EXACTLY DISTRIBUTION-FREE INFERENCE IN INSTRUMENTAL VARIABLES REGRESSION WITH POSSIBLY WEAK INSTRUMENTS

By

Donald W.K. Andrews and Vadim Marmer

March 2005

COWLES FOUNDATION DISCUSSION PAPER NO. 1501

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Exactly Distribution-Free Inference in Instrumental Variables Regression with Possibly Weak Instruments

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

Vadim Marmer
Cowles Foundation for Research in Economics
Yale University

February 2004
Revised: March 2005

[1] The authors thank Jean-Marie Dufour, Jin Hahn, Whitney Newey, Tom Rothenberg, and participants of seminars at Harvard/MIT, Toronto, and UCLA/USC for comments. The first author gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-0001706 and SES-0417911.

Abstract

This paper introduces a rank-based test for the instrumental variables regression model that dominates the Anderson-Rubin test in terms of finite sample size and asymptotic power in certain circumstances. The test has correct size for any distribution of the errors with weak or strong instruments. The test has noticeably higher power than the Anderson-Rubin test when the error distribution has thick tails and comparable power otherwise. Like the Anderson-Rubin test, the rank tests considered here perform best, relative to other available tests, in exactly-identified models.

Keywords: Aligned ranks, Anderson-Rubin statistic, categorical covariates, exact size, normal scores, rank test, weak instruments, Wilcoxon scores.

JEL Classification Numbers: C13, C30.

1 Introduction

The Anderson-Rubin (1949) (AR) test has a long history in econometrics. It was introduced over fifty years ago, but it has seen a resurgence of popularity in the last decade due to increased concern with the quality of inference in the presence of weak instruments (IVs). The AR test has the property that it has exactly correct size in the IV regression model with normally distributed errors regardless of the properties of the IVs. Few other statistics have this property. Furthermore, in exactly-identified models, the AR test is asymptotically best unbiased under weak IV asymptotics and asymptotically efficient under strong IV asymptotics.[2]

In this paper, we introduce a rank-based statistic that is similar to the AR statistic, but has improved finite sample size and asymptotic power in certain circumstances. Its size properties are improved because it has exact size for any distribution of the errors, not just normal errors. Its asymptotic power properties are improved because it has equal asymptotic power under normal errors and considerably higher power for thick-tailed error distributions. This holds under both weak and strong IV asymptotics. These advantages occur in IV regression models in which the IVs are independent of the errors (not just uncorrelated) and (i) are simple, i.e., have no covariates, (ii) have IVs that are independent of the covariates, or (iii) have categorical covariates.

Type (i) and (iii) models are used regularly in the applied literature, e.g., both are used in Angrist and Krueger (1991) and Duflo and Saez (2003). The rank tests for these models have exact size for any error distribution. In type (iii) models, the rank tests allow the error distribution to differ across the covariate categories.

Type (ii) models arise frequently in applications utilizing natural or randomized experiments, e.g., see Angrist and Krueger (1991), Levitt (1997), Angrist and Evans (1998), Duflo (2001), and Angrist, Bettinger, Bloom, King, and Kremer (2002). The tests have exact size for any error distribution and allow the errors to be conditionally heteroskedastic given the covariates. We handle covariates in these models by "aligning" the ranks. This method has been used widely in the statistics literature, e.g., see Hodges and Lehmann (1962), Koul (1970), and Hettmansperger (1984). Unlike most results in the statistics literature, however, our aligned rank tests are exactly distribution free, not just asymptotically distribution free.

In over-identified models, the conditional likelihood ratio (CLR) test of Moreira (2003) has superior power to the AR test, see Moreira (2003) and Andrews, Moreira, and Stock (2004). In such models, the CLR test also has higher power than the rank tests introduced here unless the errors are thick-tailed. Nevertheless, there are numerous applications in the natural and randomized experiments literature with exactly-identified models—typically with one IV and one endogenous regressor. For example, all the empirical papers referenced above include such model specifications. For exactly-identified models, the rank tests considered here are the most powerful that are available.

Under weak and strong IV asymptotics, we show that the rank statistics are as-

[2] The weak IV asymptotic result can be shown using the finite sample results in Moriera (2001).

ymptotically non-central chi-squared with the same non-centrality parameter as the AR statistic up to a scalar constant. This constant is the same as arises with rank tests for many other testing problems, such as two-sample problems, and with rank estimators for location and regression models. For the normal scores rank test, the noncentrality parameter is at least as large as that of the AR statistic for any symmetric error distribution and equals it for the normal distribution. Our asymptotic results make use of asymptotic results of Koul (1970) and Hájek and Sidák (1967) for general rank statistics.

We carry out some Monte Carlo power comparisons of the AR, normal scores rank, and Wilcoxon rank tests. The results indicate that the normal scores rank test essentially dominates the AR test. Its power is essentially the same as that of the AR test for symmetric non-thick-tailed distributions, slightly higher for asymmetric non-thick-tailed distributions, and considerably higher for thick-tailed distributions. The Wilcoxon rank test has power that is quite similar to that of the normal scores rank test, but is somewhat more powerful for thick-tailed distributions and a bit less powerful for non-thick-tailed distributions. The comparative power performance of the three tests is remarkably similar over different sample sizes, strengths of IVs, correlations between the errors, and numbers of IVs. Given that the normal scores test dominates the AR test in terms of power, we prefer the normal scores test to the Wilcoxon test.

The rank tests introduced above share several useful finite-sample robustness properties that the AR test enjoys. These include robustness to excluded IVs and to the specification of a model for the endogenous variables, see Dufour (2003).

The exact rank tests introduced here can be used to construct exact confidence intervals (CIs). The rank tests introduced here also yield conservative tests for subsets of the parameters on the endogenous regressors and covariates via the projection method, see Dufour and Jasiak (1991).

For the case of a simple regression model, IV rank tests have been discussed by Bekker (2002). However, Bekker (2002) does not analyze the power properties of the rank tests and does not allow for covariates in the model. Dealing with covariates with rank tests is more difficult than with the AR test. Theil (1950) considers a rank-based method of constructing CIs in the model considered here. His method delivers conservative CIs and is quite different from the method considered here.

Rosenbaum (1996, 2002), Greevy, Silber, Cnaan, and Rosenbaum (2004), and Imbens and Rosenbaum (2005) consider rank tests that are similar to the rank tests considered, but are based on randomization inference. The probabilistic set-up considered in these papers takes the IVs to be randomized and every quantity that does not depend on the randomized IVs to be fixed. In this context, the tests are exact. In contrast, the present paper considers inference based on a population model, which is typical in econometrics, and shows that the tests are exact given certain conditions on the model. When the same test statistic is considered, the two approaches yield the same asymptotic critical values, but different finite sample critical values. (For example, population model critical values depend on the IVs, whereas those based on randomization inference do not). Our population model approach allows us to compare the power of rank tests with typical tests in the econometric literature such as the AR test. No theoretical power results are given in the randomization inference papers referenced above. We view our results to be complementary to those based on randomization inference.

In ongoing research, the first author and Gustavo Soares are pursuing a rank analogue of the CLR test for over-identified models based on the rank tests introduced here. Such tests are not exact.

There is a huge literature on rank tests in statistics, e.g., see Hájek and Sidák (1967) and Hettmansperger (1984). For a review of rank tests econometrics, see Koenker (1996). An alternative approach to aligning rank tests for dealing with covariates is to use regression rank scores, see Gutenbrunner and Jurečková (1992). We do not pursue this approach here because aligned rank tests are simpler and have comparable theoretical properties.

One could construct M-estimator versions of the AR test, but such tests would have the following drawbacks: (i) their overall asymptotic power properties for non-normal errors would not be as good as for rank tests—just as in the standard regression model, (ii) their asymptotic power for normal errors would be less than that of the AR and normal scores rank tests, (iii) their size would not be exact, and (iv) they would require simultaneous estimation of scale, which would require iterative computational methods.

The paper is organized as follows. Section 2 considers aligned rank tests for models with covariates that need not be categorical. Section 3 considers within-group rank tests for models with categorical covariates. Section 4 presents Monte Carlo power results. An Appendix contains proofs.

2 IV Regression with Covariates

2.1 IV Regression Model

We consider the following linear IV regression model:

$$y_{1i} = \alpha + y_{2i}'\beta + X_i'\theta + u_i \tag{2.1}$$

for $i = 1, ..., n$, where $y_{1i} \in R$, $y_{2i} \in R^{\ell}$, and $X_i \in R^d$ are observed dependent, endogenous regressor, and covariate variables, respectively, $\alpha \in R$, $\beta \in R^{\ell}$, and $\theta \in R^d$ are unknown parameters, and u_i is an unobserved scalar error. We also observe a k-vector of IVs Z_i (that does not include elements of X_i or a constant).

The hypotheses of interest are:

$$H_0 : \beta = \beta_0 \text{ and } H_1 : \beta \neq \beta_0 \text{ for some } \beta_0 \in R^{\ell}. \tag{2.2}$$

Assumption 1. $\{(u_i, X_i) : i \geq 1\}$ are iid.

Assumption 2. $\{Z_i : i \geq 1\}$ is a fixed sequence of k-vectors.

In place of Assumption 2, one could treat the IVs as random. In this case, the IVs would be assumed to be independent of the errors and covariates. As is, Assumption

2 is consistent with random IVs provided one conditions on the IVs. Assumptions 1 and 2 are violated if the distribution of either u_i or X_i depends on the IV vector Z_i. This is a strong assumption concerning the exogeneity of the IVs.

Assumptions 1 and 2 allow for correlation between the endogenous regressor y_{2i} and the error u_i. Assumptions 1 and 2 place no restrictions on the dependence between the endogenous regressor y_{2i} and the IV Z_i. The tests and CIs introduced here have correct size and coverage probability even if the distribution of y_{2i} does not depend on Z_i. Of course, the power of the tests and the lengths of the CIs depend on whether y_{2i} and Z_i are related.

Assumptions 1 and 2 allow for the distribution of u_i to depend on that of X_i. Hence, arbitrary forms of heteroskedasticity are allowed. In fact, Assumptions 1 and 2 even allow for correlation between u_i and X_i.

