A Note on Optimal Inference in the Linear IV Model

Donald W.K. Andrews
Vadim Marmer
Zhengfei Yu

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

Part of the Economics Commons

Recommended Citation
Andrews, Donald W.K.; Marmer, Vadim; and Yu, Zhengfei, "A Note on Optimal Inference in the Linear IV Model" (2017). Cowles Foundation Discussion Papers. 2538.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/2538

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
A NOTE ON OPTIMAL INFERENCE IN THE LINEAR IV MODEL

By

Donald W. K. Andrews, Vadim Marmer, and Zhengfei Yu

January 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 2073

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

http://cowles.yale.edu/
A Note on
Optimal Inference
in the Linear IV Model

Donald W. K. Andrews  Vadim Marmer
Cowles Foundation  Vancouver School of Economics
Yale University  University of British Columbia

Zhengfei Yu
Faculty of Humanities and Social Sciences
University of Tsukuba

October 23, 2015
Revised: January 30, 2017*

*Andrews gratefully acknowledges the research support of the National Science Foundation via grant number SES-1355504. Marmer gratefully acknowledges the research support of the SSHRC via grant numbers 410-2010-1394 and 435-2013-0331.
Abstract

This paper considers tests and confidence sets (CS’s) concerning the coefficient on the endogenous variable in the linear IV regression model with homoskedastic normal errors and one right-hand side endogenous variable. The paper derives a finite-sample lower bound function for the probability that a CS constructed using a two-sided invariant similar test has infinite length and shows numerically that the conditional likelihood ratio (CLR) CS of Moreira (2003) is not always “very close” to this lower bound function. This implies that the CLR test is not always very close to the two-sided asymptotically-efficient (AE) power envelope for invariant similar tests of Andrews, Moreira, and Stock (2006) (AMS).

On the other hand, the paper establishes the finite-sample optimality of the CLR test when the correlation between the structural and reduced-form errors, or between the two reduced-form errors, goes to 1 or -1 and other parameters are held constant, where optimality means achievement of the two-sided AE power envelope of AMS. These results cover the full range of (non-zero) IV strength.

The paper investigates in detail scenarios in which the CLR test is not on the two-sided AE power envelope of AMS. Also, the paper shows via theory and numerical work that the CLR test is close to having greatest average power, where the average is over a grid of concentration parameter values and over pairs alternative hypothesis values of the parameter of interest, uniformly over pairs of alternative hypothesis values and uniformly over the correlation between the structural and reduced-form errors.

The paper concludes that, although the CLR test is not always very close to the two-sided AE power envelope of AMS, CLR tests and CS’s have very good overall properties.

Keywords: Conditional likelihood ratio test, confidence interval, infinite length, linear instrumental variables, optimal test, weighted average power, similar test.

JEL Classification Numbers: C12, C36.
1 Introduction

The linear instrumental variables (IV) regression model is one of the most widely used models in economics. It has been widely studied and considerable effort has been made to develop good estimation and inference methods for it. In particular, following the recognition that standard two stage least squares $t$ tests and confidence sets (CS’s) can perform quite poorly under weak IV’s (see Dufour (1997), Staiger and Stock (1997), and references therein), inference procedures that are robust to weak IV’s have been developed, e.g., see Kleibergen (2002) and Moreira (2003, 2009). The focus has been on models with one right-hand side (rhs) endogenous variable, because this arises most frequently in applications, and on over-identified models, because Anderson and Rubin (1949) (AR) tests and CS’s are robust to weak IV’s and perform very well in exactly-identified models.

Andrews, Moreira, and Stock (2006) (AMS) develop a finite-sample two-sided AE power envelope for invariant similar tests concerning the coefficient on the rhs endogenous variable in the linear IV model under homoskedastic normal errors and known reduced-form variance matrix. They show via numerical simulations that the conditional likelihood ratio (CLR) test of Moreira (2003) has power that is essentially (i.e., up to simulation error) on the power envelope. Chernozhukov, Hansen, and Jansson (2009) (CHJ) show that this power envelope also applies to non-invariant tests provided the envelope is for power averaged over certain direction vectors in a unit sphere. CHJ also shows that the invariant similar tests that generate the two-sided AE power envelope are $\alpha$-admissible and $d$-admissible. Mikusheva (2010) provides approximate optimality results for CLR-based CS’s that utilize the testing results in AMS. Chamberlain (2007), Andrews, Moreira, and Stock (2008), and Hillier (2009) provide related results.

It is shown in Dufour (1997) that any CS with correct size $1 - \alpha$ must have positive probability of having infinite length at every point in the parameter space. The AR and CLR CS’s have this property. In fact, simulation results show that in some over-identified contexts the AR CS has a lower probability of having an infinite length than the CLR CS does. For example, consider a model with one rhs endogenous variable, $k$ IV’s, a concentration parameter $\lambda_v$ (which is a measure of the strength of the IV’s), homoskedastic normal errors, a correlation $\rho_{uv}$ between the structural-equation error and the reduced-form error (for the first-stage equation) equal to zero, and no covariates. When $(k, \lambda_v)$ equals $(2, 7)$, $(5, 12)$, $(10, 15)$, $(20, 20)$, and $(40, 20)$, the differences between the probabilities that the 95% CLR and AR CS’s have infinite length are .016, .029, .037, .044, and .038, respectively. In fact, one obtains positive differences for all combinations of $(k, \lambda_v)$ for $k = 2, 5, 10, 20, 40$ and $\lambda_v = 1, 5, 10, 15, 20$. Hence, in these over-identified scenarios the AR CS

\footnote{See Table SM-I in the Supplemental Material.}
outperforms the CLR CS in terms of its infinite-length behavior, which is an important property for CS’s. Similarly, one obtains positive (but smaller) differences also when $\rho_{uv} = .3$ for the same range of $(k, \lambda_v)$ values. On the other hand, for $\rho_{uv} = .5, .7, \text{and} .9$, the differences are negative over the same range of $(k, \lambda_v)$ values.

The AR and CLR CS’s are based on inverting AR and CLR tests that fall into the class of invariant similar tests considered in AMS. Hence, the simulation results for $\rho_{uv} = .0$ and .3 raise the question: how can these results be reconciled with the near optimal CLR test and CS results described above? In this paper, we answer this question and related questions concerning the optimality of the CLR test and CS.

The contributions of the paper are as follows. First, the paper shows that the probability that an invariant similar CS has infinite length for a fixed true parameter value $\beta_*$ equals one minus the power against $\beta_*$ of the test used to construct the CS as the null value $\beta_0$ goes to $\infty$ or $-\infty$. This leads to explicit formulae for the probabilities that the AR and CLR CS’s have infinite length.

Second, the paper uses the first result to determine a finite-sample lower bound function on the probabilities that a CS has infinite length for CS’s based on invariant similar tests. The lower bound function is found to be very simple. It is a function only of $|\rho_{uv}|, \lambda_v, \text{and} k$. These results allow one to compare the probabilities that the AR and CLR CS’s have infinite length with the lower bound.

Third, simulation results show that the AR and CLR CS’s are not always close to the lower bound. This is not surprising for the AR CS, but it is surprising for the CLR CS in light of the AMS results. The probabilities that the CLR CS has infinite length are found to be off the lower bound function by a magnitude that is decreasing in $|\rho_{uv}|$, increasing in $k$, and are maximized over $\lambda_v$ at values that correspond to somewhat weak IV’s, but not irrelevant IV’s. For $\rho_{uv} = 0$, the paper shows (analytically) that the AR test achieves the lower bound function. Hence, for $\rho_{uv} = 0$, the probabilities that the CLR CS has infinite length exceed the lower bound by the same amounts as reported above for the difference between the infinite length probabilities of the CLR and AR CS’s for several $(k, \lambda_v)$ values. On the other hand, for values of $|\rho_{uv}| \geq .7$, the CLR CS has probabilities of having infinite length that are close to the lower bound function, .010 or less and typically much less, for all $(k, \lambda_v)$ combinations considered. For values of $|\rho_{uv}| \geq .7$, the AR CS has probabilities of having infinite length that are often far from the lower bound. For $|\rho_{uv}| = .9$ and certain values of $\lambda_v$, they are as large as .084, .196, .280, .353, and .422 for $k = 2, 5, 10, 20, \text{and} 40$, respectively.

Fourth, the paper derives new optimality properties of the CLR test when $\rho_{uv} \to \pm 1$ or $\rho_\Omega \to \pm 1$ with other parameters fixed at any values (with non-zero concentration parameter), where $\rho_\Omega$.

---

2 See Table SM-I in the SM.
denotes the correlation between the reduced-form errors. Optimality here is in the class of invariant similar tests or similar tests and employs the two-sided AE power envelope of AMS. These results are consistent with the numerical results that show that the CLR test is very close to the power envelope when $|\rho_{uv}|$ is large.

Fifth, we simulate power differences (PD’s) between the two-sided AE power envelope of AMS and the power of the CLR test for a fixed alternative value $\beta_*$ and a range of finite null values $\beta_0$ (rather than the PD’s as $\beta_0 \rightarrow \pm\infty$ discussed above). These PD’s are equivalent to the false coverage probability differences between the CLR CS and the corresponding infeasible optimal CS for a fixed true value $\beta_*$ at incorrect values $\beta_0$. We consider a wide range of $(\beta_0, \lambda_v, \rho_{uv}, k)$ values. The maximum (over $\beta_0$ and $\lambda_v$ values) PD’s range between $.016,.061$ over the $(\rho_{uv}, k)$ values considered. On the other hand, the average (over $\beta_0$ and $\lambda$ values) PD’s only range between $.002,.016$. This indicates that, although there are some $(\beta_0, \lambda)$ values at which the CLR test is noticeably off the power envelope, on average the CLR test’s power is not far from the power envelope. The maximum PD’s over $(\beta_0, \lambda)$ are found to increase in $k$ and decrease in $|\rho_{uv}|$. The $\lambda_v$ values at which the maxima are obtained are found to (weakly) increase with $k$ and decrease in $|\rho_{uv}|$. The $|\beta_0|$ values at which the maxima are obtained are found to be independent of $k$ and decrease in $|\rho_{uv}|$.

Sixth, the paper considers a weighted average power (WAP) envelope with a uniform weight function over a grid of concentration parameter values $\lambda_v$ and the same two-point AE weight function over $(\beta, \lambda)$ as in AMS. We refer to this as the WAP2 envelope. We determine numerically how close the power of the CLR test is to the WAP2 envelope. We find that the difference between the WAP2 envelope and the average power of the CLR test is in the range of $.001,.007$ over all of the $(\beta_0, \beta_*, \rho_{uv}, k)$ values that we consider. Hence, the average power of the CLR test is quite close to the WAP2 envelope.

Other papers in the literature that consider WAP include Wald (1943), Andrews and Ploberger (1994), Andrews (1998), Moreira and Moreira (2013), Elliott, Müller, and Watson (2015), and papers referenced above. The WAP2 envelope considered here is closest to the WAP envelopes in Wald (1943), AMS, and CHJ because the other papers listed put a weight function over all of the parameters in the alternative hypothesis, which yields a single weighted alternative density. In contrast, the WAP2 envelope, Wald (1943), AMS, and CHJ consider a family of weight functions over disjoint sets of parameters in the alternative hypothesis, which yields a WAP envelope.

In conclusion, based on our findings, we recommend use of the CLR test and CS. More specifically, we recommend using heteroskedasticity-robust versions of these procedures that have the same asymptotic properties as these procedures under homoskedasticity. For example, such tests
are given in Andrews, Moreira, and Stock (2004) and Andrews and Guggenberger (2015). The CLR CS has higher probability of having infinite length than the AR CS in some scenarios, and the CLR test is not a UMP two-sided invariant similar test. But, no such UMP test exists and the CLR CS is close to the two-sided AE power envelope for invariant similar tests when $|\rho_{uv}|$ is not close to zero and is close to the WAP2 envelope for all values of $|\rho_{uv}|$.

Finally, we point out that the results of this paper illustrate a point that applies more generally than in the linear IV model. In weak identification scenarios, where CS’s may have infinite length (or may be bounded only due to bounds on the parameter space), good test performance at a priori implausible parameter values is important for good CS performance at plausible parameter values. More specifically, the probability under an a priori plausible parameter value $\beta_*$ that a CS has infinite length depends on the power of the test used to construct the CS against $\beta_*$ when the null value $|\beta_0|$ is arbitrarily large, which may be an a priori implausible null value.

