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: October 10, 2004.

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 the participants at the Boston College-Harvard-MIT, Boston University and UCLA seminars, and at the *Semiparametrics in Rio* conference organized by FGV/EPGE. 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

It is well-known that size-adjustments based on Edgeworth expansions for the t-statistic perform poorly when instruments are weakly correlated with the endogenous explanatory variable. This paper shows, however, that the lack of Edgeworth expansions and bootstrap validity are not tied to the weak instrument framework, but instead depends on which test statistic is examined. In particular, Edgeworth expansions are valid for the score and conditional likelihood ratio approaches, even when the instruments are uncorrelated with the endogenous explanatory variable. Furthermore, there is a belief that the bootstrap method fails when instruments are weak, since it replaces parameters with inconsistent estimators. Contrary to this notion, we provide a theoretical proof that guarantees the validity of the bootstrap for the score test, as well as the validity of the conditional bootstrap for many conditional tests. Monte Carlo simulations show that the bootstrap actually decreases size distortions in both cases.

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

JEL Classification: C12, C31.

1 Introduction

Inference in the linear simultaneous equations model with weak instruments has recently received considerable attention in the econometrics literature. It is now well understood that standard first-order asymptotic theory breaks down when the instruments are weakly correlated with the endogenous regressor; 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, 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, these procedures appear to be unreliable in weak-instrument cases; cf., Hahn, Hausman, and Kuersteiner (2002), and Horowitz (2001), and Rothenberg (1984).

In this paper, we show that the validation of Edgeworth expansions and bootstrap is not tied to the weak-instrument framework generally, but instead depends upon the statistic examined. In particular, our results work for the test of Anderson and Rubin (1949), the score test proposed by Kleibergen (2002) and Moreira (2001), and the conditional likelihood ratio test of Moreira (2003). To our knowledge, these results contain the first formal proofs of the validity of Edgeworth expansions and the bootstrap for cases where some parameters are not identified. At the outset, this exercise appears to face several potential pitfalls. First, the statistics are typically non-regular when the instruments are uncorrelated with the endogenous explanatory variable. Since a general theory of higher-order expansions for non-regular statistics is unavailable, it is *a priori* unclear whether the statistics we examine admit such expansions; see Bhattacharya and Ghosh (1978), Chambers (1967), Phillips (1977), Sargan (1976), and Wallace (1958). Second, in many known non-regular cases the usual bootstrap method fails, *even in the first-order*; cf. Andrews (2000), Horowitz (2001), and Shao (1994). Thus, the non-regularity characterizing the unidentified case poses a potential threat to even first-order validity of the bootstrap. Third, the bootstrap replaces parameters with estimators that are inconsistent in the weak-instrument case. Hence, the empirical distribution function of the residuals may differ considerably from their true cumulative distribution function, which runs counter to the usual argument for bootstrap success.

To show the existence of higher-order expansions, we augment the standard Bhattacharya and Ghosh approach by breaking the proof into two simple steps. We first provide an Edgeworth expansion for certain sufficient statistics, and then we find an approximation to the distribution of the score and the conditional likelihood ratio statistics. As a result, we obtain higher-order expansions for any fixed value of π, including the unidentified case $\pi = 0$. We also propose an expansion approach developed in Cavanagh (1983) and Rothenberg (1984) when errors are normal. Although this method does not provide a formal proof of high-order expansions for the score, it can be used to compute Edgeworth expansions in those cases. The fact the score statistic is non-regular leads to a non-standard result: the higher-order terms are in general not continuous functions of the nuisance parameters at the unidentified case. Thus, the empirical Edgeworth expansion approach of replacing unknown parameters by consistent estimators can perform poorly in the weak-instrument case.

Perhaps more unexpectedly, we show the validity up to the first-order of the bootstrap for the score, and of two conditional bootstrap methods for the conditional likelihood ratio test. These simulation methods, however, do not generally provide higher-order improvements in the unidentified case. Nevertheless, Monte Carlo simulations indicate that the (conditional) bootstrap tends to outperform the first-order asymptotic approximation for the score and conditional likelihood ratio tests. Recently there has been some related work on the bootstrap in weak-instrument settings. Work by Inoue (2002) and Kleibergen (2003) also presents Monte Carlo results suggesting that the usual bootstrap may work when applied to the Anderson-Rubin statistic and score statistics. In the present paper, we 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 justification for the simulation findings of Inoue (2002) and Kleibergen (2003). 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, we present the model and establish some notation. In Section 3, we summarize some folk theorems showing the size improvements based on Edgeworth expansion or the bootstrap for the Wald, score and likelihood ratio tests under the standard asymptotics. In Section 4, we present the main results. We show the validity of Edgeworth expansions for the score and conditional likelihood ratio test statistics when instruments are unrelated to the endogenous explanatory variable. We also establish the validity of the bootstrap for the score test and of two conditional bootstrap methods for the conditional likelihood ratio test up to first order. 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. In Section 6, we conclude and point out some extensions.

2 The Model

We begin by introducing the notation for the instrumental variable specification considered. Throughout the paper, we remark on the extension of the results to other versions of this specification. The structural equation of interest is

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

where y_1 and y_2 are $n \times 1$ vectors of observations on two endogenous variables, u is an $n \times 1$ unobserved disturbance vector, and β is an unknown scalar parameter. This equation is assumed to be part of a larger linear simultaneous equations model, which implies that y_2 is correlated with u.

The complete system contains exogenous variables that can be used as instruments for conducting inference on β. Specifically, it is assumed that the reduced form for $Y = [y_1, y_2]$ can be written as

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

where Z is an $n \times k$ matrix of exogenous variables having full column rank k and π is a $k \times 1$ vector. The n rows of Z are i.i.d., and F is the distribution of each row of Z and $V = [v_1, v_2]$. Unless stated otherwise, we consider the case where Z is independent of V. The n rows of the $n \times 2$ matrix of the reduced-form errors V are i.i.d. with mean zero and 2×2 nonsingular covariance matrix $\Omega = [\omega_{i,j}]$. For ease of exposition in the main body of the paper, we consider statistics designed for the case in which the covariance matrix Ω is assumed to be known. In the proofs in the Appendix we relax this assumption. In what follows, X_n is the n-th observation of some random vector X, and $\overline{X}_n$ is the sample mean of the first n observations of X. The subscript n is typically omitted in what follows, unless it helps exposition. Finally, let $N_A = A\,(A'A)^{-1}\,A'$ and $M_A = I - N_A$ for any conformable matrix A, and let $b_0 = (1, -\beta_0)'$ and $a_0 = (\beta_0, 1)'$.

Tests for the null hypothesis $H_0 : \beta = \beta_0$ play an important role in our results. The commonly used Wald test rejects H_0 for large values of the Wald statistic

$$
W = \frac{\left(\widehat{\beta}_{2SLS} - \beta_0\right)\sqrt{y_2' N_Z y_2}}{\hat{\sigma}_u},
$$

where $\widehat{\beta}_{2SLS} = (y_2' N_Z y_2)^{-1} y_2' N_Z y_1$ and $\hat{\sigma}_u^2 = [1, -\widehat{\beta}_{2SLS}]\Omega[1, -\widehat{\beta}_{2SLS}]'$. The Wald statistic has some important limitations, and it is now well-understood that it may have important size distortions when the instruments may be weak. In particular, under the weak-instrument asymptotics of Staiger and Stock (1997), the limiting distribution of the Wald statistic is not standard normal. Other testing statistics designed for H_0 are based on the Anderson-

Rubin (AR), score (LM), and likelihood ratio (LR) statistics:

$$\begin{aligned} AR &= S'S, \\ LM &= S'T/\sqrt{T'T}, \\ LR &= \frac{1}{2}\left(S'S - T'T + \sqrt{(S'S + T'T)^2 - 4(S'S \cdot T'T - (S'T)^2)}\right), \end{aligned}$$

where $S = (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\Omega b_0)^{-1/2}$ and $T = (Z'Z)^{-1/2}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2}$. The test of Anderson and Rubin (1949) rejects the null if the AR statistic is larger than the $1-\alpha$ quantile of the chi-square-k distribution. The (two-sided) score test proposed by Kleibergen (2002) and Moreira (2001) rejects the null if the LM^2 statistic is larger than the $1-\alpha$ quantile of the chi-square-one distribution. The conditional likelihood ratio test of Moreira (2003) rejects the null if the LR statistic is larger then the $1-\alpha$ conditional quantile of its null distribution conditional on T. All three of these tests are similar if the errors are normal with known variance Ω, since the AR and LM statistics are pivotal and the LR statistic is pivotal conditionally on T.

When the covariance matrix Ω is unknown, we can replace it with the consistent estimator $\tilde{\Omega} = Y'M_Z Y/n$. For example,

$$\begin{aligned} \widetilde{S} &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\tilde{\Omega}b_0)^{-1/2}, \\ \widetilde{T} &= (Z'Z)^{-1/2}Z'Y\tilde{\Omega}^{-1}a_0 \cdot (a_0'\tilde{\Omega}^{-1}a_0)^{-1/2}, \\ \widetilde{LM} &= \widetilde{S}'\widetilde{T}/\sqrt{\widetilde{T}'\widetilde{T}}. \end{aligned}$$

With unknown error distribution, the Anderson-Rubin, score and conditional likelihood ratio tests are no longer similar.[3] However, unlike the Wald test, these three tests are asymptotically similar under both the weak-instrument and standard asymptotics. This important feature allows us to derive the validity of higher-order expansions and the bootstrap regardless of the degree of identification.

[3] An exception occurs with the Anderson-Rubin test, which is similar for normal errors and unknown covariance matrix.

3 Preliminary Results

In this section, we review the good-instrument case for Edgeworth expansions and the bootstrap. Some of the results are already known, and those that are new follow from standard results. The results in this section provide a foundation for the weak-instrument results to be presented in Section 4.

For any symmetric $\ell \times \ell$ matrix A, let $vech(A)$ denote the $\ell(\ell+1)/2$-column vector containing the column by column vectorization of the non-redundant elements of A. The test statistics given in the previous section can be written as functions of

$$\begin{aligned} R_n &= vech\left((Y_n', Z_n')' (Y_n', Z_n')\right) \\ &= \left(f_1 (Y_n', Z_n'), ..., f_\ell (Y_n', Z_n')\right) \end{aligned}$$

for suitably chosen f_i, $i = 1, ..., \ell$, where $\ell = (k+2)(k+3)/2$. In this section, we focus on one-sided tests based on W and LM statistics, which can be written in the form

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

where the gradient of H evaluated at $\mu = E(R_n)$ differs from zero. At the end of this section, we briefly address two-sided tests based on AR and LR statistics. These statistics can be written in the form

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

for suitably chosen functions H whose gradient evaluated at $\mu = E(R_n)$ equals zero, and the Hessian matrix L and variance V of R_n satisfy $LVL' = L$.