2.2 Aligned Rank IV Tests and CIs

The rank statistics that we consider are based on a sample covariance k-vector:

$$S_n = n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n) \varphi(R_i/(n+1)), \tag{2.3}$$

where R_i is the rank of $y_{1i} - y_{2i}'\beta_0 - X_i'\widehat{\theta}_n$ in $\{y_{11} - y_{21}'\beta_0 - X_1'\widehat{\theta}_n, ..., y_{1n} - y_{2n}'\beta_0 - X_1'\widehat{\theta}_n\}$, $\widehat{\theta}_n$ is a null-restricted estimator of θ, $\overline{Z}_n = n^{-1} \sum_{i=1}^{n} Z_i$, and $\varphi : [0,1) \to R$ is a non-stochastic score function.[3] The ranks $\{R_i : i \leq n\}$ are referred to as *aligned* ranks due to the aligning by the term $X_i'\widehat{\theta}_n$. We consider the null-restricted least squares (LS) estimator of θ:

$$\widehat{\theta}_n = \left(\sum_{i=1}^{n} (X_i - \overline{X}_n)(X_i - \overline{X}_n)' \right)^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)(y_{1i} - y_{2i}'\beta_0). \tag{2.4}$$

Estimators other than the LS estimator could be considered, but the LS estimator is convenient because it is easy to compute.

Different score functions $\varphi : (0,1) \to R$ yield different rank statistics. The two of greatest interest are the normal (or van der Waerden) score function and the Wilcoxon score function:

$$\varphi^{NS}(x) = \Phi^{-1}(x) \text{ and } \varphi^{WS}(x) = x, \tag{2.5}$$

where $\Phi^{-1}(\cdot)$ is the inverse standard normal distribution function (df).

The rank test statistic, B_n, is a quadratic form in S_n:

$$\begin{aligned} B_n &= n S_n{}' W_n S_n, \text{ where} \\ W_n &= \left(n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n)(Z_i - \overline{Z}_n)' \cdot \int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx \right)^{-1} \end{aligned} \tag{2.6}$$

[3] If there are any ties in the ranks, then we determine a unique ranking by randomization. For example, if $y_{1i} - y_{2i}'\beta_0 - X_i'\widehat{\theta}_n = y_{1j} - y_{2j}'\beta_0 - X_j'\widehat{\theta}_n$ for some $i \neq j$ and these observations are the ℓ-th largest in the sample, then $R_i = \ell$ with probability 0.5, $R_i = \ell + 1$ with probability 0.5, $R_j = \ell + 1$ if $R_i = \ell$, and $R_j = \ell$ if $R_i = \ell + 1$. Ties only occur with positive probability if F is not absolutely continuous. In consequence, in practice one is not likely to have to deal with ties very often.

and $\overline{\varphi} = \int_0^1 \varphi(x)dx$. For normal scores, the statistic B_n^{NS} is

$$B_n^{NS} = \left(\sum_{i=1}^n (Z_i - \overline{Z}_n) \Phi^{-1} \left(\frac{R_i}{n+1} \right) \right)' \left(\sum_{i=1}^n (Z_i - \overline{Z}_n)(Z_i - \overline{Z}_n)' \right)^{-1} \times \left(\sum_{i=1}^n (Z_i - \overline{Z}_n) \Phi^{-1} \left(\frac{R_i}{n+1} \right) \right). \quad (2.7)$$

For Wilcoxon scores, the definition of B_n^{WS} is the same but with the multiplicative constant 12 added and with $\Phi^{-1}(\cdot)$ deleted.[4] In contrast to alternative statistics, such as the AR, LM, and LR statistics, the rank statistic B_n does not require any error variance estimation.

The rank test rejects H_0 if B_n exceeds a critical value c_τ, defined below. The intuition behind the test is as follows. If the null hypothesis is true, $\{Z_i - \overline{Z}_n : i \leq n\}$ are not be related to the ranks $\{R_i : i \leq n\}$ because the ranks depend on $y_{1i} - y_{2i}'\beta_0 - X_i'\widehat{\theta}_n = u_i - X_i'(\widehat{\theta}_n - \theta)$ and the distribution of (u_i, X_i) does not depend on the IVs. Hence, S_n should be close to the zero vector under H_0. On the other hand, under the alternative, if $\{Z_i - \overline{Z}_n : i \leq n\}$ are related to $\{y_{2i} : i \leq n\}$, then $\{Z_i - \overline{Z}_n : i \leq n\}$ are related to $\{u_i + y_{2i}'(\beta - \beta_0) - X_i'(\widehat{\theta}_n - \theta) : i \leq n\}$ and to their scored ranks $\{\varphi(R_i/(n+1)) : i \leq n\}$. (Here β denotes the true value of the parameter.) In this case, the test will have power greater than its size under H_1.

Under H_0, we have

$$\begin{aligned} \eta_i &= y_{1i} - y_{2i}'\beta_0 - X_i'\widehat{\theta}_n \\ &= \alpha + u_i - X_i' \left(\sum_{j=1}^n (X_j - \overline{X}_n)(X_j - \overline{X}_n)' \right)^{-1} \sum_{j=1}^n (X_j - \overline{X}_n) u_j \end{aligned} \quad (2.8)$$

and $\{\eta_i : i \leq n\}$ are exchangeable. The ranks of exchangeable random variables have the same distribution as the ranks of iid random variables because the probability of the ranks taking on any given vector is the same for all vectors and, hence, equals $1/n!$. This leads to the following result.

Theorem 1 *Suppose Assumptions* 1 *and* 2 *hold. Then, under the null hypothesis, the distribution of* B_n *does not depend on* α, θ, β_0, *the distribution of* (u_i, X_i), *or the distribution of the endogenous variables* $\{y_{2i} : i \leq n\}$*. The null distribution of* B_n *is the same when covariates* X_i *appear in the model and the ranks are aligned as in the same model but with no covariates* X_i *and no aligning of the ranks.*

Comments. 1. Theorem 1 indicates that the test statistic B_n is exactly pivotal under H_0 (and, hence, yields a similar test) for any underlying distribution of (u_i, X_i).

[4] The scalar constant $\int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx$ in the definition of the weight matrix W_n of the statistic B_n is a convenient normalization because with this constant included B_n has a χ_k^2 distribution under the null hypothesis, see Section 2.3 below. If desired, this constant can be omitted when exact finite sample critical values or p-values are employed.

Hence, the null behavior of the statistic is completely robust to thick-tailed, thin-tailed, and skewed errors. In contrast, the AR statistic is exactly pivotal under H_0 only with homoskedastic normally distributed errors. The CLR test is only asymptotically pivotal under H_0 and for this it requires finite variance errors.

2. The statistic B_n is exactly pivotal under H_0 without any requirement on how the endogenous variables $\{y_{2i} : i \leq n\}$ are related to the IVs $\{Z_i : i \geq 1\}$. They could be unrelated or related in a linear or nonlinear way.

3. The conditional distribution of B_n given X_i and the fixed IVs is not pivotal. Whether one considers this a drawback is a philosophical issue. In any event, the conditional distribution is asymptotically pivotal.

4. Under the assumptions, aligning of the ranks is not necessary for the statistic B_n to be exactly pivotal under H_0. But, if the ranks are not aligned the power of the test typically suffers, see below.

The significance level τ rank test based on B_n rejects H_0 if

$$B_n > c_\tau, \tag{2.9}$$

where c_τ is chosen so that the test has significance level $\tau \in (0, 1)$.[5] When the observed test statistic takes the value b_{ob}, the exact p-value, p, of the test is defined by $P(B_n > b_{ob}) = p$.

The exact critical value, c_τ, and p-value, p, depend on the IVs, $\{Z_i : i \leq n\}$, and, hence, need to be generated on a case by case basis. This can be done easily and quickly by simulation. One simulates n iid uniform (0,1) random variables, say $\{u_{ri} : i = 1, ..., n\}$, and calculates

$$B_{nr} = nS'_{nr}W_nS_{nr}, \text{ where } S_{nr} = n^{-1}\sum_{i=1}^{n}(Z_i - \overline{Z}_n)\varphi(R_{ri}/(n+1)) \tag{2.10}$$

and R_{ri} is the rank of u_{ri} among $\{u_{r1}, ..., u_{rn}\}$.[6] One repeats this for $r = 1, ..., R_S$. The simulated critical value $c_{sim,\tau}$ is the $1-\tau$ sample quantile of $\{B_{nr} : r = 1, ..., R_S\}$. The simulated p-value is $p = R_S^{-1}\sum_{r=1}^{R_S} 1(B_{nr} > b_{ob})$.

The matrix programming languages GAUSS and Matlab have very fast built-in procedures for finding the ranks of a given vector. For example, the GAUSS procedure *rankindx* can compute a critical value using 40,000 simulation repetitions in a matter of seconds for sample sizes n up to 500 and numbers of IVs k up to 10 using a typical PC. The computation time increases with n roughly proportionally and much less than proportionally in k. Hence, even for data sets with sample sizes in the thousands, computation of critical values is fast and accurate.

---

[5] Because B_n has a discrete distribution, it may not be possible to find c_τ such that the test has exact significance level τ for arbitrary values of τ. In practice, this is not a serious problem because the discrete distribution of B_n is very nearly continuous for values of n that typically arise in practice. The probability of any given value of B_n is $1/n!$.

[6] Because the distribution of the ranks under H_0 does not depend on F, we can compute the critical value by simulating from any distribution that is convenient, such as uniform $(0, 1)$.

We construct exactly distribution-free CIs (or confidence regions if $\ell > 1$) for β by inverting the test statistic B_n. For clarity, we write the rank statistic B_n for testing $H_0 : \beta = \beta_0$ as $B_n(\beta_0)$. The CI is given by

$$CI_{n,1-\tau} = \{\beta_0 : B_n(\beta_0) \leq c_\tau\}. \tag{2.11}$$

Because the critical value c_τ does not depend on β_0, one does not have to compute a new critical value for each value of β_0. To compute $CI_{n,1-\tau}$, one just needs to compute $B_n(\beta_0)$ for a grid of values β_0 and compute c_τ once.

Rank tests for testing $H_0 : \beta = \beta_0$ also apply to the nonlinear model:

$$g(y_{1i}, y_{2i}, \beta) + \alpha + X_i'\theta = u_i, \tag{2.12}$$

where $g(\cdot,\cdot,\cdot)$ is a known function. In this case, $\{R_i : i \leq n\}$ are the ranks of $\{g(y_{1i}, y_{2i}, \beta_0) + X_i'\widehat{\theta}_n : i \leq n\}$ and $\widehat{\theta}_n$ is defined as in (2.4) but with $y_{1i} - y_{2i}'\beta_0$ replaced by $g(y_{1i}, y_{2i}, \beta_0)$. Otherwise, B_n and its critical value or p-value are the same as above. Theorem 1 holds with $y_{1i} - y_{2i}'\beta_0 - X_i'\widehat{\theta}_n$ replaced by $g(y_{1i}, y_{2i}, \beta_0) + X_i'\widehat{\theta}_n$.

