SCISPACE
formerly Typeset

Open access • Journal Article • DOI:10.1111/1468-0262.00353

Pivotal statistics for testing structural parameters in instrumental variables regression — Source link

Frank Kleibergen

Institutions: University of Amsterdam

Published on: 01 Sep 2002 - Econometrica (Tinbergen Institute)

Topics: Ancillary statistic, Pivotal quantity, Test statistic, PRESS statistic and Statistic

Related papers:

- A Conditional Likelihood Ratio Test for Structural Models
- Instrumental variables regression with weak instruments
- Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations
- A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments
- Some impossibility theorems in econometrics with applications to structural and dynamic models

Share this paper:

View more about this paper here: https://typeset.io/papers/pivotal-statistics-for-testing-structural-parameters-in-55dx7z55y1

TI 2000-055/4
Tinbergen Institute Discussion Paper

Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression

Frank Kleibergen

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at
http://www.tinbergen.nl

Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression

Frank Kleibergen*

June 29, 2000

Abstract

We propose a novel statistic for testing the structural parameters in Instrumental Variables Regression. The statistic is straightforward to compute and has a limiting distribution that is pivotal with a degrees of freedom parameter that is equal to the number of tested parameters. It therefore differs from the Anderson-Rubin statistic, whose limiting distribution is pivotal but has a degrees of freedom parameter that is equal to the number of instruments, and the Likelihood based, Wald, Likelihood Ratio and Lagrange Multiplier, statistics, whose limiting distributions are not pivotal. We analyze the relationship between the statistic and the concentrated likelihood of the structural parameters and show that its' limiting distribution is not affected by weak instruments. We discuss examples of the non-standard shapes of the asymptotically pivotal confidence sets that can be constructed using the statistic and investigate its power properties. To show its importance for practical purposes, we apply the statistic to the Angrist-Krueger (1991) data and find similar results as in Staiger and Stock (1997).

1 Introduction

Instrumental variables regression is commonly applied. Appropriately conducting inference on the parameters of this model has therefore traditionally been an important research topic. Initially, the focus of this research was on constructing the distribution of estimators of the structural parameters, see *e.g.* Mariano and Sawa (1972), Anderson *et. al.* (1979,1983) and Phillips (1983,1989). More recently, this focus has shifted more towards testing the structural parameters, see *e.g.* Staiger and Stock (1997), Dufour (1997), Nelson and Startz (1990) and Zivot, Nelson and Startz (1998), although also the early literature contains contributions to this topic, see, for example, Anderson and Rubin (1949). This paper belongs to the second stream of papers.

The problem with conducting inference on the structural parameters in Instrumental variable regression results as they appear as a product in the (restricted) reduced form. Hence, the inference on the structural parameters is conditional on the other parameters in the product which are the reduced form parameters of the second (set of) endogenous variable(s).

*I thank Peter Boswijk, Geert Ridder and participants of the Tinbergen Institute Labor Economics Seminar for helpful comments and discussion. Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Email:kleiberg@fee.uva.nl. Homepage: http://www.fee.uva.nl/ke/kleibergen.htm.

When these latter parameters are equal to zero, inference on the structural parameters breaks down and estimators converge to random variables, see Phillips (1989). Statistics that test hypothezes on the structural parameters are also hampered by these problems such that, for example, the limiting distributions of the Likelihood based, Wald, Likelihood Ratio and Lagrange Multiplier, statistics assume a fixed non-zero full rank value of the reduced form parameter of the second (set of) endogenous variable(s), see *e.g.* Staiger and Stock (1997) and Dufour (1997). Staiger and Stock (1997), see also Wang and Zivot (1998), therefore propose an alternative testing procedure that is based on the dependence of the estimator of the structural parameters on a statistic that indicates the significance of the reduced form parameters, *i.e.* the concentration parameter, and use Bonferonni's inequality. The Anderson-Rubin statistic, see Anderson and Rubin (1949), can also be used for conducting tests on the structural parameters. Its limiting distribution is pivotal but has a degrees of freedom parameter that is equal to the number of instruments and is therefore larger than or equal to the number of structural parameters. This affects the power of the statistic. In the paper we propose a novel statistic for testing the structural parameters whose limiting distribution is pivotal and has a degrees of freedom parameter that is equal to the number of structural parameters. The statistic results from the specification of the maximum likelihood estimators in Instrumental Variable Regression as invertible functions of orthogonal statistics from Kleibergen (2000).

The paper is organized as follows. In section 2, we propose the statistic for conducting tests on the structural parameters. We construct its limiting distribution and show that is pivotal and has a degrees of freedom parameter that is equal to the number of structural parameters. Its' limiting distribution is therefore not affected by weak instruments. In section 4, we discuss the relationship of the statistic with the concentrated likelihood of the structural parameters. Section 5 shows the different shapes, as discussed by *e.g.* Dufour (1997) and Staiger and Stock (1997), of the confidence sets that can be constructed using the asymptotically pivotal statistic. In section 6, we conduct a power comparison of different statistics for conducting tests on the structural parameters. Section 7 contains an application to the Angrist-Krueger (1991) data. Finally, the eight section concludes.

2 Instrumental Variable Regression Model

The instrumental variables regression model in *structural form* can be represented as a limited information simultaneous equation model, see *e.g.* Hausman (1983) and Kleibergen and Zivot (1998),

$$\begin{aligned} y_1 &= Y_2\beta + Z\gamma + \varepsilon_1, \\ Y_2 &= X\Pi + Z\Gamma + V_2, \end{aligned} \tag{1}$$

where y_1 and Y_2 are a $T \times 1$ and $T \times (m-1)$ matrix of endogenous variables, respectively, Z is a $T \times k_1$ matrix of included exogenous variables, X is a $T \times k_2$ matrix of excluded exogenous variables (or instruments), ε_1 is a $T \times 1$ vector of structural errors and V_2 is a $T \times (m-1)$ matrix of reduced form errors. The $(m-1) \times 1$ and $k_1 \times 1$ parameter vectors β and γ contain the structural parameters. The variables in X and Z are assumed to be of full column rank, uncorrelated with ε_1 and V_2, and to be weakly exogenous for β and Π, see Engle *et. al.* (1983). The error terms ε_{1t} and V_{2t}, where ε_{1t} denotes the t-th observation on ε_1 and V_{2t} is a column vector denoting the t-th row of V_2, are assumed to be uncorrelated over time, to have finite moments up to at least the fourth order and the finite $m \times m$ (unconditional) covariance matrix is represented by

$$\Sigma = var \begin{pmatrix} \varepsilon_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \tag{2}$$

and is assumed to be unknown. The degree of endogeniety of Y_2 in (1) is determined by the vector of correlation coefficients defined by $\rho = \Sigma_{22}^{-1/2\prime}\Sigma_{21}\sigma_{11}^{-1/2}$ and the quality of the instruments is captured by Π.

Substituting the reduced form equation for Y_2 into the structural equation for y_1 gives the non-linearly *restricted reduced form* specification

$$Y = X\Pi B + Z\Psi + V, \tag{3}$$

where $Y = \begin{pmatrix} y_1 & Y_2 \end{pmatrix}$, $B = \begin{pmatrix} \beta & I_{m-1} \end{pmatrix}$, $\Psi = \Gamma B + \begin{pmatrix} \gamma & 0 \end{pmatrix}$, $V = \begin{pmatrix} v_1 & V_2 \end{pmatrix}$, $v_1 = \varepsilon_1 + V_2\beta$ and, hence, $(v_{1t}\ V_{2t}')'$ has covariance matrix

$$\Omega = var \begin{pmatrix} v_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{pmatrix}, \tag{4}$$

Note that Ψ is a unrestricted $k_1 \times m$ matrix.

The *unrestricted reduced form* of the model expresses each endogenous variable as a linear function of the exogenous variables and is given by

$$Y = X\Phi + Z\Psi + V, \tag{5}$$

where $\Phi : k_2 \times m$, $\Phi = \begin{pmatrix} \varphi_1 & \Phi_2 \end{pmatrix}$, $\varphi_1 : k_2 \times 1$, $\Phi_2 : k_2 \times (m-1)$. Since the unrestricted reduced form is a multivariate linear regression model, all of the reduced form parameters are identified. It is assumed that $k_2 \geq m-1$ so that the structural parameter vector β is "apparently" identified by the order condition. We call the model just-identified when $k_2 = m-1$ and the model over-identified when $k_2 > m-1$. $k_2 - m + 1$ is therefore the degree of over-identification. β is identified if and only if $rank(\Pi) = m-1$. The extreme case in which β is totally unidentified occurs when $\Pi = 0$ and, hence, $rank(\Pi) = 0$, see Phillips (1989). The case of "weak instruments", as discussed by Nelson and Startz (1990), Staiger and Stock (1997), Wang and Zivot (1998), and Zivot, Nelson and Startz (1998), occurs when Π is close to zero or, as discussed by Kitamura (1994), Dufour and Khalaf (1997) and Shea (1997) when Π is close to having reduced rank.

The parameter β is typically the focus of the analysis. We can therefore simplify the presentation of the results without changing their implications by setting $\gamma = 0$ and $\Gamma = 0$ ($\Psi = 0$) so that Z drops out of the model. In what follows, let $k = k_2$ denote the number of instruments. We note that the form of the analytical results for β in this simplified case carry over to the more general case where $\gamma \neq 0$ and $\Gamma \neq 0$ by interpreting all data matrices as residuals from the projection on Z.

