numerical evidence and an empirical example for the proposed bootstrap approach.

4.1 Monte Carlo Analysis

In the data generating process, a simple just-identified linear IV model is used:

$$\begin{cases} Y_i = X_i \cdot 0 + U_i \\ X_i = Z_i \cdot \Pi + V_i \end{cases} \quad with\ Z_i \sim NID(0,1), \begin{bmatrix} U_i \\ V_i \end{bmatrix} \sim NID\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

where $\rho = 0.99$ and $\rho = 0.50$ for severe and medium endogeneity, respectively; $i = 1, ..., n$, with $n = 1000$; $\Pi = \sqrt{\mu^2/n}$, in order to control the strength of instruments.

For each concentration parameter μ^2 listed in Table 1, the corresponding KS distance between the standardized TSLS estimator and the standard normal variate is reported.[9] As μ^2 increases, Table 1 shows that KS decreases. The comparison of Panel A with Panel B indicates that KS also decreases when endogeneity is less severe. In particular, when $\rho = 0.99$ and $\mu^2 = 35$, $KS \approx 5\%$. These findings are consistent with the Edgeworth expansion in Corollary 1.

The subsequent three columns of Table 1 report the sample median, mean and standard deviation of the bootstrap-based KS^*. These statistics result from 1000 Monte Carlo

[8]It might be interesting to further investigate the power of the bootstrap-based test under all the existing many instruments asymptotics (e.g., Bekker 1994, Chao and Swanson 2005, 2006). Depending on the interaction of μ^2, k and n, it is known that the TSLS estimator could be consistent or inconsistent, asymptotically normal or non-normal. I leave this investigation for future research.

[9]The reported KS results from simulation: after the standardized TSLS estimator is simulated 100,000 times, its empirical distribution is constructed and the resulting KS distance is reported.

Table 1: Monte Carlo outcome of the bootstrap

Panel A: $\rho = 0.99$								
		KS^*			Coverage of C.I.		Rejection Freq.	
μ^2	KS	median	mean	s.d.	90%	95%	5%	10%
5	0.111	0.109	0.123	0.069	0.938	0.968	0.962	0.987
10	0.088	0.091	0.093	0.029	0.899	0.960	0.850	0.926
20	0.065	0.067	0.069	0.016	0.898	0.951	0.458	0.596
30	0.053	0.056	0.057	0.011	0.907	0.963	0.140	0.233
35	0.050	0.052	0.053	0.010	0.905	0.956	0.058	0.112
40	0.047	0.049	0.050	0.009	0.902	0.940	0.039	0.066
50	0.042	0.044	0.045	0.007	0.909	0.957	0.008	0.019
Panel B: $\rho = 0.50$								
		KS^*			Coverage of C.I.		Rejection Freq.	
μ^2	KS	median	mean	s.d.	90%	95%	5%	10%
5	0.079	0.075	0.091	0.073	0.921	0.958	0.217	0.324
10	0.060	0.057	0.063	0.037	0.896	0.953	0.090	0.161
20	0.040	0.040	0.044	0.019	0.899	0.951	0.015	0.035
30	0.032	0.033	0.035	0.013	0.903	0.949	0.003	0.008
35	0.029	0.030	0.032	0.011	0.899	0.948	0.000	0.003
40	0.027	0.028	0.030	0.010	0.909	0.956	0.000	0.000
50	0.024	0.026	0.026	0.008	0.905	0.957	0.000	0.000

Note: μ^2 is the concentration parameter defined in Corollary 1; KS is the Kolmogorov-Smirnov distance between the distribution of the standardized TSLS estimator and the standard normal distribution defined in (4); KS^* is the bootstrap approximation to KS defined in (11); the C.I. is the two-sided bootstrap confidence interval of KS, with the nominal coverage 90% or 95%; the rejection frequency is from the one-sided test for $H_0 : KS \leq 5\%$ vs. $H_1 : KS > 5\%$, at the 5% or 10% level; ρ is the correlation coefficient of errors in the linear IV model. The detailed description of the d.g.p. for this table is provided in the main text. The number of Monte Carlo replications is 1000.

replications, while the number of bootstrap replications is $B = 10{,}000$ in each Monte Carlo replication. Table 1 shows KS^* approximates KS well, especially when μ^2 is large. This is consistent with Theorem 3, which establishes that under strong instruments, KS^* is the super-consistent proxy for KS. When μ^2 is small, similar to KS, the mean and median of KS^* appear large; in addition, the standard error of KS^* is also sizable. This is consistent with Theorem 4, which shows μ^{*2} does not accumulate to infinity under weak instruments, so the bootstrap distribution differs from the normal distribution. In addition, the limiting random behavior of μ^{*2} in Theorem 4 helps explain the large variation of KS^* under weak instruments.

