3-1-2006

Rank Tests for Instrumental Variables Regression with Weak Instruments

Donald W.K. Andrews

Gustavo Soares

RANK TESTS FOR INSTRUMENTAL VARIABLES REGRESSION WITH WEAK INSTRUMENTS

By

Donald W.K. Andrews and Gustavo Soares

March 2006

COWLES FOUNDATION DISCUSSION PAPER NO. 1564

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

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

Rank Tests for Instrumental Variables Regression with Weak Instruments

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

Gustavo Soares
Department of Economics
Yale University

December 2004
Revised: March 2006

[1] Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-0417911.

Abstract

This paper considers tests in an instrumental variables (IVs) regression model with IVs that may be weak. Tests that have near-optimal asymptotic power properties with Gaussian errors for weak and strong IVs have been determined in Andrews, Moreira, and Stock (2006a). In this paper, we seek tests that have near-optimal asymptotic power with Gaussian errors and improved power with non-Gaussian errors relative to existing tests. Tests with such properties are obtained by introducing rank tests that are analogous to the conditional likelihood ratio test of Moreira (2003). We also introduce a rank test that is analogous to the Lagrange multiplier test of Kleibergen (2002) and Moreira (2001).

Keywords: Asymptotically similar tests, conditional likelihood ratio test, instrumental variables regression, Lagrange multiplier test, power of test, rank tests, thick-tailed distribution, weak instruments.

JEL Classification Numbers: C12, C30.

1 Introduction

This paper is concerned with inference in the standard linear instrumental variable (IV) regression model with possibly weak IVs. We start by giving a brief account of recent developments in the literature on weak IVs in order to explain the contribution of this paper to the literature. It has been documented in the weak IV literature that standard methods, such as two-stage least squares-based tests and confidence intervals (CIs), perform poorly when IVs are weak, especially when endogeneity is moderate to strong. Specifically, such tests have size well in excess of their nominal level and corresponding CIs have size well below their nominal level. See the review papers of Stock, Wright, and Yogo (2002), Dufour (2003), and Andrews and Stock (2005).

The well-known Anderson and Rubin (1949) (AR) test does not exhibit size distortions due to weak IVs. Hence, Staiger and Stock (1997) and Dufour (1997) propose basing inference on the AR test. AR-based CIs can be constructed by inverting AR tests. The AR test has good power properties when the model is just identified, see Moreira (2001) and Andrews, Moreira, and Stock (2006a) (AMS1) for some optimality properties for the case of Gaussian errors. However, the AR test sacrifices power when the model is over-identified. This leads to excessively long AR-based CIs.

In consequence, considerable effort has been expended recently to develop new tests that circumvent this problem. Such tests are of interest in their own right and because they can be used to construct CIs by inversion. Kleibergen (2002) and Moreira (2001) introduce an LM test whose size is robust to weak IVs and whose power exceeds that of the AR test in many cases when the model is over-identified. However, this test has somewhat quirky power properties. For example, its power function can be non-monotonic, see AMS1 and Andrews, Moreira, and Stock (2006b) (AMS2).

Subsequently, Moreira (2003) showed that any test can be made robust to weak IVs asymptotically by using a conditional critical value function that conditions on a statistic that is complete and sufficient under the null hypothesis. Using this method, he introduced the conditional likelihood ratio (CLR) test. AMS investigate the power properties of the CLR test in the case of a single right-hand side endogenous variable and show that its power is essentially on the asymptotic power envelope for two-sided invariant similar tests under the assumption of Gaussian errors. This is true under both the "weak IV asymptotics" introduced in Staiger and Stock (1997), in which the coefficient on the IVs in the first-stage regression shrinks to zero as the sample size goes to infinity, and under the standard "strong IV asymptotics." Andrews and Stock (2005) show that these optimality properties extend to the "many weak IV asymptotic scenario," in which the number of IVs increases with the sample size. Hence, the CLR test has the desirable features of having size that is robust to weak IVs and near-optimal power properties with Gaussian errors.[2]

In this paper, we aim to further improve the power properties of weak IV tests by constructing a test that has the same asymptotic behavior as the CLR test with

[2] We note that the CLR test reduces to the AR test when the model is just-identified, so that the optimality properties mentioned in this paragraph are consistent with those mentioned above for the AR test.

Gaussian errors, but improved power with non-Gaussian errors. To do this, we construct a *rank* analogue of the CLR test, denoted RCLR. We also construct a rank analogue of the LM test of Kleibergen (2002) and Moreira (2001), denoted RLM. As is well-known from location and regression models, rank estimators and tests have more robust efficiency properties than least squares-based procedures, see Hettmansperger (1984). For example, Chernoff and Savage (1958) have shown that the asymptotic relative efficiency (ARE) of the normal scores rank test to the analogous least-squares t test is greater than or equal to one for all symmetric error distributions with equality at the Gaussian. This holds in both location and regression models, and it also holds for estimators. This suggests that for the linear IV model rank-based tests whose size is robust to weak IVs may exhibit similarly desirable power properties under non-normality.

Andrews and Marmer (2004) develop a rank analogue of the AR test, denoted RAR. This test has exact finite sample size under Gaussian and non-Gaussian errors under certain circumstances. Its asymptotic power properties improve on those of the AR test and are excellent for just-identified models. However, as with the non-rank AR test, the RAR test sacrifices power in over-identified models. The RCLR and RLM tests developed here substantially improve the power properties of the RAR test in over-identified models.

We now summarize the results of the paper. The model considered is a linear IV regression model with a single structural equation with m right-hand side endogenous variables and p exogenous variables coupled with m reduced-form equations for the rhs endogenous variables. The null hypothesis is $H_0 : \beta = \beta_0$, where β is the m-dimensional coefficient on the m rhs endogenous variables. The alternative hypothesis is $H_1 : \beta \neq \beta_0$.

First, we introduce rank analogues of the CLR and LM tests. This is more difficult than for the AR test because the LR and LM statistics are more complicated functions of the data than is the AR statistic. A hybrid rank/linear test statistic is required to obtain power properties of RCLR and RLM tests that are analogous to those of the CLR and LM tests under Gaussianity and superior for other distributions.

Second, we obtain the weak IV asymptotic distributions of the rank statistics under the null and fixed alternatives. These results are used to show that under Gaussian errors the normal scores (NS) RCLR and RLM tests have the same null and alternative asymptotic behavior as the non-rank versions of these tests. The same is true for the Wilcoxon scores (WS) rank and non-rank CLR and LM tests under uniform errors. Furthermore, these asymptotic distributions allow one to compare the weak IV asymptotic power of the rank to non-rank tests under different error distributions. It is shown that the same AREs for the rank versus non-rank LM and AR tests arise in the weak IV context as in the location and regression models. Hence, the Chernoff-Savage result also applies to these tests. That is, the NS-RLM test (weakly) dominates the LM test in terms of power for all symmetric error distributions and the same is true for the NS-RAR test versus the AR test.

For the rank versus non-rank CLR tests, the weak IV asymptotic power comparison is more complicated. However, numerical calculation of the asymptotic powers of these tests shows the same pattern that is typical for rank versus non-rank procedures in other contexts. In particular, the NS-RCLR test has noticeably higher asymptotic power for thin-tailed (uniform) and thick-tailed (t_3 and difference of independent log normals (DLN)) errors than the non-rank CLR test and equal asymptotic power for Gaussian errors. The WS-RCLR test has asymptotic power that is close to that of the CLR test for Gaussian and uniform errors and substantially higher power for t_3 and DLN errors.

Third, we establish the strong IV asymptotic distributions of the rank statistics under the null and local alternatives. These results show that the RCLR and RLM tests are asymptotically equivalent under strong IV asymptotics. This is also true of the non-rank versions of these tests. The results also show that the ARE of the rank to the non-rank versions of these tests under strong IV asymptotics is the same as the standard ARE that arises in location and regression models for tests and estimators. Hence, the Chernoff-Savage result applies under strong IV asymptotics to both the NS-RCLR test and the NS-RLM test. In consequence, the NS-RCLR test (weakly) dominates the CLR test in terms of power for symmetric errors under strong IV asymptotics.

The proofs of the weak and strong IV asymptotic results make use of results and arguments given in Hájek and Sidák (1967) and Koul (1969, 1970).

Fourth, we carry out finite sample size and power comparisons of the WS-RCLR, NS-RCLR, CLR, LM, and AR tests. For brevity, we do not report results for the RLM and RAR tests, because they are found to be inferior (both asymptotically and in finite sample experiments) to those of the RCLR tests. We compare the tests for a variety of scenarios that differ according to the degree of endogeneity, strength of the IVs, number of IVs, and size of the sample. For each scenario we consider Gaussian, uniform, t_1, t_2, t_3, and DLN errors. The two RCLR tests perform noticeably better in terms of size than the non-rank CLR, LM, and AR tests. The finite sample power comparisons reflect the asymptotic power comparisons discussed above fairly closely. Specifically, the NS-RCLR test has similar power to the CLR test for Gaussian errors and higher power for non-Gaussian errors. The WS-RCLR test does not perform as well as the NS-RCLR test with uniform errors, but it performs better with thick-tailed errors.

Based on the asymptotic and finite sample results, we recommend the NS-RCLR test over the WS-RCLR, CLR, LM, and AR tests. The WS-RCLR test also has good overall properties, but we prefer the NS-RCLR test because of its excellent power performance for both thin-tailed and thick-tailed errors.

The main drawback of the RCLR tests is that they are not robust to heteroskedasticity of the errors. That is, their size may be distorted by heteroskedasticity. This is also true of the CLR test. However, it is possible to robustify the CLR test to heteroskedasticity, see Andrews, Moriera, and Stock (2004) and Kleibergen (2005). It is not possible to robustify the RCLR tests to heteroskedasticity. Hence, there is a trade-off between power for non-Gaussian errors and robustness to heteroskedasticity for these tests. If heteroskedasticity is a possible problem, then the robustified CLR test is preferred to the NS-RCLR or WS-RCLR tests. If not, then the rank tests are preferred.

There is a vast literature on rank procedures in statistics, e.g., see Hájek and Sidák (1967), Hettmansperger (1984), Puri and Sen (1985), and Hájek, Sidák, and Sen (1999). Rank procedures have been used in both cross-section and time series econometrics. For a review, see Koenker (1996). Some more recent econometric references include Hasan and Koenker (1997), Cavanagh and Sherman (1998), Abrevaya (1999), Chen (2000, 2002), and Thompson (2004).

The remainder of this paper is organized as follows. Section 2 defines the model. Section 3 introduces the rank analogues of the CLR, LM, and AR tests. Sections 4 and 5 provide asymptotic results for these tests under weak IV and strong IV asymptotics, respectively. These sections also give asymptotic power comparisons of rank and non-rank tests. Section 6 provides finite sample size and power comparisons of rank and non-rank tests. An Appendix contains proofs of the results.

All limits are taken as $n \to \infty$. $vec(\cdot)$ is the column by column vec operator.

2 Model

We consider the following model, which consists of a single structural equation and m reduced-form equations:

$$\begin{aligned} y_{1i} &= \beta' y_{2i} + \gamma_1' X_i + u_i, \\ y_{2i} &= \Pi' \widetilde{Z}_i + \xi_1' X_i + v_{2i}, \end{aligned} \tag{2.1}$$

where $y_{1i} \in R$, $y_{2i} \in R^m$, $X_i \in R^p$, and $\widetilde{Z}_i \in R^k$ are observed variables; $u_i \in R$ and $v_{2i} \in R^m$ are unobserved errors; and $\beta \in R^m$, $\Pi \in R^{k \times m}$, $\gamma_1 \in R^p$, and $\xi_1 \in R^{p \times m}$ are unknown parameters.

Our interest is in testing the hypotheses

$$H_0 : \beta = \beta_0 \text{ and } H_1 : \beta \neq \beta_0. \tag{2.2}$$

Let $\widetilde{Z}$ and X denote the $n \times k$ IV and $n \times p$ regressor matrices whose i-th rows are $\widetilde{Z}_i'$ and X_i', respectively. We transform the IV matrix $\widetilde{Z}$ so that the transformed IV matrix, Z, and the regressor matrix, X, are orthogonal:

$$\begin{aligned} Z &= M_X \widetilde{Z},\ M_X = I_n - P_X,\ P_X = X(X'X)^{-1}X', \text{ and} \\ y_{2i} &= \Pi' Z_i + \xi' X_i + v_{2i}, \end{aligned} \tag{2.3}$$

where Z_i is the i-th row of Z written as a column and $\xi = \xi_1 + (X'X)^{-1}X'\widetilde{Z}\Pi$. By construction, $Z'X = 0$.

Substituting the reduced-form equations for y_{2i} into the structural equation for y_{1i} yields $m+1$ reduced-form equations:

$$\begin{aligned} y_{1i} &= \beta'\Pi' Z_i + \gamma' X_i + v_{1i} \text{ and} \\ y_{2i} &= \Pi' Z_i + \xi' X_i + v_{2i}, \text{ where} \\ v_{1i} &= u_i + \beta' v_{2i}, \end{aligned} \tag{2.4}$$

and $\gamma = \gamma_1 + \xi\beta$. The $m+1$ reduced-form equations also can be written as

$$\begin{aligned} y_i &= A\Pi' Z_i + \eta' X_i + v_i, \text{ where} \\ y_i &= (y_{1i}, y_{2i}')' \in R^{m+1}, \; v_i = (v_{1i}, v_{2i}')' \in R^{m+1}, \\ A &= \begin{bmatrix} \beta' \\ I_m \end{bmatrix} \in R^{(m+1)\times m}, \text{ and } \eta = [\gamma : \xi] \in R^{p\times(m+1)}. \end{aligned} \tag{2.5}$$

Let Y and Y_2 denote the $n \times (m+1)$ and $n \times m$ matrices whose i-th rows are y_i' and y_{2i}', respectively.

We make the following basic assumptions about the model. (Additional assumptions are given below.)

Assumption 1. (a) $\{(u_i, v_{2i}) : i \geq 1\}$ are iid random variables with mean zero.
(b) v_{2i} has nonsingular variance matrix $\Omega_{22} \in R^{m\times m}$.

Assumption 2. (a) $\{(\widetilde{Z}_i, X_i) : i \geq 1\}$ are fixed (i.e., non-random).
(b) The first element of X_i is 1 for all i.
(c) $n^{-1}\sum_{i=1}^n (\widetilde{Z}_i', X_i')'(\widetilde{Z}_i', X_i') \to D > 0$.
(d) $\max_{i\leq n}(||\widetilde{Z}_i||^2 + ||X_i||^2)/n \to 0$.

The combination of Assumptions 1 and 2(a) implies that the distribution of the errors $\{(u_i, v_{2i}) : i \geq 1\}$ does not depend on the IVs or regressors. In place of Assumption 2(a), one could treat the IVs and regressors as random. In this case, the IVs and regressors would be assumed to be independent of the errors. As is, Assumption 2(a) is consistent with random IVs and regressors provided one conditions on these variables.

Assumption 2(b) requires that the structural and reduced-form equations include an intercept. Given that $Z'X = 0$, this implies that $n^{-1}\sum_{i=1}^n Z_i = 0$. Assumptions 2(c) and 2(d) are standard assumptions concerning the behavior of IVs and regressors. They hold with probability one if $\{(\widetilde{Z}_i, X_i) : i \geq 1\}$ is a realization of an iid sequence with pd variance matrix and $2+\delta$ moments finite for some $\delta > 0$, see Lemma 12 in the Appendix.

We now define the CLR test of Moreira (2003), the LM test of Kleibergen (2002) and Moreira (2001), and the AR test. The CLR test depends on an LR test statistic coupled with a "conditional" critical value defined below. The LR, LM, and AR test statistics are based on the following statistics:[1]

$$\begin{aligned} S_n &= (Z'Z)^{-1/2} Z'Y b_0 \cdot (b_0'\widehat{\Omega}_n b_0)^{-1/2} \in R^k \text{ and} \\ T_n &= (Z'Z)^{-1/2} Z'Y \widehat{\Omega}_n^{-1} A_0 (A_0'\widehat{\Omega}_n^{-1} A_0)^{-1/2} \in R^{k\times m}, \text{ where} \\ b_0 &= \begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} \in R^{m+1}, \; A_0 = \begin{bmatrix} \beta_0' \\ I_m \end{bmatrix} \in R^{(m+1)\times m}, \\ \widehat{\Omega}_n &= (n-k-p)^{-1} Y' M_{[Z:X]} Y, \text{ and } M_{[Z:X]} = I_n - P_Z - P_X. \end{aligned} \tag{2.6}$$

[1] The statistics S_n and T_n are denoted $\overline{S}$ and $\overline{T}$, respectively, in Moreira (2003).

Note that $\widehat{\Omega}_n$ is an estimator of the variance matrix $\Omega = Ev_i v_i'$, which needs to be well-defined and positive definite in order for S_n and T_n to be well-behaved asymptotically. After proper centering, the statistics S_n and T_n have a joint multivariate normal asymptotic distribution with zero covariance under weak IV asymptotics under the null and the alternative. Hence, S_n and T_n are asymptotically independent.

The LR, LM, and AR test statistics depend on (S_n, T_n) in the following way:

$$\begin{aligned} LR_n &= S_n' S_n - \lambda_{\min}([S_n\!:\!T_n]'\,[S_n\!:\!T_n]), \\ LM_n &= S_n' T_n (T_n' T_n)^{-1} T_n' S_n, \text{ and} \\ AR_n &= S_n' S_n / k, \end{aligned} \tag{2.7}$$

where $\lambda_{\min}(C)$ denotes the minimum eigenvalue of the matrix C. When $m = 1$, LR_n can be written as

$$\begin{aligned} LR_n &= \frac{1}{2}\left(Q_{Sn} - Q_{Tn} + \sqrt{(Q_{Sn} - Q_{Tn})^2 + 4Q_{STn}^2}\right), \text{ where} \\ Q_{Sn} &= S_n' S_n,\ Q_{Tn} = T_n' T_n, \text{ and } Q_{STn} = S_n' T_n, \end{aligned} \tag{2.8}$$

see Moreira (2003) and Andrews and Stock (2005).[2]

The CLR test with asymptotic level α rejects the null hypothesis when

$$LR_n > \kappa_{LR,\alpha}(Q_{Tn}, k, m), \tag{2.9}$$

where $\kappa_{LR,\alpha}(\cdot, k, m)$ is a critical value function defined such that the CLR test has asymptotic null rejection rate α under weak IV asymptotics (under the assumptions above and $Eu_i^2 < \infty$). See (3.10) below for the definition of $\kappa_{LR,\alpha}(\cdot, k, m)$.

The LM statistic has a chi-squared asymptotic null distribution with m degrees of freedom, denoted χ_m^2, under weak and strong IVs (under the assumptions above and $Eu_i^2 < \infty$). Hence, the critical value for the asymptotic level α LM test is the $1 - \alpha$ quantile of a χ_m^2 distribution.

The AR statistic times k has a chi-squared asymptotic null distribution under weak and strong IVs with k $(\geq m)$ degrees of freedom (under the assumptions above and $Eu_i^2 < \infty$). Under the assumption of normal errors $\{v_i : i \geq 1\}$, it has an exact $F_{k,n-k-p}$ distribution. Thus, use of the $1 - \alpha$ quantile of an $F_{k,n-k-p}$ distribution as the critical value for the level α AR test is justified asymptotically for non-normal errors and yields an exact test for normal errors.

3 Rank CLR, LM, and AR Tests

In this section, we introduce rank analogues, S_n^φ and T_n^φ, of the statistics S_n and T_n, where φ is a score function defined below. By design, S_n^φ and T_n^φ are asymptotically independent. Given S_n^φ and T_n^φ, we define rank statistics that are analogous

[2] The statistic LR_n is the likelihood ratio statistic for the case of normal errors v_i with known covariance matrix Ω and with $\widehat{\Omega}_n$ plugged in for Ω. One can also consider the likelihood ratio statistic for the case of normal errors and unknown covariance matrix, see Moreira (2003).

to the CLR, LM, and AR statistics defined above. We show that for normal scores, i.e., $\varphi = \varphi^{NS}$, and multivariate normal errors (u_i, v_{2i}), S_n^φ and T_n^φ are asymptotically equivalent to S_n and T_n under weak IV and strong IV asymptotics under the null and the alternative. For non-normal errors, the rank tests have power advantages.

The statistic S_n depends on the inner product of Z and a vector of null-restricted residuals from the structural equation (2.1):

$$Z'Yb_0 = \sum_{i=1}^{n} Z_i(y_{1i} - \beta_0' y_{2i}) = \sum_{i=1}^{n} Z_i(y_{1i} - \beta_0' y_{2i} - \widehat{\gamma}_{1n}' X_i), \tag{3.1}$$

where $\widehat{\gamma}_{1n}$ is some estimator of γ_1 and the second equality holds because $Z'X = 0$. The rank analogue of S_n that we consider depends on the inner product of Z with the vector of ranks of $\{y_{1i} - \beta_0' y_{2i} - \widehat{\gamma}_{1n}' X_i : i \leq n\}$.

Let $\widehat{\gamma}_n(\beta_0)$ be some "null-restricted" estimator of γ_1. For example, one could use the least squares (LS) null-restricted estimator:

$$\widehat{\gamma}_n^{LS}(\beta_0) = (X'X)^{-1}X'Y(1, -\beta_0')'. \tag{3.2}$$

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

Let $\widehat{R}_i(\beta_0)$ be the rank of $y_{1i} - \beta_0' y_{2i} - \widehat{\gamma}_n(\beta_0)' X_i$ in $\{y_{1j} - \beta_0' y_{2j} - \widehat{\gamma}_n(\beta_0)' X_j : j = 1, ..., n\}$. The ranks $\{\widehat{R}_i(\beta_0) : i \leq n\}$ are referred to as *aligned* ranks.[3,4]

Let $\varphi : [0, 1) \to R$ be a non-stochastic score function. Different score functions φ lead to different rank statistics. Of primary interest are: (a) the normal (or van der Waerden) score function and (b) the Wilcoxon score function:

$$\text{(a) } \varphi^{NS}(x) = \Phi^{-1}(x) \text{ and (b) } \varphi^{WS}(x) = x, \tag{3.3}$$

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

$$c_\varphi = \int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx > 0, \text{ where } \overline{\varphi} = \int_0^1 \varphi(x) dx. \tag{3.4}$$

For normal scores, $c_\varphi = 1$. For Wilcoxon scores, $c_\varphi = 1/12$.