3 Pivotal Statistic for the Structural Parameters

3.1 Motivation

The first order condition (foc) for β in the restricted reduced from (3) reads as

$$\Pi' X'(y_1 - X\Pi\beta) = 0. \tag{6}$$

We can replace the unknown Π by its least squares estimate such that the foc becomes

$$\hat{\Phi}_2' X'X(\hat{\varphi}_1 - \hat{\Phi}_2\beta) = 0 \tag{7}$$

where $\hat{\Phi} = (\hat{\varphi}_1 \; \hat{\Phi}_2) = (X'X)^{-1}X'Y$, and which results in the 2SLS estimator for β. Under $H_0 : \beta = 0$ and for a full rank value of Π, we then obtain that

$$\frac{1}{s_{11}} \hat{\varphi}_1' X'X\hat{\Phi}_2 \left(\hat{\Phi}_2' X'X\hat{\Phi}_2\right)^{-1} \hat{\Phi}_2' X'X\hat{\varphi}_1 \Rightarrow \chi^2(m-1), \tag{8}$$

where $s_{11} = \frac{1}{T-K} y_1' M_X y_1$, $M_V = I_T - V(V'V)^{-1}V'$. In the extreme case of a zero value of Π, however,

$$\sqrt{T}\hat{\Phi} \Rightarrow \begin{pmatrix} u_1 & U_2 \end{pmatrix} \sim N\left(0, \Omega \otimes Q^{-1}\right), \tag{9}$$

with $u_1 : k \times 1$, $U_2 : k \times (m-1)$ and $Q = p\lim_{T\to\infty} \frac{X'X}{T}$. We can then specify u_1 as

$$u_1 = U_2 \Omega_{22}^{-1} \omega_{21} + u_{1.2}$$

where $\omega_{11.2} = \omega_{11} - \omega_{12}\Omega_{22}^{-1}\omega_{21}$, $u_{1.2} \sim N(0, \omega_{11.2} \otimes Q^{-1})$ and stochastic independent of U_2, such that

$$\begin{aligned} \tfrac{1}{s_{11}} \hat{\varphi}_1' X'X\hat{\Phi}_2 \left(\hat{\Phi}_2' X'X\hat{\Phi}_2\right)^{-1} \hat{\Phi}_2' X'X\hat{\varphi}_1 \Rightarrow \; & \left(U_2\Omega_{22}^{-1}\omega_{21} + u_{1.2}\right)' QU_2 \left(U_2' QU_2\right)^{-1} U_2' Q \\ & \left(U_2\Omega_{22}^{-1}\omega_{21} + u_{1.2}\right) \end{aligned} \tag{10}$$

which is not equal to a $\chi^2(m-1)$ random variable. This shows that the 2SLS t-statistic doesnot have a standard normal limiting distribution in case of invalid instruments, see *e.g.* Phillips (1989). A similar argument as the above one can also be used in case of weak instruments, see Staiger and Stock (1997), and to show that also the limiting distributions of the Likelihood based, Wald, Likelihood Ratio and Lagrange Multiplier, statistics differ in case of weak or invalid instruments, see *e.g.* Dufour (1997).

Instead of projecting $\hat{\varphi}_1$ on $\hat{\Phi}_2$, with which it is (asymptotically) correlated such that the limiting distribution in (8) breaks down in case of weak or invalid instruments, we project $\hat{\varphi}_1$ on a specific variable that is, under H_0, both a consistent estimator of Π and (asymptotically) stochastic independent of $\hat{\varphi}_1$. This variable orginates from the specification of the maximum likelihood estimators of the parameters of the restricted and unrestricted reduced forms as invertible functions of orthogonal statistics that is constructed in Kleibergen (2000). We use it instead of $\hat{\Phi}_2$ in (8). The specification of the variable is

$$\hat{\Phi}_2 - (\hat{\varphi}_1 - \Pi\beta)\omega_{11}^{-1}\omega_{12} \tag{11}$$

such that it is under $H_0 : \beta = 0$ equal to

$$\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12} \tag{12}$$

and we can use the consistent estimator $\hat{\Theta} = \hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1} s_{12}$ instead of $\hat{\Phi}_2$ in (8), with $s_{12} = \frac{1}{T-k} y_1' M_X Y_2$.

3.2 Asymptotically Pivotal Statistic

A statistic with an asymptotic distribution that is independent of nuisance parameters can be constructed for conducting tests on the structural parameter β. We construct this statistic to test the hypothesis $H_0 : \beta = 0$ against the alternative hypothesis $H_1 : \beta \neq 0$ but the statistic can as well be used to test for other values of β.

Under H_0 and the assumptions made for the disturbances, the joint limiting distribution of the least squares estimators $\hat{\varphi}_1$ and $\hat{\Phi}_2$ is normal

$$\sqrt{T}\left(\hat{\Phi} - (0\ \Pi)\right) \Rightarrow N\left(0, \Omega \otimes Q^{-1}\right), \tag{13}$$

where " $\Rightarrow$ " stands for weak convergence, see Billingsley (1986), and $Q = p\lim_{T\to\infty} \frac{X'X}{T}$. This implies the marginal and conditional limiting distributions of $\hat{\varphi}_1$ and $\hat{\Phi}_2$, see Kleibergen (2000),

$$\begin{aligned} \sqrt{T}\hat{\varphi}_1 &\Rightarrow N(0, \omega_{11} \otimes Q^{-1}), \\ \sqrt{T}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12} - \Pi\right) &\Rightarrow N(0, \Omega_{22.1} \otimes Q^{-1}), \end{aligned} \tag{14}$$

where $\Omega_{22.1} = \Omega_{22} - \omega_{21}\omega_{11}^{-1}\omega_{12}$, and $\sqrt{T}\hat{\varphi}_1$ and $\sqrt{T}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12}\right)$ are asymptotically stochastic independent. The expression in (14) contains the unobserved parameters ω_{11} and ω_{12} which we need to replace by observable ones. When S is the (consistent) estimator of the covariance matrix Ω,

$$S = \frac{1}{T-k}Y'M_XY,\ S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & S_{22} \end{pmatrix}, \tag{15}$$

where $M_V = I_T - V(V'V)^{-1}V'$, $V = X$; $s_{11} : 1 \times 1$, s_{12}, $s_{21}' : 1 \times (m-1)$ and $S_{22} : (m-1) \times (m-1)$, S is (asymptotically) stochastic independent from $\hat{\Phi}$ and $s_{11}^{-1}s_{12}$ is a consistent estimator of $\omega_{11}^{-1}\omega_{12}$. We can therefore replace $\omega_{11}^{-1}\omega_{12}$ by $s_{11}^{-1}s_{12}$ in (14) and obtain that, see Kleibergen (2000),

$$\sqrt{T}\left(\hat{\Theta} - \Pi\right) \Rightarrow N(0, \Omega_{22.1} \otimes Q^{-1}) \tag{16}$$

where $\hat{\Theta} = \hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}$ and $\sqrt{T}\hat{\Theta}$ is asymptotically stochastic independent of $\sqrt{T}\hat{\varphi}_1$.[1] Consequently, since s_{11} is a consistent estimator of ω_{11}

$$\left(\hat{\Theta}'X'X\hat{\Theta}\right)^{-\frac{1}{2}} \hat{\Theta}'X'X\hat{\varphi}_1 s_{11}^{-\frac{1}{2}} \Rightarrow N(0, I_{m-1}) \tag{17}$$

and

$$\begin{aligned} F(H_0|H_1) &= \tfrac{1}{(m-1)s_{11}} y_1'X(X'X)^{-1}X'(Y_2 - y_1 s_{11}^{-1}s_{12})\left[(Y_2 - y_1 s_{11}^{-1}s_{12})'X(X'X)^{-1}X'\right. \\ &\quad \left.(Y_2 - y_1 s_{11}^{-1}s_{12})\right]^{-1}(Y_2 - y_1 s_{11}^{-1}s_{12})'X(X'X)^{-1}X'y_1 \\ &= \frac{\hat{\varphi}_1'X'X\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\right)\left(\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\right)'X'X\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\right)\right)^{-1}\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\right)'X'X\hat{\varphi}_1}{(m-1)s_{11}} \\ &= \frac{\hat{\varphi}_1'X'X\hat{\Theta}\left(\hat{\Theta}'X'X\hat{\Theta}\right)^{-1}\hat{\Theta}'X'X\hat{\varphi}_1}{(m-1)s_{11}} \\ &\Rightarrow \frac{\chi^2(m-1)}{m-1}, \end{aligned} \tag{18}$$

[1](16) results as $\sqrt{T}\hat{\varphi}_1 \Rightarrow N(0, I_k)$ and $s_{11}^{-1}s_{12} \Rightarrow \omega_{11}^{-1}\omega_{12}$. Hence, $\sqrt{T}\left(\hat{\Theta} - \Pi\right) = \sqrt{T}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12} - \Pi\right) + \left(\sqrt{T}\hat{\varphi}_1\right)\left(\omega_{11}^{-1}\omega_{12} - s_{11}^{-1}s_{12}\right) \Rightarrow \sqrt{T}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12} - \Pi\right)$ as $\left(\sqrt{T}\hat{\varphi}_1\right)\left(\omega_{11}^{-1}\omega_{12} - s_{11}^{-1}s_{12}\right) \Rightarrow 0$ since $\hat{\varphi}_1$ and $s_{11}^{-1}s_{12}$ are (asymptotically) stochastic independent and $\sqrt{T}\hat{\varphi}_1 \Rightarrow N(0, I_k)$ and $s_{11}^{-1}s_{12} \Rightarrow \omega_{11}^{-1}\omega_{12}$.

which shows that the asymptotic distribution of (18) is completely characterized by the parameters under H_0 and doesnot depend on unobserved nuisance parameters. The difference between the limiting distribution in (18) and the limiting distributions of Likelihood Ratio, Wald and Lagrange Multiplier statistics, is that the limiting distribution (18) is independent of nuisance parameters. The limiting distribution of the other statistics is based on the assumption of a fixed full rank value of Π, see *e.g.* Dufour (1997), Staiger and Stock (1997) and Wang and Zivot (1998). Another asymptotically pivotal statistic that can be used to test H_0 is the Anderson-Rubin statistic, see Anderson and Rubin (1949). The degrees of freedom of the limiting distribution of (18) is exactly equal to the number of parameters pre-specified in H_0 while it is equal to the sum of this number of parameters and the degree of over-identification, $k-m+1$, for the Anderson-Rubin statistic. The difference results as the Anderson-Rubin statistic conducts a joint test of H_0 and of the restricted reduced form against the unrestricted reduced form and affects the power of the statistic when it is used as a statistic to only test H_0. In section 6, we therefore analyze the power of the Anderson-Rubin statistic and statistic (41) for a few simulated datasets.

In the just-identified case k is equal to $m-1$ and $\hat{\Theta}$ in (18) is invertible. The statistic (18) and the Anderson-Rubin statistic are then identical. This further shows that statistic (18) is the appropriate generalization of the Anderson-Rubin statistic from the just-identified to the over-identified case as a statistic to only test H_0.

