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

BY

DONALD W. K. ANDREWS and VADIM MARMER

COWLES FOUNDATION PAPER NO. 1253

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

2008

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

Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 142 (2008) 183–200
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Exactly distribution-free inference in instrumental variables regression with possibly weak instruments

Donald W.K. Andrews[a], Vadim Marmer[b,*]

[a]Cowles Foundation for Research in Economics, Yale University, P.O. Box 208281, New Haven, CT 06520-8281, USA
[b]Department of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, B.C., Canada V6T 1Z1

Available online 24 May 2007

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.

JEL: C13; C30

Keywords: Aligned ranks; Anderson–Rubin statistic; Exact size; Rank test; Weak instruments

1. Introduction

The Anderson and Rubin (1949) (AR) test has a long history in econometrics. It was introduced over 50 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.[1]

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. This is established using exchangeability. Its asymptotic power properties are improved because it has equal asymptotic power under normal errors and

*Corresponding author. Tel.: +1 604 822 8217; fax: +1 604 822 5915.
E-mail addresses: donald.andrews@yale.edu (D.W.K. Andrews), vmarmer@interchange.ubc.ca (V. Marmer).

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

0304-4076/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2007.05.012

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. These conditions can be restrictive, but they are satisfied in a variety of applications of interest.

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 et al. (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 et al. (2006). 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 asymptotically most powerful tests that are available.

Under weak and strong IV asymptotics, we show that the rank statistics are asymptotically 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.

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 et al. (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 power results are given in the randomization inference papers. We view our results to be complementary to those based on randomization inference.

Andrews and Soares (2007) develop 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 in econometrics, see Koenker (1997). The closest paper in the econometrics literature to the present paper is McCabe (1989), which considers of aligned rank tests of misspecification in a linear regression model and uses exchangeability to show that the test is exact for normal and non-normal error distributions. 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. Appendix A 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 $i = 1, \ldots, n$, 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

$$\mathrm{H}_0 : \beta = \beta_0 \quad \text{and} \quad \mathrm{H}_1 : \beta \neq \beta_0 \quad \text{for some } \beta_0 \in R^{\ell}. \tag{2.2}$$

(Note that (2.1) and (2.2) also cover tests of $\mathrm{H}_0 : \beta = \beta_0$ and $\theta = \theta_0$ by absorbing X_i into y_{2i}.)

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

Assumption 2. $\{Z_i : i \geqslant 1\}$ 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 (conditional) heteroskedasticity are allowed. In fact, Assumptions 1 and 2 even allow for correlation between u_i and X_i. This is possible because of the strong exogeneity assumption on the IVs.

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 $\{y_{11} - y_{21}'\beta_0 - X_1'\widehat{\theta}_n, \ldots, y_{1n} - y_{2n}'\beta_0 - X_n'\widehat{\theta}_n\}$, $\widehat{\theta}_n$ 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.[2] The ranks $\{R_i : i \leqslant n\}$ 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:

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

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:

$$B_n = nS_n'W_nS_n, \text{ where}$$

$$W_n = \left(n^{-1}\sum_{i=1}^{n}(Z_i - \overline{Z}_n)(Z_i - \overline{Z}_n)' \cdot \int_0^1[\varphi(x) - \overline{\varphi}]^2\,\mathrm{d}x\right)^{-1} \tag{2.6}$$

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

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

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.[3] 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.

[2] 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 ℓth 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.

[3] The scalar constant $\int_0^1[\varphi(x) - \overline{\varphi}]^2\,\mathrm{d}x$ 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.

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, $\{Z_i - \overline{Z}_n : i \leqslant n\}$ are not related to the ranks $\{R_i : i \leqslant n\}$ 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 $\{Z_i - \overline{Z}_n : i \leqslant n\}$ are related to $\{y_{2i} : i \leqslant n\}$, then $\{Z_i - \overline{Z}_n : i \leqslant n\}$ are related to $\{u_i + y_{2i}'(\beta - \beta_0) - X_i'(\widehat{\theta}_n - \theta) : i \leqslant n\}$ and to their scored ranks $\{\varphi(R_i/(n+1)) : i \leqslant n\}$. (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

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

and $\{\eta_i : i \leqslant n\}$ 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 $\{y_{2i} : i \leqslant n\}$. 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). 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 $\{y_{2i} : i \leqslant n\}$ are related to the IVs $\{Z_i : i \geqslant 1\}$. 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.

