ROBUST CONFIDENCE SETS IN THE PRESENCE OF WEAK INSTRUMENTS

By Anna Mikusheva [1],

*MIT, Department of Economics*

**Abstract**

This paper considers instrumental variable regression with a single endogenous variable and the potential presence of weak instruments. I construct confidence sets for the coefficient on the single endogenous regressor by inverting tests robust to weak instruments. I suggest a numerically simple algorithm for finding the Conditional Likelihood Ratio (CLR) confidence sets. Full descriptions of possible forms of the CLR, Anderson- Rubin (AR) and Lagrange Multiplier (LM) confidence sets are given. I show that the CLR confidence sets have nearly shortest expected arc length among similar symmetric invariant confidence sets in a circular model. I also prove that the CLR confidence set is asymptotically valid in a model with non-normal errors.

`Key Words`: weak instruments, confidence set, uniform asymptotics

JEL Classifications: C30

# 1 Introduction

This paper considers confidence sets for the coefficient $\beta$ on the single endogenous regressor in an instrumental variable (IV) regression. A confidence set provides information about a range of parameter values compatible with the data. A good confidence set should adequately describe sampling uncertainty observed in the data. In particular, a confidence set should be large, possibly infinite (in the case of unbounded parameter space), if the data contains very little or no information about a parameter. In many empirically relevant situations, the correlation between the instruments and the endogenous regressor is almost indistinguishable from zero (so called weak instruments case), and little or no information about $\beta$ can be extracted. When instruments can be arbitrary weak a confidence set with correct coverage prob-

[1] Dept. of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142, U.S.A.; (617)324-5459, fax-(617) 253-1330; *amikushe@mit.edu* .

ability must have an infinite length with positive probability (Gleser and Hwang (1987), Dufour (1997)). Most empirical applications use the conventional Wald confidence interval, which is always finite. As a result, the Wald confidence interval has a low coverage probability (Nelson and Startz (1990)) and should not be used when instruments are weak (Dufour (1997)).

To construct a confidence set robust to weak instruments, one can invert a test which has the correct size even when instruments are weak (Lehmann (1986)). Namely, a confidence set with correct coverage can be constructed as the set of $\beta_0$ for which the hypothesis $H_0 : \beta = \beta_0$ is accepted. The idea of inverting robust tests in the context of IV regression was first proposed by Anderson and Rubin (1949) and has recently been used by many authors, including Moreira (2002), Stock, Wright and Yogo (2002), Dufour, Khalaf and Kichian (2005), and Kleibergen and Mavroeidis (2009). The class of tests robust to weak identification includes but is not limited to the Anderson and Rubin (1949) (AR) test, the Lagrange multiplier (LM) test proposed by Kleibergen (2002) and Moreira (2002), and the Conditional Likelihood Ratio (CLR) test suggested by Moreira (2003).

This paper has three main goals. The first is to compare the CLR, AR, and LM confidence sets using accuracy and length as criteria. The second goal is to provide a practitioner with simple and fast algorithms for obtaining these confidence sets; currently a fast inversion algorithm exists for AR but not for the CLR or the LM. Last but not least, I prove that the confidence sets mentioned above have asymptotically correct coverage; this entails a non-trivial extension of point-wise validity arguments in the literature to uniform validity.

Accuracy of a confidence set is defined as the probability of excluding false values of the parameter of interest. A uniformly most accurate (UMA) confidence set maximizes the probability of excluding each false value. A UMA confidence set corresponds to a uniformly most powerful (UMP) test and vice versa. Practitioners are usually more interested in another criterion, the expected length. According to Pratt's (1961) theorem (also see Ghosh (1961)), the expected length of a confidence set equals the integral over false values of the probability that each false value is included. If the expected length is *finite*, then a UMA confidence set is of the shortest expected length.

Andrews, Moreira and Stock (2006) show that the CLR test is nearly UMP in the class of *two-sided* similar tests invariant with respect to orthogonal transformations of instruments. This suggests that a confidence set corresponding to the CLR test may possess some optimality properties with respect to length. There are, however, two obstacles in applying Pratt's theorem directly. First, the expected length of a confidence set with correct coverage in the case of weak instruments must be infinite. Second, the CLR does not maximize power at every point, rather it nearly maximizes the average power at two points lying on different sides of the true value. The locations of the points depend on each other, but they are not symmetric, at least in the native parametrization of the IV model.

The reasons stated above prevent establishing "length optimality" of the CLR confidence set in the native parametrization. However, in a circular version (re-parametrization into spherical coordinates) of the simultaneous equation model considered in the statistics literature by Anderson (1976), and Anderson, Stein and Zaman (1985) and suggested in the present context by Hillier (1990) and Chamberlain (2005), the CLR sets have some near optimality properties. In spherical coordinates the parameter of interest, $\phi$, lies on a one-dimensional unit circle. This parameter, $\phi$, is in one-to-one correspondence with the coefficient, $\beta$, on the endogenous regressor. Inferences on $\phi$ can be easily translated to inferences on $\beta$ and vice versa. This circular model has two nice features. First, the length of the parameter space for $\phi$ is finite, which makes every confidence set for $\phi$ finite (a confidence interval of length $Pi$ for $\phi$ corresponds to a confidence set for $\beta$ equal to the whole line). Second, a circular model possesses additional symmetry and invariance properties. In particular, the 2-sidedness condition corresponds to a *symmetry* on the circle. I show that the CLR confidence set has nearly minimal arc length among symmetric similar invariant confidence sets in a simultaneous equation model formulated in spherical coordinates.

I use simulations to examine the distribution of the lengths of the CLR, AR, and LM confidence sets for $\beta$ in linear coordinates. I also compute their expected lengths over a fixed finite interval. I find that the distribution of the length of the CLR confidence set is first order stochastically dominated by the distribution of the length of the LM confidence set. It is, therefore, not advisable to use the LM confidence set in practice.

If one compares the length of the CLR and AR sets over a fixed finite interval, then the CLR confidence set is usually shorter. The distributions of length of the AR and CLR confidence sets, however, do not dominate one another in a stochastic sense. The reason is that the AR confidence set can be empty with non-zero probability. In other words, the distribution of length of the AR confidence set has a mass point at zero. This peculiarity of the AR confidence set can be quite confusing for applied researchers, since an empty interval makes inferences impractical.

This paper also addresses the practical problem of inverting the CLR, LM and AR tests. One way of inverting a test is to perform grid testing, namely, to perform a series of tests $H_0 : \beta = \beta_0$, where $\beta_0$ belongs to a fine grid. This procedure, however, is numerically cumbersome. Due to the simple form of the AR and LM tests, it is relatively easy to invert them by solving polynomial inequalities (this is known for the AR, but apparently not for the LM). The problem of inverting the CLR test is more difficult, since both the LR statistic and a critical value are complicated functions of $\beta_0$. I find a very fast way to numerically invert the CLR test without using grid testing. I also characterize all possible forms of the CLR confidence region.

The third main result of this paper is a proof of asymptotic validity of the CLR confidence set. Moreira (2003) showed that if the reduced form errors are normally distributed with zero mean and known covariance matrix, then the CLR test is similar, and the CLR confidence set has exact coverage. Andrews, Moreira and Stock (2006) showed that without these assumptions a feasible version of the CLR test has asymptotically correct rejection rates both in weak instrument asymptotics and in strong instrument (classical) asymptotics. I add to their argument by proving that a feasible version of the CLR has asymptotically correct coverage *uniformly* over the whole parameter space (including nuisance parameters).

The paper is organized as follows. Section 2 contains a brief overview of the model and definitions of the CLR, AR, and LM tests. Section 3 defines the circular model and establishes its relation to the linear model. It also discusses a correspondence between properties of tests and properties of confidence sets. Section 4 gives algorithms for inverting the CLR, AR and LM tests. Section 5 provides the results of simulations comparing the length of the CLR, AR, and LM confidence sets. Section 6 contains a proof of a theorem about a uniform asymptotic coverage of the CLR confidence set.

# 2 The model and notation.

In this section I introduce notation and give a brief overview of the tests used in this paper for confidence set construction. I keep the same notation as in Andrews, Moreira and Stock (2006) for the simultaneous equations model in linear coordinates and try to stay close to the notation of Chamberlain (2005) for the model written in spherical coordinates (the circular model).

We start with a model containing structural and reduced form equations with a single endogenous regressor:

$$y_1 = y_2\beta + X\gamma_1 + u; \tag{1}$$

$$y_2 = Z\pi + X\xi + v_2. \tag{2}$$

Vectors $y_1$ and $y_2$ are $n \times 1$ endogenous variables, $X$ is $n \times p$ matrix of exogenous regressors, $Z$ is $n \times k$ matrix of instrumental variables, $\beta$ is the coefficient of interest. To make linear and circular models equivalent I assume that $\beta \in \mathbb{R}\bigcup\{\infty\}$. There are also some additional unknown parameters $\gamma_1, \xi \in \mathbb{R}^p$ and $\pi \in \mathbb{R}^k$. The $n \times 2$ matrix of errors $[u, v_2]$ consists of independent identically distributed ($i.i.d.$) rows, and each row is normally distributed with mean zero and a non-singular covariance matrix.

Without loss of generality, I assume that $Z'X = 0$. If the orthogonality condition $Z'X = 0$ is not satisfied, one can change variables by considering $\widetilde{Z} = (I - X(X'X)^{-1}X')Z$ instead of initial instruments. This will change the nuisance coefficient $\xi$ to $\widetilde{\xi} = \xi + (X'X)^{-1}X'Z\pi$.

I also consider a system of two reduced form equations obtained by substituting equation (2) into equation (1):

$$y_1 = Z\pi\beta + X\gamma + v_1; \tag{3}$$

$$y_2 = Z\pi + X\xi + v_2,$$

where

$$\gamma = \gamma_1 + \xi\beta; \quad v_1 = u + \beta v_2.$$

The reduced form errors are assumed to be i.i.d. normal with zero mean and positive-definite covariance matrix $\Omega$. Assume $\Omega$ to be known. The last two assumptions will be relaxed in Section 6.

It is well-known that all optimal inference procedures depend on the data only through sufficient statistics. So, without loss of generality, we can concentrate our attention on a set of sufficient statistics for coefficients $(\beta, \pi)$:

$$\zeta = (\Omega^{-1/2} \otimes (Z'Z)^{-1/2}Z') \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}.$$

Using these sufficient statistics, the simultaneous equations model (1) and (2) is reduced to the following which I will call a linear model:

$$\zeta \sim N((\Omega^{-1/2}a) \otimes ((Z'Z)^{1/2}\pi), I_{2k}), \tag{4}$$

where $a = (\beta, 1)'$.

I also consider what I call a circular model, which is a re-parametrization of linear model (4) in spherical coordinates. Following Chamberlain (2005), let $S^i = \{x \in \mathbb{R}^{i+1} : \|x\| = 1\}$ be an $i$-dimensional sphere in $\mathbb{R}^{i+1}$. Two elements $x_1$ and $x_2 \in S^1$ are equivalent if $x_1 = x_2$ or $x_1 = -x_2$. Let $S^1_+$ be the space of equivalence classes. Define vectors $\phi = \Omega^{-1/2}a/\|\Omega^{-1/2}a\| \in S^1_+$, and $\omega = (Z'Z)^{1/2}\pi/\|(Z'Z)^{1/2}\pi\| \in S^{k-1}$ and a real number $\rho = \|\Omega^{-1/2}a\| \cdot \|(Z'Z)^{1/2}\pi\|$. Then the circular model is given by

$$\zeta \sim N(\rho\phi \otimes \omega, I_{2k}). \tag{5}$$

The vector $\phi$ has the same direction as the vector $\Omega^{-1/2}(\beta, 1)'$, but the former is normalized to the unit length. Since, $\Omega$ is known, there is one-to-one correspondence between $\beta \in \mathbb{R}\bigcup\{\infty\}$ and $\phi \in S^1_+$. As a result, all inferences about $\phi$ can be translated into inferences about $\beta$ and vice versa. The one-dimensional parameter $\rho$ characterizes the strength of instruments.