2.3 Asymptotic Power of Aligned Rank IV Tests

In this section, we determine the asymptotic power of the B_n rank test and compare it to that of the AR test. We consider two asymptotic frameworks. One consists of $1/n^{1/2}$ local alternative parameter values coupled with strong IVs, which is the standard asymptotic set-up. The other consists of fixed alternatives coupled with weak IVs, which is the weak IV set-up of Staiger and Stock (1997).

The score function φ is required to satisfy the following mild assumption.

Assumption 3. **(a)** $\varphi : (0, 1) \rightarrow R$ is absolutely continuous and bounded with two derivatives that exist almost everywhere and are bounded.
(b) $\int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx > 0$.

Assumption 3 holds for normal scores with $\int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx = 1$. It holds for Wilcoxon scores with $\int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx = 1/12$. Assumption 3(b) holds provided $\varphi(x)$ is not constant almost everywhere on $[0, 1)$.

Next, we state the assumptions concerning the IVs $\{Z_i : i \geq 1\}$.

Assumption 4. **(a)** $n^{-1}\sum_{i=1}^n (Z_i - \overline{Z}_n)(Z_i - \overline{Z}_n)' \rightarrow \Sigma_Z$ pd as $n \rightarrow \infty$.
(b) $\max_{1 \leq i \leq n} ||Z_i - \overline{Z}_n||^2/n \rightarrow 0$ as $n \rightarrow \infty$.
(c) $\sum_{i=1}^\infty ||Z_i - \overline{Z}_n||^{1+\delta}/i^{1+\delta} < \infty$ for some $\delta > 0$.

Assumption 4 holds with probability one if $\{Z_i : i \geq 1\}$ is a realization of an iid sequence with pd variance matrix and $2 + \delta$ moments finite for some $\delta > 0$, see Lemma 2 of the Appendix. Hence, Assumption 4 is not very restrictive.

We assume that the covariates $\{X_i : i \geq 1\}$ satisfy:

Assumption 5. **(a)** $E||X_i||^{2+\delta} < \infty$ for some $\delta > 0$.
(b) $\Sigma_X = E(X_i - EX_i)(X_i - EX_i)'$ is pd.

Next we state the assumptions concerning the alternative data generating process.

Assumption 6. **(a)** $y_{1i} = \alpha + y_{2i}'\beta_n + X_i'\theta + u_i$ for $i \geq 1$, where $\beta_n \in R^\ell$ is a constant for $n \geq 1$.

(b) $y_{2i} = \mu + \pi_n Z_i + \Lambda X_i + v_i$ for $i \geq 1$, where π_n is an $\ell \times k$ matrix of constants for $n \geq 1$, μ is an ℓ-vector of constants, Λ is an $\ell \times d$ matrix of constants, and v_i is an ℓ-vector of random variables.

(c) $\{u_i : i \geq 1\}$ are independent of $\{X_i : i \geq 1\}$, and $Eu_i^2 < \infty$.

The parameter π_n indexes the strength of the IVs relation to the endogenous regressors. The difference $\beta_n - \beta_0$ indexes the distance of the alternative from the null. These parameters differ in the weak and strong IV cases, as specified below.

Let $I(f)$ denote Fisher's information of an absolutely-continuous density f. That is, $I(f) = \int [f'(x)/f(x)]^2 f(x)dx$, where f' denotes the derivative of f.

For weak IVs, we consider fixed alternatives and π_n that is local to zero.

Assumption 7W. **(a)** $\beta_n = \beta_0 + \gamma$ for some $\gamma \in R^\ell$.

(b) $\pi_n = C/n^{1/2}$ for some $\ell \times k$-matrix of constants C.

(c) $\{(v_i, u_i) : i \geq 1\}$ are iid and independent of $\{X_i : i \geq 1\}$, and $E\|v_i\|^2 < \infty$.

(d) $v_i'\gamma + u_i$ has an absolutely-continuous strictly-increasing df G and an absolutely-continuous and bounded density g that satisfies $I(g) < \infty$.

For strong IVs, we consider local alternatives and a fixed value of π_n.

Assumption 7S. **(a)** $\beta_n = \beta_0 + \gamma/n^{1/2}$ for some $\gamma \in R^\ell$.

(b) $\pi_n = \pi$ for all n for some $\ell \times k$-matrix of constants π.

(c) $v_i = \varepsilon_i + \rho u_i$ for $i \geq 1$, where ε_i is a random ℓ-vector and $\rho \in R^\ell$ is a vector of constants.

(d) $\{\varepsilon_i : i \geq 1\}$ are iid and independent of $\{u_i : i \geq 1\}$, and $0 < E\|\varepsilon_i\|^{2+\delta} < \infty$ for some $\delta > 0$.

(e) u_i has an absolutely-continuous strictly-increasing df F and an absolutely-continuous and bounded density f that satisfies $I(f) < \infty$.

Assumption 7W allows for arbitrary dependence between v_i and u_i. Assumption 7S allows for arbitrary dependence between X_i and ε_i. Assumption 7S(c) and 7S(d) make explicit the form of the dependence between the main equation error u_i and the reduced form error v_i. This facilitates the determination of the asymptotic non-null properties of B_n.

For a score function φ and a density f, define

$$\xi(\varphi, f) = \frac{\left(\int_0^1 \varphi(x)\varphi(x,f)dx\right)^2}{\int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx}, \text{ where } \varphi(x,f) = -\frac{f'(F^{-1}(x))}{f(F^{-1}(x))} \tag{2.1}$$

for $x \in (0,1)$. For normal and Wilcoxon scores,

$$\xi(\varphi^{NS}, f) = \left(\int \frac{f^2(x)}{\phi(\Phi^{-1}(F(x)))}dx\right)^2 \text{ and } \xi(\varphi^{WS}, f) = 12\left(\int f^2(x)dx\right)^2. \tag{2.2}$$

where ϕ and Φ denote the standard normal density and df, respectively, and $F' = f$.

Let $\chi^2_k(\delta)$ denote a noncentral chi-squared distribution with k degrees of freedom and noncentrality parameter δ.

The following theorem establishes the asymptotic distribution of B_n in the weak IV/fixed alternative and strong IV/local alternative scenarios.

Theorem 2 (**a**) *Under Assumptions* 1-6 *and* 7W,

$$B_n \to_d \chi^2_k(\delta_W), \ \textit{where } \delta_W = \gamma' C \Sigma_Z C' \gamma \xi(\varphi, g).$$

(**b**) *Under Assumptions* 1-6 *and* 7S,

$$B_n \to_d \chi^2_k(\delta_S), \ \textit{where } \delta_S = \gamma' \pi \Sigma_Z \pi' \gamma \xi(\varphi, f).$$

Comments. 1. The results of Theorem 2 show that the statistic B_n, which is based on aligned ranks using the estimator $\widehat{\theta}_n$, has the same asymptotic distribution as when the true value θ is used in place of $\widehat{\theta}_n$.

2. The results of the Theorem continue to hold when the restricted LS estimator $\widehat{\theta}_n$ is replaced by any estimator θ^*_n that satisfies $n^{1/2}(\theta^*_n - \theta - \Lambda'(\beta_n - \beta_0)) = O_p(1)$.[7]

3. If the statistic B_n is constructed without aligning the ranks, then its asymptotic distribution is given by Theorem 2, but with g and f being the densities of $X_i'\theta_0 + v_i'\gamma + u_i$ and $X_i'\theta_0 + u_i$, respectively. Typically this increases the constants $\xi(\varphi, g)$ and $\xi(\varphi, f)$ because the addition of $X_i'\theta_0$ increases the dispersion of the random variables. Note that $\xi(\varphi, \sigma^{-1} f(\cdot\sigma^{-1})) = \sigma^{-2}\xi(\varphi, f)$ for all f, see Hájek and Sidák (1967, Lemma I.2.4e, p. 21). Hence, a scale increase by σ reduces the noncentrality parameter by the factor σ^{-2}. For example, if X_i, v_i, and u_i are jointly normal and the addition of $X_i'\theta_0$ doubles the variance, then the noncentrality parameter is reduced by a factor of two. This multiplicative effect is the same for the AR and rank tests. In sum, aligning the ranks typically increases the power of tests.

For the AR statistic, $AR_n \times k \to_d \chi^2_k(\delta^{AR}_W)$ and $AR_n \times k \to_d \chi^2_k(\delta^{AR}_S)$ under weak and strong IV asymptotics, respectively, where

$$\begin{aligned} \delta^{AR}_W &= C'\Sigma_Z C \gamma^2 / \sigma^2_g, \\ \delta^{AR}_S &= \pi'\Sigma_Z \pi \gamma^2 / \sigma^2_f, \end{aligned} \tag{2.3}$$

and σ^2_g and σ^2_f denote the variances corresponding to the densities g and f, respectively, under the assumptions above plus $Eu_i^2 < \infty$.

Hence, the noncentrality parameters of the rank IV tests can be compared to those of the AR test by comparing $\xi(\varphi, g)$ to $1/\sigma^2_g$ for weak IVs and $\xi(\varphi, f)$ to $1/\sigma^2_f$ for strong IVs. Specifically, the *asymptotic relative efficiency* (ARE) of the rank IV test to the AR test is given by

$$\begin{aligned} ARE_f(B_n, AR) &= \xi(\varphi, g)\sigma^2_g \text{ for weak IVs and} \\ ARE_f(B_n, AR) &= \xi(\varphi, f)\sigma^2_f \text{ for strong IVs.} \end{aligned} \tag{2.4}$$

[7] The term $\Lambda'(\beta_n - \beta_0)$ arises here because $y_{1i} - y_{2i}'\beta_0 = \alpha + y_{2i}'(\beta_n - \beta_0) + X_i'\theta + u_i = \alpha + Z_{2i}'\pi_n'(\beta_n - \beta_0) + X_i'(\theta + \Lambda'(\beta_n - \beta_0)) + v_i'(\beta_n - \beta_0) + u_i$ by Assumption 6.

(An ARE greater than one means that the rank IV test has higher power than the AR test.)

Comparisons of this type have been considered extensively in the literature because they are exactly the same comparisons that arise when computing the ARE of a rank test compared to the usual t-test in a simple location model with error density g or f. They are also the same as the comparisons that arise when comparing the ARE of a rank estimator with the sample mean in the location model. Note that the ARE's considered here are all independent of the location and scale of g or f.

