Bootstrap Validity for the Score Test When Instruments May Be Weak [1]

Marcelo J. Moreira
Harvard University and NBER

Jack R. Porter
University of Wisconsin

Gustavo A. Suarez [2]
Harvard University

First Version: April 2003.
Current Version: July 2005.

 This paper is the first part of a joint project entitled "Bootstrap and Higher-Order Expansion Validity When Instruments May Be Weak"; see Moreira, Porter, and Suarez (2004).

[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 Boston College, Boston University, Cornell, Harvard-MIT, Maryland, Montreal, Rice, Texas A&M, Texas Austin, and USC-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 bootstrapping the t-statistic perform poorly when instruments are weakly correlated with the endogenous explanatory variable. In this paper, we provide a theoretical proof that guarantees the validity of the bootstrap for the score statistic. This theory does not follow from standard results, since the score statistic is not a smooth function of sample means and some parameters are not consistently estimable when the instruments are uncorrelated with the explanatory variable. This finding should also serve as a reminder that the bootstrap may work for problems of statistical inference where the statistics are not smooth or the parameters are not consistently estimable.

Keywords: bootstrap, t-statistic, score statistic, identification, non-regular case, Edgeworth expansion, instrumental variable regression.

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). It is then natural to apply the bootstrap to decrease size distortions of the Wald statistic (also known as the t-statistic), since the bootstrap is valid under some regularity conditions. However, these conditions, which rely on the statistics being smooth functions of sample moments and the parameters being consistently estimable, break down for the Wald statistic in the weak-instrument case. In fact, the bootstrap does not perform well to decrease size distortions of the Wald statistic; cf., Horowitz (2001).

In this paper, we show that it is valid to bootstrap the score statistic even in the weak-instrument case. Although the score is well-behaved with weak instruments, showing the validity of the bootstrap in the unidentified case has several potential pitfalls. First, the bootstrap replaces parameters with inconsistent estimators. 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. Second, the score statistic is not a smooth function of sample means. In many known non-regular cases[3] the usual bootstrap method fails, *even in the first-order*; cf. Shao (1994) and Andrews (2000). We point out here three non-regular cases in which the bootstrap fails (we use * to indicate the bootstrap sample and, with some abuse of notation, we use the same symbols for the parameters and random variables in all examples):

[3]A statistic is said to be regular if, when written as function of sample moments, the first derivative of this function evaluated at the populational mean exists and is different from zero.

Example 1 [null first-order derivatives]: Let $\theta = g(\mu)$ and $\widehat{\theta}_n = g\left(\overline{X}_n\right)$, where μ is the population mean and $\overline{X}_n$ is the sample average of $X_i \overset{i.i.d.}{\sim} N(\mu, 1)$. The function g is second-order continuously differentiable at μ with $\nabla g(\mu) = 0$ and $\nabla^2 g(\mu) \neq 0$. The continuity of g guarantees that $\widehat{\theta}_n$ is a consistent estimator of θ, and the statistic $n(\widehat{\theta}_n - \theta)$ has a limiting distribution (in fact, $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{p} 0$ due to the null first-order derivative). However, Shao (1994) shows that the bootstrapped statistic $n(\widehat{\theta}_n^* - \widehat{\theta}_n)$ does not have a limiting distribution for almost all X_1, $X_2, ...$

Example 2 [estimation on the boundary of the parameter space]: Let $\widehat{\theta}_n = \max\{\overline{X}_n, 0\}$ be the MLE of the population mean of $X_i \overset{i.i.d.}{\sim} N(\mu, 1)$ with $\mu \geq 0$. If μ is on the boundary of the parameter space, $\sqrt{n}(\max\{\overline{X}_n, 0\} - \mu) \xrightarrow{d} \max\{Z, 0\}$, where $Z \sim N(0, 1)$. However, for $\mu = 0$, Andrews (2000) shows that $\sqrt{n}(\max\{\overline{X}_n^*, 0\} - \max\{\overline{X}_n, 0\})$ does not converge in distribution to $\max\{Z, 0\}$ for almost all $X_1, X_2, ...$