3.3 Distribution Functions

To illustrate that the statistic (18) is an asymptotically pivotal statistic under H_0, we computed its empirical distribution function for two data generating processes with weak instruments and strong endogeneity. Figure 1 contains the distribution functions for a data generating process with $m=2$ and $m=3$ in figure 2. Figures 1 and 2 contain the empirical distribution functions of (18) jointly with the asymptotic distribution function and also the empirical distribution function of the likelihood ratio statistic. As (18) is an asymptotically pivotal statistic, its empirical distribution function coincides with its asymptotic distribution function. Figures 1 and 2 also show that the distribution of the likelihood ratio statistic differs from its asymptotic distribution as its' limiting distribution depends on nuisance parameters.

3.4 Weak Instruments

To functionalize the in practice frequently observed combination of a large sample size and a small but significant "F-statistic" for instrument relavence, Staiger and Stock, see Staiger and Stock (1997), specify the reduced form parameter matrix Π such that it decreases with the sample size, *i.e.* $\Pi = \frac{1}{\sqrt{T}}\Psi$ with Ψ a fixed full rank parameter matrix. The limiting expression in (14) then changes to

$$\begin{aligned} \sqrt{T}\hat{\varphi}_1 &\Rightarrow N(0, \omega_{11} \otimes Q^{-1}) \\ \sqrt{T}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12}\right) &\Rightarrow N(\Psi, \Omega_{22.1} \otimes Q^{-1}) \end{aligned} \tag{19}$$

and $\sqrt{T}\hat{\varphi}_1$ and $\sqrt{T}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12}\right)$ are asymptotically stochastic independent. When we replace $\omega_{11}^{-1}\omega_{12}$ by $s_{11}^{-1}s_{12}$, (19) still holds

$$\begin{aligned} \sqrt{T}\hat{\varphi}_1 &\Rightarrow N(0, \omega_{11} \otimes Q^{-1}), \\ \sqrt{T}\hat{\Theta} &\Rightarrow N(\Psi, \Omega_{22.1} \otimes Q^{-1}) \end{aligned} \tag{20}$$

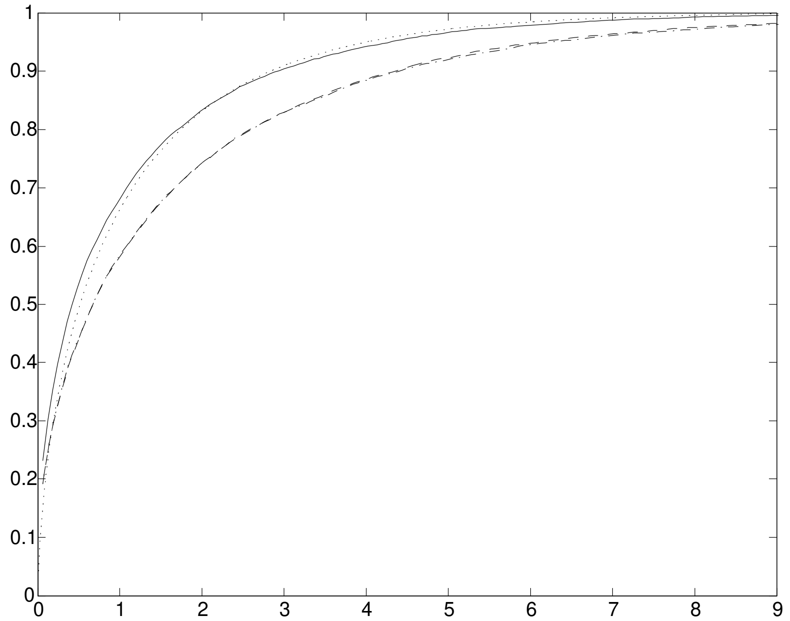

Figure 1: Distribution pivotal statistic (18), $m = 2$, $k = 20$, $T = 100$, weak instruments and strong endogeneity, asymp. dis. (..), pivotal stat. (-), LR stat (-.-).

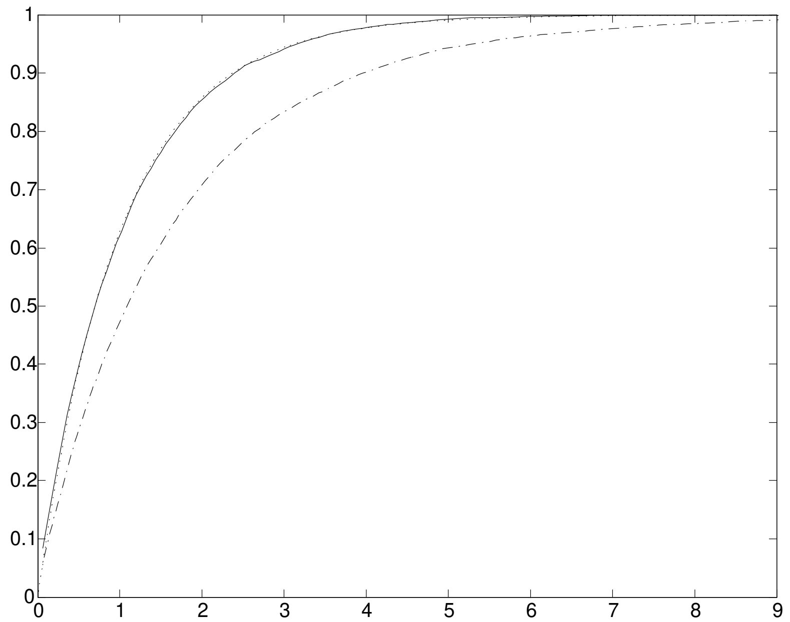

Figure 2: Distribution pivotal statistic (18), $m = 3$, $k = 20$, $T = 100$, weak instruments and strong endogeneity, asymp. dis. (..), pivotal stat. (-), LR stat/2. (-.-).

where $\hat{\Theta} = \hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1} s_{12}$, with asymptotically stochastic independent $\sqrt{T}\hat{\varphi}_1$ and $\sqrt{T}\hat{\Theta}$ and

$$\left(\hat{\Theta}' X' X \hat{\Theta}\right)^{-\frac{1}{2}} \hat{\Theta}' X' X \hat{\varphi}_1 s_{11}^{-\frac{1}{2}} \Rightarrow N(0, I_{m-1}). \tag{21}$$

The limiting distribution of (18) is therefore not affected by the specification of Π which further shows that, its' limiting distribution is independent of unobserved nuisance parameter as, it is an asymptotically pivotal statistic.

4 Relationship Pivotal Statistic and LIML estimator

By using $y_1^* = y_1 - Y_2\beta_0$ instead of y_1 in all the elements of the statistic (18) that contain y_1, (18) can also be used to test the hypothesis $H_0^* : \beta = \beta_0$ for various values of β_0. The expression for (18) then becomes

$$\begin{aligned} F(H_0^*|H_1) = & \tfrac{1}{(m-1)s_{11}^*} y_1^{*\prime} X(X'X)^{-1}X'(Y_2 - y_1^* s_{11}^{*-1} s_{12}^*) \left[(Y_2 - y_1^* s_{11}^{*-1} s_{12}^*)' X(X'X)^{-1}X' \right. \\ & \left. (Y_2 - y_1^* s_{11}^{*-1} s_{12}^*)\right]^{-1} (Y_2 - y_1^* s_{11}^{*-1} s_{12}^*)' X(X'X)^{-1}X' y_1^* \\ = & \tfrac{T-k}{(m-1)(y_1 - Y_2\beta_0)' M_X (y_1 - Y_2\beta_0)} (y_1 - Y_2\beta_0)' P_{X(X'X)^{-1}X'(Y_2 - (y_1 - Y_2\beta_0)s_{11}^{*-1} s_{12}^*)} (y_1 - Y_2\beta_0), \end{aligned} \tag{22}$$

where $s_{11}^* = \frac{1}{T-k}(y_1 - Y_2\beta_0)' M_X (y_1 - Y_2\beta_0)$, $s_{12}^* = \frac{1}{T-k}(y_1 - Y_2\beta_0)' M_X Y_2 = s_{12} - \beta_0' S_{22}$ and $P_V = V(V'V)^{-1}V'$, $V = X(X'X)^{-1}X'(Y_2 - (y_1 - Y_2\beta_0)\, s_{11}^{*-1} s_{12}^*)$. Instead of maximizing the likelihood function to obtain the limited information maximum likelihood (liml) estimator of β, we also obtain an estimator of β by minimizing (22) over β_0. We analyze the similarities and differences between the resulting estimators. We therefore first briefly discuss the construction of the liml estimator.

The mle of β, $\hat{\beta}$, is obtained from the concentrated log-likelihood, under independently normal distributed disturbances with a fixed covariance matrix, that results when we have concentrated out Π and Σ from the log-likelihood of the parameters of model (1), see *e.g.* Hausman (1983),

$$\begin{aligned} \log(L(\beta|X,Y)) &= \tfrac{1}{2}T \log \left| \tfrac{(y_1 - Y_2\beta)' M_X (y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'(y_1 - Y_2\beta)} \right| \\ &= \tfrac{1}{2}T \log \left| 1 - \tfrac{(y_1 - Y_2\beta)' X(X'X)^{-1}X'(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'(y_1 - Y_2\beta)} \right| \\ &= \tfrac{1}{2}T \log |1 - \eta|, \end{aligned} \tag{23}$$

where $\eta = \frac{(y_1 - Y_2\beta)' X(X'X)^{-1}X'(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'(y_1 - Y_2\beta)}$. Since the concentrated log-likelihood of β is a monotonic decreasing function of η, maximizing with respect to β is identical to finding the minimal value of η,

$$\eta = \min_{\beta} \left[\frac{(y_1 - Y_2\beta)' X(X'X)^{-1}X'(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'(y_1 - Y_2\beta)} \right], \tag{24}$$

which is identical to solving the eigenvalue problem,

$$\begin{aligned} |\eta Y'Y - Y'X(X'X)^{-1}X'Y| = 0 \Leftrightarrow \\ |\eta I_m - (Y'Y)^{-1}\hat{\Phi} X' X \hat{\Phi}| = 0, \end{aligned} \tag{25}$$

and to use the smallest root of (25), see Anderson and Rubin (1949) and Hood and Koopmans (1953). The liml estimator of β, $\hat{\beta}$, is then contructed such that the eigenvector associated with η equals $a(1\ -\hat{\beta}')'$, where a is the first element of the eigenvector associated with η.