Let us reshape the $2k \times 1$- vector $\zeta = (\zeta_1', \zeta_2')'$ into the $k \times 2$-matrix $D(\zeta) = (\zeta_1, \zeta_2)$ and define the $2\times2$-matrix $A(\zeta) = D'(\zeta)D(\zeta)$, which is the maximal invariant statistic for a group of orthogonal transformations as discussed in Section 3.2. Also consider the $2 \times 2$- matrix $Q(\zeta, \beta) = J'A(\zeta)J$, where $J = \left[\frac{\Omega^{1/2}b}{\|\Omega^{1/2}b\|}, \frac{\Omega^{-1/2}a}{\|\Omega^{-1/2}a\|}\right]$ is a $2 \times 2$ matrix, and $b = (1, -\beta)'$. Note that $J = [\phi^\perp, \phi]$, where $\phi^\perp$ is orthogonal to $\phi$: $\phi'\phi^\perp = 0$. Properties of matrix $Q(\zeta, \beta)$ are discussed in Andrews, Moreira and Stock (2006). In particular, if $\beta_0$ is the true value of the coefficient of interest, then $Q_{11}(\zeta, \beta_0)$, the upper left element of $Q(\zeta, \beta_0)$, is $\chi^2_k$-distributed. The lower right element, $Q_{22}(\zeta, \beta_0)$, has a non-central $\chi^2$-distribution with non-centrality parameter that depends on the strength of the instruments. The distribution of the off-diagonal element $Q_{12}(\zeta, \beta_0)$ can be found in Andrews, Moreira and Stock (2006).

This paper considers three tests: the Anderson - Rubin (1949) AR test, the LM test proposed by Kleibergen (2002) and Moreira (2002), and Moreira's (2003) CLR test. I define each of them below for a linear model. The corresponding definitions for a circular model are obvious. All three tests have the exact size $\alpha$ independent of the strength of instruments.

The AR test rejects the null $H_0 : \beta = \beta_0$ if the statistic

$$AR(\beta_0) = \frac{Q_{11}(\zeta, \beta_0)}{k}$$

exceeds the $(1 - \alpha)$- quantile of a $\chi^2_k$ distribution.

The LM test accepts the null if the statistic

$$LM(\beta_0) = \frac{Q^2_{12}(\zeta, \beta_0)}{Q_{22}(\zeta, \beta_0)}$$

is less than the $(1 - \alpha)$- quantile of a $\chi^2$ distribution with 1 degree of freedom.

The CLR test is based on the conditional approach proposed by Moreira (2003). He suggested a whole class of tests using critical values that are functions of the data.

The CLR test uses the LR statistic:

$$LR = \frac{1}{2}\left(Q_{11} - Q_{22} + \sqrt{(Q_{11} + Q_{22})^2 - 4(Q_{11}Q_{22} - Q_{12}^2)}\right)$$

and critical value $m_\alpha(Q_{22})$ which is a function of $Q_{22}$. For every $\alpha$, the critical value function, $m_\alpha(q_{22})$, is chosen in such a way that the conditional probability of the LR statistic exceeding $m_\alpha(q_{22})$ given that $Q_{22} = q_{22}$ equals $\alpha$:

$$P\left\{LR > m_\alpha(q_{22}) | Q_{22} = q_{22}\right\} = \alpha.$$

The CLR test accepts the null $H_0 : \beta = \beta_0$ if $LR(\beta_0) < m_\alpha(Q_{22}(\beta_0))$.

# 3 Relation between properties of a test and properties of a confidence set.

This section describes how the properties of tests are translated into properties of the corresponding confidence sets. Let $\zeta$ be a random variable satisfying a linear model (4) (or a circular model (5)). I intend to construct a confidence set for parameter $\beta$ (for parameter $\phi$) which is only a part of the parameter vector $\theta = (\beta, \pi)$ ($\theta = (\phi, \omega, \rho)$).

**Definition 1** *A set $C(\zeta)$ is a confidence set for $\beta$ at confidence level $1 - \alpha$ if for all values of $\beta$ and $\pi$*

$$P_{\beta,\pi}\{\beta \in C(\zeta)\} \geq 1 - \alpha. \tag{6}$$

According to Lehmann (1986, p.90), there is a one-to-one correspondence between testing a series of hypotheses of the form $H_0 : \beta = \beta_0$ and constructing confidence sets for $\beta$. In particular, if $C(\zeta)$ is a confidence set at confidence level $1 - \alpha$, then a test that accepts $H_0 : \beta = \beta_0$ if and only if $\beta_0 \in C(\zeta)$, is an $\alpha$-level test. And, vice versa, if $A_{\beta_0}$ is an acceptance region for testing $\beta_0$, then $C(\xi) = \{\beta_0 : \xi \in A_{\beta_0}\}$ is a confidence set. A confidence set is similar if statement (6) holds with equality. Similar tests correspond to similar confidence sets and vice versa.

### 3.1 Power vs. accuracy and expected length.

Accuracy of a confidence set is the ability to not cover false values of the parameter of interest. A uniformly most accurate (UMA) confidence set maximizes for each false value the probability of not including it. A UMA confidence set corresponds to a uniformly most powerful (UMP) test and vice versa.

Practitioners are usually more interested in another criterion, the expected length. According to Pratt's theorem (1961), the expected length of a confidence set (if it is finite) equals the integral over false values of the probability each false value is included. In fact, the statement is more general: "length" can be treated as a length with respect to any measure. Namely,

$$E_{\beta_0,\pi}\int_{\beta\in C(\zeta)}\mu(d\beta)=\int_{-\infty}^{\infty}P_{\beta_0,\pi}\{\beta\in C(\zeta)\}\mu(d\beta),$$

for any measure $\mu$ as long as both sides of the equality are finite. As a consequence, a UMA set has the shortest expected length as long as the expected length is finite.

Andrews, Moreira and Stock (2006) show that the CLR test is nearly a UMP test in a class of two-sided invariant similar tests. The invariance here is an invariance with respect to orthogonal transformations $O_k$ of instruments, defined below. This result suggests that the CLR confidence set might possess some optimality properties with respect to the length. However, Pratt's theorem cannot be applied directly. First, a confidence set with correct coverage in the case of weak instruments must be infinite with positive probability. As a result, the expected length of such an interval is infinite. Second, the CLR does not maximize power at every point; rather, it nearly maximizes the average power at two points lying on different sides of the true value. This is the ways of imposing two-sidedness condition. The location of the points depend on each other, but they are not symmetric in the linear sense. For the model written in spherical coordinates, I prove that the CLR confidence set for parameter $\phi$ will have nearly shortest expected arc length.

## 3.2 Invariance with respect to $O_k$.

Consider a group of orthogonal transformations $O_k$ on the sample space:

$$O_k = \left\{ g_F : g_F(\zeta) = \begin{pmatrix} F\zeta_1 \\ F\zeta_2 \end{pmatrix} = (I_2 \otimes F)\zeta; F \textit{ is } k \times k \textit{ orthogonal matrix} \right\}.$$

The corresponding group of transformations on the parameter space of a linear model does not change the value of $\beta$, but it does rotate the nuisance parameter $\pi$:

$$O_k^l = \{g_F^l : g_F^l(\beta, \pi) = (\beta, (Z'Z)^{-1/2} F (Z'Z)^{1/2}\pi); F \textit{ is } k \times k \textit{ orthogonal matrix}\}.$$

For a circular model (the model written in spherical coordinates), the corresponding group of transformations on the parameter space is:

$$O_k^c = \{g_F^c : g_F^c(\phi, \omega, \rho) = (\phi, F\omega, \rho); F \textit{ is } k \times k \textit{ orthogonal matrix}\}.$$

A confidence set $C(\zeta)$ for $\beta$ (for $\phi$) is invariant with respect to the group of transformations $O_k$ if $C(\zeta) = C(g_F(\zeta))$ for all $g_F \in O_k$. Invariant tests correspond to invariant confidence sets. Andrews, Moreira and Stock(2006) showed that $A(\zeta)$ is maximal invariant for the group of transformations $O_k$. It means that confidence sets (linear and circular) invariant with respect to $O_k$ can depend on $\zeta$ only through statistics $A(\zeta)$. That is, for any $O_k$-invariant confidence set $C(\zeta)$ there is a function $f$ such that:

$$C(\zeta) = \{\phi_0 : F(\phi_0, A(\zeta)) \geq 0\} = \{\phi_0 : f(\phi_0, Q(\zeta, \phi_0)) \geq 0\}.$$

If we restrict our attention to decision rules that are invariant with respect to $O_k$, then the risks for invariant loss functions (for example, rejection rates and power for tests; coverage probability, accuracy, and expected length for sets) depend only on a lower-dimensional parameter, $\left(\beta, \lambda = \frac{\pi' Z' Z \pi}{k}\right)$ in a linear model and $(\phi, \rho)$ in a circular model.

## 3.3 Two-sided tests and symmetry in a circular model.

Andrews, Moreira and Stock (2006) discuss different ways of constructing 2-sided power envelopes. One approach is to maximize the average power at two alternatives on different sides of the null by choosing these alternatives in such a way that the maximizer is an asymptotically efficient test under strong instruments asymptotics. Consider some value of the null, $\beta_0$, and an alternative, $(\beta^*, \lambda^*)$. Then there is another alternative, $(\beta_2^*, \lambda_2^*)$, on the other side of $\beta_0$ such that a test maximizing average power at these two points is asymptotically efficient (formula for $(\beta_2^*, \lambda_2^*)$ is given in Andrews, Moreira and Stock (2006)).

In general, there is no linear symmetry between alternatives: $\beta^* - \beta_0 \neq \beta_0 - \beta_2^*$. However, the way of imposing two-sidedness stated above gives symmetry of alternatives in a circular model. Namely, let $(\phi^*, \rho^*)$ correspond to $(\beta^*, \lambda^*)$ and $(\phi_2^*, \rho_2^*)$ correspond to $(\beta_2^*, \lambda_2^*)$. Then $\rho^* = \rho_2^*$ and $\phi^*$ is symmetric (on the circle) to $\phi_2^*$ with respect to $\phi_0$; that is, $\phi_0'\phi^* = \phi_0'\phi_2^*$ and $(\phi_0^\perp)'\phi^* = -(\phi_0^\perp)'\phi_2^*$.