Let R_φ denote the n-vector whose i-th element is $\varphi(\widehat{R}_i(\beta_0)/(n+1))$. The rank analogue of S_n is

$$S_n^\varphi = (Z'Z)^{-1/2} Z' R_\varphi c_\varphi^{-1/2} \in R^k. \tag{3.5}$$

[3] If there are any ties in the ranks, then we determine a unique ranking by randomization. For example, if $y_{1i} - \beta' y_{2i} - \widehat{\gamma}_n(\beta)' X_i = y_{1j} - \beta' y_{2j} - \widehat{\gamma}_n(\beta)' X_j$ for some $i \neq j$ and these observations are the ℓ-th largest in the sample, then $\widehat{R}_i(\beta) = \ell$ with probability 0.5, $\widehat{R}_i(\beta) = \ell + 1$ with probability 0.5, $\widehat{R}_j(\beta) = \ell + 1$ if $\widehat{R}_i(\beta) = \ell$, and $\widehat{R}_j(\beta) = \ell$ if $\widehat{R}_i(\beta) = \ell + 1$. Ties only occur with positive probability if the distribution of $y_{1i} - \beta' y_{2i} - \widehat{\gamma}_n(\beta)' X_i$ is not absolutely continuous. In consequence, in practice ties rarely occur.

[4] The matrix programming languages GAUSS and Matlab have very fast built-in procedures for finding the ranks of a given vector. The GAUSS procedure is called *rankindx.*

The rank statistic S_n^φ replaces $Yb_0 \cdot (b_0'\widehat{\Omega}_n b_0)^{-1/2}$ in S_n by $R_\varphi c_\varphi^{-1/2}$. We want the rank analogue of T_n to do the same, as well as to be asymptotically independent of S_n^φ. In consequence, to construct a rank analogue of T_n, it is helpful to rewrite T_n as follows:

$$
\begin{aligned}
T_n &= (Z'Z)^{-1/2}Z'[Yb_0\widehat{\sigma}_n^{-1} : Y_2]\widehat{\Omega}_{*n}^{-1}H(H'\widehat{\Omega}_{*n}^{-1}H)^{-1/2}, \text{ where } \widehat{\sigma}_n^2 = b_0'\widehat{\Omega}_n b_0,\\
H &= \begin{bmatrix} 0_m' \\ I_m \end{bmatrix} \in R^{(m+1)\times m},\ \widehat{\Omega}_{*n} = [b_0\widehat{\sigma}_n^{-1} : H]'\widehat{\Omega}_n[b_0\widehat{\sigma}_n^{-1} : H] = \begin{bmatrix} 1 & \widehat{\nu}_n' \\ \widehat{\nu}_n & \widehat{\Omega}_{22n} \end{bmatrix},\\
\widehat{\Omega}_{22n} &= H'\widehat{\Omega}_n H = (n-k-p)^{-1}Y_2{}'M_{[Z:X]}Y_2 \in R^{m\times m}, \text{ and } \widehat{\nu}_n = H'\widehat{\Omega}_n b_0\widehat{\sigma}_n^{-1} \in R^m.
\end{aligned} \tag{3.6}
$$

(See (7.81) in the Appendix for a proof of (3.6).) As defined, $\widehat{\Omega}_{*n}$ is an estimator of the asymptotic variance matrix, Ω_*, of $n^{-1/2}\sum_{i=1}^n[b_0\widehat{\sigma}_{gn}^{-1} : H]'y_i = n^{-1/2}\sum_{i=1}^n(b_0'y_i\widehat{\sigma}_{gn}^{-1}, y_{2i}')'$. The definition of $\widehat{\Omega}_{*n}$ is chosen to yield asymptotic independence of S_n and T_n.

The rank analogue of T_n is[5]

$$
\begin{aligned}
T_n^\varphi &= (Z'Z)^{-1/2}Z'[R_\varphi c_\varphi^{-1/2}:Y_2]\widehat{\Omega}_{\varphi n}^{-1}H(H'\widehat{\Omega}_{\varphi n}^{-1}H)^{-1/2} \in R^{k\times m}, \text{ where}\\
\widehat{\Omega}_{\varphi n} &= \begin{bmatrix} 1 & \widehat{\nu}_{\varphi n}' \\ \widehat{\nu}_{\varphi n} & \widehat{\Omega}_{22n} \end{bmatrix} \text{ and } \widehat{\nu}_{\varphi n} = n^{-1}Y_2{}'M_{[Z:X]}R_\varphi c_\varphi^{-1/2} \in R^m.
\end{aligned} \tag{3.7}
$$

Note that $\widehat{\Omega}_{\varphi n}$ is an estimator of the asymptotic variance matrix of $n^{-1/2}\sum_{i=1}^n (\varphi(\widehat{R}_i(\beta_0)/(n+1))c_\varphi^{-1/2}, y_{2i}')'$. The definition of $\widehat{\Omega}_{\varphi n}$ ensures that S_n^φ and T_n^φ are asymptotically independent.

We define the rank LR, LM, and AR statistics to be

$$
\begin{aligned}
RLR_n^\varphi &= S_n^{\varphi\prime}S_n^\varphi - \lambda_{\min}([S_n^\varphi:T_n^\varphi]'\,[S_n^\varphi:T_n^\varphi]),\\
RLM_n^\varphi &= S_n^{\varphi\prime}T_n^\varphi(T_n^{\varphi\prime}T_n^\varphi)^{-1}T_n^{\varphi\prime}S_n^\varphi, \text{ and}\\
RAR_n^\varphi &= S_n^{\varphi\prime}S_n^\varphi/k.
\end{aligned} \tag{3.8}
$$

For $m=1$, the RLR_n^φ statistic simplifies as in (2.8) with $(S_n^\varphi, T_n^\varphi)$ in place of (S_n, T_n).

Notice that when $k = m$ (i.e., the structural equation is just identified), $k \cdot RAR_n^\varphi = S_n^{\varphi\prime}S_n^\varphi = RLM_n^\varphi = RLR_n^\varphi$.[6] That is, the rank CLR, LM, and AR tests are equivalent when $k=m$.

---

[5] The definition of T_n^φ uses the ranks R_φ of $\{y_{1j} - \beta_0'y_{2j} - \widehat{\gamma}_n(\beta_0)'X_j : j = 1, ..., n\}$, but is linear in Y_2 (or equivalently, in $Y_2 - P_XY_2$ since $Z'P_X = 0$). One might think that it is more natural to replace Y_2 in the definition of T_n^φ by the ranks of $Y_2 - P_XY_2$. We do not do this for the following reason. For power purposes one wants the Y_2 term in the definition of T_n^φ to be (asymptotically) linear in its mean $Z\Pi$. If one replaces Y_2 by the ranks of $Y_2 - P_XY_2$, then (asymptotic) linearity does not hold under strong IV asymptotics, defined in Section 5 below, because $Z\Pi$ is not an $n^{-1/2}$-perturbation from the zero vector, see Lemma 6 in the Appendix. Hence, one does not obtain the desired power properties under strong IV asymptotics. Under weak IV asymptotics, defined in Section 4 below, (asymptotic) linearity holds because $Z\Pi = ZCn^{-1/2}$ for some matrix C and the latter is an $n^{-1/2}$-perturbation from the zero vector. Hence, power problems with this alternative definition of T_n^φ only arise under strong IV asymptotics.

[6] The second equality holds because T_n^φ is a square invertible matrix when $k = m$. The last equality holds because $[S_n:T_n]'\,[S_n:T_n]$ is positive semi-definite and singular, which implies that $\lambda_{\min}([S_n:T_n]'\,[S_n:T_n]) = 0$. Singularity holds because $[S_n:T_n]'$ is an $(m+1)\times m$ matrix and $[S_n:T_n]'\,[S_n:T_n]$ is $(m+1)\times(m+1)$ when $k=m$.

The rank CLR, LM, and AR tests use the same critical values as the non-rank versions of these tests. Hence, the rank LM and AR tests with asymptotic significance level α have critical values given by the $1-\alpha$ quantiles of the χ^2_m and $F_{k,n-k-p}$ distributions, respectively.

The rank CLR test rejects the null hypothesis if

$$RLR^{\varphi}_n > \kappa_{LR,\alpha}(T^{\varphi\prime}_n T^{\varphi}_n, k, m), \tag{3.9}$$

where $\kappa_{LR,\alpha}(\cdot, k, m)$ is defined as follows. For $t \in R^{k\times m}$, define $\kappa_{LR,\alpha}(t't, k, m)$ via

$$\begin{aligned} &P\left(LR_\infty(S_0, t) > \kappa_{LR,\alpha}(t't, k, m)\right) = \alpha, \text{ where} \\ &S_0 \sim N(0, I_k) \text{ and } LR_\infty(s, t) = s's - \lambda_{\min}([s : t]'[s : t]) \end{aligned} \tag{3.10}$$

for $s \in R^k$. Note that $\kappa_{LR,\alpha}(\cdot, k, m)$ depends on k (the dimension of Z_i) and m (the dimension of y_{2i}). AMS2 provides detailed tables of $\kappa_{LR,\alpha}(\tau, k, m)$ for $m = 1$ and a variety of values of τ and k. Andrews, Moreira, and Stock (2006c) provide a GAUSS program for computing p-values of the CLR test for $m = 1$ and arbitrary k. This program also can be used for the rank CLR test by replacing AMS1's $\widehat{LR}_n$ and $\widehat{Q}_{T,n}$ statistics by RLR^{φ}_n and $T^{\varphi\prime}_n T^{\varphi}_n$, respectively.

For $m > 1$, the critical value function $\kappa_{LR,\alpha}(\cdot, k, m)$ can be simulated quite easily by simulating $S_0(r) \sim iid\ N(0, I_k)$ for $r = 1, ..., Reps$ and taking $\kappa_{LR,\alpha}(t't, k, m)$ to be the $1-\alpha$ sample quantile of $\{LR_\infty(S_0(r), t) : r = 1, ..., Reps\}$, where $Reps$ is a large integer, such as 25,000.

4 Weak IV Asymptotic Results

4.1 Weak IV Asymptotic Distributions of Rank Statistics

In this section, we establish the weak IV asymptotic distributions of the RLR^{φ}_n, RLM^{φ}_n, and RAR^{φ}_n test statistics under the null and fixed alternatives.

We assume the score function φ satisfies:

Assumption 3. **(a)** $\varphi : [0, 1) \to R$ is absolutely continuous and bounded with two derivatives that exist almost everywhere and are bounded.
(b) $0 < c_\varphi < \infty$ for c_φ defined in (3.4).

Assumption 3 holds for normal scores and Wilcoxon scores.

Under weak IVs, the asymptotic variance matrix, $\Omega_{\varphi g}$, of $n^{-1/2}\sum_{i=1}^n(\varphi(\widehat{R}_i(\beta_0)/(n+1)), y'_{2i})'$ is defined by

$$\begin{aligned} \Omega_{\varphi g} &= Var\begin{pmatrix} \varphi(U_{gi})c_\varphi^{-1/2} \\ y_{2i} \end{pmatrix} = \begin{bmatrix} 1 & \nu'_{\varphi g} \\ \nu_{\varphi g} & \Omega_{22} \end{bmatrix} \in R^{(m+1)\times(m+1)}, \text{ where} \\ U_{gi} &= G(u_i + (\beta-\beta_0)'v_{2i}) \in R,\ \nu_{\varphi g} = Cov(y_{2i}, \varphi(U_{gi})c_\varphi^{-1/2}) \in R^m, \end{aligned} \tag{4.1}$$

G is the df of $u_i + (\beta - \beta_0)'v_{2i}$, and g is the density corresponding to G.[7]

[7] $Var(\varphi(U_{gi})) = c_\varphi$ because U_{gi} has a $U[0, 1]$ distribution.

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

The weak IV assumption is the first part of the following assumption.

Assumption 4W. **(a)** $\Pi = Cn^{-1/2}$ for some matrix $C \in R^{k\times m}$.
(b) β does not depend on n.
(c) $u_i + (\beta - \beta_0)'v_{2i}$ has an absolutely-continuous strictly-increasing df G and an absolutely-continuous and bounded density g that satisfies $I(g) < \infty$.
(d) $(u_i + (\beta - \beta_0)'v_{2i}, v_{2i})$ has an absolutely-continuous bounded joint density with partial derivative with respect to its first argument that is bounded over both arguments.
(e) $\Omega_{\varphi g}$ is positive definite.
(f) $n^{1/2}(\widehat{\gamma}_n(\beta_0) - \gamma_1 - \xi_1(\beta - \beta_0)) = O_p(1)$.

Assumption 4W(b) implies that the data-generating process satisfies the null hypothesis or a fixed alternative. Assumptions 4W(c) and (d) require that $(u_i + (\beta - \beta_0)'v_{2i}, v_{2i})$ is absolutely-continuous, but otherwise are not very restrictive. Note that Assumptions 1-3 and 4W place no moment restrictions on u_i.

Assumption 4W(f) requires the null-restricted estimator $\widehat{\gamma}_n(\beta_0)$ to be well-behaved. It is satisfied by the LS estimator under the assumptions above if $Eu_i^2 < \infty$:

Lemma 1 *Under Assumptions* 1, 2, 4W(a), *and* 4W(b) *and* $Eu_i^2 < \infty$, $\widehat{\gamma}_n^{LS}(\beta_0)$ *satisfies Assumption* 4W(f).

We show that S_n^{φ} and T_n^{φ} converge in distribution to independent random quantities $S_\infty^{\varphi} \in R^k$ and $T_\infty^{\varphi} \in R^{k\times m}$, respectively, that are defined as follows. Let $D_Z \in R^{k\times k}$ be the probability limit of $n^{-1}Z'Z$:

$$D_Z = D_{11} - D_{12}D_{22}^{-1}D_{21},\ D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}, \tag{4.2}$$

where $D_{11} \in R^{k\times k}$, $D_{12} \in R^{k\times p}$, and $D_{22} \in R^{p\times p}$.

For a score function φ and a density f, define

$$\begin{aligned} \xi(\varphi, f) &= \frac{\left(\int_0^1 \varphi(x,f)\varphi(x)dx\right)^2}{\int_0^1 [\varphi(x) - \overline{\varphi}]^2 dx}, \text{ where} \\ \varphi(x,f) &= -\frac{f'(F^{-1}(x))}{f(F^{-1}(x))} \text{ for } x \in [0,1] \end{aligned} \tag{4.3}$$

and f' denotes the derivative of f. For normal and Wilcoxon scores,

$$\xi(\varphi^{NS}, f) = \left(\int \frac{f^2(x)}{\phi(\Phi^{-1}(F(x)))}dx\right)^2 \text{ and } \xi(\varphi^{WS}, f) = 12\left(\int f^2(x)dx\right)^2, \tag{4.4}$$

respectively, where ϕ and Φ denote the standard normal density and df and $F' = f$.[8]

[8] The expressions for $\xi(\varphi, f)$ for normal and Wilcoxon scores are established by change of variables and integration by parts.

Let $[N_\varphi : N_2]$ be a $k \times (m+1)$ multivariate normal matrix with

$$\begin{aligned} EN_\varphi &= D_Z C \ell^\varphi_{g,\beta-\beta_0} \in R^k, \text{ where } \ell^\varphi_{g,\beta-\beta_0} = (\beta-\beta_0)\xi^{1/2}(\varphi, g) \in R^m, \\ EN_2 &= D_Z C \in R^{k\times m}, \text{ and} \\ Var(vec([N_\varphi : N_2])) &= \Omega_{\varphi g} \otimes D_Z, \end{aligned} \tag{4.5}$$

where g is the density of $u_i + (\beta-\beta_0)' v_{2i}$, see Assumption 4W(c). Now, define

$$\begin{aligned} S^\varphi_\infty &= D_Z^{-1/2} N_\varphi \sim N(D_Z^{1/2} C \ell^\varphi_{g,\beta-\beta_0}, I_k) \in R^k, \\ T^\varphi_\infty &= D_Z^{-1/2}[N_\varphi : N_2]\Omega^{-1}_{\varphi g} H(H'\Omega^{-1}_{\varphi g} H)^{-1/2} \in R^{k\times m}, \text{ and} \\ vec(T^\varphi_\infty) &\sim N(vec(D_Z^{1/2} C[\ell^\varphi_{g,\beta-\beta_0} : I_m]\Omega^{-1}_{\varphi g} H(H'\Omega^{-1}_{\varphi g} H)^{-1/2}), I_{km}). \end{aligned} \tag{4.6}$$

Under H_0, S^φ_∞ has mean zero, but T^φ_∞ does not. It is shown below that the covariance of S^φ_∞ and T^φ_∞ is zero and, hence, these normal random variates are independent (under H_0 and H_1).

The following result holds under the null hypothesis and fixed (i.e., non-local) alternative hypotheses.

Theorem 1 *Under Assumptions* 1-3 *and* 4W,
(a) $(S^\varphi_n, T^\varphi_n) \to_d (S^\varphi_\infty, T^\varphi_\infty)$, *where* S^φ_∞ *and* T^φ_∞ *are independent,*
(b) $RLR^\varphi_n \to_d LR^\varphi_\infty := S^{\varphi\prime}_\infty S^\varphi_\infty - \lambda_{\min}([S^\varphi_\infty : T^\varphi_\infty]'[S^\varphi_\infty : T^\varphi_\infty])$,
(c) $RLM^\varphi_n \to_d S^{\varphi\prime}_\infty T^\varphi_\infty (T^{\varphi\prime}_\infty T^\varphi_\infty)^{-1} T^{\varphi\prime}_\infty S^\varphi_\infty$, *and*
(d) $RAR^\varphi_n \to_d S^{\varphi\prime}_\infty S^\varphi_\infty / k$.

Comments. 1. Theorem 1(d) shows that $k \cdot RAR^\varphi_n$ has an asymptotic χ^2_k distribution under the null and a $\chi^2_k(\delta^\varphi_{AR,W})$ distribution under fixed alternatives, where

$$\delta^\varphi_{AR,W} = (\beta-\beta_0)C' D_Z C(\beta-\beta_0) \cdot \xi(\varphi, g). \tag{4.7}$$

This justifies using the $1-\alpha$ quantile of the $F_{k,n-k-p}$ distribution as the critical value for the test based on RAR^φ_n because $F_{k,n-k-p} \to_d \chi^2_k/k$ as $n \to \infty$.

2. Theorem 1(a) and (c) imply that RLM^φ_n has an asymptotic χ^2_m distribution under the null hypothesis (because $S^\varphi_\infty \sim N(0_k, I_k)$ under the null implies that $S^{\varphi\prime}_\infty T^\varphi_\infty (T^{\varphi\prime}_\infty T^\varphi_\infty)^{-1} T^{\varphi\prime}_\infty S^\varphi_\infty$ has a χ^2_m distribution conditional on T^φ_∞ and, hence, an unconditional χ^2_m distribution as well). Under the alternative, *conditional on* $P_{T^\varphi_\infty}$ $(= T^\varphi_\infty (T^{\varphi\prime}_\infty T^\varphi_\infty)^{-1} T^{\varphi\prime}_\infty)$, RLM^φ_n has a noncentral chi-squared distribution, $\chi^2_m(\delta^\varphi_{LM,W})$, with m degrees of freedom and noncentrality parameter

$$\delta^\varphi_{LM,W} = (\beta-\beta_0)C' D_Z^{1/2} T^\varphi_\infty (T^{\varphi\prime}_\infty T^\varphi_\infty)^{-1} T^{\varphi\prime}_\infty D_Z^{1/2} C(\beta-\beta_0) \cdot \xi(\varphi, g). \tag{4.8}$$

The random projection matrix $P_{T^\varphi_\infty}$ equals $P_{T^\varphi_\infty M}$, where M is any random or nonrandom nonsingular $m \times m$ matrix. In consequence, $P_{T^\varphi_\infty}$ has the same distribution as $P_{T^*_\infty}$, where $vec(T^*_\infty) \sim N(vec(D_Z^{1/2} C), I_{km})$. Note that the distribution of T^*_∞ does not depend on φ or g. Hence, the asymptotic distribution of RLM^φ_n only depends on (φ, g) through the distribution of S^φ_∞.

3. The statistics RLR_n^{φ} and RLM_n^{φ} and their asymptotic distributions depend on $(S_n^{\varphi}, T_n^{\varphi})$ and $(S_{\infty}^{\varphi}, T_{\infty}^{\varphi})$ only through $Q_n^{\varphi} = [S_n^{\varphi} : T_n^{\varphi}]'[S_n^{\varphi} : T_n^{\varphi}]$ and $Q_{\infty}^{\varphi} = [S_{\infty}^{\varphi} : T_{\infty}^{\varphi}]'[S_{\infty}^{\varphi} : T_{\infty}^{\varphi}]$, respectively. Given the multivariate normal distribution of $[S_{\infty}^{\varphi} : T_{\infty}^{\varphi}]$, Q_{∞}^{φ} has a noncentral Wishart distribution. It depends on unknown parameters only through

$$\begin{aligned}
&[ES_{\infty}^{\varphi} : ET_{\infty}^{\varphi}]'[ES_{\infty}^{\varphi} : ET_{\infty}^{\varphi}], \text{ where} \\
&[ES_{\infty}^{\varphi} : ET_{\infty}^{\varphi}] = D_Z^{1/2} C \left[\ell_{g,\beta-\beta_0}^{\varphi} : [\ell_{g,\beta-\beta_0}^{\varphi} : I_m] \Omega_{\varphi g}^{-1} H (H' \Omega_{\varphi g}^{-1} H)^{-1/2} \right]. \qquad (4.9)
\end{aligned}$$

The following Corollary uses Theorem 1(a) and (b) to show that the use of $\kappa_{LR,\alpha}(\tau, k, m)$ (defined in (3.10)) as the critical value function for the RLR_n^{φ} statistic yields a test with asymptotic null rejection rate α under weak IV asymptotics.