Maximizing the log-likelihood (23) is identical to minimizing minus the log-likelihood

$$\begin{aligned} -\log(L(\beta|X,Y)) &= \tfrac{1}{2}T\log\left|\tfrac{(y_1-Y_2\beta)'(y_1-Y_2\beta)}{(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)}\right| \\ &= \tfrac{1}{2}T\log\left|1+\tfrac{(y_1-Y_2\beta)'X(X'X)^{-1}X'(y_1-Y_2\beta)}{(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)}\right| \end{aligned} \tag{26}$$

which is again identical to minimizing

$$K(\beta)=\frac{(y_1-Y_2\beta)'\,A(y_1-Y_2\beta)}{(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)} \tag{27}$$

with $A=P_X=X(X'X)^{-1}X'$. The statistic (22) is identical to (27) multiplied by $\frac{T-k}{m-1}$ and using the specification of $A=P_{XD^*}$ with $D^*=(X'X)^{-1}X'(Y_2-(y_1-Y_2\beta)\,s_{11}^{*-1}s_{12}^*)$. P_{XD^*} is a projection on XD^* which is a sub-space of X on which P_X projects. It shows that the liml estimator of β and the estimator that results by minimizing (22) are closely related. This can be further verified by analyzing the derivative of $K(\beta)$ (27) with respect to β

$$\begin{aligned} \tfrac{\partial K(\beta)}{\partial\beta'} = \;& 2K(\beta)\left[\tfrac{(y_1-Y_2\beta)'M_XY_2}{(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)}-\tfrac{(y_1-Y_2\beta)'AY_2}{(y_1-Y_2\beta)'A(y_1-Y_2\beta)}\right]+ \\ & \tfrac{1}{(y_1-Y_2\beta)'M_X(y_1-Y_2\beta)}\left((y_1-Y_2\beta)'\otimes(y_1-Y_2\beta)'\right)\tfrac{\partial vec(A)}{\partial\beta'}. \end{aligned} \tag{28}$$

Since A is a projection matrix, for $A=P_{XD^*}$, $\frac{\partial vec(A)}{\partial\beta'}$ is very small and is equal to zero for $A=P_X$. As a consequence, the value of β that is such that the first part of (28) is equal to zero, which approximately holds for the liml estimator since $A=P_{XD^*}$ is a projection on a sub-space of X, is also the value that satisfies the first order condition. Thus the liml estimator more or less coincides with the estimator that minimizes the statistic (22).

5 Confidence Regions

By using $y_1^*=y_1-Y_2\beta_0$ instead of y_1 in all the elements of the statistic (18) that contain y_1, (18) can also be used to test the hypothesis $H_0^*:\beta=\beta_0$ for various values of β_0. By specifying a grid of values for β_0, we can then construct a $\alpha\%$ asymptotic confidence set for β that is independent of the value of the other parameters. These asymptotic pivotal confidence sets can have peculiar shapes that differ from the standard symmetric asymptotic $\alpha\%$ confidence sets, see *e.g.* Dufour (1997) and Zivot *et. al.* (1998). Possible shapes that can occur are infinite confidence sets, discontinuous confidence sets and empty confidence sets. We show examples of these kind of asymptotically pivotal confidence sets by simulating data from pre-specified Data Generating Processes (DGPs). As shown before, the minimal value of the statistic (18) lies at the liml estimator. An empty confidence set can therefore occur when the true DGP doesnot correspond with the estimated model and the statistic (18) is even significant when it is used to test the hypothesis that β is equal to its maximum likelihood estimate.

For each of the asymptotic p-value plots that we show in figures 3-8, we have artificially sampled data from the model

$$\begin{aligned} y_1 &= \beta y_2+\varepsilon_1 \\ y_2 &= X\pi+v_2, \end{aligned} \tag{29}$$

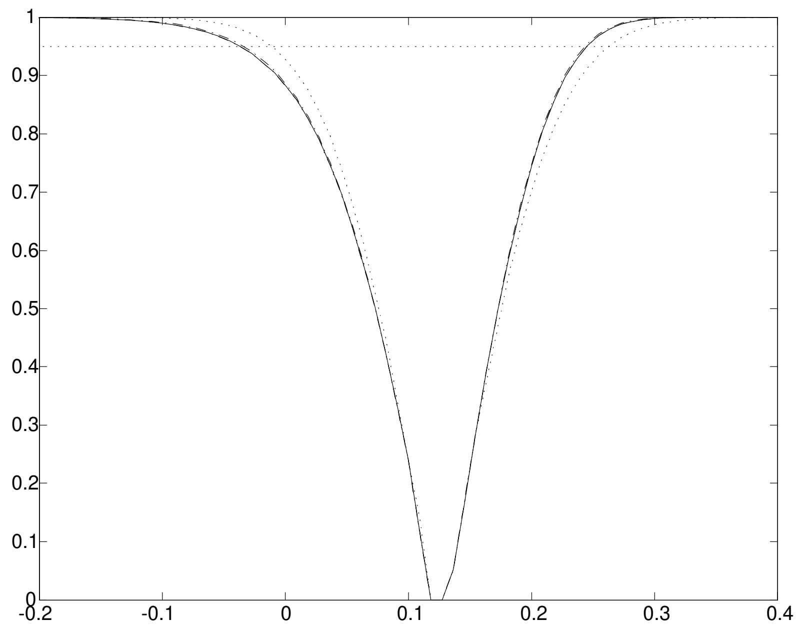

Figure 3: p-value plots of statistics that test the hypothesis $H_0 : \beta = \beta_0$ for dataset with $\pi_1 = 1$, $k = 1$. statistic (18) (-), Anderson-Rubin (- -), LR (-.-), 2sls t-statistic (..).

where y_1, $y_2 : T \times 1$, $X : T \times k$, $(\varepsilon_1 \ v_2) \sim N(0, \Sigma \otimes I_T)$; $X \sim N(0, I_k \otimes I_T)$, $T = 100$, $\pi : k \times 1$, $\pi = (\pi_1 ... \pi_k)'$, $\pi_2 = ... = \pi_k = 0$, $\beta = 0$, $\Sigma = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}$; for values of π_1 equal to 0.1, *i.e.* a weak instrument, and 1, *i.e.* a valid instrument, and values of k equal to 1, 5 and 20. The data that are simulated from the DGP therefore only differ over the value of π_1 and are the same for the different values of k. Hence, for a fixed π_1, we only add superfluous instruments to the model when we increase k and the generated endogenous variables stay the same. In this manner, we visualize the robustness of the statistics to adding superfluous instruments.

$\mathbf{k = 1}$ Since m is equal to 2, the model is just-identified when k is equal to one. Hence, the Anderson-Rubin statistic and statistic (18) are identical. Also the 2SLS and liml estimators in this case coincide. Figures 3 and 4 show asymptotic p-value plots of the 2SLS t-statistic, statistic (18), the Anderson-Rubin statistic and the Likelihood Ratio statistic that test the hypothesis $H_0 : \beta = \beta_0$ for artificial datasets with $\pi_1 = 1$, figure 3, and $\pi_1 = 0.1$, figure 4. The figures also contain a straight line at 0.95 that enables us to construct the 95% asymptotic confidence set in a straightforward way. The instrument in figure 3 is thus a valid instrument while it is a weak one in figure 4.

Because the Anderson-Rubin statistic is identical to statistic (18) in the just-identified case, the p-value plots of both statistics are indistinguishable. Also the p-value plot of the Likelihood Ratio statistic is hard to distinguish from the p-value plot of the Anderson-Rubin statistic. These statistics are closely related in the just-identified case as the test on β is essentially a linear test then. The 2SLS and liml estimators are identical in the exact identified case which explains why the different p-value plots have values of β in common at the zero p-value. The 2SLS t-statistic and LR-AR-pivotal statistics p-value plots are very similar for the valid instrument case but very different for the weak instrument case. The 95% confidence set that results from the LR-AR-pivotal statistics is even discontinuous and infinite for the latter case

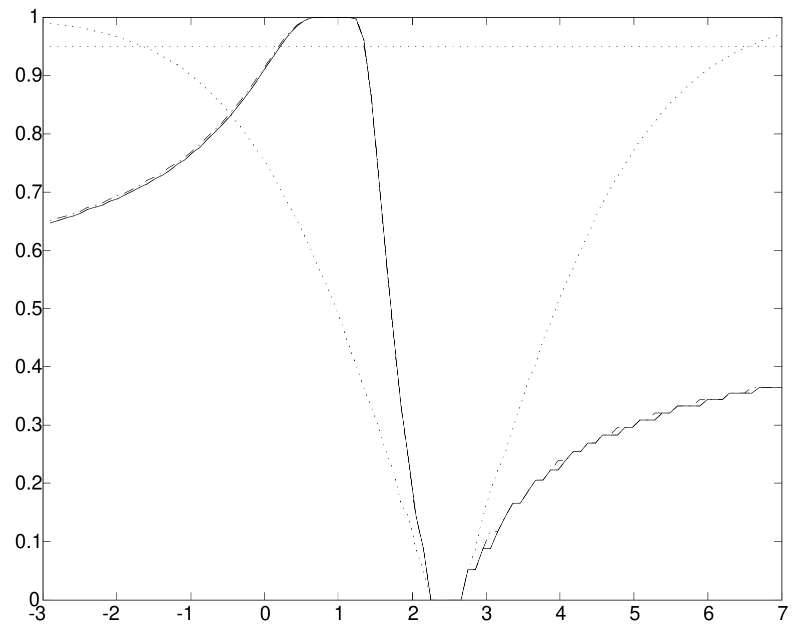

Figure 4: p-value plots of statistics that test the hypothesis $H_0 : \beta = \beta_0$ for dataset with $\pi_1 = 0.1$, $k = 1$. statistic (18) (-), Anderson-Rubin (- -), LR (-.-), 2sls t-statistic (..).

while it is (always) finite and symmetric for the 2SLS t-statistic.

k = 5 For a number of instruments that is equal to 5, $k = 5$, figures 5 and 6 show asymptotic p-value plots of the 2SLS t-statistic, statistic (18), the Anderson-Rubin statistic and the Likelihood Ratio statistic that test the hypothesis $H_0 : \beta = \beta_0$ for artificial datasets with $\pi_1 = 1$, figure 5, and $\pi_1 = 0.1$, figure 6.