For normal scores, $\varphi^{NS}(x) = \Phi^{-1}(x)$, the ARE is

$$ARE_f(NS, AR) = \sigma^2(f) \left(\int \frac{f^2(x)}{\phi(\Phi^{-1}(F(x)))} dx \right)^2. \qquad (2.5)$$

A result due to Chernoff and Savage states that $ARE_f(NS, AR) \geq 1$ for all symmetric distributions f (about some point not necessarily zero), see Hettmansperger (1984, Thm. 2.9.2, p. 110). Hence, the asymptotic power of the normal scores rank IV test is greater than or equal to that of the AR test for any symmetric distribution for weak or strong IVs.

For Wilcoxon scores, $\varphi^{WS}(x) = x$ and a density f, the ARE of the rank IV test to the AR test is

$$ARE_f(WS, AR) = 12\sigma_f^2 \left(\int f^2(x)dx \right)^2. \qquad (2.6)$$

For the normal distribution ϕ, $ARE_\phi(WS, AR) = .955$. For the double exponential distribution f_{de}, $ARE_{f_{de}}(WS, AR) = 1.50$. For a contaminated normal distribution $f_\varepsilon(x) = (1-\varepsilon)\phi(x) + \varepsilon\phi(x/3)/3$, $ARE_{f_\varepsilon}(WS, AR) = 1.196$, 1.373, and 1.497 for $\varepsilon = .05$, .10, and .15, respectively, see Hettmansperger (1984, pp. 71-2). A result due to Hodges and Lehmann states that $ARE_f(WS, AR) \geq .864$ for all symmetric distributions f (about some point not necessarily zero), see Hettmansperger (1984, Thm. 2.6.3, p. 72). Hence, the noncentrality parameter of the Wilcoxon scores rank IV test is almost as large as that of the AR test for the normal distribution, is significantly larger than that of the AR test for heavier tailed distributions, and is not much smaller for any symmetric distribution.

For any densities f_1 and f_2 symmetric about zero, $ARE_{f_1}(WS, NS) \leq ARE_{f_2}(WS, NS)$ whenever the tails of f_1 are lighter than the tails of f_2 in the sense that $F_2^{-1}(F_1(x))$ is convex for $x \geq 0$, see Thm. 2.9.5 of Hettmansperger (1984, p. 116). Hence, the comparative power of Wilcoxon scores to normal scores tests increases as the tail thickness of the distribution increases. For any symmetric density f, $ARE_f(WS, NS) \in (0, 1.91)$, see Hettsmansperger (1984, Thm. 2.9.3, p. 115).

3 IV Regression with Categorical Covariates

3.1 Model and Test

In this section, we consider a regression model with categorical covariates. In contrast to the model in Section 2, the covariates and IVs may be related. The model is

$$y_{1i} = D_i'\alpha + y_{2i}'\beta + u_i \tag{3.1}$$

for $i = 1, \ldots, n$, where y_{1i} is an observed scalar dependent variable, y_{2i} is an observed ℓ-vector of endogenous variables, $D_i = (D_{i1}, \ldots, D_{iJ})'$ is an observed J-vector of dummy variables, and $\alpha = (\alpha_1, \ldots, \alpha_J)'$ and $\beta \in R^\ell$ are unknown parameters. We also observe a k-vector of IVs Z_i (that does not include elements of D_i or a constant). The dummy variable D_{ij} equals 1 if observation i is in group j; otherwise, it equals 0. We assume that $\sum_{j=1}^J D_{ij} = 1$ for all $i = 1, \ldots, n$.

The basic assumptions of the model are:

Assumption C1. $\{u_i : i \geq 1\}$ are independent random variables with $u_i \sim F_j$ when $D_{ij} = 1$ for some df's $\{F_j : j = 1, \ldots, J\}$.

Assumption C2. $\{D_i : i \geq 1\}$ is a fixed sequence of J-vectors.

Assumption C1 allows for different error distributions across the J groups.

In addition, we assume that Assumption 2 holds, i.e., the IVs $\{Z_i : i \geq 1\}$ are fixed k-vectors. As above, random IVs can be treated by conditioning on the IVs. In this case, the distribution of the IVs can differ across covariate categories and, hence, the IVs and covariates can be related.

We want to test $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$ while leaving α unspecified. Or, more generally, we might be interested in the alternative hypothesis where β may differ across groups and is different from β_0 in at least one group. The distribution of a test statistic based on the ranks of $y_{1i} - y_{2i}'\beta_0$ in the entire sample depends on the nuisance parameters $(\alpha_1, \ldots, \alpha_J)$, which is problematic. However, one can divide the sample into the J homogeneous sub-samples and use the ranks within the sub-samples to achieve invariance with respect to the nuisance parameters. This approach was used in the two sample location problem (without IVs) by van Elteren (1960).

It is convenient to rewrite the model in (3.1) as follows. Let n_j be the size of group j (i.e., $n_j = \sum_{i=1}^n D_{ij}$). As defined, $\sum_{j=1}^J n_j = n$. Next, define $y_{1,ij}$, $y_{2,ij}$, and Z_{ij} to be the dependent, endogenous regressor, and instrumental variables, respectively, that belong to group j for $i = 1, ..., n_j$, for $j = 1, ..., J$. Then, the model in (3.1) can be rewritten as

$$y_{1,ij} = \alpha_j + y_{2,ij}'\beta + u_{ij}. \tag{3.2}$$

Let R_{ij} denote the rank of $y_{1,ij} - y_{2,ij}'\beta_0$ in $\{y_{1,1j} - y_{2,1j}'\beta_0, \ldots, y_{1,n_jj} - y_{2,n_jj}'\beta_0\}$ for $j = 1, \ldots, J$.

We introduce the following test statistic:

$$\begin{aligned}
B_{Cn} &= nS_{Cn}{}'W_{Cn}S_{Cn}, \text{ where} \\
S_{Cn} &= \sum_{j=1}^J w_{nj}S_{Cnj},\ S_{Cnj} = n^{-1}\sum_{i=1}^{n_j}(Z_{ij} - \overline{Z}_{nj})\varphi(R_{ij}/(n_j+1)),\ \overline{Z}_{nj} = n_j^{-1}\sum_{i=1}^{n_j} Z_{ij}, \\
W_{Cn} &= \left(n^{-1}\sum_{j=1}^J\sum_{i=1}^{n_j} w_{nj}^2(Z_{ij} - \overline{Z}_{nj})(Z_{ij} - \overline{Z}_{nj})'\int_0^1[\varphi(x) - \overline{\varphi}]^2dx\right)^{-1},
\end{aligned} \tag{3.3}$$

where $\{w_{n1}, \ldots, w_{nJ}\}$ are non-random weights assigned to the J groups. (The weights may depend on n and at least one weight must be non-zero.) Optimal weighting schemes are discussed in Section 3.2 below. In general, we recommend using constant weights: $w_{nj} = 1$ for all $j = 1, \ldots, J$.

The ranks $\{R_{ij} : i = 1, \ldots, n_j\}$ are the ranks of $\{\alpha_j + y'_{2,ij}(\beta - \beta_0) + u_{ij} : i = 1, \ldots, n_j\}$ (where β denotes the true value). Since ranks are invariant under location shifts and we rank the errors within the homogenous location groups, the ranks are not affected by the unknown nuisance parameters $(\alpha_1, \ldots, \alpha_J)$. This holds under H_0 and H_1.

In consequence, if the null hypothesis is true, $\{R_{ij} : i = 1, \ldots, n_j\}$ equal the ranks of $\{u_{ij} : i = 1, \ldots, n_j\}$ for each j. Hence, the null distributions of $S_{Cn1}, \ldots, S_{CnJ}$, and S_{Cn} do not depend on $(\alpha_1, ..., \alpha_J)$, β_0, or the distribution of $\{y_{2,ij} : i \leq n_j, j = 1, ..., J\}$. Furthermore, Assumptions C1 and C2 and randomization in the case of ties in ranks, combined with the exchangeability argument given in Section 2, imply that the distributions of $S_{Cn1}, \ldots, S_{CnJ}$, and S_{Cn} do not depend on $\{F_1, \ldots, F_J\}$ under H_0. The null distributions of these statistics do depend on the IVs and the group structure, but both of these are observed.

The following analogue of Theorem 1 holds.

Theorem 3 *Suppose Assumptions* C1, C2, *and* 2 *hold. Then, under* H_0, *the distribution of* B_{Cn}, *defined in* (3.3), *does not depend on* $\{\alpha_j : j = 1, ..., J\}$, β_0, $\{F_j : j = 1, ..., J\}$, *or the distribution of the endogenous variables* $\{y_{2,i} : i \leq n_j, j = 1, ..., J\}$.

One rejects the null if B_{Cn} is sufficiently large. The desired exact critical value can be calculated by simulation. First, one simulates n iid uniform (0,1) random variables, say $\{u_{r,ij} : i = 1, \ldots, n_j, j = 1, \ldots, J\}$ and calculates

$$S_{Cn,r} = \sum_{j=1}^{J} w_{nj} n^{-1} \sum_{i=1}^{n_j} (Z_{ij} - \overline{Z}_{nj}) \varphi(R_{r,ij}/(n_j + 1)), \tag{3.4}$$

where $R_{r,ij}$ is the rank of $u_{r,ij}$ among $\{u_{r,1j}, \ldots, u_{r,n_j j}\}$ for $j = 1, \ldots, J$. Next, one computes $B_{Cn,r}$ and repeats the process for $r = 1, \ldots, R_S$. The critical value for significance level τ is the $1 - \tau$ sample quantile of $\{B_{Cn,r} : r = 1, \ldots, R_S\}$. Given an observed value, b_{ob}, of the test statistic, the p-value is $p = R_S^{-1} \sum_{r=1}^{R_S} 1(B_{Cn,r} > b_{ob})$.

As in (2.11), CIs can be constructed by inverting the test based on B_{Cn}.

The B_{Cn} test for $H_0 : \beta = \beta_0$ generalizes to nonlinear models of the form $g_j(y_{1,ij}, y_{2,ij}, \beta) + \alpha_j = u_{ij}$. For this model, $\{R_{ij} : i \leq n\}$ are defined to be the ranks of $\{g_j(y_{1i}, y_{2i}, \beta_0) : i \leq n\}$. Otherwise, the test statistic B_{Cn} and its critical value are the same as above.

3.2 Asymptotic Power

In this section, we study the power properties of the B_{Cn} rank test in the model with categorical covariates. As in Section 2.3, we consider fixed alternatives combined with weak IVs and local alternatives combined with strong IVs.