An equivalent way of imposing the 2-sidedness is imposing a sign-invariance condition. This condition is specific to the null value $\phi_0$ tested. Consider a statistic $S = D(\zeta)\phi_0^\perp$. If $\phi_0$ is the true value, then $S$ has a $k$-dimensional normal distribution with zero mean and identity covariance matrix, otherwise the mean of $S$ is nonzero. Consider a group of transformations on the sample space which contains two transformations: $S \mapsto -S$ and $S \mapsto S$. One can check that the corresponding group of transformations on the parameter space consists of two transformations: $\phi^* \mapsto \phi_2^*$ and $\phi^* \mapsto \phi^*$. The null hypothesis $H_0 : \phi = \phi_0$ is invariant to the group of sign transformations. Let a vector $|Q| = (Q_{11}, |Q_{12}|, Q_{22})$ contain the absolute values of elements of $Q(\zeta, \phi_0)$. An $O_k$-invariant test for testing $H_0 : \phi = \phi_0$ is invariant to the group of sign transformations if it depends on $|Q(\zeta, \phi_0)|$ only. I call a confidence set $C(\zeta)$ symmetric if

$$C(\zeta) = \{\phi_0 : f(\phi_0, |Q(\zeta, \phi_0)|) \geq 0\}.$$

By applying Pratt's theorem to the result of Andrews, Moreira and Stock (2006), I receive the following statement:

**Lemma 1** *In an IV model with a single endogenous regressor (1), (2) with homoscedastic normal error terms with known covariance matrix $\Omega$, the CLR confidence set has nearly uniformly shortest expected arc length among similar symmetric $O_k$-*

*invariant confidence sets for $\phi$, where $\phi$ is the parameter of interest in the circular formulation of the model (5).*

### 3.4 Invariance with respect to $O_2$.

Another type of invariance introduced in Chamberlain (2005) is invariance with respect to rotations of vector $(y_1, y_2)'$. This type of invariance is quite cumbersome to deal with in a linear model, but it is very natural in spherical coordinates.

Let me consider a group of transformations on the sample space:

$$O_2 = \{G_F : G_F(\zeta) = (F \otimes I_k)\zeta;\ F \textit{ is } 2 \times 2 \textit{ orthogonal matrix}\}.$$

The corresponding group of transformations in the parameter space of a circular model is a group of rotations of vector $\phi$:

$$O_2^c = \{G_F^c : G_F^c(\phi, \omega, \rho) = (F\phi, \omega, \rho); F \textit{ is } 2 \times 2 \textit{ orthogonal matrix}\}.$$

The confidence set $C^c(\zeta)$ for $\phi$ in a circular model (5) is invariant with respect to the group of transformations $O_2$ if $C^c(G_F(\zeta)) = F(C^c(\zeta))$ for all $G_F \in O_2$, here $F(C) = \{\phi : F^{-1}\phi \in C\}$ stays for the corresponding rotation of the set over the unit circle.

**Lemma 2** *A confidence set $C(\zeta)$ for $\phi$ in a circular model (5) is invariant with respect to group $O_2 \times O_k$ if and only if there exists a function $f$ such that*

$$C(\zeta) = \{\phi : f(Q(\zeta, \phi)) \geq 0\}.$$

**Corollary 1** *Confidence sets obtained by inverting the CLR, AR, and LM tests are invariant with respect to $O_2 \times O_k$.*

**Corollary 2** *The expected arc length of confidence sets for $\phi$ obtained by inverting the CLR, AR, and LM tests depend only on $\rho$ and $k$.*

# 4 Algorithms for constructing CLR, AR and LM confidence sets.

In this section I describe an easy way to invert the CLR, AR, and LM tests and find an analytical description of the three confidence sets. I should emphasize that the general description of the AR sets as well as the algorithm for finding them is well known. The descriptions of the other two sets as well as algorithms for finding them are new.

## 4.1 Confidence sets based on the CLR test.

This section describes an algorithm for constructing a confidence set for the coefficient on the single endogenous regressor, $\beta$, by inverting the CLR test.

One way to invert the CLR test is to perform a series of tests $H_0 : \beta = \beta_0$ over a fine grid of $\beta_0$ using the CLR testing procedure. However, such an algorithm is numerically cumbersome. The main difficulty with finding an analytically tractable way of inverting the CLR test is that both the test statistic(LR) and the critical value function $m_\alpha(Q_t)$ depend not only on the data, but on the null value of the parameter $\beta_0$. In both cases the dependence on $\beta_0$ is quite complicated. I transform both sides to make the dependence simpler.

Let $M(\beta_0) = maxeval(Q(\beta_0))$ be the maximum eigenvalue of the matrix $Q(\beta_0)$, then

$$M = \frac{1}{2}\left(Q_{11} + Q_{22} + \sqrt{(Q_{11} + Q_{22})^2 - 4(Q_{11}Q_{22} - Q_{12}^2)}\right).$$

As a result, the LR statistic can be written as

$$LR(\beta_0) = M(\zeta, \beta_0) - Q_{22}(\zeta, \beta_0).$$

Recall that $Q(\zeta, \beta_0) = J'A(\zeta)J$. Since $J'J = I_2$, $M = maxeval(Q(\zeta, \beta_0)) = maxeval(A(\zeta))$ does not depend on the null value $\beta_0$. That is, $LR(\beta_0) = M(\zeta) - Q_{22}(\zeta, \beta_0)$.

The confidence set based on the CLR test is the set

$$C_\alpha^{CLR}(\zeta) = \{\beta_0 : M(\zeta) - Q_{22}(\zeta, \beta_0) < m_\alpha\left(Q_{22}(\zeta, \beta_0)\right)\}$$

$$= \{\beta_0 : M(\zeta) < Q_{22}(\zeta, \beta_0) + m_\alpha (Q_{22}(\zeta, \beta_0))\},$$

where $m_\alpha(q_{22})$ is the critical value function for the CLR test.

**Lemma 3** *For any $\alpha \in (0, 1)$, the function $f(q_{22}) = q_{22} + m_\alpha(q_{22})$ is strictly increasing. There exists a strictly increasing inverse function $f^{-1}$.*

It follows from Lemma 3 that the CLR confidence set is

$$C_\alpha^{CLR}(\zeta) = \{\beta_0 : Q_{22}(\zeta, \beta_0) > C(\zeta)\},$$

where $C(\zeta) = f^{-1}(M)$ depends on the data only, but not on the null value $\beta_0$. Since

$$Q_{22}(\zeta, \beta_0) = \frac{a_0'\Omega^{-1/2}A(\zeta)\Omega^{-1/2}a_0}{a_0'\Omega^{-1}a_0},$$

the problem of finding the CLR confidence set can be reduced to solving an ordinary quadratic inequality:

$$a_0' \left(\Omega^{-1/2}A(\zeta)\Omega^{-1/2} - C\Omega^{-1}\right) a_0 > 0.$$

**Theorem 1** *Assume that we have model (1) and (2) written in linear coordinates. Then the CLR confidence region $C_\alpha^{CLR}(\zeta)$ can have one of three possible forms:*
*1) a finite interval $C_\alpha^{CLR}(\zeta) = (x_1, x_2)$;*
*2) a union of two infinite intervals $C_\alpha^{CLR}(\zeta) = (-\infty, x_1) \cup (x_2, +\infty)$;*
*3) the whole line $C_\alpha^{CLR}(\zeta) = (-\infty, +\infty)$.*

The form of the interval might seem to be a little bit strange. However, one should keep in mind that the interval with correct coverage under the weak instrument assumptions should be infinite with positive probability. For a model written in spherical coordinates one has:

$$C_\alpha^{CLR}(\zeta) = \{\phi_0 : \phi_0'(A(\zeta) - C)\phi_0 > 0\}.$$

The second case described in the theorem corresponds to the arc containing point $\phi = \frac{\Omega^{-1/2}e_2}{\sqrt{e_2'\Omega^{-1}e_2}}$, where $e_2 = (0, 1)'$.

***More on technical implementation.*** I suggest a numerically simple way of finding the inverse function of $f$. Let $C = f^{-1}(M)$, that is, $m_\alpha(C) + C = M$, or

$m_\alpha(C) = M - C$. Andrews, Moreira and Stock (2007) defined the conditional p-value of the CLR test as the following function:

$$p(lr; q_{22}) = P\{LR > lr | Q_{22} = q_{22}\}.$$

An alternative way to perform CLR test is to compare $p(LR; Q_{22})$ with the significance level $\alpha$. Andrews, Moreira and Stock (2007) wrote the function $p(lr; q_{22})$ as an integral of an analytic function and suggested a numerical way of computing it. It is easy to see that finding $C$ for any given $M$ is equivalent to solving an equation $p(M - C; C) = \alpha$. We now have:

**Lemma 4** *For any fixed $M > 0$ the function $l(C) = p(M - C; C)$ is monotonic in $C$ for $0 < C < M$.*

Since $l(C)$ is monotonic, and $C$ belongs to an interval $[0, M]$, I can find $C$ such that $l(C) = \alpha$ using a binary search algorithm. Given that the calculation of $p(lr; q_{22})$ is fast, finding $C$ with any reasonable accuracy will be fast as well. Mikusheva and Poi (2006) describe a Stata software program implementing the suggested procedure.

## 4.2 AR confidence set.

The results of this subsection are not new; I summarize them for the sake of completeness. The idea of inverting the AR test goes back to Anderson and Rubin (1949). A similar argument is also used in Dufour and Taamouti (2005)

According to its definition, the AR confidence set is a set $C_\alpha^{AR}(\zeta) = \{\beta_0 : Q_{11}(\zeta, \beta_0) < k\chi^2_{\alpha,k}\}$, which can be found by solving a quadratic inequality.

**Lemma 5** *Assume that we have model (1) and (2). Then the AR confidence region $C_\alpha^{AR}(\zeta)$ can have one of four possible forms:*
*1) a finite interval $C_\alpha^{AR}(\zeta) = (x_1, x_2)$;*
*2) a union of two infinite intervals $C_\alpha^{AR}(\zeta) = (-\infty, x_1) \cup (x_2, +\infty)$;*
*3) the whole line $C_\alpha^{AR}(\zeta) = (-\infty, +\infty)$;*
*4) an empty set $C_\alpha^{AR}(\zeta) = \emptyset$.*

### 4.3 The LM confidence set.

Inverting the LM test is easier than inverting the CLR test because the LM statistic is a relatively simple function of $\beta_0$ and critical values are fixed. Finding the LM region is equivalent to solving an inequality of the fourth power, which always has a solution in radicals due to Cardano's formula. Solving an arbitrary polynomial inequality of the fourth order can be cumbersome. I rewrite the LM statistic in a way that allows us to solve two quadratic inequalities instead. The new formula also reveals new peculiarities of the LM test.

Let $N = mineval(Q)$ be the minimum eigenvalue of the matrix $Q$. The value of $N$ depends on the data only, but not on the null value tested. As shown before

$$\frac{1}{2}\left(Q_{11} - Q_{22} + \sqrt{(Q_{11} + Q_{22})^2 - 4(Q_{11}Q_{22} - Q_{12}^2)}\right) = M - Q_{22}. \qquad (7)$$

Similarly,

$$\frac{1}{2}\left(Q_{11} - Q_{22} - \sqrt{(Q_{11} + Q_{22})^2 - 4(Q_{11}Q_{22} - Q_{12}^2)}\right) = N - Q_{22}. \qquad (8)$$

By multiplying (7) and (8) I obtain:

$$Q_{12}^2(\beta_0) = -(M - Q_{22}(\beta_0))(N - Q_{22}(\beta_0)).$$

As a result, the LM statistic has the following form:

$$LM(\beta_0) = -\frac{(M - Q_{22}(\beta_0))(N - Q_{22}(\beta_0))}{Q_{22}(\beta_0)}.$$

The LM confidence region is a set

$$C_\alpha^{LM}(\zeta) = \left\{\beta_0 : -\frac{(M(\zeta) - Q_{22}(\zeta,\beta_0))(N(\zeta) - Q_{22}(\zeta,\beta_0))}{Q_{22}(\zeta,\beta_0)} < \chi_{1,\alpha}^2\right\}.$$

Obtaining the LM confidence set can be done in two steps. As the first step, one solves for the values of $Q_{22}(\zeta,\beta_0)$ satisfying the inequality above, which is an ordinary quadratic inequality with respect to $Q_{22}$. Then, one finds the LM confidence set for $\beta_0$ by solving inequalities of the form $\{\beta_0 : Q_{22}(\zeta,\beta_0) < s_1\} \cup \{\beta_0 : Q_{22}(\zeta,\beta_0) > s_2\}$.