For the computation of CLR CS’s, see Mikusheva (2010). For a formula for the power of the CLR test, see Hillier (2009).

The paper is organized as follows. Section 2 specifies the model. Section 3 defines the class of invariant similar tests. Section 4 provides a formula for the probability that a CS has infinite length. Section 5 derives a lower bound on the probability that a CS constructed using two-sided invariant similar tests has infinite length. Section 6 reports differences between the probability that the CLR CS has infinite length and the lower bound derived in the previous section. Section 7 proves the optimality results for the CLR test described above. Section 8 reports differences between the power of CLR tests and the two-sided AE power bound of AMS for a wide range of parameter configurations. Section 9 provides comparisons of the power of the CLR test to the WAP2 power envelope described above. Proofs and additional theoretical and numerical results are given in the Supplemental Material (SM).

2 Model

We consider the same model as in Andrews, Moreira, and Stock (2004, 2006) (AMS04, AMS) but, for simplicity and without loss of generality (wlog), without any exogenous variables. The model has one rhs endogenous variable, $k$ instrumental variables (IV’s), and normal errors with known reduced-form error variance matrix. The model consists of a structural equation and a reduced-form equation:

$$y_1 = y_2 \beta + u \quad \text{and} \quad y_2 = Z\pi + v_2, \quad (2.1)$$
where \( y_1, y_2 \in \mathbb{R}^n \) and \( Z \in \mathbb{R}^{n \times k} \) are observed variables; \( u, v_2 \in \mathbb{R}^n \) are unobserved errors; and \( \beta \in \mathbb{R} \) and \( \pi \in \mathbb{R}^k \) are unknown parameters. The IV matrix \( Z \) is fixed (i.e., non-stochastic) and has full column rank \( k \). The \( n \times 2 \) matrix of errors \([u;v_2]\) is i.i.d. across rows with each row having a mean zero bivariate normal distribution.

The two corresponding reduced-form equations are

\[
Y := [y_1 : y_2] := [Z\pi \beta + v_1 : Z\pi + v_2] = Z\pi a' + V, \quad \text{where} \\
V := [v_1 : v_2] = [u + v_2 \beta : v_2], \quad \text{and} \quad a := (\beta, 1)'.
\]

The distribution of \( Y \in \mathbb{R}^{n \times 2} \) is multivariate normal with mean matrix \( Z\pi a' \), independence across rows, and reduced-form variance matrix \( \Omega \in \mathbb{R}^{2 \times 2} \) for each row. For the purposes of obtaining exact finite-sample results, we suppose \( \Omega \) is known. As in AMS, asymptotic results for unknown \( \Omega \) and weak IV’s are the same as the exact results with known \( \Omega \). The parameter space for \( \theta = (\beta, \pi')' \) is \( \mathbb{R}^{k+1} \).

We are interested in tests of the null hypothesis \( H_0 : \beta = \beta_0 \) and CS’s for \( \beta \).

As shown in AMS, \( Z'Y \) is a sufficient statistic for \( (\beta, \pi')' \). As in Moreira (2003) and AMS, we consider a one-to-one transformation \([S : T]\) of \( Z'Y \):

\[
S := (Z'Z)^{-1/2}Z'Y b_0 \cdot (b_0'\Omega b_0)^{-1/2} \sim N(c_\beta(\beta_0, \Omega) \cdot \mu_{\pi}, I_k) \quad \text{and} \\
T := (Z'Z)^{-1/2}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \sim N(d_\beta(\beta_0, \Omega) \cdot \mu_{\pi}, I_k), \quad \text{where} \\
b_0 := (1, -\beta_0)', \quad a_0 := (\beta_0, 1)', \quad \mu_{\pi} := (Z'Z)^{1/2}\pi \in \mathbb{R}^k, \\
c_\beta(\beta_0, \Omega) := (\beta - \beta_0) \cdot (b_0'\Omega b_0)^{-1/2} \in \mathbb{R}, \\
d_\beta(\beta_0, \Omega) := b'\Omega b_0 \cdot (b_0'\Omega b_0)^{-1/2} \det(\Omega)^{-1/2} \in \mathbb{R}, \quad \text{and} \quad b = (1, -\beta)').
\]

As defined, \( S \) and \( T \) are independent. Note that \( S \) and \( T \) depend on the null hypothesis value \( \beta_0 \).

### 3 Invariant Similar Tests

As in Hillier (1984) and AMS, we consider tests that are invariant to orthonormal transformations of \([S : T]\), i.e., \([S : T] \rightarrow [FS : FT]\) for a \( k \times k \) orthogonal matrix \( F \). The \( 2 \times 2 \) matrix \( Q \) is a maximal invariant, where

\[
Q = [S:T]'[S:T] = \begin{bmatrix}
S'S & S'T \\
S'T & T'T
\end{bmatrix} = \begin{bmatrix}
Q_S & Q_ST \\
Q_ST & Q_T
\end{bmatrix} \quad \text{and} \quad Q_1 = \begin{pmatrix}
S'S \\
S'T
\end{pmatrix} = \begin{pmatrix}
Q_S \\
Q_ST
\end{pmatrix},
\]


5
e.g., see Theorem 1 of AMS. Note that $Q_1$ is the first column of $Q$ and the matrix $Q$ depends on the null value $\beta_0$.

The statistic $Q$ has a non-central Wishart distribution because $[S:T]$ is a multivariate normal matrix that has independent rows and common covariance matrix across rows. The distribution of $Q$ depends on $\pi$ only through the scalar

$$\lambda := \pi'Z'Z\pi \geq 0.$$  \hfill (3.2)

Leading examples of invariant identification-robust tests in the literature include the AR test, the LM test of Kleibergen (2002) and Moreira (2009), and the CLR test of Moreira (2003). The latter test depends on the standard LR test statistic coupled with a “conditional” critical value that depends on $Q_T$. The LR, LM, and AR test statistics are

$$LR := \frac{1}{2} \left( QS - QT + \sqrt{(QS - QT)^2 + 4Q^2_{ST}} \right),$$

$$LM := Q^2_{ST}/QT = (S'T)^2/T'T, \text{ and } AR := QS/k = S'S/k.$$  \hfill (3.3)

The critical values for the $LM$ and $AR$ tests are $\chi^2_{1,1-\alpha}$ and $\chi^2_{k,1-\alpha}/k$, respectively, where $\chi^2_{m,1-\alpha}$ denotes the $1 - \alpha$ quantile of the $\chi^2$ distribution with $m$ degrees of freedom.

A test based on the maximal invariant $Q$ is similar if its null rejection rate does not depend on the parameter $\pi$ that determines the strength of the IV’s $Z$. As in Moreira (2003), the class of invariant similar tests is specified as follows. Let the $[0,1]$-valued statistic $\phi(Q)$ denote a (possibly randomized) test that depends on the maximal invariant $Q$. An invariant test $\phi(Q)$ is similar with significance level $\alpha$ if and only if $E_{\beta_0}(\phi(Q)|Q_T = q_T) = \alpha$ for almost all $q_T > 0$ (with respect to Lebesgue measure), where $E_{\beta_0}(\cdot|Q_T = q_T)$ denotes conditional expectation given $Q_T = q_T$ when $\beta = \beta_0$ (which does not depend on $\pi$).

The CLR test rejects the null hypothesis when

$$LR > \kappa_{LR,\alpha}(Q_T),$$  \hfill (3.4)

where $\kappa_{LR,\alpha}(Q_T)$ is defined to satisfy $P_{\beta_0}(LR > \kappa_{LR,\alpha}(Q_T)|Q_T = q_T) = \alpha$ and the conditional distribution of $Q_1$ given $Q_T$ is specified in AMS and in (11.3) in the SM.

The invariance condition discussed above is a rotational invariance condition. In some cases, we also consider a sign invariance condition. A test that depends on $[S:T]$ is sign invariant if it is invariant to the transformation $[S:T] \rightarrow [-S:T]$. A rotation invariant test is also sign invariant if it depends on $Q_{ST}$ only through $|Q_{ST}|$. Tests that are sign invariant are two-sided tests. In fact,
AMS shows that the two-sided AE power envelope is identical to the power envelope generated by
sign and rotation invariant tests, see (4.11) in AMS.

For simplicity, we will use the term invariant test to mean a rotation invariant test and the term
sign and rotation invariant test to describe a test that satisfies both invariance conditions.

The paper also provides some results that apply to tests that satisfy no invariance properties.
A test \(\phi([S : T])\) (that is not necessarily invariant) is similar with significance level \(\alpha\) if and only if
\[ E_{\beta_0}(\phi([S : T])|T = t) = \alpha \quad \text{for almost all } t \quad (\text{with respect to Lebesgue measure}), \]
denotes conditional expectation given \(T = t\) when \(\beta = \beta_0\) (which does not depend on \(\pi\)), see
Moreira (2009).

4 Probability That a Confidence Set Has Infinite Length

In this section, we show that the probability that a confidence set (CS) has infinite length is
given by one minus the power of the test used to construct the CS as the null value \(\beta_0\) of the test
goes to \(\infty\) or \(-\infty\). This provides motivation for interest in the power of tests as \(\beta_0 \to \pm \infty\). It shows
why high power against distant null hypotheses is highly desirable.

We sometimes make the dependence of \(Q, S, T\) on \(Y\) and \(\beta_0\) explicit and write
\[ Q = Q_{\beta_0}(Y) = [S_{\beta_0}(Y) : T_{\beta_0}(Y)]' [S_{\beta_0}(Y) : T_{\beta_0}(Y)]. \quad (4.1) \]

We denote the \((1, 1), (1, 2),\) and \((2, 2)\) elements of \(Q_{\beta_0}(Y)\) by \(Q_{S, \beta_0}(Y), Q_{ST, \beta_0}(Y),\) and \(Q_{T, \beta_0}(Y),\)
respectively.

Let \(\phi(Q_{\beta_0}(Y)) = 1(T(Q_{\beta_0}(Y)) > cv(Q_{T, \beta_0}(Y)))\)
be a (nonrandomized) invariant similar level \(\alpha\) test for testing \(H_0 : \beta = \beta_0\) for fixed known \(\Omega,\)
where \(T(Q_{\beta_0}(Y))\) is a test statistic and \(cv(Q_{T, \beta_0}(Y))\) is a (possibly data-dependent) critical value.
Examples include the AR, LM, and CLR tests in (3.3). Let \(CS_{\phi}\) be the level \(1 - \alpha\) CS corresponding
to \(\phi\). That is,
\[ CS_{\phi}(Y) = \{\beta_0 : \phi(Q_{\beta_0}(Y)) = 0\}. \quad (4.3) \]

We say \(CS_{\phi}(Y)\) has right (or left) infinite length, which we denote by \(RLength(CS_{\phi}(Y)) = \infty\)
(or \(LLength(CS_{\phi}(Y)) = \infty\)), if
\[ \exists K(Y) < \infty \quad \text{such that } \beta \in CS_{\phi}(Y) \quad \forall \beta \geq K(Y) \quad (\text{or } \forall \beta \leq -K(Y)). \quad (4.4) \]
We say \( CS_\phi(Y) \) has infinite length, which we denote by \( \text{Length}(CS_\phi(Y)) = \infty \), if it has right and left infinite lengths. A CS with infinite length contains a set of the form \( (-\infty, K_1(Y)) \cup (K_2(Y), \infty) \) for some \( -\infty < K_1(Y) \leq K_2(Y) < \infty \).

Let \( P_{\beta_*,\pi,\Omega}() \) denote probability for events determined by \( Y \) when \( Y \) has a multivariate normal distribution with means matrix \([\pi\beta_* : \pi] \in R^2\), independence across rows, and variance matrix \( \Omega \) for each row. Let \( P_{\beta_*,\beta_0,\lambda,\Omega}() \) denote probability for events determined by \( Q \) when \( Q := [S : T]'[S : T] \) and \( [S : T] \) has the multivariate normal distribution in \([2.3]\) with \( \beta = \beta_* \) and \( \lambda = \mu_\pi^*\mu_\pi \). In this case, \( Q \) has a noncentral Wishart distribution whose density is given in \([11.2]\) in the SM.