We make the following assumptions concerning the model.

Assumption C3. For all $j = 1, ..., J$,

(a) $n_j^{-1} \sum_{i=1}^{n_j} (Z_{ij} - \overline{Z}_{nj})(Z_{ij} - \overline{Z}_{nj})' \to \Sigma_{Zj}$ pd as $n \to \infty$.

(b) $\max_{1 \le i \le n_j} ||Z_{ij} - \overline{Z}_{nj}||^2 / n_j \to 0$ as $n \to \infty$.

Assumption C4. **(a)** $y_{1,ij} = \alpha_j + y_{2,ij}' \beta_{nj} + u_{ij}$, where $\beta_{nj} \in R^\ell$ is a constant for all n and j.

(b) $y_{2,ij} = \mu_j + \pi_{nj} Z_{ij} + v_{ij}$, where π_{nj} is an $\ell \times k$ matrix of constants, μ_j is an ℓ-vector of constants, and v_{ij} is a random ℓ-vector.

Assumption C5. $\lim_{n \to \infty} n_j / n = b_j > 0$ for all $j = 1, \ldots, J$.

Assumption C5 guarantees that all groups are non-negligible in the limit.

We assume the weights are chosen to satisfy:

Assumption C6. $\lim_{n \to \infty} w_{nj} = w_j$ for all $j = 1, \ldots, J$ for some constants $\{w_j : j = 1, ..., J\}$ at least one of which is non-zero.

For the case of weak IVs, we assume:

Assumption C7W. For all $j = 1, ..., J$,

(a) $\beta_{nj} = \beta_0 + \gamma_j$ for all n for some $\gamma_j \in R^\ell$.

(b) $\pi_{nj} = C_j / n^{1/2}$ for some $C_j \in R^{\ell \times k}$.

(c) (v_{ij}, u_{ij}) are iid across $i = 1, \ldots, n_j$ and independent across $j = 1, \ldots, J$.

(d) $v_{ij}' \gamma_j + u_{ij}$ has absolutely-continuous strictly-increasing df G_j and absolutely-continuous and bounded density g_j that satisfies $I(g_j) < \infty$.

Assumption C7W(a) allows the true β vector to vary across groups, which covers more general alternatives than $H_1 : \beta \neq \beta_0$. For the alternative $H_1 : \beta \neq \beta_0$, one has $\gamma_j = \gamma$ for all $j = 1, ..., J$ for some $\gamma \in R^\ell$. Assumption C7W(b) implies that the correlation between the covariates and the IVs may vary across the groups, but it is of the same order of magnitude for all $j = 1, \ldots, J$. Assumption C7W places no restriction on the dependence between the main equation error u_{ij} and the reduced form error v_{ij}.

For the case of strong IVs, we assume:

Assumption C7S. For all $j = 1, ..., J$,

(a) $\beta_{nj} = \beta_0 + \gamma_j / n^{1/2}$ for some $\gamma_j \in R^\ell$.

(b) $\pi_{nj} = \pi_j$ for all n and some $\pi_j \in R^{\ell \times k}$.

(c) $v_{ij} = \varepsilon_{ij} + \rho_j u_{ij}$, where ε_{ij} is a random ℓ-vector and ρ_j is a constant ℓ-vector.

(d) $\{\varepsilon_{ij} : i \ge 1\}$ are iid across $i \ge 1$, independent across $j = 1, \ldots, J$, and independent of $\{u_{ij} : i \ge 1, j \le J\}$ and $E||\varepsilon_{ij}||^{2+\delta} < \infty$ for some $\delta > 0$.

(e) $\sum_{i=1}^{\infty} ||Z_{ij} - \overline{Z}_{nj}||^{1+\delta} / i^{1+\delta} < \infty$ for some $\delta > 0$.

(f) u_{ij} has an absolutely-continuous strictly-increasing df F_j and an absolutely-continuous and bounded density f_j that satisfies $I(f_j) < \infty$.

Assumption C7S(a) implies that the distance from the null γ_j may vary with j, but its order of magnitude is the same for all groups.

The proof of Theorem 2 can be used to obtain the following:

Corollary 1 (**a**) *Under Assumptions* C1-C6, C7W, 2, *and* 3, $B_{Cn} \to_d \chi_k^2(\delta_W)$, *where*

$$\delta_W = \left\| \left(\sum_{j=1}^{J} w_j^2 b_j \Sigma_{Zj} \right)^{-1/2} \left(\sum_{j=1}^{J} w_j b_j \Sigma_{Zj} C_j' \gamma_j \xi(\varphi, g_j) \right) \right\|^2 .$$

(**b**) *Under Assumptions* C1-C6, C7S, 2, *and* 3, $B_{Cn} \to_d \chi_k^2(\delta_S)$, *where*

$$\delta_S = \left\| \left(\sum_{j=1}^{J} w_j^2 b_j \Sigma_{Zj} \right)^{-1/2} \left(\sum_{j=1}^{J} w_j b_j \Sigma_{Zj} \pi_j' \gamma_j \xi(\varphi, f_j) \right) \right\|^2 .$$

Comments. 1. Corollary 1 shows that the statistic B_{Cn} has a non-central χ_k^2 distribution under the alternative with non-centrality parameter that depends on the weights w_j, the group sizes b_j, the variability of the IVs Σ_{Zj}, the strength of the IVs C_j or π_j, the score function φ, and the density g_j of $v_{ij}'\gamma_j + u_i$ or the density f_j of u_{ij}.

2. When g_j does not depend on j, then $\xi(\varphi, g_j)$ scales out of δ_W and the difference between the noncentrality parameters of the B_{Cn} rank test and an analogous "categorical AR" test is the same as in Section 2.3. Hence, ARE of a categorical rank test versus the categorical AR test is the same as the ARE of the rank and AR tests given in Section 2.3. The same is true under strong IV asymptotics if f_j does not depend on j.

One can compute the asymptotically optimal weights by maximizing δ_W or δ_S with respect to the weights $w_1, \ldots, w_J$. In general, the optimal weights depend on the values of Σ_{Zj}, γ_j, b_j, C_j, and $\xi(\varphi, g_j)$ with weak IVs and Σ_{Zj}, γ_j, b_j, π_j, and $\xi(\varphi, f_j)$ with strong IVs. But, if $k = \ell = 1$, the optimal weights do not depend on Σ_{Zj} or b_j and equal $w_j = C_j'\gamma_j\xi(\varphi, g_j)$ with weak IVs and $w_j = \pi_j'\gamma_j\xi(\varphi, f_j)$ with strong IVs. If these quantities do not depend on j, then equal weights are optimal.

In the absence of information about how the quantities in the previous paragraph vary with j, as is usually the case in practice, it is reasonable to employ a test that is invariant to permutations of these quantities across groups. This leads to taking $w_{nj} = 1$ for all $j = 1, \ldots, J$. This is the weighting scheme that we recommend.

4 Monte Carlo Results

4.1 Experimental Design

In this section, we report simulated power comparisons of the B_n^W, B_n^{NS}, and AR tests. We take the model to be essentially as in Assumption 6 and 7S(c) with β being a scalar ($\ell = 1$):

$$\begin{aligned} y_{1i} &= \alpha + y_{2i}\beta + X_i'\theta + u_i, \\ y_{2i} &= Z_i'\pi + X_i'\Lambda + (1-\rho^2)^{1/2}\varepsilon_i + \rho u_i, \end{aligned} \tag{4.1}$$

for $i = 1, ..., n$, where $Z_i = (Z_{i1}, ..., Z_{ik})'$, $X_i = (X_{i1}, ..., X_{ip})'$, and $Z_{ij}, X_{is}, u_i, \varepsilon_i$ are iid with distribution F for all $j = 1, ..., k$, $s = 1, ..., p$, and $i = 1, ..., n$.

The test statistics considered are invariant with respect to α, θ, Λ, and the location and scale of F. Hence, without loss of generality we take α, θ, and Λ to be zero and we take F to have mean zero (if its mean is well defined), center of symmetry zero (if it is symmetric), and variance one (if its variance is well defined).

The parameter vector π, which determines the strength of the IVs, is taken to be proportional to a k-vector of ones:

$$\pi = \frac{\rho_{IV}}{k^{1/2}(1-\rho_{IV}^2)^{1/2}}(1, ..., 1)' \text{ for some } \rho_{IV} \in [-1, 1], \tag{4.2}$$

where, by construction, ρ_{IV} is the correlation between the reduced form regression function, $Z_i'\pi$, and the endogenous variable y_{2i} (provided F has a finite variance). The parameter ρ_{IV} can be related to a parameter λ which directly measures the strength of the IVs (and is closely related to the so-called concentration parameter):

$$\lambda = \frac{n\rho_{IV}^2}{1-\rho_{IV}^2} = n\pi E Z_i Z_i' \pi \approx \pi' Z' Z \pi, \tag{4.3}$$

where the first equality defines λ, the second equality holds provided Z_i has a finite variance, and $\approx$ means "is approximately equal for large n."

As above, the hypotheses of interest are $H_0 : \beta = \beta_0$ and $H_1 : \beta \neq \beta_0$. The true parameter β is taken so that the AR test with significance level .05 has power around .4 for the given choice of λ, ρ, n, k, p, and $F = \Phi$.

We provide results for selected subsets of the cases for which $n = 50$, 100, 200; $k = 1$, 5, 10; $p = 0$, 5; and F is normal, t_r with r degrees of freedom (df) for $r = 1-10$, difference of independent log-normals (DLN), uniform, absolute value of a normal, logistic, double exponential (DE), and log-normal (LN). The t distributions exhibit heavy tails for small values df (e.g., $r = 1$ yields the Cauchy distribution) as do the DLN and LN distributions and to a lesser extent the DE distribution. The uniform distribution exhibits thin tails. The absolute value of a normal and LN distributions exhibit skewness.

4.2 Power Comparisons

We compare the power of the .05 significance level rank tests B_n^W and B_n^{NS} to the AR test for a variety of cases. We report size-corrected power for the AR test when the errors are non-normal, where the size-correcting critical values are obtained using $10,000$ simulation repetitions. The power results are based on $5,000$ Monte Carlo simulations.

We first consider a Base Case in which $\lambda = 9$, $\rho = .75$, $\beta - \beta_0 > 0$, $n = 100$, $k = 1$, $p = 5$, and F equals the normal, t_1, t_2, t_3, t_{10}, or DLN distribution. This case exhibits moderately weak IVs, moderately strong endogeneity, and exact identification. Then, we consider a number of variations of the Base Case to illustrate the effect of changes in the distribution F, strength of IVs λ, level of endogeneity ρ, sign of $\beta - \beta_0$, sample size n, and number of IVs k.