**Theorem 2** *Assume that we have model (1) and (2) with $k > 1$. Then the LM confidence region $C_\alpha^{LM}(\zeta)$ can have one of three possible forms:*

*1) a union of two finite intervals* $C_\alpha^{LM}(\zeta) = (x_1, x_2) \cup (x_3, x_4)$*;*

*2) a union of two infinite intervals and one finite interval*

$C_\alpha^{LM}(\zeta) = (-\infty, x_1) \cup (x_2, x_3) \cup (x_4, +\infty)$*;*

*3) the whole line* $C_\alpha^{LM}(\zeta) = (-\infty, +\infty)$*.*

The LM confidence sets for $\beta$ in general correspond to two arcs on the circle in spherical coordinates. Case 2) takes place when one of the arcs covers the point $\phi = \frac{\Omega^{-1/2} e_2}{\sqrt{e_2' \Omega^{-1} e_2}}$.

## 4.4 Comparison of the CLR, AR and LM confidence sets.

There are several observations one can make based on the above descriptions of the CLR, AR, and LM confidence sets.

First, all three confidence sets can be infinite, and even equal to the whole line. A good confidence set is supposed to correctly describe the measure of uncertainty about the parameter contained in the data. Infinite confidence sets appear mainly when instruments are weak. In these cases, we have little or no information about the parameter of interest, which is correctly pointed out by these confidence sets. The confidence sets might be infinite but not equal to the whole line (two rays - for AR and the CLR, or an interval and two rays for the LM). One should interpret these as cases with very limited information where, nevertheless, one can reject some values of the parameter.

Second, the LM confidence set has a more complicated structure than the AR and CLR sets. In general, the LM set corresponds to two arcs on the unit circle, whereas the AR and CLR correspond to one arc. This makes the LM sets more difficult to explain in practice. I will discuss this point more in the next subsection.

My third observation is that the AR confidence set is empty with non-zero probability. This is due to the fact that the AR test rejects the null not only when $\beta_0$ seems to be different from the true value of the parameter, but also when the exclusion restrictions for the IV model seem to be unreliable. When the AR confidence set is empty, it means that the data rejects the model. This is not a problem from a theoretical point of view since false rejections happen in less than 5% of cases (significance level). However, receiving an empty confidence set can be quite confusing for empirical researchers.

Fourth, there is no strict order among the length of intervals valid in all realizations. Despite the fact that the CLR test possesses better power properties than the AR test, one cannot claim that an interval produced by the CLR test is always shorter than one produced by the AR test. More than that, it is possible that AR set is empty while the CLR set is the whole line. This would happen, if $N > \chi^2_{k,\alpha}$, and the difference between two eigenvalues of the matrix $Q$ is small, in particular, if $f^{-1}(M) < N$.

## 4.5 Point estimates.

This section points out one peculiarity of LM confidence sets, namely that it concentrated around two points, one of which is "wrong".

For each test, one may find parameter value(s) that would be the last to disappear from the corresponding confidence set, or in other words, the limit of a confidence set when the confidence level decreases. This is equivalent to finding the value of $\beta_0$ which maximizes the p-value of the AR and LM tests and the conditional p-value of the CLR test. This idea was previously suggested in Dufour et al. (2005).

**Lemma 6** *Assume that we have model (1) and (2) with iid error terms $(u_i, v_{2,i})$ that are normally distributed with zero mean and non-singular covariance matrix. Let $\widehat{\beta}_{LIML}$ be the Limited Information Likelihood Maximum (LIML) estimator of $\beta$. Let us also introduce a statistic $\widetilde{\beta}$, such that $\widetilde{\beta} = \arg\min_{\beta_0} Q_{22}(\zeta, \beta_0)$. Then*

*1) $\widehat{\beta}_{LIML}$ is the maximizer of both the p-value for AR test and the conditional p-value for the CLR test. Maximum of the p-value for the LM test is achieved at two points $\widehat{\beta}_{LIML}$ and $\widetilde{\beta}$:*

$$\widehat{\beta}_{LIML} = \arg\max_{\beta_0} P\{\chi^2_k > AR(\zeta, \beta_0)\} = \arg\max_{\beta_0} p(LR(\zeta, \beta_0); Q_{22}(\zeta, \beta_0));$$

$$\{\widehat{\beta}_{LIML}, \widetilde{\beta}\} = \arg\max_{\beta_0} P\{\chi^2_1 > LM(\zeta, \beta_0)\}.$$

*2) $\widetilde{\beta}$ is the minimizer of both the p-value for AR test and the conditional p-value for the CLR test*

$$\widetilde{\beta} = \arg\min_{\beta_0} P\{\chi_k^2 > \frac{Q_{11}(\zeta, \beta_0)}{k}\} = \arg\min_{\beta_0} p(LR(\zeta, \beta_0); Q_{22}(\zeta, \beta_0)).$$

A part of Lemma 6 is known; in particular, Moreira (2002, 2003) noted that the LIML always belongs to the LM and CLR confidence sets, and the LM statistics has two zeros.

The p-value of the LM test reaches its maximum at two points, the LIML and $\widetilde{\beta}$. It is interesting to notice that $\widetilde{\beta}$ is the point where the conditional p-value of the CLR test achieves its minimum (the "worst" point from the standpoint of AR and the CLR)! The non-desirable point $\widetilde{\beta}$, and a neighborhood around it always belong to the LM confidence set (remember that the LM set corresponds to two arcs in a circular model). This observation can be treated as an argument against using the LM test in practice.

# 5 Simulations.

According to Andrews, Moreira and Stock (2006), the CLR test is nearly optimal in the class of two-sided similar tests that are invariant to orthogonal transformations. It has higher power than the AR and LM tests for a wide range of parameters. A more powerful test tends to produce a shorter confidence set. As I showed in Section 3, the CLR confidence set has nearly shortest expected arc length among similar symmetric $O_k$ - invariant confidence sets. In this section, I assess the magnitude of the differences among the expected arc length of the three confidence sets. I also compare lengths of different confidence sets in linear coordinates.

I start with comparing the expected arc length of the CLR, AR and LM confidence sets. Recall from Section 3 that all three confidence sets are $O_2 \times O_k$ - invariant. As a result, their expected lengths depend only on the number of instruments $k$ and a parameter $\rho$, which characterizes the strength of instruments. I compute the expected length using simulations for $k = 2, 3, 5, 10$ and for $\rho$ ranging from 0.5 to 10 with a step of 0.5. All results are based on 1000 simulations. The results of these simulations are given in Figure 1. The expected arc length of the CLR set is always smaller than that of the AR confidence set, in accordance with the results from Section 3. The difference between the expected arc lengths of the CLR and AR confidence sets, however, is relatively small. Both sets significantly outperform the LM confidence set when the number of instruments is big.

[Figure 1 goes here.]

Although the CLR confidence set has the nearly shortest arc length, few practitioners value this property; instead, most prefer to have a short confidence set in linear coordinates. That is why I compare the linear length of confidence sets. One of the problems, though, is that valid confidence sets are infinite with a positive probability, and as a result, the expected length is infinite. I do two types of experiments: 1) I simulate the *distribution* of confidence set length for different tests in a linear model; 2) I find the average linear length of sets *over a fixed bounded interval*; that is, the expected length of the intersection of a confidence set with a fixed interval.

I check whether the distributions of the length of the AR and LM confidence sets first order stochastically dominate the distribution of the CLR confidence set. By applying a linear transformation to model (1) and (2), one can always assume that the true value of $\beta$ equals zero and $\Omega = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$. The distributions of lengths of the CLR, AR and LM confidence sets depend on the number of instruments $k$, the strength of instruments $\lambda$ $(= \frac{1}{k}\pi' Z' Z \pi)$ and the correlation between errors $r$.

As a base case I use the same setup ($k = 5$, $\lambda = 8$, $r = 0$) as in Andrews, Moreira and Stock (2006). I also compute the results for $k = 2, 3, 5, 10$; $\lambda = 1, 2, 4, 8$; $r = 0, 0.2, 0.5, 0.95$. Coverage probability for all sets is 95%. Representative results are reported in Figure 2 and Table 1.

[Figure 2 goes here.]

[Table 1 goes here.]

Several conclusions can be made. First, the distribution of length of the LM confidence set first order dominates the one of the CLR confidence set. This result is robust over the range of parameters I checked. This shows the relative inaccuracy of the LM confidence set. Based on my simulation results, I recommend not using the LM confidence set in practice.

Second, one cannot say that the distribution of the length of the AR confidence set first order dominates that of the CLR confidence set. The opposite order does not hold either. The lack of ordering can be partially explained by the fact that the distribution of the length of the AR confidence set has a mass point at zero due to "false" rejection of the model. Furthermore, the cdfs for the length of the AR and the CLR sets cross. Crossing of the cdfs occurs before the cdfs reach the 10% level.

Another way to compare the length of different confidence sets is to compute the expected length of intersection of confidence sets with a fixed finite interval. It corresponds to a situation when a practitioner can restrict the parameter space to be a fixed finite interval. The expected length would depend on $k, \rho$, $\beta_0$ and the interval. I performed simulations for $\beta_0 = 0$ and symmetric intervals [-1,1], [-3,3], [-5,5], [-10,10], [-100,100] and [-500,500]. The results are in Figure 3. As the interval length becomes bigger (a researcher puts weaker restrictions on the parameter space) the expected lengths of the CLR and AR sets become closer to each other. One reason for that is related to the fact that the length of the AR confidence set has a mass point at zero due to false rejection of the model. For large intervals the LM set performs poorly. When a practitioner has really good prior information and can restrict the parameter space to a small interval, the CLR outperforms the two other sets.

[Figure 3 goes here.]

To summarize, in many setups the CLR confidence set looks more attractive in terms of its length. The LM confidence set possesses some unfavorable properties (such as always including $\widetilde{\beta}$) and tends to be longer. I would not recommend using the LM confidence sets in practice.

# 6 Asymptotic validity

In previous sections I assumed that the reduced form errors $[v_1, v_2]$ are i.i.d. normal with zero mean and known covariance matrix $\Omega$. Then the CLR, AR and LM testing procedures and confidence sets are exact; that is

$$\inf_{\beta_0,\pi} P_{\beta_0,\pi}\{ \textit{ hypothesis } H_0 : \beta = \beta_0 \textit{ is accepted}\} = 1 - \alpha,$$

when any of the three tests is used.