Furthermore, Table 1 reports the actual coverage rates of the *two-sided* bootstrap confidence intervals of KS at the nominal 90%, 95%, respectively. These confidence intervals are constructed by the double-bootstrap procedure described in Section 3.3. None of the actual coverage rates reported in Table 1 are severely below the nominal ones, so the bootstrap confidence intervals do not appear to suffer under-coverage in the simulation. Table 1, however, suggests the existence of over-coverage, particularly when instruments are weak so the bootstrapped KS^* and KS^{**} have large variations. In order to establish the correct size for the bootstrap-based test, the coverage rates under strong instruments should be close to the nominal ones, which is supported by Table 1. On the other hand, for the power of the test under weak instruments, the coverage rates being close to the nominal ones are not required, i.e., as long as the bootstrap confidence intervals mostly contain values larger than 5%, the bootstrap-based test will have the power to reject the null.

Finally, the size and power of the bootstrap-based test for $H_0 : KS \leq 5\%$ against $H_1 : KS > 5\%$ are illustrated by the last two columns of Table 1. In particular, when $\mu^2 = 35$ in Panel A with $KS \approx 5\%$, the reported rejection frequencies are close to the nominal 5% and 10%, so the actual size of the test is as expected. When $KS > 5\%$ under small μ^2's, Table 1 reports large rejection frequencies: e.g., the rejection frequencies are close to 100% when $\mu^2 = 5$ and $\rho = 0.99$, which suggest the test has good power under large values of KS.

As shown in Table 1, the value of ρ affects KS and thus the performance of the bootstrap-based test; by contrast, the first stage F-statistic in (7) is independent of ρ. In other words, although both the F-test and the bootstrap-based test help evaluate the strength of instruments, the F-test is driven by μ^2, while the bootstrap-based test focuses on KS. Therefore, these two tests do not simply substitute for each other, nor are they directly comparable.

Overall, the simulation outcome in Table 1 shows that the proposed bootstrap approach works as expected.

4.2 Application

For further illustration, this subsection employs a classical example from Card (1995). The same data set as in Card (1995) is used.

By using the IV method, Card (1995) investigates the return to education, or specifically, how much more an individual can earn if he/she completes an extra year of schooling. The data set is taken from the National Longitudinal Survey of Young Men between 1966-1981 with 3010 observations. There are two variables in the data set that measure college proximity: $nearc2$ and $nearc4$, both of which are dummy variables, and are equal to 1 if there is a two-year or four-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 the following wage equation:

$$lwage = \beta_0 + \theta \cdot edu + W'\boldsymbol{\beta} + u \tag{16}$$

where $lwage$ is the log of wage; edu is the years of schooling; the covariate vector W contains the control variables; u is the error term.[10] The endogenous edu is to be instrumented by $nearc2$ or $nearc4$. Among the set of parameters $(\beta_0, \theta, \boldsymbol{\beta})$, the parameter θ measuring the percentage change in wage due to education is of interest.

The strength of instruments is examined by the first stage F-test. If $nearc2$ is used as the instrument for edu, F-statstic $= 0.54$; while if $nearc4$ is used as the instrument for edu, F-statistic $= 10.52$. These two F-statistics are reported in Table 2. According to the F-statistic > 10 rule of thumb, they suggest that $nearc2$ is a weak instrumental variable, while $nearc4$ is strong.

[10] The basic specification in Card (1995) uses five control variables: experience, the square of experience, a dummy for race, a dummy for living in the south and a dummy for living in the standard metropolitan statistical area. To bypass the issue that experience is also endogenous, Davidson and MacKinnon (2010) replace experience and the square of experience with age and the square of age. Following Davidson and MacKinnon (2010), this paper uses age, square of age and the three dummy variables as control variables. Therefore, edu is considered to be the only endogenous regressor. While Davidson and MacKinnon (2010) simultaneously use the available instruments, this paper uses $nearc2$ and $nearc4$ one by one as the sole instrument for edu, in order to illustrate the proposed bootstrap approach.

Table 2: Estimation outcome of return to education and diagnostic statistics

	IV: *nearc*2	**IV: *nearc*4**
Return to education		
TSLS estimate $\hat{\theta}_n$	0.508	0.094
95% C.I. by t/Wald	(-0.813, 1.829)	(-0.004, 0.191)
95% C.I. by $AR/CLR/K$	$(-\infty, -0.1750] \cup [0.0867, +\infty)$	[0.0009, 0.2550]
Diagnostic statistics		
F-statistic	0.54	10.52
KS^*	0.229	0.054
90% C.I. of KS	$(0.094, 0.471)$	$(0.028, 0.094)$
95% C.I. of KS	$(0.088, 0.486)$	$(0.026, 0.107)$

Note: This table presents the estimate $\hat{\theta}_n$ and confidence intervals of return to education using the data of Card (1995). The first stage F-statistic is reported for the two instrumental variables, $nearc2$ and $nearc4$, which are used one by one for the endogenous years of schooling. The included control variables are age, square of age, and dummy variables for race, living in the south and living in the standard metropolitan statistical area. KS^* is the Kolmogorov-Smirnov distance between the bootstrap distribution of the standardized TSLS estimator and the standard normal distribution. The 90% and 95% C.I.'s of KS are constructed by the double-bootstrap procedure. The number of bootstrap replications is $B = 10,000$.