Table I reports the results. The results of the Base Case show that for the normal distribution the power of the normal scores (NS) test is within simulation error of equaling that of the AR test, whereas the power of the Wilcoxon scores (WS) test is slightly lower. For thick-tailed non-normal distributions, on the other hand, the NS and WS tests are much more powerful than the AR test. For thick-tailed distributions, the NS and WS tests have quite similar power, although that of the WS test is somewhat higher, especially for the DLN distribution. For the t_{10} distribution, which has moderate tails, the NS, WS, and AR tests have similar power.

Case 2 differs from the Base Case only in terms of the distribution F. Case 2 shows that for a t_4 distribution the rank tests have higher power than the AR test, but for a t_6 distribution the three tests have roughly equal power. For the uniform distribution, which has thin tails, the NS and AR tests have essentially equal power, whereas that of the WS test is somewhat lower. For the absolute value of a normal distribution, which is highly skewed, the NS test is somewhat more powerful than the WS and NS tests. The results for the DE distribution are quite similar to those for the t_4 distribution. The rank tests have higher power than the AR test. For the log-normal distribution, which is both skewed and thick-tailed, the rank tests outperform the AR test and the WS test outperforms the NS test.

These results, combined with those of the Base Case, suggest that NS and WS tests have considerably higher power than the AR test for thick-tailed distributions, but the tails have to be quite thick for this advantage to appear. For non-thick-tailed distributions, the NS test has power that is at least as high as that of the AR test and the WS test has power that is equal to or close to the power of the AR test.

Cases 3-11 exhibit power comparisons for variations of the Base Case. The general pattern exhibited in the Base Case, as discussed above, is observed in all of these additional cases to a remarkable degree. Hence, the general pattern is found to be robust to negative deviations $\beta - \beta_0$ (Case 3), strong IVs (Case 4), weak IVs (Case 5), high endogeneity (Case 6), no endogeneity (Case 7), five IVs (Case 8), ten IVs (Case 9), smaller sample size, $n = 50$ (Case 10), and larger sample size, $n = 200$ (Case 11).

To conclude, the power simulations reported above show that the NS rank test, B^{NS}, essentially dominates the AR test in terms of finite sample power. It has much higher power for thick-tailed distributions and essentially equal power (or in some cases slightly higher power) for non-thick-tailed distributions. The WS rank test, B^{W}, has finite sample power quite similar to that of the NS test, but it is slightly more powerful for thick-tailed distributions and often slightly less powerful for non-thick-tailed distributions. Hence, the WS test does not dominate the AR test, but is close to doing so.

5 Appendix of Proofs

The asymptotic results of the paper are proved using the following Lemma. Part (a) of the Lemma is an extension of Theorem 2.1 and Lemma 2.3 of Koul (1970) from scalar constants c_i and d_i to vectors. As Koul (1970, p. 1280) notes, his proof of these results goes through for this extension with virtually no changes. Part (b) of the Lemma follows from part (a). Part (c) of the Lemma is a standard result giving the asymptotic normality of a suitably normalized weighted average of rank scores based on iid random variables, e.g., see Theorem V.1.6a of Hájek and Sidák (1967, p. 163) (extended from scalar constants c_i to vectors using the Cramér-Wold device). Condition (V.1.6.2) of Hájek and Sidák (1967, p. 163) holds under Assumption 3.

Lemma 1 *Let* $\Psi_n(t) = n^{-1}\sum_{i=1}^{n}(c_i - \overline{c}_n)\varphi(r_i(t)/(n+1))$, *where* (i) $r_i(t)$ *is the rank of* $Q_i - d_i't$ *among* $\{Q_j - d_j't : 1 \leq j \leq n\}$ *for a constant vector* $t \in R^{\delta_d}$, (ii) $\{Q_i : i \geq 1\}$ *is a sequence of iid random variables with absolutely-continuous strictly-increasing df* H *and absolutely-continuous and bounded density* h *that satisfies* $I(h) < \infty$, (iii) $\{c_i : i \geq 1\}$ *and* $\{d_i : i \geq 1\}$ *are fixed sequences of* δ_c*-vectors and* δ_d*-vectors, respectively, that satisfy* $\lim_{n\to\infty}\max_{1\leq i\leq n}||c_i - \overline{c}_n||^2/\sum_{i=1}^{n}||c_i - \overline{c}_n||^2 = 0$ *and* $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}||c_i - \overline{c}_n||^2 < \infty$ *and likewise with* $c_i - \overline{c}_n$ *replaced by* $d_i - \overline{d}_n$, *where* $\overline{c}_n = n^{-1}\sum_{i=1}^{n} c_i$ *and* $\overline{d}_n = n^{-1}\sum_{i=1}^{n} d_i$, *and* (iv) *the score function* φ *satisfies Assumption 3. Then,* (**a**) *for all* $\varepsilon > 0$ *and* $b > 0$,

$$\lim_{n\to\infty} P\left(\sup_{||t||\leq b} n^{1/2}\left|\Psi_n(tn^{-1/2}) - \Psi_n(0) - n^{-1/2}\dot{A}_n(0)t\right| > \varepsilon\right) = 0,$$

where

$$\dot{A}_n(0) = -n^{-1}\sum_{i=1}^{n}(c_i - \overline{c}_n)(d_i - \bar{d}_n)'\int_0^1 \varphi(x,h)\varphi(x)dx,$$

(**b**) *for any sequence of random* δ_d*-vectors* $\{\widehat{\tau}_n : n \geq 1\}$ *for which* $n^{1/2}\widehat{\tau}_n = O_p(1)$,

$$n^{1/2}\Psi_n(\widehat{\tau}_n) = n^{1/2}\Psi_n(0) + \dot{A}_n(0)n^{1/2}\widehat{\tau}_n + o_p(1),$$

(**c**) *provided* $\Sigma_c = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(c_i - \overline{c}_n)(c_i - \overline{c}_n)'$ *is pd,*

$$n^{1/2}\Psi_n(0) \to_d N(0, \Sigma_c\int_0^1[\varphi(x) - \overline{\varphi}]^2 dx).$$

Comments. 1. The expression for $\dot{A}_n(0)$ on p. 1277 of Koul (1970) is correct, but the expression for $\dot{A}_n(0)$ given on p. 1278 (which is of the form given above) contains a typo—a minus sign is missing. Also, the proof of Theorem 2.1 of Koul (1970) contains a typo that could be confusing to the reader. The term $\varphi(q_n)$ that appears at the end of the expression on the first two lines of the first equation on p. 1276 should be $\varphi'(q_n)$ in both places.

2. We do not require φ to satisfy the second condition of (i) on p. 1274 of Koul (1970) because this is a normalization condition that implies that $\varphi(1/2) = 0$ which is not needed for his Theorem 2.1 or Lemma 2.3. It is needed for his $n^{1/2}S_n(0)$ to have an asymptotic normal distribution. We do not require it for $n^{1/2}\Psi_n(0)$ to have an asymptotic normal distribution because we consider demeaned constant vectors $c_i - \overline{c}_n$, which yields $n^{1/2}\Psi_n(0)$ invariant to additive constants in φ, whereas Koul (1970) does not.

The following Lemma gives sufficient conditions for an iid sequence to satisfy Assumption 4(b) a.s.

Lemma 2 *Suppose $\{\xi_i : i \geq 1\}$ is an iid sequence of non-negative random variables with $E\xi_i^{1+\delta} < \infty$ for some $\delta > 0$. Then,* (**a**) $\sum_{i=1}^{\infty} \xi_i^{1+\delta}/i^{1+\delta} < \infty$ *a.s. and* (**b**) $\max_{i \leq n} \xi_i/n \to 0$ *a.s.*

Proof of Lemma 2. Part (a) holds because $E\sum_{i=1}^{\infty} \xi_i^{1+\delta}/i^{1+\delta} = E\xi_1^{1+\delta}\sum_{i=1}^{\infty} i^{-(1+\delta)} < \infty$ implies that $\sum_{i=1}^{\infty} \xi_i^{1+\delta}/i^{1+\delta} < \infty$ a.s. Part (b) holds because the result of part (a) and Kronecker's Lemma (e.g., see Chow and Teicher (1978, p. 111)) imply that $n^{-1-\delta}\sum_{i=1}^{n} \xi_i^{1+\delta} \to 0$ a.s. Hence, $n^{-1-\delta}\max_{i\leq n} \xi_i^{1+\delta} \leq n^{-1-\delta}\sum_{i=1}^{n} \xi_i^{1+\delta} \to 0$ a.s. In turn, this gives $n^{-1}\max_{i \leq n} \xi_i \to 0$ a.s. $\square$

Proof of Theorem 2. We prove part (a) first. It suffices to show that

$$\lim_{n\to\infty} P\left(n^{1/2}S_n + \Sigma_Z C'\gamma \int_0^1 \varphi(x,g)\varphi(x)dx \leq z\right) = P\left(G^* \leq z\right), \qquad (5.1)$$

for all $z \in R$, where $G^* \sim N(0, \Sigma_Z \int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx)$. We show that (5.1) holds conditional on an $\{X_i : i \geq 1\}$ sequence that satisfies certain properties, and that $\{X_i : i \geq 1\}$ sequences satisfy these properties with probability one. Because conditional probabilities are bounded by zero and one, this implies that (5.1) also holds unconditionally by the bounded convergence theorem.

We condition on a sequence $\{X_i : i \geq 1\}$ for which

$$\lim_{n\to\infty} \max_{1\leq i\leq n} \left\|X_i - \overline{X}_n\right\|^2 / \sum_{i=1}^{n} \left\|X_i - \overline{X}_n\right\|^2 = 0, \qquad (5.2)$$

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} (X_i - \overline{X}_n)(X_i - \overline{X}_n)' = \Sigma_X, \text{ and} \qquad (5.3)$$

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \left(Z_i - \overline{Z}_n\right)\left(X_i - \overline{X}_n\right)' = 0. \qquad (5.4)$$

Such sequences occur with probability one (a.s.). Conditions (5.2) and (5.3) hold a.s. under Assumptions 1 and 5 by Lemma 2(b) and the Kolmogorov strong LLN. Condition (5.4) holds a.s. under Assumptions 1, 4(c), and 5(a) by a strong LLN due to Loève, see Thm. 5.2.1 of Chow and Teicher (1978, p. 121).