The assumption of normality can be taken away and the matrix $\Omega$ (if unknown) can be replaced with an estimator of $\Omega$ at the cost of obtaining asymptotically valid rather than exactly valid tests and confidence sets. Due to the presence of the nuisance parameter $\pi$, there are several notions of asymptotic validity. I concentrate on *uniform* asymptotic validity:

$$\lim_{n\to\infty} \inf_{\beta_0,\pi} P_{\beta_0,\pi}\{ \textit{ hypothesis } H_0 : \beta = \beta_0 \textit{ is accepted}\} = 1 - \alpha, \tag{9}$$

rather than the weaker notion of point-wise asymptotic validity(often called strong instrument asymptotics):

$$\lim_{n\to\infty} \inf_{\beta_0} P_{\beta_0,\pi}\{ \textit{ hypothesis } H_0 : \beta = \beta_0 \textit{ is accepted}\} = 1 - \alpha, \quad \forall\, \pi. \tag{10}$$

The difference between the two is that (10) allows the speed of convergence to depend on $\pi$ and be very slow for $\pi$ very close to zero (weak identification). This may lead to the test having finite sample size far from the declared asymptotic size for some part of the parameter space. For example, it is known that the standard TSLS t-test is point-wise asymptotically valid for $\pi \in \mathbb{R}^k \setminus \{0\}$, but not uniformly asymptotically valid. Another example, Andrews and Guggenberger (2005) showed that a subsampling TSLT t-test is also point-wise, but not uniformly asymptotically valid. In both cases the sizes of the mentioned tests are very misleading if $\pi$ is close to 0.

Another way to look at the differences between (9) and (10) is that (9) requires convergence of $P_{\beta_0,\pi}\{\dots\}$ along all sequences of $\pi_n$, whereas for (10) it is sufficient to check this for constant sequences $\pi_n = \pi$. Andrews, Moreira, and Stock (2006) showed that the CLR test has point-wise asymptotically correct size (10) and has an asymptotically correct size along local to unity sequences $\pi_n = C/\sqrt{n}$:

$$\lim_{n\to\infty} \inf_{\beta_0} P_{\beta_0, \pi_n = C/\sqrt{n}}\{ \textit{ hypothesis } H_0 : \beta = \beta_0 \textit{ is accepted}\} = 1 - \alpha, \qquad (11)$$

for all non-stochastic $C$. Statement (11) is called weak instrument asymptotics. Theorems about asymptotic size by Andrews and Guggenberger (2005) are suggestive that statements (10) and (11) may indicate that the CLR satisfies (9). However, to the best of my knowledge no proof of uniform asymptotic validity of the CLR exists.

Some suggestions on how to prove uniform asymptotic validity (9) were stated by Moreira(2003). I use a different approach. I prove asymptotic validity of the CLR test (confidence set) by using a strong approximation principle. The idea of the proof is to put some sample statistics with normal errors and with non-normal errors on a common probability space in such a way that they are almost surely close to each other.

I use the Representation Theorem from Pollard (1984, chapter IV.3):

**Lemma 7** *Let $\{P_n\}$ be a sequence of probability measures on a metric space weakly converging to a probability measure $P$. Let $P$ concentrate on a separable set of completely regular points. Then there exist random elements $X_n$ and $X$, where $P_n = \mathcal{L}(X_n)$, and $P = \mathcal{L}(X)$, such that $X_n \to X$ almost surely.*

## 6.1 Assumptions

Assume we have a structural IV model (1), (2), that leads to reduced form model (2), (3). Let us introduce two $n \times 2$-dimensional matrices $Y = [y_1, y_2]$ and $v = [v_1, v_2]$, and let $V_i = [v_{1,i}, v_{2,i}]$ be a $1 \times 2$-matrix for every $i = 1, ..., n$. We drop the initial assumptions that $V_i$ are iid normally distributed with known variance matrix $\Omega$. Instead, we use the following high-level assumptions, previously introduced and discussed in Andrews, Moreira, and Stock (2006).

*Assumption 1.* $\frac{1}{\sqrt{n}} vec(Z'v) \to^d N(0, \Phi)$ for some positive-definite $2k \times 2k$ matrix $\Phi$.

*Assumption 2.* $\frac{1}{n}Z'Z \to^p D$ for some positive-definite $k \times k$ matrix $D$;

*Assumption 3.* $\frac{1}{n}v'v \to^p \Omega$ for some positive-definite $2 \times 2$ matrix $\Omega$;

*Assumption 4.* $\Phi = \Omega \otimes D$.

Assumptions 2 and 3 hold under suitable conditions for a weak Law of Large Numbers. Assumption 1 is satisfied if the corresponding Central Limit Theorem holds. Assumption 4 is consistent with some form of conditional homoscedastisity. As a simple example, all assumptions are satisfied if $\{V_i, Z_i\}_{i=1}^n$ are iid with $E(V_i|Z_i) = 0$, $E(V_i'V_i|Z_i) = \Omega$ and the second moments of $V_i, Z_i$ and $V_i'Z_i$ are finite.

## 6.2 $\Omega$ is known.

For the moment, I assume that $\Omega$ is known. Let me consider the following statistics, the properties of which are discussed in Moreira (2003):

$$S = (Z'Z)^{-1/2}Z'Yb_0(b_0'\Omega b_0)^{-1/2};$$

$$T = (Z'Z)^{-1/2}Z'Y\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2},$$

where $b_0 = (1, -\beta_0)'$, $a_0 = (\beta_0, 1)'$. The CLR test can be performed by calculating the LR statistic

$$LR(S,T) = \frac{1}{2}\left(S'S - T'T + \sqrt{(S'S + T'T)^2 - 4(S'ST'T - (S'T)^2)}\right),$$

and comparing the conditional p-value function $P(S,T)$ with $\alpha$:

$$P(S,T) = p(m = LR(S,T); q_{22} = T'T).$$

I will track the dependence of the statistics on $\pi$ explicitly. Under the null one has:

$$S(\pi) = (Z'Z)^{-1/2}Z'vb_0(b_0'\Omega b_0)^{-1/2} = S;$$

$$T(\pi) = (Z'Z)^{1/2}\pi(a_0'\Omega^{-1}a_0)^{1/2} + (Z'Z)^{-1/2}Z'v\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2}.$$

The Representation Theorem applied to Assumptions 1,2 and 4 implies that there exist random variables on a common probability space such that $(Z'Z)^{-1/2}Z'v \to \xi = [\xi_1 : \xi_2]$ *a.s.*, where $vec(\xi) \sim N(0, \Omega \otimes I_k)$. Let us define a pair of variables

$$(S^*(\pi), T^*(\pi)) = (\xi b_0(b_0'\Omega b_0)^{-1/2}, (Z'Z)^{1/2}\pi(a_0'\Omega^{-1}a_0)^{1/2} + \xi\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2}).$$

Then

$$\sup_{\pi}\left(|S^*(\pi)-S(\pi)|+|T^*(\pi)-T(\pi)|\right)=|((Z'Z)^{-1/2}Z'v-\xi)b_0(b_0'\Omega b_0)^{-1/2}|+$$

$$+|((Z'Z)^{-1/2}Z'v-\xi)\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2}|\to 0 \;\; a.s. \tag{12}$$

Let $\varepsilon=[\varepsilon_1:\varepsilon_2]$ be $2\times n$ normal random variables. Assume that they are i.i.d. across rows with each row having a bivariate normal distribution with mean zero and covariance matrix $\Omega$. Let me define statistics in a model with normal errors:

$$S^N(\pi)=(Z'Z)^{-1/2}Z'\varepsilon b_0(b_0'\Omega b_0)^{-1/2};$$

$$T^N(\pi)=(Z'Z)^{1/2}\pi(a_0'\Omega^{-1}a_0)^{1/2}+(Z'Z)^{-1/2}Z'\varepsilon\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2}.$$

Then the pair of variables $(S^N(\pi),T^N(\pi))$ is distributionally equivalent to the pair $(S^*(\pi),T^*(\pi))$. Since the CLR test is exact under normality assumptions:

$$P\{P(S^N(\pi),T^N(\pi))>\alpha\}=1-\alpha \;\text{ for all }\; \pi,$$

the analogous statement for $(S^*(\pi),T^*(\pi))$ is true:

$$P\{P(S^*(\pi),T^*(\pi))>\alpha\}=1-\alpha \;\text{ for all }\; \pi. \tag{13}$$

Now I note that the conditional p-value function is a Lipschitz function with respect to $S$ and $T$.

**Lemma 8** *The function $P(S,T)$ is Lipshitz with respect to $S$ and $T$. In particular, there exists a constant $C$ such that for all $(S,T)$ and $(\widetilde{S},\widetilde{T})$*

$$|P(S,T)-P(\widetilde{S},\widetilde{T})|\le C\left(\|T-\widetilde{T}\|+\|S-\widetilde{S}\|\right).$$

Combining together equations (12), (13) and Lemma 8, I end up with the following theorem about the asymptotic validity of the CLR confidence set:

**Theorem 3** *If assumptions 1)-4) are satisfied then the CLR test is asymptotically valid:*

$$\lim_{n\to\infty}\inf_{\pi}P_\pi\left\{P(S(\pi),T(\pi))>\alpha\right\}=1-\alpha.$$

### 6.3 $\Omega$ is unknown.

I showed how to construct a strong approximation when the covariance matrix of reduced form errors $\Omega$ is known. When $\Omega$ is unknown, one can substitute for it with an estimate $\Omega_n = (n-k-p)^{-1}\widehat{V}'\widehat{V}$, where $\widehat{V} = Y - P_Z Y - P_X Y$. Andrews, Moreira, and Stock (2006) show that under Assumptions 1-4 $\widehat{\Omega}_n$ is a consistent estimate of $\Omega$, and the convergence holds uniformly with respect to $\pi$. The feasible versions of the statistics are:

$$S(\pi) = (Z'Z)^{-1/2}Z'Yb_0(b_0'\widehat{\Omega}b_0)^{-1/2} = (Z'Z)^{-1/2}Z'Vb_0(b_0'\widehat{\Omega}b_0)^{-1/2},$$

and

$$\begin{aligned} T(\pi) &= (Z'Z)^{-1/2}Z'Y\widehat{\Omega}^{-1}a_0(a_0'\widehat{\Omega}^{-1}a_0)^{-1/2} \\ &= (Z'Z)^{1/2}\pi(a_0'\widehat{\Omega}^{-1}a_0)^{1/2} + (Z'Z)^{-1/2}Z'v\widehat{\Omega}^{-1}a_0(a_0'\widehat{\Omega}^{-1}a_0)^{-1/2}. \end{aligned}$$

Let $(S^*(\pi), T^*(\pi))$ be defined as before. Then

$$\sup_\pi (|S^*(\pi) - \delta S(\pi)| + |T^*(\pi) - \delta T(\pi)|) \to 0 \ \ a.s.,$$

where $\delta = \sqrt{\frac{a_0'\Omega^{-1}a_0}{a_0'\widehat{\Omega}^{-1}a_0}}$. One can note that $\delta \to 1$ a.s., and the convergence holds uniformly with respect to $\pi$. From the Lipshitz property I have:

$$\sup_\pi |P(S^*(\pi), T^*(\pi)) - P(\delta S(\pi), \delta T(\pi))| \to 0 \quad a.s.,$$

which implies

$$\sup_\pi |p(S^*(\pi), T^*(\pi)) - p(S(\pi), T(\pi))| \to 0 \quad a.s.$$

I conclude that the CLR test is asymptotically correct.

## 7 Discussion and Conclusion.

The paper shows that CLR confidence sets possess some optimality property in a class of similar symmetric invariant confidence set. The class includes well-known AR and LM confidence sets. The paper also provides a fast algorithm for constructing CLR set and proves that the CLR has uniform asymptotic coverage under quite general assumptions.

There are some restrictive assumptions maintained in the paper, namely, normality of error terms, homoscedasticity of errors, and presence of a single endogenous regressor. Below I discuss whether these assumptions can be dropped.

