Bootstrap and Higher-Order Expansion Validity When Instruments May Be Weak [1]

Marcelo J. Moreira
Harvard University and NBER

Jack R. Porter
University of Wisconsin

Gustavo A. Suarez[2]
Harvard University

First Version: April 2003
Current Version: July 2005

Comments are welcome.

[2]Discussions with Tom Rothenberg were very important for this paper, and we are indebted for his insights and generosity. We also thank Silvia Goncalves, the participants at the Boston College, Boston University, Cornell, Harvard-MIT, Maryland, Montreal, Rice, Texas A&M, Texas Austin, and UCLA seminars, and at *Conference of Econometrics and Mathematical Economics* (CEME) and *Semiparametrics in Rio* conferences. Moreira and Porter gratefully acknowledge the research support of the National Science Foundation via grant numbers SES-0418268 and SES-0112095, respectively. Suarez acknowledges support from the Los Andes-Harvard Fund and Banco de la Republica.

Abstract

In this paper, we provide a framework for the bootstrap in the weak-instrument case. This framework yields a formal proof for the bootstrap validity of the Anderson-Rubin statistic, and bootstrap *invalidity* of the likelihood ratio (and Wald) statistic. Additionally, we propose two conditional bootstrap methods for the conditional likelihood ratio (CLR) test. Monte Carlo simulations show that the (conditional) bootstrap yields higher-order improvements with good instruments and is first-order valid with weak instruments. Finally, we provide a framework to derive asymptotic expansions in the weak-instrument setting. In particular, we show that an empirical Edgeworth expansion for the Anderson-Rubin test provides an $o(n^{-1})$ improvement to the chi-square-k distribution.

Keywords: Instrumental variables regression, weak instruments, t-statistic, Edgeworth expansion, bootstrap, conditional likelihood ratio test.

JEL Classification: C12, C31.

1 Introduction

The challenge posed to first-order asymptotics by weak instruments in the traditional simultaneous equations model has been explored extensively in the econometrics literature; cf. Bound, Jaeger, and Baker (1995), Dufour (1997), Nelson and Startz (1990), Staiger and Stock (1997), and Wang and Zivot (1998). In particular, the 2SLS estimator is biased, and the size of the Wald test is larger than the nominal significance level. Under standard asymptotics (with good instruments), empirical Edgeworth expansions show that the bootstrap actually provides asymptotic refinements. It is then natural to apply either higher-order asymptotics or the bootstrap to decrease the bias of the 2SLS estimator and the size distortions of the Wald test. However, in weak-instrument cases, these procedures appear to be unreliable; cf., Hahn, Hausman, and Kuersteiner (2002), Horowitz (2001), and Rothenberg (1984).

In this paper, we provide a theoretical proof for the validity of the bootstrap for the Anderson-Rubin statistic, and invalidity for the likelihood ratio (and Wald) statistic. In addition, we propose two conditional bootstrap methods for the conditional likelihood ratio (CLR) test. These simulation methods do not generally provide higher-order improvements in the unidentified case, since the bootstrap replaces parameters with inconsistent estimators. Although other methods, such as the m-out-of-n bootstrap (or, of course, using the chi-square critical value), also provide first-order asymptotic approximations, the usual bootstrap method has the advantage of providing a higher-order approximation in the good-instrument case. Thus, bootstrapping such statistics should outperform competing methods overall. A Monte Carlo study for the (conditional) bootstrap supports our theoretical results.

We also propose a theoretical two-step framework to study asymptotic expansions in the weak-instrument setting that augments the standard Bhattacharya and Ghosh approach. We first provide an Edgeworth expansion for certain sufficient statistics, and then we find an approximation of the distribution of the test statistic as a function of the sufficient statistics. As a result, we show that an empirical Edgeworth expansion for the Anderson-

Rubin test provides an $o(n^{-1})$ improvement to the chi-square-k distribution. This theory may also be helpful to study asymptotic expansions for non-differentiable statistics, since a general theory of higher-order expansions for non-smooth statistics is unavailable; see Bhattacharya and Ghosh (1978), Chambers (1967), Phillips (1977), Sargan (1976), and Wallace (1958).

Other papers consider bootstrapping in weak-instrument settings. In related work, Moreira, Porter, and Suarez (2005) focused exclusively on the score statistic and point out that it is non-differentiable in the unidentified case. That work emphasizes the importance of recentering residuals for bootstrapping the score statistic, and shows that the higher-order terms are in general not continuous functions of the nuisance parameters in the (locally) unidentified case. The current paper focuses on a different, yet complementary, set of issues: the bootstrap for a smooth statistic (e.g, Anderson-Rubin), conditional bootstrap methods, and developing an approach to Edgeworth expansions for test statistics under weak instruments. Work by Inoue (2002) and Kleibergen (2003) presents Monte Carlo results suggesting that the usual bootstrap may work when applied to the Anderson-Rubin statistic and score statistics. However, they do not provide formal proofs for the validity of Edgeworth and the bootstrap that work in the unidentified case. Our theoretical results can in principle be extended to the GMM and GEL contexts and provide a formal validation of their simulation results. This can be done by replicating our results on the higher-order expansion and bootstrap behavior of the GMM and GEL versions of the statistics considered in the simple simultaneous equations model analyzed here.

The remainder of this paper is organized as follows. In Section 2, the traditional instrumental variables model is presented and corresponding notation established. In Section 3, folk theorems are summarized which give the size improvements for the likelihood ratio and Anderson-Rubin tests based on Edgeworth expansion or the bootstrap under the standard asymptotics. Section 4 presents the main results. We provide a formal result on the validity of the bootstrap for the Anderson-Rubin test, and of two conditional bootstrap methods for the CLR test. Additionally, we provide a framework for Edgeworth expansions when instruments are unrelated to the endogenous explanatory variable. In Section 5, we present Monte Carlo simulations that suggest that the bootstrap methods may lead to improvements, although in general they do not lead to higher-order adjustments in the weak-instrument case. Section 6 concludes.

2 The Model

In this section, we introduce notation for the standard instrumental variable specification. The following structural equation is assumed to be part of a larger linear simultaneous equations model

$$y_1 = y_2\beta + u. \tag{1}$$

This notation represents a vector of n observations, so that the two endogenous variables, y_1 and y_2, and the unobserved disturbance, u, are each $n \times 1$ vectors. The key object of interest is the unknown scalar parameter, β. It is assumed that y_2 and u are correlated, and the reduced form for $Y = [y_1, y_2]$ is given as follows

$$\begin{aligned} y_1 &= Z\pi\beta + v_1 \\ y_2 &= Z\pi + v_2, \end{aligned} \tag{2}$$

where the $n \times k$ matrix of exogenous variables Z has full column rank k and π is a $k \times 1$ parameter vector. We assume the rows of Z and $V = [v_1, v_2]$ are i.i.d. The rows of the $n \times 2$ matrix of the reduced-form errors V are mean zero with 2×2 nonsingular covariance matrix $\Omega = [\omega_{i,j}]$. For ease of exposition, we consider statistics tailored to the known Ω case. In the proofs, we indicate how to handle the unknown Ω case. Let F denote the joint distribution of an observation from Z and V. We will focus on the case where Z is independent of V. In what follows, X_n is the n-th observation of some random matrix X, and $\overline{X}_n$ is the sample mean of the first n observations of X. The subscript n is typically omitted unless necessary. Finally, for any conformable matrix A, let $N_A = A\,(A'A)^{-1}\,A'$ and $M_A = I - N_A$. Also, set $b_0 = (1, -\beta_0)'$ and $a_0 = (\beta_0, 1)'$.

Tests for the null hypothesis $H_0 : \beta = \beta_0$ play an important role in our results. Two common statistics designed for H_0 are the score (LM) and Wald (W) statistics:

$$\begin{aligned} LM &= S'T/\sqrt{T'T} \text{ and} \\ W &= \left(\widehat{\beta}_{2SLS} - \beta_0\right) \sqrt{y_2' N_Z y_2}/\hat{\sigma}_u, \end{aligned}$$

where $S = (Z'Z)^{-1/2} Z'Y b_0 \cdot (b_0' \Omega b_0)^{-1/2}$, $T = (Z'Z)^{-1/2} Z'Y \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2}$, $\widehat{\beta}_{2SLS} = (y_2' N_Z y_2)^{-1} y_2' N_Z y_1$, and $\hat{\sigma}_u^2 = [1, -\widehat{\beta}_{2SLS}] \Omega [1, -\widehat{\beta}_{2SLS}]'$,

Staiger and Stock (1997) show that under weak-instrument asymptotics the limiting distribution of the Wald statistic is not standard normal and hence may have large size distortions in the weak-instrument case. The score test used by Kleibergen (2002) and Moreira (2001) is a discontinuous function of the observations at $\pi = 0$. The bootstrap behavior in this special circumstance is examined in separate work, by Moreira, Porter, and Suarez (2005).

Other frequently-used test statistics include the Anderson-Rubin (AR) and likelihood ratio (LR) statistics:

$$\begin{aligned} AR &= S'S \text{ and} \\ LR &= \frac{1}{2} \left(S'S - T'T + \sqrt{(S'S + T'T)^2 - 4[S'T \cdot T'T - (T'T)^2]} \right). \end{aligned}$$

The test of Anderson and Rubin (1949) rejects the null hypothesis if the AR statistic is larger than the $1 - \alpha$ quantile of the chi-square-k distribution. The CLR test of Moreira (2003) rejects the null if the LR statistic is larger than the $1 - \alpha$ conditional quantile of its null distribution conditional on T. Both of these tests are similar if the errors are normal with known variance Ω, since the AR statistic is pivotal and the LR statistic is pivotal conditional on T.

In this paper, we focus primarily on the test based on the AR statistic and conditional tests based on the LR statistic. These two statistics seem to be the most relevant for applied work. The AR statistic is robust to misspecifications of the structural equation for y_2,[3] and the CLR test has optimality properties.

[3] This robustness property of the AR statistic was pointed out by Jean-Marie Dufour.

3 Preliminary Results

In this section, we review Edgeworth expansion and bootstrap results for the good-instrument case. These results are already known or follow from standard results. The results in this section parallel results for the Wald and score tests given in Moreira, Porter, and Suarez (2005) and provide a basis of comparison for the weak-instrument results to come in the next section.