Corollary 1 *Under the null hypothesis, $H_0 : \beta = \beta_0$, and Assumptions* 1-3 *and* 4W, $\lim_{n \to \infty} P\left(RLR_n^{\varphi} > \kappa_{LR,\alpha}(T_n^{\varphi\prime} T_n^{\varphi}, k, m)\right) = \alpha$.

4.2 Weak IV Asymptotic Distributions of Non-Rank Statistics

To enable comparisons of the power of rank and non-rank tests, we now provide the null and non-null weak IV asymptotic distributions of the non-rank statistics S_n and T_n under the assumption that $\Omega = E v_i v_i'$ is well-defined and positive definite. The results given here extend results in AMS1 from $m = 1$ to $m \geq 1$. They are not covered by Moreira (2003), because Moreira (2003) only provides asymptotic results under the null hypothesis.

To make comparisons of rank and non-rank tests more transparent, we write the asymptotic distributions of the non-rank tests in a form that is analogous to that of S_{∞}^{φ} and T_{∞}^{φ}, which differs from the form given in AMS1. Define

$$\begin{aligned}
\Omega_g &= Var \begin{pmatrix} y_i' b_0 \sigma_g^{-1} \\ y_{2i} \end{pmatrix} = Var \begin{pmatrix} (u_i + (\beta - \beta_0)' v_{2i}) \sigma_g^{-1} \\ v_{2i} \end{pmatrix} \\
&= \left[b_0 \sigma_g^{-1} : H \right]' \Omega \left[b_0 \sigma_g^{-1} : H \right] = \begin{bmatrix} 1 & \nu_g' \\ \nu_g & \Omega_{22} \end{bmatrix}, \\
\sigma_g^2 &= Var(y_i' b_0) = Var(u_i + (\beta - \beta_0)' v_{2i}) = b_0' \Omega b_0, \text{ and} \\
\nu_g &= Cov(y_{2i}, (u_i + (\beta - \beta_0)' v_{2i}) \sigma_g^{-1}) = H' \Omega b_0 \sigma_g^{-1}. \qquad (4.10)
\end{aligned}$$

Let $[N_1 : N_2]$ be a $k \times (m+1)$ multivariate normal matrix with N_2 as above,

$$\begin{aligned}
&EN_1 = D_Z C (\beta - \beta_0) \sigma_g^{-1} \in R^k, \text{ and} \\
&Var(vec([N_1 : N_2])) = \Omega_g \otimes D_Z. \qquad (4.11)
\end{aligned}$$

Next, define

$$\begin{aligned}
S_{\infty} &= D_Z^{-1/2} N_1 \sim N(D_Z^{1/2} C \ell_{g,\beta-\beta_0}, I_k), \\
T_{\infty} &= D_Z^{-1/2} [N_1 : N_2] \Omega_g^{-1} H (H' \Omega_g^{-1} H)^{-1/2} \in R^{k \times m}, \\
vec(T_{\infty}) &\sim N(vec(D_Z^{1/2} C [\ell_{g,\beta-\beta_0} : I_m] \Omega_g^{-1} H (H' \Omega_g^{-1} H)^{-1/2}), I_{km}), \text{ and} \\
\ell_{g,\beta-\beta_0} &= (\beta - \beta_0) \sigma_g^{-1} \in R^m. \qquad (4.12)
\end{aligned}$$

Lemma 2 *Under Assumptions* 1-3 *and* 4W *and* $\Omega > 0$,
(a) $(S_n, T_n) \to_d (S_\infty, T_\infty)$, *where* S_∞ *and* T_∞ *are independent,*
(b) $LR_n \to_d S_\infty' S_\infty - \lambda_{\min}([S_\infty : T_\infty]' [S_\infty : T_\infty])$,
(c) $LM_n \to_d S_\infty' T_\infty (T_\infty' T_\infty)^{-1} T_\infty' S_\infty$, *and*
(d) $AR_n \to_d S_\infty' S_\infty / k$.

Comments. 1. Lemma 2(d) shows that $k \cdot AR_n$ has an asymptotic χ^2_k distribution under the null and a $\chi^2_k(\delta_{AR,W})$ distribution under fixed alternatives, where

$$\delta_{AR,W} = (\beta - \beta_0) C' D_Z C (\beta - \beta_0) \cdot \sigma_g^{-2}. \tag{4.13}$$

2. Lemma 2(a) and (c) imply that LM_n has an asymptotic χ^2_m distribution under the null hypothesis. Under the alternative, *conditional on* P_{T_∞}, LM_n has an asymptotic noncentral chi-squared distribution, $\chi^2_m(\delta_{LM,W})$, with m degrees of freedom and noncentrality parameter

$$\delta_{LM,W} = (\beta - \beta_0) C' D_Z^{1/2} T_\infty (T_\infty' T_\infty)^{-1} T_\infty' D_Z^{1/2} C (\beta - \beta_0) \cdot \sigma_g^{-2}. \tag{4.14}$$

4.3 Weak IV Power Comparisons: Rank Versus Non-rank Tests

In this section, we compare the weak IV asymptotic power of the rank AR, LM, and CLR tests to that of the non-rank versions of these tests. We consider the AR and LM tests first because the comparison is simpler for these tests.

4.3.1 Anderson-Rubin and Lagrange Multiplier Tests

The RAR_n^φ and AR_n statistics have noncentral chi-squared distributions under weak IV asymptotics by Comment 1 to Theorem 1 and Comment 1 to Lemma 2(d). Their noncentrality parameters, given in (4.7) and (4.13), respectively, differ only by the multiplicative constants $\xi(\varphi, g)$ and σ_g^{-2}. In consequence, for weak IVs, the *asymptotic relative efficiency*[9] (ARE) of the rank AR test to the (non-rank) AR test is

$$ARE_g(RAR_n^\varphi, AR_n) = \xi(\varphi, g) \sigma_g^2. \tag{4.15}$$

(An ARE greater than one means that the rank AR test has higher power than the AR test.) Note that the ARE in (4.15) is independent of the location and scale of g.

[9] The ARE of one test to another is usually defined, roughly speaking, to be the limit of the ratio of the sample sizes of the second test to the first required for the two tests to have the same power, see Lehmann (1986, p. 321). In standard scenarios—in which the two tests have noncentral chi-square asymptotic distributions—the ARE reduces to the ratio of the (asymptotic) noncentrality parameter of the first test to the second. In the present section, which involves non-standard weak IV asymptotics—in which the power of a test does not necessarily increase with the sample size—we adopt the ratio of the (asymptotic) noncentrality parameters to be the definition of the ARE. That is, by definition, the ARE of one test to another is the ratio of the noncentrality parameter of the asymptotic distribution of the first test to that of the second test provided this ratio is nuisance parameter free and the two tests have noncentral chi-square asymptotic distributions or mixed noncentral chi-square asymptotic distributions (and the ratio of the noncentrality parameters is the same for all values of the mixing variable).

When $k = m$, the ARE in (4.15) also applies to the rank versus non-rank CLR and LM tests because they are the same as the AR tests.

The RLM_n^{φ} and LM_n statistics have noncentral chi-squared distributions under weak IV asymptotics conditional on $P_{T_\infty^\varphi}$ and P_{T_∞}, respectively, by Comment 2 to Theorem 1 and Comment 2 to Lemma 2(d). Note that the distributions of P_{T_∞} and $P_{T_\infty^\varphi}$ are equal by the argument given in Comment 2 to Theorem 1. In consequence, the ARE of the RLM_n^{φ} test to the LM_n test is the same as that of the rank to non-rank AR test given in (4.15).

The literature on rank tests contains extensive calculations of the ARE in (4.15) because exactly the same ARE arises when comparing a rank test with the usual t-test in a simple location model with error density g. In addition, it is the same as the ARE of a rank estimator with the sample mean in the location model.

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

$$\begin{aligned} ARE_g^{NS} &= \xi(\varphi^{NS}, g)\sigma_g^2 = \sigma^2(g)\left(\int \frac{g^2(x)}{\phi(\Phi^{-1}(G(x)))}dx\right)^2, \text{ where} \\ ARE_g^{NS} &= ARE_g(RAR_n^{NS}, AR_n) = ARE_g(RLM_n^{NS}, LM_n) \end{aligned} \quad (4.16)$$

and $G(\cdot)$ of the df of g. A result due to Chernoff and Savage (1958) implies that $ARE_g^{NS} \geq 1$ for all symmetric distributions g (about some point not necessarily zero). Hence, the asymptotic power under weak IVs of the normal scores rank AR (LM) test is greater than or equal to that of the non-rank AR (LM) test for any symmetric distribution.

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

$$\begin{aligned} ARE_g^{WS} &= \xi(\varphi^{WS}, g)\sigma_g^2 = 12\sigma_g^2\left(\int g^2(x)dx\right)^2, \text{ where} \\ ARE_g^{WS} &= ARE_g(RAR_n^{WS}, AR_n) = ARE_g(RLM_n^{WS}, LM_n). \end{aligned} \quad (4.17)$$

For the normal distribution, i.e., $g = \phi$, $ARE_\phi^{WS} = .955$. For the double exponential distribution g_{de}, $ARE_{g_{de}}^{WS} = 1.50$. For a contaminated normal distribution $\phi_\varepsilon(x) = (1-\varepsilon)\phi(x) + \varepsilon\phi(x/3)/3$, $ARE_{\phi_\varepsilon}^{WS} = 1.196$, 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 $ARE_g^{WS} \geq .864$ for all symmetric distributions g (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 g_1 and g_2 symmetric about zero (with df's G_1 and G_2), $ARE_{g_1}(RAR_n^{WS}, RAR_n^{NS}) \leq ARE_{g_2}(RAR_n^{WS}, RAR_n^{NS})$ whenever the tails of g_1 are lighter than the tails of g_2 in the sense that $G_2^{-1}(G_1(x))$ is convex for $x \geq 0$, see Thm. 2.9.5 of Hettmansperger (1984, p. 116). (The same is true with AR replaced by LM.) Thus, the comparative power of Wilcoxon scores to normal scores tests increases as the tail thickness of the distribution increases. For any symmetric density g, $ARE_g(RAR_n^{WS}, RAR_n^{NS}) \in (0, 1.91)$, see Hettsmansperger (1984, Thm. 2.9.3, p. 115).

4.3.2 Conditional Likelihood Ratio Test

Next, we compare the weak IV asymptotic power of the rank CLR and (non-rank) CLR tests. Analytical comparisons are difficult because of the complicated form of the asymptotic distributions. However, the power of these tests comes primarily from the magnitude of the means of S_∞^φ and S_∞, respectively, see (4.6) and (4.12). Hence, when $u_i + (\beta - \beta_0)'v_{2i}$ has relatively heavy tails, the rank CLR test should have higher power. Furthermore, as discussed in Andrews and Stock (2005), the CLR test is a data-dependent combination of the AR and LM tests and, hence, the advantage of the rank versions of the latter tests when $u_i + (\beta - \beta_0)'v_{2i}$ has relatively heavy or thin tails should carry over to that of the CLR rank test.

These conjectures are shown to hold (in the scenarios considered) by numerical comparisons of the asymptotic power of the $RCLR_n$ and CLR_n tests using the asymptotic results of Theorem 1(b), Corollary 1, and Lemma 2(b). Table I reports the weak IV asymptotic powers of the WS-RCLR, NS-RCLR, and CLR tests. For comparative purposes, asymptotic powers of the LM and AR tests also are given in Table I.

The cases considered in Table I include a Base Case and several variations of it. The Base Case has $m = 1$ (i.e., β is a scalar), $\lambda = C'D_ZC = 10$ (which corresponds to moderately weak IVs), $k = 5$ (i.e., five IVs), $\rho_{uv_2} = Corr(u_i, v_{2i}) = 0.75$ (which corresponds to moderately strong endogeneity), and $\beta_0 = 0$ (without loss of generality). Two values of β are considered viz., $\beta = 1$ and $\beta = -0.43$. These values are selected so that the CLR test has asymptotic power 0.40 with normal errors (u_i, v_{2i}). A "High Endogeneity" case is the same as the Base Case except that $\rho_{uv_2} = Corr(u_i, v_{2i}) = 0.95$ and $\beta = 1.1$ or $\beta = -0.37$. A "Weaker IV" case is the same as the Base Case except that $\lambda = 4.0$ and $\beta = 5.0$ or $\beta = -0.7$. A "Ten IV" case is the same as the Base Case except $k = 10$. In each variation of the Base Case, the values of β considered are chosen so that the CLR test has asymptotic power approximately equal to 0.40 with normal errors.

In all cases considered, the structural error u_i and a latent variable ε_i are taken to be independent with distribution F. We consider four distributions F, viz., standard normal, uniform $[-2\sqrt{3}, 2\sqrt{3}]$, t_3, and difference of independent log-normals (DLN). The uniform distribution exhibits thin tails, whereas the t_3 and DLN distributions exhibit thick tails. The reduced-form error v_{2i} is defined to be the following function of u_i and ε_i:

$$v_{2i} = (1 - \rho_{uv_2}^2)^{1/2}\varepsilon_i + \rho_{uv_2}u_i. \qquad (4.18)$$

By construction, $Corr(u_i, v_{2i}) = \rho_{uv_2}$. The distribution G, upon which the asymptotic properties of the tests depend, is the distribution of $u_i + (\beta - \beta_0)'v_{2i}$ when u_i and ε_i are independent with distribution F. When F has thin or thick tails, so does G.

Details concerning the computation of the asymptotic power reported in Table I are given in the Appendix.

Table I indicates that for the normal distribution F the WS-RCLR, NS-RCLR, and CLR tests have roughly equal asymptotic power in all cases. (This is analogous to the result in Section 4.3.1 that $ARE_{\phi}^{NS} = 1$ and $ARE_{\phi}^{WS} = 0.955$.) For the (thin-tailed) uniform distribution, the NS-RCLR test has higher power than the CLR test, whereas the WS-CLR test has lower power in all cases. (The former is analogous to the result in Section 4.3.1 that $ARE_{g}^{NS} \geq 1$ for all symmetric distributions g. The latter is analogous to the result in Section 4.3.1 that $ARE_g\ (RAR_n^{WS}, RAR_n^{NS}) \leq ARE_{\phi}(RAR_n^{WS}, RAR_n^{NS})$ for any distribution g that has thinner tails than ϕ.) For the (thick-tailed) t_3 and DLN distributions, the WS-RCLR and NS-RCLR tests have noticeably higher power than the CLR test except in one case (viz. the "Weaker IV" case with positive β and t_3 distribution). In the Base Case, for the t_3 distribution, the rank CLR tests' powers are 33% higher or more than the non-rank CLR test. In the Base Case, for the DLN distribution, the rank tests' powers are more than 50% higher. (This is analogous to the results in Section 4.3.1 that $ARE_{g}^{NS} \geq 1$ for all symmetric distributions g and $ARE_g\ (RAR_n^{WS}, RAR_n^{NS}) \geq ARE_{\phi}(RAR_n^{WS}, RAR_n^{NS})$ for any distribution g that has thicker tails than ϕ.)

Table I shows that the NS-RCLR and WS-RCLR tests cannot be rank ordered in an overall sense because the NS-RCLR test has noticeably higher power for the uniform distribution, but lower power for the t_3 and DLN distributions. Table I also shows that the AR test has lower asymptotic power than the other tests considered (because $k = 5 > m = 1$ or $k = 10 > m = 1$). Also, the LM test has comparable asymptotic power to the CLR test in the scenarios considered except the "Weaker IV" case with negative β, in which case it has lower power.

We conclude from Table I that the WS-RCLR and NS-RCLR tests have weak IV asymptotic power advantages over the CLR test. For the NS-CLR test, this is true both for thin- and thick-tailed distributions. Furthermore, there is little or no cost asymptotically for using the WS-RCLR or NS-RCLR test in place of the CLR test for the normal distribution. Since it is shown in AMS1 that the CLR test is nearly asymptotically UMP in the class of invariant similar tests under normality, the results suggest that the NS-CLR test also inherits this property.

4.3.3 Asymptotic Equivalence

We now provide a result that establishes when the rank and non-rank versions of the CLR, LM, and AR tests are asymptotically equivalent. We show that for a given score function $\varphi(x)$ there is a distribution G of $u_i + (\beta - \beta_0)'v_{2i}$ (and vice versa) such that the rank and non-rank versions of these tests are asymptotically equivalent under weak IV asymptotics.

Lemma 3 *Let $L(\cdot)$ be some df with finite variance. Suppose $(u_i + (\beta - \beta_0)'v_{2i})\kappa \sim L(\cdot)$ for some $\kappa > 0$ and $\varphi(x) = L^{-1}(x)$, then*
(a) $\varphi(U_{gi})c_{\varphi}^{-1/2} = (u_i + (\beta - \beta_0)'v_{2i})\sigma_g^{-1}$,
(b) $\Omega_{\varphi g} = \Omega_g$,

(c) $\int_0^1 \varphi(x,g)\varphi(x)dx \cdot c_\varphi^{-1/2} = \sigma_g^{-1}$, *and*
(d) $N_\varphi \sim N_1$, $S_\infty^\varphi \sim S_\infty$, *and* $T_\infty^\varphi \sim T_\infty$, *where "$\sim$" denotes "has the same distribution as."*

Comment. Lemmas 2 and 3 and Theorem 1 imply that if $u_i + (\beta - \beta_0)' v_{2i}$ has a normal distribution, then the normal score function leads to asymptotic equivalence between the rank and non-rank versions of the CLR, LM, and AR tests. Likewise, if $u_i + (\beta - \beta_0)' v_{2i}$ has a uniform $[-a, a]$ distribution for some $a > 0$, then the Wilcoxon score function leads to asymptotic equivalence between the rank and non-rank versions of these statistics.

5 Strong IV Asymptotic Results

5.1 Strong IV Asymptotic Distributions of Rank Statistics

In this section, we provide the asymptotic distributions of the RLR_n^φ, RLM_n^φ, and RAR_n^φ test statistics under standard strong IV asymptotics under the null hypothesis and local alternatives.

In place of Assumption 4W, we use the following assumption. The first part of this assumption is the local alternative assumption.

Assumption 4S. (a) $\beta = \beta_0 + Bn^{-1/2}$ for some vector $B \in R^m$.
(b) Π does not depend on n and is full column rank m.
(c) $v_{2i} = \varepsilon_i + \rho u_i$ for $i \geq 1$, where ε_i is a random m-vector and $\rho \in R^m$ is a vector of constants.
(d) $\{\varepsilon_i : i \geq 1\}$ are iid and independent of $\{u_i : i \geq 1\}$, and $E\|\varepsilon_i\|^{2+\delta} < \infty$ 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 $I(f) < \infty$.
(f) (u_i, v_{2i}) has an absolutely-continuous bounded joint density with partial derivative with respect to its first argument that is bounded over both arguments.
(g) $\Omega_{\varphi f}$ is positive definite.
(h) $\sum_{i=1}^\infty ||\widetilde{Z}_i||^2/i^2 < \infty$ and $\sum_{i=1}^\infty ||X_i||^2/i^2 < \infty$.
(i) $n^{1/2}(\widehat{\gamma}_n(\beta_0) - \gamma_1) = O_p(1)$.

Assumption 4S(c) allows for dependence between the structural error u_i and the reduced-form error v_{2i}, but it must be of a special form. The special form is needed to make the asymptotic results for the rank statistic S_n^φ tractable. Assumption 4S(h) is not very restrictive.[10] Assumption 4S(i) holds for the null-restricted LS estimator under Assumptions 1, 2, and 4S(a)-4S(c).[11] The combination of Assumptions 1 and 4S(c) implies that $Eu_i^2 < \infty$.

[10] A sufficient condition for Assumption 4S(h) is the same condition with 2 replaced by $1+\delta$ and the latter holds with probability one for sequences $\{(\widetilde{Z}_i, X_i) : i \geq 1\}$ that are realizations of iid random vectors with finite $1+\delta$ moments, see Lemma 12 in the Appendix.

[11] The proof is the same as for Lemma 1 except that in place of (7.69) we have $n^{1/2}(\xi - \xi_1)(\beta - \beta_0) = O(1)$ because $\beta - \beta_0 = O(n^{-1/2})$ by Assumption 4S(a) and $\xi - \xi_1 = O(1)$ by Assumptions 2(c) and 4S(b).

Under strong IV asymptotics, S_n^φ has a nondegenerate asymptotic distribution given by that of $S_{f\infty}^\varphi$, and $n^{-1/2}T_n^\varphi$ converges in probability to a constant $\alpha_T^\varphi \neq 0$, where

$$\begin{aligned} S_{f\infty}^\varphi &\sim N(\alpha_S^\varphi, I_k),\ \alpha_S^\varphi = D_Z^{1/2}\Pi\ell_{f,B}^\varphi \in R^k, \\ \alpha_T^\varphi &= D_Z^{1/2}\Pi(H'\Omega_{\varphi f}^{-1}H)^{1/2} \in R^{k\times m}, \\ \Omega_{\varphi f} &= Var\left(\begin{array}{c} \varphi(F(u_i))c_\varphi^{-1/2} \\ y_{2i} \end{array}\right) = \left[\begin{array}{cc} 1 & \nu_{\varphi f}' \\ \nu_{\varphi f} & \Omega_{22} \end{array}\right] \in R^{(m+1)\times(m+1)},\ \text{and} \\ \nu_{\varphi f} &= Cov(y_{2i}, \varphi(F(u_i))c_\varphi^{-1/2}) \in R^m. \end{aligned} \tag{5.1}$$

Note that $S_{f\infty}^\varphi$ differs from S_∞^φ only in that $\ell_{f,B}^\varphi$ replaces $\ell_{g,\beta-\beta_0}^\varphi$ (both of which are defined in (4.12)) in its mean.