Table 2 also reports the point estimates and 95% confidence intervals of θ (i.e., return to education) by TSLS and the associated t/Wald-test, using $nearc2$ or $nearc4$ as the instrument. In addition, the conditional likelihood ratio (CLR) test in Moreira (2003) is also employed to construct a robust confidence interval of θ for comparison. Since the model under consideration is just-identified, the CLR test is equivalent to the AR test in Anderson and Rubin (1949) and the K test in Kleibergen (2002). Under the weak $nearc2$, the robust $AR/CLR/K$ test produces a much wider confidence interval than the one by inverting the t/Wald-test. By contrast, under the strong $nearc4$, confidence intervals by the t/Wald-test and the robust $AR/CLR/K$ test are comparable; however, the difference of these two intervals is still substantial, although F-statistic > 10 occurs.[11]

As suggested by this paper, the bootstrap helps evaluate the strength of instruments, e.g., $nearc2$ and $nearc4$ in this application. Using each instrument, the TSLS estimator of return to education is thus computed by bootstrapping $B = 10,000$ times. After standardization,

[11] As reported in Table 2, the robust confidence interval by the $AR/CLR/K$ test under $nearc4$ rejects no return to education at the 5% level, while the t/Wald-test fails to reject it. In addition, the robust confidence interval by the $AR/CLR/K$ test is over 30% wider than the one by the t/Wald-test.

Figure 1 in the introduction section presents the p.d.f. of the resulting bootstrap distribution against the p.d.f. of the standard normal distribution.

nearc2 as IV, for Panel (a) of Figure 1: It is obvious in Panel (a) that the bootstrap distribution is far from the normal distribution, suggesting $nearc2$ is a weak instrument. The KS^* statistic is found to be 0.229, as reported in Table 2, which is well above the threshold 0.05 for weak instruments. In addition, the *two-sided* 90%, 95 % C.I.'s of KS constructed by double-bootstrap are found to be $(0.094, 0.471)$ and $(0.088, 0.486)$, which contain no points below 0.05. Furthermore, since $0.094 > 0.05$, the null hypothesis that $nearc2$ is strong is rejected at the 5% level by the *one-sided* test proposed in this paper. The KS^* statistic, the *two-sided* C.I.'s and the *one-sided* test, all suggest that $nearc2$ is a weak instrument.

nearc4 as IV, for Panel (b) of Figure 1: The bootstrap distribution does not appear too far from the normal distribution by eyeballing Panel (b), as reflected by 0.054 for KS^* in Table 2. The *two-sided* 90%, 95% bootstrap C.I.'s of KS are found to be $(0.028, 0.094)$ and $(0.026, 0.107)$. Given $0.028 < 0.05$, the null hypothesis that $nearc4$ is strong is not rejected at the 5% level by the proposed test. Although the null is not rejected, the 90% and 95% C.I.'s indicate the departure of the TSLS estimator from normality is still likely to be sizable.

To summarize, no matter whether $nearc2$ or $nearc4$ is used, diagnostic statistics by bootstrap suggest that KS is not minor. Consequently, the confidence intervals of return to education by the t/Wald-test, whose validity relies on normality of TSLS, differ substantially from those by the robust $AR/CLR/K$ test. As indicated by the studied example, the bootstrap provides helpful information on the deviation of the TSLS estimator from the normal variate, through the graph and statistics that are straightforward to interpret.

5 Conclusions

When instruments are not weak, the difference between the bootstrap distribution of the standardized TSLS estimator and the standard normal distribution is minor. For practical purposes, researchers can thus detect weak instruments by simply examining whether these two distributions are sufficiently close. A bootstrap-based test for testing the null of strong instruments is also developed in this paper. In addition, the empirical relevance of the proposed bootstrap approach is emphasized by an example adopted from Card (1995).

The bootstrap-based test is found to have power under weak instruments and many instruments in the simulation study, while the reason is two-fold. Under weak instruments, the bootstrap distribution of the TSLS estimator is not asymptotically normal. Under many instruments with the malfunction of conventional asymptotics, the bootstrap distribution of the conventionally standardized TSLS estimator also differs substantially from the standard normal distribution, although the TSLS estimator itself can be asymptotically normal.

Last but not least, an appealing feature of bootstrap resampling is that it is commonly applicable. Therefore, the bootstrap has the potential to serve as a simple and intuitive tool that detects whether weak identification exists in various settings, including but not limited to the one analyzed in this paper.

Appendix

Appendix A: Proof of Corollary 1

The proof uses Theorem 2.2 in Hall (1992).