Hereinafter, we use the following high-level assumptions:

Assumption 1. π is fixed and different from zero.

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

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

Assumption 1 is related to the standard good-instrument asymptotics. Assumption 2 holds if $E\left\|(Y_n', Z_n')\right\|^{2s} < \infty$. This minimum moment assumption seems too strong at first glance, but note that test statistics involve quadratic functions of (Y_n', Z_n'). Assumption 3 is the commonly used Cramér's condition. The following result by Bhattacharya (1977) provides a sufficient condition for Assumption 3.

Lemma 1 (Bhattacharya (1977)) *Let (Y_n', Z_n') be a random vector with values in $\mathbb{R}^{k+2}$ whose distribution has a nonzero absolutely continuous component G (relative to the Lebesgue measure on $\mathbb{R}^{k+2}$). Assume that there exists an open ball B of $\mathbb{R}^{k+2}$ in which the density of G is positive almost everywhere. If, in B, the functions $1, f_1, ..., f_\ell$ are linearly independent, then Assumption 3 holds.*

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. As a simple application of Theorem 2 of Bhattacharya and Ghosh (1978), we obtain the following result:

Theorem 2 *Under Assumptions 1-3, the null distributions of W_n and LM_n statistics can be uniformly approximated (in x) by Edgeworth expansions:*
(a) $\left\|P\left(LM_n \leq x\right) - \left[\Phi\left(x\right) + \sum_{i=1}^{s-2} n^{-i/2} p_{LM}^i\left(x; F, \beta_0, \pi\right)\phi\left(x\right)\right]\right\|_\infty = o\left(n^{-(s-2)/2}\right),$
(b) $\left\|P\left(W_n \leq x\right) - \left[\Phi\left(x\right) + \sum_{i=1}^{s-2} n^{-i/2} p_{W}^i\left(x; F, \beta_0, \pi\right)\phi\left(x\right)\right]\right\|_\infty = o\left(n^{-(s-2)/2}\right),$
where p_W^i and p_{LM}^i, $i = 1, 2$, are polynomials in x with coefficients depending on moments of R_n, β_0 and π.

We now turn to the bootstrap. For each bootstrap sample, a test statistic is computed, which in turn generates a simulated empirical distribution for the Wald or score statistics. This distribution can then be used to provide new critical values for the test. Importantly, the bootstrap sample is generated based on an estimate of β, and likewise the null hypothesized value of β is replaced by that estimate in forming the bootstrap test statistics. Given consistent estimates $\widehat{\beta}$ and $\widehat{\pi}$, the residuals from the reduced-form equations are obtained as

$$\begin{aligned} \widehat{v}_1 &= y_1 - Z\widehat{\pi}\widehat{\beta} \\ \widehat{v}_2 &= y_2 - Z\widehat{\pi}. \end{aligned}$$

These residuals are re-centered to yield $(\widetilde{v}_1, \widetilde{v}_2)$. Then Z^* and (v_1^*, v_2^*) are drawn independently from the empirical distribution function of Z and $(\widetilde{v}_1, \widetilde{v}_2)$. Next, we set

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

We want to stress here that the simulation method above is exactly equivalent to simulating directly from the structural model

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

where Z^* and (u^*, v_2^*) are drawn independently from the empirical distribution function of Z and $(\widetilde{u}, \widetilde{v}_2)$, where $\widetilde{u} = \widetilde{v}_1 - \widetilde{v}_2\widehat{\beta}$. Also, the probability under the empirical distribution function (conditional on the sample) will be denoted P^* in what follows. Finally, the fact that Z^* is randomly drawn reflects the fact that we are interested in the correlated case. We do not consider the fixed Z case here, although this can be done by establishing conditions similar to those by Navidi (1989) and Qumsiyeh (1990, 1994) in the simple regression model. Of course, this entails different Edgeworth expansions and bootstrap methods.

The following result shows that the bootstrap approximates the empirical Edgeworth expansion up to the $o\left(n^{-(s-2)/2}\right)$ order.

Theorem 3 *Under Assumptions 1-3,*
(a) $\left\| P^* \left(LM_n^* \leq x \right) - [\Phi \left(x \right) + \sum_{i=1}^{s-2} n^{-i/2} p_{LM}^i (x; F_n, \widehat{\beta}, \widehat{\pi}) \phi \left(x \right)] \right\|_\infty = o \left(n^{-(s-2)/2} \right)$,
(b) $\left\| P^* \left(W_n^* \leq x \right) - [\Phi \left(x \right) + \sum_{i=1}^{s-2} n^{-i/2} p_W^i (x; F_n, \widehat{\beta}, \widehat{\pi}) \phi \left(x \right)] \right\|_\infty = o \left(n^{-(s-2)/2} \right)$,
a.s. as $n \to \infty$.

The error based on the bootstrap simulation is of order $n^{-1/2}$ 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 3 shows that the bootstrap offers a better approximation than the standard normal approximation.

Finally, if one is interested in the problem of two-sided hypothesis testing, one could reject H_0 for large values of $|W|$ and $|LM|$. Using the fact that the polynomials $p_W^1 \left(x \right)$ and $p_{LM}^1 \left(x \right)$ are even, one can show that the $n^{-1/2}$-term for the expansion of $|W|$ and $|LM|$ vanishes. Hence, the approximation error based on the bootstrap for two-sided Wald and score tests is of order n^{-1}. For the Anderson-Rubin and likelihood ratio statistics, one could use the results of Chandra and Ghosh (1979) to get (empirical) Edgeworth expansions for their density function of the form

$$\kappa_v \left(x \right) \sum_{r=0}^{m} n^{-r} q_r \left(x \right),$$

where $\kappa_v \left(x \right)$ is the density function of a chi-square-v variable and $q_r \left(x \right)$ are polynomials of x with $q_0 \left(x \right) = 1$. Here, the order of the expansion m is a function of the largest s such that $E \left\| R_n \right\|^s < \infty$.

4 Main Results

In the previous section, we considered the good-instrument case in which the structural parameter β is identified. Our results are threefold: the null distribution of the Wald and score statistics can be approximated by an Edgeworth expansion up to the $n^{-(s-2)/2}$ order, for some integer s; the bootstrap estimate and the $(s-1)$ −term empirical Edgeworth expansion for both statistics are asymptotically equivalent up to the $n^{-(s-2)/2}$ order; and, the error of estimation of the bootstrap is of order $n^{-1/2}$ for one-sided versions and of order n^{-1} for two-sided versions of the Wald and score tests. However, the three results in Section 3 depend crucially on Assumption 1. First, the commonly used (and to our knowledge, the only) proof of the existence of Edgeworth expansions for statistics in the form (3) is given by Bhattacharya and Ghosh (1978), and crucially depends upon the assumption that derivatives of functions evaluated at $\mu = E(R_n)$ are defined and different from zero (regular case). However, if the instruments are uncorrelated with the endogenous variables, the score and Wald statistics do not satisfy this requirement. Hence, in the unidentified case, it is not obvious whether these statistics actually admit second-order expansions, and, if they exist, how to prove their existence. Second, and more importantly, the hypothesized value β_0 is replaced by an inconsistent estimator $\widehat{\beta}$. Consequently, it is not clear whether the bootstrap actually provides valid approximations *even in the first-order*. In fact, similar versions of Theorems 2 and 3 have been considered to fix size distortions of the Wald test in the weak-instrument case. However, when instruments are weak, it is well-known that this method does not lead to substantial size improvements.

In this section, we address the issues above that arise in a weak-instrument setting. We show that the score test admits a standard higher-order expansion, and that the conditional likelihood ratio test admits a higher-order expansion whose leading term is nuisance-parameter-free. In addition, we prove that the bootstrap does provide a valid first-order approximation to the null distribution of the score test, and that conditional bootstrap methods provide a valid first-order approximation to the null distribution of the conditional likelihood ratio test. Finally, we point out that these bootstrap simulations generally do not provide higher-order approximations in the weak-instrument case due to the inconsistency of any estimator of β.

4.1 Edgeworth Expansions

Here we show that the score and conditional likelihood ratio tests admit higher-order expansions for the unidentified case. Formally, we replace Assumption 1 by the following assumption in our main Edgeworth expansion results:

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

To motive parts of the discussion, it will also be useful to have a statement of the weak-instrument asymptotics:

Assumption 1B. (locally unidentified case) $\pi = c/n^{1/2}$ for some nonstochastic k-vector c.

To illustrate why the score test admits an Edgeworth expansion, it is worth considering a stochastic expansion following the work by Nagar (1959). To compute the approximate bias of the $2SLS$ estimator, we can expand its formula into a power series,

$$\widehat{\beta}_{2SLS} = X_n + \frac{P_n}{\sqrt{n}} + \frac{Q_n}{n} + O_p\left(n^{-3/2}\right), \tag{4}$$

where X_n, P_n, and Q_n are sequences of random variables with limiting distributions as n tends to infinity. More specifically, we can arrange the expression for the $2SLS$ estimator such that:

$$\widehat{\beta}_{2SLS} = \beta + \frac{(\pi' Z' u + v_2' N_Z u)/n}{\frac{\pi' Z' Z\pi}{n}[1 + \frac{2}{\sqrt{n}}\frac{(\pi' Z' v_2)/\sqrt{n}}{(\pi' Z' Z\pi)/n} + \frac{1}{n}\frac{v_2' N_Z v_2}{(\pi' Z' Z\pi)/n}]}.$$

For fixed, nonzero π and a large enough sample size, we can do a power series expansion in the denominator to get (4). Taking expectations based on the terms up to the order n^{-1} we obtain:

$$E(\widehat{\beta}_{2SLS}) - \beta = (k-2)\frac{\sigma_{u,v_2}}{\pi' Z' Z\pi} + o(n^{-1}), \tag{5}$$

where σ_{u,v_2} is the covariance between the disturbances u and v_2. The derivation of (5) depends on showing that the terms $\pi' Z v_2$ and $v' Z(Z'Z)^{-1}Z' v$ are asymptotically negligible relative to $\pi' Z' Z \pi$. However, with weakly correlated instruments, $\pi' Z' Z \pi$ is close to zero, so that in finite samples the other terms may be just as important to the bias as $\pi' Z' Z \pi$. Hence, equation (5) may not provide a good approximation to the finite sample bias of the 2SLS estimator when instruments are weak. The Wald statistic presents the same limitation: its stochastic expansion assumes that some terms are asymptotically negligible, an assumption that breaks down with weak instruments. Close inspection, however, shows that a stochastic expansion for the score test is valid for any value of π, including zero. Recall the connection between stochastic expansions and Edgeworth expansions, as conjectured by Wallace (1958) and proved by Bhattacharya and Ghosh (1978) for regular cases. Although this connection has not been proved for non-regular cases, a valid stochastic expansion for any value of π illustrates an important feature of the score statistic.