For any symmetric $\ell \times \ell$ matrix A, $vech(A)$ will denote the $\ell(\ell+1)/2$-column vector containing the column by column vectorization of the non-redundant elements of A. In this section, we focus on two-sided tests based on AR and LR statistics, which can be written in the form

$$2n\left(H\left(\overline{R}_n\right) - H\left(\mu\right)\right), \tag{3}$$

where $R_n = vech\left((Y_n', Z_n')'(Y_n', Z_n')\right)$. Also, H has gradient evaluated at $\mu = E\left(R_n\right)$ equal to zero, and the Hessian matrix L and variance Ξ of R_n satisfy $L\Xi L = L$.

We use the following high-level assumptions:

Assumption 1. π is fixed and different from zero.

Assumption 2. $E\left\|R_n\right\|^s < \infty$ for some $s \geq 3$.

Assumption 3. $\limsup_{\|t\| \to \infty} \left|E\exp\left(it'R_n\right)\right| < 1$.

Assumption 1 corresponds to standard good-instrument asymptotics. Since the test statistics involve quadratic functions of (Y_n', Z_n'), Assumption 2 holds if $E\left\|(Y_n', Z_n')\right\|^{2s} < \infty$. Assumption 3 is the commonly used Cramér's condition. A primitive sufficient condition for this assumption is given in Bhattacharya (1977).

In the identified case in which π is fixed and different from zero, not only is the 2SLS estimator consistent for β, but both Wald and score statistics also admit second-order Edgeworth expansions under mild conditions. Application of Bhattacharya and Ghosh (1978) and Chandra and Ghosh (1979) yields the following good-instrument Edgeworth expansion result for the AR and LR statistics:

Theorem 1 *Under Assumptions 1-3, the null distributions of AR_n and LR_n statistics can be uniformly approximated (in x) by the following expansions:*
(a) $P(AR_n \leq x) = G_k(x) + \sum_{i=1}^{r} n^{-i} p_{AR}^{i}(x; F, \beta_0, \pi)\, g_k(x) + o(n^{-r})$,
(b) $P(LR_n \leq x) = G_1(x) + \sum_{i=1}^{r} n^{-i} p_{LR}^{i}(x; F, \beta_0, \pi)\, g_1(x) + o(n^{-r})$,
where $\Phi(x)$ and $\phi(x)$ are the cdf and pdf of a chi-square-one variable; $G_v(x)$ and $g_v(x)$ are the cdf and pdf of a chi-square-v variable; p_{AR}^{i} and p_{LR}^{i}, $i = 1, 2$, are polynomials in x with coefficients depending on moments of R_n, β_0, and π; and r is the greatest integer less than or equal to $(s-2)/2$.

Proof. See Appendix A.

A corresponding bootstrap result is also available. First, we describe the natural method for obtaining bootstrap samples in this model. Given a bootstrap sample, each bootstrapped test statistic value can be computed. These bootstrapped test statistics have an empirical distribution which can be approximated through simulation and used to give critical values for the tests.

We start with a sample from which consistent estimates $\widehat{\beta}$ and $\widehat{\pi}$ for β and π are obtained. The reduced-form residuals are then given by $\widehat{v}_1 = y_1 - Z\widehat{\pi}\widehat{\beta}$ and $\widehat{v}_2 = y_2 - Z\widehat{\pi}$. Next, re-center these residuals to obtain $\widetilde{v}_j = \widehat{v}_j - \bar{\widehat{v}}_j$ for $j = 1, 2$. Now, construct the bootstrap sample for Z^* and (v_1^*, v_2^*) from independent draws of the empirical distribution functions of Z and $(\widetilde{v}_1, \widetilde{v}_2)$ (with distribution denoted by F_n). Finally, set

$$\begin{aligned} y_1^* &= Z^*\widehat{\pi}\widehat{\beta} + v_1^* \\ y_2^* &= Z^*\widehat{\pi} + v_2^*. \end{aligned}$$

Note that an alternative *exactly equivalent* method for generating bootstrap samples is to simulate directly from the structural model. The re-centered structural residual is $\widetilde{u} = \widetilde{v}_1 - \widetilde{v}_2\hat{\beta}$. From the empirical distribution functions of Z and $(\widetilde{u}, \widetilde{v}_2)$ take independent draws to form Z^* and (u^*, v_2^*).

Then obtain $Y^* = [y_1^*, y_2^*]$ from

$$\begin{aligned} y_1^* &= y_2^* \widehat{\beta} + u^* \\ y_2^* &= Z^* \widehat{\pi} + v_2^*. \end{aligned}$$

Let P^* denote the probability under the empirical distribution function (conditional on the sample). Note also that our results are for the "correlated case" corresponding to random draws on Z^*. Different Edgeworth expansions and bootstrap approximations would result from the fixed Z case. We will not consider this case here, although one could handle it by establishing conditions similar to those by Navidi (1989) and Qumsiyeh (1990, 1994) in the simple regression model.

The following result shows that the bootstrap approximates the empirical Edgeworth expansion up to some order of the sample size n.

Theorem 2 *Under Assumptions 1-3,*
(a) $P^*(AR_n^* \leq x) = G_k(x) + \sum_{i=1}^{r} n^{-i} p_{AR}^i(x; F_n, \widehat{\beta}, \widehat{\pi}) g_k(x) + o(n^{-r}),$
(b) $P^*(LR_n^* \leq x) = G_1(x) + \sum_{i=1}^{r} n^{-i} p_{LR}^i(x; F_n, \widehat{\beta}, \widehat{\pi}) g_1(x) + o(n^{-r}),$
a.s. as $n \to \infty$.

Proof. See Appendix A.

Comments: 1. The error based on the bootstrap simulation is of order n^{-1} for two-sided tests due to the fact that the conditional moments of R_n^* converge almost surely to those of R_n, and that $\widehat{\beta}$ and $\widehat{\pi}$ converge almost surely to β and π. Consequently, Theorem 2 shows that the bootstrap offers a better approximation than the standard normal and chi-square approximations.

2. Similar results holds for the one-sided W and LM statistics; see Moreira, Porter, and Suarez (2005).

4 Main Results

The previous section yielded three findings for the good-instrument case. First, higher-order expansion approximations are available for the null distribution of the likelihood ratio and Anderson-Rubin statistics. Second, the bootstrap distribution of these tests approximates their empirical Edgeworth expansions up to the n^{-r} order. Third, for the Anderson-Rubin and likelihood ratio tests, the error of the bootstrap is $o(n^{-1})$.

The good-instrument condition, given in Assumption 1, is critical to these results. When instruments are weak, versions of Theorems 1 and 2 have been unsuccessful in fixing size distortions of the Wald test. Three issues that arise in the weak-instrument setting help explain the poor performance of the bootstrap and Edgeworth expansions.

First, to our knowledge, the only approach to proving existence of Edgeworth expansions (Bhattacharya and Ghosh (1978), Chandra and Ghosh (1979)) depends upon the existence of derivatives of the function, defining the given statistic, evaluated at $\mu = E(R_n)$. However many statistics (likelihood ratio, Wald, score) are not smooth when instruments are uncorrelated with the endogenous variables.

Second, many statistics (such as likelihood ratio and Wald) have a discontinuous limiting distribution. They are pivotal under standard asymptotics, but have non-standard distributions that depend on parameters that are not consistently estimable under weak-instrument asymptotics.

Third, the null hypothesized value β_0 is replaced by an estimator $\widehat{\beta}$ that is inconsistent with weak instruments.

In this section, we study asymptotic expansions and the (conditional) bootstrap for the likelihood ratio and Anderson-Rubin statistics in the weak-instrument setting.[4] Formally, Assumption 1 will be replaced by one of the following poor-instrument assumptions:

Assumption 1A (unidentified case). $\pi = 0$.

[4] All stated results are for tests designed for the known covariance matrix case. Analogous results hold when we replace Ω with the consistent estimator $\widetilde{\Omega}$.

Assumption 1B (locally unidentified case). $\pi = c/\sqrt{n}$ for some non-stochastic k-vector c.

In Section 4.1, we present the main results of this paper. Using Theorem 3 and Corollary 4, we prove that: the bootstrap provides a valid first-order approximation to the null distribution of the Anderson-Rubin (and score) test; we show that the bootstrap does not provide a valid first-order approximation to the null distribution of the likelihood ratio (and Wald) test; and we propose two conditional bootstrap methods to approximate the null distribution of the CLR test. Our bootstrap results for the weak-instrument case are of first-order nature given that the bootstrap replaces β_0 by the (inconsistent) estimator $\widehat{\beta}$.

In Section 4.2, we derive the first results on higher-order expansions in the weak-instrument setting. These results allow us to show formally that the Anderson-Rubin statistic admits a standard asymptotic expansion with continuous higher-order terms even at $\pi = 0$. In particular, an empirical Edgeworth approach would be able to provide an $o(n^{-1})$ improvement to the chi-square-k approximation. We also note how the provided Edgeworth expansions for S and T can be used to find asymptotic expansions for other statistics, like the CLR (and the score).

4.1 Bootstrap

A priori, it is not clear whether the empirical distribution generating the bootstrap sample is close to the distribution of the data under the null when instruments are weak. In particular, the residuals are based on inconsistent parameter estimates and so may not provide a good approximation to the corresponding distribution of errors. This line of intuition underlies the doubt concerning bootstrap usefulness when instruments are weak.

However, we note that for any fixed value of π (including the unidentified case $\pi = 0$), we typically have $\widehat{\pi} \overset{a.s.}{\to} \pi$ and $\widehat{\pi}\widehat{\beta} \overset{a.s.}{\to} \pi\beta$. For $\pi = 0$, this convergence occurs despite an inconsistent estimator, $\widehat{\beta}$. Since the residuals depend on $\widehat{\pi}$ and $\widehat{\pi}\widehat{\beta}$ (and not $\widehat{\beta}$ alone), this convergence suggests that the bootstrap might be useful even in the unidentified case. This basic insight will underlie our first-order bootstrap results below.

We note that the test statistics of interest here are functions of the statistics S and T. Hence, we first derive bootstrap asymptotics for S and T and use this result to obtain the bootstrap asymptotic results for each test. As noted above, we discuss the known Ω case for expositional ease. In this case, Ω is replaced by the estimator $\widetilde{\Omega}$ (based on $(\widetilde{v}_1, \widetilde{v}_2)$) in the bootstrapped test statistics. The bootstrapped versions of S and T are:

$$\begin{aligned} S^* &= (Z^{*\prime}Z^*)^{-1/2}Z^{*\prime}Y^*\hat{b}\cdot(\hat{b}'\tilde{\Omega}\hat{b})^{-1/2}, \\ T^* &= (Z^{*\prime}Z^*)^{-1/2}Z^{*\prime}Y^*\tilde{\Omega}^{-1}\hat{a}\cdot(\hat{a}'\tilde{\Omega}^{-1}\hat{a})^{-1/2}, \end{aligned}$$

where $\hat{a} = (\widehat{\beta}, 1)'$ and $\hat{b} = (1, -\widehat{\beta})'$.