In both figures the Anderson-Rubin statistic leads to larger asymptotic confidence sets than statistic (18). This results as the degrees of freedom parameter of the limiting distribution of the Anderson-Rubin statistic is equal to the number of instruments for the Anderson-Rubin statistic while it is equal to the number of tested parameters for statistic (18). The figures also show that the confidence sets that result from statistic (18) are larger than those that result from the Likelihood Ratio statistic and that they are closely related as discussed in section 4.

Figure 6 nicely shows that the 95% confidence set that results from the 2SLS t-statistic has hardly any relationship with the other confidence sets anymore when the instruments are weak. The 95% confidence set that results from the 2SLS t-statistic is finite and symmetric while the other statistics lead to discontinuous infinite 95% confidence sets. In case of valid instruments, as in figure 5, the confidence sets that result from the different statistics are quite similar.

k = 20 For a number of instruments that is equal to 20, $k = 20$, figures 7 and 8 show asymptotic p-value plots of the 2SLS t-statistic, statistic (18), the Anderson-Rubin statistic and the Likelihood Ratio statistic that test the hypothesis $H_0 : \beta = \beta_0$ for artificial datasets with $\pi_1 = 1$, figure 7, and $\pi_1 = 0.1$, figure 8.

For the valid instrument case, figure 7, the Anderson-Rubin statistic leads to a much larger 95% confidence set than the other statistics. This results from the larger degrees of freedom parameter of its limiting distribution. Since the 2SLS estimator converges to the least

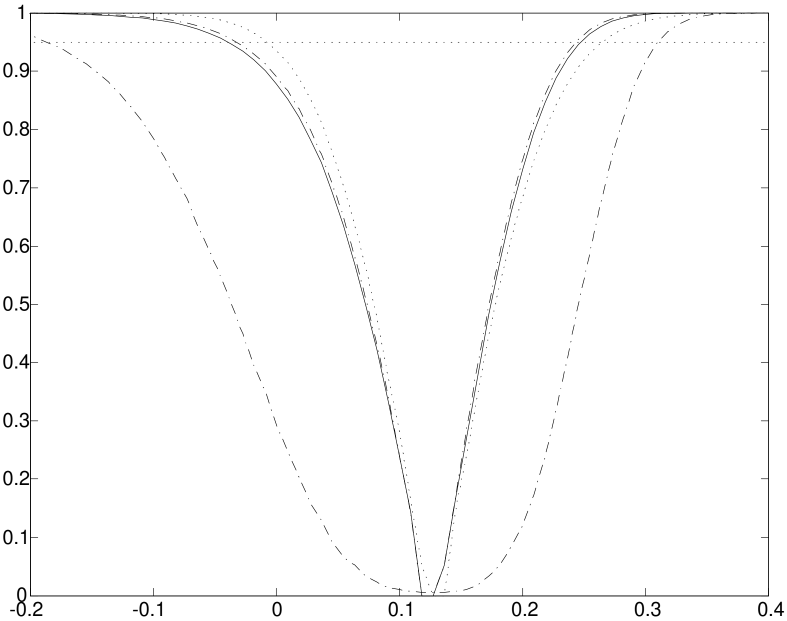

Figure 5: p-value plots of statistics that test the hypothesis $H_0 : \beta = \beta_0$ for dataset with $\pi_1 = 1$, $k = 5$. statistic (18) (-), Anderson-Rubin (- -), LR (-.-), 2sls t-statistic (..).

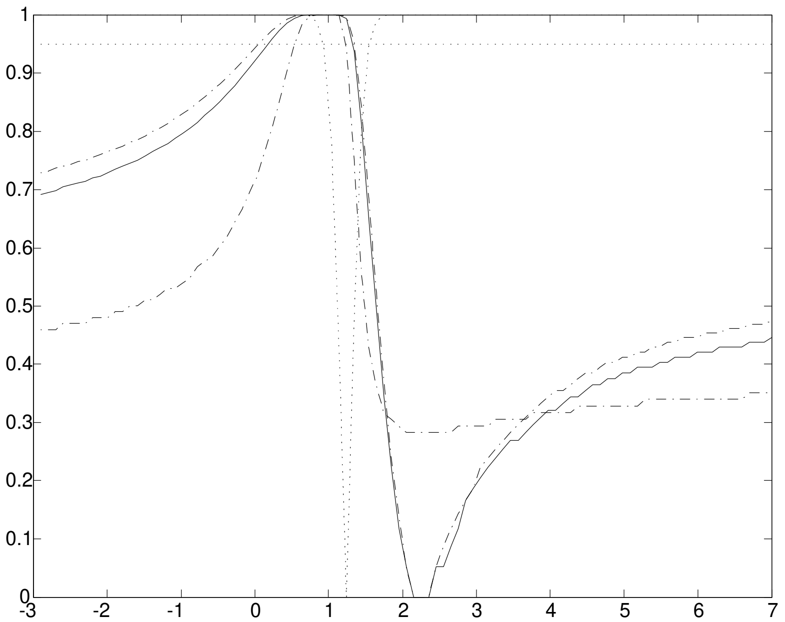

Figure 6: p-value plots of statistics that test the hypothesis $H_0 : \beta = \beta_0$ for dataset with $\pi_1 = 0.1$, $k = 5$. statistic (18) (-), Anderson-Rubin (- -), LR (-.-), 2sls t-statistic (..).

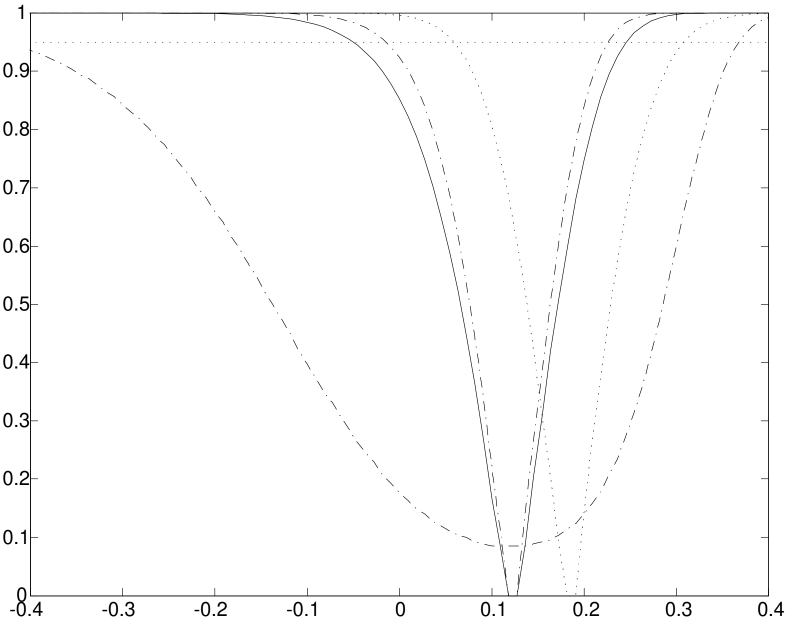

Figure 7: p-value plots of statistics that test the hypothesis $H_0 : \beta = \beta_0$ for dataset with $\pi_1 = 1$, $k = 20$. statistic (18) (-), Anderson-Rubin (- -), LR (-.-), 2sls t-statistic (..).

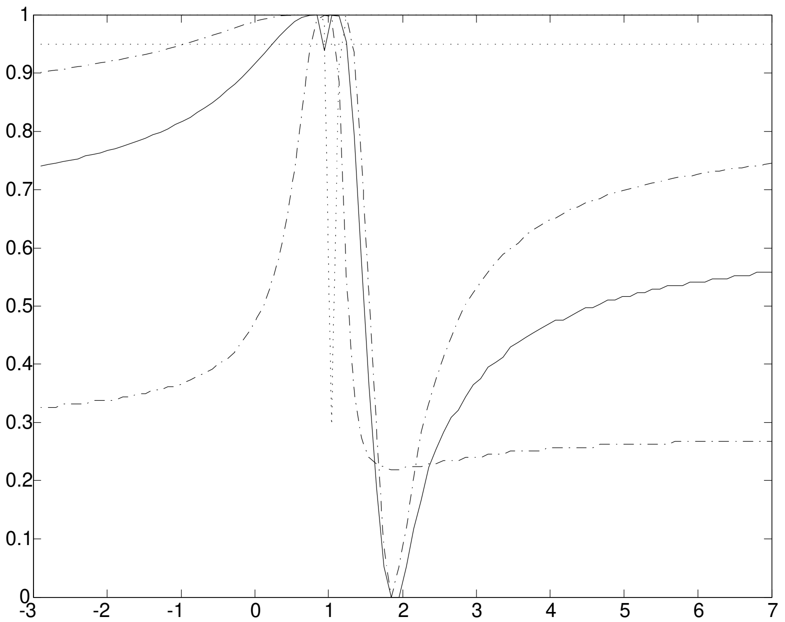

Figure 8: p-value plots of statistics that test the hypothesis $H_0 : \beta = \beta_0$ for dataset with $\pi_1 = 0.1$, $k = 20$. statistic (18) (-), Anderson-Rubin (- -), LR (-.-), 2sls t-statistic (..).

squares estimator when the number of instruments increases, see Nelson and Startz (1990), the difference between the liml estimator and the 2SLS estimator has increased compared to the previous figures. This can be concluded from the areas with zero p-values of the different p-value plots. Also the difference between the confidence sets that result from statistic (18) and the Likelihood Ratio statistic has increased compared to the previous figures.

For the weak instrument case, figure 8, there is again a strong difference between the confidence sets that result from the 2SLS t-statistic, finite and symmetric, and the confidence sets that result from the other procedures, discontinuous and infinite. The confidence sets that result from statistic (18) are also distinctly larger than those that result from the Likelihood Ratio statistic but smaller, as expected, than those that result from the Anderson-Rubin statistic. Also a small decrease of the p-values from statistic (18) occurs at the location of the non-significant p-values from the 2SLS t-statistic.

When we compare the p-value plots from figures 3-4 with those from figures 5-8, we see the robustness of the statistical inference from the different statistics to adding superfluous instruments. The p-value plots from statistic (18) are clearly the most robust while those from the 2SLS t-statistics are the least robust.

6 Power Comparison

Using DGP (29), we conducted a power comparison of a few different statistics that can be used to test the hypothesis $H_0 : \beta = 0$. We therefore simulated datasets from DGP (29) for various values of β and computed the frequency of rejecting H_0 using the 95% asymptotic critical value of the statistic under H_0. The involved statistics are (18), the Anderson-Rubin statistic, the Likelihood Ratio statistic and the 2SLS t-statistic. The instruments are kept fixed over the (5000) simulations and we used the same values of k and π_1 as in figures 3-8, *i.e.* $k = 1$, 5 and 20 and $\pi_1 = 1$ and 0.1.