Following Bhattacharya and Ghosh (1978), the Wald and score statistics can still be written as functions of averages for various moments in 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. In fact, in the locally unidentified case, asymptotic approximations for the Wald statistic based on Edgeworth expansions break down. In this case, the leading term is not the c.d.f. of a standard normal. In fact, its limiting distribution depends on nuisance parameters that are not consistently estimable, as we can see using the results of Staiger and Stock (1997).

Proposition 4 *Under Assumptions 1B and 2,*

$$\widehat{\beta}_{2SLS} \Rightarrow \mathcal{B} = \frac{(\lambda + z_{v_2})' \left(\lambda \beta_0 + (\omega_{11}/\omega_{22})^{1/2} z_{v_1}\right)}{(\lambda + z_{v_2})' (\lambda + z_{v_2})},$$

where

$$\lambda = \omega_{22}^{-1/2} E\left(Z_n Z_n'\right)^{1/2} c, \text{ and } \left(z_{v_1}', z_{v_2}'\right)' \sim N\left(0, \Xi \otimes I_K\right),$$

where Ξ is a 2×2 matrix with diagonal elements being equal to one, and off-diagonal elements equal to $\omega_{12}/\sqrt{\omega_{11}\omega_{22}}$. Moreover, under the null hypothesis,

$$W_n \Rightarrow \frac{\omega_{22}^{1/2}}{\sigma_{\mathcal{B}}} \cdot \frac{[\lambda + z_{v_2}]' \left[(\omega_{11}/\omega_{22})^{1/2} z_{v_1} - z_{v_2}\beta_0 \right]}{\left([\lambda + z_{v_2}]' [\lambda + z_{v_2}] \right)^{1/2}},$$

where $\sigma_{\mathcal{B}}^2 = [1, -\mathcal{B}]\Omega[1, -\mathcal{B}]'$.

Like the Wald statistic, the score statistic is not differentiable at $\pi = 0$, making a Bhattacharya and Ghosh (1978)'s expansion method unavailable. Unlike the Wald procedure, however, the score test does admit a second-order Edgeworth expansion with a normal c.d.f. as the leading term, even in the unidentified case. Since we cannot apply Theorem 2 of Bhattacharya and Ghosh (1978) directly, we break the proof of the existence of an Edgeworth expansion for the score statistic into two simple steps. First, we present the following intermediate result:

Lemma 5 *Under Assumptions 1 or 1A, 2 and 3, the joint null distribution of S and T can be uniformly approximated by an Edgeworth expansion:*

$$\left\| P\left(S_n \leq x_1, T_n \leq t_n\right) - \left[\mathbf{\Phi_A}\left(x\right) + \sum_{i=1}^{s-2} n^{-i/2} \mathbf{q}^i\left(x\right) \boldsymbol{\phi}_A\left(x\right)\right] \right\|_\infty = o\left(n^{-(s-2)/2}\right),$$

where $t_n = \sqrt{n}\Omega_{ZZ}^{1/2}\pi \cdot (a_0'\Omega^{-1}a_0)^{1/2} + x_2$ and and Φ_A denotes the c.d.f. of a mean zero normal distribution with variance A.

An explicit expression for A will be given in section 4.2. 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, note that the end-point $t_n = \sqrt{n}\Omega_{ZZ}^{1/2}\pi \cdot (a_0'\Omega^{-1}a_0)^{1/2} + x_2$ changes with the sample size. This adjustment is due to the fact that the mean of T drifts off to infinity in the case $\pi \neq 0$, 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

$$\Omega_{ZZ}^{1/2} c \cdot (a_0' \Omega^{-1} a_0)^{1/2} + x_2.$$

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, it guarantees that the score test admits an Edgeworth expansion even when $\pi = 0$:

Theorem 6 *Under Assumptions 1A, 2 and 3, the null distribution of LM can be approximated by an Edgeworth expansion:*

$$\left\| P\left(LM_n \leq x\right) - \left[\Phi\left(x\right) + \sum_{i=1}^{s-2} n^{-i/2} q_{LM}^{i}\left(x; F, \beta_0\right) \phi\left(x\right)\right] \right\|_{\infty} = o\left(n^{-(s-2)/2}\right).$$

Note that the leading term in the expansion for the score test is the c.d.f. of a standard normal. Therefore, we extend previous results in the literature, which show that the score test is asymptotically normal even in the unidentified case; c.f., Kleibergen (2002) and Moreira (2001). Theorems 2(b) and 6 show that the null rejection probability of the score test can be approximated by a second-order Edgeworth expansion pointwise in the nuisance parameters π. Unfortunately, the score test does not present very good power properties. In particular, this test is dominated in practice by the conditional likelihood ratio test; cf., Moreira (2003) and Andrews, Moreira, and Stock (2003). Like the Wald test, the null distribution of the likelihood ratio statistic is not nuisance-parameter-free. Hence, we focus here on obtaining an expansion for the conditional null distribution of the likelihood ratio statistic. We follow Barndorff-Nielsen and Cox (1979), and begin by providing expansions for the unconditional probabilities of S and T:

$$P\left(S_n \leq x_1, T_n \leq t_n\right) \text{ and } P\left(T_n \leq t_n\right) \tag{6}$$

where the end-point $t_n = \sqrt{n}\Omega_{ZZ}^{1/2}\pi\cdot(a_0'\Omega^{-1}a_0)^{1/2}+x_2$ changes with the sample size. Of course, under the same conditions as given in Lemma 5, we can also obtain an Edgeworth expansion for the marginal distribution $P(T_n \leq t_n)$.

By obtaining Edgeworth expansions for (6), we can approximate the null conditional distribution of S up to a $o\left(n^{-(s-2)/2}\right)$ term:

$$P(S_n \leq x_1 | T_n = t_n) = [\mathbf{\Phi}(x_1) + \sum_{i=1}^{s} n^{-i/2}\mathbf{p}^i(x_1|x_2)\boldsymbol{\phi}(x_1)].$$

By using this approximation, we can compute Edgeworth expansions for the conditional distribution of a statistic $\psi(S,T)$. In particular, we can obtain an approximation for the conditional distribution of the likelihood ratio statistic for the known Ω case:

$$LR = \frac{1}{2}\left(S'S - T'T + \sqrt{(S'S+T'T)^2 - 4(S'S\cdot T'T - (S'T)^2)}\right).$$

The leading term of the conditional distribution in this expansion is nuisance-parameter-free, but not of a standard normal random variable. Thus, although we may be able to achieve size improvements by considering higher-order terms in this expansion, it may prove difficult to do in practice. In fact, the leading term has not even been computed, and in practice is approximated by simulation methods; cf., Moreira (2003).

Although all stated results are for tests designed for the known covariance matrix case, analogous results hold when we replace Ω with its consistent estimator $\widetilde{\Omega}$. In particular, the $\widetilde{LM}$ and $\widetilde{W}$ statistics also admit Edgeworth expansions, but with different polynomials in the higher-order terms (see Appendix A). Of course, the Edgeworth expansion breaks down for the $\widetilde{W}$ statistic in the unidentified case.

We finish this section presenting an alternative way to find second-order Edgeworth expansions for the $\widetilde{LM}$ statistic when Ω is unknown but the errors are normal. Applying the results in Cavanagh (1983) and Rothenberg (1988), Proposition 7 computes the second-order Edgeworth distribution for $\widetilde{LM}$ based on a stochastic expansion:

$$\widetilde{LM} = LM + n^{-1/2}P_n + n^{-1}Q_n + O_p\left(n^{-3/2}\right),$$

where P and Q are stochastically bounded with conditional moments

$$p_n(x) = E(P_n | LM_n = x), \; q_n(x) = E(Q_n | LM_n = x), \; v_n(x) = V(P_n | LM_n = x).$$

Cavanagh's method re-writes the statistic of interest to include the normalization provided by the denominator in the numerator. What remains in the denominator can then be expanded. This approach allows us to avoid the division by zero problem and the non-differentiability at $\pi = 0$.

Proposition 7 *If the errors are jointly normally distributed, and $\widetilde{LM}$ admits a second-order Edgeworth expansion, $P\left(\widetilde{LM}_n \leq x\right)$ can be approximated by*

$$\Phi\left[x - n^{-1/2} p_n(x) + 0.5 \cdot n^{-1}\left[2p_n(x)\, p_n'(x) - 2q_n(x) + v_n'(x) - x v_n(x)\right]\right]$$

up to a $o(n^{-1})$ term.

Comment: The terms $p_n(x)$, $q_n(x)$, and $v_n(x)$ can be approximated by functions such that the terms in the higher-order expansion are expressed exactly as powers of $n^{-1/2}$; see Rothenberg (1988).

Recall that under normality the LM statistic is $N(0,1)$ under H_0, but the $\widetilde{LM}$ statistic is not. Therefore, Proposition 7 provides a second-order correction for the $\widetilde{LM}$ statistic using conditional moments on the LM statistic. In FGLS examples, Edgeworth expansions are known to correct for skewness and kurtosis due to an estimated error covariance matrix; cf. Horowitz (2001) and Rothenberg (1988). We find that this behavior carries over to the IV setting as well. Finally, unlike Theorems 2(a) and 6, Proposition 7 does not prove the existence of second-order Edgeworth expansions. It only states that if such an expansion exists (as shown in Theorems 2(a) and 6), it can be found by computing some moments conditional on LM, the score statistic for known Ω. In principle, this technique can also be applied to the multivariate case and, consequently, to the conditional tests; see Appendix B.

In practice, we do not know π and Ω, and need replace them with consistent estimators in the high-order terms. As long as the high-order polynomials are continuous functions of the parameters, empirical Edgeworth expansions lead to high-order improvements. However, the continuity of the high order terms cannot be taken for granted in the weak-instrument case due to the possible non-differentiability of statistics at the unidentified case. For example, suppose that $E(v_i|z_i) = 0$, $E(v_i v_i'|z_i) = \Omega$, and $\mu_{zz} = E\left(z_i z_i'\right) < \infty$. Tedious algebraic manipulations show that, for $\pi \neq 0$, the polynomial of the first-order term for the score is given by $[\alpha_2 + (\alpha_1 - \alpha_2)x^2]$, where

$$\alpha_1 = \frac{1}{2}\frac{E[(z'\pi)(v'b)^3]}{(b'\Omega b)^{3/2}(\pi'\mu_{zz}\pi)^{1/2}} \quad \text{and} \quad \alpha_2 = \frac{1}{6}\frac{E[(z'\pi)^3(v'b)^3]}{(b'\Omega b)^{3/2}(\pi'\mu_{zz}\pi)^{3/2}}.$$

This higher order term in general cannot be extended to be continuous at $\pi = 0$. Thus, the empirical Edgeworth expansion approach may not provide a $n^{-1/2}$ correction and can perform poorly at the unidentified case. This finding need not apply to other statistics. For instance, the Anderson-Rubin statistic can be written as a function of sample moments which has higher order derivatives even in the unidentified case. Thus, the Anderson-Rubin statistic has continuous higher order terms, and its empirical Edgeworth expansion would provide higher order corrections even at $\pi = 0$.