To derive the asymptotic distributions of S^* and T^*, we must re-center T^* by subtracting the term

$$t_n^* = \sqrt{n}\left(\frac{Z'Z}{n}\right)^{1/2}\widehat{\pi}\sqrt{\hat{a}'\tilde{\Omega}^{-1}\hat{a}}.$$

We can then consider the joint limiting distribution of $(S^*, T^* - t_n^*)$, where

$$T^* - t_n^* = \sqrt{n}\left[\left(\frac{Z^{*\prime}Z^*}{n}\right)^{1/2} - \left(\frac{Z'Z}{n}\right)^{1/2}\right]\widehat{\pi}\sqrt{\hat{a}'\tilde{\Omega}^{-1}\hat{a}} + \sqrt{n}\frac{\left(\frac{Z^{*\prime}Z^*}{n}\right)^{-1/2}\frac{Z^{*\prime}V^*}{n}\tilde{\Omega}^{-1}\hat{a}}{\sqrt{\hat{a}'\tilde{\Omega}\hat{a}}}.$$

To describe this limiting distribution, we require some additional notation, Liapunov's Central Limit Theorem and the Delta method,

$$\sqrt{n}[(Z'Z/n)^{1/2} - E(Z'Z/n)^{1/2}]\pi \xrightarrow{d} N(0, \Sigma),$$

where Σ depends directly on π. In particular, define $\Sigma = 0$ when $\pi = 0$. For $\pi = 0$, $\sqrt{n}\widehat{\pi}$ is bounded in probability and $(Z^{*\prime}Z^*/n)^{1/2} - (Z'Z/n)^{1/2}$ has zero conditional probability limit almost surely. Hence, the first term of $T^* - t_n^*$ is asymptotically negligible, and the second term has a joint normal limit distribution with S^*.

Lemmas in Appendix A show asymptotic normality of the following expression under various assumptions including good- and weak-instrument cases:

$$\begin{pmatrix} \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right)\frac{\widehat{b}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}} \\ \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right)\frac{\widehat{\Omega}^{-1}\widehat{a}}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}} \\ \sqrt{n}\left(\frac{W^{*\prime}\imath}{n}-\frac{W'\imath}{n}\right) \end{pmatrix}$$

where $w_i = vech\,(Z_i Z_i') \in R^{k(k+1)/2}$, $W = (w_1, \ldots, w_n)$, $w_i^* = vech\,(Z_i^* Z_i^{*\prime})$, $W^* = (w_1^*, \ldots, w_n^*)$, and $\imath$ denotes an $n \times 1$ vector of ones. Since (S^*, T^*) is a function of the above expression, the next result on the asymptotic distribution of these bootstrapped statistics follows.

Theorem 3 *Suppose that, for some $\delta > 0$, $E\|Z_i\|^{4+\delta}, E\|v_{\cdot i}\|^{4+\delta} < \infty$, where $v_{\cdot i} = [v_{1,i}, v_{2,i}]'$, is the i-th observation of the reduced-form residuals. Let $\widehat{\pi}$ and $\widehat{\beta}$ be estimators satisfying either*
(i) Assumption 1, $\widehat{\beta} \stackrel{a.s.}{\to} \beta$, $\widehat{\pi} - \pi \stackrel{a.s.}{\to} 0$; or
(ii) Assumption 1B (or 1A), $\widehat{\pi}\widehat{\beta} - \pi_n\beta \stackrel{a.s.}{\to} 0$, $\widehat{\pi} - \pi_n = O_p(n^{-1/2})$, $\widehat{\beta} = O_p(1)$.
Then, the following result holds:

$$\begin{pmatrix} S^* \\ T^* - t_n^* \end{pmatrix} |\mathcal{X}_n \stackrel{d}{\to} \begin{pmatrix} \mathcal{S} \\ \mathcal{T} \end{pmatrix} \quad a.s.\ ,$$

where $\mathcal{X}_n = \{(Y_1', Z_1'), \ldots, (Y_n', Z_n')\}$ and $(\mathcal{S}', \mathcal{T}')'$ is $N(0, A)$ with

$$A = \begin{pmatrix} I_k & 0 \\ 0 & I_k + \Sigma a_0{}'\Omega^{-1}a_0 \end{pmatrix}.$$

Proof. See Appendix A.

Comments: 1. Lemma A in Moreira, Porter, and Suarez (2005) shows that the assumption of almost sure convergence of $\hat{\pi}$ and $\hat{\pi}\hat{\beta}$ is the norm even in the unidentified case. However, if instead $\hat{\pi}$ and $\hat{\pi}\hat{\beta}$ only converge in probability, then the proof of Theorem 3 still works but the weak convergence in the conclusion of the theorem occurs with probability approaching one rather than almost surely. Both almost-sure and in-probability conclusions correspond to modes of convergence that have been proposed for the bootstrap; cf. Efron (1979) and Bickel and Freedman (1981).

2. This result provides a first-order approximation. In principle this result could be extended to include higher-order terms. However, the bootstrap in the weak-instrument case does not provide a second-order approximation because the higher-order terms depend on $\widehat{\beta}$ separately from the term $\hat{\pi}\hat{\beta}$. In other words, second-order improvements based on the bootstrap may worsen as π approaches zero.

3. Of course, other bootstrap methods are also possible. One alternative method would replace generate bootstrap samples using β_0 rather than replacing it with $\widehat{\beta}$. While this method avoids the use of an inconsistent estimator (when instruments are weak), it also possibly leads to power losses (under the alternative hypothesis, the true β will differ from β_0 and so the e.d.f. of the residuals will not be close to their c.d.f.). Another method would be to ignore the "structural restrictions" on the parameters and just perform OLS on the reduced-form model. However, then we no longer have the interpretation of bootstrapping from the structural form residuals (in the over-identified model).

The joint distribution of S^* and T^* can be used to provide first-order validity of the bootstrapped Anderson-Rubin even with weak instruments. It also yields a proof for the invalidity of the bootstrap in approximating the asymptotic distribution of the likelihood ratio (and Wald) statistic with weak instruments. This is rather expected, given that the likelihood ratio (and Wald) statistic is not well-behaved with weak instruments. However, we are unaware of any *formal* proof in the literature of bootstrap failure with weak instruments.[5] Finally, following the discussion of Section 2, conditioning can be used to provide asymptotically similar tests, as is the case with the likelihood ratio statistic. These tests rely on a theoretically constructed

[5] The only negative result is by Dufour (1997) on Wald-type confidence regions having zero confidence level even using bootstrapped critical values. Dufour (1997)'s impossibility result does not hold for the likelihood ratio statistic.

(and typically Monte Carlo-simulated) critical value function. Theorem 3 provides another way of obtaining a critical value for conditional tests. We could exploit the (first-order) independence of S^* and $T^* - t_n^*$ from Theorem 3 by fixing T at its observed value and obtaining S^* from bootstrap samples. These results are presented in the following corollary:

Corollary 4 *Suppose that, for some $\delta > 0$, $E\|Z_i\|^{4+\delta}, E\|v_{\cdot i}\|^{4+\delta} < \infty$, where $v_{\cdot i} = [v_{1,i}, v_{2,i}]'$ is the i-th observation of the reduced-form residuals. Then, the following results hold:*
(a) Under Assumption 1A, $\widehat{\pi\beta} - \pi_n\beta \xrightarrow{a.s.} 0$, $\widehat{\pi} - \pi_n = O_p(n^{-1/2})$, $\widehat{\beta} = O_p(1)$,

$$LR^*|\mathcal{X}_n \xrightarrow{d} \frac{1}{2}\left(\mathcal{S}'\mathcal{S} - \mathcal{T}_n'\mathcal{T}_n + \sqrt{(\mathcal{S}'\mathcal{S} + \mathcal{T}_n'\mathcal{T}_n)^2 - 4[\mathcal{S}'\mathcal{T}_n \cdot \mathcal{T}_n'\mathcal{T}_n - (\mathcal{T}_n'\mathcal{T}_n)^2]}\right),$$

almost surely, where $\mathcal{T}_n = \mathcal{T} + t_n^*$.
(b) Either under Assumption 1, $\widehat{\beta} \xrightarrow{a.s.} \beta$, $\widehat{\pi} - \pi \xrightarrow{a.s.} 0$, *or under Assumption 1B,* $\widehat{\pi\beta} - \pi_n\beta \xrightarrow{a.s.} 0$, $\widehat{\pi} - \pi_n = O_p(n^{-1/2})$, $\widehat{\beta} = O_p(1)$,

i) $AR^*|\mathcal{X}_n \xrightarrow{d} \chi^2(q)$

ii) $LR(S^*, t)|\mathcal{X}_n \xrightarrow{d} \frac{1}{2}\left(\mathcal{S}'\mathcal{S} - t't + \sqrt{(\mathcal{S}'\mathcal{S} + t't)^2 - 4[\mathcal{S}'t \cdot t't - (t't)^2]}\right)$

almost surely.

Comments: 1. The term $t_n^* = O_p(1)$ present in the asymptotic bootstrap distribution of LR^* affects the ability of the bootstrap in approximating the asymptotic distribution of LR when $\pi = 0$:

$$LR \xrightarrow{d} \frac{1}{2}\left(\mathcal{S}'\mathcal{S} - \mathcal{T}'\mathcal{T} + \sqrt{(\mathcal{S}'\mathcal{S} + \mathcal{T}'\mathcal{T})^2 - 4[\mathcal{S}'\mathcal{T} \cdot \mathcal{T}'\mathcal{T} - (\mathcal{T}'\mathcal{T})^2]}\right).$$

A similar argument can be used to show that the bootstrap does not work with the Wald statistic.

2. Under Assumption 1, the term t_n^* drifts off to infinity. Following the proof of Theorem 6(c), of Andrews, Moreira, and Stock (2006), we can show that the asymptotic bootstrap distribution of LR^* is chi-square-one. This result is just the first-order version of Theorem 2(c).