For fixed true value \( \beta_* \) and reduced-form variance matrix \( \Omega \), let \( \Sigma_* \) denote the corresponding structural variance matrix of each row of \([u : v_2]\). Let \( \rho_{uv} \) denote the correlation between the structural and reduced-form errors, i.e., the correlation corresponding to \( \Sigma_* \). Some calculations show that

\[
\rho_{uv} = \frac{\omega_{12} - \omega_2^2 \beta_*}{(\omega_1^2 - 2\omega_{12}\beta_* + \omega_2^2\beta_*^2)^{1/2}\omega_2} \quad \text{and} \quad \Sigma_* = \begin{bmatrix} \sigma_u^2 & \sigma_u\sigma_v\rho_{uv} \\ \sigma_u\sigma_v\rho_{uv} & \sigma_v^2 \end{bmatrix} = \begin{bmatrix} \omega_1^2 - 2\omega_{12}\beta_* + \omega_2^2\beta_*^2 & \omega_{12} - \omega_2^2\beta_* \\ \omega_{12} - \omega_2^2\beta_* & \omega_2^2 \end{bmatrix},
\]

where \( \omega_1^2, \omega_2^2, \) and \( \omega_{12} \) denote the \((1,1), (2,2), \) and \((1,2)\) elements of \( \Omega \), respectively, see \((11.9)\) in the SM.

It is shown in Lemma \([15.1]\) in the SM that the limit as \( \beta_0 \to \pm\infty \) of \( Q_{\beta_0}(Y) \) is

\[
Q_{\pm\infty}(Y) := \begin{bmatrix} e_2'Y'P_ZY e_2 \cdot \frac{\sigma_2^2}{\sigma_v^2} & e_2'Y'P_ZY\Omega^{-1}e_1 \cdot \frac{\xi_{1,2}}{\sigma_v} \\ e_2'Y'P_ZY\Omega^{-1}e_1 \cdot \frac{\xi_{1,2}}{\sigma_v} & e_1'\Omega^{-1}Y'P_ZY\Omega^{-1}e_1 \cdot \frac{1 - \rho_{uv}^2}{\sigma_u^2} \end{bmatrix},
\]

where \( P_Z := Z(Z'Z)^{-1}Z' \), \( e_1 := (1,0)' \), and \( e_2 := (0,1)' \). Let \( Q_{T,\pm\infty}(Y) \) denote the \((2,2)\) element of \( Q_{\pm\infty}(Y) \). It is also shown in Lemma \([15.1]\) in the SM that \( Q_{\pm\infty}(Y) \) has a noncentral Wishart distribution with means matrix \( \xi_\pi(1/\sigma_v, \rho_{uv}/(\sigma_v(1 - \rho_{uv}^2)^{1/2})) \in R^{k \times 2} \) and identity variance matrix.

**Theorem 4.1** Suppose \( CS_\phi(Y) \) is a CS based on invariant level \( \alpha \) tests \( \phi(Q_{\beta_0}(Y)) \) whose test statistic and critical value functions, \( T(q) \) and \( cv(q_T) \), respectively, are continuous at all positive definite \( 2 \times 2 \) matrices \( q \) and positive constants \( q_T \). \( P_{\beta_*,\pi,\Omega}(T(Q_{\phi}(Y)) = cv(Q_{T,\phi}(Y))) = 0 \) for \( c = +\infty \) in parts (a) and (c) below and \( c = -\infty \) in part (b) below. Then, for all \( (\beta_*, \lambda, \Omega) \),

(a) \( P_{\beta_*,\pi,\Omega}(\text{Length}(CS_\phi(Y)) = \infty) = 1 - \lim_{\beta_0 \to \infty} P_{\beta_*,\beta_0,\lambda,\Omega}(\phi(Q) = 1), \)

\[\text{The density of this distribution is given in \([11.4]\) in the SM.}\]
(b) \( P_{\beta_0,\pi,\Omega}(\text{LLength}(CS_\phi(Y)) = \infty) = 1 - \lim_{\beta_0 \to -\infty} P_{\beta_0,\lambda,\Omega}(\phi(Q) = 1), \) and

c) if the tests are sign invariant, i.e., \( T(Q) \) depends on \( Q_{ST} \) only through \(|Q_{ST}|\), then
\( P_{\beta_0,\pi,\Omega}(\text{Length}(CS_\phi(Y)) = \infty) = 1 - \lim_{\beta_0 \to -\infty} P_{\beta_0,\lambda,\Omega}(\phi(Q) = 1). \)

Comments. (i). For the AR, LM, and LR tests, the continuity conditions on \( T(q) \) and \( cv(q_T) \) hold given their simple functional forms in \( (3.3) \) using the assumption that \( q_T > 0 \) for the LM statistic and using the continuity of \( \kappa_{LR,\alpha}(q_T) \), which holds by the argument in the proof of Thm. 10.1 in Andrews and Guggenberger (2016). We have with respect to Lebesgue measure. For the CLR test, the AR and LM tests because that they are equal).

provides formulae for the quantities
sign invariant. Hence, parts (a)-(c) of Theorem 4.1 apply to these tests. Theorem 5.4(a)-(c) below yields a lower bound on the probability that a CS
invariant similar level
yields a lower bound on the probability that a CS
invariant similar level
is a constant and \( T(Q_{T,\pm \infty}(Y)) \) is absolutely continuous with respect to Lebesgue measure. For the CLR test, \( P_{\beta_0,\pi,\Omega}(T(Q_{T,\pm \infty}(Y)) = cv(Q_{T,\pm \infty}(Y))) = 0 \) by the argument given in the proof of Theorem 5.4 in the SM. The AR, LM, and CLR test statistics are sign invariant. Hence, parts (a)-(c) of Theorem 4.1 apply to these tests. Theorem 5.4(a)-(c) below provides formulae for the quantities \( \lim_{\beta_0 \to -\infty} P_{\beta_0,\lambda,\Omega}(\phi(Q) = 1) \), which appear in Theorem 4.1 for the AR, LM, and CLR tests.

(ii). Comment (iii) to Theorem 5.2 below provides a lower bound on \( 1 - \lim_{\beta_0 \to -\infty} P_{\beta_0,\lambda,\Omega}(\phi(Q) = 1) \) over all sign and rotation invariant similar level \( \alpha \) tests. Combining this with Theorem 4.1(c) yields a lower bound on the probability that a CS \( CS_\phi(Y) \) based on such tests has \( \text{Length} = \infty \).

The lower bound on the probability that \( \text{Length} = \infty \) is greater than the lower bound on the probability that \( \text{RLength} = \infty \) (or that \( \text{LLength} = \infty \) unless \( \rho_{uv} = 0 \) (in which case it turns out that they are equal).

Theorem 12.1 in the SM provides lower bounds on \( 1 - \lim_{\beta_0 \to -\infty} P_{\beta_0,\lambda,\Omega}(\phi(Q) = 1) \) over all invariant similar level \( \alpha \) tests. Combining these with Theorem 4.1(a) and (b) yields lower bounds on the probabilities that a CS \( CS_\phi(Y) \) has \( \text{RLength} = \infty \) based on \( \beta_0 \to \infty \) and \( \text{LLength} = \infty \) based on \( \beta_0 \to -\infty \).

(iii). The results of Theorem 4.1(a) and (b) also hold for a \( CS_\phi(Y) \) that is based on level \( \alpha \) tests that are not invariant. Denote such tests by \( \phi(S_{\beta_0}(Y), T_{\beta_0}(Y)) \) and suppose their test statistic and critical value functions, \( T(s, t) \) and \( cv(t) \), respectively, are continuous at all \( k \times 2 \) matrices \( [s : t] \) and \( k \) vectors \( t \) and satisfy \( P_{\beta_0,\pi,\Omega}(T(S(Y), T(Y)) = cv(T(Y))) = 0 \) for \( c = +\infty \), where \( S_{\pm \infty}(Y) := (Z'Z)^{-1/2}Z'Ye_2/\sigma_v \) and \( T_{\pm \infty}(Y) := \pm(Z'Z)^{-1/2}Z'Y\Omega^{-1}e_1 \cdot (1 - \rho_{uv}^2)^{1/2}\sigma_u. \) In this case, \( P_{\beta_0,\pi,\Omega}(\text{RLength}(CS_\phi(Y)) = \infty) = 1 - \lim_{\beta_0 \to -\infty} P_{\beta_0,\pi,\Omega}(\phi(\{S : T\}) = 1) \) and likewise with \( \text{LLength}(\cdot) \), \( \beta_0 \to -\infty \), and \( c = -\infty \) in place of \( \text{RLength}(\cdot) \), \( \beta_0 \to \infty \), and \( c = +\infty \).

(iv). By Dufour (1997), all CS’s for \( \beta \) with correct size must have positive probability of having infinite length (assuming \( \pi \) is not bounded away from 0). In consequence, expected CS length, which is a standard measure of the performance of a CS, is infinite for all identifiability-robust CS’s. Due to this, Mikusheva (2010) compares CS’s based on their expected truncated
lengths for various truncation values. The result of Theorem 5.2 below implies that, for two CS’s where the rhs of Theorem 5.2(c) is smaller for the first CS than the second, the first CS has smaller expected truncated length than the second for sufficiently large truncation values.

5 Power Bound as $\beta_0 \to \pm \infty$

In this section, we provide two-sided AE power bounds for invariant similar tests as $\beta_0 \to \pm \infty$ for fixed $\beta_\pi$. The power bounds also apply to the larger class of similar tests for which invariance is not imposed, provided power is averaged over $\mu_\pi / \| \mu_\pi \|$ vectors using the uniform distribution on the unit sphere in $\mathbb{R}^k$, as in CHJ.

Using Theorem 4.1 these results are used to obtain bounds on the probabilities that CS’s constructed using sign and rotation invariant similar tests have infinite length. They also are used to obtain bounds on certain average probabilities that similar invariant tests and similar tests have infinite right (or left) length.

This section also determines the power of the AR, LM, and CLR tests as $\beta_0 \to \pm \infty$ and the probabilities that AR, LM, and CLR CS’s have infinite length.

5.1 Density of $Q$ as $\beta_0 \to \pm \infty$

The density of $Q := [S : T]'[S : T]$ when $[S : T]$ has the multivariate normal distribution in (2.3) only depends on $\pi$ through $\lambda := \mu_\pi^T \mu_\pi$. Let $f_Q(q; \beta_\pi, \beta_0, \lambda, \Omega)$ denote this density when $\beta = \beta_\pi$. It is a noncentral Wishart density with means matrix of rank one and identity covariance matrix, which was first derived by Anderson (1946, eqn. (6)). An explicit expression for $f_Q(q; \beta_\pi, \beta_0, \lambda, \Omega)$ is given in (11.2) in the SM.

Now, we determine the limit of the density $f_Q(q; \beta_\pi, \beta_0, \lambda, \Omega)$ as $\beta_0 \to \pm \infty$. Define

$$r_{uv} := \frac{\rho_{uv}}{(1 - \rho_{uv}^2)^{1/2}} \text{ and } \lambda_v := \lambda / \sigma_v^2 = \frac{\mu_\pi^T \mu_\pi}{\sigma_v^2}. \quad (5.1)$$

Note that $\lambda_v$ is the concentration parameter, which indexes the strength of the IV’s. Let $f_Q(q; r_{uv}, \lambda_v)$ denote the density of $Q := [S : T]'[S : T]$ when $[S : T]$ has a multivariate normal distribution with means matrix

$$\mu_\pi \cdot (1/\sigma_v, r_{uv}/\sigma_v) \in \mathbb{R}^{k \times 2}, \quad (5.2)$$

all variances equal to one, and all covariances equal to zero. This density also is a noncentral Wishart density with means matrix of rank one and identity covariance matrix. The density depends on
Comment. Lemma 5.1 is proved by showing: \( \lim_{\beta_0 \to -\infty} \varphi_{\beta_0}(\beta_0, \Omega) = \frac{1}{\sigma_v} \) and \( \lim_{\beta_0 \to +\infty} \varphi_{\beta_0}(\beta_0, \Omega) = \frac{\rho_{uw}}{\sigma_v^2} \), see Lemma 14.1 in the SM. When expressed in terms of \( \Sigma_* \), the latter limit only depends on \( \rho_{uw}, \sigma_u, \) and \( \sigma_v \) and its functional form is of a relatively simple multiplicative form.