Example 3 [Non-Differentiable Functions]: Let $\theta = g(\mu)$ and $\widehat{\theta}_n = g\left(\overline{X}_n\right)$, where $X_i \overset{i.i.d.}{\sim} N(\mu, 1)$. The function g is non-differentiable at μ. In general, $\sqrt{n}(\widehat{\theta}_n - \theta)$ is not asymptotically normal, and it may not have a limiting distribution when both $g'(\mu^+)$ and $g'(\mu^-)$ are nonzero. Moreover, Shao (1994) shows the inconsistency of the bootstrap estimator $\sqrt{n}(\widehat{\theta}_n^* - \widehat{\theta}_n)$ for almost all $X_1, X_2, ...$

These stylized examples summarize some of the limitations of the bootstrap in the econometrics and statistical literature. A commonly used fix is to use the m out of n bootstrap. However, this method has two limitations. First, in practice it gives quite different results for different choices of the bootstrap sample m. Second, it does not provide asymptotic refinements if we are in a regular case. In the context of Examples 1-3, the statistics are non-regular for one value of the parameter μ. For all other parameter values, the statistics are typically regular and the usual bootstrap is not only valid but also provides second-order improvements. Hence, there is a trade-off between robustness (m out of n bootstrap) and refinements (usual bootstrap).

For statistical and econometric problems related to Example 1, more elegant solutions are readily available. For example, in regression models, we are often willing to make the additional assumption that the population mean of the disturbances is zero. In this case, re-centering the residuals (or moment conditions) can often solve the problem of bootstrapping statistics (with null first-order derivatives); e.g., see Lahiri (1992), Shao (1994), Hall and Horowitz (1996), and Brown and Newey (2004).

However, to our knowledge, there is no available method to fix problems related to non-smooth statistics (Examples 2 and 3). This paper aims to fill this gap in the literature, by providing the first proof of the validity of the bootstrap for a statistic that is not smooth or when the parameters are not identified. Our proof is quite simple and some steps can be generalized to a variety of problems in which the statistics are not smooth and some parameters are not consistently estimable. In light of all the negative results on the bootstrap in non-regular cases, our finding should also serve as a reminder that the bootstrap may work for problems in which the statistics are not smooth or the parameters are not consistently estimable.

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 the bootstrap for the Wald and score tests under standard asymptotics. In Section 4, we present the main results. We establish the validity of the bootstrap for the score statistic, and show that the bootstrap will not in general provide second-order improvements in the unidentified case. 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 with probability one (w.p.1) 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 $N_A = A\left(A'A\right)^{-1}A'$ 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 (two-sided) Wald test rejects H_0 for large values (of the square) 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. An alternative statistic is the score (LM) used by Kleibergen (2002) and Moreira (2001):

$$LM = S'T/\sqrt{T'T},$$

where $S = (Z'Z)^{-1/2} Z'Y b_0 \cdot (b_0' \Omega b_0)^{-1/2}$ and $T = (Z'Z)^{-1/2} Z'Y \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2}$. The (two-sided) score test rejects the null if the LM^2 statistic is larger than the $1 - \alpha$ quantile of the chi-square-one distribution. This test is similar if the errors are normal with known variance Ω, since the LM statistic is pivotal.

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

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

With unknown error distribution, the score test is no longer similar. However, unlike the Wald test, the score test is asymptotically similar under both weak-instrument and standard asymptotics.

3 Preliminary Results

In this section, we summarize some folk theorems for the good-instrument case. 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 matrix A, let $vech(A)$ denote the 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 $\ell = (k+2)(k+3)/2$. Both W and LM statistics 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.