3. The fixed-T bootstrap method of item (b), part (ii), can be extended to other conditional tests. However, this conditional bootstrap method does not work for the conditional Wald testing procedure, since the Wald statistic depends on β_0 not only through S and T.

Theorem 3 is also connected to the conditional bootstrap method as proposed by Booth, Hall, and Wood (1992). When conditioning on the observed value of T, we make use of the bootstrap samples for which T^* is close to T. The fixed-T bootstrap has a significant, computational efficiency advantage over the non-parametric method proposed by Booth, Hall, and Wood (1992). This non-parametric method also depends crucially on bandwidth choice, which may prove problematic in practice. In addition, T^* is a random vector with dimension k, and non-parametric methods are known to perform poorly for high dimensions. The high-dimensionality problem can be avoided for the class of invariant similar tests analyzed by Andrews, Moreira, and Stock (2003). These tests depend exclusively on $S'S$, $S'T$ and $T'T$, which allows us to consider modified versions of the fixed-T and non-parametric conditional bootstraps. For example, Moreira (2003) shows that the LR statistic can be rewritten as

$$LR = \frac{1}{2}\left[Q_1 + Q_{k-1} - T'T + \sqrt{(Q_1 + Q_{k-1} + T'T)^2 - 4Q_{k-1}T'T}\right], \tag{4}$$

where $Q_1 = S'T(T'T)^{-1}T'S$ and $Q_{k-1} = S'[I - T(T'T)^{-1}T']S$. Conditional on $T'T = \tau$, Q_1 and Q_{k-1} are asymptotically independent, and, under the null hypothesis, have limiting chi-square distributions with one and $k-1$ degrees of freedom, respectively.

The first conditional bootstrap method adapted to similar tests exploits the asymptotic first-order independence of $Q = (S'S, S'T/\sqrt{T'T})$ and $T'T$. For each bootstrap sample, the bootstrap version of the statistic Q, denoted Q^*, is generated. The bootstrap critical value is then the $1-\alpha$ quantile of the empirical distribution of $LR(Q^*, T'T)$. Note that $T'T$ is fixed at its observed value here. The second conditional bootstrap procedure is based on the non-parametric method described in Booth, Hall, and Wood (1992). Suppose B bootstrap samples are generated. Let Q_j^* and $T_j^{*\prime}T_j^*$ denote the values of Q

and $T'T$ in the j-th bootstrap sample. Booth, Hall, and Wood (1992) suggest using a standard non-parametric kernel estimate of the desired conditional distribution based on these bootstrap samples. Therefore, the problem of finding the critical value of the LR statistic conditional on $T'T = \tau$ boils down to determining the value $x(\tau)$ such that

$$\frac{\frac{1}{B}\sum_{j=1}^{B} \mathbf{1}\left[LR(Q_j^*, T_j^{*\prime}T_j^*) \leq x(\tau))\right] \phi\left(\frac{T_j^{*\prime}T_j^* - \tau}{h}\right)}{\frac{1}{B}\sum_{j=1}^{B} \phi\left(\frac{T_j^{*\prime}T_j^* - \tau}{h}\right)} = 1 - \alpha,$$

where $\mathbf{1}[\cdot]$ is an indicator function, $\phi(\cdot)$ is a kernel function, and h is a bandwidth parameter. In section 5, each of these bootstrap procedures is implemented and compared in a Monte Carlo exercise.

4.2 Edgeworth Expansions

The Anderson-Rubin, likelihood ratio, Wald, and score statistics can all be written as functions of averages for various moments of the data. For the Wald statistic, this function includes a division by zero under Assumption 1A when evaluated at the expected values of the averages. Hence, the results in Bhattacharya and Ghosh (1978) are unavailable for the Wald statistic. A similar problem holds for the score and likelihood ratio statistics. In particular, for one-sided statistics of the form $\sqrt{n}\left(H\left(\overline{R}_n\right) - H\left(\mu\right)\right)$, Bhattacharya and Ghosh (1978)'s approach uses

$$\begin{aligned} g_n\left(z\right) &= n^{1/2}\left[H\left(\mu + n^{-1/2}z\right) - H\left(\mu\right)\right] \text{ and} \\ h_{s-1}\left(z\right) &= \sum_i l_i z^{(i)} + \frac{1}{2}n^{-1/2}\sum_{i,j} l_{i,j} z^{(i)} z^{(j)} + \cdots \\ &\quad + \frac{1}{(s-1)!} n^{-(s-2)/2} \sum_{i_1,\cdots,i_{s-1}} l_{i_1,\cdots,i_{s-1}} z^{(i_1)} \cdots z^{(i_{s-1})}, \end{aligned}$$

where $l_{i_1,\cdots,i_p} = \left(D_{i_1}\cdots D_{i_p}H\right)(\mu)$ and D_i denotes differentiation with respect to the i-th coordinate, to show that

$$\begin{aligned} P\left(\sqrt{n}\left(H\left(\overline{R}_n\right) - H\left(\mu\right)\right) \in B\right) &= \int_{\{g_n(z)\in B\}} \xi_{s,n}\left(z\right) dz + o\left(n^{-(s-2)/2}\right) \\ &= \int_{\{h_{s-1}(z)\in B\}} \xi_{s,n}\left(z\right) dz + o\left(n^{-(s-2)/2}\right) \end{aligned}$$

for a suitably chosen Edgeworth expansion $\xi_{s,n}(z)$. However, the one-sided score statistic is not differentiable at $\pi = 0$, and so we cannot obtain $h_{s-1}(z)$. This problem makes Bhattacharya and Ghosh (1978)'s expansion method unavailable; cf. Moreira, Porter, and Suarez (2005). A similar argument holds for the likelihood ratio statistic with respect to the results in Chandra and Ghosh (1979). More specifically, the likelihood ratio statistic is not second-order differentiable at $\pi = 0$, making Chandra and Ghosh (1979)'s expansion method unavailable.

In fact, little is known about expansions when the statistic is not smooth. In the words of Wallace (1958), "The assumption $H'(\mu) \neq 0$ and its equivalent for functions of several moments rule out many interesting functions for which no general theory of asymptotic expansions is known." Yet, there has not been any work in the last decades on expansions for non-smooth statistics. The next result provides a first step towards developing a framework to analyze asymptotic expansions for statistics in the weak-instrument setting. This is done by re-writing statistics as functions of S and T.

Lemma 5 *Let $\mathcal{B}$ be any class of Borel Sets satisfying*

$$\sup_{B\in\mathcal{B}} \int_{(\partial B)^{\varepsilon}} \boldsymbol{\phi}_A(v)\, dv = O(\epsilon) \quad \text{as } \epsilon \downarrow 0, \tag{5}$$

where $\boldsymbol{\phi}_A$ is the pdf of a mean zero normal distribution with variance A, ∂B is the boundary of B, and $(\partial B)^{\varepsilon}$ is the ε-neighborhood of A. If Assumptions 1 or 1A, 2, and 3 hold, then

$$\sup_{B\in\mathcal{B}} \left| P\left((S_n', T_n' - t_n')' \in B\right) - \int_B \psi_{s,n}(v)\, dv \right| = o\left(n^{-(s-2)/2}\right)$$

under $H_0 : \beta = \beta_0$*, where* $t_n = \sqrt{n}\Omega_{ZZ}^{1/2}\pi \cdot (a_0'\Omega^{-1}a_0)^{1/2}$, $\Omega_{ZZ} = E\left(Z_i Z_i'\right) < \infty$, *and* $\psi_{s,n}$ *is a formal Edgeworth expansion of order* $s-2$ *for* S_n *and* $T_n - t_n$*. In particular, the joint null distribution of* S *and* T *can be uniformly (in* $x = (x_1', x_2')'$*) approximated by an Edgeworth expansion*

$$P\left(S_n \leq x_1, T_n - t_n \leq x_2\right) = \mathbf{\Phi_A}\left(x\right) + \sum_{i=1}^{s-2} n^{-i/2}\mathbf{q}^i\left(x; F, \beta_0, \pi\right)\boldsymbol{\phi}_A\left(x\right)$$

up to an $o\left(n^{-(s-2)/2}\right)$ *term, where* $\mathbf{\Phi}_A$ *is the cdf of a mean zero normal distribution with variance* A*.*

Proof. The statistics S and $T - t_n$ can be written as

$$\begin{aligned} S &= \sqrt{n}\left(Z'Z/n\right)^{-1/2}\left(Z'V/n\right)b_0 \cdot (b_0'\Omega b_0)^{-1/2}, \\ T - t_n &= \sqrt{n}\left[\begin{array}{c} (Z'Z/n)^{1/2}\pi \cdot (a_0'\Omega^{-1}a_0)^{1/2} - \Omega_{ZZ}^{1/2}\pi \cdot (a_0'\Omega^{-1}a_0)^{1/2} \\ +(Z'Z/n)^{-1/2}(Z'V/n)\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \end{array}\right] \end{aligned}$$

under H_0. Therefore,

$$(S', T' - t_n')' = \sqrt{n}\left(H\left(\overline{R}_n\right) - H\left(\mu\right)\right)$$

for a measurable mapping H from $R^{(k+2)(k+3)/2}$ onto R^{2k} with derivatives of order s and less being continuous in the neighborhood of μ. Using the results for the multivariate case by Bhattacharya and Ghosh (1978, p. 437), S and $T - t_n$ admit an Edgeworth expansion:

$$\psi_n\left(s,t\right) = \left(1 + \sum_{i=1}^{s-2} n^{-i/2}P_i\left(-D : F\right)\right)\boldsymbol{\phi}_A\left(s,t\right),$$

where $\boldsymbol{\phi}_M\left(s,t\right)$ is the normal density on R^{2k} with mean zero and dispersion M, $P_i\left(-D : F\right)$ is a polynomial in p variables whose coefficients do not depend on n, and $-D = (-D_1, ..., -D_{2k})$. An analogous result holds for statistics in the unknown variance case, $\widetilde{S}$ and $\widetilde{T}$, albeit the Edgeworth expansion would have different polynomials for the higher-order terms. □