By Assumptions 6(a) and (b) and 7W(a) and (b), we have

$$\begin{aligned} & y_{1i} - y_{2i}'\beta_0 - X_i'\widehat{\theta}_n \\ = {} & \alpha + y_{2i}'\gamma - X_i'(\widehat{\theta}_n - \theta) + u_i \\ = {} & \alpha + \mu'\gamma + Z_i'C'\gamma/n^{1/2} - X_i'(\widehat{\theta}_n - \theta - \Lambda'\gamma) + v_i'\gamma + u_i, \end{aligned} \qquad (5.5)$$

using $y_{2i} = \mu + CZ_i/n^{1/2} + \Lambda X_i + v_i$. The constant $\alpha + \mu\gamma$ does not affect the ranks of the right-hand side (rhs) expression in (5.5) and can be ignored.

We apply Lemma 1 with $\Psi_n(\widehat{\tau}_n) = S_n$, $Q_i = v_i'\gamma + u_i$, $c_i = Z_i$, $d_i = (Z_i', X_i')'$, $\widehat{\tau}_n = (-\gamma' C/n^{1/2}, (\widehat{\theta}_n - \theta - \Lambda'\gamma)')'$, and $h = g$. The assumptions of Lemma 1 on c_i are satisfied by Assumption 4. The required conditions for d_i are satisfied by Assumptions 2 and 4, (5.2), and (5.3). The assumptions of Lemma 1 for Q_i are satisfied by Assumptions 1, 6(c), and 7W(c) and (d).

Next, we show that $n^{1/2}\widehat{\tau}_n = O_p(1)$. By the definition of $\widehat{\theta}_n$, we have

$$\begin{aligned}
\widehat{\theta}_n &= \left(n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)(X_i - \overline{X}_n)' \right)^{-1} n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)(\alpha + y_{2i}'\gamma + X_i'\theta + u_i) \\
&= \theta + \Lambda'\gamma + \left(n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)(X_i - \overline{X}_n)' \right)^{-1} \\
&\quad \times n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)(Z_i' C'\gamma / n^{1/2} + v_i'\gamma + u_i), \qquad (5.6)
\end{aligned}$$

using $y_{2i} = CZ_i/n^{1/2} + \Lambda X_i + v_i$. Hence, we obtain

$$\begin{aligned}
&\left(n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)(X_i - \overline{X}_n)' \right) n^{1/2} (\widehat{\theta}_n - \theta - \Lambda'\gamma) \\
&= n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}_n) Z_i' C'\gamma + n^{-1/2} \sum_{i=1}^{n} (X_i - \overline{X}_n)(v_i'\gamma + u_i). \qquad (5.7)
\end{aligned}$$

The first multiplicand on the left-hand side of (5.7) equals $\Sigma_X + o(1)$, where $\Sigma_X > 0$ by (5.3). The first term on the rhs of (5.7) is $o(1)$ by (5.3) and (5.4). Each element of the second term on the rhs of (5.7) is asymptotically normal by the Lindeberg central limit theorem using Assumptions 1, 5(b), 6(c), and 7W(c), (5.2), and (5.3). In particular, the Lindeberg condition is satisfied element by element, because (i) wlog we can suppose X_i is a scalar, (ii) by (5.3), it suffices to show that for all $\varepsilon > 0$ $\lambda_n = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 E\xi_i^2 1((X_i - \overline{X}_n)^2 \xi_i^2 > n\varepsilon) \to 0$, where $\xi_i = v_i'\gamma + u_i$, and (iii) using $(X_i - \overline{X}_n)^2 \leq \max_{j \leq n}(X_j - \overline{X}_n)^2$ in the indicator function gives $\lambda_n \leq (n^{-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2) E\xi_i^2 1(\max_{j \leq n}(X_j - \overline{X}_n)^2 \xi_i^2 > n\varepsilon) \to 0$ by (5.2), (5.3), $E\xi_i^2 < \infty$, and the dominated convergence theorem. We conclude that $n^{1/2}(\widehat{\theta}_n - \theta - \Lambda'\gamma) = O_p(1)$, $n^{1/2}\widehat{\tau}_n = O_p(1)$, and the conditions of Lemma 1 hold.

Hence, by Lemma 1(b), $n^{1/2}S_n = n^{1/2}\Psi_n(0) + \dot{A}_n(0) n^{1/2}\widehat{\tau}_n + o_p(1)$ and by Lemma 1(c), $n^{1/2}\Psi_n(0) \to_d G^*$. Next, using the definitions of c_i, d_i, and $\widehat{\tau}_n$, we have $\dot{A}_n(0) n^{1/2}\widehat{\tau}_n / \int_0^1 \varphi(x, g)\varphi(x)\, dx$ equals

$$\begin{aligned}
&n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n)(Z_i - \overline{Z}_n)' C'\gamma \\
&- n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n)(X_i - \overline{X}_n)' n^{1/2} (\widehat{\theta}_n - \theta - \Lambda'\gamma)
\end{aligned}$$

$$= \Sigma_Z C'\gamma + o_p(1), \tag{5.8}$$

where the equality uses Assumption 4, (5.4), and $n^{1/2}(\widehat{\theta}_n - \theta - \Lambda'\gamma) = O_p(1)$. These results combine to give (5.1) conditional on an $\{X_i : i \geq 1\}$ sequence that satisfies (5.2)-(5.4) and the proof of part (a) is complete.

We now prove part (b). We use the same conditioning argument as in the proof of part (a). We condition on sequences $\{(X_i, \varepsilon_i) : i \geq 1\}$ for which (5.2)-(5.4) hold and the following conditions also hold:

$$\lim_{n\to\infty} \max_{1\leq i\leq n} ||\varepsilon_i - \overline{\varepsilon}_n||^2 / \sum_{i=1}^n ||\varepsilon_i - \overline{\varepsilon}_n||^2 = 0, \tag{5.9}$$

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^n ||\varepsilon_i - \overline{\varepsilon}_n||^2 < \infty, \text{ and} \tag{5.10}$$

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \left(Z_i - \overline{Z}_n\right)\left(\varepsilon_i - \overline{\varepsilon}_n\right)' = 0. \tag{5.11}$$

Conditions (5.9) and (5.10) hold a.s. by Assumption 7S(d), Lemma 2(b), and Kolmogorov's strong LLN. Condition (5.11) is satisfied a.s. by Assumptions 4(c) and 7S(d) and the strong LLN in Thm. 5.2.1 of Chow and Teicher (1978, p. 121).

By Assumptions 6(a) and (b) and 7S(a)-(c), we have

$$\begin{aligned} & y_{1i} - y_{2i}'\beta_0 - X_i'\widehat{\theta}_n \\ =& \alpha + \mu'\gamma + Z_i'\pi'\gamma/n^{1/2} - X_i'(\widehat{\theta}_n - \theta - \Lambda'\gamma/n^{1/2}) + \varepsilon_i'\gamma/n^{1/2} + (1 + \rho'\gamma/n^{1/2})u_i \end{aligned} \tag{5.12}$$

using $y_{2i} = \mu + \pi Z_i + \Lambda X_i + \varepsilon_i + \rho u_i$.

Let $\zeta_n = (1 + \rho'\gamma/n^{1/2})^{-1}$. Since $\zeta_n > 0$ for n sufficiently large, $\{R_i : i \leq n\}$ are equal to the ranks of the iid random variables $\{u_i : i \leq n\}$ plus the terms

$$\left\{\zeta_n Z_i'\pi'\gamma/n^{1/2} - \zeta_n X_i'(\widehat{\theta}_n - \theta - \Lambda'\gamma/n^{1/2}) + \zeta_n \varepsilon_i'\gamma/n^{1/2} : i \leq n\right\}. \tag{5.13}$$

We apply Lemma 1 with $\Psi_n(\widehat{\tau}_n) = S_n$, $Q_i = u_i$, $c_i = Z_i$, $d_i = (Z_i', X_i', \varepsilon_i')'$, $\widehat{\tau}_n = (-\zeta_n\gamma'\pi/n^{1/2}, \zeta_n(\widehat{\theta}_n - \theta - \Lambda'\gamma/n^{1/2})', -\zeta_n\gamma'/n^{1/2})'$, and $h = f$. The assumptions of Lemma 1 on c_i are satisfied by Assumptions 2 and 4. The required conditions for d_i are satisfied by Assumptions 2 and 4, (5.2), (5.3), (5.9), and (5.10). The assumptions of Lemma 1 for Q_i are satisfied by Assumptions 1 and 7S(e).

Next, we show that $n^{1/2}\widehat{\tau}_n = O_p(1)$. It suffices to show that $n^{1/2}(\widehat{\theta}_n - \theta) = O_p(1)$ because $\zeta_n \to 1$. We have

$$\begin{aligned} & \left(n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)(X_i - \overline{X}_n)'\right) n^{1/2}(\widehat{\theta}_n - \theta) \\ =& n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)Z_i'\pi'\gamma + n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)X_i'\Lambda'\gamma \\ & + n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)\varepsilon_i'\gamma + \zeta_n^{-1}n^{-1/2}\sum_{i=1}^n (X_i - \overline{X}_n)u_i. \end{aligned} \tag{5.14}$$

The first multiplicand on the left-hand side of (5.14) equals $\Sigma_X + o(1)$, where $\Sigma_X > 0$. The first term on the rhs is $o(1)$ by (5.4). The second term on the rhs equals $(\Sigma_X + o(1))\Lambda'\gamma = O(1)$. The third term on the rhs has Euclidean norm bounded by

$$||\gamma|| \left(n^{-1} \sum_{i=1}^{n} \left\| X_i - \overline{X}_n \right\|^2 \right)^{1/2} \left(n^{-1} \sum_{i=1}^{n} ||\varepsilon_i - \overline{\varepsilon}_n||^2 \right)^{1/2} = O(1) \tag{5.15}$$

by the Cauchy-Schwarz inequality, (5.3), and (5.10). Finally, the fourth term on the rhs is asymptotically normal and, hence, $O_p(1)$, by the Lindeberg CLT using Assumptions 1, 5(b), and 6(c) and (5.2) and (5.3). (The Lindeberg condition is verified by the same argument as above.) Hence, $n^{1/2}(\widehat{\theta}_n - \theta) = O_p(1)$ and Lemma 1(b) and (c) apply.