5. It is apparent from the proof of Theorem 1 that Assumption 1 can be replaced by the weaker assumption of exchangeability of $\{(u_i, X_i) : i = 1, \ldots, n\}$ and the result of Theorem 1 still holds. Assumption 1 is required for the asymptotic power results given 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 $\tau \in (0, 1)$.[4] 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, $\{Z_i : i \leqslant n\}$, 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 $\{u_{ri} : i = 1, \ldots, n\}$, and calculates

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

[4]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 $1/n!$. One could randomize to obtain an exact test, but this would have little effect unless n is very small.

and R_{ri} is the rank of u_{ri} among $\{u_{r1}, \ldots, u_{rn}\}$.[5] One repeats this for $r = 1, \ldots, R_S$. The simulated critical value $c_{\text{sim},\tau}$ is the $1-\tau$ sample quantile of $\{B_{nr} : r = 1, \ldots, R_S\}$. The simulated p-value is $\text{p} = R_S^{-1}\sum_{r=1}^{R_S} 1(B_{nr} > b_{\text{ob}})$.

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.

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 $\text{H}_0 : \beta = \beta_0$ as $B_n(\beta_0)$. The CI is given by

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

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 $\text{H}_0 : \beta = \beta_0$ 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, $\{R_i : i \leqslant n\}$ are the ranks of $\{g(y_{1i}, y_{2i}, \beta_0) + X_i'\widehat{\theta}_n : i \leqslant n\}$ 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) $\varphi : (0,1) \to R$ is absolutely continuous and bounded with two derivatives that exist almost everywhere and are bounded.
(b) $\int_0^1 [\varphi(x) - \overline{\varphi}]^2\,\text{d}x > 0$.

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

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

Assumption 4. (a) $n^{-1}\sum_{i=1}^n (Z_i - \overline{Z}_n)(Z_i - \overline{Z}_n)' \to \Sigma_Z$ pd as $n \to \infty$.
(b) $\max_{1 \leqslant i \leqslant n} \|Z_i - \overline{Z}_n\|^2/n \to 0$ as $n \to \infty$.
(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 1 if $\{Z_i : i \geqslant 1\}$ is a realization of an iid sequence with pd variance matrix and $2+\delta$ moments finite for some $\delta > 0$, see Lemma 5 of Appendix A. Hence, Assumption 4 is not very restrictive.

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

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

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

[5] 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 $(0,1)$.

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

(b) $y_{2i} = \mu + \pi_n Z_i + \Lambda X_i + v_i$ for $i \geqslant 1$, where π_n is an $\ell \times k$ matrix of constants for $n \geqslant 1$, 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) $\{u_i : i \geqslant 1\}$ are independent of $\{X_i : i \geqslant 1\}$, and $\mathrm{E}u_i^2 < \infty$.

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, $I(f) = \int [f'(x)/f(x)]^2 f(x)\,\mathrm{d}x$, where f' denotes the derivative of f.

For weak IVs, we consider fixed alternatives and π_n that is local to 0.

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

(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 $i \geqslant 1$, where ε_i is a random ℓ-vector and $\rho \in R^\ell$ is a vector of constants.

(d) $\{\varepsilon_i : i \geqslant 1\}$ are iid and independent of $\{u_i : i \geqslant 1\}$, and $0 < \mathrm{E}\|\varepsilon_i\|^{2+\delta} < \infty$ 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. (These assumptions are not needed for the rank tests to have power.)