An explicit expression for A is given in section 4.1. It should be noted here that when $\pi = 0$, $A = I_{2k}$. Otherwise, A is a block diagonal matrix with upper diagonal block I_k. Also, the term $t_n = \sqrt{n}\Omega_{ZZ}^{1/2}\pi \cdot (a_0'\Omega^{-1}a_0)^{1/2}$ changes with the sample size. This adjustment is due to the fact that the mean of T drifts off to infinity when $\pi \neq 0$, and guarantees an Edgeworth expansion. Note that we can understand the weak-instrument asymptotics as if the drift term t_n were fixed at the level

$$\Omega_{ZZ}^{1/2}c \cdot (a_0'\Omega^{-1}a_0)^{1/2}.$$

Thus, Lemma 5 can be seen as a higher-order expansion to the weak-instrument asymptotics of Staiger and Stock (1997). In particular, it allows us to analyze the behavior of many tests in the unidentified case. As a direct application, we show that the Anderson-Rubin statistic admits a standard expansion even when $\pi = 0$:

Theorem 6 *Under Assumptions 1A, 2, and 3, the null distributions of AR_n can be uniformly approximated (in x) by*
$P\left(AR_n \leq x\right) = G_k\left(x\right) + \sum_{i=1}^{r} n^{-i} p_{AR}^{i}\left(x; F, \beta_0, \pi\right) g_k\left(x\right) + o\left(n^{-r}\right).$

Proof: We want to approximate

$$P\left(AR_n \leq x\right) = P\left(S_n' S_n \leq x\right)$$

uniformly in x. This expression can be written as $P\left((S_n', T_n')' \in C_x\right)$, for the convex sets

$$C_x = \left\{(s', t')' \in \mathbb{R}^{2k}; s's \leq x\right\}.$$

Using Corollary 3.2 of Bhattacharya and Rao (1976), we can show that

$$\sup_{x \in \mathbb{R}} \Phi\left(\left(\partial C_x\right)^{\varepsilon}\right) \leq d\left(k\right)\varepsilon,$$

where $d\left(k\right)$ is a function of only k and $\varepsilon > 0$. Therefore, Lemma 5 holds. Finally, for odd values of i, the terms $P_i\left(-D : F\right)\boldsymbol{\phi}\left(s, t\right)$ are even. Consequently, the terms of $n^{-i/2}$ order vanish, and we get the desired result by integration. □

Comments: 1. We could in principle apply Lemma 5 to approximate: (i) $P(LM_n \leq x)$ uniformly in x, (ii) $P(LR_n \leq x_1 | T_n \leq x_2)$ uniformly in $x = (x_1, x_2)$; and, (iii) $P(LR(S_n, t) \leq x)$ uniformly in (x, t). It may not be possible to provide an approximation for (i) and (ii) of order $o\left(n^{-(s-2)/2}\right)$ uniformly in x, since the boundary of sets $C_x = \left\{(s', t')' \in \mathbb{R}^{2k}; \psi(s, t) \leq x\right\}$ associated to a statistic $\psi(S, T)$ may not be well-behaved in the sense of (5). For (iii), we can get an approximation for (fixed-T) CLR uniformly in x (but pointwise in $t \neq 0$).

2. Moreira, Porter, and Suarez (2005) show that LM_n is not differentiable and the higher-order terms of an Edgeworth expansion for LM_n cannot in general be extended continuously at $\pi = 0$. This indicates that approximating $P(LM_n \leq x)$ by an asymptotic expansion up to an $o\left(n^{-(s-2)/2}\right)$ term uniformly in x is not be possible at $\pi = 0$. Nevertheless, $\int_{C_x} \psi_{s,n}(v)\, dv$, where $C_x = \left\{(s', t')' \in \mathbb{R}^{2k}; s't/\sqrt{t't} \leq x\right\}$, may be a better numerical approximation to the $P(LM_n \leq x)$ than $\Phi(x)$ (note that the leading term $\int_{C_x} \psi_{s,n}(v)\, dv$ is the c.d.f. of a standard normal). Using Lemma 5 to provide asymptotic expansions for LM_n (and standard CLR_n) when Bhattacharya and Ghosh (1978)'s results are unavailable will be considered elsewhere.

Close inspection of the proof of Theorems 1(c) and 6(b) shows that the higher-order polynomials for the AR statistic are continuous even at $\pi = 0$.[6] Therefore, unlike the bootstrap (it replaces β_0 by an estimator $\widehat{\beta}$), an empirical Edgeworth expansion for the AR statistic does provide an $o(n^{-1})$ improvement from a chi-square-k distribution even with weak instruments. Unfortunately, empirical Edgeworth involves extensive algebraic manipulations to find the higher-order terms.

5 Monte Carlo Simulations

Theorem 2 suggests that the bootstrap can decrease size distortions for the Anderson-Rubin and likelihood ratio tests when instruments are good. More importantly, Theorem 3 and its corollary provide theoretical support for

[6] We actually use the argument of Lemma 5 to derive Theorem 6(b).

bootstrapping the Anderson-Rubin and CLR tests, even when instruments are weak. The same validation of the bootstrap does not hold for the Wald or likelihood ratio tests. In this section, we focus on the CLR test. For bootstrap simulations results for the Wald and score statistics, see Moreira, Porter, and Suarez (2005).

In the following set of results, we compare the sizes of the CLR test when based on the two conditional bootstrap methods for computing the critical value function. We calculate actual rejection probabilities of nominal 5% tests based on these two methods using 1000 simulations. We follow designs I and II of Staiger and Stock (1997). We simulate the simple model introduced in equations (1) and (2). The true value of the structural parameter, β, is assumed to be zero. We assume that the n rows of $[u, v_2]$ are i.i.d. random variables with mean zero, unit variance, and correlation coefficient ρ. The correlation coefficient represents the degree of endogeneity of y_2. The first column of the matrix of instruments, Z, is a vector of ones and the other $k-1$ columns are drawn from independent standard normal distributions, which are independent from $[u, v_2]$.[7] We consider bootstrap performance under three different degrees of identification, corresponding to three different values of $\pi'(nI_k)\pi/k$, the population first-stage F-statistic: the good instrument case ($\pi'(nI_k)\pi/k = 10$), the weak-instrument case ($\pi'(nI_k)\pi/k = 1$), and the completely unidentified case ($\pi'(nI_k)\pi/k = 0$). Note that our population first-stage F-statistic corresponds to the concentration parameter $\lambda'\lambda/k$ in Staiger and Stock (1997). For design I, we assume that u_t and v_{2t} are normally distributed with unit variance and correlation ρ. For design II, we assume that $u_t = (\xi_{1t}^2 - 1)/\sqrt{2}$ and $v_{2t} = (\xi_{2t}^2 - 1)/\sqrt{2}$, where ξ_{1t} and ξ_{2t} are standard normal random variables with correlation $\sqrt{\rho}$.

Table I shows rejection rates computed using the fixed-T conditional bootstrap. The rejection probabilities using bootstrap critical values are considerably smaller than those using the critical value function of Moreira (2003). The size distortions obtained by the bootstrap are particularly important when instruments are weak. This seems to hold for different values of ρ,

[7] There is a slight difference between Moreira's (2001, 2003) design and ours. Our design takes Z as being random whereas Moreira's (2001, 2003) design takes Z as being fixed.

sample sizes ($n = 20$ or 80), and error distributions (normal or Wishart).[8] The non-parametric conditional bootstrap method can in principle work even better than the fixed-T conditional bootstrap. The non-parametric bootstrap offers second-order improvements at least in the good-instrument case. Tables II and III summarize the results for the non-parametric bootstrap with Gaussian kernel for different sample sizes ($n = 20$ or 80) and error distributions (normal or Wishart). In general, the non-parametric bootstrap offers size improvements over the critical value function, but its performance is below the fixed-T bootstrap. The nonparametric procedure is not very sensitive to the choice of bandwidth, although an intermediate value of the bandwidth parameter tends to outperform extreme choices. Finally, we consider other kernels, such as the Epanechnikov and truncated types. Simulations not reported here suggest that our results are not very sensitive to the choice of kernel function.

6 Conclusions and Extensions

It is well-known that the Wald statistic admits higher-order Edgeworth expansions *under some regularity conditions*. Replacing the unknown parameters by consistent estimators and using the continuity of the polynomials in the higher-order terms guarantee that empirical Edgeworth expansions lead to smaller size distortions than those found when using the chi-square-one critical value. Computing the critical value with the bootstrap also leads to size improvements given the asymptotic equivalence between the bootstrap and the empirical Edgeworth expansion up to higher-order terms. However, when the instruments are uncorrelated with the endogenous explanatory variable, these regularity conditions break down. The consequences of this breakdown are threefold. First, the Wald statistic no longer admits a standard higher-order Edgeworth expansion. Second, the Wald statistic is a

[8] We chose a small sample size to have meaningful size comparisons using chi-square and bootstrapped critical values with normal and Wishart error distributions. It is well-known that size distortions with large samples can happen in the presence of either different error distributions or heretoskedasticity.

non-differentiable function of sample means and, consequently, non-regular. Third, the bootstrap and the empirical Edgeworth expansion approaches replace unknown parameters by estimators that are inconsistent in the unidentified model.

Our contribution to the weak-instrument literature is as follows. First, we show that the bootstrap is valid for the Anderson-Rubin test. Second, we provide a theoretical proof that the bootstrap does not work for the likelihood ratio (and Wald) statistic. This is rather expected, given that the likelihood ratio statistic is not well-behaved with weak instruments. However, we are unaware of any *formal* proof in the literature of bootstrap failure with weak instruments. Third, we propose two conditional bootstrap methods for the CLR test. Fourth, we provide a theoretical framework to derive asymptotic expansions in the weak-instrument setting. As a direct result, we show that an empirical Edgeworth expansion for the Anderson-Rubin statistic provides an $o(n^{-1})$ improvement to chi-square-k distribution. We can also obtain asymptotic expansions for other statistics (e.g., fixed-T CLR).

Finally, our results can, in principle, be extended to the GEL and GMM contexts; cf. Guggenberger and Smith (2003) and Stock and Wright (2000). Simulations in Inoue (2002) and Kleibergen (2003) indicate that the bootstrap can lead to size improvements for the unidentified case in the GMM context as well. By extending our two-step approach to analogous sufficient statistics for the GMM and GEL cases, a formal proof of the validity of the bootstrap and Edgeworth expansions in the (locally) unidentified case should be available.

References

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2003): "Optimal Invariant Similar Tests for Instrumental Variables Regression," *Un-*

published Manuscript, Yale University.

——— (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74, forthcoming.

Babu, J., and K. Singh (1984): "On one term Edgeworth correction by Effron's bootstrap," *Sankhya, Series A*, 46, 219–32.