Let \( P_{\beta_*\beta_0,\lambda,\Omega}(\cdot) \) and \( P_{\rho_{uw}\lambda,\Omega}(\cdot) \) denote probabilities under the alternative hypothesis densities \( f_Q(q; \beta_*; \beta_0, \lambda, \Omega) \) and \( f_Q(q; \rho_{uw}, \lambda_v) \), respectively, defined above.

### 5.2 Two-Sided AE Power Bound as \( \beta_0 \to \pm \infty \)

AMS provides a two-sided power envelope for invariant similar tests based on maximizing average power against two points in the alternative hypothesis: \((\beta_*, \lambda)\) and \((\beta_2, \lambda_2)\). AMS refers to this as the two-sided AE power envelope because given one point \((\beta_*, \lambda)\), the second point \((\beta_2, \lambda_2)\) is the unique point such that the test that maximizes average power against these two points is a two-sided AE efficient test under strong IV asymptotics. This power envelope is a function of \((\beta_*, \lambda)\).

Given \((\beta_*, \lambda)\), the second point \((\beta_2, \lambda_2)\) satisfies

\[
\beta_2 = \beta_0 - \frac{d_{\beta_0}(\beta_* - \beta_0)}{d_{\beta_0} + 2r_{\beta_0}(\beta_* - \beta_0)} \text{ and } \lambda_2 = \lambda \left( \frac{d_{\beta_0} + 2r_{\beta_0}(\beta_* - \beta_0)}{d_{\beta_0}} \right)^2,
\]

where \( r_{\beta_0} := e_1^T \Omega^{-1} a_0 \cdot (a_0^T \Omega^{-1} a_0)^{-1/2} \), see (4.2) of AMS. We let \( POIS_2(Q; \beta_0; \beta_*, \lambda) \) denote the optimal average-power test statistic for testing \( \beta = \beta_0 \) against \((\beta_*, \lambda)\) and \((\beta_2, \lambda_2)\). Its conditional critical value is denoted by \( \kappa_{2,\beta_0}(Q_T) \). For brevity, the formulas for \( POIS_2(Q; \beta_0; \beta_*, \lambda) \) and \( \kappa_{2,\beta_0}(Q_T) \) are given in Section 16 in the SM.

The limit as \( \beta_0 \to \pm \infty \) of the \( POIS_2(Q; \beta_0; \beta_*, \lambda) \) statistic is shown in (11.4) in the SM to be

\[
POIS_2(Q; \infty, \rho_{uw}, \lambda_v) := \frac{\psi(Q; \rho_{uw}, \lambda_v) + \psi(Q; -\rho_{uw}, \lambda_v)}{2\psi_2(Q_T; \rho_{uw}, \lambda_v)}, \quad \text{where}
\]

\[
\psi(Q; \rho_{uw}, \lambda_v) := \exp(-\lambda_v(1 + r_{uw}^2)/2)(\lambda_v \xi(Q; \rho_{uw}))^{-(k-2)/2}I_{(k-2)/2}(\sqrt{\lambda_v} \xi(Q; \rho_{uw})),
\]

\[
\psi_2(Q_T; \rho_{uw}, \lambda_v) := \exp(-\lambda_v r_{uw}^2/2)(\lambda_v r_{uw}^2 Q_T)^{-(k-2)/2}I_{(k-2)/2}(\sqrt{\lambda_v} r_{uw}^2 Q_T), \text{ and}
\]

\[
\xi(Q; \rho_{uw}) := QS + 2r_{uw} QST + r_{uw}^2 Q_T,
\]

(5.4)
where $Q$, $Q_S$, $Q_{ST}$, and $Q_T$ are defined in (3.1), $\rho_{uv}$ is defined in (4.5), $r_{uv}$ and $\lambda_v$ are defined in (5.1), and $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind of order $\nu$ (e.g., see Comment (ii) to Lemma 3 of AMS for more details regarding $I_\nu(\cdot)$).

Let $\kappa_{2,\infty}(q_T)$ denote the conditional critical value of the $POIS2(Q; \infty, |\rho_{uv}|, \lambda_v)$ test statistic. That is, $\kappa_{2,\infty}(q_T)$ is defined to satisfy

$$P_{Q_1|Q_T}(POIS2(Q; \infty, |\rho_{uv}|, \lambda_v) > \kappa_{2,\infty}(q_T)|q_T) = \alpha$$

(5.5)

for all $q_T \geq 0$, where $P_{Q_1|Q_T}(\cdot|q_T)$ denotes probability under the null density $f_{Q_1|Q_T}(\cdot|q_T)$, which is specified explicitly in (11.3) in the SM and does not depend on $\beta_0$.

When $\rho_{uv} = 0$, the test based on $POIS2(Q; \infty, |\rho_{uv}|, \lambda_v)$ is the AR test. This follows because $\xi(Q; 0) = Q_S$, $\psi(Q; 0, \lambda_v)$ is monotone increasing in $\xi(Q; 0)$, and $\psi_2(Q_T; 0, \lambda_v)$ is a constant. Some intuition for this is that $EQ_{ST} = 0$ under the null and $\lim_{|\beta_0| \to \infty} EQ_{ST} = 0$ under any fixed alternative $\beta_*$ when $\rho_{uv} = 0$:\footnote{We have $EQ_{ST} = ES'ET$ by independence of $S$ and $T$, $EQ_{ST} = 0$ under $H_0$ because $ES = 0$, and $\lim_{|\beta_0| \to \infty} EQ_{ST} = 0$ under $\beta_*$ because $ET = \mu_\cdot d_\cdot_\cdot(\beta_0, \Omega)$, $\lim_{|\beta_0| \to \pm \infty} d_\cdot_\cdot(\beta_0, \Omega) \to \mp r_{uv}/\sigma_v$ by Lemma 14.1(e) in the SM, and $r_{uv} = 0$ when $\rho_{uv} = 0$.} In consequence, $Q_{ST}$ is not useful for distinguishing between $H_0$ and $H_1$ when $|\beta_0| \to \infty$ and $\rho_{uv} = 0$. Furthermore, it is shown in (12.5) and Theorem 12.1 in the SM that the AR test is also the best one-sided test as $\beta_0 \to +\infty$ and as $\beta_0 \to -\infty$.

The following theorem shows that the $POIS2(Q; \infty, |\rho_{uv}|, \lambda_v)$ test provides a two-point average-power bound as $\beta_0 \to \pm \infty$ for any invariant similar test for any fixed $(\beta_*, \lambda)$ and $\Omega$.

**Theorem 5.2** Let $\{\phi_{\beta_0}(Q) : \beta_0 \to \pm \infty\}$ be any sequence of invariant similar level $\alpha$ tests of $H_0 : \beta = \beta_0$ for fixed known $\Omega$. For fixed $(\beta_*, \lambda)$, $(\beta_2, \lambda_2)$ defined (5.3), and $\Omega$, the two-sided AE power envelope test $POIS2(Q; \infty, |\rho_{uv}|, \lambda_v)$ defined in (5.4) and (5.5) satisfies

$$\limsup_{\beta_0 \to \pm \infty} \left( P_{\beta_*, \beta_0, \Omega}(\phi_{\beta_0}(Q) = 1) + P_{\beta_2, \beta_0, \lambda_2, \Omega}(\phi_{\beta_0}(Q) = 1) \right)/2$$

$$\leq P_{\rho_{uv}, \lambda_v}(POIS2(Q; \infty, |\rho_{uv}|, \lambda_v) > \kappa_{2,\infty}(Q_T))$$

$$= P_{-\rho_{uv}, \lambda_v}(POIS2(Q; \infty, |\rho_{uv}|, \lambda_v) > \kappa_{2,\infty}(Q_T)).$$

**Comments.** (i). The power bound in Theorem 5.2 only depends on $(\beta_*, \lambda)$, $(\beta_2, \lambda_2)$, and $\Omega$ through $|\rho_{uv}|$, which is the absolute magnitude of endogeneity under $\beta_*$, and $\lambda_v$, which is the concentration parameter.

(ii). The power bound in Theorem 5.2 is strictly less than one. Hence, it is informative.

(iii). For sign and rotation invariant similar tests $\phi_{\beta_0}(Q)$, the $\limsup$ on the left-hand side in Theorem 5.2 is the average of two equal quantities.
(iv). Theorem 5.2 can be extended to cover sequences of similar tests \( \{\phi_{\beta_0}(S,T) : \beta_0 \to \pm \infty\} \) that satisfy no invariance properties, using the proof of Theorem 1 in CHJ. In this case, the left-hand side (lhs) probabilities in Theorem 5.2 depend on \( \pi \) or, equivalently \( (\lambda, \mu_{\pi}/||\mu_{\pi}||) \), rather than just \( \lambda \). In this case, Theorem 5.2 holds with \( P_{\beta_1,\beta_0,\lambda,\Omega}(\phi_{\beta_0}(Q) = 1) \) replaced by \( \int P_{\beta_1,\lambda,\mu_{\pi}/||\mu_{\pi}||}(\phi_{\beta_0}(S,T) = 1) dUnif(\mu_{\pi}/||\mu_{\pi}||) \) and analogously for the term that depends on \( (\beta_{2*}, \lambda_2) \), where \( P_{\beta_1,\lambda,\mu_{\pi}/||\mu_{\pi}||}(\cdot) \) denotes probability under \( (\beta_1, \lambda, \mu_{\pi}/||\mu_{\pi}||, \Omega) \) and \( Unif(\cdot) \) denotes the uniform measure on the unit sphere in \( R^K \).

5.3 Lower Bound on the Probability That a CS Has Infinite Length

Next, we combine Theorems 4.1 and 5.2 to provide a lower bound on the probability that a sign and rotation invariant similar CS has infinite length. The same lower bound applies to the average probability over \( (\beta_1, \lambda) \) and \( (\beta_{2*}, \lambda) \) that a rotation invariant similar CS has right (left) infinite length. For a similar CS with no invariance properties, the same lower bound applies to a different average probability that the CS has right (left) infinite length.

Let \( P_{\beta_1,\lambda,\Omega}(\cdot) \) denote probability for events determined by \( (Z'Z)^{1/2}Z'Y \) that depend on \( \pi \) only through \( \lambda \), such as events that are determined by a CS based on invariant tests.

**Corollary 5.3** Suppose \( CS_{\phi}(Y) \) is a CS based on invariant similar level \( \alpha \) tests \( \phi(Q_{\beta_0}(Y)) \) that satisfy the continuity condition in Theorem 4.1. (a) For any fixed \( (\beta_1, \lambda, \Omega) \),

\[
(P_{\beta_1,\lambda,\Omega}(RLength(CS_{\phi}(Y)) = \infty) + P_{\beta_{2*},\lambda_2,\Omega}(RLength(CS_{\phi}(Y)) = \infty))/2
\]

\[
\geq 1 - P_{\rho_{uv},\lambda_v}(POIS2(Q; \infty, |\rho_{uv}|, \lambda_v) > \kappa_{2,\infty}(Q_T)) \text{ and }
\]

\[
(P_{\beta_1,\lambda,\Omega}(LLength(CS_{\phi}(Y)) = \infty) + P_{\beta_{2*},\lambda_2,\Omega}(LLength(CS_{\phi}(Y)) = \infty))
\]

\[
\geq 1 - P_{\rho_{uv},\lambda_v}(POIS2(Q; \infty, |\rho_{uv}|, \lambda_v) > \kappa_{2,\infty}(Q_T)).
\]

(b) If the tests \( \phi(Q_{\beta_0}(Y)) \) also are sign invariant, then for any fixed \( (\beta_1, \lambda, \Omega) \),

\[
P_{\beta_1,\pi,\Omega}(Length(CS_{\phi}(Y)) = \infty) \geq 1 - P_{\rho_{uv},\lambda_v}(POIS2(Q; \infty, |\rho_{uv}|, \lambda_v) > \kappa_{2,\infty}(Q_T)).
\]

**Comments.** (i). All three lower bounds in Corollary 5.3 are the same. The different parts of Corollary 5.3 specify different probabilities or average probabilities that have this lower bound.