For a score function φ and a density f, define

$$\xi(\varphi, f) = \frac{\left(\int_0^1 \varphi(x)\varphi(x,f)\,\mathrm{d}x\right)^2}{\int_0^1 [\varphi(x) - \overline{\varphi}]^2\,\mathrm{d}x}, \quad \text{where } \varphi(x,f) = -\frac{f'(F^{-1}(x))}{f(F^{-1}(x))} \tag{2.13}$$

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

$$\xi(\varphi^{NS}, f) = \left(\int \frac{f^2(x)}{\phi(\Phi^{-1}(F(x)))}\,\mathrm{d}x\right)^2 \quad \text{and}$$

$$\xi(\varphi^{WS}, f) = 12\left(\int f^2(x)\,\mathrm{d}x\right)^2, \tag{2.14}$$

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

Let $\chi_k^2(\delta)$ 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,

$$B_n \to_d \chi_k^2(\delta_{\mathrm{W}}) \quad \textit{where } \delta_{\mathrm{W}} = \gamma' C \Sigma_Z C' \gamma \xi(\varphi, g).$$

(b) *Under Assumptions* 1–6 *and* 7S,

$$B_n \to_d \chi_k^2(\delta_{\mathrm{S}}) \quad \textit{where } \delta_{\mathrm{S}} = \gamma' \pi \Sigma_Z \pi' \gamma \xi(\varphi, f).$$

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 $n^{1/2}(\theta_n^* - \theta - \Lambda'(\beta_n - \beta_0)) = O_p(1)$.[6]

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). 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 2, just as with the AR test. In sum, aligning the ranks typically increases the power of rank tests.

For the AR statistic, $AR_n \times k \to_d \chi_k^2(\delta_{\mathrm{W}}^{AR})$ and $AR_n \times k \to_d \chi_k^2(\delta_{\mathrm{S}}^{AR})$ under weak and strong IV asymptotics, respectively, where

$$\delta_{\mathrm{W}}^{AR} = C'\Sigma_Z C\gamma^2/\sigma_g^2 \quad \text{and} \quad \delta_{\mathrm{S}}^{AR} = \pi'\Sigma_Z\pi\gamma^2/\sigma_f^2, \tag{2.15}$$

and σ_g^2 and σ_f^2 denote the variances corresponding to the densities g and f, respectively, under the assumptions above plus $\mathrm{E}u_i^2 < \infty$.

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

$$\begin{aligned} ARE_f(B_n, AR) &= \xi(\varphi, g)\sigma_g^2 \text{ for weak IVs and} \\ ARE_f(B_n, AR) &= \xi(\varphi, f)\sigma_f^2 \text{ for strong IVs.} \end{aligned} \tag{2.16}$$

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

$$ARE_f(NS, AR) = \sigma^2(f)\left(\int \frac{f^2(x)}{\phi(\Phi^{-1}(F(x)))}\,\mathrm{d}x\right)^2. \tag{2.17}$$

A result due to Chernoff and Savage states that $ARE_f(NS, AR) \geqslant 1$ for all symmetric distributions f (about some point not necessarily 0), see Hettmansperger (1984, Theorem 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

$$ARE_f(WS, AR) = 12\sigma_f^2\left(\int f^2(x)\,\mathrm{d}x\right)^2. \tag{2.18}$$

For the normal distribution ϕ, $ARE_\phi(WS, AR) = .955$. For the double exponential distribution f_{de}, $ARE_{f_{\mathrm{de}}}(WS, AR) = 1.50$. 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. 71–72). A result due to Hodges and Lehmann states that $ARE_f(WS, AR) \geqslant 0.864$ for all symmetric distributions f (about some point not necessarily 0), see Hettmansperger (1984, Theorem 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

[6] 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.

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 0, $ARE_{f_1}(WS, NS) \leqslant ARE_{f_2}(WS, NS)$ 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 $x \geqslant 0$, see Hettmansperger (1984, Theorem 2.9.5, 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 Hettmansperger (1984, Theorem 2.9.3, p. 115).