$\mathbf{k = 1}$ Figures 9 and 10 show plots of the rejection frequencies (power curves) of the hypothesis $H_0 : \beta = 0$ for various values of β in DGP (29) where H_0 is tested using (18), the Anderson-Rubin statistic, the Likelihood Ratio statistic and the 2SLS t-statistic. In both figures the estimated model is just-identified, $k = m - 1 = 1$, and $\pi_1 = 1$ in figure 9 while $\pi_1 = 0.1$ in figure 10. Figure 9 is thus the case of a valid instrument while the instrument is a weak one in figure 10. As the model is just-identified, the power curves of the Anderson-Rubin and statistic (18) coincide. Also because of the just-identification, the test on β is a linear test and the Likelihood Ratio statistic is therefore closely related to the Anderson-Rubin statistic.

The power of the Anderson-Rubin statistic, statistic (18) and the Likelihood Ratio statistic are equal to the (asymptotic) size in $\beta = 0$. This doesnot hold for the 2SLS t-statistic. As expected, the different statistics have quite some power for discriminating values of β in the valid instrument case but little in the weak instrument case.

$\mathbf{k = 5}$ Figures 11 and 12 show plots of the rejection frequencies (power curves) of the hypothesis $H_0 : \beta = 0$ for various values of β in DGP (29) where H_0 is tested using (18), the Anderson-Rubin statistic, the Likelihood Ratio statistic and the 2SLS t-statistic. As $k = 5$, the estimated model is over-identified. Figure 9 is the case of a valid instrument, $\pi_1 = 1$, while the instrument is weak in figure 10, $\pi_1 = 0.1$.

Since the model is over-identified, the Anderson-Rubin statistic and statistic (18) no longer coincide. There is then also a difference between these statistics and the Likelihood Ratio

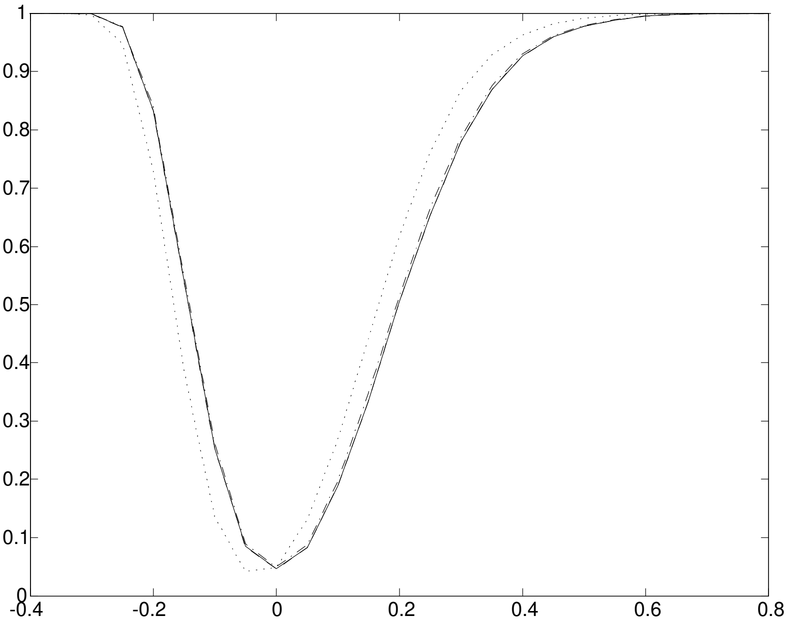

Figure 9: Power curves of statistics that test $H_0 : \beta = 0$ with 5% (asymptotic) significance for various values of β in DGP (29), $k = 1$, $\pi_1 = 1$. Pivotal statistic (-), Anderson-Rubin (- -), Likelihood Ratio (-.-), 2SLS t-statistic (..).

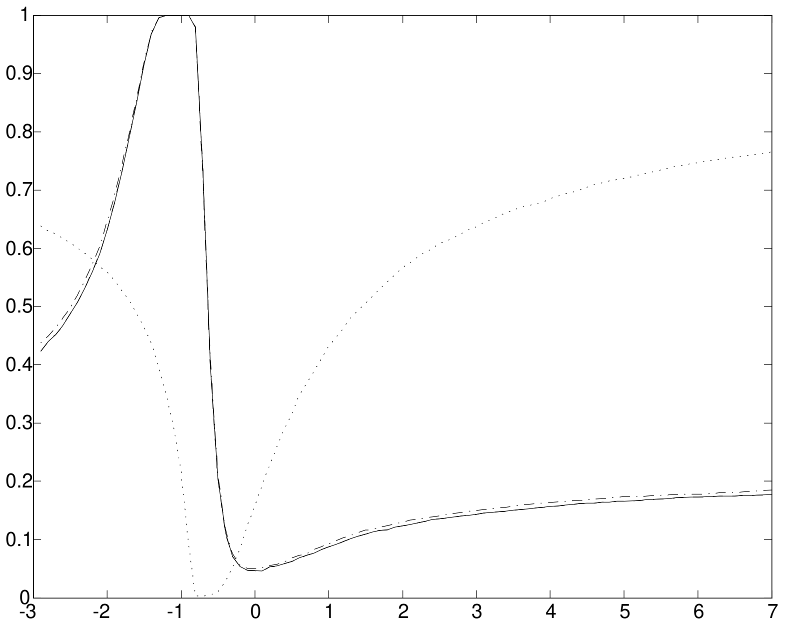

Figure 10: Power curves of statistics that test $H_0 : \beta = 0$ with 5% (asymptotic) significance for various values of β in DGP (29), $k = 1$, $\pi_1 = 0.1$. Pivotal statistic (-), Anderson-Rubin (- -), Likelihood Ratio (-.-), 2SLS t-statistic (..).

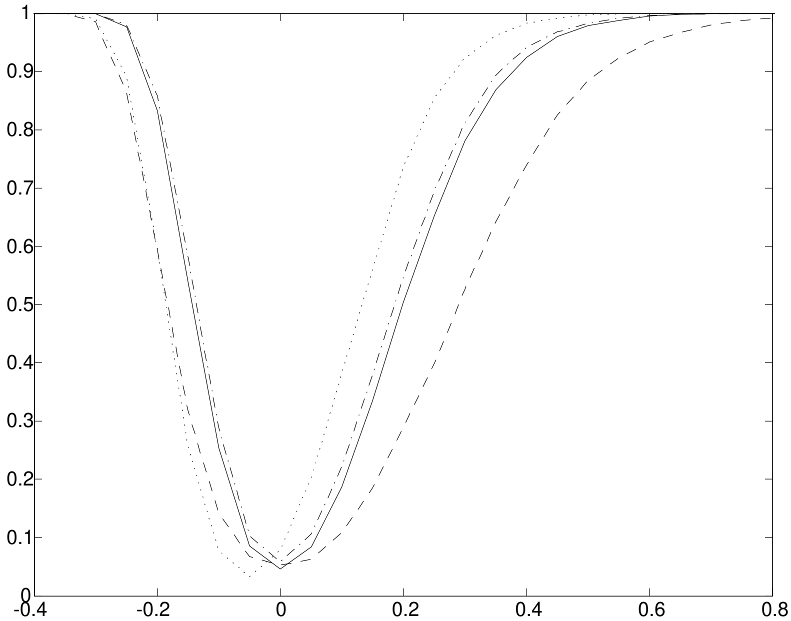

Figure 11: Power curves of statistics that test $H_0 : \beta = 0$ with 5% (asymptotic) significance for various values of β in DGP (29), $k = 5$, $\pi_1 = 1$. Pivotal statistic (-), Anderson-Rubin (- -), Likelihood Ratio (-.-), 2SLS t-statistic (..).

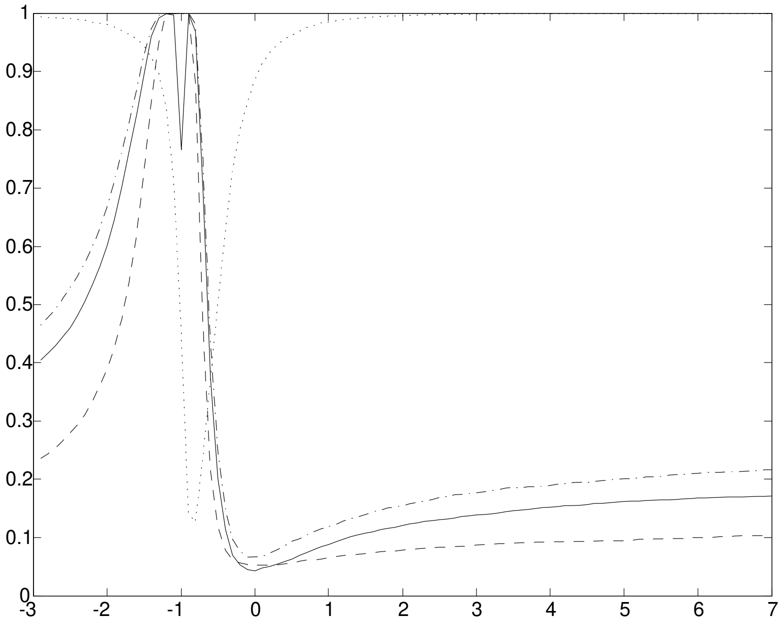

Figure 12: Power curves of statistics that test $H_0 : \beta = 0$ with 5% (asymptotic) significance for various values of β in DGP (29), $k = 5$, $\pi_1 = 0.1$. Pivotal statistic (-), Anderson-Rubin (- -), Likelihood Ratio (-.-), 2SLS t-statistic (..).