The main result of this section is the following.

Theorem 2 *Under Assumptions* 1-3 *and* 4S,
(a) $(S_n^\varphi, n^{-1/2}T_n^\varphi) \to_d (S_{f\infty}^\varphi, \alpha_T^\varphi)$,
(b) $RLR_n^\varphi \to_d S_{f\infty}^{\varphi\prime}\alpha_T^\varphi(\alpha_T^{\varphi\prime}\alpha_T^\varphi)^{-1}\alpha_T^{\varphi\prime}S_{f\infty}^\varphi \sim \chi_m^2(\delta_{LM,S}^\varphi)$, *where*
$\delta_{LM,S}^\varphi = \alpha_S^{\varphi\prime}\alpha_T^\varphi(\alpha_T^{\varphi\prime}\alpha_T^\varphi)^{-1}\alpha_T^{\varphi\prime}\alpha_S^\varphi$,
(c) $RLM_n^\varphi \to_d S_{f\infty}^{\varphi\prime}\alpha_T^\varphi(\alpha_T^{\varphi\prime}\alpha_T^\varphi)^{-1}\alpha_{T\infty}^{\varphi\prime}S_{f\infty}^\varphi \sim \chi_m^2(\delta_{LM,S}^\varphi)$, *and*
(d) $RAR_n^\varphi \to_d S_{f\infty}^{\varphi\prime}S_{f\infty}^\varphi/k \sim \chi_k^2(\delta_{AR,S}^\varphi)/k$, *where* $\delta_{AR,S}^\varphi = \alpha_S^{\varphi\prime}\alpha_S^\varphi$.

Comments. 1. Theorem 2(b) and (c) show that under strong IV asymptotics the RLR and RLM test statistics are asymptotically equivalent under the null and local alternatives for any values of k and m. (As noted above, when $k = m$, the RLR and RLM test statistics are the same, so the tests are trivially asymptotically equivalent.)

2. Theorem 2(b)-(d) shows that the RAR test statistic has a different asymptotic distribution from that of the RLR and RLM statistics when $k > m$. When $k = m$, $k \cdot RAR_n^\varphi = RLM_n^\varphi = RLR_n^\varphi$, so the three rank statistics are asymptotically equivalent.

5.2 Strong IV Asymptotic Distributions of Non-Rank Statistics

For comparative purposes, we now provide the strong IV asymptotic distributions under the null hypothesis and local alternatives of the non-rank LR_n, LM_n, and AR_n test statistics. The results for LR_n with $m > 1$ are new. (AMS1 provides the same results for $m = 1$.) Let

$$\begin{aligned} S_{f\infty} &\sim N(\alpha_S, I_k),\ \alpha_S = D_Z^{1/2}\Pi B\sigma_f^{-1} \in R^k, \\ \alpha_T &= D_Z^{1/2}\Pi(H'\Omega_f^{-1}H)^{1/2} \in R^{k\times m},\ \text{and} \\ \Omega_f &= Var((u_i\sigma_f^{-1}, v_{2i})') = \left[\begin{array}{cc} 1 & \nu_f' \\ \nu_f & \Omega_{22} \end{array}\right],\ \nu_f = Cov(y_{2i}, u_i\sigma_f^{-1}). \end{aligned} \tag{5.2}$$

Lemma 4 *Under Assumptions* 1-3 *and* 4S *and* $\Omega > 0$,
(a) $(S_n, n^{-1/2}T_n) \rightarrow_d (S_{f\infty}, \alpha_T)$,
(b) $LR_n \rightarrow_d S'_{f\infty}\alpha_T(\alpha'_T\alpha_T)^{-1}\alpha'_T S_{f\infty} \sim \chi^2_m(\delta_{LM,S})$, *where* $\delta_{LM,S} = \alpha'_S\alpha_T(\alpha'_T\alpha_T)^{-1}\alpha'_T\alpha_S$,
(c) $LM_n \rightarrow_d S'_{f\infty}\alpha_T(\alpha'_T\alpha_T)^{-1}\alpha'_T S_{f\infty} \sim \chi^2_m(\delta_{LM,S})$, *and*
(d) $AR_n \rightarrow_d S'_{f\infty}S_{f\infty}/k \sim \chi^2_k(\delta_{AR,S})/k$, *where* $\delta_{AR,S} = \alpha'_S\alpha_S$.

5.3 Strong IV Power Comparisons: Rank Versus Non-rank Tests

Theorem 2 and Lemma 4 allow calculation of the ARE of the rank and non-rank tests with strong IVs. The calculation is analogous to that given in Section 4.3.1 for weak IVs, but with three differences. The first difference is that α^φ_T and α_T are fixed in the strong IV case, whereas T^φ_∞ and T_∞ are random in the weak IV case. This does not affect the ARE calculations. The second difference is that the asymptotic distributions depend on the density f of u_i rather than the density g of $u_i + (\beta - \beta_0)'v_{2i}$. This occurs because β converges to β_0 under strong IV local alternatives and, hence, $(\beta - \beta_0)'v_{2i} \rightarrow 0$ as $n \rightarrow \infty$. The third difference is that under strong IVs the asymptotic distributions of RLR^φ_n and RLM^φ_n are the same and, analogously, those of LR_n and LM_n are the same.

Combining the results of Section 4.3.1 with these differences, we find that under strong IVs the ARE of the rank to non-rank AR tests is the same as for the rank to non-rank LM and CLR tests and is equal to the usual ARE for rank to non-rank procedures based on the density f. That is,

$$ARE_f(RAR^\varphi_n, AR_n) = ARE_f(RLM^\varphi_n, LM_n) = ARE_f(RLR^\varphi_n, LR_n) = \xi(\varphi^{NS}, f)\sigma^2_f, \tag{5.3}$$

where $\xi(\varphi^{NS}, f)\sigma^2_f$ is given in (4.16) and (4.17) for normal and Wilcoxon scores, respectively, with f in place of g.[12]

In sum, all of the statements in Section 4.3.1 concerning (4.15) apply to the ARE of the rank to non-rank versions of the AR, LM, and CLR tests under strong IVs, but with f in place of g.

5.4 Asymptotic Equivalence

The next result establishes when the rank and non-rank versions of the CLR, LM, and AR tests are asymptotically equivalent under strong IV asymptotics.

Lemma 5 *Let* $L(\cdot)$ *be some df with finite variance. Suppose* $u_i\kappa \sim L(\cdot)$ *for some* $\kappa > 0$ *and* $\varphi(x) = L^{-1}(x)$, *then*
(a) $\varphi(F(u_i))c_\varphi^{-1/2} = u_i\sigma_f^{-1}$,

[12] The AREs discussed in this section can be defined by the usual method involving the limit of ratios of sample sizes or in terms of the ratio of noncentrality parameters—see the footnote in Section 4.3.1 regarding these definitions. Under strong IV asymptotics, the two definitions are equivalent for the tests considered here.

(b) $\Omega_{\varphi f} = \Omega_f$,
(c) $\int_0^1 \varphi(x,f)\varphi(x)dx \cdot c_\varphi^{-1/2} = \sigma_f^{-1}$,
(d) $S_{f\infty}^{\varphi} \sim S_{f\infty}$, *and* $\alpha_T^{\varphi} = \alpha_T$.

Comments. 1. Lemmas 4 and 5 and Theorem 2 imply that if u_i has a normal distribution, then the normal score function leads to asymptotic equivalence between the rank and non-rank versions of the CLR, LM, and AR tests. Likewise, if u_i has a uniform $[-a, a]$ distribution for some $a > 0$, then the Wilcoxon score function leads to asymptotic equivalence between these statistics.

2. For the case of normal errors, the (non-rank) CLR and LM tests are asymptotically efficient under strong IV asymptotics, see AMS1. This combined with Comment 1 implies that the normal scores rank CLR and LM tests also are asymptotically efficient under normal errors and strong IV asymptotics. When $k > m$, the rank AR statistic has a different asymptotic distribution from that of the rank LR and LM statistics (see Comment 2 to Theorem 2) and, hence, it is not asymptotically efficient.

6 Finite Sample Results

In this section, we report simulation results concerning the finite sample size of some of the rank and non-rank tests discussed above. We also provide power comparisons of size-corrected versions of these tests.

We consider the Wilcoxon scores rank CLR test, denoted $RCLR_n^{WS}$, and the normal scores CLR rank test, denoted $RCLR_n^{NS}$. For comparative purposes, we also consider the CLR, LM, and AR tests. We do not report results for the rank LM and rank AR tests both for brevity and for the following reasons. First, when the model is over-identified, the AR test has distinctly lower power than the CLR test, see AMS1 and AMS2, and the same is true for the rank versions of these tests. Second, the LM test has quirky power properties in parts of the parameter space, e.g., see AMS1 and AMS2, and the rank LM test inherits these properties.

6.1 Experimental Design

We take the model to be as in (2.1) with y_{2i} and β being scalars ($m = 1$) and v_{2i} defined as in (4.18), where $\rho_{uv_2} \in [-1, 1]$. Let $\widetilde{Z}_i = (\widetilde{Z}_{i1}, ..., \widetilde{Z}_{ik})'$ and $X_i = (1, X_{i2}, ..., X_{ip})'$. We take $\widetilde{Z}_{ij}, X_{is}, u_i, \varepsilon_i$ to be iid with distribution F for all $j = 1, ..., k$, $s = 2, ..., p$, and $i = 1, ..., n$.[13]

The test statistics considered are invariant with respect to γ_1, ξ_1, and the location and scale of F. Hence, without loss of generality we take γ_1 and ξ_1 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).

[13] Thus, we consider a model with random exogenous variables and IVs. The tests considered have the correct size asymptotically both conditionally and unconditionally on the exogenous variables and IVs.

The parameter vector $\pi \in R^k$ determines the strength of the IVs. It is taken to be proportional to a k-vector of ones:

$$\pi = \frac{\rho_{IV}}{k^{1/2}(1-\rho_{IV}^2)^{1/2}}(1, ..., 1)' \text{ for some } \rho_{IV} \in [-1, 1], \tag{6.1}$$

where ρ_{IV} is the correlation between the reduced-form regression function, $Z_i'\pi$, and the endogenous variable y_{2i} (when 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):

$$\lambda = \frac{n\rho_{IV}^2}{1-\rho_{IV}^2} = n\pi' E\widetilde{Z}_i\widetilde{Z}_i'\pi \approx \pi'\widetilde{Z}'\widetilde{Z}\pi, \tag{6.2}$$

where the first equality defines λ, the second equality holds provided $\widetilde{Z}_i$ has a finite variance, and $a_n \approx b_n$ means $a_n/b_n \to_p 1$ as $n \to \infty$.

The hypotheses of interest are $H_0 : \beta = \beta_0$ and $H_1 : \beta \neq \beta_0$. Without loss of generality, we take $\beta_0 = 0$.[14]

For both the size and power results, we first consider a Base Case with moderately weak IVs $\lambda = 10$ (equivalently, $\rho_{IV} = .302$ when $n = 100$), moderately strong endogeneity $\rho_{uv_2} = .75$, sample size $n = 100$, number of IVs $k = 5$, no exogenous variables beyond a constant $p = 1$, and distribution F equal to the normal, uniform, t_1, t_2, t_3, or difference of independent log-normals (DLN). The uniform distribution exhibits thin tails, and the t distributions exhibit heavy tails (e.g., t_1 is the Cauchy distribution) as does the DLN distribution. For the power results, both positive and negative true β values are considered. The β values are selected so that the level .05 CLR test has power around .4 for the given choice of λ, ρ_{uv_2}, n, k, and p when F is normal.

We also consider a number of variations of the Base Case to illustrate the effect of changes in the level of endogeneity: $\rho_{uv_2} = 0, .95$; strength of IVs: $\lambda = 4, 20$; number of IVs: $k = 1, 10$, and sample size: $n = 50, 200$. In each variation of the Base Case, only one of these parameters is different from the Base Case. In the Base Case, we find that when F is normal the power of the normal scores rank CLR test is slightly higher than that of the non-rank CLR test, but the opposite is true for negative β. (These differences disappear asymptotically under weak and strong IV asymptotics.) In consequence, to maintain fair comparisons and for brevity, in each variation of the Base Case, we report average power for two β values—one positive and one negative—each of which is chosen so that the CLR test has power approximately equal to .4 when F is normal.[15]

For the power results, the tests are all size-corrected. The size-correcting critical values are obtained via simulation with $100,000$ simulation repetitions. The number of simulation repetitions is $20,000$ for the size results and $5,000$ for the power results.

[14] There is no loss of generality in taking $\beta_0 = 0$ because the structural equation $y_{1i} = y_{2i}\beta + \gamma_1' X_i + u_i$ and hypothesis $H_0 : \beta = \beta_0$ can be transformed into $\widetilde{y}_{1i} = y_{2i}\widetilde{\beta} + \gamma_1' X_i + u_i$ and $H_0 : \widetilde{\beta} = 0$, where $\widetilde{y}_{1i} = y_{1i} - y_{2i}\beta_0$ and $\widetilde{\beta} = \beta - \beta_0$.

[15] The reported power of the CLR test for the case where λ or n is small is less than .4 because it is not possible to choose β so that the CLR test has power as high as .4.

Note that the size results for the AR test are invariant to ρ_{uv_2} and λ.

6.2 Size Results

Table II presents the size results. The two rank CLR tests perform noticeably better in terms of size than the non-rank CLR, LM, and AR tests. Nine different cases are considered with six different distributions for each case. Over the 54 trials, the range of null rejection rates for each test is WS-RCLR: $[.027, .052]$; NS-CLR: $[.033, .051]$; CLR: $[.047, .091$; LM: $[.042, .070]$; and AR: $[.049, .127]$. For the two rank tests, the majority of rejection rates are in the desired $[.040, .050]$ range, which corresponds to no over-rejection and sufficiently small under-rejection as to minimize the power loss. (In particular, 42/54 for WS-CLR and 38/54 for NS-CLR are in this range.) In contrast, for the non-rank tests a small number of rejection rates are in this desired range: 1/54 for CLR, 3/54 for LM, and 11/54 for AR. Not surprisingly, the largest over-rejections for the non-rank tests occur for the thickest-tailed distributions.

6.3 Power Comparisons

Table III presents the power results. The general pattern of finite sample power in Table III reflects that of asymptotic power given in Table I. In particular, the NS-RCLR and CLR tests have comparable power for the normal distribution, the NS-RCLR test has noticeably higher power than the CLR test for the uniform distribution and much higher power for the thick-tailed distributions. This occurs in the Base Case and in the variations of the Base Case. For example, in the Base Case with two β values the (average) power of the NS-RCLR test for t_2 distribution is 0.67 compared to 0.46 for the CLR test. The WS-RCLR and NS-RCLR tests have similar power with the NS-RCLR test having slightly higher power for the normal distribution, noticeably higher power for the uniform distribution, and slightly worse power for the thick-tailed distributions. The LM test has similar power to the CLR test, but with lower power in the weaker IVs case with normal distribution and slightly higher power for the heavy-tailed distributions. The AR test has significantly lower power than the other tests except in the case with $k = 1$.

In sum, the NS-RCLR test has power that essentially dominates that of the (non-rank) CLR, LM, and AR tests. Its power is comparable to that of the CLR and LM tests for the normal distribution and higher for the other distributions, especially the thick-tailed ones. The power of the WS-RCLR test is similar to that of the NS-RCLR test.

7 Appendix of Proofs

The proofs of Lemmas 1-5 and Corollary 1 are given at the end of the Appendix, as is the description of the numerical calculation of asymptotic power under weak IVs.

The proofs of Theorem 1(a) and 2(a) rely on the following Lemmas. The first Lemma follows from results of Koul (1970) and Hájek and Sidák (1967).

Lemma 6 *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) $\{c_i : i \leq n, n \geq 1\}$ *and* $\{d_i : i \leq n, n \geq 1\}$ *are triangular arrays of non-random* δ_c*-vectors and* δ_d*-vectors, respectively* (*with dependence of* c_i *and* d_i *on* n *suppressed for brevity*) *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* $b < \infty$,

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

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

(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) $n^{1/2}\Psi_n(0) = n^{-1/2}\sum_{i=1}^n (c_i - \overline{c}_n)\varphi(H(Q_i)) + o_p(1)$.

Comments. 1. Lemma 6(a) 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. Lemma 6(b) follows from part (a). Lemma 6(c) follows from the proofs of Hájek and Sidák's (1967) Thm. V.1.5a, p. 160, Thm. VI.1.6a, p. 163, and Lem. VI.1.6a, p. 164, which show that in the scalar c_i case $E(n^{-1/2}\sum_{i=1}^n (c_i - \overline{c}_n)\varphi(H(Q_i)) - n^{-1/2}\sum_{i=1}^n (c_i - \overline{c}_n)a_n^{\varphi}(i))^2 = o(1)$ and $E(n^{-1/2}\sum_{i=1}^n (c_i - \overline{c}_n)a_n^{\varphi}(i) - n^{1/2}\Psi_n(0))^2 = o(1)$, respectively, where $a_n^{\varphi}(i) = E(\varphi(H(Q_1))|r_1(0) = i)$.

2. 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.

3. 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 next lemma is used to establish the probability limit of $\widehat{\nu}_{\varphi n}$.

Lemma 7 *Suppose* (i) $\{(Q_{1i}, Q_{2i}) : i \geq 1\}$ *is an iid sequence of random* $(m+1)$-*vectors with* $Q_{1i} \in R$, (ii) (Q_{1i}, Q_{2i}) *has an absolutely-continuous and bounded joint df* H_{Q_1,Q_2} *that satisfies* $\sup_{(q_1,q_2)} |\partial H_{Q_1,Q_2}(q_1, q_2)/\partial q_1| < \infty$, (iii) $E||Q_{2i}|| < \infty$, (iv) $r_i(t)$ *is the rank of* $Q_{1i} - d_i't$ *among* $\{Q_{1j} - d_j't : j \leq n\}$, *where* $t \in R^{\delta_d}$, (v) $\{d_i : i \leq n, n \geq 1\}$ *is a triangular array of non-random* δ_d-*vectors that satisfies* $\overline{\lim}_{n\to\infty} n^{-1}\sum_{i=1}^n ||d_i|| < \infty$, *and* (vi) *the score function* φ *satisfies Assumption 3. Then,*
(a) *for all* $b < \infty$,

$$\sup_{t:||t||\leq b} \left| n^{-1}\sum_{i=1}^n \varphi\left(\frac{r_i(tn^{-1/2})}{n+1}\right) Q_{2i} - n^{-1}\sum_{i=1}^n \varphi\left(\frac{r_i(0)}{n+1}\right) Q_{2i} \right| = o_p(1),$$

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

$$n^{-1}\sum_{i=1}^n \varphi\left(\frac{r_i(\widehat{\tau}_n)}{n+1}\right) Q_{2i} = n^{-1}\sum_{i=1}^n \varphi\left(\frac{r_i(0)}{n+1}\right) Q_{2i} + o_p(1),$$

(c) $n^{-1}\sum_{i=1}^n \varphi\left(\frac{r_i(0)}{n+1}\right) Q_{2i} = E\varphi(H_{Q_1}(Q_{1i}))Q_{2i} + o_p(1)$, *where* H_{Q_1} *is the df of* Q_{1i}.

Comment. Lemma 7(a) follows from arguments similar to the ones used to prove Lemma 2.2 in Koul (1970), which was originally proved, under different assumptions, as Theorem 3.1 in Koul (1969). The result established in Lemma 7(a) is different from the results established in Koul (1969, 1970), but the idea of the argument is essentially the same. The results in Koul (1969, 1970) are for a linear regression model with deterministic regressors. Hence, using our notation, the results in Koul (1969, 1970) are restricted to the case where $\{Q_{2i} : i \leq n\}$ are nonrandom real numbers, and $\{(Q_{1i}, Q_{2i}) : i \leq n\}$ and $\{d_i : i \leq n\}$ satisfy the relation imposed by a linear regression equation. Hence, the conditions in Lemma 7(a) generalize those in Lemma 2.2 of Koul (1970). On the other hand, Lemma 2.2 of Koul (1970) establishes that the lhs in Lemma 7(a) is $o_p(n^{-1/2})$, which is a stronger result than that given in Lemma 7(a).

Let Φ be the n-vector with i-th element given by $\varphi(U_{gi}) = \varphi(G(u_i + (\beta - \beta_0)'v_{2i}))$.

Lemma 8 *Under Assumptions* 1-3 *and* 4W,
(a) $n^{-1/2}Z'R_\varphi = n^{-1/2}Z'(\Phi + ZC\ell^\varphi_{g,\beta-\beta_0}c_\varphi^{1/2}n^{-1/2}) + o_p(1)$,
(b) $S_n^\varphi = (Z'Z)^{-1/2}Z'(\Phi c_\varphi^{-1/2} + ZC\ell^\varphi_{g,\beta-\beta_0}n^{-1/2}) + o_p(1)$,
(c) $n^{-1}Z'Z \to D_Z > 0$, *and*
(d) $n^{-1/2}Z'[(\Phi c_\varphi^{-1/2} + ZC\ell^\varphi_{g,\beta-\beta_0}n^{-1/2}) : Y_2] \to_d [N_\varphi : N_2]$.

Lemma 9 *Under Assumptions* 1-3 *and* 4W, (a) $\widehat{\nu}_{\varphi n} \to_p \nu_{\varphi g}$ *and* (b) $\widehat{\Omega}_{22n} \to_p \Omega_{22}$.