3. IV Regression with categorical covariates

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. $\{u_i : i \geqslant 1\}$ 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. $\{D_i : i \geqslant 1\}$ 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 $\{Z_i : i \geqslant 1\}$ 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 $i = 1, \ldots, n_j$, for $j = 1, \ldots, J$. 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:

$$B_{Cn} = nS_{Cn}'W_{Cn}S_{Cn},$$

where

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

$$\overline{Z}_{nj} = n_j^{-1} \sum_{i=1}^{n_j} Z_{ij},$$

$$W_{Cn} = \left(n^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n_j} (Z_{ij} - \overline{Z}_{nj})(Z_{ij} - \overline{Z}_{nj})' \int_0^1 [\varphi(x) - \overline{\varphi}]^2 \, dx \right)^{-1}. \tag{3.3}$$

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 $(\alpha_1, \ldots, \alpha_J)$, β_0, or the distribution of $\{y_{2,ij} : i \leqslant n_j, j = 1, \ldots, J\}$. 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* $\{\alpha_j : j = 1, \ldots, J\}$, β_0, $\{F_j : j = 1, \ldots, J\}$, *or the distribution of the endogenous variables* $\{y_{2,i} : i \leqslant n_j, j = 1, \ldots, J\}$.

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

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

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 simulated 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 $p = R_S^{-1} \sum_{r=1}^{R_S} 1(B_{Cn,r} > b_{ob})$.

In Andrews and Marmer (2005), we show that the B_{Cn} statistic has a non-central chi-square asymptotic distribution using weak IV asymptotics/fixed alternatives and strong IV asymptotics/local alternatives under assumptions that are similar to those in Section 2.3. Under strong IV asymptotics, the ARE of a categorical rank test to an analogous categorical AR test is the same as in Section 2.3 provided F_j does not depend on j. Under weak IV asymptotics, the ARE is the same provided the densities $\{g_j : j \leqslant J\}$, which are analogous to g in Assumption 7W, do not depend on j. In Andrews and Marmer (2005), we consider non-random weights assigned to the J statistics $\{S_{Cnj} : j \leqslant J\}$ in the definition of S_{Cn}. We determine optimal weighting schemes in terms of asymptotic power. Under certain conditions, the constant weights employed here are optimal.

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, $\{R_{ij} : i \leqslant n\}$ are defined to be the ranks of $\{g_j(y_{1i}, y_{2i}, \beta_0) : i \leqslant n\}$. Otherwise, the test statistic B_{Cn} and its critical value are the same as above.

4. Monte Carlo results

4.1. Experimental design

In this section, we report simulated power comparisons of the B_n^{WS}, 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 ($\ell = 1$):

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

$$y_{2i} = Z_i'\pi + X_i'\Lambda + (1-\rho^2)^{1/2}\varepsilon_i + \rho u_i, \tag{4.1}$$

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

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 0 and we take F to have mean 0 (if its mean is well defined), center of symmetry 0 (if it is symmetric), and variance 1 (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:

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

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):

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

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 $\mathrm{H}_0 : \beta = \beta_0$ and $\mathrm{H}_1 : \beta \neq \beta_0$. 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, d, and $F = \Phi$.

We provide results for selected subsets of the cases for which $n = 50$, 100, 200; $k = 1$, 5, 10; $d = 0$, 5; and F is normal, t_r with r degrees of freedom 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 degrees of freedom 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^{WS} 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 5000 Monte Carlo simulations.

We first consider a Base Case in which $\lambda = 9$, $\rho = .75$, $\beta - \beta_0 > 0$, $n = 100$, $k = 1$, $d = 5$, 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 1 reports the results for five cases. Andrews and Marmer (2005) report results for six additional cases. 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

Table 1
Finite sample power of Wilcoxon scores B_n^{WS}, normal scores B_n^{NS}, and (size-corrected) Anderson–Rubin tests of significance level $\alpha = 0.05$