(ii). Corollary 5.3 (a) also holds for a similar CS that does not satisfy any invariance properties. In this case, \( P_{\beta_1,\lambda,\Omega}(RLength(CS_{\phi}(Y)) = \infty) \) is replaced by \( \int P_{\beta_1,\lambda,\mu_{\pi}/||\mu_{\pi}||}(RLength(CS_{\phi}(Y)) = \infty) dUnif(\mu_{\pi}/||\mu_{\pi}||) \) and analogously for the other three lhs terms that depend on \( LLength(CS_{\phi}(Y)) \).
and/or \( (\beta_2, \lambda_2) \). This holds provided the similar level \( \alpha \) tests \( \phi(S_{\beta_0}(Y), T_{\beta_0}(Y)) \) that define the CS satisfy the conditions in Comment (iii) to Theorem 4.1.

5.4 Power of the AR, LM, and CLR Tests as \( \beta_0 \rightarrow \pm \infty \)

Here, we provide the power of the AR, LM, and CLR tests as \( \beta_0 \rightarrow \pm \infty \) for fixed \((\beta_s, \Omega)\).

Theorem 5.4 For fixed true \((\beta_s, \lambda, \Omega)\), the AR, LM, and CLR tests satisfy

(a) \( \lim_{\beta_0 \rightarrow \pm \infty} P_{\beta_s, \beta_0, \lambda, \Omega}(AR > \chi_{k,1-\alpha}^2/k) = P_{\rho_{uv}, \lambda_v}(AR > \chi_{k,1-\alpha}^2/k) = P(\chi_{k}^2(\lambda_v) > \chi_{k,1-\alpha}^2) \),

(b) \( \lim_{\beta_0 \rightarrow \pm \infty} P_{\beta_s, \beta_0, \lambda, \Omega}(LM > \chi_{1,1-\alpha}^2) = P_{\rho_{uv}, \lambda_v}(LM > \chi_{1,1-\alpha}^2) \), and

(c) \( \lim_{\beta_0 \rightarrow \pm \infty} P_{\beta_s, \beta_0, \lambda, \Omega}(LR > \kappa_{LR,\alpha}(QT)) = P_{\rho_{uv}, \lambda_v}(LR > \kappa_{LR,\alpha}(QT)) \),

where \( AR, LM, \) and \( LR \) are defined as functions of \( Q \) in (3.3), \( \chi_{m,1-\alpha}^2 \) is the \( 1 - \alpha \) quantile of the \( \chi_m^2 \) distribution, and \( \chi_{m}^2(\lambda_v) \) is a noncentral \( \chi_m^2 \) random variable with noncentrality parameter \( \lambda_v \).

Comment. By Theorem 4.1(c), Theorem 5.4 provides the probabilities that the AR, LM, and CLR CS’s have infinite length when the true parameters are \((\beta_s, \lambda, \Omega)\). These probabilities depend only on \((|\rho_{uv}|, \lambda_v)\). For the AR CS, they only depend on \( \lambda_v \).

6 Comparisons of Probabilities That Confidence Sets Have Infinite Length

Next, we investigate how close are the probabilities the CLR CS has infinite length to the lower bound in Corollary 5.3. Let POIS2 refer to the tests that generates the two-sided AE power envelope of AMS. These tests depend on the alternative \((\beta_s, \lambda)\) considered and \( \Omega \). Let POIS2\( \infty \) refer to the tests in (5.4), which are the limits as \( \beta_0 \rightarrow \pm \infty \) of the POIS2 tests. These tests depend on \( \beta_s \) (through \( |\rho_{uv}| \)) and \( \lambda_v \). Let POIS2 and POIS2\( \infty \) CS’s refer to the CS’s constructed by inverting the POIS2 and POIS2\( \infty \) tests. These CS’s are infeasible because they depend on knowing \((\beta_s, \lambda)\).

Table I reports differences in simulated probabilities that the CLR and POIS2\( \infty \) CS’s have infinite lengths. The latter provide a lower bound on infinite-length probabilities for CS’s based on sign and invariant tests, such as the CLR CS, by Corollary 5.3(b). Hence, these differences are necessarily nonnegative. The results cover \( k = 2,5,10,20,40 \), a range of \( \lambda \) values between 1 and 60 depending on the value of \( k \), and \( \rho_{uv} = 0, 3, 5, 7, 9 \). Table I also reports the probabilities that the CLR CS has infinite length for the same \( k \) and \( \lambda \) values and a subset of the \( \rho_{uv} \) values, viz., 0, 7, 9. The true value of \( \beta_s \) is taken to be 0 wlog by Section 2 in the SM. The results for negative and positive \( \rho_{uv} \) values are the same by Section 2 in the SM, and hence, results for negative \( \rho_{uv} \).
are not reported. The number of simulation repetitions employed is 50,000. The critical values are determined using 100,000 simulation repetitions.

The results show that the CLR CS is not close to optimal in some parameter scenarios. In particular, the differences in probabilities of infinite length (DPIL’s) between the CLR and the POIS2∞ CS’s are positive for numerous combinations of \((k, \lambda, \rho_{uv})\). The DPIL’s are increasing in \(k\), decreasing in \(|\rho_{uv}|\), and maximized in the middle of the range of \(\lambda\) values considered. For example, for \((k, \rho_{uv}) = (2, 0)\), DPIL ∈ \([0.002, 0.016]\) over the \(\lambda\) values considered, whereas for \((k, \rho_{uv}) = (5, 0)\), DPIL ∈ \([0.003, 0.031]\) and for \((k, \rho_{uv}) = (40, 0)\), DPIL ∈ \([0.002, 0.049]\)\(^5\). Hence, \(k\) has a noticeable effect on the magnitude of non-optimality of the CLR CS with larger values of \(k\) leading to larger non-optimality.

For \((k, \lambda) = (5, 10)\), we have DPIL ∈ \([0.002, 0.031]\) over the \(\rho_{uv}\) values considered, and for \((k, \lambda) = (20, 15)\), we have DPIL ∈ \([0.001, 0.046]\) over the \(\rho_{uv}\) values considered. Hence, \(|\rho_{uv}|\) also has a noticeable effect on the magnitude of non-optimality of the CLR CS in terms of DPIL’s with non-optimality greatest at \(\rho_{uv} = 0\)\(^6\).

7 Optimality of CLR and LM Tests as \(\rho_{uv} \to \pm 1\) or \(\rho_\Omega \to \pm 1\)

The results of Table I show that the magnitude of non-optimality of the CLR CS decreases as \(|\rho_{uv}|\) increases to 1. This raises the question of whether CLR tests are optimal in some sense in the limit as \(|\rho_{uv}| \to 1\). In this section, we show that this is indeed the case, not just for power as \(\beta_0 \to \pm \infty\), but uniformly over all \((\beta_0, \beta_\lambda)\) parameter values in a two-sided AE power sense.

Let \(\rho_\Omega\) denote the correlation parameter corresponding to the reduced-form variance matrix \(\Omega\), i.e., \(\rho_\Omega := \omega_{12}/(\omega_1 \omega_2)\).

In this section, we provide parameter configurations under which the CLR and LM tests have optimality properties. The results cover the case of strong and semi-strong identification (where \(\lambda \to \infty\)). They cover the cases where \(\rho_{uv} \to \pm 1\) or \(\rho_\Omega \to \pm 1\) for (almost) any fixed values of the other parameters, which includes weak identification of any strength. And, they cover the cases where \((\rho_{uv}, \beta_0) \to (\pm 1, \pm \infty)\) or \((\rho_\Omega, \beta_0) \to (\pm 1, \pm \infty)\) and the other parameters are fixed at (almost) any values, which also includes weak identification.

In somewhat related results, CHJ show that the CLR and LM tests can be written as the limits of certain WAP LR tests, which indicates that they are at least close to being admissible.\(^5\) The simulation standard deviations of the DPIL’s are in the range of \([0.0000, 0.014]\) with most being in the range of \([0.0004, 0.012]\), see Table SM-I in the SM.\(^6\) Table SM-I in the SM shows that the differences in probabilities that the AR and POIS2 CS’s have infinite length are very large for large \(\rho_{uv}\) values for some \(\lambda\) values. For example, for \(\rho_{uv} = .9\), they are as large as \(.084\), \(.196\), \(.280\), \(.353\), \(.422\) for \(k = 2, 5, 10, 20, 40\), respectively, for some \(\lambda\) values. As shown above, \(\text{AR} = \text{POIS2}\) when \(\rho_{uv} = 0\), so the differences are zero in this case and they increase in \(|\rho_{uv}|\) for given \((k, \lambda)\).
Let $d^2_{\beta_*} := d^2_{\beta_*}(\beta_0, \Omega)$ and $c^2_{\beta_*} := c^2_{\beta_*}(\beta_0, \Omega)$. As in Section 5.2, let POIS2($Q; \beta_0, \beta_*, \lambda$) and $\kappa_{2,\beta_0}(Q_T)$ denote the optimal average-power test statistic and its data-dependent critical value. Let $\chi_1^2(c^2_{\infty})$ denote a noncentral $\chi_1^2$ random variable with noncentrality parameter $c^2_{\infty}$.

**Theorem 7.1** Consider any sequence of null parameters $\beta_0$ and true parameters $(\beta_*, \lambda, \Omega)$ such that $\lambda d^2_{\beta_*} \to \infty$ and $\lambda^{1/2}c_{\beta_*} \to c_\infty \in R\setminus\{0\}$. Then, as $\lambda d^2_{\beta_*} \to \infty$ and $\lambda^{1/2}c_{\beta_*} \to c_\infty$,

(a) $P_{\beta_0, \beta_*, \lambda, \Omega}(POIS2(Q; \beta_0, \beta_*, \lambda) > \kappa_{2,\beta_0}(Q_T)) \to P(\chi_1^2(c^2_{\infty}) > \chi_{1,1-\alpha}^2),$

(b) $P_{\beta_0, \beta_*, \lambda, \Omega}(LR > \kappa_{LR,0}(Q_T)) \to P(\chi_1^2(c^2_{\infty}) > \chi_{1,1-\alpha}^2),$ and

(c) $P_{\beta_0, \beta_*, \lambda, \Omega}(LM > \chi_{2,1-\alpha}^2) \to P(\chi_1^2(c^2_{\infty}) > \chi_{1,1-\alpha}^2).$

**Comments.** (i). Theorem 7.1 shows that the CLR and LM tests have the same limit power as the POIS2 test. Theorem 7.1 provides both finite-sample limiting optimality results, where $n$ is fixed and the limits are determined by sequences of parameters $(\beta_0, \beta_*, \lambda, \Omega)$, and large-sample limiting optimality results, where the limits are determined by sequences of sample sizes $n$ and parameters $(\beta_0, \beta_*, \lambda, \Omega)$.

(ii). By Corollary 1 of AMS, for any invariant similar test $\phi(Q)$, for any $(\beta_*, \beta_0, \lambda, \Omega)$,

\[
\frac{1}{2}(P_{\beta_0, \beta_*, \lambda, \Omega}(\phi(Q) = 1) + P_{\beta_2, \beta_0, \lambda, \Omega}(\phi(Q) = 1) \leq P_{\beta_0, \beta_*, \lambda, \Omega}(POIS2(Q; \beta_0, \beta_*, \lambda)). \tag{7.1}
\]

That is, the POIS2 test determines the two-sided AE average power envelope of AMS for invariant similar tests, where the average is over $(\beta_*, \lambda)$ and $(\beta_2, \lambda_2)$. A fortiori, by Theorem 1 of CHJ, for any similar test $\phi([S : T])$ (that is not necessarily invariant), for any $(\beta_*, \beta_0, \lambda, \Omega)$, (7.1) holds with $P_{\beta_0, \beta_*, \lambda, \Omega}(\phi(Q) = 1)$ replaced by the power average $\int P_{\beta_0, \beta_*, \lambda, \Omega}(\phi([S : T]) = 1)dUnif(\mu_\pi/|\|\mu_\pi||))$ and likewise for the second lhs summand in (7.1). Hence, the POIS2 test also determines this average power envelope for similar tests.

