Journal of Econometrics 157 (2010) 236–247

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Robust confidence sets in the presence of weak instruments

Anna Mikusheva *

MIT, Department of Economics, USA

ARTICLE INFO

Article history:
Received 16 June 2008
Received in revised form
10 December 2009
Accepted 19 December 2009
Available online 7 January 2010

JEL classification:
C30

Keywords:
Weak instruments
Confidence set
Uniform asymptotics

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 the 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.

1. Introduction

This paper considers confidence sets for the coefficient β 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, 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 β can be extracted. When instruments can be arbitrarily weak, a confidence set with correct coverage probability 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. Namely, a confidence set with correct coverage can be constructed, as the set of β_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 et al. (2002), Dufour et al. (2004) 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 provide a practitioner with simple and fast algorithms for obtaining the CLR, AR and LM confidence sets; currently a fast inversion algorithm exists for AR but not for the CLR or the LM. The second goal is to compare these confidence sets using the expected length as a criterion. 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.

The paper addresses the practical problem of inverting the CLR, LM and AR tests. One way of inverting a test is to do grid testing, namely, to perform a series of tests $H_0 : \beta = \beta_0$, where β_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 β_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 paper also compares the three above mentioned confidence sets in terms of expected length and attempts to establish

* Corresponding address: Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142, USA. Tel.: +1 617 324 5459; fax: +1 617 2531330.

E-mail address: amikushe@mit.edu.

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.12.003

optimality of the CLR confidence set. According to Pratt's (1961) theorem (also see Ghosh (1961)), the uniformly most powerful (UPM) test produces the confidence set with the shortest expected length if the expected length is *finite*. Andrews et al. (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 et al. (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, ϕ, lies on a one-dimensional unit circle. This parameter, ϕ, is in one-to-one correspondence with the coefficient, β, on the endogenous regressor. This circular model has two nice features. First, the length of the parameter space for ϕ is finite, which makes every confidence set for ϕ finite (a confidence interval of length *Pi* for ϕ corresponds to a confidence set for β equal to the whole line). Second, a circular model possesses additional symmetry and invariance properties. In particular, the two-sidedness condition corresponds to a *symmetry* on the circle. I show that the CLR confidence set has a 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 β 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.

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 et al. (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 gives algorithms for inverting the CLR, AR and LM tests. Section 4 discusses a correspondence between properties of tests and properties of confidence sets. 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. Section 7 concludes.

2. The model and notation

In this section I introduce notation and give a brief overview of the tests used in this paper for the confidence set construction. I keep the same notation as in Andrews et al. (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:

$$y_1 = y_2\beta + X\gamma_1 + u; \quad (1)$$

$$y_2 = Z\pi + X\xi + v_2. \quad (2)$$

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, β 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 ξ to $\widetilde{\xi} = \xi + (X'X)^{-1}X'Z\pi$.

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

$$y_1 = Z\pi\beta + X\gamma + v_1; \quad (3)$$

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

where $\gamma = \gamma_1 + \xi\beta$; $v_1 = u + \beta v_2$. The reduced form errors are assumed to be i.i.d. normal with zero mean and positive-definite covariance matrix Ω. Assume Ω 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 (β, π):

$$\zeta = (\Omega^{-1/2} \otimes (Z'Z)^{-1/2}Z')\,(y_1, y_2) = \left(\zeta_1', \zeta_2'\right)'.$$

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

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

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

I also consider 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

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