Case[a]	λ	ρ	$\beta-\beta_0$	n	k	d	F	B_n^{WS}	B_n^{NS}	AR
1. Base case	9	0.75	0.95	100	1	5	Norm	0.36	0.37	0.38
							t_1	0.81	0.79	0.45
							t_2	0.62	0.59	0.41
							t_3	0.50	0.48	0.39
							t_{10}	0.38	0.38	0.37
							DLN	0.60	0.56	0.39
2. Other distributions	9	0.75	0.95	100	1	5	t_4	0.45	0.44	0.39
							t_6	0.42	0.41	0.39
							Unif	0.34	0.40	0.38
							Abs Norm	0.40	0.43	0.39
							Logistic	0.40	0.40	0.39
							DE	0.45	0.43	0.39
							Log Norm	0.70	0.64	0.39
3. Negative $\beta-\beta_0$	9	0.75	**−0.40**	100	1	5	Norm	0.39	0.41	0.42
							t_1	0.82	0.79	0.45
							t_2	0.63	0.60	0.42
							t_3	0.50	0.48	0.39
							t_{10}	0.40	0.40	0.38
							DLN	0.62	0.57	0.41
4. Strong IVs	**20**	0.75	**0.37**	100	1	5	Norm	0.37	0.38	0.39
							t_1	0.83	0.81	0.47
							t_2	0.66	0.62	0.43
							t_3	0.53	0.50	0.40
							t_{10}	0.40	0.39	0.37
							DLN	0.64	0.60	0.41
5. Weaker IVs	**4**	0.75	**4.3**	100	1	5	Norm	0.37	0.39	0.40
							t_1	0.79	0.77	0.43
							t_2	0.61	0.58	0.41
							t_3	0.49	0.47	0.39
							t_{10}	0.38	0.38	0.37
							DLN	0.57	0.53	0.39

[a] λ = Strength of IVs, ρ = correlation of errors, $\beta-\beta_0$ = deviation from null, n = sample size, k = number of IVs, d = number of covariates, and F = error/IV/covariate distribution.

highly skewed, the NS test is somewhat more powerful than the WS and AR 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 provide 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. (For brevity, only the results of Cases 3–5 are reported in Table 1.) 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), 10 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_n^{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_n^{WS}, 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.

Acknowledgments

The authors thank two referees, 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 Nos. SES-0001706 and SES-0417911.

Appendix A. 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 Hájek and Sidák (1967, Theorem V.1.6a, 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 4. *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* $\{Q_j - d_j't : 1 \leqslant j \leqslant n\}$ *for a constant vector* $t \in R^{\delta_d}$, (ii) $\{Q_i : i \geqslant 1\}$ *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) $\{c_i : i \geqslant 1\}$ *and* $\{d_i : i \geqslant 1\}$ *are fixed sequences of* δ_c*-vectors and* δ_d*-vectors, respectively, that satisfy the conditions* $\lim_{n\to\infty}\max_{1\leqslant i\leqslant n}\|c_i - \overline{c}_n\|^2/\sum_{i=1}^n\|c_i - \overline{c}_n\|^2 = 0$ *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* $b > 0$,

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

where

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

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

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

(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*,

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

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 yield $n^{1/2}\Psi_n(0)$ 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 5. *Suppose* $\{\xi_i : i \geqslant 1\}$ *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) $\max_{i \leqslant n} \xi_i / n \to 0$ *a.s.*

Proof of Lemma 5. 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, $n^{-1-\delta}\max_{i \leqslant n} \xi_i^{1+\delta} \leqslant n^{-1-\delta} \sum_{i=1}^{n} \xi_i^{1+\delta} \to 0$ a.s. In turn, this gives $n^{-1}\max_{i \leqslant n} \xi_i \to 0$ a.s. □

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

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

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

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

$$\lim_{n\to\infty} \max_{1\leqslant i \leqslant n} \|X_i - \overline{X}_n\|^2 \Big/ \sum_{i=1}^{n} \|X_i - \overline{X}_n\|^2 = 0, \tag{5.2}$$

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

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

Such sequences occur with probability 1 (a.s.). Conditions (5.2) and (5.3) hold a.s. under Assumptions 1 and 5 by Lemma 5(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 Chow and Teicher (1978, Theorem 5.2.1, p. 121).