Lemma 10 *Under Assumptions* 1-3 *and* 4S,
(a) $n^{-1/2}Z'R_\varphi = n^{-1/2}Z'(\Phi + Z\Pi\ell^\varphi_{f,B}c_\varphi^{1/2}n^{-1/2}) + o_p(1)$,
(b) $S_n^\varphi = (Z'Z)^{-1/2}Z'(\Phi c_\varphi^{-1/2} + Z\Pi\ell^\varphi_{f,B}n^{-1/2}) + o_p(1)$, *and*
(c) $n^{-1}Z'[\Phi : Y_2] \to_p D_Z[0_k : \Pi]$.

Lemma 11 *Under Assumptions* 1-3 *and* 4S, (a) $\widehat{\nu}_{\varphi n} \to_p \nu_{\varphi f}$ *and* (b) $\widehat{\Omega}_{22n} \to_p \Omega_{22}$.

The following Lemma gives sufficient conditions for an iid sequence to satisfy Assumption 2(d) and 4S(h) a.s.

Lemma 12 *Suppose* $\{\psi_i : i \geq 1\}$ *is an iid sequence of non-negative random variables with* $E\psi_i^{1+\delta} < \infty$ *for some* $\delta > 0$. *Then,* (a) $\sum_{i=1}^\infty \psi_i^{1+\delta}/i^{1+\delta} < \infty$ *a.s. and* (b) $\max_{i \leq n} \psi_i/n \to 0$ *a.s.*

The last Lemma is a Glivenko-Cantelli Theorem for triangular arrays of random variables, which is used in the proof of Lemma 7. It is proved by verifying the conditions in Pollard (1990, Thm. 8.3).

Lemma 13 *Suppose* (i) $\{(Q_{1i}, Q_{2i}) : i \geq 1\}$ *is an iid sequence of random* $(m+1)$-*vectors with* $Q_{1i} \in R$, *and* (ii) $\{d_i : i \geq 1\}$ *is any sequence of non-random* δ_d-*vectors. Then, for any* $b < \infty$,

$$\sup_{(q_1,q_2)\in R^{m+1}} \sup_{t \in R^{\delta_d}: ||t|| \leq b} \left| n^{-1} \sum_{i=1}^n [h_{ni}(q_1, q_2, t) - Eh_{ni}(q_1, q_2, t)] \right| \to 0 \text{ a.s., where}$$
$$h_{ni}(q_1, q_2, t) = 1(Q_{1i} \leq q_1 + d_i' t n^{-1/2}, Q_{2i} \leq q_2).$$

The proofs of Lemmas 7-13 are given after the proofs of Theorems 1-2.

Proof of Theorem 1. Lemma 9 and Assumption 4W(e) imply that

$$\widehat{\Omega}_{\varphi n} \to_p \Omega_{\varphi g} \text{ and } \widehat{\Omega}_{\varphi n}^{-1}H(H'\widehat{\Omega}_{\varphi n}^{-1}H)^{-1/2} \to_p \Omega_{\varphi g}^{-1}H(H'\Omega_{\varphi g}^{-1}H)^{-1/2}. \tag{7.1}$$

This, Lemma 8, the continuous mapping theorem, and the definitions of $(S_n^\varphi, T_n^\varphi)$ and $(S_\infty^\varphi, T_\infty^\varphi)$ combine to establish part (a).

Independence of S_∞^φ and T_∞^φ is implied by zero covariance between the normal variates N_φ and $[N_\varphi : N_2]\Omega_{\varphi g}^{-1}H$. The latter holds by the following argument. Let

$N_{\varphi,j}$, $N_{2,\ell}$, and $D_{Z,j\ell}$ denote the j-th element of N_φ, the ℓ-th row of N_2, and the (j,ℓ) element of D_Z, respectively. Let e_1 denote an $m+1$ vector of ones. The covariance between $N_{\varphi,j}$ and the ℓ-th row of $[N_\varphi : N_2]\Omega_{\varphi g}^{-1}H$ for $j,\ell = 1,...,k$ is

$$\begin{aligned} & Cov(N_{\varphi,j}, [N_{\varphi,\ell} : N_{2,\ell}]\Omega_{\varphi g}^{-1}H) \\ = \ & Ee_1' \begin{pmatrix} N_{\varphi,j} - EN_{\varphi,j} \\ N_{2,j}' - EN_{2,j}' \end{pmatrix} [N_{\varphi,\ell} : N_{2,\ell}]\Omega_{\varphi g}^{-1}H = D_{Z,j\ell} \cdot e_1'\Omega_{\varphi g}\Omega_{\varphi g}^{-1}H = 0. \quad (7.2) \end{aligned}$$

Parts (b)-(d) of the Theorem follow immediately from part (a) and the continuous mapping theorem. □

Proof of Theorem 2. The result $S_n^\varphi \to_d S_{f\infty}^\varphi$ of part (a) follows from Lemma 10(b), Lemma 8(c) (which does not rely on Assumption 4W), and the Lindeberg CLT applied to $n^{-1/2}Z'\Phi c_\varphi^{-1/2}$. The CLT applies by the same argument as given in the proof of Lemma 8(d) below.

The result $n^{-1/2}T_n^\varphi \to_d \alpha_T^\varphi$ (or $n^{-1/2}T_n^\varphi \to_p \alpha_T^\varphi$) is established as follows:

$$\begin{aligned} n^{-1/2}T_n^\varphi &= n^{-1/2}(Z'Z)^{-1/2}Z'[R_\varphi c_\varphi^{-1/2} : Y_2]\widehat{\Omega}_{\varphi n}^{-1}H(H'\widehat{\Omega}_{\varphi n}^{-1}H)^{-1/2} \\ &= (n^{-1}Z'Z)^{-1/2}[n^{-1}Z'R_\varphi c_\varphi^{-1/2} : n^{-1}Z'Y_2]\Omega_{\varphi f}^{-1}H(H'\Omega_{\varphi f}^{-1}H)^{-1/2} + o_p(1) \\ &= D_Z^{-1/2}[n^{-1}Z'(\Phi c_\varphi^{-1/2} + Z\Pi\ell_{f,B}^\varphi n^{-1/2}) : n^{-1}Z'Y_2]\Omega_{\varphi f}^{-1}H(H'\Omega_{\varphi f}^{-1}H)^{-1/2} + o_p(1) \\ &= D_Z^{1/2}[0_k : \Pi]\Omega_{\varphi f}^{-1}H(H'\Omega_{\varphi f}^{-1}H)^{-1/2} + o_p(1) \\ &= D_Z^{1/2}\Pi(H'\Omega_{\varphi f}^{-1}H)^{1/2} + o_p(1), \end{aligned} \quad (7.3)$$

where the second equality holds because Lemma 11 and Assumption 4S(g) imply that $\widehat{\Omega}_{\varphi n}^{-1} \to_p \Omega_{\varphi f}^{-1}$, the third equality holds by Lemma 8(c) and Lemma 10(b), the fourth equality holds by Lemma 10(c), and the fifth equality holds because $[0_k : \Pi] = \Pi[0_m : I_m] = \Pi H'$. The convergence of $(S_n^\varphi, n^{-1/2}T_n^\varphi)$ holds jointly because α_T^φ is a constant.

Parts (c) and (d) follow immediately from part (a) using the continuous mapping theorem noting that $\alpha_T^{\varphi\prime}\alpha_T^\varphi$ is pd by Assumptions 2(c), 4S(b), and 4S(g).

We now prove part (b). Given the definition of RLR_n^φ in (3.8) and the result of Theorem 2(c), it suffices to show that

$$\begin{aligned} \lambda_{\min}([S_n^\varphi : T_n^\varphi]'\,[S_n^\varphi : T_n^\varphi]) \ &= \ S^{\perp\prime}S^\perp + o_p(1), \text{ where} \\ S^\perp \ &= \ S_n^\varphi - T_n^\varphi(T_n^{\varphi\prime}T_n^\varphi)^{-1}T_n^{\varphi\prime}S_n^\varphi. \end{aligned} \quad (7.4)$$

For notational simplicity, let $[S:T]$ denote $[S_n^\varphi : T_n^\varphi]$ and let $T_j \in R^{m+1}$ denote the jth column of T for $j = 1,...,m$. We rotate $[S:T]$ by an orthogonal matrix $B \in R^{(m+1)\times(m+1)}$ whose first column, b_1, is designed to be such that $[S:T]\,b_1 = d_1 S^\perp$, where d_1 is a positive scalar that equals $1 + o_p(1)$. Then, we have

$$\lambda_{\min}([S:T]'\,[S:T]) = \lambda_{\min}(B'\,[S:T]'\,[S:T]\,B) \quad (7.5)$$

and the $(1,1)$ element of the matrix on the right-hand side equals $\lambda_1^2 d_1^2 S^{\perp\prime}S^\perp$.

Let b_j denote the jth column of B and b_{ij} denote the (i,j)th element of B. Define

$$b_1 = d_1 \left(\begin{array}{c} 1 \\ -(T'T)^{-1}T'S \end{array} \right) \in R^{m+1}, \tag{7.6}$$

where d_1 is a constant such that $b_1'b_1 = 1$. Next, we define the orthogonal vectors $\{b_j : j = 2, ..., m+1\}$ via the Gramm-Schmidt procedure applied to the vectors $b_1, e_2, ..., e_{m+1}$, where e_j is the jth elementary vector (whose jth element is one and whose other elements are zero). We have

$$\begin{aligned} b_2 &= d_2(e_2 - (e_2'b_1)b_1) = d_2(e_2 - b_{12}b_1), \\ b_3 &= d_3(e_3 - (e_3'b_2)b_2 - (e_3'b_1)b_1) = d_3(e_3 - b_{23}b_2 - b_{13}b_1), \end{aligned} \tag{7.7}$$

and so on, where d_j is the constant that yields $||b_j|| = 1$ for $j = 1, ..., m$.

The constants $\{d_j : j = 1, ..., m+1\}$ satisfy

$$\begin{aligned} d_1 &= (1 + n^{-1}(n^{-1/2}S'T)(n^{-1}T'T)^{-2}(n^{-1/2}T'S))^{-1/2} = 1 + o_p(1), \\ d_2 &= (1 - b_{12}^2)^{-1/2} = 1 + o_p(1), \\ d_3 &= (1 - b_{23}^2 - b_{13}^2)^{-1/2} = 1 + o_p(1), \end{aligned} \tag{7.8}$$

and so on, using Theorem 2(a) and the fact that

$$\begin{aligned} b_{1j} &= n^{-1/2}[-d_1(n^{-1}T'T)^{-1}n^{-1/2}T'S]_j = O_p(n^{-1/2}) \text{ for } j = 2, ..., m, \\ b_{2j} &= d_2(-b_{12}b_{1j}) = O_p(n^{-1}) \text{ for } j = 3, ..., m, \\ b_{3j} &= d_3(-b_{23}b_{2j} - b_{13}b_{1j}) = O_p(n^{-1}) \text{ for } j = 4, ..., m, \end{aligned} \tag{7.9}$$

and so on.

Let $\lambda = (\lambda_1, ..., \lambda_{m+1})' = (\lambda_1, \widetilde{\lambda}_2')' \in R^{m+1}$ be such that $||\lambda|| = 1$. Then, we have

$$\begin{aligned} &\lambda_{\min}(B' [S{:}T]' [S{:}T] B) = \inf_{\lambda \in R^{m+1}: ||\lambda|| = 1} J(\lambda), \text{ where} \\ &J(\lambda) := || [S{:}T] B\lambda||^2 = \lambda_1^2 d_1^2 S^{\perp\prime} S^{\perp} + 2\lambda_1 d_1 S^{\perp\prime} [S{:}T] [b_2 \cdots b_{m+1}]\widetilde{\lambda}_2 + J_3(\lambda), \\ &J_3(\lambda) := \widetilde{\lambda}_2'[b_2 \cdots b_{m+1}]' [S{:}T]' [S{:}T] [b_2 \cdots b_{m+1}]\widetilde{\lambda}_2. \end{aligned} \tag{7.10}$$

The cross-product summand of $J(\lambda)$ in (7.10) equals

$$2\lambda_1 d_1 \left[S^{\perp\prime} S{:}0_{1\times m} \right] [b_2 \cdots b_{m+1}]\widetilde{\lambda}_2 = O_p(||\widetilde{\lambda}_2||), \tag{7.11}$$

using $S^{\perp\prime}T = 0$, $(S^{\perp\prime}S)^2 \leq (S^{\perp\prime}S^{\perp})S'S \leq (S'S)^2 = O_p(1)$, $|b_{ij}| \leq 1$, and $d_1 = 1 + o_p(1)$. For the third summand $J_3(\lambda)$ of $J(\lambda)$, we have

$$\begin{aligned} &[S{:}T] [b_2 \cdots b_{m+1}] \\ &= \left[d_2(T_1 - b_{12}S^{\perp}){:}d_3(T_2 - b_{23}d_2(T_1 - b_{12}S^{\perp}) - b_{13}S^{\perp}){:}\cdots \right]. \end{aligned} \tag{7.12}$$

Combining this with (7.8), (7.9), $S^{\perp\prime}T = 0$, $S^{\perp} = O_p(1)$, and $n^{-1/2}T \to_p \alpha_T^{\varphi}$ (by part (a) of the Theorem), we obtain

$$0 \leq J_3(\lambda) = n\widetilde{\lambda}_2'(\alpha_T^{\varphi\prime}\alpha_T^{\varphi} + o_p(1))\widetilde{\lambda}_2, \tag{7.13}$$

where $\alpha_T^{\varphi\prime}\alpha_T^{\varphi}$ is pd by Assumptions 2(c), 4S(b), and 4S(g).

Let $\lambda^* = (\lambda_1^*, ..., \lambda_{m+1}^*)' = (\lambda_1^*, \widetilde{\lambda}_2^{*\prime})' \in R^{m+1}$ be an $m+1$ vector that minimizes $J(\lambda)$ over $\lambda \in R^{m+1}$ such that $||\lambda|| = 1$. If $||\widetilde{\lambda}_2^*|| = o_p(n^{-1})$, then

$$J(\lambda^*) = S^{\perp\prime}S^{\perp} + o_p(1) \tag{7.14}$$

by (7.10)-(7.13) and $S^{\perp\prime}S^{\perp} = O_p(1)$ by part (a) of the Theorem.

On the other hand, suppose $||\widetilde{\lambda}_2^*|| = o_p(1)$ and $||\widetilde{\lambda}_2^*|| \neq o_p(n^{-1})$, then $|\lambda_1^*| = 1 + o_p(1)$,

$$\begin{aligned} J(\lambda^*) &= S^{\perp\prime}S^{\perp} + o_p(1) + J_3(\lambda^*) \text{ and} \\ 0 &\leq J_3(\lambda^*) = n\widetilde{\lambda}_2^{*\prime}(\alpha_T^{\varphi\prime}\alpha_T^{\varphi} + o_p(1))\widetilde{\lambda}_2^* \neq o_p(1). \end{aligned} \tag{7.15}$$

This contradicts the assumption that λ^* minimizes $J(\lambda)$ over λ such that $||\lambda|| = 1$ because a different choice of λ, viz. λ such that $||\widetilde{\lambda}_2|| = o_p(n^{-1})$ yields a smaller value $J(\lambda)$ as indicated in (7.14).

Next, suppose $||\widetilde{\lambda}_2^*|| \neq o_p(1)$. Then,

$$\begin{aligned} J(\lambda^*) &= O_p(1) + J_3(\lambda^*), \\ 0 &\leq J_3(\lambda^*) \neq o_p(n), \text{ and } J(\lambda^*) \neq O_p(1) \end{aligned} \tag{7.16}$$

by (7.10)-(7.13). In particular, for some $\varepsilon > 0$ and some (infinite) subsequence $\{\ell_n\}$ of $\{n\}$, $P(J_3(\lambda^*) > \ell_n\varepsilon) > \varepsilon$ when the sample size is ℓ_n for all $n \geq 1$. Again this is a contradiction, because a different choice of λ, viz., λ such that $||\widetilde{\lambda}_2|| = o_p(n^{-1})$ yields a smaller value $J(\lambda)$, viz. one that is $O_p(1)$ as indicated in (7.14). We conclude that $||\widetilde{\lambda}_2^*||$ must satisfy $||\widetilde{\lambda}_2^*|| = o_p(n^{-1})$ and, hence, (7.14) in conjunction with (7.4), (7.5), and (7.10) combine to establish the result of part (b). □

Proof of Lemma 7. Because $E||Q_{2i}|| < \infty$, given any $\varepsilon > 0$, there exists a constant $c_\varepsilon < \infty$ such that

$$E||Q_{2i}||1(||Q_{2i}|| > c_\varepsilon) < \varepsilon. \tag{7.17}$$

Hence, using the boundedness of φ, say by C, and Markov's inequality, we have: for any $\eta > 0$ and $\varepsilon > 0$,

$$\begin{aligned} & P\left(\sup_{t:||t||\leq b}\left|n^{-1}\sum_{i=1}^{n}\varphi\left(\frac{r_i(tn^{-1/2})}{n+1}\right)Q_{2i}1(||Q_{2i}|| > c_\varepsilon)\right| > \eta\right) \\ \leq & \frac{C}{\eta}E||Q_{2i}||1(||Q_{2i}|| > c_\varepsilon) < \frac{C\varepsilon}{\eta}. \end{aligned} \tag{7.18}$$

Therefore, without loss of generality, we can assume that Q_{2i} is bounded.

Define

$$\begin{aligned} L_{1n}(q_1, t) &= n^{-1}\sum_{i=1}^{n}1(Q_{1i} - d_i't \leq q_1) \text{ and} \\ L_{12n}(q_1, q_2, t) &= n^{-1}\sum_{i=1}^{n}1(Q_{1t} - d_i't \leq q_1, Q_{2i} \leq q_2). \end{aligned} \tag{7.19}$$

Note that

$$
\begin{aligned}
EL_{1n}(q_1,t) &= n^{-1}\sum_{i=1}^{n} H_{Q_1}(q_1+d_i't) \text{ and} \\
EL_{12n}(q_1,q_2,t) &= n^{-1}\sum_{i=1}^{n} H_{Q_1,Q_2}(q_1+d_i't,q_2). \qquad (7.20)
\end{aligned}
$$

Now, we have

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\varphi\left(\frac{r_i(t)}{n+1}\right)Q_{2i} \\
&= n^{-1}\sum_{i=1}^{n}\varphi\left(\frac{1}{n+1}\sum_{j=1}^{n}1(Q_{1j}-d_j't\le Q_{1i}-d_i't)\right)Q_{2i} \\
&= n^{-1}\sum_{i=1}^{n}\varphi\left(\frac{nL_{1n}(Q_{1i}-d_i't,t)}{n+1}\right)Q_{2i} \\
&= \int\int\varphi\left(\frac{nL_{1n}(q_1,t)}{n+1}\right)q_2dL_{12n}(q_1,q_2,t) \\
&= \int\int\left[\varphi\left(\frac{nL_{1n}(q_1,t)}{n+1}\right)-\varphi\left(\frac{nEL_{1n}(q_1,t)}{n+1}\right)\right]q_2dL_{12n}(q_1,q_2,t) \\
&\quad +\int\int\varphi\left(\frac{nEL_{1n}(q_1,t)}{n+1}\right)q_2dL_{12n}(q_1,q_2,t). \qquad (7.21)
\end{aligned}
$$

Therefore, using the triangle inequality,

$$
\begin{aligned}
& \sup_{t:||t||\le b}\left|n^{-1}\sum_{i=1}^{n}\varphi\left(\frac{r_i(tn^{-1/2})}{n+1}\right)Q_{2i}-n^{-1}\sum_{i=1}^{n}\varphi\left(\frac{r_i(0)}{n+1}\right)Q_{2i}\right| \\
& \le A_{1n}(b)+A_{1n}(0)+A_{2n}, \qquad (7.22)
\end{aligned}
$$

where, for $b\ge 0$,

$$
\begin{aligned}
A_{1n}(b) &= \sup_{t:||t||\le b}\left|\int\int\left[\varphi\left(\frac{nL_{1n}(q_1,tn^{-1/2})}{n+1}\right)-\varphi\left(\frac{nEL_{1n}(q_1,tn^{-1/2})}{n+1}\right)\right]\right. \\
&\qquad \left.\times q_2dL_{12n}(q_1,q_2,tn^{-1/2})\right| \text{ and} \\
A_{2n} &= \sup_{t:||t||\le b}\left|\int\int\varphi\left[\frac{nEL_{1n}(q_1,tn^{-1/2})}{n+1}\right]q_2dL_{12n}(q_1,q_2,tn^{-1/2})\right. \\
&\qquad \left.-\int\int\varphi\left[\frac{nEL_{1n}(q_1,0)}{n+1}\right]q_2dL_{12n}(q_1,q_2,0)\right|. \qquad (7.23)
\end{aligned}
$$

Now, by Lemma 13,

$$\sup_{q_1 \in R} \sup_{t:||t|| \leq b} \left| L_{1n}(q_1, tn^{-1/2}) - EL_{1n}(q_1, tn^{-1/2}) \right| \tag{7.24}$$
$$= \sup_{q_1 \in R} \sup_{t:||t|| \leq b} \left| n^{-1} \sum_{i=1}^{n} \left[1(Q_{1i} \leq q_1 + d_i' tn^{-1/2}) - H_{Q_1}(q_1 + d_i' tn^{-1/2}) \right] \right| \to_p 0.$$

This implies that $A_{1n}(b) \to_p 0$ and $A_{1n}(0) \to_p 0$, because φ is absolutely continuous, Q_{2i} is bounded, and $0 \leq A_{1n}(0) \leq A_{1n}(b)$.