4.2 Bootstrap

The usual intuition for the bootstrap requires that the empirical distribution from which the bootstrap sample is drawn is close to the distribution of the data under the null. For the model given in equations (1) and (2), the empirical distribution used in bootstrap sampling depends on the residuals from these equations. When instruments are weak, these residuals depend on inconsistent parameter estimates, so it is not clear *a priori* that the empirical distribution will be close to the distribution of the errors. However, we typically have

$$\widehat{\pi} \xrightarrow{a.s.} \pi \text{ and } \widehat{\pi\beta} \xrightarrow{a.s.} \pi\beta$$

for any fixed value of π, including the important $\pi = 0$ case; see Lemma A in the appendix for an example. Since the reduced-form residuals depend on the parameter estimates only through $\widehat{\pi}$ and $\widehat{\pi}\widehat{\beta}$, this result suggests that the estimated residuals $(\hat{v}_1, \hat{v}_2)$ are close to (v_1, v_2) in the reduced-form model. This is a simple but important insight for the results of this section.

As an additional complication, the null hypothesized value of $\beta = \beta_0$ is replaced by the estimator $\widehat{\beta}$ in the corresponding bootstrap test statistics. Recall that $\widehat{\beta}$ is not a consistent estimator under Assumptions 1A or 1B. Also, as before, we treat the known Ω case here for expositional ease. So, Ω will be replaced by the estimator $\tilde{\Omega}$ based on $(\tilde{v}_1, \tilde{v}_2)$ in the bootstrap test statistics. Therefore, we have:

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

where $\hat{a} = (\widehat{\beta}, 1)'$ and $\hat{b} = (1, -\widehat{\beta})'$. In particular, the bootstrap score test statistic is given by

$$LM^* = \frac{S^{*\prime}T^*}{\sqrt{T^{*\prime}T^*}}.$$

To derive the asymptotic distribution of the bootstrap versions of the score, we must re-center T^* in (bootstrap) analogy with Lemma 5 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

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

To describe this limiting distribution, we require some additional notation, namely Lyapunov'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, the second term has a joint normal limit distribution with S^*, and the bootstrap score has the expected distribution.

Theorem 8 *Suppose that, for some $\delta > 0$, $E\|z_i\|^{2+\delta}, E\|v_i\|^{2+\delta} < \infty$, and let $\widehat{\pi}$ and $\widehat{\beta}$ be estimators such that $\widehat{\pi} \xrightarrow{a.s.} \pi$ and $\widehat{\pi}\widehat{\beta} \xrightarrow{a.s.} \pi\beta$. Under Assumptions 1 or 1A, we have*

$$LM^*|\mathcal{X}_n \xrightarrow{d} N(0,1) \quad a.s. \ ,$$

where $\mathcal{X}_n = \{(Y_1', Z_1'), \ldots, (Y_n', Z_n')\}$.

Theorem 8 yields first-order validity of the bootstrap score test regardless of instrument weakness. The validity of the bootstrap in approximating the asymptotic distribution of the score test in the unidentified case is notable. Unfortunately, the bootstrap in the weak-instrument case does not provide a second-order approximation, because higher-order terms depend on $\widehat{\beta}$ separately from the term $\widehat{\pi}\widehat{\beta}$. In other words, second-order improvements for the score test based on the bootstrap may worsen as π approaches zero. An alternative bootstrap method could be pursued by not replacing β_0 with $\widehat{\beta}$. This avoids the problem of replacing the structural parameter with the inconsistent estimator $\widehat{\beta}$, yet it possibly entails power losses (recall that the e.d.f. of the residuals will not be close to their c.d.f. when the true β is different from the hypothesized value β_0).

Lemma A in the Appendix 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, we note that the proof of Theorem 8 (and 9) also works for the case where $\hat{\pi}$ and $\hat{\pi}\hat{\beta}$ converge in probability. Then 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).

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 (and typically Monte Carlo-simulated) critical value function. The bootstrap provides another way of obtaining a critical value for conditional tests. We provide a parallel result to Lemma 5, which gives the joint limiting behavior of $(S^*, T^* - t_n^*)$.

Theorem 9 *Suppose that, for some $\delta > 0$, $E\|z_i\|^{4+\delta}, E\|v_i\|^{2+\delta} < \infty$, and let $\widehat{\pi}$ and $\widehat{\beta}$ be estimators such that $\widehat{\pi} \xrightarrow{a.s.} \pi$ and $\widehat{\pi}\widehat{\beta} \xrightarrow{a.s.} \pi\beta$. Under Assumptions 1 or 1A,*

$$\left(\begin{array}{c} S^* \\ T^* - t_n^* \end{array} \right) \Bigg| \mathcal{X}_n \xrightarrow{d} N(0, A) \quad a.s. \ ,$$

where

$$A = \left(\begin{array}{cc} I_k & 0 \\ 0 & I_k + \Sigma {a_0}' \Omega^{-1} a_0 \end{array} \right).$$

The joint distribution of S^* and T^* can therefore be used to derive the bootstrapped distribution of the score test in the unidentified case, but it requires stronger moment conditions than Theorem 8. More importantly, Theorem 9 suggests two bootstrap methods for the conditional tests[4]. The first method exploits the (first-order) independence of S^* and $T^* - t_n^*$ from Theorem 9 by fixing T at its observed value and obtaining S^* from bootstrap samples. The second method is 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 first method has a significant, computational efficiency advantage over the non-parametric proposed by Booth, Hall, and Wood (1992), but its ability to provide asymptotic refinements depends on higher-order independence of S^* and T^*. Consequently, it may entail more size distortions than the non-parametric method, at least in the good-instrument case. On the other hand, the second method depends crucially on bandwidth choice,

[4]Unfortunately, the conditional bootstrap methods do not work for the conditional Wald testing procedure, since the Wald statistic depends on β not only through S and T.

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-dimension 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, the LR statistic can be rewritten as

$$(7) \qquad LR = \frac{1}{2}\left[Q_1 + Q_{k-1} - T'T + \sqrt{(Q_1 + Q_{k-1} + T'T)^2 - 4Q_{k-1}T'T}\right],$$

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 the next section, each of these bootstrap procedures is implemented and compared in a Monte Carlo exercise.

5 Monte Carlo Simulations

Theorem 3 suggests that the bootstrap can decrease size distortions for the score and Wald tests when instruments are good. More importantly, Theorems 8 and 9 provide a theoretical support for bootstrapping the score test and the conditional likelihood ratio test, even when instruments are weak. The same validation of the bootstrap does not hold for the Wald test. This crucial difference has implications for the ability of the bootstrap to improve inference for each of these tests. In this section, we present Monte Carlo simulations that support our theoretical results. We first compare the performance of the bootstrap for the score test and the Wald test. Then we provide results of simulations for the two conditional bootstrap methods that are applied to the conditional likelihood ratio test.

Following 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 $[u, v_2]$. To examine the performance of the bootstrap under various degrees of identification, we consider three different values of the population first-stage F-statistic, $\pi'(nI_k)\pi/k$. The first-stage F-statistic corresponds to the concentration parameter $\lambda'\lambda/k$ in the notation of Staiger and Stock (1997). In particular, we consider the completely unidentified case ($\lambda'\lambda/k = 0$), the weak-instrument case ($\lambda'\lambda/k = 1$), and the good-instrument case ($\lambda'\lambda/k = 10$). 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 $\sqrt{\rho}$. In these simulations, we are considering two-sided versions of the score and Wald tests. For each specification, 1000 pseudo-data sets are generated under the null hypothesis ($\beta = 0$). For each pseudo-data set, we consider the score and Wald statistics using chi-square-one and bootstrap critical values at 5% significance level.

Table I and II report null rejection probabilities for the score and Wald tests when the sample size equals 20 and 80, respectively. The bootstrap yields null rejection probabilities for the score test fairly close to the nominal 5% level. Perhaps more important, bootstrapping the score test instead of using the first-order asymptotic approximation always takes actual rejection rates closer to the nominal size, including the case $\lambda'\lambda/k = 0$.[5] By contrast, bootstrapping the Wald test offers improvements over first-order asymptotics only when instruments are good. In fact, when $\lambda'\lambda/k = 0$ and ρ are small, the bootstrap can be even worse than first-order asymptotics. The poor behavior of the bootstrap for the Wald test is explained by its dependence on π. For small values of π, the null distribution of the Wald statistic is quite sensitive to π in the weak-instrument case. Consequently, the bootstrap is likely to give very different answers depending on the initial estimation of this parameter. The sensitivity is considerably reduced for large values of π. On the other hand, the null asymptotic distribution of the score does not depend on π asymptotically. Hence, the bootstrap procedure exhibits little sensitivity to its initial estimate of π.

In the following set of results, we compare the sizes of the conditional likelihood ratio 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 again designs I and II of Staiger and Stock (1997). Table III shows rejection rates computed using the fixed-T conditional bootstrap. The rejection probabilities using bootstrap critical values are considerably smaller than the ones using the critical value function used in 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, sample sizes ($n = 20$ or 80), and error distributions (normal or Wishart).

[5] We have also done some simulations using the empirical Edgeworth expansion for the one-sided score test. Results not reported here indicate that this approximation method is outperformed by the bootstrap.

The non-parametric conditional bootstrap method can in principle work even better than the fixed-T conditional bootstrap. Recall that the non-parametric bootstrap offers second-order improvements at least in the good-instrument case. Tables IV and V 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 h, although an intermediate value of the bandwidth parameter tends to outperform extreme choices. Finally, we considered 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, score and likelihood ratio statistics admit higher-order Edgeworth expansions *under some regularity conditions*. Replacing the unknown parameters by consistent estimators and using the continuity of the polynomials in the high-order terms guarantee that empirical Edgeworth expansions leads 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, those regularity conditions break down. The consequences of this break down are threefold. First, the Wald statistic no longer admits a standard high-order Edgeworth Expansion. Second, the Wald statistic is a 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.