Normality of error terms is needed for optimality property only, since Section 6 proves that without normality the CLR is uniformly asymptotically similar. Cattaneo, Crump and Jansson (2009) show that optimality of the CLR is lost once the normality assumption is relaxed. In particular, they show that if errors are i.i.d and belong to some class of smooth distributions, the CLR can be improved. However, since the CLR is a "nearly optimal" test for the limiting problem, according to Müller(2008), the CLR possesses asymptotically "nearly optimal" performance among "robust" tests (in a sub-class of asymptotically similar, symmetric, invariant tests). Müller (2008) defines the robustness of a test as the ability to provide the correct asymptotic size for a wide range of data generating processes satisfying very mild assumptions about weak convergence (Assumptions 1-4 in our case). Müller (2008) argues that any test with a higher asymptotic power necessarily lacks robustness.

Homoscedasticity of error terms is a restrictive assumption. Andrews, Moreira and Stock (2006) provide versions of AR, LM and CLR tests robust to heteroscedasticity. This robustness, however, comes at the cost of the CLR losing its optimality. The proof of asymptotic uniform coverage of the CLR after slight modifications remains valid.

The generalization of the results to the case of more than one endogenous regressor seems to be extremely hard. The likelihood ratio statistic in the case of $m$ endogenous regressors ($m > 1$) is a function of the smallest eigenvalue of the appropriately defined $(m+1) \times (m+1)$-dimensional matrix $Q$. In the case of $m = 2$, this leads to a formula for cubic roots, for $m > 3$ no explicit formula is available. Besides the fact that a formula for the LR statistic is cumbersome for $m = 2, 3$ and unavailable for $m > 3$, we would also have to find its conditional quantiles given the $m$-dimensional sufficient statistic for the strength of instruments. Hiller(2006) made significant progress in creating a numerically feasible procedure for performing the CLR test for $m > 1$. However, it is not obvious how one would invert the CLR test in such a situation without a grid search. To the best of my knowledge, no results on optimality of any test exists in a setting with more than one endogenous regressor.

APPENDIX: Proofs

**Proof of Lemma 1.** The proof is a direct corollary of Lemma 9, stating the claim more formally. I want to point out, that according to Andrews, Moreira and Stock (2006), the CLR is "nearly" optimal. The optimality of CLR has not been proven, instead it was shown in simulations that CLR power curve cannot be distinguished from the power envelope in all cases in a wide simulation study Andrews, Moreira and Stock undertook. The word "nearly" in the lemma above, as well as parameter $\varepsilon$ in the lemma below, stay for available simulation accuracy.

**Lemma 9** *Let $K(\beta_0;\beta^*,\lambda^*)$ be a two-sided power envelope for invariant similar tests described in section 4 of Andrews, Moreira and Stock (2006), that is,*

$$K(\beta_0;\beta^*,\lambda^*) = \max_{\widetilde{\varphi}\in\Psi}\left(E_{\beta^*,\lambda^*}\widetilde{\varphi}(Q(\beta_0)) + E_{\beta_2^*,\lambda_2^*}\widetilde{\varphi}(Q(\beta_0))\right),$$

*where $\Psi$ is a class of similar tests invariant with respect to $O_k$.*

*Let $\varphi(\beta_0,|Q(\beta_0)|)$ be a similar test for testing $H_0:\beta=\beta_0$ (or equivalently for testing $H_0:\phi=\phi_0$) invariant with respect to $O_k$ such that for some $\varepsilon>0$ we have*

$$\left(E_{\beta^*,\lambda^*}\varphi(\beta_0,|Q(\beta_0)|) + E_{\beta_2^*,\lambda_2^*}\varphi(\beta_0,|Q(\beta_0)|)\right) \geq K(\beta_0;\beta^*,\lambda^*) - \varepsilon, \quad \text{for all } \beta_0,\beta^*,\lambda^*.$$

*Let $C_\varphi$ be a confidence set for $\phi$ corresponding to $\varphi$. Then for all similar symmetric $O_k$ - invariant confidence sets $C(\zeta)$ for $\phi$, we have the following statement about the expected arc length:*

$$E_{\phi,\rho}(\text{arc length } C(\zeta)) \geq E_{\phi,\rho}(\text{arc length } C_\varphi(\zeta)) - \varepsilon Pi,$$

*where $Pi = 3.1416...$*

**Proof of Lemma 9.** For every symmetric $O_k$-invariant similar test $\widetilde{\varphi}$, its power at points $(\phi^*,\rho^*)$ and $(\phi_2^*,\rho^*)$ is the same (since power is an invariant risk function):

$$E_{\phi^*,\rho^*}\widetilde{\varphi}(|Q(\zeta,\phi_0)|) = E_{\phi_2^*,\rho^*}\widetilde{\varphi}(|Q(\zeta,\phi_0)|) \leq K(\phi_0;\phi^*,\rho^*).$$

For a similar symmetric $O_k$-invariant set $C(\zeta)$, $C(\zeta) = \{\phi_0 : f(\phi_0, |Q(\zeta, \phi_0)|) \geq 0\}$. As a result,

$$E_{\phi,\rho}(arc\ length\ C(\zeta)) = E_{\phi,\rho} \int_0^{Pi} I\{f(\phi_0, |Q(\zeta, \phi_0)|) \geq 0\} d\phi_0 =$$

$$= \int_0^{Pi} E_{\phi,\rho} \left(1 - \widetilde{\varphi}(\phi_0, |Q(\zeta, \phi_0)|)\right) d\phi_0 \geq \int_0^{Pi} \left(1 - K(\phi_0; \phi, \rho)\right) d\phi_0.$$

For the test $\varphi$ we have

$$E_{\phi,\rho}(arc\ length\ C_\varphi(\zeta)) = \int_0^{Pi} E_{\phi,\rho} \left(1 - \varphi(\phi_0, |Q(\zeta, \phi_0)|)\right) d\phi_0 \leq$$

$$\leq \int_0^{Pi} \left(1 - K(\phi_0; \phi, \rho) + \varepsilon\right) d\phi_0.$$

**Proof of Lemma 2.** Any $O_k$ - invariant confidence set can be written as $C(\zeta) = \{\phi_0 : f(\phi_0, Q(\zeta, \phi_0)) \geq 0\}$. The statement of the lemma follows from two facts: 1) $Q(G_F(\zeta), F\phi_0) = Q(\zeta, \phi_0)$ for all orthogonal $2 \times 2$ matrices $F$; 2) for any $\phi_0, \phi \in S^1_+$ there exists an orthogonal $2 \times 2$ matrix $F$ such that $\phi_0 = F\phi$.

**Proof of Lemma 3.** From Andrews, Moreira and Stock (2007) it is known that the conditional p-value $p(m; q_{22})$ has the following form

$$p(m; q_{22}) = 1 - 2 \int_0^1 P\left\{\chi^2_k < \frac{q_{22} + m}{1 + q_{22}s_2^2/m}\right\} K_4(1 - s_2^2)^{(k-3)/2} ds_2 =$$

$$= 1 - \int_0^1 F_{\chi^2_k}\left(\frac{q_{22} + m}{1 + q_{22}s_2^2/m}\right) g(s_2) ds_2,$$

where $F_{\chi^2_k}(x) = P\left\{\chi^2_k < x\right\}, g(x) = 2K_4(1 - x^2)^{(k-3)/2}$.

Let $f_{\chi^2_k}(x)$ be a derivative of $F_{\chi^2_k}(x)$. Let us also denote $h(m; q_{22}) = \frac{q_{22}+m}{1+q_{22}s_2^2/m} = m\frac{q_{22}+m}{m+q_{22}s_2^2}$; then the implicit function theorem implies that:

$$\frac{dm(q_{22})}{dq_{22}} = -\frac{\partial p(m; q_{22})}{\partial q_{22}} / \frac{\partial p(m; q_{22})}{\partial m} = -\frac{\int_0^1 f_{\chi^2_k}\left(h(m; q_{22})\right) g(s_2) \frac{\partial h(m;q_{22})}{\partial q_{22}} ds_2}{\int_0^1 f_{\chi^2_k}\left(h(m; q_{22})\right) g(s_2) \frac{\partial h(m;q_{22})}{\partial m} ds_2},$$

$$\frac{d(m(q_{22}) + q_{22})}{dq_{22}} = \frac{\int_0^1 f_{\chi^2_k}\left(h(m; q_{22})\right) g(s_2) \left(-\frac{\partial h(m;q_{22})}{\partial q_{22}} + \frac{\partial h(m;q_{22})}{\partial m}\right) ds_2}{\int_0^1 f_{\chi^2_k}\left(h(m; q_{22})\right) g(s_2) \frac{\partial h(m;q_{22})}{\partial m} ds_2}.$$

Now, we notice that

$$\frac{\partial h(m; q_{22})}{\partial q_{22}} = m\frac{(m + q_{22}s_2^2) - s_2^2(m + q_{22})}{(m + q_{22}s_2^2)^2} = \frac{m^2(1 - s_2^2)}{(m + q_{22}s_2^2)^2},$$

$$\frac{\partial h(m;q_{22})}{\partial m} = \frac{(2m+q_{22})(m+q_{22}s_2^2) - m(q_{22}+m)}{(m+q_{22}s_2^2)^2} = \frac{m^2+2q_{22}ms_2^2+q_{22}^2s_2^2}{(m+q_{22}s_2^2)^2},$$

$$-\frac{\partial h(m;q_{22})}{\partial q_{22}} + \frac{\partial h(m;q_{22})}{\partial m} = \frac{m^2+2q_{22}ms_2^2+q_{22}^2s_2^2 - m^2(1-s_2^2)}{(m+q_{22}s_2^2)^2} = \frac{m^2s_2^2+2q_{22}ms_2^2+q_{22}^2s_2^2}{(m+q_{22}s_2^2)^2}.$$

So we have that $\frac{\partial h(m;q_{22})}{\partial m} > 0$ and $-\frac{\partial h(m;q_{22})}{\partial q_{22}} + \frac{\partial h(m;q_{22})}{\partial m} > 0$. Since $f_{\chi_k^2}(h(m;q_{22}))\, g(s_2)$ is also always positive, it follows that $f'(q_{22}) > 0$.

**Proof of Theorem 1.** We know that the confidence set is a set of values $\beta_0$ such that the vector $a_0 = (\beta_0, 1)'$ satisfies the following inequality:

$$a_0' \left(\Omega^{-1/2} A(\zeta) \Omega^{-1/2} - C\Omega^{-1}\right) a_0 > 0.$$

Let

$$\mathcal{A} = \Omega^{-1/2} A(\zeta) \Omega^{-1/2} - C\Omega^{-1} = (\alpha_{i,j}), \quad D = -4det(\mathcal{A}).$$

Let $x_{1,2} = \frac{-2\alpha_{12} \pm \sqrt{D}}{2\alpha_{11}}$. There are 4 different cases depending on the signs of $D$ and $\alpha_{11}$:

1. If $\alpha_{11} < 0$ and $D < 0$ the confidence set is empty.
2. If $\alpha_{11} < 0$ and $D > 0$ then the confidence set is an interval $[x_1, x_2]$.
3. If $\alpha_{11} > 0$ and $D < 0$ then the confidence set is the whole line $(-\infty, \infty)$.
4. If $\alpha_{11} > 0$ and $D > 0$ then the confidence set is a union of two intervals $(-\infty, x_2]$ and $[x_1, \infty)$.

Moreira (2002) stated that a CLR interval always contains the LIML point estimate, and as a result, is never empty. All other cases 2-4 could be observed in practice.