When $k = 1$, $\hat{\theta}_n$ reduces to $(Z'X)^{-1}Z'Y$.[12] Re-define a random vector R_i, its population mean μ_R and its sample mean $\bar{R}$ for this proof:

$$\begin{aligned} R_i &= (X_i Z_i', Y_i Z_i')' \\ \mu_R = \mathbb{E}(R_i) &= (\Pi' Q_{zz}, \Pi' Q_{zz}\theta)' \\ \bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i &= \frac{1}{n}(X'Z, Y'Z)' \end{aligned}$$

Consider a function $A : \mathcal{R}^2 \to \mathcal{R}$ that operates on a vector of the dimension of R_i:

$$\begin{aligned} A(R_i) &= \frac{(Z_i X_i)^{-1} Z_i Y_i - \theta}{\sigma} \\ A(\bar{R}) &= \frac{(Z'X)^{-1} Z'Y - \theta}{\sigma} \end{aligned}$$

so $A(\mu_R) = 0$ and $\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} = \sqrt{n}A(\bar{R})$.

For the vector R_i, define:

$$\mu_{i_1,\ldots,i_j} = \mathbb{E}\left[(R_i - \mu_R)^{(i_1)} \ldots (R_i - \mu_R)^{(i_j)}\right]$$

where $(R_i - \mu_R)^{(i_j)}$ denotes the i_j-th element in $R_i - \mu_R$. Thus $\mu_{i_1,\ldots,i_j}$ is the centered moment of elements in R_i.

For the function A, define:

$$a_{i_1,\ldots,i_j} = \frac{\partial^j}{\partial x^{(i_1)} \ldots \partial x^{(i_j)}} A(x)|_{x=\mu_R}$$

[12] Π is now a scalar, though Π' is still occasionally used instead of Π, to be consistent with the notation used in the main text.

where $x^{(i_j)}$ is the i_j-th element in x. Thus $a_{i_1,\ldots,i_j}$ is the derivative of the function A taking values at μ_R.

With the notation above, by Theorem 2.2 in Hall (1992),

$$\begin{aligned} p(c) &= -A_1 - \frac{1}{6}A_2(c^2-1), \ \textit{where} \\ A_1 &= \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}\mu_{ij} \\ A_2 &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{2} a_i a_j a_l \mu_{ijl} + 3\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{2}\sum_{m=1}^{2} a_i a_j a_{lm}\mu_{il}\mu_{jm} \end{aligned}$$

For A_1: $A_1 = -\frac{\rho}{\sqrt{\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}}}$, which is derived as follows.

$$\begin{aligned} A_1 &= \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} a_{ij}\mu_{ij} \\ &= a_{12}\mu_{12} + \frac{1}{2}a_{11}\mu_{11} \\ &= a_{12}Q_{zz}\rho\sigma_u\sigma_v \\ &= -\frac{\rho}{\sqrt{\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}}} \end{aligned}$$

by using $a_{11} = -2\theta a_{12}$, $a_{12} = -\frac{1}{\Pi^2 Q_{zz}^2 \sigma}$, $a_{22} = 0$, $\mu_{12} = \theta\mu_{11} + Q_{zz}\rho\sigma_u\sigma_v$, which result from the definitions of $a_{i_1,\ldots,i_j}$ and $\mu_{i_1,\ldots,i_j}$.

For A_2: $A_2 = \frac{-6\rho}{\sqrt{\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}}}$, which is derived as follows.

Firstly, $\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{2} a_i a_j a_l \mu_{ijl} = 0$. This is because $a_1a_1a_1\mu_{111} + a_1a_1a_2\mu_{112} = 0$, using $\mu_{111} = \mathbb{E}(Z_iX_i - \Pi Q_{zz})^3$, $\mu_{112} = \mathbb{E}(Z_iX_i - \Pi Q_{zz})^2(Z_iY_i - \Pi Q_{zz}\theta) = \theta\mu_{111} + \mathbb{E}[(Z_iX_i - \Pi Q_{zz})^2 Z_iU_i] = \theta\mu_{111}$, $a_1 = -\theta a_2$. Similarly, $a_1a_2a_1\mu_{121} + a_1a_2a_2\mu_{122} = 0$, $a_2a_1a_1\mu_{211} + a_2a_1a_2\mu_{212} = 0$, $a_2a_2a_1\mu_{221} + a_2a_2a_2\mu_{222} = 0$.

Secondly, using $a_{22} = 0$, $a_1 = -\theta a_2$, $a_{11} = -2\theta a_{12}$, $\mu_{12} = \theta\mu_{11} + Q_{zz}\rho\sigma_u\sigma_v$, $\mu_{22} = \theta^2\mu_{11} + Q_{zz}\sigma_u^2 + 2\theta Q_{zz}\rho\sigma_u\sigma_v$, $a_{12} = -\frac{1}{\Pi^2 Q_{zz}^2\sigma}$, $a_2 = \frac{1}{\Pi Q_{zz}\sigma}$ to get:

$$
\begin{aligned}
& \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{2}\sum_{m=1}^{2} a_i a_j a_{lm}\mu_{il}\mu_{jm} \\
= \;& a_1a_1a_{11}\mu_{11}\mu_{11} + a_1a_1a_{12}\mu_{11}\mu_{12} + a_1a_1a_{21}\mu_{12}\mu_{11} + a_1a_2a_{11}\mu_{11}\mu_{21} + a_1a_2a_{12}\mu_{11}\mu_{22} \\
& + a_1a_2a_{21}\mu_{12}\mu_{21} + a_2a_1a_{11}\mu_{21}\mu_{11} + a_2a_1a_{12}\mu_{21}\mu_{12} + a_2a_1a_{21}\mu_{22}\mu_{11} + a_2a_2a_{11}\mu_{21}\mu_{21} \\
& + a_2a_2a_{12}\mu_{21}\mu_{22} + a_2a_2a_{21}\mu_{22}\mu_{21} \\
= \;& a_1^2a_{11}\mu_{11}^2 + a_2^2a_{11}\mu_{12}^2 + 2(a_1^2a_{12}\mu_{11}\mu_{12} + a_1a_2a_{11}\mu_{11}\mu_{12} + a_1a_2a_{12}\mu_{11}\mu_{22} + a_1a_2a_{12}\mu_{12}^2 \\
& + a_2^2a_{12}\mu_{12}\mu_{22}) \\
= \;& -2\theta^3a_2^2a_{12}\mu_{11}^2 - 2\theta a_2^2a_{12}\mu_{12}^2 + 2\theta^2a_2^2a_{12}\mu_{11}\mu_{12} + 4\theta^2a_2^2a_{12}\mu_{11}\mu_{12} - 2\theta a_2^2a_{12}\mu_{11}\mu_{22} \\
& - 2\theta a_2^2a_{12}\mu_{12}^2 + 2a_2^2a_{12}\mu_{12}\mu_{22} \\
= \;& -2\theta^3a_2^2a_{12}\mu_{11}^2 + 6\theta^2a_2^2a_{12}\mu_{11}\mu_{12} - 4\theta a_2^2a_{12}\mu_{12}^2 - 2\theta a_2^2a_{12}\mu_{11}\mu_{22} + 2a_2^2a_{12}\mu_{12}\mu_{22} \\
= \;& 2a_2^2a_{12}(-\theta^3\mu_{11}^2 + 3\theta^2\mu_{11}\mu_{12} - 2\theta\mu_{12}^2 - \theta\mu_{11}\mu_{22} + \mu_{12}\mu_{22}) \\
= \;& 2a_2^2a_{12}[-\theta^3\mu_{11}^2 + 3\theta^3\mu_{11}^2 + 3\theta^2\mu_{11}Q_{zz}\rho\sigma_u\sigma_v - 2\theta(\theta^2\mu_{11}^2 + Q_{zz}^2\rho^2\sigma_u^2\sigma_v^2 + 2\theta Q_{zz}\rho\sigma_u\sigma_v\mu_{11}) \\
& + \theta^2\mu_{11}Q_{zz}\rho\sigma_u\sigma_v + Q_{zz}^2\sigma_u^2\rho\sigma_u\sigma_v + 2\theta Q_{zz}^2\rho^2\sigma_u^2\sigma_v^2] \\
= \;& 2a_2^2a_{12}Q_{zz}^2\rho\sigma_u^3\sigma_v
\end{aligned}
$$

and

$$
\begin{aligned}
A_2 &= \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{2} a_i a_j a_l \mu_{ijl} + 3\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{l=1}^{2}\sum_{m=1}^{2} a_i a_j a_{lm}\mu_{il}\mu_{jm} \\
&= 6a_2^2a_{12}Q_{zz}^2\rho\sigma_u^3\sigma_v \\
&= \frac{-6\rho}{\sqrt{\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}}}
\end{aligned}
$$

With $A_1 = -\frac{\rho}{\sqrt{\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}}}$ and $A_2 = \frac{-6\rho}{\sqrt{\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}}}$ from above, $p(c) = -A_1 - \frac{1}{6}A_2(c^2-1) = \frac{\rho c^2}{\sqrt{\frac{\Pi' Q_{zz}\Pi}{\sigma_v^2}}}$. Plugging in $p(c)$ to the expansion in Theorem 1 and using the definition of μ^2 yield the expansion result in Corollary 1.