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 $E\|(Y_n', Z_n')\|^{2s} < \infty$. 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 are the 2SLS and LIML estimators consistent for β, but also both Wald and score statistics admit second-order Edgeworth expansions under mild conditions (see Sargan (1980) and Park and Phillips (1988)). 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) $\left\| P\left(LM_n \leq x\right) - [\Phi\left(x\right) + \sum_{i=1}^{s-2} n^{-i/2} p_{LM}^i\left(x; F, \beta_0, \pi\right)\phi\left(x\right)] \right\|_\infty = o\left(n^{-(s-2)/2}\right),$
(b) $\left\| P\left(W_n \leq x\right) - [\Phi\left(x\right) + \sum_{i=1}^{s-2} n^{-i/2} p_{W}^i\left(x; F, \beta_0, \pi\right)\phi\left(x\right)] \right\|_\infty = o\left(n^{-(s-2)/2}\right),$
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) $\left\| P^* \left(LM_n^* \leq x \right) - \left[\Phi(x) + \sum_{i=1}^{s-2} n^{-i/2} p_{LM}^i (x; F_n, \widehat{\beta}, \widehat{\pi}) \phi(x) \right] \right\|_\infty = o\left(n^{-(s-2)/2}\right)$,
(b) $\left\| P^* \left(W_n^* \leq x \right) - \left[\Phi(x) + \sum_{i=1}^{s-2} n^{-i/2} p_{W}^i (x; F_n, \widehat{\beta}, \widehat{\pi}) \phi(x) \right] \right\|_\infty = o\left(n^{-(s-2)/2}\right)$,
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.

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. In this section, we address the issues above that arise in a weak-instrument setting. Formally, we replace Assumption 1 by one of the following poor-instrument assumptions:

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

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

Under Assumption 1A, β_0 is replaced by an inconsistent estimator $\widehat{\beta}$ and the score and Wald statistics are non-smooth functions of sample means. However, the proofs of bootstrap validity for statistics in the form (3) crucially depends upon the assumption that the derivatives of functions evaluated at $\mu = E(R_n)$ are defined and different from zero (regular case). In light of Examples 1-3, 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 perform well; cf. Horowitz (2001).

4.1 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

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

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 $\widehat{\Omega}$ based on $(\tilde{v}_1, \tilde{v}_2)$ in the bootstrap test statistics. Therefore, we have:

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

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

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

To our knowledge, the next result contains the first formal proof of the bootstrap for cases where some parameters are not identified or the statistic is not smooth.

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

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

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

Theorem 4 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 and weak-instrument cases is notable due to lack of differentiability and inconsistency of estimators.

Comments: 1. Validity of the bootstrap in the (locally) unidentified case is the main result in this paper. For completeness, we also show the first-order validity under the same moment conditions for the good instrument case. Of course, second-order improvements with good instruments are available under stronger assumptions; see Theorem 3(a).

2. Two alternative bootstrap methods could also be pursued. The first alternative amounts to 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). The second alternative amounts to doing OLS regressions in the reduced-form model ignoring the nonlinear constraints of the reduced-form coefficients. However, the interpretation from bootstrapping from the structural form residuals is no longer valid (in the over-identified model).

3. 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 4 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).

4.2 Edgeworth Expansions

Theorem 3(a) shows that the bootstrap provides higher-order improvements for the score (and Wald) statistic in the identified case. Theorem 4 shows that the bootstrap provides a valid first-order approximation for the score even in the unidentified case. In this section, we show that the bootstrap typically does not deliver higher-order improvements (in the usual sense). The reason is twofold. First, the higher-order terms typically depend on $\widehat{\beta}$ separately from the term $\widehat{\pi}\widehat{\beta}$. Second, the higher-order terms are not necessarily continuous functions of the parameters in the unidentified case.

To be more specific, as noted in section 3, under the null, the score statistic can be written as a function of sample means

$$LM = \sqrt{n}\left(H\left(\overline{R}_n\right) - H\left(\mu\right)\right). \tag{4}$$

This function is non-differentiable in the unidentified case. Using the fact that $E\left[v_{\cdot i}\right] = 0$, we find that

$$LM = \sqrt{n}H\left(\overline{R}_n\right) \quad \left(= S'T/\sqrt{T'T}\right). \tag{5}$$

Expressions (4) and (5) suggest two possible bootstrap methods. The first one would use $\overline{R}_n$ and $\overline{R}_n^*$, the sample means and bootstrap sample means based on residuals (without re-centering). The bootstrap statistic would be

$$LM^* = \sqrt{n}\left(H\left(\overline{R}_n^*\right) - H\left(\overline{R}_n\right)\right). \tag{6}$$

Because $H\left(\overline{R}_n\right)$ is not necessarily zero here, we would need to expand (6). Since $H(\cdot)$ is not smooth, this bootstrap would be problematic. The second, more commonly-used bootstrap method would be based on $\overline{\widetilde{R}}_n$ and $\overline{\widetilde{R}}_n^*$, respectively, the sample means and bootstrap sample means using re-centered residuals. Then the bootstrapped score statistic, as defined in section 4.1, can be written as

$$LM^* = \sqrt{n}\left(H\left(\overline{\widetilde{R}}_n^*\right) - H\left(\overline{\widetilde{R}}_n\right)\right). \tag{7}$$