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

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

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 4 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 4 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 4 for Q_i are satisfied by Assumptions 1, 6(c), and 7W(c) and (d).

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

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

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 $\Sigma_X + \mathrm{o}(1)$, 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) without loss of generality we can suppose X_i is a scalar, (ii) by (5.3), it suffices to show that for all $\varepsilon > 0$ $\lambda_n = n^{-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \mathrm{E}\xi_i^2 1((X_i-\overline{X}_n)^2\xi_i^2 > n\varepsilon) \to 0$, where $\xi_i = v_i'\gamma + u_i$, and (iii) using $(X_i - \overline{X}_n)^2 \leqslant \max_{j\leqslant n}(X_j - \overline{X}_n)^2$ in the indicator function gives $\lambda_n \leqslant (n^{-1}\sum_{j=1}^{n}(X_j-\overline{X}_n)^2)\mathrm{E}\xi_i^2 1(\max_{j\leqslant n}(X_j-\overline{X}_n)^2\xi_i^2 > n\varepsilon) \to 0$ by (5.2), (5.3), $\mathrm{E}\xi_i^2 < \infty$, and the dominated convergence theorem. We conclude that $n^{1/2}(\widehat{\theta}_n - \theta - \Lambda'\gamma) = \mathrm{O}_p(1)$, $n^{1/2}\widehat{\tau}_n = \mathrm{O}_p(1)$, and the conditions of Lemma 4 hold.

Hence, by Lemma 4(b), $n^{1/2}S_n = n^{1/2}\Psi_n(0) + \dot{A}_n(0)n^{1/2}\widehat{\tau}_n + \mathrm{o}_p(1)$ and by Lemma 4(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 $\dot{A}_n(0)n^{1/2}\widehat{\tau}_n / \int_0^1 \varphi(x,g)\varphi(x)\,\mathrm{d}x$ equals

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}(Z_i-\overline{Z}_n)(Z_i-\overline{Z}_n)'C'\gamma \\
&\quad - n^{-1}\sum_{i=1}^{n}(Z_i-\overline{Z}_n)(X_i-\overline{X}_n)'n^{1/2}(\widehat{\theta}_n-\theta-\Lambda'\gamma) \\
&= \Sigma_Z C'\gamma + \mathrm{o}_p(1), \qquad (5.8)
\end{aligned}$$

where the equality uses Assumption 4, (5.4), and $n^{1/2}(\widehat{\theta}_n - \theta - \Lambda'\gamma) = \mathrm{O}_p(1)$. These results combine to give (5.1) conditional on an $\{X_i : i \geqslant 1\}$ 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 $\{(X_i, \varepsilon_i) : i \geqslant 1\}$ for which (5.2)–(5.4) hold and the following conditions also hold:

$$\lim_{n\to\infty}\max_{1\leqslant i\leqslant n}\|\varepsilon_i-\overline{\varepsilon}_n\|^2 \Big/ \sum_{i=1}^{n}\|\varepsilon_i-\overline{\varepsilon}_n\|^2 = 0, \qquad (5.9)$$

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

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

Conditions (5.9) and (5.10) hold a.s. by Assumption 7S(d), Lemma 5(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 Chow and Teicher (1978, Theorem 5.2.1, p. 121).

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

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

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, $\{R_i : i\leqslant n\}$ are equal to the ranks of the iid random variables $\{u_i : i\leqslant n\}$ plus the terms

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

We apply Lemma 4 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 4 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 4 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

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

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

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

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 4(b) and (c) apply.

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

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

$$
\begin{aligned}
&+ \zeta_n n^{-1} \sum_{i=1}^{n} (Z_i - \overline{Z}_n)\, \varepsilon_i' \gamma \\
&= \Sigma_Z \pi' \gamma + \mathrm{o}_p(1), \qquad (5.16)
\end{aligned}
$$

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

Hence, by Lemma 4(b) and (c) and (5.16), we have

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

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

References

Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. The Annals of Mathematical Statistics 20 (1), 46–63.