Bhattacharya, R., and R. Rao (1976): *Normal Approximation and Asymptotic Expansions*, Wiley Series in Probability and Mathematical Analysis.

Bhattacharya, R. N. (1977): "Refinements of the Multidimensional Central Limit Theorem and Applications," *Annals of Probability*, 5, 1–27.

——— (1987): "Some Aspects of Edgeworth Expansions in Statistics and Probability," in *New Perspectives in Theoretical and Applied Statistics*, ed. by M. L. Puri, J. P. Vilaplana, and W. Wertz, pp. 157–170. New York: John Wiley and Sons.

Bhattacharya, R. N., and J. Ghosh (1978): "On the Validity of the Formal Edgeworth Expansion," *Annals of Statistics*, 6, 434–451.

Bickel, P. J., and D. A. Freedman (1981): "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics*, 9, 1196–1217.

Booth, J., P. Hall, and A. Wood (1992): "Bootstrap Estimation of Conditional Distributions," *The Annals of Statistics*, 20, 1594–1610.

Bound, J., D. A. Jaeger, and R. M. Baker (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of American Statistical Association*, 90, 443–450.

Chambers, J. M. (1967): "On Methods of Asymptotic Approximation for Multivariate Distributions," *Biometrika*, 54, 367–383.

Chandra, T., and J. Ghosh (1979): "Valid Asymptotic Expansions for the Likelihood Ratio Statistic and Other Perturbed Chi-Square Variables," *Sankhya*, 41, 22–47.

Chibishov, D. M. (1972): "An asymptotic expansion for the distribution of a statistic admitting an asymptotic expansion," *Theory of Probability and its Applications*, 17, 620–630.

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

Efron, B. (1979): "Bootstrap Methods: Another Look at the Jacknife," *The Annals of Statistics*, 7, 1–26.

Guggenberger, P., and R. Smith (2003): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak and Strong Identification," *CEMAP Working Paper*, CWP08/03.

Hahn, J., J. Hausman, and G. Kuersteiner (2002): "Estimation with Weak Instruments: Accuracy of Higher Order Bias and MSE Approximations," *Unpublished Manuscript*, MIT.

Horowitz, J. (2001): "The Bootstrap," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5, chap. 52, pp. 3159–3228. New York: North-Holland.

Inoue, A. (2002): "A Bootstrap Approach to Moment Selection," *Unpublished Manuscript*, North Caroline State University.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

Kleibergen, F. (2003): "Expansions of GMM statistics that indicate their properties under weak and/or many instruments and the bootstrap," *Unpublished Manuscript*, University of Amsterdam.

Moreira, M. J. (2001): "Tests with Correct Size when Instruments Can Be Arbitrarily Weak," *Center for Labor Economics Working Paper Series*, 37, UC Berkeley.

——— (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

Moreira, M. J., J. R. Porter, and G. A. Suarez (2005): "Bootstrap Validity for the Score Test when Instruments May Be Weak," *Unpublished Manuscript*, Harvard University.

Navidi, W. (1989): "Edgeworth Expansions for Bootstrapping Regression Models," *The Annals of Statistics*, 17, 1472–1478.

Nelson, C. R., and R. Startz (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

Phillips, P. C. B. (1977): "A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Econometric Estimators," *Econometrica*, 45, 1517–1534.

Qumsiyeh, M. B. (1990): "Edgeworth Expansion in Regression Models," *Journal of Multivariate Analysis*, 35, 86–101.

——— (1994): "Bootstrapping and Empirical Edgeworth Expansions in Multiple Linear Regression Models," *Communication in Statistics - Theory and Methods*, 23, 3227–3239.

Rothenberg, T. J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. Intriligator, vol. 2, chap. 15, pp. 881–935. Elsevier Science, Amsterdam.

Sargan, J. D. (1976): "Econometric Estimators and the Edgeworth Approximation," *Econometrica*, 44, 421–448.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Stock, J. H., and J. Wright (2000): "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

Wallace, D. (1958): "Asymptotic Approximations to Distributions," *The Annals of Mathematical Statistics*, 29, 635–654.

Wang, J., and E. Zivot (1998): "Inference on a Structural Parameter in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389–1404.

Appendix A - Proofs

Proof of Theorem 1. To prove part (a), we can also rely on Bhattacharya and Ghosh (1978). The S statistic can be written as

$$S = \sqrt{n}\left(H\left(\overline{R}_n\right) - H\left(\mu\right)\right)$$

under H_0, where H is a Borel measurable function that maps onto $\mathbb{R}^k$ such that $H(\mu) = 0$. All derivatives of H of order s and less are continuous in the neighborhood of μ. This guarantees an Edgeworth expansion for S,

$$\psi_n(s) = \left(1 + \sum_{i=1}^{s-2} n^{-i/2} P_i\left(-D : F\right)\right) \boldsymbol{\phi}(s),$$

where D is the differential operator and $P_i(-D : F)$ are suitably chosen polynomials. We can use Chibishov (1972):

$$\sup_{C \in \mathcal{C}} \left| P\left(S_n \in C\right) - \int_C \psi_{n,s}(u)\, du \right| = o\left(n^{-(s-2)/2}\right)$$

holds uniformly over every $\mathcal{C}$ of measurable convex sets. In particular, this holds for the class $\{C_x; x \in \mathbb{R}\}$, where

$$C_x = \left\{s \in \mathbb{R}^2; s's \leq x\right\}.$$

For odd values of i, the terms $P_i(-D:F)\,\boldsymbol{\phi}_A(s,t)$ are even. Consequently, the terms of $n^{-i/2}$ order vanish. The Anderson-Rubin statistic for unknown variance, $\widetilde{AR}$, also admits an Edgeworth expansion by first finding an expansions for $\widetilde{S}$ and then proceeding as above.

Part (b) is a direct result of Chandra and Ghosh (1979). □

Proof of Theorem 2. Let F be the distribution of

$$R_n = vech\left((Y_n', Z_n')'(Y_n', Z_n')\right)$$

and let F_n be the distribution of

$$\widetilde{R}_n^* = vech\left((\widetilde{Y}_n^{*\prime}, Z_n^{*\prime})'(\widetilde{Y}_n^{*\prime}, Z_n^{*\prime})\right)$$

conditional on $\mathcal{X}_n = \{(Y_1', Z_1'), \ldots, (Y_n', Z_n')\}$. Here, Z_n^* has probability $1/n$ in taking the values Z_n, and Y_n^* has probability $1/n$ in taking the values

$$\widetilde{Y}_n = Z_n\widehat{\pi}\widehat{a} + \widetilde{V}_n = Z_n\widehat{\pi}(\widehat{\beta}, 1) + \widetilde{V}_n.$$

The re-sampling mechanism for $\widetilde{Y}_n$ and Z_n and the re-centering procedure for $\widehat{V}$ of subtracting sample means reflect the fact that Z and V are independent. If Z and V were uncorrelated, it would entail different drawing mechanisms and re-centering procedures. But the essence of the proofs for the bootstrap presented here would remain the same.

Let $\widehat{F}_n$ be the Fourier transform of F_n and

$$\widetilde{R}_n = vech\left((\widetilde{Y}_n', Z_n')'(\widetilde{Y}_n', Z_n')\right).$$

Following Lemma 2 of Babu and Singh (1984), there exists for each $d > 0$ positive numbers ϵ and δ such that

$$\limsup_{n\to\infty} \sup_{d\le\|t\|\le e^{n\delta}} \left|\widehat{F}_n(t)\right| \le 1 - \epsilon \text{ a.s.}$$

Let D be the differential operator, ∂B be the boundary of B, and $(\partial B)^\varepsilon$ be the ε-neighborhood of B. Since the rows $\widetilde{R}_n^*$ are i.i.d. (conditionally given

$\mathcal{X}_n$) with common distribution F_n, one can proceed as in Bhattacharya (1987) to show that for suitably chosen polynomials $P_i\left(-D : F_n\right)$

$$\sup_{B\in\mathcal{B}}\left|P^*\left(\sqrt{n}\left(\widetilde{\overline{R}}_n^* - \widetilde{\overline{R}}_n\right)\in B\right) - \int_B\left[1+\sum_{i=1}^{s-2} n^{-i/2}P_i\left(-D : F_n\right)\right]\boldsymbol{\phi}_V\left(x\right)dx\right|$$

is $o\left(n^{-(s-2)/2}\right)$ a.s. as $n \to \infty$ for every class $\mathcal{B}$ of Borel subsets of $\mathbb{R}^\ell$ satisfying, for some $\theta > 0$,

$$\sup_{B\in\mathcal{B}}\boldsymbol{\Phi}_V\left(\left(\partial B\right)^\varepsilon\right) = O\left(\varepsilon^\theta\right) \text{ as } \varepsilon \downarrow 0.$$

Reduction of the expansion of $\sqrt{n}\left(\widetilde{\overline{R}}_n^* - \widetilde{\overline{R}}_n\right)$ to AR^* and LR^* statistics follows as in the proof of Theorem 1. □

Lemma A *Suppose* $\widehat{\pi}\widehat{\beta} - \pi_n\beta \overset{a.s.}{\to} 0$. *If, for some* $\delta > 0$, $E\|Z_i\|^{2+\delta} < \infty$, $E\|v_i\|^{2+\delta} < \infty$, *then for* $j1,\ldots,k$ and $m = 1,2$, $E^*[|Z^*_{j,i}v^*_{m,i}|^{2+\delta}]$ *is bounded a.s.*

Proof. By independence, $E^*[|Z^*_{j,i}v^*_{m,i}|^{2+\delta}] = E^*[|Z^*_{j,i}|^{2+\delta}]E^*[|v^*_{m,i}|^{2+\delta}]$. For $j = 1,\ldots,k$,

$$E^*[|Z^*_{j,i}|^{2+\delta}] = \frac{1}{n}\sum_i |Z_{j,i}|^{2+\delta} \overset{a.s.}{\to} E[|Z_{j,i}|^{2+\delta}].$$

Let $\overline{v}_m = \frac{1}{n}\sum_{i=1}^n v_{m,i}$, $m = 1,2$, and $\overline{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$. For $m = 1,2$,

$$\begin{aligned}
E^*[|v^*_{m,i}|^{2+\delta}] &= n^{-1}\sum_{i=1}^n |\tilde{v}_{m,i}|^{2+\delta} \\
&= n^{-1}\sum_{i=1}^n |v_{m,i} - \overline{v}_m - \left(Z_i - \overline{Z}\right)'(\widehat{\pi}\widehat{\beta} - \pi_n\beta)|^{2+\delta} \\
&\leq C\left\{n^{-1}\sum_{i=1}^n |v_{m,i} - \overline{v}_m|^{2+\delta} + n^{-1}\sum_{i=1}^n |\left(Z_i - \overline{Z}\right)'(\widehat{\pi}\widehat{\beta} - \pi_n\beta)|^{2+\delta}\right\} \\
&\leq C\left\{n^{-1}\sum_{i=1}^n |v_{m,i} - \overline{v}_m|^{2+\delta} + \left\|\widehat{\pi}\widehat{\beta} - \pi_n\beta\right\|^{2+\delta} n^{-1}\sum_{i=1}^n \left\|Z_i - \overline{Z}\right\|^{2+\delta}\right\}.
\end{aligned}$$

for a large enough constant C. The first inequality follows from Minkowski's inequality, and the second inequality uses Cauchy-Schwartz inequality.