Due to re-centered residuals, $H\left(\overline{\widetilde{R}}_n\right) = 0$, expression (7) can be further simplified as above:

$$LM^* = \sqrt{n}H\left(\overline{\widetilde{R}}_n^*\right) \quad \left(= S^{*\prime}T^*/\sqrt{T^{*\prime}T^*}\right). \tag{8}$$

Equation (8) provides intuition for the first order validity of the second bootstrap without parameter identification. The re-centered residuals, allows us to rely on the continuous mapping theorem, applied to (8), to yield first-order validity.

Although the standard bootstrap provides a first-order approximation, the lack of differentiability in (7) means that the standard expansion arguments in Bhattacharya and Ghosh (1978) break down in the unidentified case. Because standard expansion arguments rely on smoothness, higher-order improvement results for empirical Edgeworth expansions (or the bootstrap) may not hold here. For example, consider the problem of finding second-order Edgeworth expansions for the $\widetilde{LM}$ statistic when Ω is unknown but the errors are normal.[4] We can compute the high-order terms using standard methods. Atlernatively, we can adapt the results in Cavanagh (1983) and Rothenberg (1988) to compute the second-order Edgeworth distribution for $\widetilde{LM}$ based on a stochastic expansion:

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

[4]Although the 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).

where P and Q are stochastically bounded with conditional moments

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

Proposition 5 *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*

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

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

Comments: 1. 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).

2. Recall that under normality the LM statistic is $N(0,1)$ under H_0, but the $\widetilde{LM}$ statistic is not. Therefore, Proposition 5 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.

The higher-order terms for the score statistic typically depend on π and Ω. In practice, we do not know π and Ω, and need to replace them with consistent estimators in the high-order terms. As long as the higher-order polynomials are continuous functions of the parameters, empirical Edgeworth expansions (or the bootstrap) leads 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 the score statistic at the unidentified case. For example, suppose that $E(v_i|Z_i) = 0$, $E(v_i v_i'|Z_i) = \Omega$, and $\Omega_{ZZ} = E(Z_i Z_i') < \infty$. 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

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

This higher-order term in general cannot be extended to be continuous at $\pi = 0$ (pick $\pi = c \cdot n^{-1/2} \to 0$ for different vectors c). Thus, the empirical Edgeworth expansion and bootstrap[5] approaches typically do not provide a $n^{-1/2}$ correction and can perform poorly in the unidentified case.

5 Monte Carlo Simulations

In this section, we use simulation to examine the size performance of the bootstrap for Wald and score test statistics. The basic simulation model is described by equations (1) and (2). The n rows of $[u, v_2]$ are i.i.d. with mean zero, unit variance, and correlation ρ. The correlation coefficient, ρ, represents the degree of endogeneity of y_2, and the distribution for these disturbances will vary depending on the design, as described below. We take the first column of the matrix of instruments, Z, to be a vector of ones; the remaining $k-1$ columns are distributed $N(0, I_{k-1})$, independently from $[u, v_2]$.[6]

To examine the performance of the tests under various degrees of identification, we consider three different values of the population first-stage F-statistic, $\pi'(nI_k)\pi/k$.[7,8] In particular, we consider the completely unidentified case ($\pi'(nI_k)\pi/k = 0$); a weak-instruments case ($\pi'(nI_k)\pi/k = 1$); and a good-instruments case ($\pi'(nI_k)\pi/k = 10$).

[5] In addition, the bootstrap has the problem of replacing β with the inconsistent estimator $\widehat{\beta}$.

[6] There is a slight difference between Moreira's (2001, 2003) design and ours. In perfect analogy with our results, our design takes Z as being random whereas Moreira's (2001, 2003) design takes Z as being fixed.

[7] The first-stage F-statistic corresponds to the concentration parameter $\lambda'\lambda/k$ in the notation of Staiger and Stock (1997).

[8] Note that $E(Z'Z) = nI_k$.