Andrews, D.W.K., Marmer, V., 2005. Exactly distribution-free inference in instrumental variables regression with possibly weak instruments. Cowles Foundation Discussion Paper 1501, Yale University.

Andrews, D.W.K., Moreira, M.J., Stock, J.H., 2006. Optimal invariant similar tests for instrumental variables regression. Econometrica 74 (3), 715–752.

Andrews, D.W.K., Soares, G., 2007. Rank tests for instrumental variables regression with weak instruments. Econometric Theory 23 (6) forthcoming.

Angrist, J.D., Evans, W.N., 1998. Children and their parents' labor supply: evidence from exogenous variation in family size. American Economic Review 88 (3), 450–477.

Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect schooling and earnings? The Quarterly Journal of Economics 106 (4), 979–1014.

Angrist, J.D., Bettinger, E., Bloom, E., King, E., Kremer, M., 2002. Vouchers for private schooling in Colombia: evidence from a randomized natural experiment. American Economic Review 92 (5), 1535–1558.

Bekker, P.A., 2002. Symmetry-based inference in an instrumental variable setting. Department of Economics, University of Groningen (unpublished manuscript).

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

Duflo, E., 2001. Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment. American Economic Review 91 (4), 795–813.

Duflo, E., Saez, E., 2003. The role of information and social interaction in retirement plan decisions: evidence from a randomized experiment. The Quarterly Journal of Economics 118 (3), 815–842.

Dufour, J.-M., 2003. Identification, weak instruments and statistical inference in econometrics. Canadian Journal of Economics 36 (4), 767–808.

Greevy, R., Silber, J.H., Cnaan, A., Rosenbaum, P.R., 2004. Randomization inference with imperfect compliance in the ACE-inhibitor after anthracycline randomized trial. Journal of the American Statistical Association 99 (465), 7–16.

Gutenbrunner, C., Jurečková, J., 1992. Regression rank scores and regression quantiles. The Annals of Statistics 20 (1), 305–330.

Hájek, J., Sidák, Z., 1967. Theory of Rank Tests. Academic Press, New York.

Hettmansperger, T.P., 1984. Statistical Inference Based on Ranks. Wiley, New York.

Hodges, J.L., Lehmann, E.L., 1962. Rank methods for combination of independent experiments in analysis of variance. The Annals of Mathematical Statistics 33 (2), 482–497.

Imbens, G.W., Rosenbaum, P.R., 2005. Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. Journal of the Royal Statistical Society Series A 168 (1), 109–126.

Koenker, R., 1997. Rank tests for linear models. In: Rao, C.R., Maddala, G.S. (Eds.), Handbook of Statistics, vol. 15. Elsevier, New York, pp. 175–199 (Chapter 8).

Koul, H.L., 1970. A class of ADF tests for subhypothesis in the multiple linear regression. The Annals of Mathematical Statistics 41 (4), 1273–1281.

Levitt, S.D., 1997. Using electoral cycles in police hiring to estimate the effect of police on crime. American Economic Review 87 (3), 270–290.

McCabe, B.P.M., 1989. Misspecification tests in econometrics based on ranks. Journal of Econometrics 40 (2), 261–278.

Moreira, M.J., 2001. Tests with correct size when instruments can be arbitrarily weak. Department of Economics, University of California, Berkeley (unpublished manuscript).

Moreira, M.J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71 (4), 1027–1048.

Rosenbaum, P.R., 1996. Identification of causal effects using instrumental variables: comment. Journal of the American Statistical Association 91 (434), 465–468.

Rosenbaum, P.R., 2002. Covariance adjustment in randomized experiments and observational studies. Statistical Science 17 (3), 286–327.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557–586.

Theil, H., 1950. A rank invariant method of linear and polynomial regression analysis. In: Proceedings of the Royal Netherlands Academy of Sciences, vol. 53 part I: 386–392, part II: 521–525, part III: 1397–1412.

van Elteren, P.H., 1960. On the combination of independent two-sample tests of Wilcoxon. Bulletin of the International Institute of Statistics 37 (3), 351–361.