These results and Theorem 7.1 show that the CLR and LM tests achieve these average power envelopes for all $(\beta_*, \beta_0, \lambda, \Omega)$ asymptotically when $\lambda d^2_{\beta_*} \to \infty$ and $\lambda^{1/2}c_{\beta_*} \to c_\infty \neq 0$.

(iii) The power envelopes in Comment (ii) translate immediately into false coverage probability (FCP) lower bounds for CS’s based on invariant similar tests and similar tests. Specifically, one minus the lhs in (7.1), which equals the average FCP of the point $\beta_0$ by the CS based on $\phi(Q)$, where the average is over the truth being $(\beta_*, \lambda)$ and $(\beta_2, \lambda_2)$, is greater than or equal to one minus the rhs in (7.1). In the case of non-invariant similar tests, the bound is on the average of the FCP’s of the CS with averaging over $(\beta_*, \lambda)$ and $(\beta_2, \lambda_2)$ and $\mu_\pi/|\|\mu_\pi||$ in the unit sphere in $R^k$. Thus, Theorem 7.1 shows that the CLR and LM CS’s have optimal average FCP properties asymptotically when $\lambda d^2_{\beta_*} \to \infty$ and $\lambda^{1/2}c_{\beta_*} \to c_\infty \neq 0$. 

16
(iv). Theorem 7.1 does not apply when the IV’s are completely irrelevant, i.e., \( \lambda = 0 \), because \( \lambda = 0 \) implies that \( c_\infty = 0 \). However, Theorem 7.1 does cover some cases where the IV’s can be arbitrarily weak, see Theorem 7.2 below.

Next, we provide conditions under which \( \lambda d_{\beta_*}^2 \to \infty \) and \( \lambda^{1/2} c_{\beta_*} \to c_\infty \in R \setminus \{0\} \), as is assumed in Theorem 7.1. First, if \( \beta_0 \) and \( \Omega \) are fixed, \( \Omega \) is nonsingular, and \( (\beta_*, \lambda) \) satisfy \( \lambda \to \infty \) and

\[
\lambda^{1/2} (\beta_* - \beta_0) \to L \in R \text{ as } \lambda \to \infty, \tag{7.2}
\]

then \( \lambda d_{\beta_*}^2 \to \infty \) and \( \lambda^{1/2} c_{\beta_*} \to c_\infty \in R \setminus \{0\} \) with \( c_\infty = L(b_0^T \Omega b_0)^{-1/2} \). Here \( L \) indexes the local alternatives against which the tests have nontrivial power. This result covers the usual strong IV case in which \( \pi \) is fixed, \( Z'Z \) depends on \( n \), and \( \lambda = \pi' Z' Z \pi \to \infty \) as \( n \to \infty \).

The scenario in (7.2) also covers cases where \( \pi = \pi_n \to 0 \) as \( n \to \infty \), but sufficiently slowly that \( \lambda = \pi_n' Z' Z \pi_n \to \infty \) as \( n \to \infty \), which covers “semi-strong” identification. As far as we are aware, this is the only optimality property in the literature for tests under semi-strong identification. The scenario in (7.2) also covers finite-sample, i.e., fixed \( n \), cases in which \( Z'Z \) is fixed, \( \pi \) diverges, i.e., \( ||\pi|| \to \infty \), and \( \lambda_{\min}(Z'Z) > 0 \). In these cases, \( \lambda = \pi' Z' Z \pi \to \infty \) as \( ||\pi|| \to \infty \).

The most novel cases in which Theorem 7.1 applies are when \( \rho_{uv} \to \pm 1 \) or \( \rho_\Omega \to \pm 1 \). The next result shows that \( \lambda d_{\beta_*}^2 \to \infty \) and \( \lambda^{1/2} c_{\beta_*} \to c_\infty \in R \setminus \{0\} \) when \( \rho_{uv} \to \pm 1 \) or \( \rho_\Omega \to \pm 1 \) and the other parameters are fixed at (almost) any values. It also shows that this holds when \( (\rho_{uv}, \beta_0) \to (1, \pm \infty) \) or \( (-1, \pm \infty) \) or \( (\rho_\Omega, \beta_0) \to (1, \pm \infty) \) or \( (-1, \pm \infty) \) and the other parameters are fixed at (almost) any values.

**Theorem 7.2** (a) Suppose the parameters \( \beta_0, \beta_* \), \( \sigma_u > 0, \sigma_v > 0 \), and \( \lambda > 0 \) are fixed, \( \rho_{uv} \in (-1,1) \), and \( \rho_{uv} \to \pm 1 \). Then, (i) \( \lim_{\rho_{uv} \to \pm 1} \lambda^{1/2} c_{\beta_*} = \lambda^{1/2}(\beta_* - \beta_0)/|\sigma_u \pm (\beta_* - \beta_0)\sigma_v| \) and (ii) \( \lim_{\rho_{uv} \to \pm 1} \lambda d_{\beta_*}^2 = \infty \) provided \( \beta_* - \beta_0 \neq \mp \sigma_u/\sigma_v \).

(b) Suppose the parameters \( \beta_0, \beta_* \), \( \omega_1 > 0, \omega_2 > 0 \), and \( \lambda > 0 \) are fixed, \( \rho_\Omega \in (-1,1) \), and \( \rho_\Omega \to \pm 1 \). Then, (i) \( \lim_{\rho_\Omega \to \pm 1} \lambda^{1/2} c_{\beta_*} = \lambda^{1/2}(\beta_* - \beta_0)/|\omega_1 \pm \omega_2\beta_0| \) provided \( \beta_0 \neq \pm \omega_1/\omega_2 \) and (ii) \( \lim_{\rho_\Omega \to \pm 1} \lambda d_{\beta_*}^2 = \infty \) provided \( \beta_0 \neq \pm \omega_1/\omega_2 \) and \( \beta_* \neq \pm \omega_1/\omega_2 \).

(c) Suppose the parameters are as in part (a) except \( (\rho_{uv}, \beta_0) \to (1, \pm \infty) \) or \( (-1, \pm \infty) \). Then, (i) \( \lim_{(\rho_{uv}, \beta_0) \to (1, \pm \infty)} \lambda^{1/2} c_{\beta_*} = \lim_{(\rho_{uv}, \beta_0) \to (1, \pm \infty)} \lambda^{1/2} c_{\beta_*} = \pm \lambda^{1/2}/\sigma_v \) and (ii) \( \lim_{(\rho_{uv}, \beta_0) \to (1, \pm \infty)} \lambda d_{\beta_*}^2 = \lim_{(\rho_{uv}, \beta_0) \to (1, \pm \infty)} \lambda d_{\beta_*}^2 = \infty \).

(d) Suppose the parameters are as in part (b) except \( (\rho_\Omega, \beta_0) \to (1, \pm \infty) \) or \( (-1, \pm \infty) \). Then, (i) \( \lim_{(\rho_\Omega, \beta_0) \to (1, \pm \infty)} \lambda^{1/2} c_{\beta_*} = \lim_{(\rho_\Omega, \beta_0) \to (1, \pm \infty)} \lambda^{1/2} c_{\beta_*} = \pm \lambda^{1/2}/\omega_2 \) and (ii) \( \lim_{(\rho_\Omega, \beta_0) \to (1, \pm \infty)} \lambda d_{\beta_*}^2 = \lim_{(\rho_\Omega, \beta_0) \to (1, \pm \infty)} \lambda d_{\beta_*}^2 = \infty \) provided \( \beta_* \neq \pm \omega_1/\omega_2 \) and \( \lim_{(\rho_\Omega, \beta_0) \to (-1, \pm \infty)} \lambda d_{\beta_*}^2 = \infty \) provided \( \beta_* \neq \omega_1/\omega_2 \).
Comments. (i). Combining Theorems 7.1 and 7.2 provides analytic finite-sample limiting optimality results for the CLR and LM tests and CS’s as \( \rho_{uv} \to \pm 1 \) or \( \rho_\Omega \to \pm 1 \) with \( \beta_0 \) fixed or jointly with \( \beta_0 \to \pm \infty \) for (almost) any fixed values of the other parameters. These results apply for any strength of the IV’s except \( \lambda = 0 \). These results are much stronger than typical weighted average power (WAP) results because they hold for (almost) any fixed values of the parameters \( \beta_0, \beta_*, \sigma_1, \sigma_v, \) and \( \lambda > 0 \) when \( \rho_{uv} \to \pm 1 \) and (almost) any fixed values of the parameters \( \beta_0, \beta_*, \omega_1, \omega_2, \) and \( \lambda > 0 \) when \( \rho_\Omega \to \pm 1 \).

(ii). The cases \( \rho_{uv} \to \pm 1 \) and \( \rho_\Omega \to \pm 1 \) are closely related because \( (1 - \rho_\Omega^2)^{1/2} \omega_1 = (1 - \rho_{uv}^2)^{1/2} \sigma_u \) by (15.10) in the SM. Thus, \( \rho_{uv} \to \pm 1 \) implies \( |\rho_\Omega| \to 1 \) and/or \( \omega_1 \to 0 \). And, \( \rho_\Omega \to \pm 1 \) implies \( |\rho_{uv}| \to 1 \) and/or \( \sigma_u \to 0 \).

8 General Power/False-Coverage-Probability Comparisons

By Theorem 4.1, the results in Table I equal power differences (PD’s) between the POIS2 and CLR tests as the null value \( \beta_0 \to \pm \infty \) for fixed true value \( \beta_* = 0 \). Here, we consider PD’s between the POIS2 and CLR tests for finite \( \beta_0 \) values, rather than PD’s as \( \beta_0 \to \pm \infty \). Specifically, Table II reports maximum and average PD’s over \( \beta_0 \in R \) and \( \lambda > 0 \) for a fixed true value \( \beta_* = 0 \) for a range of values of \( (\rho_{uv}, k) \). As above, the choice of \( \beta_* = 0 \) (and \( \omega_1^2 = \omega_2^2 = 1 \)) is wlog. These PD’s are equivalent to false coverage probability differences (FCPD’s) between the CLR and POIS2 CS’s for a fixed true value \( \beta_* \) at incorrect values \( \beta_0 \). They are necessarily nonnegative.

The \( \lambda \) values considered are 1, 3, 5, 7, 10, 15, 20, as well as 22, 25 when \( k = 20 \) and 40, and .7, .8, .9 when \( k = 2 \) and 5 and \( \rho_{uv} = .9 \). The positive and negative \( \beta_0 \) values considered are those with \( |\beta_0| \in \{.25, .5, \ldots, 3.75, 4, 5, 7.5, 10, 50, 100, 1000, 10000\} \).

The number of simulation repetitions employed is 5,000. The critical values are determined using 100,000 simulation repetitions. For example, the simulation standard deviations for the PD’s for \( (\rho_{uv}, k) = (0, 20) \) and any fixed \( (\beta_0, \lambda) \) value range from [.0013, .0040] across different \( (\beta_0, \lambda) \) values, which compares to simulated averages of the PD’s over \( (\beta_0, \lambda) \) values that are of the .014 order of magnitude.

Tables II(a) and II(b) contain the same numbers, but are reported differently to make the patterns in the table more clear. Table II(a) shows variation across \( k \) for fixed \( \rho_{uv} \), whereas Table II(b) shows variation across \( \rho_{uv} \) for fixed \( k \). The third and fourth columns in each table report the values of \( \lambda \) and \( \beta_0 \) at which the maximum PD is obtained. The fifth column in each table reports \( \rho_{uv,0} \), which is the correlation between the structural-equation and reduced-form errors when \( \beta_0 \) is the true value (based on the assumption that the consistently-estimable reduced-form variance...
matrix is the same whether the truth is $\beta_0$ or $\beta_*$). In contrast, $\rho_{uv}$ is the same correlation, but when $\beta_*$ is the true value—which is the true $\beta$ value in the PD simulations. The sixth column in the tables reports the power of the CLR test at the $(\beta_0, \lambda)$ values that maximize the PD for given $(\rho_{uv}, k)$, i.e., at $(\beta_{0,\text{max}}, \lambda_{\text{max}})$.