**Proof of Lemma 4.**

$$l'(C) = \left( \frac{\partial p(m;q_{22})}{\partial q_{22}} - \frac{\partial p(m;q_{22})}{\partial m} \right)\Bigg|_{m=M-C,\ q_{22}=C} =$$

$$= \left( \int_0^1 f_{\chi_k^2}(h(m;q_{22}))\, g(s_2) \left( -\frac{\partial h(m;q_{22})}{\partial q_{22}} + \frac{\partial h(m;q_{22})}{\partial m} \right) ds_2 \right)\Bigg|_{m=M-C,\ q_{22}=C}.$$

We already proved that the last expression is always positive.

**Proof of Lemma 5.** Let us denote $Y = [y_1, y_2]$ and

$$\Gamma = Y'Z(Z'Z)^{-1}Z'Y - k\chi_{k,\alpha}\Omega = (\gamma_{i,j}).$$

The value of $\beta_0$ belongs to the confidence set if and only if $b_0 = (1, -\beta_0)'$ satisfies an inequality $b_0'\Gamma b_0 < 0$. Let $D_2 = det(\Gamma)$. Let $x_{1,2} = \frac{b_{12} \mp \sqrt{D_2}}{b_{22}}$. There are 4 cases:

1. If $\gamma_{22} > 0$ and $D_2 < 0$ then the confidence set is empty.

2. If $\gamma_{22} > 0$ and $D_2 > 0$ then the confidence set is an interval $[x_1, x_2]$.

3. If$\gamma_{22} < 0$ and $D_2 < 0$ then the confidence set is the full line.

4. If $\gamma_{22} < 0$ and $D_2 > 0$ then the confidence set is a union of two intervals $(-\infty, x_2] \cup [x_1, \infty)$.

**Proof of Theorem 2.** As the first step we solve for the values of $Q_{22}(\beta_0)$ satisfying the inequality

$$-\frac{(M - Q_{22}(\beta_0))(N - Q_{22}(\beta_0))}{Q_{22}(\beta_0)} < \chi^2_{1,\alpha}.$$

We have an ordinary quadratic inequality with respect to $Q_{22}$. If $D_1 = (M + N - \chi^2_{1,\alpha})^2 - 4MN \leq 0$, then there are no restrictions placed on $Q_{22}$, and the confidence region for $\beta$ is the whole line $(-\infty, \infty)$.

If $D_1 = (M + N - \chi^2_{1,\alpha})^2 - 4MN > 0$, then $Q_T \in [N, M] \setminus (s_1, s_2)$, where $s_{1,2} = \frac{M+N-\chi^2_{1,\alpha} \mp \sqrt{D_1}}{2}$.

As the second step we solve for the confidence set of $\beta_0$. The confidence set is a union of two non-intersecting confidence sets: $\{\beta_0 : Q_T(\beta_0) < s_1\} \cup \{\beta_0 : Q_T(\beta_0) > s_2\}$.

Let us denote

$$\mathcal{A}_1 = \Omega^{-1/2} A(\zeta) \Omega^{-1/2} - s_1 \Omega^{-1} = (\alpha^1_{i,j}),$$

$$\mathcal{A}_2 = \Omega^{-1/2} A(\zeta) \Omega^{-1/2} - s_2 \Omega^{-1} = (\alpha^2_{i,j}).$$

The confidence set contains $\beta_0$ if and only if $a_0'\mathcal{A}_1 a_0 < 0$ or $a_0'\mathcal{A}_2 a_0 > 0$. Since $s_1, s_2 \in (N, M)$, the quadratic equations $a_0'\mathcal{A}_1 a_0 = 0$ and $a_0'\mathcal{A}_2 a_0 = 0$ have two zeros each. Also note that since $s_1 < s_2$, then $\alpha^1_{11} > \alpha^2_{11}$. As a result, we have 3 different cases:

1. If $\alpha^1_{11} > 0$ and $\alpha^2_{11} > 0$, then the confidence set is a union of two infinite intervals and one finite interval.

2. If $\alpha^1_{11} > 0$ and $\alpha^2_{11} < 0$, then the confidence set is a union of two finite intervals.

3. If $\alpha_{11}^1 < 0$ and $\alpha_{11}^2 < 0$, then the confidence set is a union of two infinite intervals and one finite interval.

**Proof of Lemma 6.** First, we note that according to Lemma 4 we have

$argmax_{\beta_0} p(LR(\beta_0); Q_{22}(\beta_0)) = argmax_{\beta_0} p(M - Q_{22}(\beta_0); Q_{22}(\beta_0)) = argmax_{\beta_0} Q_{22}(\beta_0)$.

It is easy to see that

$$argmax_{\beta_0} P\{\chi_k^2 > Q_{11}(\beta_0)\} = argmin_{\beta_0} Q_{11}(\beta_0).$$

As a second step we prove that

$$(\beta = argmax_{\beta_0} Q_{22}(\beta_0)) \Leftrightarrow (\beta = argmin_{\beta_0} Q_{11}(\beta_0))$$

and

$$(\beta = argmin_{\beta_0} Q_{22}(\beta_0)) \Leftrightarrow (\beta = argmax_{\beta_0} Q_{11}(\beta_0))\,.$$

Let $x = \Omega^{1/2} b_0$, $y = \Omega^{-1/2} a_0$. Then $Q_{11}(\beta_0) = \frac{x' A(\zeta) x}{x'x}$, $Q_{22} = \frac{y' A(\zeta) y}{y'y}$, and $x'y = 0$; that is, $x$ and $y$ are orthogonal to each other. Because the matrix $A(\zeta)$ is positively definite, it has two eigenvectors that are orthogonal to each other. If $\beta_0 = argmax_{\beta_0} Q_{22}(\beta_0)$, then $x$ is the eigenvector of $A(\zeta)$ corresponding to the largest eigenvalue. Then $y$ is the eigenvector of $A(\zeta)$ corresponding to the smallest eigenvalue, and $\beta_0 = argmin_{\beta_0} Q_{11}(\beta_0)$. The second statement has a similar proof.

Since $\widehat{\beta}_{LIML} = argmax_{\beta_0} Q_{22}(\beta_0)$, we have that $\widehat{\beta}_{LIML}$ maximizes the p-value of the AR test and the conditional p-value of the CLR. From the definition of $\widetilde{\beta}$, we can see that it minimizes the p-value of the AR test and the conditional p-value of the CLR.

It is easy to notice that the LM statistic takes the value of 0 in two cases when $Q_{22}(\beta_0) = M$ and when $Q_{22}(\beta_0) = N$. But we know that $M = \max_{\beta_0} Q_{22}(\beta_0)$ and $N = \min_{\beta_0} Q_{22}(\beta_0)$, that is,

$$Argmax_{\beta_0} P\{\chi_1^2 > LM(\beta_0)\} = \{\widehat{\beta}_{LIML}, \widetilde{\beta}\}.$$

**Proof of Lemma 8.** I check that $sup_{t,s} \left|\frac{\partial P(s,t)}{\partial s}\right| < \infty$ and $sup_{t,s} \left|\frac{\partial P(s,t)}{\partial t}\right| < \infty$. Let $h(s_2) = \frac{q_{22}+m}{1+q_{22}s_2^2/m} = m\frac{q_{22}+m}{m+q_{22}s_2^2}$, where $m = LR(S,T)$. Then

$$\frac{\partial p(m(S,T); T'T)}{\partial S} = -2K \int_0^1 f_{\chi_k^2}(h(s_2))\,(1-s_2^2)^{(k-3)/2} \left(\frac{\partial h}{\partial m} \cdot \frac{\partial m}{\partial S}\right) ds_2.$$

Note that $h(s_2) \geq m$ for all $s_2$. Also note that

$$m \geq \frac{1}{2}\left(S'S - T'T + (S'S + T'T) - 2\sqrt{(S'ST'T - (S'T)^2)}\right) \geq S'S.$$

The pdf of a $\chi^2_k$ distribution has an exponential decay, and the term $\frac{\partial h}{\partial m} \cdot \frac{\partial m}{\partial S}$ is a polynomial with respect to $S$ and $T$. As a result, $\left|\frac{\partial P(s,t)}{\partial s}\right| \to 0$ as $s \to \infty$, and we can bound it above for $S'S > C_1 = const$. Let us consider $S'S < C_1$. It is easy to check that $\left|\frac{\partial P(s,t)}{\partial s}\right| \to const$ as $t \to \infty$. So we can choose $C_2$ such that $\left|\frac{\partial P(s,t)}{\partial s}\right|$ is bounded if $S'S < C_1$ and $T'T > C_2$. Since $\left|\frac{\partial P(s,t)}{\partial s}\right|$ is a continuous function of $s$ and $t$, it is bounded on the set $S'S < C_1, T'T < C_2$. This proves the first statement. The proof of the second one is totally analogous.

**Proof of Theorem 3.** From statement (12) and Lemma 8 we have that

$$\sup_\pi |P(S^*(\pi), T^*(\pi)) - P(S(\pi), T(\pi))| \to 0 \quad a.s.$$

Since pairs $(S^*(\pi), T^*(\pi))$ and $(S^N(\pi), T^N(\pi))$ have the same distribution, it follows that:

$$\sup_\pi P_\pi \left\{P(S(\pi), T(\pi)) \leq \alpha\right\} \leq P_\pi \left\{P(S^N(\pi), T^N(\pi)) \leq \alpha + \epsilon\right\} +$$
$$+ \sup_\pi P_\pi \left\{|P(S^*(\pi), T^*(\pi)) - P(S(\pi), T(\pi))| \leq \epsilon\right\} \leq$$
$$\leq \alpha + \epsilon + P_\pi \left\{\sup_\pi |P(S(\pi), T(\pi)) - P(S(\pi), T(\pi))| \leq \epsilon\right\} \to \alpha + \epsilon$$

The last line relies on the fact that the method is exact for a model with normal errors.

*I thank Yury Baskakov, Gary Chamberlain, Victor Chernozhukov, JB Doyle, Jerry Hausman, Whitney Newey, a co-editor, and two anonymous referees for valuable comments. I am grateful to Jim Stock for his initial idea, his support and attention.*

REFERENCES

Anderson, T., 1976, Estimation of Linear Functional Relationships: Approximate Distributions and Connections with Simulteneous Equations in Econometrics, Journal of the Royal Statistical Society, Series B, 38, N 1, 1-20.

Anderson, T., 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.

Anderson, T., Ch. Stein and A. Zaman, 1985, Best Invariant Estimation of a Direction Parameter, Annals of Statistics, 13, N 2, 526-533.

Andrews, D.W.K. and P. Guggenberger, 2005, Applications of Subsampling, Hybrid, and Size-Correction Methods, Cowles Foundation Discussion Paper 1608.

Andrews, D.W.K. , M. Moreira and J. Stock, 2006, Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression, Econometrica, 74(3), 715-752.

Andrews, D.W.K. , M. Moreira and J. Stock, 2007, Performance of Conditional Wald Tests in IV Regression with Weak Instruments, Journal of Econometrics, 139(1), 116-132.

M. Cattaneo, R. K. Crump and M. Jansson, 2009, Optimal Inference for Instrumental Variables Regression with non-Gaussian Errors, working paper.

Chamberlain, G., 2005, Decision Theory Applied to an Instrumental Variables Model, Working paper, Harvard University.

Dufour, J-M., 1997, Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models, Econometrica, 65, 1365-1388.