Appendix B: Proof of Theorem 3

$KS^* = \sup_{-\infty<c<\infty} \left| P(\sqrt{n}\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}} \leq c|\mathcal{X}_n) - \Phi(c) \right|$, $KS = \sup_{-\infty<c<\infty} \left| P(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} \leq c) - \Phi(c) \right|$, by the triangle property:

$$KS^* - KS \leq \sup_{-\infty<c<\infty} \left| P(\sqrt{n}\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}} \leq c|\mathcal{X}_n) - P(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} \leq c) \right|$$

To complete the proof, it suffices to show $p^*(c) = p(c) + O_p(n^{-1/2})$, since from Theorem 1 and Theorem 2:

$$\begin{aligned} P(\sqrt{n}\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}} \leq c|\mathcal{X}_n) &= \Phi(c) + n^{-1/2}p^*(c)\phi(c) + O(n^{-1}) \\ P(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} \leq c) &= \Phi(c) + n^{-1/2}p(c)\phi(c) + O(n^{-1}) \\ P(\sqrt{n}\frac{\hat{\theta}_n^* - \hat{\theta}_n}{\hat{\sigma}} \leq c|\mathcal{X}_n) - P(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} \leq c) &= n^{-1/2}\left[p^*(c) - p(c)\right]\phi(c) + O(n^{-1}) \end{aligned}$$

For the random vector R_i, define its population mean μ_R and its sample mean $\bar{R}$:

$$\begin{aligned} R_i &= (X_i Z_i', Y_i Z_i', vec(Z_i Z_i')')' \\ \mu_R = \mathbb{E}(R_i) &= (\Pi' Q_{zz}, \Pi' Q_{zz}\theta, vec(Q_{zz})')' \\ \bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i &= \frac{1}{n}(X'Z, Y'Z, vec(Z'Z)')' \end{aligned}$$

Consider a function $A : \mathcal{R}^{2k+k^2} \to \mathcal{R}$ that operates on a vector of the dimension of R_i:

$$\begin{aligned} A(R_i) &= \frac{\left[X_i Z_i'(Z_i Z_i')^{-1} Z_i X_i\right]^{-1} X_i Z_i'(Z_i Z_i')^{-1} Z_i Y_i - \theta}{\sigma} \\ A(\bar{R}) &= \frac{\left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}Z'Y - \theta}{\sigma} \end{aligned}$$

so $A(\mu_R) = 0$, and $\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma} = \sqrt{n}A(\bar{R})$.

For the vector R_i, define $\mu_{i_1,\ldots,i_j} = \mathbb{E}\left[(R_i - \mu_R)^{(i_1)} \ldots (R_i - \mu_R)^{(i_j)}\right]$, where $(R_i - \mu_R)^{(i_j)}$ denotes the i_j-th element in $R_i - \mu_R$.

For the function A, define $a_{i_1,\ldots,i_j} = \frac{\partial^j}{\partial x^{(i_1)}\ldots\partial x^{(i_j)}} A(x)|_{x=\mu_R}$, where $x^{(i_j)}$ is the i_j-th element in x.

With the notation above, by Theorem 2.2 in Hall (1992),

$$\begin{aligned} p(c) &= -A_1 - \frac{1}{6}A_2(c^2-1), \ \textit{where} \\ A_1 &= \frac{1}{2}\sum_{i=1}^{2k+k^2}\sum_{j=1}^{2k+k^2} a_{ij}\mu_{ij} \\ A_2 &= \sum_{i=1}^{2k+k^2}\sum_{j=1}^{2k+k^2}\sum_{l=1}^{2k+k^2} a_i a_j a_l \mu_{ijl} + 3\sum_{i=1}^{2k+k^2}\sum_{j=1}^{2k+k^2}\sum_{l=1}^{2k+k^2}\sum_{m=1}^{2k+k^2} a_i a_j a_{lm}\mu_{il}\mu_{jm} \end{aligned}$$

and similarly for the bootstrap counterparts,

$$\begin{aligned} p^*(c) &= -A_1^* - \frac{1}{6}A_2^*(c^2-1), \ \textit{where} \\ A_1^* &= \frac{1}{2}\sum_{i=1}^{2k+k^2}\sum_{j=1}^{2k+k^2} a_{ij}^*\mu_{ij}^* \\ A_2^* &= \sum_{i=1}^{2k+k^2}\sum_{j=1}^{2k+k^2}\sum_{l=1}^{2k+k^2} a_i^* a_j^* a_l^* \mu_{ijl}^* + 3\sum_{i=1}^{2k+k^2}\sum_{j=1}^{2k+k^2}\sum_{l=1}^{2k+k^2}\sum_{m=1}^{2k+k^2} a_i^* a_j^* a_{lm}^*\mu_{il}^*\mu_{jm}^* \end{aligned}$$

with

$$\begin{aligned} R_i^* &= (X_i^* Z_i^{*\prime}, Y_i^* Z_i^{*\prime}, vec(Z_i^* Z_i^{*\prime})')' \\ \mu_R^* = \mathbb{E}^*(R_i^*) &= (\hat{\Pi}_n' \frac{Z'Z}{n}, \hat{\Pi}_n' \frac{Z'Z}{n}\hat{\theta}_n, vec(\frac{Z'Z}{n})')' \\ \mu_{i_1,\ldots,i_j}^* &= \mathbb{E}^*\left[(R_i^* - \mu_R^*)^{(i_1)}\ldots(R_i^* - \mu_R^*)^{(i_j)}\right] \\ a_{i_1,\ldots,i_j}^* &= \frac{\partial^j}{\partial x^{(i_1)}\ldots\partial x^{(i_j)}} A(x)|_{x=\mu_R^*} \end{aligned}$$

If $a_{i_1,\ldots,i_j}^* \xrightarrow{p} a_{i_1,\ldots,i_j}$ and $\mu_{i_1,\ldots,i_j}^* = \mu_{i_1,\ldots,i_j} + O_p(n^{-1/2})$, then $A_i^* = A_i + O_p(n^{-1/2}), i = 1, 2$, and consequently, $p^*(c) = p(c) + O_p(n^{-1/2})$.