Like the Wald statistic, the score statistic is non-regular in the unidentified model. The standard Bhattacharya and Ghosh argument breaks down, so that there is no guarantee that the score statistic admits a high-order Edgeworth expansion for each fixed value of π including zero. To show its existence, we write the score statistic as a function of two statistics that admit Edgeworth expansions, but do not approximate this function by the Generalized Delta method. Unlike standard situations, the high-order terms for this expansion are not necessarily continuous when the correlation between the instruments and the explanatory variable is zero. Consequently, the empirical Edgeworth expansion approach does not necessarily lead to high-order improvements from the standard first-order asymptotic theory.

Our second striking result is the validity of the bootstrap. Given previous warnings in the literature concerning the bad performance of the bootstrap in approximating the null distribution of the Wald statistic, there has been a perception that the bootstrap as a general simulation method fails in the unidentified model. This argument seems justified since the bootstrap replaces unknown parameters with estimators that are not consistent and the statistics are non-regular in the unidentified case. Nevertheless, we show theoretically that the bootstrap actually provides a correct approximation for the score statistic up to first-order. Although other methods, such as the m-out-of-n bootstrap (or, of course, using the chi-square-one critical value), also provide first-order asymptotic approximations for the score statistic, the usual bootstrap method has the advantage of providing a higher-order approximation in the good instrument case. We also consider two conditional bootstrap methods to approximate the conditional null distribution of the conditional likelihood ratio statistic. The first conditional bootstrap fixes the value of the statistic we are conditioning on, and bootstraps the remaining statistic(s). The second conditional bootstrap is based on a non-parametric estimation of a conditional probability, and is proposed by Booth, Hall, and Wood (1992).

To assess the performance of the (conditional) bootstrap, we provide some Monte Carlo simulations for the score and conditional likelihood ratio statistics. Even without a guarantee that the standard bootstrap and the two conditional bootstrap methods provide improvements in the unidentified model, our simulations show that they outperform the previous methods based on first-order (weak-instrument) asymptotics. This raises the question as to why the bootstrap performs so remarkably well, but is beyond the scope of this paper. In fact, there is a lack of general theoretical justifications of why the bootstrap outperforms second-order empirical Edgeworth expansions even in the standard regular cases.

Finally, our results for the unidentified case can in principle be extended to the GEL and GMM contexts; cf. Guggenberger and Smith (2003) and Stock and Wright (2000). Inoue (2002) and Kleibergen (2003) present some simulations which indicate that the bootstrap can lead to size improvements for the unidentified case also in the GMM context. However, there is a lack of formal theoretical results that show the validity of the bootstrap and Edgeworth expansions in the (locally) unidentified case. Our theoretical results can then be adapted to those cases by analyzing GMM and GEL versions of the two sufficient statistics for the simple simultaneous equations model analyzed here.

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. (2000): "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space," *Econometrica*, 68, 399–405.

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

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

Barndorff-Nielsen, O., and D. R. Cox (1979): "Edgeworth and Saddle-Point Approximations with Statistical Applications," *Journal of the Royal Statistical Society, Series B (Methodological)*, 41, 279–312.

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.

Cavanagh, C. (1983): "Hypothesis Testing in Models with Discrete Dependent Variables," Ph.D. thesis, UC Berkeley.

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.

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.

NAGAR, A. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of Parameters in Simultaneous Equations," *Econometrica*, 27, 575–595.

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.

REDNER, R. (1981): "Note on the Consistency of the Maximum Likelihood Estimate for Nonidentifiable Distributions," *The Annals of Statistics*, 9, 225–228.

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.

——— (1988): "Approximate Power Functions for Some Robust Tests of Regression Coefficients," *Econometrica*, 56, 997–1019.

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

Shao, J. (1994): "Bootstrap Sample Size in Nonregular Cases," *Proceedings of the American Mathematical Society*, 122, 1251–1262.

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.

Wald, A. (1949): "Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics*, 20, 595–601.

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 2. First, we prove part (a). Under H_0,

$$LM = \sqrt{n}\frac{b_0'\left(Y'Z/n\right)\left(Z'Z/n\right)^{-1}\left(Z'Y/n\right)\Omega^{-1}a_0/\sqrt{b_0'\Omega b_0}}{\sqrt{a_0'\Omega^{-1}\left(Y'Z/n\right)\left(Z'Z/n\right)^{-1}\left(Z'Y/n\right)\Omega^{-1}a_0}}$$

can be re-written as

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

where H is a real-valued Borel measurable function on $R^{(k+2)(k+3)/2}$ such that $H\left(\mu\right) = 0$. All the derivatives of H of order s and less are continuous in the neighborhood of μ. Using Assumptions 2 and 3, the result follows Theorem 2 of Bhattacharya and Ghosh (1978). For the unknown Ω case, note that

$$\widetilde{\Omega} = Y'Y/n - \left(Y'Z/n\right)\left(Z'Z/n\right)^{-1}\left(Z'Y/n\right).$$

Hence, $\widetilde{LM}$ statistic can also be written as

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

under H_0 for a real-valued Borel measurable function H such that $H\left(\mu\right) = 0$. Therefore, by Theorem 2 of Bhattacharya and Ghosh (1978), $\widetilde{LM}$ admits an Edgeworth expansion up to the second term.

The proof for part (b) is analogous to the proof for part (a). The Wald statistic equals

$$W = \sqrt{n}\frac{(y_2'Z/n\left(Z'Z/n\right)^{-1}Z'y_2/n)^{-1/2}y_2'Z/n\left(Z'Z/n\right)^{-1}Z'\left(y_1 - y_2\beta_0\right)/n}{\sqrt{[1, -\widehat{\beta}_{2SLS}]\Omega[1, -\widehat{\beta}_{2SLS}]'}},$$

where

$$\widehat{\beta}_{2SLS} = (y_2'Z/n\left(Z'Z/n\right)^{-1}Z'y_2/n)^{-1}y_2'Z/n\left(Z'Z/n\right)^{-1}Z'y_1/n.$$

Like the score statistic, the Wald statistic can be written as

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

under H_0, where H is a real-valued Borel measurable function such that $H\left(\mu\right) = 0$. All the derivatives of H of order s and less are continuous in the neighborhood of μ. The result then follows by Theorem 2 of Bhattacharya and Ghosh (1978). The Wald statistic for unknown variance, $\widetilde{W}$, also admits an Edgeworth expansion by proceeding as it was done for the $\widetilde{LM}$ statistic. □

Proof of Theorem 3. Let F be the distribution of

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

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 resampling mechanism for $\widetilde{Y}_n$ and Z_n and the recentering procedure for $\widehat{V}$ of subtracting samples means reflect the fact that Z and V are independent. If Z and V were uncorrelated, it would entail different drawing mechanisms and recentering 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. Following Lemma 2 of Babu and Singh (1984), there exists for each $d > 0$ positive numbers ϵ and δ such that

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

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

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

is $o(n^{-1})$ a.s. as $n \to \infty$ for every class $\mathcal{A}$ of Borel subsets of $\mathbb{R}^\ell$ satisfying, for some $\theta > 0$,

$$\sup_{A \in \mathcal{A}} \Phi_V\left((\partial A)^\varepsilon \right) = O\left(\varepsilon^\theta \right) \text{ as } \varepsilon \downarrow 0.$$

Reduction of the expansion of $n^{1/2} \left(\overline{\widetilde{R}}_n^* - \overline{\widetilde{R}}_n \right)$ to LM^* follows as in Bhattacharya and Ghosh (1978) once we realize that

$$LM^* = \sqrt{n} \left(H\left(\overline{\widetilde{R}}_n^* \right) - H\left(\overline{\widetilde{R}}_n \right) \right)$$

with $H\left(\overline{R}_n\right) = 0$ (due to recentered residuals).

□

Proof of Proposition 4. The Wald statistic for known covariance matrix Ω can be re-written as

$$W = \frac{(y_2' N_Z y_2)^{-1/2} y_2' N_Z (y_1 - y_2 \beta_0)}{\hat{\sigma}}.$$

Using the fact that $\pi = c/\sqrt{n}$, we have

$$\begin{aligned}
(Z'Z)^{-1/2} Z'y_2 &= (Z'Z/n)^{1/2} c + (Z'Z/n)^{-1/2} Z'v_2/\sqrt{n} \\
&\Rightarrow \omega_{22}^{1/2} [\lambda + z_{v_2}], \text{ and} \\
(Z'Z)^{-1/2} Z'y_1 &= (Z'Z/n)^{1/2} c\beta_0 + (Z'Z/n)^{-1/2} Z'v_1/\sqrt{n} \\
&\Rightarrow \omega_{22}^{1/2} \left[\lambda\beta_0 + (\omega_{11}/\omega_{22})^{1/2} z_{v_1}\right].
\end{aligned}$$

The LLN and CLT holds here since $E\|(Y_n, Z_n)\|^s < \infty$ for some $s \geq 3$.

Thus, under the null hypothesis,

$$\widehat{\beta}_{2SLS} \Rightarrow \mathcal{B} = \frac{(\lambda + z_{v_2})' (\lambda\beta_0 + (\omega_{11}/\omega_{22})^{1/2} z_{v_1})}{(\lambda + z_{v_2})' (\lambda + z_{v_2})}.$$

and, consequently,

$$W \Rightarrow \frac{\omega_{22}^{1/2}}{\sigma_{\mathcal{B}}} \cdot \frac{[\lambda + z_{v_2}]' \left[(\omega_{11}/\omega_{22})^{1/2} z_{v_1} - z_{v_2}\beta_0\right]}{\left([\lambda + z_{v_2}]' [\lambda + z_{v_2}]\right)^{1/2}},$$

Finally, note that the Wald statistic for unknown Ω, $\widetilde{W}$, has the same asymptotic distribution as the W statistic, since $\widetilde{\Omega}$ converges in probability to Ω.

□

Proof of Lemma 5. The statistics S and $T - t_n$ can be written as

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

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

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

where $\phi_M(s,t)$ is the normal density on R^{2k} with mean zero and dispersion M, $P_i(-D:F)$ is a polynomial in p variables whose coefficients do not depend on n, and $-D = (-D_1, ..., -D_{2k})$. 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.

□

Proof of Theorem 6. Note that

$$\sup_{B\in\mathcal{B}} \left| P\left((S_n', T_n')' \in B\right) - \int\int_B \psi_n(s,t)\, ds dt \right| = o\left(n^{-1}\right)$$

holds uniformly over every class $\mathcal{B}$ satisfying

$$\sup_{B\in\mathcal{B}} \int_{(\partial B)^\epsilon} \phi(x)\, dx = O(\varepsilon) \text{ as } \varepsilon \downarrow 0.$$

In particular this holds for the class $\{B_x; x \in R\}$, where

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

□

In Lemma A, we show that we typically have $\left(\widehat{\pi}, \widehat{\pi}\widehat{\beta}\right)$ converging almost surely to the zero vector $\mathbf{0}_{2k}$ when $\pi = 0$. In particular, this result holds for the maximum likelihood estimator $\widehat{\theta}_{MLE} = (\widehat{\pi}_{MLE}, \widehat{\beta}_{MLE})$. This lemma assumes some conditions that are satisfied if their equivalent conditions hold in the reduced-form model that ignores the constraint in the parameters. Almost sure convergence of $\widehat{\pi}_{MLE}$ and $\widehat{\pi}_{MLE}\widehat{\beta}_{MLE}$ to π and $\pi\beta$ trivially holds for any fixed value $\pi \neq 0$ under the regularity conditions in Wald (1949), and is not shown here.