Using the triangle inequality again, we have $A_{2n} \leq B_{1n} + B_{2n}$, where

$$B_{1n} = \sup_{t:||t|| \leq b} \left| \int \int \varphi \left[\frac{nEL_{1n}(q_1, tn^{-1/2})}{n+1} \right] q_2 dL_{12n}(q_1, q_2, tn^{-1/2}) \right.$$
$$\left. - \int \int \varphi \left[\frac{nEL_{1n}(q_1, 0)}{n+1} \right] q_2 dL_{12n}(q_1, q_2, tn^{-1/2}) \right| \text{ and } \tag{7.25}$$
$$B_{2n} = \sup_{t:||t|| \leq b} \left| \int \int \varphi \left[\frac{nEL_{1n}(q_1, 0)}{n+1} \right] q_2 d\{L_{12n}(q_1, q_2, tn^{-1/2}) - L_{12n}(q_1, q_2, 0)\} \right|.$$

To bound B_{1n} and B_{2n}, we write

$$\sup_{(q_1, q_2) \in R^{m+1}} \sup_{t:||t|| \leq b} \left| L_{12n}(q_1, q_2, tn^{-1/2}) - L_{12n}(q_1, q_2, 0) \right|$$
$$\leq \sup_{(q_1, q_2) \in R^{m+1}} \sup_{t:||t|| \leq b} \left| L_{12n}(q_1, q_2, tn^{-1/2}) - EL_{12n}(q_1, q_2, tn^{-1/2}) \right|$$
$$+ \sup_{(q_1, q_2) \in R^{m+1}} \sup_{t:||t|| \leq b} \left| EL_{12n}(q_1, q_2, tn^{-1/2}) - EL_{12n}(q_1, q_2, 0) \right|$$
$$+ \sup_{(q_1, q_2) \in R^{m+1}} \sup_{t:||t|| \leq b} \left| EL_{12n}(q_1, q_2, 0) - L_{12n}(q_1, q_2, 0) \right|. \tag{7.26}$$

The first and last terms on the rhs converge to zero a.s. by Lemma 13. The second term on the rhs converges to zero because it equals

$$\sup_{(q_1, q_2)} \sup_{t:||t|| \leq b} \left| n^{-1} \sum_{i=1}^{n} H_{Q_1, Q_2}(q_1 + d_i' tn^{-1/2}, q_2) - H_{Q_1, Q_2}(q_1, q_2) \right|$$
$$= \sup_{(q_1, q_2)} \sup_{t:||t|| \leq b} \left| n^{-1} \sum_{i=1}^{n} \frac{\partial H_{Q_1, Q_2}(q_1 + d_i' t^* n^{-1/2}, q_2)}{\partial q_1} d_i' tn^{-1/2} \right| = o(1), \tag{7.27}$$

where t^* lies between 0 and t, the first equality holds by a mean-value expansion around $t = 0$, and the second equality holds because $\partial H_{Q_1, Q_2} / \partial q_1$ is bounded (Assumption 4W(d)) and $\overline{\lim}_{n \to \infty} \sum_{i=1}^{n} ||d_i|| < \infty$. Therefore, using the boundedness of φ and Q_{2i}, we have $B_{2n} \to_p 0$.

Equation (7.27) and a mean-value expansion yield $B_{1n} \to_p 0$ because φ has a bounded first derivative by Assumption 3(a). In consequence, $A_{2n} \to_p 0$, which completes the proof of part (a).

Part (b) of the Lemma follows from part (a) using a standard argument.

To prove part (c), as in part (a), we can assume that Q_{2i} is bounded without loss of generality. We have

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\varphi\left(\frac{r_i(0)}{n+1}\right)Q_{2i} &= \int\int\varphi\left(\frac{nL_{1n}(q_1,0)}{n+1}\right)q_2 dL_{12n}(q_1,q_2,0) \\ &= \int\int\varphi\left(\frac{nEL_{1n}(q_1,0)}{n+1}\right)q_2 dL_{12n}(q_1,q_2,0)+o_p(1) \\ &= \int\int\varphi\left(\frac{nH_{Q_1}(q_1)}{n+1}\right)q_2 dH_{Q_1,Q_2}(q_1,q_2)+o_p(1) \\ &= E\varphi(H_{Q_1}(Q_{1i}))Q_{2i}+o_p(1), \end{aligned} \tag{7.28}$$

where the first equality holds by (7.21) with $t=0$, the second equality holds because $A_{1n}(0)\rightarrow_p 0$, the third equality holds by (7.20), and the fourth equality holds because $n/(n+1)\rightarrow 1$, Q_{2i} is bounded, and φ has a bounded derivative. □

Proof of Lemma 8. We prove part (a) first. Using (2.1) and Assumption 4W(a),

$$\begin{aligned} & y_{1i}-\beta_0' y_{2i}-\widehat{\gamma}_n(\beta_0)'X_i \\ &= (\beta-\beta_0)'y_{2i}-(\widehat{\gamma}_n(\beta_0)-\gamma_1)'X_i+u_i \\ &= (\beta-\beta_0)'C'\widetilde{Z}_i n^{-1/2}-(\widehat{\gamma}_n(\beta_0)-\gamma_1-\xi_1(\beta-\beta_0))'X_i+u_i+(\beta-\beta_0)'v_{2i}. \end{aligned} \tag{7.29}$$

In consequence, we apply Lemma 6 with

$$\begin{aligned} \Psi_n(\widehat{\tau}_n) &= n^{-1}Z'R_\varphi,\ Q_i=u_i+(\beta-\beta_0)'v_{2i},\ c_i=Z_i,\ d_i=(\widetilde{Z}_i',X_i')', \\ \widehat{\tau}_n &= \begin{pmatrix} -C(\beta-\beta_0)n^{-1/2} \\ \widehat{\gamma}_n(\beta_0)-\gamma_1-\xi_1(\beta-\beta_0) \end{pmatrix},\ \text{and } h=g. \end{aligned} \tag{7.30}$$

Note that $\overline{c}_n=\overline{Z}_n=0$ because X_i contains an intercept by Assumption 2(b) and $Z'X=0$ by construction. The required conditions of Lemma 6 on d_i are satisfied by Assumption 2. The assumptions on Q_i are satisfied by Assumptions 1(a), 4W(b), and 4W(c). The condition $n^{1/2}\widehat{\tau}_n=O_p(1)$ holds by Assumption 4W(f).

We now verify the conditions of Lemma 6 on $c_i=Z_i$. By construction, $Z_i=\widetilde{Z}_i-\widetilde{Z}'X(X'X)^{-1}X_i$, where $\widetilde{Z}'X(X'X)^{-1}\rightarrow_p D_{12}D_{22}^{-1}$ by Assumption 2(c). In consequence, by standard arguments using Assumption 2(c) and 2(d), we obtain $\lim_{n\rightarrow\infty}n^{-1}\sum_{i=1}^{n}||Z_i||^2<\infty$ and $\lim_{n\rightarrow\infty}\max_{1\leq i\leq n}||Z_i||^2/n=0$. Hence, all of the conditions of Lemma 6 hold.

Now, using (7.30) and $\overline{Z}_n=0$, $\dot{A}_n(0)n^{1/2}\widehat{\tau}_n(\int_0^1\varphi(x,g)\varphi(x)\,dx)^{-1}$ equals

$$n^{-1}\sum_{i=1}^{n}Z_i\widetilde{Z}_i'C(\beta-\beta_0)-n^{-1}\sum_{i=1}^{n}Z_iX_i'n^{1/2}(\widehat{\gamma}_n(\beta_0)-\gamma_1-\xi_1(\beta-\beta_0)). \tag{7.31}$$

The second summand is zero because $Z'X=0$. The first summand equals $n^{-1}Z'ZC(\beta-\beta_0)$ because $Z'\widetilde{Z}=Z'M_X\widetilde{Z}=Z'Z$. Hence, by Lemma 6(b), we have

$$n^{-1/2}Z'R_\varphi=n^{1/2}\Psi_n(0)+n^{-1}Z'ZC(\beta-\beta_0)\int_0^1\varphi(x,g)\varphi(x)\,dx+o_p(1). \tag{7.32}$$

(By definition, $n^{1/2}\Psi_n(0) = n^{-1/2}Z'R^0_\varphi$, where R^0_φ is the n-vector whose i-th element is $\varphi(R_i/(n+1))$ and R_i is the rank of $u_i + (\beta-\beta_0)'v_{2i}$ in $\{u_j + (\beta-\beta_0)'v_{2j} : j \leq n\}$.)

Finally, Lemma 6(c) implies that

$$n^{1/2}\Psi_n(0) = n^{-1/2}Z'\Phi + o_p(1). \tag{7.33}$$

Combining (7.32), (7.33), and the definition of $\ell^\varphi_{g,\beta-\beta_0}$ in (4.5) establishes Lemma 8(a).

Lemma 8(b) follows from part (a) and Assumption 2(c).

Lemma 8(c) follows from Assumption 2(c) and $Z'Z = \widetilde{Z}'\widetilde{Z} - \widetilde{Z}'X(X'X)^{-1}X'\widetilde{Z}$. Positive definiteness of D_Z follows from that of D.

Lemma 8(d) follows from the Lindeberg CLT for triangular arrays applied to $n^{-1/2}Z'[\Phi c_\varphi^{-1/2} \colon Y_2 - EY_2]$ plus the facts that

$$\begin{aligned}
&n^{-1/2}Z'(ZC\ell^\varphi_{g,\beta-\beta_0}n^{-1/2}) = D_Z C\ell^\varphi_{g,\beta-\beta_0} + o(1),\\
&n^{-1/2}Z'EY_2 = Z'ZCn^{-1} = D_Z C + o(1), \text{ and}\\
&Var(n^{-1/2}\mu_1'Z'[\Phi c_\varphi^{-1/2} \colon Y_2 - EY_2]\mu_2) \to \mu_2'\Omega_{\varphi g}\mu_2 \cdot \mu_1' D_Z \mu_1,
\end{aligned} \tag{7.34}$$

for arbitrary fixed non-zero vectors $\mu_1 \in R^k$ and $\mu_2 \in R^{m+1}$. Note that $EZ'\Phi = 0$ because $\overline{Z}_n = 0$ and $E\varphi(U_{gi})$ does not depend on i.

The Lindeberg condition is verified for $n^{-1/2}\mu_1'Z'[\Phi c_\varphi^{-1/2} \colon Y_2 - EY_2]\mu_2$ (for μ_1 and μ_2 as above), as follows. Let $\zeta_i = (\varphi^2(U_{gi})c_\varphi^{-1/2}, v_{2i}')\mu_2 \in R$. For any $\varepsilon > 0$,

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n (\mu_1'Z_i)^2 E\zeta_i^2 1((\mu_1'Z_i)^2\zeta_i^2 > n\varepsilon)\\
&\leq n^{-1}\sum_{j=1}^n (\mu_1'Z_j)^2 \cdot E\zeta_i^2 1(\max_{j\leq n}(\mu_1'Z_j)^2\zeta_i^2 > n\varepsilon) \to 0,
\end{aligned} \tag{7.35}$$

where the inequality uses $(\mu'Z_i)^2 \leq \max_{j\leq n}(\mu'Z_j)^2$ in the indicator function and the convergence to zero holds by Assumption 2, $E||v_{2i}||^2 < \infty$ (by Assumption 1(b)), $E\varphi^2(U_{gi}) < \infty$ (by Assumption 3), and the dominated convergence theorem. □

Proof of Lemma 9. We prove part (a) first. Let V_2 be the $n \times m$ matrix whose i-th row is v_{2i}'. Using $Z'X = 0$, we have

$$\begin{aligned}
\widehat{\nu}_{\varphi n} = \; & n^{-1}V_2'R_\varphi c_\varphi^{-1/2} - n^{-1}V_2'Z(n^{-1}Z'Z)^{-1}n^{-1}Z'R_\varphi c_\varphi^{-1/2}\\
& - n^{-1}V_2'X(n^{-1}X'X)^{-1}n^{-1}X'R_\varphi c_\varphi^{-1/2}.
\end{aligned} \tag{7.36}$$

We have $n^{-1}V_2'Z \to_p 0$ and $n^{-1}V_2'X \to_p 0$ because they have mean zero and variance $O(n^{-1})$ by Assumptions 1, 2(a), and 2(c). Assumption 2(c) implies that $(n^{-1}Z'Z)^{-1}$ and $(n^{-1}X'X)^{-1}$ are $O_p(1)$. Lemma 8(a) and (d) implies that $n^{-1/2}Z'R_\varphi = O_p(1)$. These results combine to show that the second term on the right-hand side (rhs) of (7.36) is $o_p(1)$. Next, we have

$$n^{-1}||X'R_\varphi|| = n^{-1}||\sum_{i=1}^n X_i \varphi\left(\frac{\widehat{R}_i(\beta_0)}{n+1}\right)|| \leq Cn^{-1}\sum_{i=1}^n ||X_i|| = O(1) \tag{7.37}$$

for some constant $C < \infty$, using the triangle inequality, the boundedness of φ, and Assumption 2(c). This result and the others above imply that the third term on the rhs of (7.36) is $o_p(1)$. Hence, $\widehat{\nu}_{\varphi n} = n^{-1}V_2' R_\varphi c_\varphi^{-1/2} + o_p(1)$.

We apply Lemma 7 with $Q_{1i} = u_i + (\beta - \beta_0)' v_{2i}$, $Q_{2i} = v_{2i}$, and d_i and $\widehat{\tau}_n$ as in (7.30) to get

$$\begin{aligned} & n^{-1}V_2' R_\varphi c_\varphi^{-1/2} = E[\varphi(G(u_i + (\beta - \beta_0)' v_{2i}))v_{2i}]c_\varphi^{-1/2} + o_p(1) \\ & = Cov[\varphi(G(u_i + (\beta - \beta_0)' v_{2i})), y_{2i}]c_\varphi^{-1/2} + o_p(1) = \nu_{\varphi g} + o_p(1). \end{aligned} \tag{7.38}$$

The conditions of Lemma 7 on $\widehat{\tau}_n$, d_i, and (Q_{1i}, Q_{2i}) hold by Assumptions 4W(f), 2(c), and 4W(d), respectively.

Next, we prove part (b). For simplicity, we replace $n-k-p$ by n in the definition of $\widehat{\Omega}_{22n}$. We have

$$\begin{aligned} \widehat{\Omega}_{22n} &= n^{-1}Y_2(I_n - P_Z - P_X)' Y_2 \\ &= n^{-1}(V_2' V_2 - V_2' P_Z V_2 - V_2' P_X V_2) \to_p \Omega_{22}, \end{aligned} \tag{7.39}$$

where V_2 denotes the $n \times m$ matrix whose i-th row is v_{2i}', $n^{-1}V_2'V_2 \to_p \Omega_{22}$ by Kolmogorov's LLN for iid random variables, and $n^{-1}Z'V_2 \to_p 0$ and $n^{-1}X'V_2 \to_p 0$ because they have mean zero and variances that are $O(n^{-1})$ by Assumptions 1, 2(a), and 2(c). □

Proof of Lemma 10. We prove part (a) first. It suffices to show that $(S_n^\varphi, n^{-1/2}T_n^\varphi) \to_d (S_{f\infty}^\varphi, \alpha_T^\varphi)$ conditional on an $\{\varepsilon_i : i \geq 1\}$ sequence that satisfies certain properties, and that $\{\varepsilon_i : i \geq 1\}$ sequences satisfy these properties with probability one. Because conditional probabilities are bounded by zero and one, this implies that $(S_n^\varphi, n^{-1/2}T_n^\varphi) \to_d (S_{f\infty}^\varphi, \alpha_T^\varphi)$ unconditionally by the bounded convergence theorem. The desired properties are

$$\lim_{n\to\infty} \max_{1 \leq i \leq n} ||\varepsilon_i - \overline{\varepsilon}_n||^2 / \sum_{i=1}^n ||\varepsilon_i - \overline{\varepsilon}_n||^2 = 0, \tag{7.40}$$

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^n ||\varepsilon_i - \overline{\varepsilon}_n||^2 < \infty, \tag{7.41}$$

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \widetilde{Z}_i \varepsilon_i' = 0, \text{ and} \tag{7.42}$$

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^n X_i \varepsilon_i' = 0. \tag{7.43}$$

Conditions (7.40) and (7.41) hold a.s. by Assumption 4S(d), Lemma 12(b), and Kolmogorov's strong LLN. Conditions (7.42) and (7.43) hold a.s. by Assumptions 4S(d) and 4S(h) and the strong LLN of Thm. 5.2.1 of Chow and Teicher (1978, p. 121) applied with $\alpha_n = 2$. Consequently, sequences $\{\varepsilon_i : i \geq 1\}$ that satisfy (7.40)-(7.43) occur with probability one.

Using (2.1) and Assumptions 4S(a)-(c), we have

$$\begin{aligned} &y_{1i} - \beta_0' y_{2i} - \widehat{\gamma}_n(\beta_0)' X_i \\ &= B'\Pi'\widetilde{Z}_i n^{-1/2} - (\widehat{\gamma}_n(\beta_0) - \gamma_1 - \xi_1 B n^{-1/2})' X_i + B'\varepsilon_i n^{-1/2} + (1 + \rho' B n^{-1/2}) u_i. \end{aligned} \tag{7.44}$$

Let $\zeta_n = (1 + \rho' B n^{-1/2})^{-1}$. Since $\zeta_n > 0$ for n sufficiently large, $\{\widehat{R}_i(\beta_0) : i \leq n\}$ are equal to the ranks of the iid random variables $\{u_i : i \leq n\}$ plus the terms

$$\left\{ \zeta_n B'\Pi'\widetilde{Z}_i n^{-1/2} - \zeta_n(\widehat{\gamma}_n(\beta_0) - \gamma_1 - \xi_1 B n^{-1/2})' X_i + \zeta_n B'\varepsilon_i n^{-1/2} : i \leq n \right\}. \tag{7.45}$$

Hence, we apply Lemma 6, conditional on an $\{\varepsilon_i : i \geq 1\}$ sequence that satisfies (7.40)-(7.43), with

$$\begin{aligned} &\Psi_n(\widehat{\tau}_n) = n^{-1} Z' R_\varphi, \ Q_i = u_i, \ c_i = Z_i, \\ &d_i = \begin{pmatrix} \widetilde{Z}_i \\ X_i \\ \varepsilon_i \end{pmatrix}, \ \widehat{\tau}_n = \begin{pmatrix} -\zeta_n \Pi B n^{-1/2} \\ \zeta_n(\widehat{\gamma}_n(\beta_0) - \gamma_1 - \xi_1 B n^{-1/2}) \\ -\zeta_n B n^{-1/2} \end{pmatrix}, \text{ and } h = f. \end{aligned} \tag{7.46}$$

The assumptions of Lemma 6 on Q_i are satisfied by Assumptions 1 and 4S(e). The required conditions for c_i are verified by the same argument as in the proof of Theorem 1. The assumptions on d_i are satisfied by Assumption 2, (7.40), and (7.41). The assumptions on $\widehat{\tau}_n$ are satisfied by Assumption 4S(i) because $\zeta_n \to 1$.

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

$$\begin{aligned} &\zeta_n n^{-1} \sum_{i=1}^n Z_i \widetilde{Z}_i' \Pi B - \zeta_n n^{-1} \sum_{i=1}^n Z_i X_i' n^{1/2} (\widehat{\gamma}_n(\beta_0) - \gamma_1 - \xi_1 B n^{-1/2}) \\ &+ \zeta_n n^{-1} \sum_{i=1}^n Z_i \varepsilon_i' B. \end{aligned} \tag{7.47}$$

The first term in (7.47) equals $Z'Z\Pi B + o(1)$ because $\zeta_n \to 1$. The second term is zero because $Z'X = 0$. The third term equals

$$\zeta_n n^{-1} \sum_{i=1}^n \widetilde{Z}_i \varepsilon_i' B - \zeta_n (n^{-1} \widetilde{Z}' X)(n^{-1} X' X)^{-1} n^{-1} \sum_{i=1}^n X_i \varepsilon_i' B = o(1), \tag{7.48}$$

using (7.42), (7.43), and Assumption 2(c). Hence, by Lemma 6(b) and (c), we have

$$n^{-1/2} Z' R_\varphi = n^{-1/2} Z' \Phi + Z' Z \Pi B \int_0^1 \varphi(x, f) \varphi(x)\, dx + o_p(1), \tag{7.49}$$

which establishes part (a).

Lemma 10(b) follows from part (a) and Assumption 2(c).

To establish Lemma 10(c), we have

$$n^{-1} Z' Y_2 = n^{-1} Z'(Z\Pi + X\xi + V_2) = n^{-1} Z' Z \Pi + n^{-1} Z' V_2 \to_p D_Z \Pi, \tag{7.50}$$

where V_2 denotes the $n \times m$ matrix whose i-th row is v_{2i}' and using $n^{-1}Z'V_2 \to_p 0$ because its mean is zero and its variance is $O(n^{-1})$ by Assumptions 1, 2(a), and 2(c).

In addition, we have

$$n^{-1}Z'\Phi = n^{-1}Z'(\Phi - E\Phi) \to_p 0, \tag{7.51}$$

where the equality holds because $E\Phi$ is proportional to 1_n and $Z'1_n = 0$ and the convergence to 0 holds by the strong LLN referenced in the previous paragraph. $\square$

Proof of Lemma 11. We prove part (a) first. By the same argument as in the proof of Lemma 9(a), but with Lemma 8 replaced by Lemma 10, we have $\widehat{\nu}_{\varphi n} = n^{-1}V_2'R_\varphi c_\varphi^{-1/2} + o_p(1)$. As in the proof of Lemma 10(a), it suffices to show the result conditional on an $\{\varepsilon_i : i \geq 1\}$ sequence that satisfies certain properties and that $\{\varepsilon_i : i \geq 1\}$ sequences satisfy these properties with probability one. In the present case, we need the property

$$\overline{\lim_{n\to\infty}} n^{-1} \sum_{i=1}^{n} ||\varepsilon_i|| < \infty. \tag{7.52}$$

Condition (7.52) holds a.s. by Kolmogorov's SLLN using Assumption 4S(d).