Dufour, J-M., L. Khalaf and M. Kichian, 2004, Inflation dynamics and the New Keynesian Phillips Curve: an identification robust econometric analysis, Discussion Paper, CIRANO and CIREQ, Universite de Montreal.

Dufour, J-M., and M. Taamouti, 2005, Projection-based Statistical Inference in Linear Structural Models with Possibly Weak Instruments, Econometrica, 73, 1351-1365.

Gleser, L., and J. Hwang, 1987, The Non-existence of $100(1-\alpha)\%$ Confidence Sets of Finite Expected Diameter in Error-in-variables and Related Models, Annals of Statistics, 15, 1351-62.

Ghosh, J.K., 1961, On the Relation Among Shortest Confidence Intervals of Different Types, Calcutta Statistical Association Bulletin, 147-152.

Hillier, G., 1990, On the Normalization of Structural Equations: Properties of

Direction Estimators, Econometrica, 58, 1181-1194.

Hillier, G., 2006, Exact Critical Value and Power Functions for the Conditional Likelihood Ratio and Related Tests in the IV Regression Model with Known Covariance, CeMMAP working paper #23/06

Kleibergen, F., 2002, Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression, Econometrica, 70, 1781-1803.

Kleibergen, F. and S. Mavroeidis, 2009, Weak Instrument Robust Tests in GMM and the New Phillips Curve, Journal of Business and Economic Statistics (forthcoming)

Lehmann, E., 1986, Testing Statistical Hypothesis, 2nd edition, Wiley Series in Probability and Mathematical Statistics.

Mikusheva, A. and B. Poi, 2006, Tests and confidence sets with correct size when instruments are potentially weak, Stata Journal, 6(3), 335-47.

Moreira, M., 2002, Tests with Correct Size in the Simultaneous Equations Model, PhD Thesis, UC Berkeley.

Moreira, M., 2003, A Conditional Likelihood Ratio Test for Structural Models, Econometrica, 71 (4), 1027-1048.

Müller, U.K., 2008, Efficient Tests under a Weak Convergence Assumption, working paper.

Nelson, C. and R. Startz, 1990, Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator, Econometrica, 58, 967-976.

Pollard, D., 1984, Convergence of Stochastic Processes, New York: Springer-Verlag.

Pratt, J., 1961, Length of Confidence Intervals, Journal of the American Stat. Assoc., v. 56, N. 295, 549-567.

Stock, J., M. Yogo and J. Wright, 2002, A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, Journal of Business and Economic Statistics, 20, 518-529.

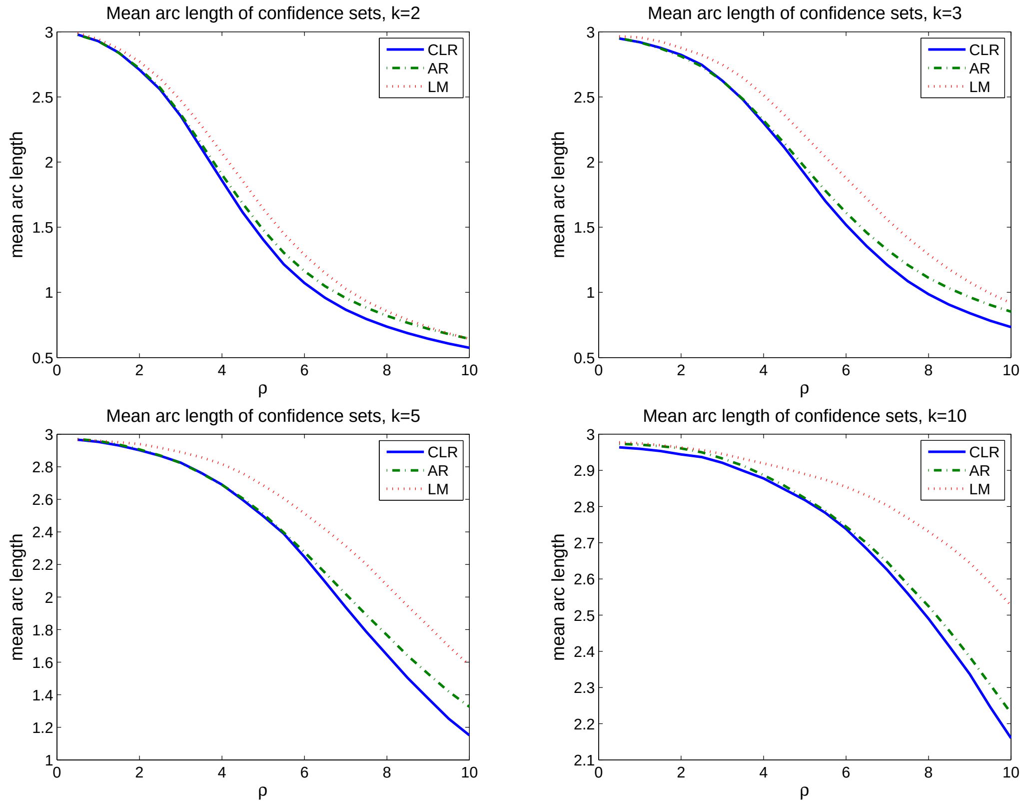


**Figure 1.** The expected arc length of the CLR, AR, and LM confidence sets for $k = 2, 3, 5, 10$.

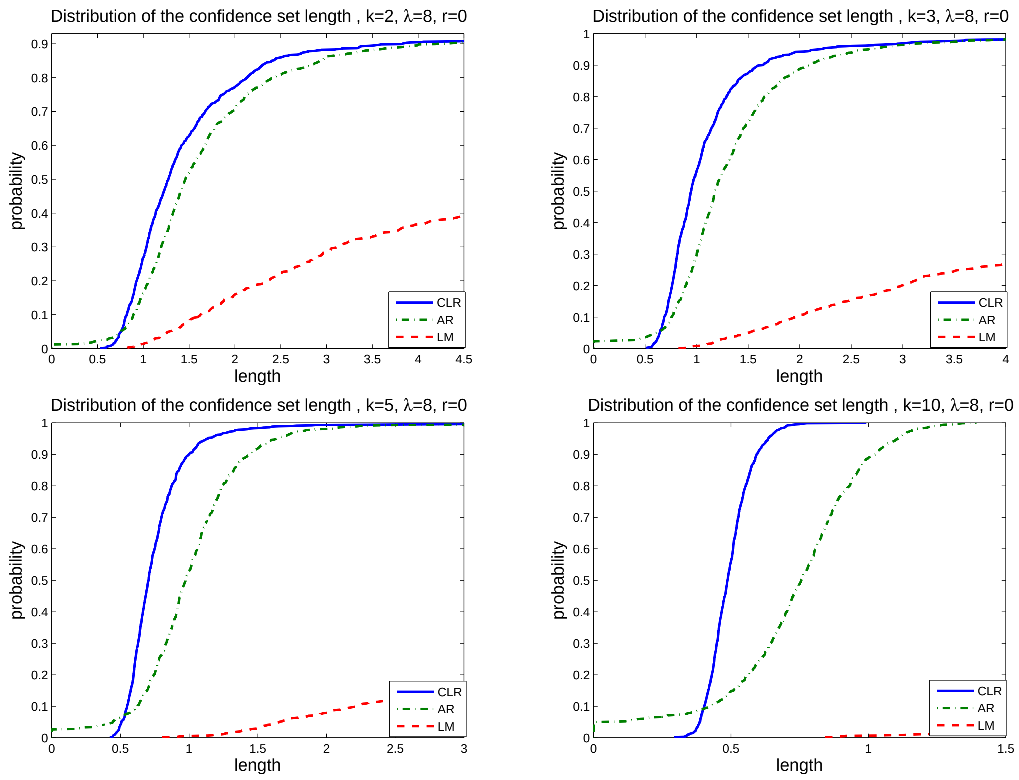


**Figure 2.** Distribution of the length of the CLR, AR, and LM confidence sets for $\lambda = 8, r = 0, k = 2, 3, 5, 10$.

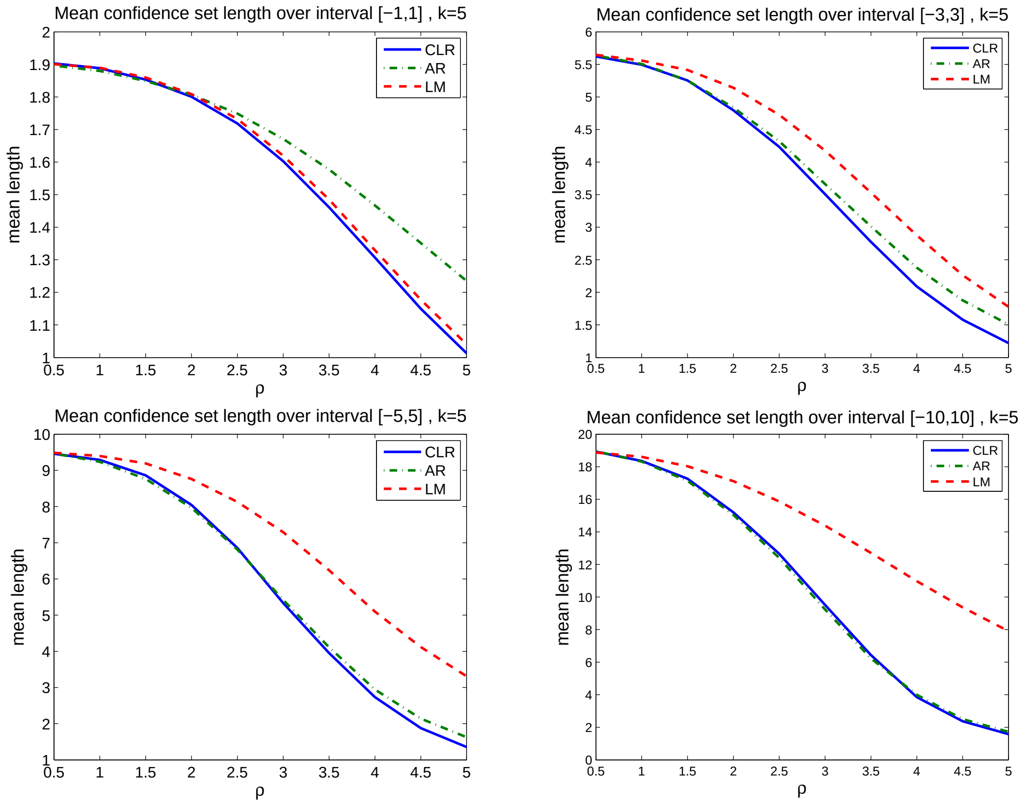


**Figure 3.** The expected length of intersection of the CLR, AR, and LM confidence sets with fixed finite intervals for $k = 5, \beta_0 = 0$.

| $\lambda$=8, $r$=0 | k=2 | k=3 | k=5 | k=10 |
|---|---|---|---|---|
| $P\{C^{AR} = \emptyset\}$ | 0.007 | 0.018 | 0.026 | 0.033 |
| $P\{lenghth(C^{AR}) = \infty\}$ | 0.048 | 0 | 0 | 0 |
| $P\{lenghth(C^{LM}) = \infty\}$ | 0.35 | 0.40 | 0.44 | 0.48 |
| $P\{lenghth(C^{CLR}) = \infty\}$ | 0.056 | 0.008 | 0 | 0 |

**Table 1.** Probability of having an empty or unbounded confidence set for the CLR, AR, and LM tests. $\lambda = 8, r = 0, k = 2, 3, 5, 10$.