Lemma A *Let* $\mathcal{L} = \{(0,\beta) \in \Pi \times \mathbb{B}\}$ *be the set of unidentified points; that is,* $f(X;\theta)$ *is the same for any* $\theta = (\pi,\beta) \in \mathcal{L}$*. Let* W *be any closed subset of the parameter space* $\Theta = \Pi \times B$ *which does not intersect* $\mathcal{L}$*. Let*

$$\begin{aligned} f(X;\theta,\rho) &= \sup_{\left|\widetilde{\theta}-\theta\right| \leq \rho} f\left(x,\widetilde{\theta}\right), \\ \varphi(X;r_0) &= \sup_{\widetilde{\theta}\in W; \left|\widetilde{\theta}\right| > r_0} f\left(x,\widetilde{\theta}\right) \end{aligned}$$

for a density function $f(x,\theta)$ *that is absolutely continuous with respect to the Lebesgue measure or counting measure. Suppose that the following holds:*

$$\begin{aligned} &\textit{i)}\ E_{\theta_0} \ln f(X;\theta) < E_{\theta_0} \ln f(X;\theta_0) \text{ for any } \theta_0 \in \mathcal{L}, \theta \notin \mathcal{L}, \\ &\textit{ii)}\ \lim_{\rho\to 0} E_{\theta_0} \ln f(X;\theta,\rho) = E_{\theta_0} \ln f(X;\theta) \text{ for any } \theta_0 \in \mathcal{L}, \theta \in \Theta, \\ &\textit{iii)}\ E_{\theta_0} \ln \varphi(X;r_0) < E_{\theta_0} \ln f(X;\theta_0) \text{ for any } \theta_0 \subset \mathcal{L}, \text{some } r_0 \in \mathbb{R}^+. \end{aligned}$$

Finally, let $\widehat{\theta}_n(x_1,...,x_n) = \left(\widehat{\pi},\widehat{\beta}\right)$ *be a function of the observations such that for all* $\theta_0 \in \mathcal{L}$

$$\text{(8)} \qquad \frac{\prod_{\alpha=1}^{n} f\left(x_\alpha;\widehat{\theta}_n\right)}{\prod_{\alpha=1}^{n} f(x_\alpha;\theta_0)} \geq \gamma > 0 \text{ for all } n \text{ and } x_1,...,x_n.$$

Then:

$$\begin{aligned} P_{\theta_0}\left(\lim_{n\to\infty} \sup_{\theta\in W} \frac{\prod_{\alpha=1}^{n} f(X_\alpha;\theta)}{\prod_{\alpha=1}^{n} f(X_\alpha;\theta_0)} = 0\right) &= 1, \\ P_{\theta_0}\left(\lim_{n\to\infty}\left(\widehat{\pi},\widehat{\pi}\widehat{\beta}\right) = \mathbf{0}_{2k}\right) &= 1. \end{aligned}$$

Proof. This proof is essentially a proof by Redner (1981), which augments Theorems 1 and 2 of Wald (1949). Let W_1 be the subset of W consisting of all points $\theta \in W$ for which $|\theta| \leq r_0$. Conditions (i) and (ii) guarantee that, for each point $\theta \in W_1$, there exists a positive value ρ_θ such that

$$E_{\theta_0} \ln f(X;\theta,\rho_\theta) < E_{\theta_0} \ln f(X;\theta_0) \text{ for any } \theta_0 \in \mathcal{L}.$$

Since the set W_1 is compact, there exists a finite number of points $\theta_1, ..., \theta_h$ such that the balls centered at θ_i and with radius $\rho_{\theta i}$, $B(\theta_i, \rho_{\theta i})$, $i = 1, ..., h$, cover W_1. Now,

$$0 \leq \sup_{\theta \in W} \prod_{\alpha=1}^{n} f(x_\alpha; \theta) \leq \prod_{\alpha=1}^{n} \varphi(x_\alpha; r_0) + \sum_{i=1}^{h} \prod_{\alpha=1}^{n} f(x_\alpha; \theta_i, \rho_{\theta i}).$$

Therefore, the first part of Lemma A is proved if the following holds:

$$\begin{aligned} P_{\theta_0}\left(\lim_{n\to\infty} \frac{\prod_{\alpha=1}^{n} f(X_\alpha; \theta_i, \rho_{\theta i})}{\prod_{\alpha=1}^{n} f(X_\alpha; \theta_0)} = 0\right) &= 1, \; i = 1, ..., h. \\ P_{\theta_0}\left(\lim_{n\to\infty} \frac{\prod_{\alpha=1}^{n} \varphi(X_\alpha; r_0)}{\prod_{\alpha=1}^{n} f(X_\alpha; \theta_0)} = 0\right) &= 1. \end{aligned}$$

This can be shown by taking logarithms and using the Strong Law of Large Numbers.

For the second part of Lemma A, it suffices to show that, for any $\epsilon > 0$, the probability is one that all limit points $\overline{\theta} = \left(\overline{\pi}, \overline{\beta}\right)$ of the sequence $\{\theta_n\}$ satisfy the inequality $\left|\left(\overline{\pi}, \overline{\pi}\overline{\beta}\right)\right| \leq \epsilon$. The event that there exists a limit point $\overline{\theta}$ such that $\left|\left(\overline{\pi}, \overline{\pi}\overline{\beta}\right)\right| > \epsilon$ implies that

$$\sup_{\theta; |(\pi, \pi\beta)| \geq \epsilon} \prod_{\alpha=1}^{n} f(x_\alpha; \theta) \geq \prod_{\alpha=1}^{n} f(x_\alpha; \theta_n)$$

for infinitely many n. But then

$$\sup_{|(\pi, \pi\beta)| \geq \epsilon} \frac{\prod_{\alpha=1}^{n} f(x_\alpha; \theta)}{\prod_{\alpha=1}^{n} f(x_\alpha; \theta_0)} \geq \gamma > 0$$

for infinitely many n. However, by the first part of this lemma, this event has probability zero.

□

Comments: The maximum likelihood estimator $\widehat{\theta}_{MLE} = (\widehat{\pi}_{MLE}, \widehat{\beta}_{MLE})$, if it exists, satisfies (8) with $\gamma = 1$.

Note also that this lemma does not assume compactness, but if $\mathbb{B}$ is compact, then trivially $\widehat{\beta}_{MLE} \xrightarrow{a.s.} 0$ for $\pi = 0$.

The following lemma holds regardless of the weakness of the instruments.

Lemma B *Suppose* $\widehat{\pi} \xrightarrow{a.s.} \pi$ *and* $\widehat{\pi}\widehat{\beta}_{2sls} \xrightarrow{a.s.} \pi\beta$*. If, for some* $\delta > 0$, $E\|z_i\|^{2+\delta} < \infty$, $E\|v_i\|^{2+\delta} < \infty$ *, then*

$$\left(\begin{array}{c} \frac{(Z^*)'V_1^*}{\sqrt{n}} \\ \frac{(Z^*)'V_2^*}{\sqrt{n}} \end{array} \right) \Bigg| \mathcal{X}_n \xrightarrow{d} N(0, \Omega \times E(z_i z_i')) \quad \text{a.s.}$$

Proof. Using the Cramér-Wald device, let $c = (c_1', c_2')'$ be a nonzero vector. Let

$$X_{n,i} = \frac{c'}{\sqrt{n}} \left(\begin{array}{c} z_i^* v_{1i}^* \\ z_i^* v_{2i}^* \end{array} \right).$$

To prove the result, we just need to verify the conditions of the Lyapunov Central Limit Theorem:
(i) $E^* X_{n,i} = 0$.
(ii) $E^* X_{n,i}^2$ finite.
(iii) $\lim_{n \longrightarrow \infty} \sum_{i=1}^n \frac{1}{s_n^{2+\delta}} E^*[|X_{n,i}|^{2+\delta}] = 0$, where $s_n^2 = \sum_{i=1}^n E^* X_{n,i}^2$.

(i) First note

$$E^* v_{1i}^* = \frac{1}{n} \sum_{j=1}^n \left(\hat{v}_{1j} - \frac{1}{n} \sum_{l=1}^n \hat{v}_{1l} \right) = 0,$$

and similarly $E^* v_{2i}^* = 0$.

$$\begin{aligned} E^* X_{n,i} &= \frac{c_1'}{\sqrt{n}} E^* z_i^* v_{1i}^* + \frac{c_2'}{\sqrt{n}} E^* z_i^* v_{2i}^* \\ &= \frac{c_1'}{\sqrt{n}} E^* z_i^* E^* v_{1i}^* + \frac{c_2'}{\sqrt{n}} E^* z_i^* E^* v_{2i}^* \\ &= 0. \end{aligned}$$

(ii) Note that

$$E^* X_{n,i}^2 = \frac{c'}{n} E^* \left[\left(\begin{array}{cc} {v_{1i}^*}^2 & v_{1i}^* v_{2i}^* \\ v_{1i}^* v_{2i}^* & {v_{2i}^*}^2 \end{array} \right) \otimes z_i^* {z_i^*}' \right] c = \frac{c'}{n} \left[\tilde{\Omega} \otimes \left(\frac{Z'Z}{n} \right) \right] c$$

is finite a.s. Now,

$$s_n^2 = c' \left[\tilde{\Omega} \otimes \left(\frac{Z'Z}{n} \right) \right] c.$$