For $a_{i_1,\ldots,i_j}^*$: $a_{i_1,\ldots,i_j}^* = \frac{\partial^j}{\partial x^{(i_1)}\ldots\partial x^{(i_j)}} A(x)|_{x=\mu_R^*}$. Under Assumption 1,

$$\mu_R^* = (\hat{\Pi}_n' \frac{Z'Z}{n}, \hat{\Pi}_n' \frac{Z'Z}{n} \hat{\theta}_n, vec(\frac{Z'Z}{n})')' \xrightarrow{p} \mu_R = (\Pi' Q_{zz}, \Pi' Q_{zz}\theta, vec(Q_{zz})')'$$

hence by continuous mapping, $a^*_{i_1,\ldots,i_j} \xrightarrow{p} a_{i_1,\ldots,i_j}$.

For $\mu^*_{i_1,\ldots,i_j}$: apply the central limit theorem for the sample mean, provided the moments exist:

$$\begin{aligned}
\mu^*_{i_1,\ldots,i_j} &= \mathbb{E}^* \left[(R_i^* - \mu_R^*)^{(i_1)} \ldots (R_i^* - \mu_R^*)^{(i_j)} \right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \left[(R_i^* - \mu_R^*)^{(i_1)} \ldots (R_i^* - \mu_R^*)^{(i_j)} \right] \\
&= \mu_{i_1,\ldots,i_j} + O_p(n^{-1/2})
\end{aligned}$$

Combining these pieces above, $p^*(c) = p(c) + O_p(n^{-1/2})$ holds to complete the proof.

Appendix C: Proof of Theorem 4

Under Assumption 2:

$$\begin{aligned}
\mu^2 &= n^{1-2\delta} \frac{\Pi_0' Q_{zz} \Pi_0}{\sigma_v^2} \\
&= O_p(n^{1-2\delta}) \\
\mu^{*2} &= \frac{[\frac{\Pi_0}{n^\delta} + (Z'Z)^{-1} Z'V]' Z' Z [\frac{\Pi_0}{n^\delta} + (Z'Z)^{-1} Z'V]}{\tilde{V}'\tilde{V}/n}
\end{aligned}$$

where $\sqrt{n}(Z'Z)^{-1}Z'V \xrightarrow{d} Q_{zz}^{-1}\Psi_{zv}$, $Z'Z/n \xrightarrow{p} Q_{zz}$ and $\tilde{V}'\tilde{V}/n \xrightarrow{p} \sigma_v^2$, so $(Z'Z)^{-1}Z'V = O_p(n^{-1/2})$.

1. If $0 \le \delta < \frac{1}{2}$, then $\mu^2 = O_p(n^{1-2\delta}) \to \infty$. $\frac{\Pi_0}{n^\delta}$ has a larger magnitude than $(Z'Z)^{-1}Z'V$, so $\mu^{*2} = O_p(n^{1-2\delta}) \to \infty$.

2. If $\delta = \frac{1}{2}$, then μ^2 reduces to $\frac{\Pi_0' Q_{zz} \Pi_0}{\sigma_v^2}$. $\frac{\Pi_0}{n^\delta}$ has the same magnitude as $(Z'Z)^{-1}Z'V$, so $\mu^{*2} \xrightarrow{d} \frac{\Pi_0' Q_{zz}\Pi_0 + 2\Pi_0'\Psi_{zv} + \Psi_{zv}' Q_{zz}^{-1}\Psi_{zv}}{\sigma_v^2}$. The asymptotic difference of μ^{*2} and μ^2 is thus $\frac{2\Pi_0'\Psi_{zv} + \Psi_{zv}' Q_{zz}^{-1}\Psi_{zv}}{\sigma_v^2}$.

3. If $\frac{1}{2} < \delta < \infty$, $1 - 2\delta < 0$, then μ^2 converges to zero, so $\mu^2 = o_p(1)$. $\frac{\Pi_0}{n^\delta}$ has a smaller magnitude than $(Z'Z)^{-1}Z'V$, so μ^{*2} asymptotically reduces to

$$\frac{[(Z'Z)^{-1}Z'V]'Z'Z[(Z'Z)^{-1}Z'V]}{\tilde{V}'\tilde{V}/n} \xrightarrow{d} \frac{\Psi_{zv}'Q_{zz}^{-1}\Psi_{zv}}{\sigma_v^2} \sim \chi_k^2$$

The asymptotic difference of μ^{*2} and μ^2 is thus χ_k^2.