Next, using the definitions of c_i, d_i, and $\widehat{\tau}_n$, we have $\dot{A}_n(0) n^{1/2} \widehat{\tau}_n / \int_0^1 \varphi(x,f)\varphi(x)\,dx$ equals

$$\begin{aligned}
& \zeta_n n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n)(Z_i - \overline{Z}_n)'\pi'\gamma \\
& -\zeta_n n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n)(X_i - \overline{X}_n)' n^{1/2}(\widehat{\theta}_n - \theta - \Lambda'\gamma/n^{1/2}) \\
& +\zeta_n n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n)\varepsilon_i'\gamma \\
= {} & \Sigma_Z \pi'\gamma + o_p(1),
\end{aligned} \tag{5.16}$$

where the equality holds by Assumption 4, (5.4), $n^{1/2}(\widehat{\theta}_n - \theta) = O_p(1)$, (5.11), and $\zeta_n \to 1$.

Hence, by Lemma 1(b) and (c) and (5.16), we have

$$\begin{aligned}
n^{1/2} S_n &= \Sigma_Z \pi'\gamma \int_0^1 \varphi(x,f)\varphi(x)\,dx + n^{1/2}\Psi_n(0) + o_p(1) \\
&\to_d \Sigma_Z \pi'\gamma \int_0^1 \varphi(x,f)\varphi(x)\,dx + G^*
\end{aligned} \tag{5.17}$$

conditional on a sequence $\{(X_i, \varepsilon_i) : i \geq 1\}$ that satisfies (5.2)-(5.4) and (5.9)-(5.11), which completes the proof of part (b). □

Proof of Corollary 1. In the case of weak IVs, we use the proof of Theorem 2(a) with the considerable simplification that no X_i and $\widehat{\theta}_n$ appear and, hence, no conditioning on $\{X_i : i \geq 1\}$ sequences is required. Under Assumptions C1-C6, C7W, 2, and 3, this proof yields: For all $j = 1, \ldots, J$,

$$\begin{aligned}
& n_j^{-1/2} n S_{Cnj} \to_d \Sigma_{Zj} C_j \gamma_j b_j^{1/2} \int_0^1 \varphi(x)\varphi(x, g_j)dx + G_j^*, \text{ where} \\
& G_j^* \sim N\left(0, \Sigma_{Zj} \int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx\right).
\end{aligned} \tag{5.18}$$

The convergence in (5.18) is joint for $j = 1, \ldots, J$ by independence. Hence, $n^{1/2} S_{Cn}$ is asymptotically normal:

$$n^{1/2} S_{Cn} \to_d \sum_{j=1}^{J} w_j b_j \Sigma_{Zj} C_j \gamma_j \int_0^1 \varphi(x) \varphi(x, g_j) dx + G^*, \text{ where}$$
$$G^* \sim N\left(0, \sum_{j=1}^{J} w_j^2 b_j \Sigma_{Zj} \int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx\right). \tag{5.19}$$

Similarly, in the case of strong IVs, it follows from the proof of Theorem 2(b) that under Assumptions C1-C6, C7S, 2, and 3, we have

$$n^{1/2} S_{Cn} \to_d \sum_{j=1}^{J} w_j b_j \Sigma_{Zj} \pi_j \gamma_j \int_0^1 \varphi(x) \varphi(x, f_j) dx + G^*. \tag{5.20}$$

These results lead to the asymptotic properties for B_{Cn}, defined in (3.3), stated in the Corollary. □

References

Anderson, T. W. and H. Rubin (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *Annals of Mathematical Statistics,* 21, 570-582.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression," Cowles Foundation Discussion Paper No. 1476, Yale University.

Angrist, J. D., E. Bettinger, E. Bloom, E. King, and M. Kremer (2002): "Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment," *American Economic Review,* 92, 1535-1558.

Angrist, J. D. and W. N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review,* 88, 450-477.

Angrist, J. D. and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics,* 106, 979-1014.

Bekker, P. A. (2002): "Symmetry-based Inference in an Instrumental Variable Setting," unpublished manuscript, Dept. of Economics, University of Groningen.

Chow, Y. S. and H. Teicher (1978): *Probability Theory: Independence, Interchangeability, Martingales.* New York: Springer.

Duflo, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review,* 91, 795-813.

Duflo, E. and E. Saez (2003): "The Role of Information and Social Interaction in Retirement Plan Decisions: Evidence from a Randomized Experiment," *Quarterly Journal of Economics,* 118, 815-842.

Dufour, J.-M. (2003): "Identification, Weak Instruments, and Statistical Inference in Econometrics," *Canadian Journal of Economics,* 36, 767-808.

Dufour, J.-M. and J. Jasiak (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, 42, 815-843.

Greevy, R., J. H. Silber, A. Cnaan, and P. R. Rosenbaum (2004): "Randomization Inference with Imperfect Compliance in the ACE-Inhibitor after Anthracycline Randomized Trial," *Journal of the American Statistical Association*, 99, 7-15.

Gutenbrunner C. and J. Jurečková (1992): "Regression Rank-Scores and Regression Quantiles," *Annals of Statistics,* 20, 305-330.

Hájek, J. and Z. Sidák (1967): *Theory of Rank Tests.* New York: Academic Press.

Hettmansperger, T. P. (1984): *Statistical Inference Based on Ranks.* New York: Wiley.

Hodges, J. L. and E. L. Lehmann (1962): "Rank Methods for Combination of Independent Experiments in Analysis of Variance," *Annals of Mathematical Statistics,* 33, 482-497.

Imbens, G. W. and P. R. Rosenbaum (2005): "Robust, Accurate Confidence Intervals with a Weak Instrument: Quarter of Birth and Education," *Journal of the Royal Statistical Society, Series A,* 168, 109-126.

Koenker, R. W. (1996): "Rank Tests for Linear Models," in *Handbook of Statistics,* 15, ed. by C. R. Rao and G. S. Maddala. New York: Elsevier.

Koul, H. L. (1970): "A Class of ADF Tests for Subhypothesis in the Multiple Linear Regression," *Annals of Mathematical Statistics,* 41, 1273-1281.

Levitt, S. D. (1997): "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review,* 87, 270-290.

Moreira, M. J. (2001): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," unpublished manuscript, Dept. of Economics, University of California, Berkeley.

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

Rosenbaum, P. R. (1996): "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association,* 91, 465-468.

——— (2002): "Covariance Adjustment in Randomized Experiments and Observational Studies," *Statistical Science*, 17, 286-327.

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

Theil, H. (1950): "A Rank Invariant Method of Linear and Polynomial Regression Analysis," *Proceedings of the Royal Netherlands Academy of Sciences,* 53, part I: 386-392, part II: 521-525, part III: 1397-1412. Reprinted in *Henri Theil's Contributions to Economics and Econometrics,* Vol. I, ed. by B. Raj and J. Koerts, 1992, Dordrecht: Kluwer.

van Elteren, P. (1960): "On the Combination of Independent Two Sample Tests of Wilcoxon," *Bulletin of the International Institute of Statistics,* 37, 351-361.

TABLE I. Finite Sample Power of Wilcoxon Scores B^W, Normal Scores B^{NS}, and (size-corrected) Anderson-Rubin Tests of Significance Level $\alpha = .05$

Case*	λ	ρ	$\beta - \beta_0$	n	k	p	F	B^W	B^{NS}	AR
1. Base Case	9	.75	.95	100	1	5	Norm	.36	.37	.38
							t_1	.81	.79	.45
							t_2	.62	.59	.41
							t_3	.50	.48	.39
							t_{10}	.38	.38	.37
							DLN	.60	.56	.39
2. Other Distributions	9	.75	.95	100	1	5	$\mathbf{t}_4$	.45	.44	.39
							$\mathbf{t}_6$	.42	.41	.39
							Unif	.34	.40	.38
							Abs Norm	.40	.43	.39
							Logistic	.40	.40	.39
							DE	.45	.43	.39
							Log Norm	.70	.64	.39
3. Negative $\beta - \beta_0$	9	.75	**-.40**	100	1	5	Norm	.39	.41	.42
							t_1	.82	.79	.45
							t_2	.63	.60	.42
							t_3	.50	.48	.39
							t_{10}	.40	.40	.38
							DLN	.62	.57	.41
4. Strong IVs	**20**	.75	**.37**	100	1	5	Norm	.37	.38	.39
							t_1	.83	.81	.47
							t_2	.66	.62	.43
							t_3	.53	.50	.40
							t_{10}	.40	.39	.37
							DLN	.64	.60	.41
5. Weaker IVs	**4**	.75	**4.3**	100	1	5	Norm	.37	.39	.40
							t_1	.79	.77	.43
							t_2	.61	.58	.41
							t_3	.49	.47	.39
							t_{10}	.38	.38	.37
							DLN	.57	.53	.39

* λ=Strength of IVs, ρ=Correlation of Errors, $\beta - \beta_0$=Deviation from Null, n=Sample Size, k=Number of IVs, p=Number of Exogenous Variables, and F=Error/IV/Covariate Distribution

TABLE I (cont..)

Case	λ	ρ	$\beta - \beta_0$	n	k	p	F	B^W	B^{NS}	AR
6. High Endogeneity	9	**.95**	**1.08**	100	1	5	Norm	.36	.38	.39
							t_1	.83	.81	.48
							t_2	.66	.62	.43
							t_3	.53	.50	.40
							t_{10}	.40	.39	.37
							DLN	.64	.60	.41
7. No Endogeneity	9	**0.0**	**.62**	100	1	5	Normal	.37	.39	.39
							t_1	.79	.77	.43
							t_2	.61	.58	.41
							t_3	.49	.47	.39
							t_{10}	.38	.38	.37
							DLN	.57	.54	.39
8. Five IVs	9	.75	**2.5**	100	**5**	5	Norm	.39	.41	.41
							t_1	.91	.92	.46
							t_2	.70	.69	.41
							t_3	.54	.51	.40
							t_{10}	.40	.40	.41
							DLN	.66	.61	.36
9. Ten IVs	9	.75	**5.8**	100	**10**	5	Norm	.38	.39	.40
							t_1	.92	.94	.50
							t_2	.72	.70	.38
							t_3	.54	.52	.38
							t_{10}	.38	.39	.39
							DLN	.65	.59	.33
10. Sample Size 50	9	.75	**3.2**	**50**	1	5	Norm	.38	.39	.41
							t_1	.62	.60	.37
							t_2	.53	.51	.42
							t_3	.47	.46	.40
							t_{10}	.40	.40	.40
							DLN	.50	.49	.39
11. Sample Size 200	9	.75	**.53**	**200**	1	5	Norm	.38	.39	.40
							t_1	.92	.90	.49
							t_2	.71	.67	.41
							t_3	.54	.51	.41
							t_{10}	.40	.41	.40
							DLN	.70	.64	.42