The Monte Carlo simulations in this section compute actual rejection rates for the two-sided version of the score and Wald tests for the hypothesis that $\beta = 0$. For each specification, 1000 pseudo-data sets are generated under the null hypothesis. For each pseudo-data set, we compute rejections using two different critical values: (a) the 5% critical value of the asymptotic distribution of the score test (i.e., chi-square one); (b) the bootstrapped 5% critical value computed with 1000 replications for each pseudo-data set.

The simulations are designed to consider various cases of disturbance distributions. The first set of simulations restrict attention to disturbances following conventional distributions with small sample sizes. For design I, we take u_t and v_{2t} to be jointly normally distributed. For design II, we consider Wishart distributed disturbances. In particular, we take $u_t = (\xi_{1t}^2 - 1)/\sqrt{2}$ and $v_{2t} = (\xi_{2t}^2 - 1)/\sqrt{2}$, where ξ_{1t} and ξ_{2t} are standard normal random variables with correlation $\sqrt{\rho}$. The next set of simulations considers various nonnormal disturbances with a large sample size. In design III, u_i and v_{2i} are distributed $\sqrt{(df-2)/df} * t(df)$ where $t(df)$ represents a Student's t distribution with df degrees of freedom, and the factor $\sqrt{(df-2)/df}$ normalizes the variance of the disturbances to 1. Under this design, disturbances exhibit excess kurtosis (including unbounded fourth moments for $df = 3, 4$). In design IV, u_i and v_{2i} follow double exponential (Laplace) distributions, with excess kurtosis of 3. Finally, in design V, the pdf's of u_i and v_{2i} are averages of the pdf's of two normal random variables.[9] Designs III and IV consider distributions with large or not defined fourth moments, while design V captures the case of disturbances with bimodal distributions.

Tables I and II report null rejection probabilities for the score and Wald tests with sample sizes of 20 and 80 observations. Bootstrapping the score test instead of using the first-order asymptotic approximation takes actual rejection rates closer to the nominal size, sometimes even in the unidentified

[9] To guarantee that u_i and v_{2i} are mean zero, the two normal random variables should have means equal in magnitude and opposite in sign. To guarantee that u_i and v_{2i} have unit variance, the mean and variance of each normal random variable must satisfy $mean = \sqrt{1 - variance}$. In our simulations, we set $variance = 0.25$.

case.[10] For the smaller sample size results, the bootstrapped score is closest to the nominal 5% level when the instruments are strongest.

The performance of the Wald test exhibits even more sensitivity to instrument strength. When instruments are not strong, the size distortions for both asymptotic and bootstrap critical values can be dramatic. On the other hand, when instruments are strong, rejection rates for the Wald test are much closer to the nominal size, and bootstrapping the Wald test offers improvements over first-order asymptotics. The poor behavior of the bootstrap for the Wald test with weak instruments is explained, as previously, by its dependence on π. For the remaining designs, we focus on the behavior of the score test.

Table III gives large sample size results for the score test under varying nonnormal disturbance distributions. Under design III with degrees of freedom equal to 3 or 4, rejection rates using the bootstrapped critical value are generally closer to the nominal 5% level, sometimes even in the (locally) unidentified case. However, for the distribution with the greatest degrees of freedom in our simulations, bootstrapped critical values do not perform systematically better than the first-order asymptotic approximation.

Under design IV (double exponential distributed disturbances), the bootstrapped critical value does not offer improvements over the first-order approximation in the unidentified case. The bootstrapped score is able to improve upon the first-order approximation only in designs with strong instruments. The two methods of computing rejections rates deliver essentially equivalent results in the weak instruments case.

Under design V (the mixture of normals distribution), the bootstrapped critical value offers some improvements in computing the size of the score test in the unidentified case with intermediate values of ρ. Just as in the previous design, rejection rates computed with the bootstrapped critical value in design V are equivalent to rejection rates obtained using the asymptotic approximation. For this design, the strong instrument case does not pro-

[10]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.

vide a systematic conclusion in comparing the bootstrap and the asymptotic critical values.

While the empirical rejection rates in Table III do not provide conclusive evidence on the bootstrap versus asymptotic critical values for the score test, these results are consistent with the theoretical findings of sections 3 and 4. In particular, the bootstrap does not yield higher order improvements for weak instrument or unidentified cases, though first order equivalence is shown in Theorem 4. For the strong instrument case, the higher order improvements yielded by the bootstrap are not readily apparent, as they were in Tables I and II.