(iii) Note that $\frac{1}{n}\sum_i |\tilde{v}_{1i}|^{2+\delta}$ is bounded a.s.:

$$\begin{aligned}
\frac{1}{n}\sum_i |\tilde{v}_{1i}|^{2+\delta} &= \frac{1}{n}\sum_i \Big| v_{1i} - z_i'(\widehat{\pi}\widehat{\beta} - \pi\beta) - \frac{1}{n}\sum_j \Big(v_{1j} - z_j'(\widehat{\pi}\widehat{\beta} - \pi\beta) \Big) \Big|^{2+\delta} \\
&\leq C_1 \left(\frac{1}{n}\sum_i |v_{1i}|^{2+\delta} + |z_i'(\widehat{\pi}\widehat{\beta} - \pi\beta)|^{2+\delta} + \left| \frac{1}{n}\sum_j \Big(v_{1j} - z_j'(\widehat{\pi}\widehat{\beta} - \pi\beta) \Big) \right|^{2+\delta} \right) \\
&\leq C_2 \left(\frac{1}{n}\sum_i |v_{1i}|^{2+\delta} + \sum_{j=1}^{k} \frac{1}{n}\sum_i |z_{ji}|^{2+\delta} |(\widehat{\pi}\widehat{\beta} - \pi\beta)_j|^{2+\delta} \right) \\
&\xrightarrow{a.s.} C_2 E|v_{1i}|^{2+\delta}.
\end{aligned}$$

for large enough constants C_1 and C_2. The first inequality follows from Minkowski's inequality, and the second inequality follows from the following reasoning

$$\Big| \frac{1}{n}\sum_j \Big(v_{1j} - z_j'(\widehat{\pi}\widehat{\beta} - \pi\beta) \Big) \Big|^{2+\delta} \leq \frac{1}{n^{1+\delta}} \frac{1}{n} \sum_j |v_{1j} - z_j'(\widehat{\pi}\widehat{\beta} - \pi\beta)|^{2+\delta}.$$

Therefore, we have:

$$\begin{aligned}
&\sum_i \frac{1}{s_n^{2+\delta}} E^* |X_{ni}|^{2+\delta} \\
&= \left[c' \left(\tilde{\Omega} \otimes \frac{Z'Z}{n} \right) c \right]^{-\left(1+\frac{\delta}{2}\right)} \sum_i n^{-\left(1+\frac{\delta}{2}\right)} E^* \left| c' \begin{pmatrix} z_i^* v_{1i}^* \\ z_i^* v_{2i}^* \end{pmatrix} \right|^{2+\delta} \\
&\leq C_3 \left[c' \left(\tilde{\Omega} \otimes \frac{Z'Z}{n} \right) c \right]^{-\left(1+\frac{\delta}{2}\right)} n^{-\frac{\delta}{2}} \sum_{j=1}^{k} \sum_{m=1}^{2} \left(|c_{1j}|^{2+\delta} \frac{1}{n}\sum_i |z_{ji}|^{2+\delta} \frac{1}{n}\sum_l |\tilde{v}_{ml}|^{2+\delta} \right) \\
&\xrightarrow{a.s.} 0,
\end{aligned}$$

for a large enough constant C_3.

Let $w_i = vech\,(z_i z_i{}')$ and $W = (w_1, \ldots, w_n)'$. Similarly, let $w_i^* = vech\,(z_i^* z_i^{*\prime})$ and $W^* = (w_1^*, \ldots, w_n^*)'$. Also let $\Omega_{WW} = \text{Var}(W_i)$ and let $\mathbf{1}$ be an $n \times 1$ vector of ones.

Lemma C *Suppose that* $\widehat{\pi} \xrightarrow{a.s.} \pi$, $\widehat{\pi}\widehat{\beta} \xrightarrow{a.s.} \pi\beta$, for some fixed π. *If, for some* $\delta > 0$, $E\|z_i\|^{4+\delta} < \infty$, $E\|v_i\|^{2+\delta} < \infty$, *then*

$$\left(\begin{array}{c} \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right) \frac{\hat{b}}{\sqrt{\hat{b}'\tilde{\Omega}\hat{b}}} \\ \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right) \frac{\tilde{\Omega}^{-1}\hat{a}}{\sqrt{\hat{a}'\tilde{\Omega}^{-1}\hat{a}}} \\ \sqrt{n}\left(\frac{W^{*\prime}\mathbf{1}}{n} - \frac{W'\mathbf{1}}{n}\right) \end{array} \right) \Bigg| \mathcal{X}_n \xrightarrow{d} N\left(0, \left(\begin{array}{cc} I_2 \otimes E(z_i z_i') & 0 \\ 0 & \Omega_{WW} \end{array} \right)\right) \quad \text{a.s.}$$

Proof. Using the Cramér-Wald device, let $d = (d_1', d_2', d_3')'$ be a nonzero vector (and $d_{12} = (d_1', d_2')'$), and let

$$X_{n,i} = \frac{d'}{\sqrt{n}} \left[\begin{array}{cc} \left(\begin{array}{c} \frac{\hat{b}'}{\sqrt{\hat{b}'\tilde{\Omega}\hat{b}}} \\ \frac{\hat{a}\tilde{\Omega}^{-1}}{\sqrt{\hat{a}'\tilde{\Omega}^{-1}\hat{a}}} \end{array} \right) \otimes I_k & 0 \\ 0 & I_{k(k+1)/2} \end{array} \right] \left(\begin{array}{c} z_i^* v_{1i}^* \\ z_i^* v_{2i}^* \\ w_i^* - \frac{1}{n}\sum_{j=1} w_j \end{array} \right).$$

The proof follows if we verify the conditions of the Lyapunov Central Limit Theorem. Similar to the proof of Lemma B, $E^* X_{n,i} = 0$. Using the fact that $\hat{b}'\hat{a} = 0$, we have:

$$E^* X_{n,i}^2 = \frac{d_{12}'}{n} \left[I_2 \otimes \left(\frac{Z'Z}{n} \right) \right] d_{12} + d_3'\hat{\Omega} d_3$$

is finite a.s., where $\hat{\Omega} = n^{-1}\sum_i (w_i - n^{-1}\sum_j w_j)^2$. Now, notice that:

$$\begin{aligned}
&\left[d_{12}'\left(I_2 \otimes \frac{Z'Z}{n}\right)d_{12} + d_3'\hat{\Omega}d_3\right]^{-\left(1+\frac{\delta}{2}\right)} \sum_i E^*|X_{ni}|^{2+\delta} \\
&\leq C\left[d_{12}'\left(I_2 \otimes \frac{Z'Z}{n}\right)d_{12} + d_3'\hat{\Omega}d_3\right]^{-\left(1+\frac{\delta}{2}\right)} n^{-\frac{\delta}{2}} \\
&\quad\cdot\Bigg\{\sum_{j=1}^{k}\Bigg[\left(\left|\frac{d_{1j}(-\widehat{\beta}_{2sls})}{\sqrt{\hat{b}'\tilde{\Omega}\hat{b}}}\right|^{2+\delta} + \left|\frac{d_{2j}(\tilde{\Omega}^{-1}\hat{a})_1}{\sqrt{\hat{a}'\tilde{\Omega}^{-1}\hat{a}}}\right|^{2+\delta}\right)E^*|z_{ji}^*v_{1i}^*|^{2+\delta} \\
&\qquad+\left(\left|\frac{d_{1j}}{\sqrt{\hat{b}'\tilde{\Omega}\hat{b}}}\right|^{2+\delta} + \left|\frac{d_{2j}(\tilde{\Omega}^{-1}\hat{a})_2}{\sqrt{\hat{a}'\tilde{\Omega}^{-1}\hat{a}}}\right|^{2+\delta}\right)E^*|z_{ji}^*v_{2i}^*|^{2+\delta}\Bigg] \\
&\quad+\sum_{l=1}^{(k+1)k/2} E^*\left|w_{li}^* - \left(\frac{1}{n}\sum_{j=1}^{n} w_{li}\right)\right|^{2+\delta}\Bigg\}.
\end{aligned}$$

First note that the denominator given by the first term $\big[d_{12}'(I_2 \otimes (Z'Z/n))\,d_{12} + d_3'\hat{\Omega}d_3\big]^{-\left(1+\frac{\delta}{2}\right)}$ is bounded away from zero almost surely since $(Z'Z/n)$ and $\hat{\Omega}$ converge *a.s.* to their positive definite limits. When $\pi = 0$, $\widehat{\beta}$ (and hence $\hat{a}$ and $\hat{b}$) has a non-degenerate limiting distribution. When $\pi \neq 0$, $\widehat{\beta}$ (and hence $\hat{a}$ and $\hat{b}$) has a finite probability limit, β. In either case the terms

$$\left|\frac{(-\widehat{\beta}_{2sls})}{\sqrt{\hat{b}'\tilde{\Omega}\hat{b}}}\right|,\ \left|\frac{(\tilde{\Omega}^{-1}\hat{a})_1}{\sqrt{\hat{a}'\tilde{\Omega}^{-1}\hat{a}}}\right|,\ \left|\frac{1}{\sqrt{\hat{b}'\tilde{\Omega}\hat{b}}}\right|, \text{and } \left|\frac{(\tilde{\Omega}^{-1}\hat{a})_2}{\sqrt{\hat{a}'\tilde{\Omega}^{-1}\hat{a}}}\right| \tag{9}$$

are always well-defined. Moreover, the terms in expression (9) are bounded by

$$\max\left\{\sqrt{\frac{\tilde{\sigma}_{11}\tilde{\sigma}_{22} - \tilde{\sigma}_{12}^2}{\tilde{\sigma}_{11}}}, \sqrt{\frac{\tilde{\sigma}_{11}\tilde{\sigma}_{22} - \tilde{\sigma}_{12}^2}{\tilde{\sigma}_{22}}}\right\}. \tag{10}$$

This bound follows from the fact that

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

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

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

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

$$\sup_{\tau} \left| \frac{\tau_1}{\sqrt{\tau' J \tau}} \right| \leq \sqrt{\frac{j_{11} j_{22} - j_{12}^2}{j_{11}}}.$$

Given the bound given by (10), the verification of the final condition in Lyapunov's Central Limit Theorem follows as in the proof of Lemma B (we also use the fact that $E\|z_i\|^{4+\delta} < \infty$ is sufficient to bound $E^* \left| w_{li}^* - (n^{-1} \sum_{j=1}^n w_{li}) \right|^{2+\delta}$ almost surely). The desired result follows. □

Corollary D *Assume π is fixed. If, for some $\delta > 0$, $E\|z_i\|^{2+\delta} < \infty$, $E\|v_i\|^{2+\delta} < \infty$, then*

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

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

Proof of Theorem 8. Under Assumption 1, the result follows from Theorem 3-a. Now, consider the case in which Assumption 1A holds, and define $t_n^* = \sqrt{n}(Z'Z/n)^{1/2} \widehat{\pi} \sqrt{\hat{a}' \tilde{\Omega}^{-1} \hat{a}}$. Then

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

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

$$\frac{Z^{*\prime} Z^*}{n} - \frac{Z'Z}{n} \Bigg| \mathcal{X}_n \xrightarrow{a.s.} 0 \ \text{ a.s.}$$

Moreover, the following holds: i) $Z'Z/n \xrightarrow{a.s.} E(z_i z_i{}')$, and so $Z^{*\prime} Z^*/n | \mathcal{X}_n \xrightarrow{a.s.} E(z_i z_i{}')$ a.s.; $Z^{*\prime} V^*/n | \mathcal{X}_n \xrightarrow{a.s.} E^*[Z^{*\prime} V^*/n] = 0$ a.s.; $\sqrt{n} \widehat{\pi}$ is bounded in probability, since $\pi = 0$; and, $\widehat{\beta}$ and $\sqrt{\hat{a}' \tilde{\Omega}^{-1} \hat{a}}$ are bounded in probability. Hence, the first term in the sum above is conditionally asymptotically negligible. It then follows from Corollary D that $(S^{*\prime}, (T^* - t_n^*)')' | \mathcal{X}_n \xrightarrow{d} N(0, I_{2k})$ a.s. The usual argument for the first order asymptotics of the score statistic in the unidentified case can then be applied to yield the desired result.