Lemma 7 applied conditional on a sequence $\{\varepsilon_i : i \geq 1\}$ that satisfies (7.52), with $(Q_{1i}, Q_{2i}) = (u_i, v_{2i})$ and d_i and $\widehat{\tau}_n$ as in (7.46), gives

$$\begin{aligned} &n^{-1}V_2'R_\varphi c_\varphi^{-1/2} = E[\varphi(F(u_i))v_{2i}]c_\varphi^{-1/2} + o_p(1) \\ &= Cov[\varphi(F(u_i)), y_{2i}]c_\varphi^{-1/2} + o_p(1) = \nu_{\varphi f} + o_p(1). \end{aligned} \tag{7.53}$$

The conditions of Lemma 7 on $\widehat{\tau}_n$, d_i, and (Q_{1i}, Q_{2i}) hold by Assumption 4S(i), condition (7.52) and Assumptions 2(c) and 2(d), and Assumption 4S(f), respectively.

The proof of part (b) is the same as for Lemma 9(b). $\square$

Proof of Lemma 12. Part (a) holds because $E\sum_{i=1}^{\infty} \psi_i^{1+\delta}/i^{1+\delta} = E\psi_1^{1+\delta}\sum_{i=1}^{\infty} i^{-(1+\delta)} < \infty$ implies that $\sum_{i=1}^{\infty} \psi_i^{1+\delta}/i^{1+\delta} < \infty$ 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 $n^{-1-\delta}\sum_{i=1}^{n} \psi_i^{1+\delta} \to 0$ a.s. Hence, $n^{-1-\delta}\max_{i\leq n} \psi_i^{1+\delta} \leq n^{-1-\delta}\sum_{i=1}^{n} \psi_i^{1+\delta} \to 0$ a.s. In turn, this gives $n^{-1}\max_{i\leq n} \psi_i \to 0$ a.s. $\square$

Proof of Lemma 13. We prove the lemma by verifying the conditions of Pollard's (1990) Thm. 8.3. To match the notation in Pollard (1990), view the sequence $\{(Q_{1i}, Q_{2i}) : i \geq 1\}$ as depending on $\omega \in \Omega$, where the probability space is $\{\Omega, \mathcal{F}, \mathbb{P}\}$, and let $(Q_{1i}, Q_{2i})(\omega)$ denote the i-th element of this sequence. Also, view the sequence of independent processes

$$\{h_{ni}(q_1, q_2, t) : (q_1, q_2, t) \in \mathcal{T} \subset R^{m+1+\delta_d}\} \tag{7.54}$$

for $i \geq 1$ as a sequence of independent processes indexed by $\tau \in \mathcal{T}$:

$$\begin{aligned} &\{h_{ni}(\omega;\tau) : \tau \in \mathcal{T}\}, \text{ where } \tau = (q_1, q_2, t) \text{ and} \\ &h_{ni}(\omega;\tau) = 1((Q_{1i}, Q_{2i})(\omega) \leq (q_1 + d_i'tn^{-1/2}, q_2)). \end{aligned} \tag{7.55}$$

Each of the processes $h_{ni}(\omega;\tau)$ has envelope $H_{ni}(\omega) = 1\ \forall\omega \in \Omega$, and these envelope functions satisfy

$$\sum_{i=1}^{\infty} \frac{EH_{ni}}{i^2} = \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty, \tag{7.56}$$

which is the first condition of Theorem 8.3 of Pollard (1990).

Now, we verify that the processes $\{h_{ni}(\omega;\tau) : \tau \in \mathcal{T}\}$ and the envelope functions $\{H_{ni}(\omega) = 1\ \forall\omega \in \Omega\}$ for $i \geq 1$ satisfy the second condition of Theorem 8.3 of Pollard (1990). For each ω, define the sets

$$\begin{aligned} \mathcal{H}_{\omega n} &= \{(h_{n1}(\omega;\tau), ..., h_{nn}(\omega;\tau)) \in R^n : \tau \in \mathcal{T}\} \text{ and} \\ \alpha \odot \mathcal{H}_{\omega n} &= \{(\alpha_1 h_{n1}(\omega;\tau), ..., \alpha_n h_{nn}(\omega;\tau)) \in R^n : \tau \in \mathcal{T}\} \end{aligned} \tag{7.57}$$

for some $\alpha \in R^n$.

Denote the largest number κ for which there exist points in a subset of a metric space T with $d(t_i, t_j) > \varepsilon$, for $i \neq j$, by $D(\varepsilon, T)$. The number $D(\varepsilon, T)$ is called the *packing number*. Denote the ℓ_1 distance in R^n by $|\alpha|_1 = \sum_{i=1}^{n} |\alpha_i|$.

By Definition 7.9 of Pollard (1990), $\{h_{ni}(\omega;\tau) : \tau \in \mathcal{T}\}$ for $i \geq 1$ is *manageable* with respect to the envelopes $\{H_{ni}(\omega) = 1\ \forall\omega \in \Omega\}$ for $i \geq 1$ if there exists a function $\lambda(\varepsilon)$ such that

1. $\int_0^1 \sqrt{\lambda(\varepsilon)} d\varepsilon < \infty$,
2. $D(\varepsilon|\alpha|_1, \alpha \odot \mathcal{H}_{\omega n}) \leq \lambda(\varepsilon)$ for $0 < \varepsilon \leq 1$, all $\omega \in \Omega$, all vectors of nonnegative weights α and all n.[16]

The second condition of Theorem 8.3 of Pollard (1990) is that $\{h_{ni}(\omega;\tau) : \tau \in \mathcal{T}\}$ for $i \geq 1$ is *manageable* with respect to the envelopes $\{H_{ni}(\omega) = 1\ \forall\omega \in \Omega\}$ for $i \geq 1$.

For any ω, the class $\mathcal{H}_{\omega n}$ belongs to a larger class of functions $\mathcal{H}$ defined by

$$\mathcal{H} = \{h | h(q_1, q_2) = 1((q_1, q_2) \in C) \text{ for } C \text{ of the type } (-\infty, c_1] \times (-\infty, c_3]^m\}. \tag{7.58}$$

The collection of all cells $(-\infty, c_1] \times (-\infty, c_2]^m$ has VC-index equal to $(m+1)+1$, which implies that the class of indicator functions $\mathcal{H}$ has VC-index equal to $(m+1)+1$ as well.[17] From Theorem 2.6.7 in van der Vaart and Wellner (1996), it follows that there exist constants A_1 and W such that

$$N(\varepsilon/2, \mathcal{H}) \leq A_1(\varepsilon/2)^{-W} \text{ for } 0 < \varepsilon \leq 2, \tag{7.59}$$

where $N(\varepsilon/2, \mathcal{H})$ is the smallest number of closed balls with radius $\varepsilon/2$ that covers $\mathcal{H}$. The number $N(\varepsilon/2, \mathcal{H})$ is called the *covering number* of $\mathcal{H}$. Since $D(\varepsilon, \mathcal{H}) \leq N(\varepsilon/2, \mathcal{H})$ and $\mathcal{H}_{\omega n} \subset \mathcal{H}$ for every n and ω, it follows that there exist constants A_2 and W,

$$D(\varepsilon, \mathcal{H}_{\omega n}) \leq A_2 \varepsilon^{-W} \text{ for } 0 < \varepsilon \leq 1. \tag{7.60}$$

[16] Because $H_{ni}(\omega) = 1\ \forall\omega$ we have that $|\alpha \odot \mathbf{H}|_1 = |\alpha|_1$, where $\mathbf{H} = (H_{ni}(\omega), ..., H_{nn}(\omega))$.

[17] See van der Vaart and Wellner (1996).

Now, using an argument similar to the one used in the proof of Theorem 4.8 of Pollard (1990), we can show that for all n and ω, there exist constants A_3 and W such that

$$D(\varepsilon|\alpha|_1, \alpha \odot \mathcal{H}_{\omega n}) \leq A_3 \varepsilon^{-W} \text{ for } 0 < \varepsilon \leq 1. \tag{7.61}$$

Take $\mathcal{H}^*_{\omega n}$ to be the set of rescaled coordinates

$$h^*_{ni} = \frac{\alpha_i h_{ni}}{2\sum_{i=1}^n \alpha_i} \text{ for } h_n \in \mathcal{H}_{\omega n}. \tag{7.62}$$

Let h^*_1 and h^*_2 in $\mathcal{H}^*_{\omega n}$ be rescaled coordinates of h_1 and h_2 in $\mathcal{H}_{\omega n}$. Then,

$$|h^*_1 - h^*_2|_1 \leq \sum_{i=1}^n \left| \frac{\alpha_i}{2\sum_{i=1}^n \alpha_i} \right| |h_{1i} - h_{2i}| \leq |h_1 - h_2|_1. \tag{7.63}$$

Hence,

$$D(\varepsilon, \mathcal{H}^*_{\omega n}) \leq A_2 \varepsilon^{-W} \text{ for } 0 < \varepsilon \leq 1. \tag{7.64}$$

Now, we have

$$\begin{aligned} |h^*_1 - h^*_2|_1 < \varepsilon/2 &\Leftrightarrow \sum_{i=1}^n \left| \frac{\alpha_i}{2\sum_{i=1}^n \alpha_i}(h_{1i} - h_{2i}) \right| < \varepsilon/2 \\ &\Leftrightarrow |\alpha \odot h_1 - \alpha \odot h_2|_1 < \varepsilon \left| \sum_{i=1}^n \alpha_i \right| < \varepsilon |\alpha|_1. \end{aligned} \tag{7.65}$$

Therefore, (7.61) holds with $A_3 = 2^W A_2$. This establishes that $\{h_{ni}(\omega; \tau) : \tau \in \mathcal{T}\}$ is *manageable* with respect to the envelopes $\{H_{ni}(\omega) = 1 \ \forall \omega \in \Omega\}$. Theorem 8.3 of Pollard (1990) then gives

$$n^{-1} \sup_{\tau \in \mathcal{T}} \left| \sum_{i=1}^n (h_{ni}(\omega; \tau) - E h_{ni}(\omega; \tau)) \right| \to 0 \text{ a.s.}, \tag{7.66}$$

which gives the result of the Lemma. □

Proof of Lemma 1. By definition of $\widehat{\gamma}_n^{LS}(\beta_0)$, we have

$$\begin{aligned} \widehat{\gamma}_n^{LS}(\beta_0) &= \left(n^{-1} X'X\right)^{-1} n^{-1} \sum_{i=1}^n X_i((\beta - \beta_0)' y_{2i} + \gamma_1' X_i + u_i) \\ &= \gamma_1 + \xi(\beta - \beta_0) + \left(n^{-1} X'X\right)^{-1} n^{-1} \sum_{i=1}^n X_i(u_i + (\beta - \beta_0)' v_{2i}), \end{aligned} \tag{7.67}$$

using $y_{2i} = \Pi' Z_i + \xi' X_i + v_{2i}$ and $X'Z = 0$. Hence, we obtain

$$\begin{aligned} n^{1/2}(\widehat{\gamma}_n^{LS}(\beta_0) - \gamma_1 - \xi_1(\beta - \beta_0)) &= \left(n^{-1} X'X\right)^{-1} n^{-1/2} \sum_{i=1}^n X_i(u_i + (\beta - \beta_0)' v_{2i}) \\ &\quad + n^{1/2}(\xi - \xi_1)(\beta - \beta_0). \end{aligned} \tag{7.68}$$

Assumption 2(c) implies that $\left(n^{-1}X'X\right)^{-1} = D_{11}^{-1} + o(1)$. The second multiplicand of the first term on the rhs of (7.68) is asymptotically normal by the Lindeberg central limit theorem using Assumptions 1 and 2 and $Eu_i^2 < \infty$. The Lindeberg condition is verified by an argument analogous to that in (7.35), where $E(u_i + (\beta - \beta_0)'v_{2i})^2 < \infty$ by Assumption 1(b) and $Eu_i^2 < \infty$. Thus, the first term on the rhs of (7.68) is $O(1)$. Next, we have

$$\xi - \xi_1 = \left(n^{-1}X'X\right)^{-1} n^{-1}X'\widetilde{Z}\Pi = \left(n^{-1}X'X\right)^{-1} n^{-1}X'\widetilde{Z}Cn^{-1/2} = O(n^{-1/2}), \tag{7.69}$$

where the first equality holds by the definition of ξ stated following (2.3), the second equality holds by Assumption 4W(a), and the last equality holds by Assumption 2(c). Assumption 4W(b) states that $\beta - \beta_0$ is a constant. Hence, $n^{1/2}(\xi - \xi_1)(\beta - \beta_0) = O(1)$, which completes the proof of the Lemma. □

Proof of Corollary 1. We have

$$\begin{aligned}
&P\left(RLR_n^{\varphi} > \kappa_{LR,\alpha}(T_n^{\varphi\prime}T_n^{\varphi}, k, m)\right) = P\left(LR_\infty(S_n^{\varphi}, T_n^{\varphi}) > \kappa_{LR,\alpha}(T_n^{\varphi\prime}T_n^{\varphi}, k, m)\right)\\
&\to P\left(LR_\infty(S_\infty^{\varphi}, T_\infty^{\varphi}) > \kappa_{LR,\alpha}(T_\infty^{\varphi\prime}T_\infty^{\varphi}, k, m)\right)\\
&= \int P\left(LR_\infty(S_\infty^{\varphi}, t) > \kappa_{LR,\alpha}(t't, k, m)\right) dF_{T_\infty^{\varphi}}(t) = \alpha,
\end{aligned} \tag{7.70}$$

where $F_{T_\infty^{\varphi}}(\cdot)$ is the df of T_∞^{φ}, the convergence holds by Theorem 1(a) and the continuous mapping theorem, the second equality holds by the independence of S_∞^{φ} and T_∞^{φ}, and the last equality holds by the definition of $\kappa_{LR,\alpha}(t't, k, m)$ in (3.10) and the fact that $S_\infty^{\varphi} \sim N(0, I_k)$ under the null by (4.6). □

Proof of Lemma 2. The proof is very similar to that of Theorem 1 with $Yb_0\widehat{\sigma}_n^{-1}$ in place of Φc_φ^{-1}. First, by the same proof as for Lemma 9(b) but with (Y_2, V_2) replaced by (Y, V), we get $\widehat{\Omega}_n \to_p \Omega$, where V is the $n \times (m+1)$ matrix with i-th row equal to v_i'. This implies $\widehat{\sigma}_n^2 \to_p b_0'\Omega b_0 = \sigma_g^2$ and $\widehat{\Omega}_{*n} \to_p \Omega_g$. Next, we need the following analogues of Lemma 8(a), (b), and (d):

$$\begin{aligned}
&n^{-1/2}Z'Yb_0 = n^{-1/2}Z'(Yb_0 - EYb_0 + ZC(\beta - \beta_0)n^{-1/2}), &(7.71)\\
&S_n = (Z'Z)^{-1/2}Z'(Yb_0\sigma_g^{-1} - EYb_0\sigma_g^{-1} + ZC(\beta - \beta_0)\sigma_g^{-1}n^{-1/2}) + o_p(1), &(7.72)\\
&n^{-1/2}Z'[(Vb_0\sigma_g^{-1} + ZC(\beta - \beta_0)\sigma_g^{-1}n^{-1/2}) : Y_2] \to_d [N_1 : N_2], &(7.73)
\end{aligned}$$

where (7.71) holds by (2.1), (2.3), Assumption 1(a), and $Z'X = 0$, (7.72) holds by (7.71) and $\widehat{\sigma}_n^2 \to_p \sigma_g^2$, and (7.73) holds by the same proof as that of Lemma 8(d) (given below) except with $\Phi c_\varphi^{-1/2}$, $\ell_{g,\beta-\beta_0}^{\varphi}$, and $\varphi(U_{gi})c_\varphi^{-1/2}$ replaced by $Vb_0\sigma_g^{-1}$, $(\beta - \beta_0)\sigma_g^{-1}$, and $v_{2i}'b_0\sigma_g^{-1}$, respectively, and with $E(v_{2i}'b_0)^2 < \infty$ by the assumption that Ω is well-defined. Given these analogues of Lemma 8(a), (b), and (d), the rest of the proof of Lemma 2 is the same as that of Theorem 1. □

Proof of Lemma 3. We establish part (a) first. Let σ_L^2 denote the variance of the df L. Since $\varphi(x) = L^{-1}(x)$, $c_\varphi = \sigma_L^2$ by change of variables. Also, we have

$$\begin{aligned}\sigma_L^2 &= Var((u_i + (\beta - \beta_0)'v_{2i})\kappa) = \sigma_g^2\kappa^2,\ \kappa = \sigma_L/\sigma_g, \text{ and}\\ G(x) &= L(\kappa x) = L(\sigma_L x/\sigma_g).\end{aligned} \tag{7.74}$$

Combining these results gives the result of part (a):

$$\begin{aligned}\varphi(U_{gi})c_\varphi^{-1/2} &= \varphi(G[(u_i + (\beta - \beta_0)'v_{2i})])c_\varphi^{-1/2}\\ &= L^{-1}(L[\sigma_L(u_i + (\beta - \beta_0)'v_{2i})/\sigma_g])\sigma_L^{-1}\\ &= u_i + (\beta - \beta_0)'v_{2i})/\sigma_g.\end{aligned} \tag{7.75}$$

Part (a) implies that $\nu_{\varphi g} = \nu_g$ and so, $\Omega_{\varphi g} = \Omega_g$ and part (b) holds.

For part (c), we have

$$\begin{aligned}&\int_0^1 \varphi(x,g)\varphi(x)dx = -\int_0^1 \frac{g'(G^{-1}(x))}{g(G^{-1}(x))}L^{-1}(x)dx = -\int_{-\infty}^{\infty} g'(y)L^{-1}(G(y))dy\\ &= -\int_{-\infty}^{\infty} g'(y)L^{-1}(L(\kappa y))dy = -\kappa\int_{-\infty}^{\infty} g'(y)ydy = \kappa = \sigma_L/\sigma_g,\end{aligned} \tag{7.76}$$

where the second equality holds by change of variables with $y = G^{-1}(x)$, the third and last equalities hold by (7.74), and the fourth equality holds by integration by parts. Combining (7.76) with $c_\varphi = \sigma_L^2$ establishes part (c).

Part (d) follows from parts (b) and (c) and the definitions of N_φ, N_1, S_∞^φ, S_∞, T_∞^φ, and T_∞. $\square$

Proof of Lemma 4. The proof is like that of Theorem 2 with $Yb_0\widehat{\sigma}_n^{-1}$ in place of $R_\varphi c_\varphi^{-1}$. By essentially the same proof as for Lemma 9(b) but with (Y_2, V_2) replaced by (Y, V), we get $\widehat{\Omega}_n - Ev_iv_i' \to_p 0$. Under Assumption 4S(a), we have

$$\begin{aligned}&v_{1i} = u_i + (\beta_0 + Bn^{-1/2})'v_{2i},\quad Ev_iv_i' \to E\begin{pmatrix} u_i + \beta_0'v_{2i}\\ v_{2i}\end{pmatrix}\begin{pmatrix} u_i + \beta_0'v_{2i}\\ v_{2i}\end{pmatrix}',\\ &b_0'E\begin{pmatrix} u_i + \beta_0'v_{2i}\\ v_{2i}\end{pmatrix}\begin{pmatrix} u_i + \beta_0'v_{2i}\\ v_{2i}\end{pmatrix}'b_0 = Eu_i^2 = \sigma_f^2, \text{ and}\\ &H'E\begin{pmatrix} u_i + \beta_0'v_{2i}\\ v_{2i}\end{pmatrix}\begin{pmatrix} u_i + \beta_0'v_{2i}\\ v_{2i}\end{pmatrix}'b_0 = Ev_{2i}u_i.\end{aligned} \tag{7.77}$$

In consequence,

$$\widehat{\sigma}_n^2 = b_0'\widehat{\Omega}_nb_0 \to_p \sigma_f^2,\ \widehat{\nu}_n = H'\widehat{\Omega}_nb_0\widehat{\sigma}_n^{-1} \to_p Ev_{2i}u_i\sigma_f^{-1} = \nu_f, \text{ and } \widehat{\Omega}_{*n} \to_p \Omega_f. \tag{7.78}$$

We need the following analogue of Lemma 10(b):

$$S_n = (Z'Z)^{-1/2}Z'(Yb_0\sigma_f^{-1} - EYb_0\sigma_f^{-1} + ZB\sigma_f^{-1}n^{-1/2}) + o_p(1), \tag{7.79}$$

which holds by (2.1), (2.3), Assumption 1(a), $Z'X = 0$, and $\widehat{\sigma}_n^2 \to_p \sigma_f^2$. Next, an analogue of (7.3) holds with $\widehat{\Omega}_{\varphi n}$, $\Omega_{\varphi f}$, and $n^{-1}Z'R_\varphi$ replaced by $\widehat{\Omega}_{*n}$, Ω_f, and $n^{-1}Z'Yb_0 = n^{-1}Z'(Vb_0 + Z\Pi B n^{-1/2})$, respectively, using the result that $\widehat{\Omega}_{*n} \to_p \Omega_f$ and the fact that $n^{-1}Z'Vb_0 \to_p 0$ because its mean is zero and its variance is $O(n^{-1})$. Given these analogues of Lemma 10(b) and (7.3), the rest of the proof of Lemma 4 is the same as that of Theorem 2. □

Proof of Lemma 5. The proofs of parts (a)-(c) are analogous to those of parts (a)-(c) of Lemma 3. Part (d) follows from parts (b) and (c) and the definitions of $S_{f\infty}^{\varphi}$, $S_{f\infty}$, α_T^{φ}, and α_T. □

Proof of an Alternative Expression for T_n:

We now provide a proof of (3.6), which gives an alternative expression for T_n from its definition in (2.6). Let $M = [b_0\widehat{\sigma}_{gn}^{-1} : H] \in R^{(m+1)\times(m+1)}$. Straightforward calculations yield

$$\begin{aligned} YM &= [Yb_0\widehat{\sigma}_{gn}^{-1} : Y_2],\ M'A_0 = H,\ \widehat{\Omega}_{*n} = M'\widehat{\Omega}_n M, \text{ and} \\ A_0'\widehat{\Omega}_n^{-1}A_0 &= A_0'M(M^{-1}\widehat{\Omega}_n^{-1}M'^{-1})M'A_0 = H'\widehat{\Omega}_{*n}^{-1}H. \end{aligned} \tag{7.80}$$

Using the definition of T_n in (2.6), we have

$$\begin{aligned} T_n &= (Z'Z)^{-1/2}Z'(YM)(M^{-1}\widehat{\Omega}_n^{-1}M'^{-1})(M'A_0)(A_0'\widehat{\Omega}_n^{-1}A_0)^{-1/2} \\ &= (Z'Z)^{-1/2}Z'[Yb_0\widehat{\sigma}_{gn}^{-1} : Y_2]\widehat{\Omega}_{*n}^{-1}H(H'\widehat{\Omega}_{*n}^{-1}H)^{-1/2}, \end{aligned} \tag{7.81}$$

where the second equality uses (7.80). The rhs of (7.81) is the expression in (3.6).