Appendix D: Power under Many Instruments

The purpose of this appendix is to illustrate the power of the bootstrap-based test, when k is large. For this purpose, consider $k = 50$. The data generating process is described in the main text, with the following modifications:

$$Z_i \sim NID(\mathbf{0}, I_k) \quad and \quad \Pi = \iota_k \cdot \sqrt{\frac{\mu^2}{n \cdot k}}$$

where I_k is the $k \times k$ identity matrix, ι_k is the $k \times 1$ vector of ones.

The simulated rejection frequencies of the bootstrap-based test for testing (13) are presented in Table 3. To facilitate comparison, the values of μ^2/k listed in Table 3 correspond to those of μ^2 in Table 1. Therefore, for each row from Table 1 to Table 3, the average strength of instruments is fixed, while the difference lies in that Table 3 considers $k = 50$ instruments instead of $k = 1$ instrument in Table 1.

Compared with its counterpart in Table 1, KS in Table 3 is found to be much larger. The increase in KS from Table 1 to Table 3 indicates the deviation of the standardized TSLS estimator from the standard normal variate is more sizable, when k jumps from 1 to 50 but μ^2/k remains unchanged. As explained in the main text, large values of KS in Table 3 are due to the malfunction of conventional standardization under many instruments.

The rejection frequencies of the bootstrap-based test are found to be 100% in Table 3. Alongside with the large values of KS, these large rejection frequencies are also driven by the fact that conventional asymptotics no longer function well under many instruments, i.e., after

Table 3: Rejection frequencies of the bootstrap-based test under $k = 50$

	Panel A: $\rho = 0.99$			Panel B: $\rho = 0.50$		
		Rejection Freq.			Rejection Freq.	
μ^2/k	KS	5%	10%	KS	5%	10%
5	0.853	100%	100%	0.511	100%	100%
10	0.709	100%	100%	0.396	100%	100%
20	0.550	100%	100%	0.294	100%	100%
30	0.465	100%	100%	0.245	100%	100%
35	0.435	100%	100%	0.228	100%	100%
40	0.410	100%	100%	0.214	100%	100%
50	0.371	100%	100%	0.193	100%	100%

Note: μ^2/k is the concentration parameter in Corollary 1 divided by k, the number of instruments; KS is the Kolmogorov-Smirnov distance between the distribution of the standardized TSLS estimator and the standard normal distribution defined in (4); ρ is the correlation coefficient of errors in the linear IV model. The detailed description of the d.g.p. for this table is provided in Appendix D and the main text. The rejection frequencies result from 200 Monte Carlo replications, with $B = 10{,}000$ in each Monte Carlo replication.

conventional standardization, the bootstrap distribution of the standardized TSLS estimator substantially differs from the standard normal distribution.

Overall, the findings in Table 3 suggest that the bootstrap-based test remains to have good power under $k = 50$ instruments, which is consistent with the discussion in Section 3.6.

References

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

Antoine, B. and E. Renault (2009). Efficient GMM with nearly-weak instruments. *Econometrics Journal 12*, S135–S171.

Bekker, P. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 657–681.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association 83*(403), 687–697.

Bhattacharya, R. and J. Ghosh (1978). On the validity of the formal Edgeworth expansion. *Annals of Statistics 6*(2), 434–451.

Bound, J., D. Jaeger, and R. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, 443–450.

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.

Chao, J. and N. Swanson (2005). Consistent estimation with a large number of weak instruments. *Econometrica 73*(5), 1673–1692.

Chao, J. and N. Swanson (2006). Asymptotic normality of single-equation estimators for the case with a large number of weak instruments. *Frontiers of Analysis and Applied Research: Essays in Honor of Peter CB Phillips, eds. D. Corbae, SN Durlauf, and BE Hansen*, 82–124.

Davidson, R. and J. MacKinnon (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. *Annals of Statistics*, 1–26.

Feir, D., T. Lemieux, and V. Marmer (2016). Weak identification in fuzzy regression discontinuity designs. *Journal of Business and Economic Statistics 34*(2), 185–196.

Guggenberger, P. (2010). The impact of a hausman pretest on the asymptotic size of a hypothesis test. *Econometric Theory 26*(2), 369.

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

Hahn, J. and G. Kuersteiner (2002). Discontinuities of weak instrument limiting distributions. *Economics Letters 75*(3), 325–331.

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

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 1029–1054.

Horowitz, J. L. (2001). The bootstrap. *Handbook of Econometrics 5*, 3159–3228.

Inoue, A. and B. Rossi (2011). Testing for weak identification in possibly nonlinear models. *Journal of Econometrics 161*(2), 246–261.

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

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

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

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

Nelson, C. and R. Startz (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica*, 967–976.

Olea, J. L. M. and C. Pflueger (2013). A robust test for weak instruments. *Journal of Business and Economic Statistics 31*(3), 358–369.

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

Sanderson, E. and F. Windmeijer (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics 190*(2), 212–221.

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

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

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

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

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