Table II shows that the maximum (over $(\beta_0, \lambda)$) PD’s between the POIS2 and CLR tests range between [.016, .061] over the $(\rho_{uv}, k)$ values. On the other hand, the average (over $(\beta_0, \lambda)$) PD’s only range between [.002, .016] over the $(\rho_{uv}, k)$ values. This indicates that, although there are some $(\beta_0, \lambda)$ values at which the CLR test is noticeably off the two-sided AE power envelope, on average the CLR test’s power is not far from the power envelope.

In contrast, the analogous maximum and average PD ranges for the AR test are [.079, .513] and [.012, .179], see Table SM-III in the SM. For the LM test, they are [.242, .784] and [.010, .203], see Table SM-IV in the SM. Hence, the power of AR and LM tests is very much farther from the POIS2 power envelope than is the power of the CLR test.

Table II(a) shows that the maximum and average (over $(\beta_0, \lambda)$) PD’s for the CLR test are clearly increasing in $k$. Table II(a) shows that for $\rho_{uv} \geq .3$, the PD’s are maximized at more or less the same $\beta_0$ regardless of the value of $k$. For $\rho_{uv} = 0$, this is also true to a certain extent, because the sign of $\beta_0$ is irrelevant (when $\rho_{uv} = 0$) and the values 50 and 10,000 are both large values. Table II(a) also shows that for each $\rho_{uv}$, the PD’s are maximized at $\lambda$ values that (weakly) increase with $k$. The increase is particularly evident going from $k = 20$ to 40.

Table II(b) shows that for $k \geq 5$, the maximum PD’s are more or less the same for $\rho_{uv} \leq .7$, but noticeably lower for $\rho_{uv} = .9$. For $k = 2$, the maximum PD’s are more or less the same for all $\rho_{uv}$ considered. Table II(b) shows that, for each $k$, the PD’s are maximized at $|\beta_0|$ values that are closer to 0 as $\rho_{uv}$ increases. Table II(b) also shows that, for each $k$, the PD’s are maximized at $\lambda$ values that are closer to 0 as $\rho_{uv}$ increases.

In sum, the maximum PD’s over $(\beta_0, \lambda)$ are found to increase in $k$ ceteris paribus and decrease in $\rho_{uv}$ ceteris paribus. The $\lambda$ values at which the maxima are obtained are found to (weakly) increase with $k$ ceteris paribus and decrease in $\rho_{uv}$ ceteris paribus. The $|\beta_0|$ values at which the maxima are obtained are found to be independent of $k$ ceteris paribus and decrease in $\rho_{uv}$ ceteris paribus.

The numerical results in this section show that the finding of AMS that the CLR test is essentially on the two-sided AE power envelope does not hold when one considers a broader range of null and alternative hypothesis values $(\beta_0, \beta_*)$ than those considered in the numerical results in AMS.

\footnote{See Table SM-II in the SM for how the maximum PD’s over $\beta_0$ vary with $\lambda$ for the $(\rho_{uv}, k)$ values in Table II.}
9 Differences between CLR Power and an Average Over $\lambda$

Power Envelope

In this section, we introduce a “WAP2” power envelope for similar tests with weight functions over: (i) a finite grid of $\lambda$ values, $\{\lambda_j > 0 : j \leq J\}$, (ii) the same two-points ($\beta_*, \lambda_j$) and ($\beta_{2*}, \lambda_{2j}$) as in AMS for each $\lambda_j$ for $j \leq J$, and (iii) the same uniform weight function over $\mu_j/||\mu_z||$ as in CHJ. In particular, we use the uniform weight function over the 36 values of $\lambda$ in $\{2.5, 5.0, ..., 90.0\}$.

The WAP2 envelope is a function of $(\beta_0, \beta_*)$. The WAP2($Q, \beta_0, \beta_*$) test statistic that generates this envelope is of the form $\sum_{j=1}^{J}(\psi(Q; \beta_0, \beta_{*j}, \lambda_j) + \psi(Q; \beta_0, \beta_{2*j}, \lambda_{2j})) / \sum_{j=1}^{J} 2\psi_2(Q_T; \beta_0, \beta_*, \lambda_j)$, where the functions $\psi(Q; \beta_0, \beta, \lambda)$ and $\psi_2(Q_T; \beta_0, \beta, \lambda)$ are as in AMS (and as in (11.5) in the SM). The WAP2($Q, \beta_0, \beta_*$) conditional critical value $\kappa_{2, \beta_0, \beta_*}(q_T)$ is defined to satisfy $P_{Q_T|Q_T}(WAP2(Q, \beta_0, \beta_*) > \kappa_{2, \beta_0, \beta_*}(q_T)|q_T) = \alpha$ for all $q_T \geq 0$, where $P_{Q_T|Q_T}(\cdot|q_T)$ denotes probability under the density $f_{Q_T|Q_T}(\cdot|q_T)$, which is specified in (11.3) in the SM.

To be consistent with Tables I and II, we report PD’s between the WAP2($Q, \beta_0, \beta_*$) and CLR tests for $\beta_*$ = 0 and a range of $\beta_0$ values. These PD’s are equivalent to the FCPD’s between the CLR and WAP2 CS’s for fixed true $\beta_*$ and varying incorrect $\beta_0$ values. The differences are necessarily nonnegative.

We consider $\rho_{uv} \in \{0, 3, .5, .7, .9, .95, .99\}$, $k = 2, 5, 10, 20, 40$, the same $\beta_0$ values as in Table II, and $\omega^2_1 = \omega^2_2 = 1$. Since $\beta_*$ = 0, $\rho_Q = \rho_{uv}$. Section [21] shows that taking $\beta_*$ = 0 and $\omega^2_1 = \omega^2_2 = 1$ is wlog provided the support of the weight function for $\lambda$ is scaled by $\omega^2_2$ when $\omega_2 \neq 1$. The number of simulation repetitions employed is 1,000 for each $\lambda_j$ value. With power averaged over the 36 $\lambda_j$ values and independence of the simulation draws across $\lambda_j$, this yields simulation SD’s that are comparable to using 36,000 simulation repetitions. The critical values are determined using 100,000 simulation repetitions for $k = 5$ and 10,000 for other values of $k$.

For brevity, Table III reports results only for $k = 5$ for a subset of the $\beta_0$ values considered. Results for all values of $k$ and $\beta_0$ considered are given in Table SM-V in the SM. Table IV reports summary results for all values of $k$. In particular, Table IV(a) provides the maxima over $\beta_0$ of the average over $\lambda$ PD’s for each $(\rho_{uv}, k)$. Table IV(b) provides the average over $\beta_0$ of the average over $\lambda$ PD’s for each $(\rho_{uv}, k)$.

Table III shows that the CLR test has power quite close to the WAP2 power envelope for $k = 5$. The PD’s for $\rho_{uv} \in \{0, .3, .5, .7\}$, we have PD $\in [.000, .005]$ and SD $\in [.0003, .0007]$ across all $\beta_0$ values. For $\rho_{uv} \in \{.9, .95, .99\}$, we have PD $\in [.000, .001]$ and SD $\in [.00000, .00003]$ across all $\beta_0$ values.

Table IV shows that PD’s between the WAP2 power envelope and the CLR power are increasing in $k$ and decreasing in $|\rho_{uw}|$. For $k = 2$, the maximum PD over $\beta_0$ and $\rho_{uv}$ values is very small:
In the worst case for CLR, which is when \((k, \rho_{uv}) = (40, 0)\), the maximum PD over \(\beta_0\) values is substantially larger: .024. The average (over \(\beta_0\) values) PD in this case is .013, which is not very large. For \(k = 40\) and \(\rho_{uv} \geq .9\), the maximum PD (over \(\beta_0\) and \(\rho_{uv}\) values) is very small: .004. This is consistent with the theoretical optimality properties of the CLR test as \(\rho_{uv} \rightarrow \pm 1\) described in Section 7. For \(k = 40\) and \(\rho_{uv} \geq .9\), the average PD (over \(\beta_0\) values and the five \(\rho_{uv}\) values) is very small: .000. The second worst case for CLR in Table V is when \((k, \rho_{uv}) = (20, 0)\). In this case, the maximum PD over \(\beta_0\) values is .013, which is noticeably lower than .024 for \((k, \rho_{uv}) = (40, 0)\).

In conclusion, the results in Tables III and IV show that the CLR test is very close to the WAP2 power envelope for most \((k, \rho_{uv}, \beta_0)\) values, but can deviate from it by as much as .024 for some \(\beta_0\) values when \((k, \rho_{uv}) = (40, 0)\).
References

Anderson, T. W. (1946): “The Non-central Wishart Distribution and Certain Problems of Multivariate Statistics,” *Annals of Mathematical Statistics*, 17, 409-431.

Anderson, T. W. and H. Rubin (1949): “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations,” *Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K. (1998): “Hypothesis Testing with a Restricted Parameter Space,” *Journal of Econometrics*, 84, 155–199.

Andrews, D. W. K., and P. Guggenberger (2015): “Identification- and Singularity-Robust Inference for Moment Condition Models,” Cowles Foundation Discussion Paper No. 1978, Yale University.

——— (2016): “Asymptotic Size of Kleibergen’s LM and Conditional LR Tests for Moment Condition Models,” *Econometric Theory*, 32, forthcoming. Also available as Cowles Foundation Discussion Paper No. 1977, Yale University.

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.

——— (2006): “Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression,” *Econometrica*, 74, 715–752.

Andrews, D. W. K. and W. Ploberger (1994): “Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative,” *Econometrica*, 63, 1383–1414.

Chamberlain, G. (2007): “Decision Theory Applied to an Instrumental Variables Model,” *Econometrica*, 75, 609–652.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): “Admissible Invariant Similar Tests for Instrumental Variables Regression,” *Econometric Theory*, 25, 806–818.

Dufour, J.-M. (1997): “Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models,” *Econometrica*, 65, 1365–1387.

Elliott, G., U. K. Müller, and M. W. Watson (2015): “Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis,” *Econometrica*, 83, 771–811.
Hillier, G. H. (1984): “Hypothesis Testing in a Structural Equation: Part I, Reduced Form Equivalence and Invariant Test Procedures,” unpublished manuscript, Dept. of Econometrics and Operations Research, Monash University.

—— (2009): “Exact Properties of the Conditional Likelihood Ratio Test in an IV Regression Model,” *Econometric Theory*, 25, 915–957.

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

Mikusheva, A. (2010): “Robust Confidence Sets in the Presence of Weak Instruments,” *Journal of Econometrics*, 157, 236–247.

Mills, B., M. J. Moreira, and L. P. Vilela (2014): “Tests Based on t-Statistics for IV Regression with Weak Instruments,” *Journal of Econometrics*, 182, 351–363.

Moreira, M. J. (2003): “A Conditional Likelihood Ratio Test for Structural Models,” *Econometrica*, 71, 1027-1048.

——— (2009): “Tests with Correct Size When Instruments Can Be Arbitrarily Weak, *Journal of Econometrics*, 152, 131–140.

Moreira, H. and M. J. Moreira (2013): “Contributions to the Theory of Optimal Tests,” unpublished manuscript, FGV/EPGE, Brasil.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press.