Using Minkowski's inequality, we get

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left\|Z_i-\overline{Z}\right\|^{2+\delta} &\leq C\left\{n^{-1}\sum_{i=1}^{n}\left\|Z_i\right\|^{2+\delta}+\left\|\overline{Z}\right\|^{2+\delta}\right\} \\ &\overset{a.s.}{\rightarrow} C\left\{E\left[\left\|Z_i\right\|^{2+\delta}\right]+\left\|E\left[Z_i\right]\right\|^{2+\delta}\right\}, \end{aligned}$$

using $\left\|\overline{Z}\right\| \overset{a.s.}{\rightarrow} \left\|E\left[Z_i\right]\right\| \leq E\left\|Z_i\right\| \leq (E\left\|Z_i\right\|^{2+\delta})^{1/(2+\delta)}$ by Jensen's inequality. Similarly, using Minkowski's inequality again, we obtain

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left|v_{m,i}-\overline{v}_m\right|^{2+\delta} &\leq C\left\{n^{-1}\sum_{i=1}^{n}\left|v_{m,i}\right|^{2+\delta}+\left|\overline{v}_m\right|^{2+\delta}\right\} \\ &\overset{a.s.}{\rightarrow} C\left\{E\left|v_{m,i}\right|^{2+\delta}\right\}, \end{aligned}$$

as $\overline{v}_m \overset{a.s.}{\rightarrow} 0$. Since $\widehat{\pi}\widehat{\beta}-\pi_n\beta \overset{a.s.}{\rightarrow} 0$, the term $n^{-1}\sum_i\left|\tilde{v}_{m,i}\right|^{2+\delta}$ is bounded a.s. □

Recall the following notation $w_i = vech\left(Z_iZ_i'\right) \in R^{k(k+1)/2}$ and $W = (w_1,\ldots,w_n)$. Similarly, let $w_i^* = vech\left(Z_i^*Z_i^{*\prime}\right)$ and $W^* = (w_1^*,\ldots,w_n^*)$. Also let $\Omega_{ww} = \mathrm{Var}(w_i)$ and let $\imath$ be an $n\times 1$ vector of ones.

Lemma B *If, for some* $\delta>0$, $E\|Z_i\|^{4+\delta}<\infty$, $E\|v_{\cdot i}\|^{2+\delta}<\infty$, *then*

$$\left(\begin{array}{c} \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right)\frac{\widehat{b}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}} \\ \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right)\frac{\widehat{\Omega}^{-1}\widehat{a}}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}} \\ \sqrt{n}\left(\frac{W^{*\prime}\imath}{n}-\frac{W'\imath}{n}\right) \end{array}\right)\Bigg|\mathcal{X}_n \overset{d}{\rightarrow} N\left(0,\left(\begin{array}{cc} I_2\otimes E(Z_iZ_i') & 0 \\ 0 & \Omega_{ww}\end{array}\right)\right) \quad \text{a.s.}$$

Proof. Let $(c',d')'$ be a nonzero vector with $c=(c_1',c_2')'\in\mathbb{R}^{2k}$ and $d\in\mathbb{R}^{k(k+1)/2}$. Define

$$X_{n,i}=\left\{c'(\widehat{J}'\otimes I_k)\left(v_{\cdot i}^*\otimes Z_i^*\right)+d'\left(w_i^*-\overline{w}\right)\right\}/\sqrt{n},$$

where $v_{\cdot i}^* = \left[v_{1,i}^*, v_{2,i}^*\right]'$ is the i-th bootstrap draw of the (recentered) reduced-form residuals, $\overline{w} = n^{-1}\sum_{i=1}^n w_i$, and

$$\widehat{J} = \left[\frac{\widehat{b}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}}, \frac{\widehat{\Omega}^{-1}\widehat{a}}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}}\right].$$

We use the Cramér-Wald device and verify the conditions of the Liapunov Central Limit Theorem.

(a) $E^*\left[X_{n,i}\right] = 0$ follows from independence and $E^*\left[v_{\cdot i}^*\right] = 0$.

(b) By independence,

$$E^*\left[X_{n,i}^2\right] = n^{-1}\left\{c'\left[I_2 \otimes \left(\frac{Z'Z}{n}\right)\right]c + d'\widehat{\Omega}_{ww}d\right\},$$

is finite a.s., where $\widehat{\Omega}_{ww} = n^{-1}\sum_{i=1}^n (w_i - \overline{w})(w_i - \overline{w})'$.

(c) Finally, we need to show that $\lim_{n\to\infty}\sum_{i=1}^n \frac{1}{s_n^{2+\delta}}E^*[|X_{n,i}|^{2+\delta}] = 0$, where $s_n^2 = \sum_{i=1}^n E^*\left[X_{n,i}^2\right]$. Note that

$$\begin{aligned}
\sum_{i=1}^n E^*\left[|X_{n,i}|^{2+\delta}\right] &\leq C_1 n^{-\frac{\delta}{2}} n^{-1}\sum_{i=1}^n E^*\left[|c'(\widehat{J}'v_{\cdot i}^* \otimes Z_i^*)|^{2+\delta} + |d'(w_i^* - \overline{w})|^{2+\delta}\right] \\
&\leq C_2 n^{-\frac{\delta}{2}} E^*\left[|c_1' Z_i^*(\widehat{J}'v_{\cdot i}^*)_1|^{2+\delta} + |c_2' Z_i^*(\widehat{J}'v_{\cdot i}^*)_2|^{2+\delta} + |d'(w_i^* - \overline{w})|^{2+\delta}\right] \\
&\leq C_3 n^{-\frac{\delta}{2}} \cdot \Bigg\{\sum_{j=1}^k \Bigg[\left(\left|\frac{c_{1,j}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}}\right|^{2+\delta} + \left|\frac{c_{2,j}(\widehat{\Omega}^{-1}\widehat{a})_1}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}}\right|^{2+\delta}\right)E^*\left[|Z_{j,i}^* v_{1,i}^*|^{2+\delta}\right] \\
&\quad + \left(\left|\frac{c_{1,j}(-\widehat{\beta})}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}}\right|^{2+\delta} + \left|\frac{c_{2,j}(\widehat{\Omega}^{-1}\widehat{a})_2}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}}\right|^{2+\delta}\right)E^*\left[|Z_{j,i}^* v_{2,i}^*|^{2+\delta}\right]\Bigg] \\
&\quad + \sum_{l=1}^{(k+1)k/2} E^*\left|w_{l,i}^* - \left(\frac{1}{n}\sum_{j=1}^n w_{j,i}\right)\right|^{2+\delta}\Bigg\}.
\end{aligned}$$

for large enough constants C_1, C_2, and C_3.

The vectors $\widehat{a}$ and $\widehat{b}$ both have one as an element, and $\widehat{\Omega}$ and $\widehat{\Omega}^{-1}$ converge almost surely to positive definite limits. So, regardless of the value of π_n or $\widehat{\beta}$, the terms

$$\left|\frac{1}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}}\right|, \left|\frac{(\widehat{\Omega}^{-1}\widehat{a})_1}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}}\right|, \left|\frac{-\widehat{\beta}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}}\right|, \text{and } \left|\frac{(\widehat{\Omega}^{-1}\widehat{a})_2}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}}\right| \tag{6}$$

are almost always well-defined. These terms are also bounded by

$$\max\left\{\sqrt{\frac{\widehat{\sigma}_{11}}{\widehat{\sigma}_{11}\widehat{\sigma}_{22}-\widehat{\sigma}_{12}^2}},\sqrt{\frac{\widehat{\sigma}_{22}}{\widehat{\sigma}_{11}\widehat{\sigma}_{22}-\widehat{\sigma}_{12}^2}}\right\}, \tag{7}$$

where $\widehat{\sigma}_{ij}$ is the (i,j)-th entry of $\widehat{\Omega}$.

This bound follows from the fact that

$$\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}=\widehat{a}'\widehat{\Omega}^{-1}\widehat{\Omega}\widehat{\Omega}^{-1}\widehat{a},$$

and the following claim (which holds regardless of the value of π). Let

$$K=\begin{pmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{pmatrix} \quad \text{and} \quad \tau=\begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix},$$

where K is a symmetric positive definite matrix. Then, the following holds:

$$\begin{aligned} \left|\frac{\tau_1}{\sqrt{\tau' K\tau}}\right| &\leq \sup_{\tau}\sqrt{\frac{\tau'\left(e_1e_1'\right)\tau}{\tau' K\tau}} \\ &= \sqrt{\frac{k_{22}}{k_{11}k_{22}-k_{12}^2}}. \end{aligned}$$

Given the bound in (7), the conclusion of Lemma A, and the fact that $E\|Z_i\|^{4+\delta}<\infty$ is sufficient to bound $E^*\left|w_{li}^*-\left(n^{-1}\sum_{j=1}^n w_{li}\right)\right|^{2+\delta}$ almost surely, the final condition of the Liapunov Central Limit Theorem now follows because

$$\left[c'\left(I_2\otimes(Z'Z/n)\right)c+d'\widehat{\Omega}_{ww}d\right]^{-\left(1+\frac{\delta}{2}\right)}$$

is bounded away from zero almost surely since $(Z'Z/n)$ and $\widehat{\Omega}_{ww}$ converge *a.s.* to their positive definite limits. □

Lemma C If, for some $\delta>0$, $E\|Z_i\|^{2+\delta}<\infty$, $E\|v_i\|^{2+\delta}<\infty$, *then*

$$\left(\begin{array}{c} \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right)\frac{\widehat{b}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}} \\ \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right)\frac{\widehat{\Omega}^{-1}\widehat{a}}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}} \end{array}\right)\Bigg|\mathcal{X}_n \xrightarrow{d} N\left(0,I_2\otimes E(Z_iZ_i')\right) \quad \text{a.s.}$$

Proof. The result is a special case of Lemma B. The main difference is that the current result has a less stringent moment condition. The result follows as a direct application of Liapunov's Central Limit Theorem, just as in the proof of Lemma B. □

Proof of Theorem 3. The result is a direct application of the Delta Method and the limiting distribution given in Lemma B (and noting the zero covariances between the three components in the normal limiting distribution).