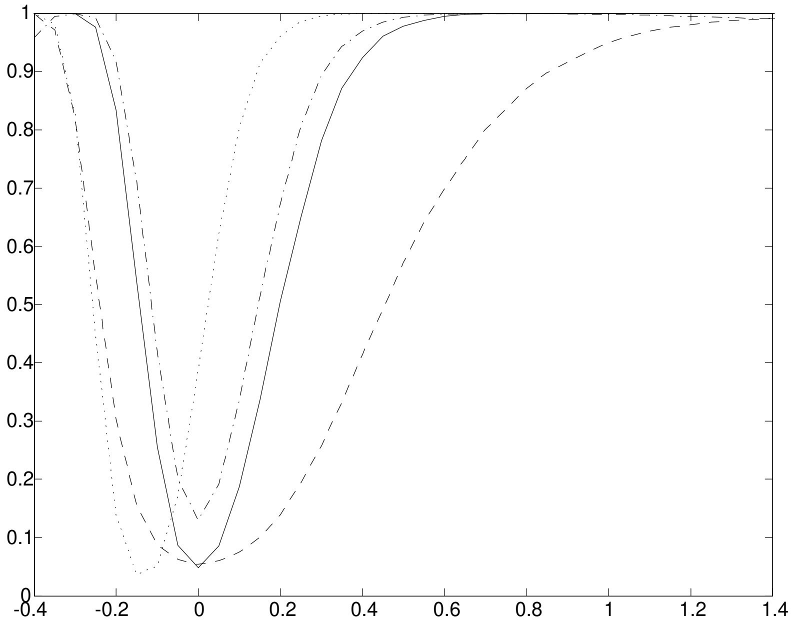

Figure 13: Power curves of statistics that test $H_0 : \beta = 0$ with 5% (asymptotic) significance for various values of β in DGP (29), $k = 20$, $\pi_1 = 1$. Pivotal statistic (-), Anderson-Rubin (- -), Likelihood Ratio (-.-), 2SLS t-statistic (..).

statistic. As the Anderson-Rubin statistic and statistic (18) are asymptotically pivotal, they have the correct size as the rejection frequency is equal to the (asymptotic) size (5%) in $\beta = 0$. The Likelihood Ratio statistic is not asymptotically pivotal and therefore has some (minor) size distortion. The size distortion of the 2SLS t-statistic is very large. Because the degrees of freedom parameter of the limiting distribution of the Anderson-Rubin statistic is larger then the number of parameters in H_0, the Anderson-Rubin statistic has less discriminatory power than statistic (18). Figures 11 and 12 clearly show this difference in discriminatory power. Figure 12 shows that the power curve of (18) has a decrease at the location of the power curve of the 2SLS t-statistic. Figure 12 also shows that the 2SLS t-statistic has a lot of discriminatory power that is completely spurious.

$\mathbf{k = 20}$ Figures 13 and 14 show plots of the rejection frequencies (power curves) of the hypothesis $H_0 : \beta = 0$ for various values of β in DGP (29) where H_0 is tested using (18), the Anderson-Rubin statistic, the Likelihood Ratio statistic and the 2SLS t-statistic. As $k = 20$, the estimated model is over-identified. Figure 13 is the case of a valid instrument, $\pi_1 = 1$, while the instrument is weak in figure 14, $\pi_1 = 0.1$.

Figures 13 and 14 show that the size distortion of the Likelihood Ratio statistic has increased while the Anderson-Rubin statistic and statistic (18) still have the correct asymptotic size. Because the degree of over-identification is substantial, there is also a distinct difference between the power curves of the Anderson-Rubin statistic and statistic (18) while both have the correct size. Figure 14 shows that both the power curves of the Likelihood Ratio and statistic (18) have a decrease at the location of the power curve of the 2SLS t-statistic. For statistic (18), this decrease is also present in figure 12 and in the p-value plot in figure 8. Also notice again the completely spurious power of the 2SLS t-statistic in figure 14. Because the DGPs in figures 9, 11 and 13 and figures 10, 12 and 14 are identical, the power of statistic (18)

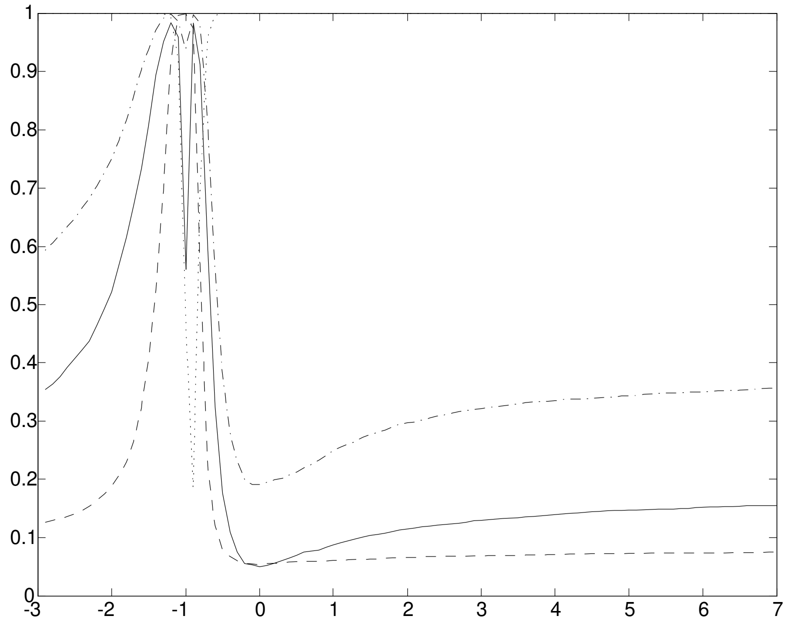

Figure 14: Power curves of statistics that test $H_0 : \beta = 0$ with 5% (asymptotic) significance for various values of β in DGP (29), $k = 20$, $\pi_1 = 0.1$. Pivotal statistic (-), Anderson-Rubin (- -), Likelihood Ratio (-.-), 2SLS t-statistic (..).

is minorly affected by adding superfluous instruments to the estimated model. The power of all the other statistics is seriously influenced by this.

7 Application to the Angrist-Krueger Data

Angrist and Krueger (1991) analyze the return of education on earnings. They use quarter of birth or quarter of birth interacted with other (dummy) variables as instruments in the earnings equation. These quarter of birth related variables can serve as instruments since the quarter of birth is randomly distributed over the population. As a result of the age at which a person enters school and state-dependent compulsory school attendance laws, the quarter of birth does, however, affect the educational attainment. We use the "men born between 1930-39" part of the Angrist and Krueger (1991) dataset which gives us 329.509 observations. This part of the dataset is also analyzed by Staiger and Stock (1997). Our dataset contains five variables, *i.e.* year of birth, state of birth, quarter of birth, years of education and log-earnings, such that we lack observations on the variables: race, standard metropolitan statistical area, region and married, that are also used by Angrist and Krueger (1991) and Staiger and Stock (1997).

The model that is used by Angrist and Krueger (1991) reads:

$$\begin{array}{rllll} w_i = & c_1 + \beta e_i & +\gamma' z_i & & +u_i \\ e_i = & c_2 & +\delta' z_i + \pi' x_i & & +v_i \end{array} \tag{30}$$

where w_i are the log-earnings of individual i, e_i the number of years of education, z_i contains the included exogenous variables and x_i contains the instruments. u_i and v_i are the disturbances. We estimated the parameters of (30) for three different specifications of the included exogenous

Estimation Method\Model specification	I	II	III
2SLS	0.0891	0.0847	0.0899
	(0.0161)	(0.0328)	(0.0107)
LIML	0.0929	0.0981	0.108
	(0.0177)	(0.0465)	(0.0147)
OLS	0.0711	0.0711	0.0673
	(0.003)	(0.0003)	(0.0003)
F-stat	4.91	1.26	1.79
	(0.00)	(0.17)	(0.00)

Table 1: Estimates of β and F-statistic for instrument relavence for three different specifications of the Angrist-Krueger model. (standard errors or p-values (F-statistic) between brackets)

variables and instruments and using three estimation procedures. The results for the parameter β are reported in table 1. Model specification I is such that z_i contains year of birth dummies and x_i contains quarter of birth interacted with year of birth dummies. Model specification II is such that z_i contains year of birth dummies, age, age^2 and x_i contains quarter of birth interacted with year of birth dummies. Model specification III is such that z_i contains year of birth dummies, state of birth dummies, age, age^2 and x_i contains quarter of birth interacted with year of birth dummies and quarter of birth interacted with state of birth dummies. The quarter of birth interacted with state of birth dummy instruments are also present in the analysis of Staiger and Stock while the Angrist and Krueger analysis only contains the quarter of birth interacted with year of birth dummy instruments. Table 1 also reports the F-statistic for instrument relavence which is the standard F-statistic that tests the hypothesis $H_0 : \pi = 0$ in the second equation of (30). The estimates for specification 1 correspond with the estimates that are reported in table 5 of Angrist and Krueger (1991).

Table 1 shows that the F-statistic that indicates instrument relavence is not significant for model specification II. The other F-statistics are also relatively small which shows that the instruments are relatively weak. This implies that we have to interpret the 2SLS and LIML t-statistics with care. We therefore computed asymptotic confidence sets using statistic (18), the Anderson-Rubin statistic, the 2SLS t-statistic and the Likelihood Ratio statistic. Figures 15 to 17 contain the p-value plots of the different model specifications and test statistics. The figures also contain a straight line at 0.95 that enables us to straightforwardly construct the 95% asymptotic confidence set.