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

Wald, A. (1943): “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large,” *Transactions of the American Mathematical Society*, 54, 426–482.
TABLE I. Differences in Probabilities of Infinite-Length CI's for the CLR and POIS2\(_\infty\) CI's, and Probabilities of Infinite-Length POIS2\(_\infty\) CI's as Functions of \(k\), \(\lambda\) and \(\rho_{uv}\)

| \(k\) | \(\lambda\) | \(\rho_{uv} = 0\) | \(.3\) | \(.5\) | \(.7\) | \(.9\) | \(\rho_{uv} = 0\) | \(.7\) | \(.9\) |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 2   | 1   | .002 | .004 | .003 | .002 | .002 | .870 | .863 | .850 |
| 2   | 3   | .007 | .003 | .002 | .001 | .001 | .680 | .656 | .613 |
| 2   | 5   | .012 | .008 | .003 | .002 | .001 | .497 | .455 | .410 |
| 2   | 7   | .016 | .055 | .001 | .001 | -.000 | .344 | .298 | .262 |
| 2   | 10  | .013 | .006 | .002 | .001 | .000 | .184 | .142 | .122 |
| 2   | 15  | .008 | .004 | .002 | .001 | .001 | .056 | .035 | .029 |
| 2   | 20  | .003 | .001 | .001 | .000 | -.000 | .014 | .008 | .007 |
| 5   | 1   | .003 | .003 | .002 | .002 | .002 | .900 | .898 | .882 |
| 5   | 3   | .010 | .010 | .006 | .002 | .004 | .778 | .751 | .671 |
| 5   | 5   | .017 | .012 | .004 | .003 | -.001 | .639 | .576 | .466 |
| 5   | 7   | .026 | .012 | .003 | .001 | -.002 | .502 | .407 | .304 |
| 5   | 10  | .031 | .016 | .005 | .004 | .002 | .323 | .221 | .144 |
| 5   | 12  | .029 | .011 | .003 | .000 | -.001 | .231 | .140 | .086 |
| 5   | 15  | .023 | .013 | .005 | .003 | .001 | .134 | .062 | .036 |
| 5   | 20  | .013 | .007 | .003 | .001 | .000 | .048 | .016 | .008 |
| 5   | 25  | .006 | .002 | .001 | .000 | -.000 | .016 | .004 | .001 |
| 10  | 1   | .001 | .002 | .002 | .001 | .000 | .919 | .916 | .904 |
| 10  | 5   | .018 | .015 | .008 | .005 | .004 | .731 | .667 | .526 |
| 10  | 10  | .032 | .014 | .003 | .005 | .001 | .459 | .320 | .176 |
| 10  | 15  | .037 | .020 | .009 | .005 | .001 | .239 | .111 | .046 |
| 10  | 17  | .035 | .017 | .008 | .003 | .000 | .176 | .069 | .025 |
| 10  | 20  | .027 | .014 | .006 | .002 | .000 | .110 | .032 | .010 |
| 10  | 25  | .017 | .008 | .003 | .001 | .000 | .045 | .008 | .002 |
| 10  | 30  | .009 | .002 | .002 | .000 | -.000 | .017 | .002 | .000 |
| 20  | 1   | .001 | .003 | .002 | .002 | .000 | .929 | .926 | .919 |
| 20  | 5   | .014 | .011 | .006 | .003 | .006 | .809 | .766 | .620 |
| 20  | 10  | .035 | .022 | .011 | .010 | .001 | .603 | .463 | .246 |
| 20  | 15  | .046 | .025 | .011 | .008 | .001 | .390 | .215 | .073 |
| 20  | 20  | .044 | .021 | .011 | .005 | .000 | .226 | .080 | .018 |
| 20  | 30  | .033 | .012 | .005 | .001 | .000 | .056 | .007 | .001 |
| 20  | 40  | .007 | .003 | .001 | .000 | -.000 | .010 | .001 | .000 |
| 40  | 1   | .002 | .001 | .001 | .001 | .000 | .937 | .937 | .934 |
| 40  | 5   | .011 | .008 | .005 | .001 | .009 | .859 | .838 | .717 |
| 40  | 10  | .028 | .020 | .009 | .009 | .002 | .721 | .614 | .354 |
| 40  | 15  | .043 | .022 | .011 | .009 | .001 | .555 | .372 | .129 |
| 40  | 20  | .049 | .029 | .013 | .009 | .001 | .394 | .186 | .038 |
| 40  | 30  | .043 | .021 | .011 | .004 | .000 | .155 | .028 | .002 |
| 40  | 40  | .022 | .012 | .005 | .001 | .000 | .046 | .003 | .000 |
| 40  | 60  | .002 | .001 | .000 | .000 | .000 | .003 | .000 | .000 |
TABLE II. Maximum and Average Power Differences over $\lambda$ and $\beta_0$ Values between POIS2 and CLR Tests for Fixed Alternative $\beta^* = 0$

(a) Across $k$ patterns for fixed $\rho_{uv}$

| $\rho_{uv}$ | $k$ | $\lambda_{\text{max}}$ | $\beta_{0,\text{max}}$ | $\rho_{uv,0}$ | POIS2-CLR | POIS2-CLR |
|-------------|-----|-----------------|-----------------|--------------|----------|----------|
|             |     |                 |                 |              | max      | average  |
| .0          | 2   | 7               | -10000.00       | 1.00         | .66      | .021     | .006     |
| .0          | 5   | 10              | -50.00          | 1.00         | .68      | .030     | .009     |
| .0          | 10  | 15              | -50.00          | 1.00         | .76      | .038     | .012     |
| .0          | 20  | 15              | 10.00           | -1.00        | .60      | .042     | .014     |
| .0          | 40  | 22              | -50.00          | 1.00         | .66      | .059     | .016     |
| .3          | 2   | 10              | 3.75            | -0.96        | .86      | .019     | .005     |
| .3          | 5   | 10              | 3.50            | -0.96        | .73      | .034     | .008     |
| .3          | 10  | 10              | 3.00            | -0.94        | .59      | .032     | .009     |
| .3          | 20  | 15              | 3.50            | -0.96        | .66      | .045     | .012     |
| .3          | 40  | 22              | 4.00            | -0.97        | .72      | .061     | .014     |
| .5          | 2   | 5               | 2.00            | -0.87        | .64      | .016     | .004     |
| .5          | 5   | 10              | 2.25            | -0.90        | .82      | .029     | .005     |
| .5          | 10  | 10              | 2.00            | -0.87        | .70      | .037     | .007     |
| .5          | 20  | 10              | 1.75            | -0.82        | .53      | .046     | .009     |
| .5          | 40  | 15              | 1.75            | -0.82        | .59      | .050     | .012     |
| .7          | 2   | 5               | 1.50            | -0.75        | .81      | .016     | .002     |
| .7          | 5   | 5               | 1.50            | -0.75        | .67      | .033     | .003     |
| .7          | 10  | 7               | 1.50            | -0.75        | .71      | .036     | .005     |
| .7          | 20  | 7               | 1.25            | -0.61        | .54      | .042     | .006     |
| .7          | 40  | 15              | 1.50            | -0.75        | .84      | .050     | .008     |
| .9          | 2   | 0.9             | 1.25            | -0.63        | .46      | .017     | .002     |
| .9          | 5   | 0.9             | 1.00            | -0.22        | .33      | .017     | .002     |
| .9          | 10  | 3               | 1.25            | -0.63        | .77      | .027     | .003     |
| .9          | 20  | 3               | 1.00            | -0.22        | .61      | .032     | .003     |
| .9          | 40  | 5               | 1.25            | -0.63        | .75      | .040     | .004     |

(b) Across $\rho_{uv}$ patterns for fixed $k$

| $k$ | $\rho_{uv}$ | $\lambda_{\text{max}}$ | $\beta_{0,\text{max}}$ | $\rho_{uv,0}$ | POIS2-CLR | POIS2-CLR |
|-----|-------------|-----------------|-----------------|--------------|----------|----------|
|     |             |                 |                 |              | max      | average  |
| 2   | .0          | 7               | -10000.00       | 1.00         | .66      | .021     | .006     |
| 2   | .3          | 10              | 3.75            | -0.96        | .86      | .019     | .005     |
| 2   | .5          | 5               | 2.00            | -0.87        | .64      | .016     | .004     |
| 2   | .7          | 5               | 1.50            | -0.75        | .81      | .016     | .002     |
| 2   | .9          | 5               | 1.25            | -0.63        | .75      | .040     | .004     |
| 5   | .0          | 10              | -50.00          | 1.00         | .68      | .030     | .009     |
| 5   | .3          | 10              | 3.50            | -0.96        | .73      | .034     | .008     |
| 5   | .5          | 10              | 2.25            | -0.90        | .59      | .032     | .009     |
| 5   | .7          | 5               | 1.50            | -0.75        | .66      | .045     | .012     |
| 5   | .9          | 9               | 1.25            | -0.63        | .72      | .061     | .014     |
| 10  | .0          | 15              | -50.00          | 1.00         | .76      | .038     | .012     |
| 10  | .3          | 10              | 3.00            | -0.94        | .59      | .032     | .009     |
| 10  | .5          | 10              | 2.00            | -0.87        | .70      | .037     | .007     |
| 10  | .7          | 7               | 1.50            | -0.75        | .71      | .036     | .005     |
| 10  | .9          | 3               | 1.25            | -0.63        | .77      | .027     | .003     |
| 20  | .0          | 15              | 10.00           | -1.00        | .66      | .042     | .014     |
| 20  | .3          | 15              | 3.50            | -0.96        | .66      | .045     | .012     |
| 20  | .5          | 10              | 1.75            | -0.82        | .53      | .046     | .009     |
| 20  | .7          | 7               | 1.25            | -0.61        | .54      | .042     | .006     |
| 20  | .9          | 3               | 1.00            | -0.22        | .61      | .032     | .003     |
| 40  | .0          | 22              | -50.00          | 1.00         | .66      | .059     | .016     |
| 40  | .3          | 22              | 4.00            | -0.97        | .72      | .061     | .014     |
| 40  | .5          | 15              | 1.75            | -0.82        | .59      | .050     | .012     |
| 40  | .7          | 15              | 1.50            | -0.75        | .84      | .050     | .008     |
| 40  | .9          | 5               | 1.25            | -0.63        | .75      | .040     | .004     |
| $\beta_0$ | $\rho_{uv,0}$ | $\rho_{uv} = 0$ | .9 | $\rho_{uv} = 0$ | .3 | .5 | .7 | .9 | .95 | .99 |
|-----------|---------------|----------------|----|----------------|----|----|----|----|-----|-----|
| -100.00   | 1.00          | 1.00           | .005 | .002          | .001       | .000 | .000 | .000 | .000 | .000 |
| -10.00    | 1.00          | 1.00           | .005 | .002          | .001       | .000 | .000 | .000 | .000 | .000 |
| -4.00     | .97           | 1.00           | .003 | .001          | .000       | .000 | .000 | .000 | .000 | .000 |
| -3.00     | .95           | .99            | .003 | .001          | .000       | .000 | .000 | .000 | .000 | .000 |
| -2.00     | .89           | .99            | .002 | .001          | .000       | .000 | .000 | .000 | .000 | .000 |
| -1.50     | .83           | .98            | .001 | .001          | .001       | .000 | .000 | .000 | .000 | .000 |
| -1.00     | .71           | .97            | .001 | .000          | .000       | .000 | .000 | .000 | .000 | .000 |
| -0.75     | .60           | .97            | .000 | .000          | .001       | .000 | .000 | .000 | .000 | .000 |
| -0.50     | .45           | .95            | -.000 | -.000      | -.001       | -.000 | -.000 | -.000 | -.000 | -.000 |
| -0.25     | .24           | .94            | -.001 | -.001      | -.001       | -.000 | -.000 | -.000 | -.000 | -.000 |
| 0.25      | -.24          | .83            | -.000 | -.001      | -.001       | -.000 | -.000 | -.000 | -.000 | -.000 |
| 0.50      | -.45          | .68            | .001 | .000          | .000       | .000 | .000 | .000 | .000 | .000 |
| 0.75      | -.60          | .33            | .000 | .001          | .001       | .000 | .000 | .000 | .000 | .000 |
| 1.00      | -.71          | -.22           | .002 | .001          | .001       | .000 | .000 | .000 | .000 | .000 |
| 1.50      | -.83          | -.81           | .001 | .002          | .003       | .003 | .001 | .000 | .000 | .000 |
| 2.00      | -.89          | -.93           | .002 | .003          | .004       | .005 | .002 | .000 | .000 | .000 |
| 3.00      | -.95          | -.98           | .003 | .005          | .005       | .005 | .001 | .000 | .000 | .000 |
| 4.00      | -.97          | -.99           | .004 | .005          | .006       | .006 | .004 | .000 | .000 | .000 |
| 10.00     | -1.00         | -1.00          | .005 | .003          | .003       | .003 | .001 | .000 | .000 | .000 |
| 100.00    | -1.00         | -1.00          | .005 | .003          | .001       | .001 | .000 | .000 | .000 | .000 |
| 10000.00  | -1.00         | -1.00          | .005 | .002          | .001       | .001 | .000 | .000 | .000 | .000 |

**TABLE III.** Average (over $\lambda$) Power Differences for $\lambda \in \{2.5, 5.0, ..., 90.0\}$ between the WAP2 and CLR Tests for $k = 5$

**TABLE IV.** Average (over $\lambda$) Power Differences between the WAP2 and CLR Tests