Notice that $E^*[Z^{*\prime}Z^*/n] = Z'Z/n$. So by the Markov Law of Large Numbers,

$$\left.\frac{Z^{*\prime}Z^*}{n} - \frac{Z'Z}{n}\right|\mathcal{X}_n \overset{a.s.}{\to} 0 \text{ a.s.}$$

Moreover, $Z'Z/n \overset{a.s.}{\to} E(Z_i Z_i{}')$, and so $Z^{*\prime}Z^*/n|\mathcal{X}_n \overset{a.s.}{\to} E(Z_i Z_i{}')$ a.s. Similarly, $Z^{*\prime}V^*/n|\mathcal{X}_n \overset{a.s.}{\to} E^*[Z^{*\prime}V^*/n] = 0$ a.s. So, under (i), $T^{*\prime}T^*/n|\mathcal{X}_n \overset{a.s.}{\to} \pi' E(Z_i Z_i{}')\pi(a'\Omega^{-1}a)$ a.s. By Lemma C,

$$\widehat{b}'(V^{*\prime}Z^*/\sqrt{n})\widehat{\pi}|\mathcal{X}_n \overset{d}{\to} N(0, \pi' E(Z_i Z_i{}')\pi(b'\Omega b)).$$

The result follows under condition (a).

Now, consider the case (ii) in which Assumption 1B (or 1A) holds, and define

$$t_n^* = \sqrt{n}(Z'Z/n)^{1/2}\widehat{\pi}\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}.$$

Therefore, we have

$$T^* - t_n^* = \sqrt{n}\left[\left(\frac{Z^{*\prime}Z^*}{n}\right)^{1/2} - \left(\frac{Z'Z}{n}\right)^{1/2}\right]\widehat{\pi}\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}} + \sqrt{n}\frac{\left(\frac{Z^{*\prime}Z^*}{n}\right)^{-1/2}\frac{Z^{*\prime}V^*}{\sqrt{n}}\widehat{\Omega}^{-1}\widehat{a}}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}}$$

The first term in the sum above is conditionally asymptotically negligible since $\sqrt{n}\widehat{\pi}$, $\widehat{\beta}$ and $\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}$ are bounded in probability. It then follows that $(S^{*\prime}, (T^* - t_n^*)')'|\mathcal{X}_n \overset{d}{\to} N(0, I_{2k})$ a.s. □

Appendix B - Tables

TABLE I

Percent Rejected Under H_0, Nominal 5%

Conditional LR Test

Number of Simulations = 1000, $k = 4$

		Normal Disturbances				Wishart Disturbances			
		$n = 20$		$n = 80$		$n = 20$		$n = 80$	
ρ	$\lambda'\lambda/k$	BS	Crit. Val. Func.	BS	Crit. Val. Func.	BS	Crit. Val. Func.	BS	Crit. Val. Func.
0	0	5.0	10.6	5.3	6.4	7.9	13.8	6.4	7.7
0	1	5.5	9.2	5.1	6.3	7.6	12.3	6.1	7.8
0	10	4.9	6.9	5.4	5.6	6.5	9.7	5.9	6.6
0.5	0	7.2	12.5	5.8	6.8	7.0	12.9	7.8	9.0
0.5	1	6.3	10.2	5.1	5.8	6.4	11.5	6.8	8.5
0.5	10	5.3	7.6	4.6	5.6	5.9	9.8	6.5	7.8
0.75	0	4.5	8.9	5.4	6.3	6.5	12.9	6.3	7.6
0.75	1	4.2	7.2	5.2	6.2	5.2	9.7	5.9	7.3
0.75	10	4.5	6.8	4.8	5.4	4.5	8.1	4.9	6.2
0.99	0	5.9	10.9	5.0	6.2	9.4	15.7	6.5	7.6
0.99	1	3.8	5.9	5.2	6.2	5.7	8.5	5.7	6.6
0.99	10	4.3	6.1	4.9	5.5	5.9	8.1	5.4	6.3

TABLE II - Panel A (Normal Disturbances)
Percent Rejected Under H_0, Nominal 5%
Conditional LR Test
Non-Parametric Bootstrap with Normal Kernel and Bandwidth h
Number of Simulations = 1000, $n = 20$, $k = 4$, $B = 5000$

		h						
ρ	$\lambda'\lambda/k$	τ	$.75\tau$	$.5\tau$	$.25\tau$	$.1\tau$	$.05\tau$	Critical Value Function
0	0	6.5	6.5	6.1	5.9	6.1	5.9	10.6
0	1	5.9	5.8	5.5	5.4	5.5	5.8	9.2
0	10	5.3	5.3	5.3	5.4	5.2	5.2	6.9
0.5	0	5.2	4.8	4.7	4.6	6.0	6.9	12.5
0.5	1	5.7	5.6	5.4	5.2	5.4	5.6	10.2
0.5	10	5.8	5.8	5.8	5.9	6.0	6.0	7.6
0.75	0	6.0	5.3	4.4	4.6	5.1	6.2	8.9
0.75	1	5.1	4.9	4.7	4.8	5.3	5.6	7.2
0.75	10	6.0	6.0	5.9	5.9	5.8	5.6	6.8
0.99	0	6.3	5.6	5.2	5.9	6.6	7.3	10.9
0.99	1	3.0	3.0	2.9	2.8	9.2	12.8	5.9
0.99	10	3.7	3.7	3.7	3.8	12.2	16.6	6.1

TABLE II - Panel B (Wishart Disturbances)
Percent Rejected Under H_0, Nominal 5%
Conditional LR Test
Non-Parametric Bootstrap with Normal Kernel and Bandwidth h
Number of Simulations = 1000, $n = 20$, $k = 4$, $B = 5000$

		h						
ρ	$\lambda'\lambda/k$	τ	$.75\tau$	$.5\tau$	$.25\tau$	$.1\tau$	$.05\tau$	Critical Value Function
0	0	10.2	9.8	9.6	9.8	9.8	9.8	13.8
0	1	9.4	9.2	9.0	8.9	9.3	9.5	12.3
0	10	7.0	7.0	7.2	7.2	7.2	7.1	9.7
0.5	0	8.9	8.4	8.3	8.4	8.9	9.5	12.9
0.5	1	7.5	7.4	7.3	7.2	7.3	7.1	11.5
0.5	10	6.2	6.3	6.4	6.3	6.3	6.3	9.8
0.75	0	9.2	8.8	8.8	8.8	9.5	10.4	12.9
0.75	1	7.2	7.2	7.1	7.5	7.6	7.5	9.7
0.75	10	5.6	5.7	5.7	5.9	5.9	5.9	8.1
0.99	0	10.1	9.6	8.9	9.6	10.2	11.3	15.7
0.99	1	5.3	5.4	5.3	4.6	10.5	14.3	8.5
0.99	10	6.9	7.1	7.1	6.4	13.0	18.1	8.1

TABLE III - Panel A (Normal Disturbances)
Percent Rejected Under H_0, Nominal 5%
Conditional LR Test
Non-Parametric Bootstrap with Normal Kernel and Bandwidth h
Number of Simulations = 1000, $n = 80$, $k = 4$, $B = 5000$

		h						
ρ	$\lambda'\lambda/k$	τ	$.75\tau$	$.5\tau$	$.25\tau$	$.1\tau$	$.05\tau$	Critical Value Function
0	0	8.4	7.8	7.6	7.2	7.3	7.3	6.4
0	1	7.9	7.9	7.5	7.4	7.7	7.7	6.3
0	10	6.4	6.3	6.3	6.3	6.4	5.9	5.6
0.5	0	7.5	7.4	6.9	7.1	7.5	7.9	6.8
0.5	1	6.3	5.9	5.9	5.6	5.9	6.2	5.8
0.5	10	6.2	6.2	6.2	6.4	6.5	6.6	5.6
0.75	0	6.0	5.8	5.5	5.4	6.3	7.0	6.3
0.75	1	4.4	4.4	4.5	4.8	4.7	4.6	6.2
0.75	10	5.0	5.0	5.1	5.2	5.2	5.0	5.4
0.99	0	6.5	6.2	5.8	5.2	5.4	6.6	6.2
0.99	1	3.0	3.0	3.0	3.0	10.6	18.5	6.2
0.99	10	4.1	4.1	4.1	4.1	8.0	17.7	5.5

TABLE III - Panel B (Wishart Disturbances)

Percent Rejected Under H_0, Nominal 5%

Conditional LR Test

Non-Parametric Bootstrap with Normal Kernel and Bandwidth h

Number of Simulations = 1000, $n = 80$, $k = 4$, $B = 5000$

		h						
ρ	$\lambda'\lambda/k$	τ	$.75\tau$	$.5\tau$	$.25\tau$	$.1\tau$	$.05\tau$	Critical Value Function
0	0	7.2	7.2	6.9	6.9	7.2	7.7	7.7
0	1	8.0	7.7	7.2	6.8	6.9	7.0	7.8
0	10	6.3	6.2	6.2	6.5	6.6	6.3	6.6
0.5	0	8.1	7.8	7.5	7.8	7.8	8.3	9.0
0.5	1	7.1	7.0	7.0	7.0	7.1	7.0	8.5
0.5	10	6.2	6.3	6.4	6.4	6.6	6.3	7.8
0.75	0	8.6	8.5	8.2	8.3	8.7	9.2	7.6
0.75	1	5.6	5.6	5.6	5.8	5.7	5.7	7.3
0.75	10	5.9	6.0	6.0	6.4	6.8	6.7	6.2
0.99	0	8.1	8.0	7.9	7.2	7.3	8.4	7.6
0.99	1	4.8	4.9	4.9	4.9	11.9	17.0	6.6
0.99	10	5.7	5.7	5.6	5.3	9.9	16.9	6.3