Asymptotic Power Calculations

Next, we describe the simulation method used to calculate the weak IV asymptotic power reported in Table I. The first step is to compute $\xi(\varphi^{NS}, g)$ and $\xi(\varphi^{WS}, g)$ when g is the density of $u_i + \beta v_{2i}$ for v_{2i} defined in (4.18) and u_i and ε_i are independent with distribution F. The idea is to use the fact that the Hodges-Lehmann estimator of location based on φ (e.g., defined in Hettmansperger (1984, eqn. (2.8.12), p. 99)) has asymptotic variance equal to $1/\xi(\varphi, g)$, see Hettmansperger (1984, Thm. 2.6.5 & eqn. (2.9.4), pp. 76 & 105). We compute the Hodges-Lehmann estimators based on φ^{NS} and φ^{WS} for 30,000 independent samples of a location model with density g and sample size 100,000. This yields 30,000 Hodges-Lehmann estimates $\widehat{\theta}^{NS}$ and $\widehat{\theta}^{WS}$. The reciprocals of the sample variances of these estimates yields estimated values of $\xi(\varphi^{NS}, g)$ and $\xi(\varphi^{WS}, g)$, denoted $\widetilde{\xi}(\varphi^{NS}, g)$ and $\widetilde{\xi}(\varphi^{WS}, g)$.

The second step is to compute the matrices $\Omega_{\varphi^{NS}g}$, $\Omega_{\varphi^{WS}g}$, and Ω_g and the scalar σ_g^2 defined in (4.1) and (4.10). The df $G(x)$ is approximated by the empirical df of 100,000 iid observations with distribution G (independent of the random variables above), call it $\widetilde{G}(x)$. Using the same observations, σ_g^2 is estimated by the sample variance, denoted $\widetilde{\sigma}_g^2$. Next, $\nu_{\varphi,g}$ is estimated by $\widetilde{\nu}_{\varphi,g} = R^{-1}\sum_{i=1}^{R}(\widetilde{v}_{2i} - \overline{\widetilde{v}}_{2R})\varphi(\widetilde{G}(\widetilde{X}_i))c_\varphi^{-1/2}$, where $R = 40,000$ for all distributions except the uniform,

$R = 100,000$ for the uniform distribution, $\overline{\widetilde{v}}_{2R} = R^{-1}\sum_{i=1}^{R}\widetilde{v}_{2i}$, $\widetilde{v}_{2i} = (1-\rho_{uv_2}^2)^{1/2}\widetilde{\varepsilon}_i + \rho_{uv_2}\widetilde{u}_i$, $\widetilde{X}_i = \widetilde{u}_i + \beta\widetilde{v}_{2i}$, $\widetilde{u}_i$ and $\widetilde{\varepsilon}_i$ are independent with distribution F, are independent of $\widetilde{G}(x)$, and are iid across $i = 1, ..., R$, and $\varphi = \varphi^{NS}, \varphi^{WS}$. ν_g is estimated by the sample covariance between $\widetilde{v}_{2i}$ and $\widetilde{X}_i\widetilde{\sigma}_g^{-1}$, denoted $\widetilde{\nu}_g$. ω_{22} is estimated by the sample variance of $\widetilde{v}_{2i}$, denoted $\widetilde{\omega}_{22}$. The matrices $\widetilde{\Omega}_{\varphi^{NS}g}$, $\widetilde{\Omega}_{\varphi^{WS}g}$, and $\widetilde{\Omega}_g$ are constructed using $\widetilde{\nu}_{\varphi^{NS},g}$, $\widetilde{\nu}_{\varphi^{WS},g}$, $\widetilde{\nu}_g$, and $\widetilde{\omega}_{22}$.

The third step is to compute 5,000 independent observations of (i) two independent k-variate normals $(\widetilde{S}_\infty^\varphi, \widetilde{T}_\infty^\varphi)$ with covariance matrices equal to I_k and means given by $\lambda^{1/2}\beta\widetilde{\xi}(\varphi, g)e_1$ and $\lambda^{1/2}(\beta\widetilde{\xi}^{1/2}(\varphi, g), 1)\widetilde{\Omega}_{\varphi g}^{-1}e_2(e_2'\widetilde{\Omega}_{\varphi g}^{-1}e_2)^{-1/2}e_1$, respectively, for $\varphi = \varphi^{NS}, \varphi^{WS}$, where $e_1 = (1, 0, ..., 0)' \in R^k$ and $e_2 = (0, 1)'$, and (ii) two independent k-variate normals $(\widetilde{S}_\infty, \widetilde{T}_\infty)$ with covariance matrices equal to I_k and means as in (i) but with $\widetilde{\sigma}_g^{-1}$ in place of $\widetilde{\xi}^{1/2}(\varphi, g)$. The same normal random variables were used for $(\widetilde{S}_\infty^{\varphi^{ws}}, \widetilde{T}_\infty^{\varphi^{ws}})$, $(\widetilde{S}_\infty^{\varphi^{ns}}, \widetilde{T}_\infty^{\varphi^{ns}})$, and $(\widetilde{S}_\infty, \widetilde{T}_\infty)$—just the means are different.

The last step is to compare each of the 5,000 WS-RCLR, NS-RCLR, CLR, LM, and AR test statistics based on $(\widetilde{S}_\infty^{\varphi^{ws}}, \widetilde{T}_\infty^{\varphi^{ws}})$, $(\widetilde{S}_\infty^{\varphi^{ns}}, \widetilde{T}_\infty^{\varphi^{ns}})$, and $(\widetilde{S}_\infty, \widetilde{T}_\infty)$ with the appropriate conditional critical value (determined by simulation) or unconditional critical value to determine whether the test rejects the null hypothesis. The fraction that reject the null hypothesis is the reported power in Table I.

References

Abrevaya, J. (1999): "Leapfrog Estimation of a Fixed-effects Model with Unknown Transformation of the Dependent Variable," *Journal of Econometrics,* 93, 203-228.

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. and V. Marmer (2004): "Exactly Distribution-free Inference in Instrumental Variables Regression with Possibly Weak Instruments," Cowles Foundation Discussion Paper No. 1501, Yale University. Available at http://cowles.econ.yale.edu.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2004): "Optimal Invariant Similar Tests for Instrumental Variables Regression with Weak Instruments," Cowles Foundation Discussion Paper No. 1476, Yale University, available at http://cowles.econ.yale.edu.

——— (2006a): "Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression with Weak Instruments," *Econometrica,* 74, forthcoming.

——— (2006b): "Supplement to 'Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression'," available on the *Econometric Society* website at www.econometricsociety.org under Supplemental Material and at http://ksghome.harvard.edu/∼ .JStock.Academic.Ksg/ams/websupp/index.htm.

——— (2006c): "Performance of Conditional Wald Tests in IV Regression with Weak Instruments," *Journal of Econometrics,* forthcoming.

Andrews, D. W. K. and J. H. Stock (2005): "Inference with Weak Instruments," Cowles Foundation Discussion Paper No. 1530, Yale University. Available at http://cowles.econ.yale.edu.

Cavanagh, C. and R. Sherman (1998): "Rank Estimators for Monotonic Index Models," *Journal of Econometrics,* 84, 351-382.

Chen, S. (2000): "Rank Estimation of a Location Parameter in the Binary Choice Model," *Journal of Econometrics,* 98, 317-334.

——— (2002): "Rank Estimation of Transformation Models," *Econometrica,* 70, 1683-1697.

Chernoff, H. and I. R. Savage (1958): "Asymptotic Normality and Efficiency of Certain Nonparametric Test Statistics," *Annals of Mathematical Statistics,* 39, 972-994.

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

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

——— (1997): "Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica,* 65, 1365-1387.

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

Hájek, J., Z. Sidák, and P. K. Sen (1999): *Theory of Rank Tests,* second edition. New York: Academic Press.

Hasan, M. N. and R. W. Koenker (1997): "Robust Rank Tests of the Unit Root Hypothesis," *Econometrica,* 65, 133-161.

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

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

——— (2005): "Testing Parameters in GMM without Assuming That They Are Identified," *Econometrica,* 73, 1103-1123.

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. (1969): "Asymptotic Behavior of Wilcoxon Type Confidence Regions in Multiple Regression," *Annals of Mathematical Statistics,* 40, 1950-1979.

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

Lehmann, E. L. (1986): *Testing Statistical Hypotheses,* second edition. New York: Wiley.

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

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

Pollard, D. (1990): *Empirical Processes: Theory and Applications.* Vol. 2 in the NSF-CBMS Regional Conference Series in Probability and Statistics. Bethesda, MD: Institute of Mathematical Statistics.

Puri, M. L. and P. K. Sen (1985): *Nonparametric Methods in General Linear Models.* New York: Wiley.

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

Stock, J. H., J. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics,* 20, 147-162.

Thompson, S. B. (2004): "Robust Tests of the Unit Root Hypothesis Should Not Be 'modified'," *Econometric Theory,* 20, 360-381.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes.* New York: Springer.

Table I. Asymptotic Power[†]

Case	Dist	WS-RCLR	NS-RCLR	CLR	LM	AR
Base Case	Normal	0.37	0.38	0.39	0.39	0.21
($\beta = 1.0$)	Uniform	0.38	0.52	0.40	0.41	0.25
	t_3	0.58	0.54	0.40	0.40	0.21
	DLN	0.69	0.66	0.39	0.39	0.20
Base Case	Normal	0.39	0.41	0.41	0.41	0.27
($\beta = -0.43$)	Uniform	0.37	0.50	0.41	0.41	0.26
	t_3	0.60	0.55	0.40	0.40	0.25
	DLN	0.78	0.68	0.41	0.40	0.25
High Endogeneity	Normal	0.37	0.38	0.39	0.39	0.21
($\rho_{uv_2} = 0.95,\ \beta = 1.1$)	Uniform	0.39	0.59	0.39	0.40	0.22
	t_3	0.59	0.55	0.38	0.38	0.20
	DLN	0.75	0.73	0.38	0.38	0.21
High Endogeneity	Normal	0.41	0.43	0.42	0.43	0.23
($\rho_{uv_2} = 0.95,\ \beta = -0.37$)	Uniform	0.40	0.61	0.42	0.42	0.24
	t_3	0.67	0.60	0.42	0.42	0.24
	DLN	0.87	0.78	0.42	0.42	0.23
Weaker IVs	Normal	0.39	0.40	0.41	0.41	0.22
($\lambda = 4,\ \beta = 5.0$)	Uniform	0.37	0.47	0.41	0.41	0.22
	t_3	0.35	0.41	0.40	0.40	0.22
	DLN	0.48	0.50	0.40	0.40	0.22
Weaker IVs	Normal	0.38	0.39	0.39	0.35	0.32
($\lambda = 4,\ \beta = -0.7$)	Uniform	0.34	0.42	0.38	0.34	0.32
	t_3	0.57	0.52	0.39	0.35	0.32
	DLN	0.73	0.63	0.40	0.35	0.33
Ten IVs	Normal	0.38	0.40	0.40	0.40	0.15
($k = 10,\ \beta = 1.0$)	Uniform	0.33	0.47	0.38	0.38	0.16
	t_3	0.59	0.56	0.41	0.41	0.16
	DLN	0.65	0.65	0.39	0.39	0.16
Ten IVs	Normal	0.35	0.37	0.37	0.37	0.19
($k = 10,\ \beta = -0.43$)	Uniform	0.33	0.44	0.36	0.36	0.19
	t_3	0.55	0.51	0.37	0.37	0.18
	DLN	0.70	0.61	0.36	0.36	0.18

[†]All cases have $\lambda = 10$, $\rho_{uv_2} = 0.75$, and $k = 5$, unless otherwise stated.

TABLE II. Finite Sample Null Rejection Rates of Nominal Level .05 Tests†

Case	Dist	WS-RCLR	NS-RCLR	CLR	LM	AR
Base Case	Normal	0.050	0.043	0.056	0.054	0.049
	Uniform	0.049	0.041	0.054	0.052	0.053
	t_1	0.032	0.045	0.073	0.610	0.108
	t_2	0.044	0.042	0.065	0.058	0.077
	t_3	0.046	0.039	0.060	0.056	0.058
	DLN	0.044	0.038	0.062	0.055	0.071
No Endogeneity	Normal	0.048	0.039	0.058	0.055	0.049
($\rho_{uv_2} = 0$)	Uniform	0.048	0.040	0.059	0.053	0.053
	t_1	0.030	0.039	0.077	0.058	0.108
	t_2	0.042	0.039	0.072	0.058	0.077
	t_3	0.046	0.037	0.063	0.057	0.058
	DLN	0.043	0.038	0.069	0.055	0.071
High Endogeneity	Normal	0.050	0.045	0.054	0.053	0.049
($\rho_{uv_2} = 0.95$)	Uniform	0.050	0.045	0.052	0.051	0.053
	t_1	0.033	0.047	0.064	0.056	0.108
	t_2	0.046	0.045	0.059	0.057	0.077
	t_3	0.047	0.042	0.056	0.055	0.058
	DLN	0.043	0.042	0.057	0.055	0.071
Weaker IVs	Normal	0.049	0.041	0.058	0.055	0.049
($\lambda = 4$)	Uniform	0.049	0.043	0.058	0.053	0.053
	t_1	0.031	0.045	0.078	0.058	0.108
	t_2	0.043	0.041	0.073	0.058	0.077
	t_3	0.046	0.041	0.064	0.056	0.058
	DLN	0.043	0.039	0.070	0.055	0.071
Stronger IVs	Normal	0.049	0.043	0.054	0.054	0.049
($\lambda = 20$)	Uniform	0.049	0.041	0.054	0.052	0.053
	t_1	0.032	0.045	0.068	0.056	0.108
	t_2	0.045	0.041	0.063	0.058	0.077
	t_3	0.047	0.041	0.057	0.055	0.058
	DLN	0.043	0.039	0.070	0.055	0.071

†All cases have $\beta = \beta_0 = 0$, $\lambda = 10$ (equivalently, $\rho_{IV} = 0.302$ for $n = 100$), $\rho_{uv_2} = 0.75$, $n = 100$, $k = 5$, and $p = 1$ (an intercept), unless otherwise stated.

Table II. (cont.)

Case	Dist	WS-RCLR	NS-RCLR	CLR	LM	AR
One IV	Normal	0.048	0.041	0.053	0.053	0.050
($k = 1$)	Uniform	0.052	0.045	0.055	0.055	0.053
	t_1	0.031	0.041	0.047	0.047	0.046
	t_2	0.041	0.041	0.055	0.054	0.053
	t_3	0.046	0.041	0.054	0.054	0.052
	DLN	0.045	0.047	0.057	0.057	0.054
Ten IVs	Normal	0.052	0.048	0.053	0.052	0.050
($k = 10, n = 200$)	Uniform	0.050	0.046	0.053	0.053	0.051
	t_1	0.032	0.051	0.058	0.044	0.127
	t_2	0.049	0.048	0.058	0.050	0.090
	t_3	0.049	0.045	0.054	0.052	0.062
	DLN	0.045	0.041	0.057	0.057	0.054
Smaller Sample Size	Normal	0.045	0.036	0.061	0.056	0.048
($n = 50$)	Uniform	0.050	0.039	0.065	0.058	0.050
	t_1	0.027	0.035	0.091	0.070	0.127
	t_2	0.039	0.033	0.078	0.064	0.078
	t_3	0.044	0.034	0.071	0.061	0.063
	DLN	0.044	0.037	0.076	0.064	0.074
Larger Sample Size	Normal	0.048	0.046	0.052	0.052	0.049
($n = 200$)	Uniform	0.049	0.044	0.052	0.052	0.053
	t_1	0.032	0.049	0.052	0.042	0.090
	t_2	0.044	0.045	0.058	0.051	0.076
	t_3	0.048	0.045	0.056	0.053	0.056
	DLN	0.050	0.046	0.056	0.054	0.067

TABLE III. Finite Sample Power of Size-corrected Level .05 Tests.†

Case	Dist	WS-RCLR	NS-RCLR	CLR	LM	AR
Base Case	Normal	0.42	0.46	0.40	0.40	0.26
($\beta = 1.35$)	Uniform	0.41	0.48	0.40	0.40	0.25
	t_1	0.92	0.95	0.56	0.61	0.40
	t_2	0.66	0.66	0.46	0.49	0.26
	t_3	0.53	0.53	0.42	0.43	0.26
	DLN	0.61	0.60	0.43	0.45	0.23
Base Case	Normal	0.35	0.35	0.39	0.38	0.25
($\beta = -0.44$)	Uniform	0.32	0.38	0.39	0.39	0.25
	t_1	0.94	0.95	0.55	0.61	0.40
	t_2	0.72	0.68	0.46	0.49	0.27
	t_3	0.53	0.50	0.41	0.43	0.25
	DLN	0.65	0.59	0.42	0.44	0.22
Base Case	Normal	0.38	0.40	0.39	0.39	0.26
($\beta = 1.35$ & $\beta = -0.44$)	Uniform	0.37	0.43	0.40	0.40	0.25
	t_1	0.93	0.95	0.55	0.61	0.40
	t_2	0.69	0.67	0.46	0.49	0.27
	t_3	0.53	0.51	0.42	0.43	0.26
	DLN	0.63	0.60	0.42	0.44	0.23
No Endogeneity	Normal	0.44	0.47	0.41	0.37	0.34
($\rho_{uv_2} = 0$,	Uniform	0.43	0.47	0.42	0.37	0.34
$\beta = 0.975$ & $\beta = -1.05$)	t_1	0.94	0.97	0.55	0.61	0.42
	t_2	0.74	0.72	0.46	0.47	0.33
	t_3	0.57	0.56	0.43	0.40	0.34
	DLN	0.67	0.64	0.41	0.42	0.30
High Endogeneity	Normal	0.38	0.39	0.41	0.41	0.22
($\rho_{uv_2} = 0.95$,	Uniform	0.39	0.46	0.41	0.41	0.22
$\beta = 0.95$ & $\beta = -1.25$)	t_1	0.93	0.96	0.61	0.64	0.41
	t_2	0.70	0.66	0.49	0.50	0.25
	t_3	0.53	0.51	0.42	0.42	0.22
	DLN	0.67	0.62	0.44	0.44	0.19

†All cases have $\beta_0 = 0$, $\lambda = 10$ (equivalently, $\rho_{IV} = 0.302$ for $n = 100$), $\rho_{uv_2} = 0.75$, $n = 100$, $k = 5$, and $p = 1$ (an intercept), unless otherwise stated.

TABLE III. (cont.)

Case	Dist	WS-RCLR	NS-RCLR	CLR	LM	AR
Weaker IVs	Normal	0.33	0.33	0.36	0.31	0.30
($\lambda = 4$,	Uniform	0.31	0.33	0.36	0.32	0.29
$\beta = 25$ & $\beta = -0.725$)	t_1	0.91	0.94	0.52	0.59	0.40
	t_2	0.61	0.58	0.41	0.43	0.28
	t_3	0.43	0.42	0.37	0.35	0.29
	DLN	0.55	0.49	0.36	0.37	0.25
Stronger IVs	Normal	0.40	0.42	0.40	0.40	0.23
($\lambda = 20$,	Uniform	0.38	0.47	0.41	0.41	0.23
$\beta = 0.62$ & $\beta = -0.325$)	t_1	0.94	0.96	0.58	0.63	0.40
	t_2	0.75	0.72	0.48	0.50	0.25
	t_3	0.57	0.54	0.42	0.43	0.23
	DLN	0.59	0.54	0.38	0.40	0.22
One IV	Normal	0.37	0.38	0.39	0.39	0.39
($k = 1$,	Uniform	0.34	0.42	0.39	0.39	0.39
$\beta = 1.05$ & $\beta = -0.41$)	t_1	0.88	0.89	0.44	0.44	0.44
	t_2	0.67	0.64	0.42	0.42	0.42
	t_3	0.54	0.51	0.42	0.42	0.42
	DLN	0.66	0.61	0.40	0.40	0.40
Ten IVs	Normal	0.38	0.39	0.41	0.41	0.24
($k = 10$,	Uniform	0.38	0.43	0.45	0.45	0.27
$\beta = 1.9$ & $\beta = -0.49$)	t_1	0.94	0.90	0.60	0.67	0.43
	t_2	0.72	0.69	0.50	0.54	0.28
	t_3	0.54	0.52	0.46	0.46	0.28
	DLN	0.64	0.60	0.46	0.48	0.24