6 Conclusions and Extensions

It is well-known that in the good-instrument case, the Wald statistic (and the score statistic) are smooth functions of sample means and that the bootstrap provides higher-order improvements. In the unidentified case, the statistics are in general non-regular and the standard proofs for the validity of the bootstrap break down. This paper provides a positive result that bootstrapping the score statistic is valid even when there is no identification and the score statistic is not smooth. In light of the negative results of Shao (1994) and Andrews (2000), our finding should also serve as a reminder that the bootstrap may work in problems in which the parameters are not consistently estimable or the statistics are not smooth.

As a negative result, the bootstrap for the score does not in general provide standard improvements in the weak-instrument case. This is due to the structural parameter not being consistently estimable and the higher-order polynomials of the Edgeworth expansions not being necessarily continuous at the unidentified case. Nevertheless, this discontinuity due to non-differentiability of the score statistic can be quite interesting, given that little is known about expansions when the statistic is not smooth. In the words of Wallace (1958): "The assumption $H'(\mu) \neq 0$ and its equivalent for functions of several moments rule out many interesting functions for which no general theory of asymptotic expansions is known." Yet, there has not been none or any work in the last decades on expansions for non-smooth statistics.

Finally, our results for the unidentified case can in principle be extended to the GEL and GMM contexts; cf. Guggenberger and Smith (2003), Stock and Wright (2000), and Brown and Newey (2004). 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 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. Moreover, in companion work (Moreira, Porter, and Suarez (2004)), we show that our proof for bootstrap validity of the score statistic can be extended to the Anderson-Rubin statistic, and generalized to accommodate two conditional bootstrap methods for the CLR test of Moreira (2003).

References

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

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

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.

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.

Brown, B., and W. Newey (2004): "GMM, Efficient Booststrapping, and Improved Inference," *Journal of Business and Economic Statistics*, forthcoming.

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

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.

Goncalvez, S., and N. Meddahi (2005): "Bootstrapping Realized Volatility," *Unpublished Manuscript.*

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

Hall, P., and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64, 891–916.

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.

INOUE, A., AND L. KILIAN (2003): “The Continuity of the Limit Distribution in the Parameter of Interest is not Essential for the Validity of the Bootstrap,” *Econometric Theory*, 19, 944–961.

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.

LAHIRI, S. N. (1992): “Bootstrapping M-Estimators of a Multiple Linear Regression Parameter,” *The Annals of Statistics*, 20, 1548–1570.

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

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

MOREIRA, M. J., J. R. PORTER, AND G. A. SUAREZ (2004): “Bootstrap and Higher-Order Expansion Validity When Instruments May Be Weak,” *NBER Working Paper*, T0302.

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.

PARK, J., AND P. PHILLIPS (1988): “On the Formulation of Wald Tests of Nonlinear Restrictions,” *Econometrica*, 56, 1065–1083.

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

Sargan, J. D. (1980): "Some Approximations to the Distribution of Econometric Criteria which are Asymptotically Distributed as Chi-Squared," *Econometrica*, 48, 1107–1138.

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 $H(\mu) = 0$. 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.

$\square$

Proof of Theorem 3. Let F be the distribution of

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

and let F_n be the distribution of

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

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

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

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

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

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

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

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

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

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

is $o\left(n^{-1}\right)$ 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 $H\left(\overline{\widetilde{R}}_n\right) = 0$ (due to re-centered residuals).