The confidence sets that result from statistic (18) nicely show that it attains its minimal value at the liml estimator since the minimal p-values lie at the location of the liml estimator. This of course also holds for the Likelihood Ratio statistic. The p-value plots also show that the 2SLS confidence sets are biased in the direction of the OLS estimator compared to the Likelihood Ratio asymptotic confidence sets and the pivotal confidence sets, see Nelson and Startz (1990). For all model specifications and all values of α, the α% confidence sets from statistic (18) exceed the asymptotic α% Likelihood Ratio based confidence set. For model specification II, the asymptotically pivotal 95% confidence sets that result from statistic (18) and the Anderson-Rubin statistic are both infinite which is not unrealistic given the non-significant value of the F-statistic for instrument relavence. The Likelihood Ratio based asymptotic 95% confidence set is finite for this model specification which shows that there is a distinct difference between statistic (18) and the Likelihood Ratio statistic. The figures also show that the Anderson-Rubin statistic leads to much larger confidence sets than statistic (18)

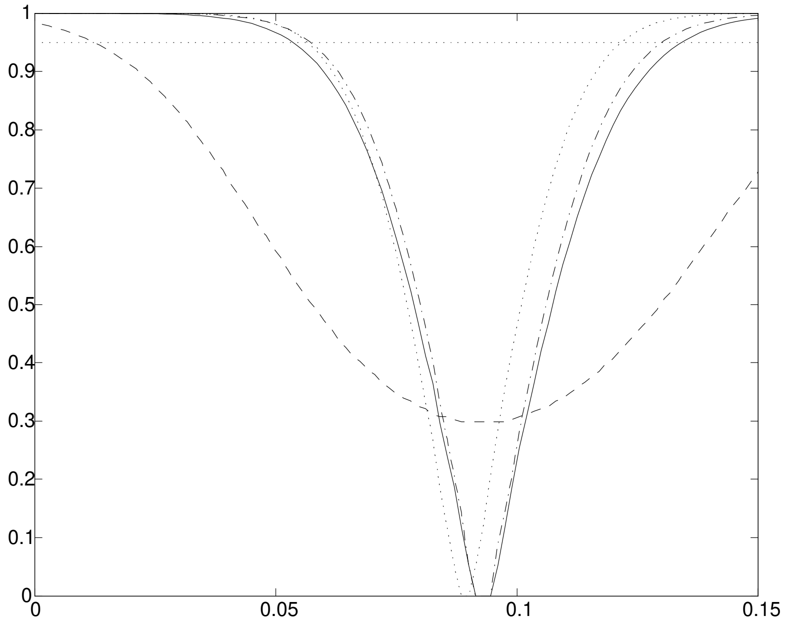

Figure 15: p-value plots of tests of the hypothesis $H_0 : \beta = \beta_0$ for model specification I using 2SLS t-statistic (..), likelihood ratio statistic (-.-), pivotal statistic (18) (-) and the Anderson-Rubin statistic (- -).

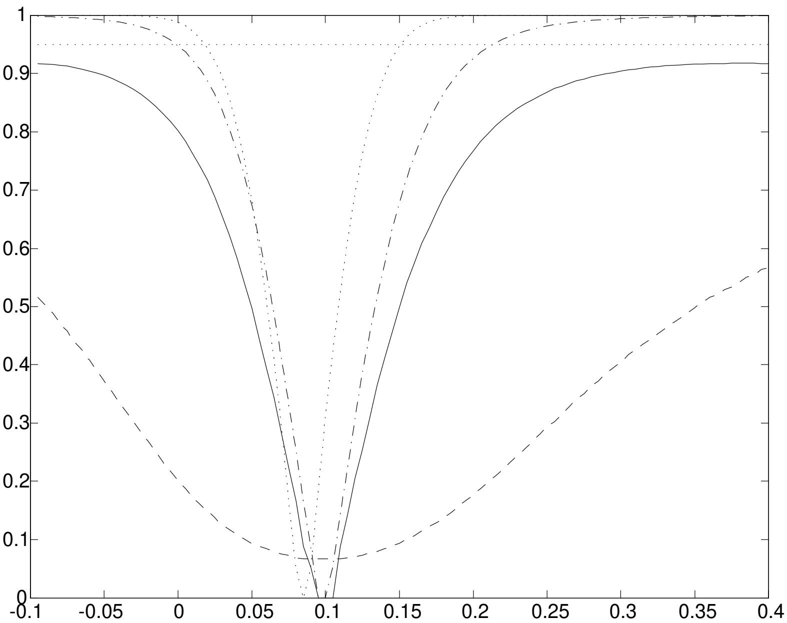

Figure 16: p-value plots of tests of the hypothesis $H_0 : \beta = \beta_0$ for model specification II using 2SLS t-statistic (..), likelihood ratio statistic (-.-), pivotal statistic (18) (-) and the Anderson-Rubin statistic (- -).

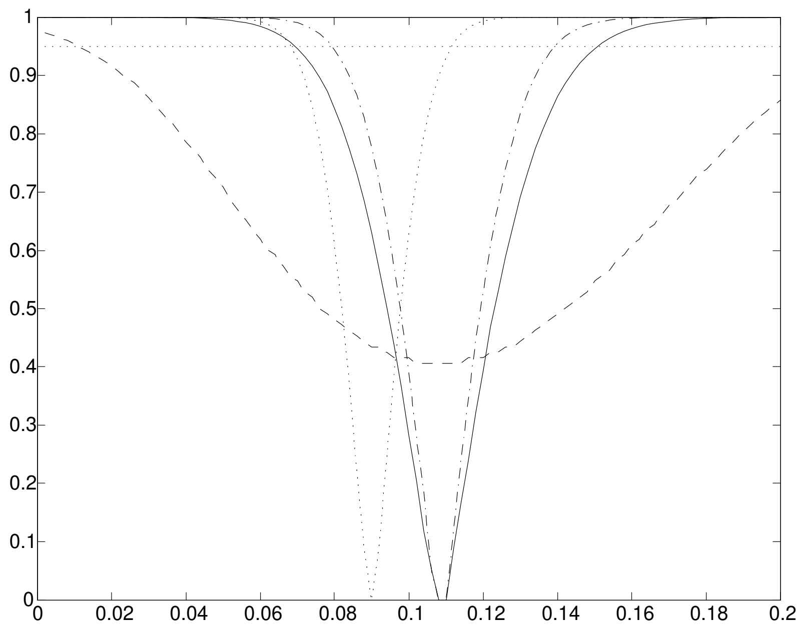

Figure 17: p-value plots of tests of the hypothesis $H_0 : \beta = \beta_0$ for model specification III using 2SLS t-statistic (..), likelihood ratio statistic (-.-), pivotal statistic (18) (-) and the Anderson-Rubin statistic (- -).

which results from the substantial degree of over-identification which is present in all of the estimated models.

The confidence sets that result from statistic (18) in figures 15-17 are smaller then or equal to the 95% Bonferroni confidence sets that are reported by Staiger and Stock (1997). The 95% Bonferroni confidence set is also infinite for specification II (which is specification III in the analysis of Staiger and Stock). The 95% Bonferroni confidence sets have a coverage ratio of at least 95% such that they are larger than or equal to the confidence sets that result from statistic (18) that are constructed here.

8 Conclusions

We developed a statistic for conducting joint tests on the structural parameters in Instrumental variables regression. The statistic is straightforward to compute and has a limiting distribution that is pivotal with a degrees of freedom parameter that is equal to the number of structural parameters. The statistic can be used to construct asymptotically pivotal confidence sets for the structural parameters. These confidence sets can have non-standard shapes. We conducted a power comparison of different statistics for conducting tests on the structural parameters and the novel statistic is favored in all cases. The statistic is applied to the Angrist-Krueger (1991) data for which we obtained similar results as in Staiger and Stock (1997) albeit with a statistic that is less complicated to construct.

The novel statistic, that can be used to conduct a joint test on all structural parameters, results from specifying the maximum likelihood estimator of the parameters of the Instrumental variables regression model as an invertible function of orthogonal statistics, see Kleibergen (2000). In Kleibergen (2000), next to (18), also a statistic that test the hypothesis of over-identification is constructed in this manner. By combining these two statistics, it is then also possible to construct a statistic that can be used to conduct tests on subsets of the structural parameters and which has a pivotal limiting distribution with a degrees of freedom parameter that is equal to the number of tested parameters, see Kleibergen (2000).

References

[1] Anderson, T.W. and H. Rubin. Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, **21**:570–582, (1949).

[2] Anderson, T.W. and T. Sawa. Evaluation of the Distribution Function of the Two Stage Least Squares Estimate. *Econometrica*, **47**:163–182, 1979.

[3] Anderson, T.W., N. Kunitomo, and T. Sawa. Comparison of the Densities of the TSLS and LIMLK estimators for Simultaneous Equations. In F.G. Adams and B.G. Hickman, editor, *Global Econometrics: Essays in Honor of Lawrence R. Klein*, pages 102–124. MIT Press, 1983.

[4] Angrist, J.D. and A.B. Krueger. Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, **106**:979–1014, 1991.

[5] Billingsley, P. *Probability and Measure*. Wiley (New York), (1986).

[6] Dufour, J.-M. Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, **65**:1365–388, 1997.

[7] Dufour, J.-M. and L. Khalaf. Simulation Based Finite and Large Sample Inference Methods in Simultaneous Equations. Working paper, CRDE, Universite de Montreal, 1997.

[8] Engle, R.F., Hendry D.F. and Richard J.-F. Exogeneity. *Econometrica*, **51**:277–304, 1983.

[9] Hausman, J.A. Specification and Estimation of Simultaneous Equations Systems. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, volume 1*. Elsevier Science (Amsterdam), (1983).

[10] Hood, W.C. and T.C. Koopmans. *Studies in Econometric Method*, volume 14 of *Cowles Foundation Monograph*. Wiley (New York), 1953.

[11] Kitamura, Y. Specification Tests with Instrumental Variables and Rank Deficiency. Working paper, Department of Economics, University of Minnesota, 1994.

[12] Kleibergen, F. Exact Test Statistics and Distributions of Maximum Likelihood Estimators that result from Orthogonal Parameters with applications to the Instrumental Variables Regression Model. Tinbergen Institute Discussion Paper TI 2000-039/4, 2000.

[13] Kleibergen F. and E. Zivot. Bayesian and Classical Approaches to Instrumental Variable Regression. Econometric Institute Report 9835/A, 1998.

[14] Mariano, R.S. and T. Sawa. The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables. *Journal of the American Statistical Association*, **67**:159–163, 1972.

[15] Muirhead, R.J. *Aspects of Multivariate Statistical Theory.* John Wiley (New York), 1982.

[16] Nelson, C.R. and R. Startz. Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator. *Econometrica*, **58**:967–976, 1990.

[17] Phillips, P.C.B. Exact Small Sample Theory in the Simultaneous Equations Model. In Z. Griliches and M.D. Intrilligator, editors, *Handbook of Econometrics, Vol.1.* North-Holland Publishing Co., Amsterdam, 1983.

[18] Phillips, P.C.B. Partially Identified Econometric Models. *Econometric Theory*, **5**:181–240, (1989).

[19] Shea, J. Instrument relevance in multivariate linear models: A simple measure. *The Review of Economics and Statistics*, **79**:348–352, 1997.

[20] Staiger, D. and J.H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, **65**:557–586, 1997.

[21] Wang, J. and E. Zivot. Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments. *Econometrica*, **66**:1389–1405, 1998.

[22] Zivot, E., R. Startz, and C. R. Nelson. Valid Confidence Intervals and Inference in the Presence of Weak Instruments. *International Economic Review*, **39**:1119–1144, 1998.