□

Proof of Theorem 9. The result is a direct application of the Delta Method and the limiting distribution given in Lemma C (and noting the zero covariances between the three components in the normal limit distribution).

□

Appendix B - Edgeworth Expansions based on Cavanagh (1983) and Rothenberg (1988) for the Multivariate Case

The Cavanagh-Rothenberg method can in principle be used for the conditional tests. Suppose

$$X_n = \bar{X}_n + \frac{P_n}{\sqrt{n}} + \frac{Q_n}{n} + O_p\left(n^{-3/2}\right), \tag{11}$$

where $\bar{X}_n$ is a $k \times 1$ vector that has distribution function F_n and density function f_n, and the variables P_n and Q_n (also $k \times 1$ vectors) possess bounded moments as n tends to infinity.

Let the conditional moments

$$p_n(x) = E\left(P_n | \bar{X}_n = x\right), \; q_n(x) = E\left(Q_n | \bar{X}_n = x\right), \; v_n(x) = V\left(P_n | \bar{X}_n = x\right)$$

be smooth functions of x. $v_n(x)$ is a $k \times k$ variance-covariance matrix, let $v_{\cdot j,n}(x)$ denote the j^{th} column of $v_n(x)$ and

$$g_n(x) = \sum_{j=1}^{k} \frac{\partial}{\partial x_j} v_{\cdot j,n}(x) + v_{\cdot j,n}(x) \frac{\partial}{\partial x_j} \ln f_n(x)$$

Next we follow Rothenberg (1988) and extend his results to the multivariate case.

$$\begin{aligned}
\psi_n(t) &= Ee^{it'X_n} \approx E\left[e^{it'\bar{X}_n}e^{it'P_n/\sqrt{n}}e^{it'Q_n/\sqrt{n}}\right] \\
&\approx E\left[e^{it'\bar{X}_n}\left(1+\frac{it'P_n}{\sqrt{n}}+\frac{it'P_nP_n'ti}{2n}\right)\left(1+\frac{it'Q_n}{n}\right)\right] \\
&\approx E\left[e^{it'\bar{X}_n}\left(1+\frac{it'P_n}{\sqrt{n}}+\frac{it'P_nP_n'ti}{2n}+\frac{it'Q_n}{n}\right)\right] \\
&= E_{\bar{X}_n}\left[e^{it'X}\left(1+\frac{it'p_n(X)}{\sqrt{n}}+\frac{it'q_n(X)}{n}+\frac{it'[v_n(X)+p_n(x)p_n(X)']ti}{2n}\right)\right] \\
&= E_{\bar{X}_n}\left[\exp\left(it'\left(X+\frac{p_n(X)}{\sqrt{n}}+\frac{q_n(X)}{n}\right)\right)\right]+\frac{1}{2n}E_{\bar{X}_n}\left[e^{it'X}it'v_n(X)ti\right]
\end{aligned}$$

For the last term, integrate by parts,

$$\begin{aligned}
&\frac{1}{2n}E_{\bar{X}_n}\left[e^{it'X}it'v_n(X)ti\right] \\
&= \frac{1}{2n}\int e^{it'x}it'v_n(x)tif_n(x)dx \\
&= \frac{1}{2n}\sum_{j=1}^{k}\int e^{it'_{-j}x_{-j}}\int e^{it'_jx_j}it'v_{\cdot j,n}(x)t_jif_n(x)dx_jdx_{-j} \\
&= \frac{1}{2n}\sum_{j=1}^{k}\int e^{it'_{-j}x_{-j}}\left[e^{it_jx_j}it'v_{\cdot j,n}(x)f_n(x)\Big|_{-\infty}^{\infty}\right. \\
&\qquad \left.-\int e^{it_jx_j}it'\left[\frac{\partial}{\partial x_j}v_{\cdot j,n}(x)+v_{\cdot j,n}(x)\frac{\partial}{\partial x_j}\ln f_n(x)\right]f_n(x)dx_j\right]dx_{-j} \\
&= -\frac{1}{2n}\int e^{it'x}it'\sum_{j=1}^{k}\left[\frac{\partial}{\partial x_j}v_{\cdot j,n}(x)+v_{\cdot j,n}(x)\frac{\partial}{\partial x_j}\ln f_n(x)\right]f_n(x)dx \\
&= -\frac{1}{2n}E\left[e^{it'\bar{X}_n}it'g_n(\bar{X}_n)\right]
\end{aligned}$$

So,

$$\psi_n(t) \approx E_{\bar{X}_n}\left[\exp\left(it'\left[X+\frac{p_n(X)}{\sqrt{n}}+\frac{2q_n(X)-g_n(X)}{2n}\right]\right)\right] = E_{\bar{X}_n}\left[e^{it'g_n(X)}\right]$$

where $h(x) = x + p_n(x)/\sqrt{n} + (2q_n(x) - g_n(x))/(2\sqrt{n})$.

We can then approximate $P\left(X_n \leq x\right) \approx P\left(h(\bar{X}_n) \leq x\right) \approx Pr(\bar{X}_n \leq h^{-1}(x))$.

Solving for $h^{-1}(x)$:

$$
\begin{aligned}
x &= h\left(x + \frac{\Delta_1(x)}{\sqrt{n}} + \frac{\Delta_2}{n}\right) \\
&= x + \frac{\Delta_1(x)}{\sqrt{n}} + \frac{\Delta_2}{n} + \frac{1}{\sqrt{n}} p_n \left(x + \frac{\Delta_1(x)}{\sqrt{n}} + \frac{\Delta_2}{n}\right) + \frac{1}{n} q_n \left(x + \frac{\Delta_1(x)}{\sqrt{n}} + \frac{\Delta_2}{n}\right) \\
&\quad - \frac{1}{2n} g_n \left(x + \frac{\Delta_1(x)}{\sqrt{n}} + \frac{\Delta_2}{n}\right) \\
&\approx x + \frac{1}{\sqrt{n}} \left(\Delta_1(x) + p(x)\right) + \frac{1}{n} \left(\Delta_2(x) + p'(x)\Delta_1(x) + q_n(x) - g_n(x)/2\right)
\end{aligned}
$$

Then,

$$
\begin{aligned}
\Delta_1(x) &= -p_n(x) \\
\Delta_2(x) &= p_n'(x) p_n(x) - q_n(x) + \frac{1}{2} g_n(x).
\end{aligned}
$$

where $p_n(x) = (p_{1,n}(x), \ldots, p_{k,n}(x))'$ is $k \times 1$ and

$$
p_n'(x) = \begin{pmatrix} \frac{\partial}{\partial x_1} p_1(x) & \cdots & \frac{\partial}{\partial x_k} p_1(x) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} p_k(x) & \cdots & \frac{\partial}{\partial x_k} p_k(x) \end{pmatrix}
$$

Then,

$$
P(X_n \leq x) = F_n \left[x - \frac{p_n\left(x\right)}{\sqrt{n}} + \frac{2p_n'\left(x\right) p_n\left(x\right) - 2q_n\left(x\right) + g_n\left(x\right)}{2n}\right].
$$

7 Appendix C - Tables

TABLE I

Percent Rejected Under Ho, Nominal 5%

Number of Simulations = 1000

$n = 20$, $k = 4$

		Normal Disturbances				Wishart Disturbances			
		LM		Wald		LM		Wald	
ρ	$\lambda'\lambda/k$	BS	3.84	BS	3.84	BS	3.84	BS	3.84
0	0	4.8	8.0	0.0	0.5	7.9	11.9	0.9	2.0
0	1	4.1	7.4	1.3	2.4	6.2	9.5	1.7	4.0
0	10	4.5	6.5	3.4	4.9	5.8	9.5	5.7	8.7
0.5	0	5.8	9.1	12.0	15.4	7.4	11.3	14.5	2.1
0.5	1	4.2	6.4	13.0	14.1	6.9	10.4	9.3	14.6
0.5	10	4.6	6.6	5.7	7.4	6.5	9.7	6.3	8.8
0.75	0	6.1	7.6	42.7	48.7	7.5	12.8	39.0	50.7
0.75	1	4.3	6.5	27.9	32.6	6.9	9.7	22.7	29.2
0.75	10	4.9	6.3	7.6	10.6	7.0	10.5	8.2	12.3
0.99	0	5.9	7.6	95.2	99.1	9.0	13.3	93.7	98.3
0.99	1	4.5	6.5	35.4	57.2	7.0	10.3	31.7	51.3
0.99	10	5.1	6.5	9.1	14.2	7.0	10.6	9.0	15.2

TABLE II

Percent Rejected Under Ho, Nominal 5%

Number of Simulations = 1000

$n = 80$, $k = 4$

		Normal Disturbances				Wishart Disturbances			
		LM		Wald		LM		Wald	
ρ	$\lambda'\lambda/k$	BS	3.84	BS	3.84	BS	3.84	BS	3.84
0	0	5.8	6.3	0.0	0.0	5.7	6.6	0.2	0.3
0	1	5.5	6.1	0.1	1.3	5.6	6.0	0.3	1.4
0	10	5.2	5.8	4.3	4.6	5.1	5.6	4.7	5.0
0.5	0	6.4	7.1	12.8	15.9	5.3	6.0	10.8	14.0
0.5	1	5.6	5.9	16.0	13.8	5.3	6.0	11.2	12.9
0.5	10	5.5	6.0	6.9	6.9	5.5	6.2	5.6	6.7
0.75	0	6.0	6.8	46.3	47.9	5.8	6.4	44.2	49.2
0.75	1	4.8	5.4	29.5	31.4	5.8	6.1	26.1	28.5
0.75	10	6.4	6.4	7.7	9.1	4.8	6.0	5.9	9.0
0.99	0	5.5	5.9	95.2	98.9	6.2	6.7	95.4	98.8
0.99	1	4.9	5.2	29.3	54.3	7.2	7.7	28.6	56.9
0.99	10	5.4	5.3	7.7	12.2	7.2	8.0	7.6	12.9

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

ρ	$\lambda'\lambda/k$	h: 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 V - Panel B (Wishart Disturbances)
Percent Rejected Under Ho, 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