□

In Lemma A, we show that we typically have$(\widehat{\pi}, \widehat{\pi\beta})$ 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:

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

$$\textit{ii)}\ \lim_{\rho \to 0} E_{\theta_0} \ln f(X; \theta, \rho) = E_{\theta_0} \ln f(X; \theta) \text{ for any } \theta_0 \in \mathcal{L}, \theta \in \Theta,$$

$$\textit{iii)}\ E_{\theta_0} \ln \varphi(X; r_0) < E_{\theta_0} \ln f(X; \theta_0) \text{ for any } \theta_0 \in \mathcal{L}, \text{some } r_0 \in \mathbb{R}^+.$$

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

$$\frac{\prod_{\alpha=1}^{n}f\left(x_\alpha;\widehat{\theta}_n\right)}{\prod_{\alpha=1}^{n}f\left(x_\alpha;\theta_0\right)}\geq\gamma>0\text{ for all }n\text{ and }x_1,...,x_n. \tag{9}$$

Then:

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

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

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

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 $\rho_{\theta i}$, $B\left(\theta_i,\rho_{\theta i}\right)$, $i=1,...,h$, cover W_1. Now,

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

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

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

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 $\left\{\widehat{\theta}_n\right\}$

satisfy the inequality $\left|\left(\overline{\pi}, \overline{\pi\beta}\right)\right| \leq \epsilon$. The event that there exists a limit point $\overline{\theta}$ such that $\left|\left(\overline{\pi}, \overline{\pi\beta}\right)\right| > \epsilon$ implies that

$$\sup_{\theta:|(\pi,\pi\beta)|\geq\epsilon} \prod_{\alpha=1}^{n} f\left(x_\alpha;\theta\right) \geq \prod_{\alpha=1}^{n} f\left(x_\alpha;\widehat{\theta}_n\right)$$

for infinitely many n. But then

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

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

$\square$

Comments: The maximum likelihood estimator $\widehat{\theta}_{MLE} = (\widehat{\pi}_{MLE}, \widehat{\beta}_{MLE})$, if it exists, satisfies (9) 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 $\pi = 0$.

Proof of Theorem 4. The following result is a key step in the proof of the theorem's conclusion:

$$(10) \qquad \left.\left(\begin{array}{c} \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right)\frac{\widehat{b}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}} \\ \left(\frac{Z^{*\prime}V^*}{\sqrt{n}}\right)\frac{\widehat{\Omega}^{-1}\widehat{a}}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}} \end{array}\right)\right| \mathcal{X}_n \xrightarrow{d} N\left(0, I_2 \otimes E(Z_i Z_i')\right) \quad \text{a.s.}$$

under either condition (i) or (ii). Equation (10) will follow by application of the Cramer-Wald device and verification of the Liapunov Central Limit Theorem conditions.

Let $c = (c_1', c_2')' \in \mathbb{R}^{2k}$ be a nonzero vector. Define

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

where $v_{\cdot i}^* = \left[v_{1,i}^*, v_{2,i}^*\right]'$ is the i-th bootstrap draw of the (re-centered) reduced-form residuals, and $\widehat{J} = \left[\frac{\widehat{b}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}}, \frac{\widehat{\Omega}^{-1}\widehat{a}}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}}\right]$.

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

(ii) By independence,

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

is finite a.s.

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

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

for large enough constants C_1, C_2, and C_3.

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

$$\left| \frac{1}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}} \right|, \left| \frac{(\widehat{\Omega}^{-1}\widehat{a})_1}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}} \right|, \left| \frac{-\widehat{\beta}}{\sqrt{\widehat{b}'\widehat{\Omega}\widehat{b}}} \right|, \text{and } \left| \frac{(\widehat{\Omega}^{-1}\widehat{a})_2}{\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}} \right| \tag{11}$$

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

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

where $\widehat{\sigma}_{ij}$ is the (i,j)-th entry of $\widehat{\Omega}$.[11]

[11] This bound follows from the fact that

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

Next we show that $E^*[|Z^*_{j,i}v^*_{m,i}|^{2+\delta}]$ is bounded a.s. for $j = 1, \ldots, k$ and $m = 1, 2$. By independence, $E^*[|Z^*_{j,i}v^*_{m,i}|^{2+\delta}] = E^*[|Z^*_{j,i}|^{2+\delta}]E^*[|v^*_{m,i}|^{2+\delta}]$. For $j = 1, \ldots, k$,

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

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

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

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

Using Minkowski's inequality, we get

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

---

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

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

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

$$\left|\frac{\tau_1}{\sqrt{\tau' K\tau}}\right| \leq \sup_\tau \sqrt{\frac{\tau'(e_1 e_1')\tau}{\sqrt{\tau' K\tau}}} = \sqrt{\frac{k_{22}}{k_{11}k_{22} - k_{12}^2}}.$$

and note $\|\overline{z}\| \overset{a.s.}{\to} \|E[Z_i]\| \le E\|Z_i\| \le (E\|Z_i\|^{2+\delta})^{1/(2+\delta)}$ by Jensen's inequality. Similarly, using Minkowski's inequality again, we get

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

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

The final condition of the Liapunov Central Limit Theorem now follows by noting that

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

is bounded away from zero almost surely since $(Z'Z/n)$ converges $a.s.$ to its positive definite limit.

$\square$

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

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

Moreover, $Z'Z/n \overset{a.s.}{\to} E(Z_iZ_i{}')$, and so $Z^{*\prime}Z^*/n|\mathcal{X}_n \overset{a.s.}{\to} E(Z_iZ_i{}')$ a.s. Similarly, $Z^{*\prime}V^*/n|\mathcal{X}_n \overset{a.s.}{\to} E^*[Z^{*\prime}V^*/n] = 0$ a.s. So, under (i), $T^{*\prime}T^*/n|\mathcal{X}_n \overset{a.s.}{\to} \pi'E(Z_iZ_i{}')\pi(a'\Omega^{-1}a)$ a.s. By (10), $\widehat{b}'(V^{*\prime}Z^*/\sqrt{n})\widehat{\pi}|\mathcal{X}_n \overset{d}{\to} N(0,\pi'E(Z_iZ_i{}')\pi(b'\Omega b))$. The result follows under condition (i).

Now, consider the case (ii) in which Assumption 1B (or 1A) holds, and define $t_n^* = \sqrt{n}(Z'Z/n)^{1/2}\widehat{\pi}\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}$. Then

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

The first term in the sum above is conditionally asymptotically negligible since $\sqrt{n}\widehat{\pi}$, $\widehat{\beta}$ and $\sqrt{\widehat{a}'\widehat{\Omega}^{-1}\widehat{a}}$ are bounded in probability. It then follows from (10) that $(S^{*\prime},(T^*-t_n^*)')'|\mathcal{X}_n \overset{d}{\to} N(0,I_{2k})$ a.s. The usual argument for the first-order asymptotics of the score statistic in the unidentified and weak instrument cases can then be applied to yield the desired result. $\square$

7 Appendix B - Tables

TABLE I

Percent Rejected Under Ho, Nominal 5%

$n = 20$, $k = 4$ (replications = 1000)

		Normal				Wishart			
		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%

$n = 80$, $k = 4$ (replications = 1000)

		Normal				Wishart			
		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 for the score test under Ho, nominal 5%

$n = 1000$, $k = 4$ (replications = 1000)

		t-Student (df degrees of freedom)						Double Exponential		Mixture of Normals	
		$df = 3$		$df = 4$		$df = 5$					
ρ	$\frac{\pi'(nI_k)\pi}{k}$	BS	3.84	BS	3.84	BS	3.84	BS	3.84	BS	3.84
0	0	5.3	5.4	5.8	6.0	5.5	6.2	5.9	5.9	6.0	5.8
0	1	6.3	6.7	7.0	7.3	5.9	6.2	5.9	5.8	5.4	4.9
0	10	5.0	5.6	5.3	5.8	4.4	4.6	6.0	6.2	4.8	5.5
0.5	0	5.7	6.5	3.5	3.9	5.8	5.8	6.4	5.7	4.9	3.9
0.5	1	5.2	5.8	4.6	4.6	3.1	3.1	5.2	5.2	6.8	6.9
0.5	10	4.9	5.0	3.6	3.9	4.2	4.2	5.0	4.6	5.3	4.7
0.75	0	4.8	5.0	4.9	5.7	5.3	5.8	5.3	5.0	5.4	4.3
0.75	1	4.9	5.4	5.2	5.7	6.4	6.3	5.3	5.2	4.0	4.0
0.75	10	4.1	4.2	5.6	5.6	4.1	4.4	5.9	5.7	5.9	5.1
0.99	0	4.9	4.8	5.5	5.8	5.3	5.4	6.9	5.8	5.0	5.1
0.99	1	5.2	5.3	5.7	5.8	5.3	5.3	5.3	4.6	5.2	5.4
0.99	10	5.7	5.7	5.4	5.3	5.0	4.3	5.5	4.0	5.3	4.5