TI 2000-88/4
Tinbergen Institute Discussion Paper

Pivotal Statistics for Testing Subsets of Structural Parameters in the IV Regression Model

Frank R. Kleibergen

Pivotal Statistics for testing Subsets of Structural Parameters in the IV Regression Model

Frank Kleibergen*

Tinbergen Institute Discussion Paper, TI 2000-088/4

October 20, 2000

Abstract

We construct a novel statistic to test hypothezes on subsets of the structural parameters in an Instrumental Variables (IV) regression model. We derive the χ^2 limiting distribution of the statistic and show that it has a degrees of freedom parameter that is equal to the number of structural parameters on which the hypothesis is specified. The statistic has this limiting distribution regardless of the quality of the instruments for the endogenous variables associated with these structural parameters. The instruments have to be valid for the endogenous variables associated with the remaining structural parameters. We analyze the relationship of the novel statistic with the Lagrange Multiplier, the Likelihood Ratio and the GMM over-identification statistic from Stock and Wright (2000). χ^2 limiting distributions for the first two statistics only hold when the instruments are valid for all endogenous variables. A χ^2 limiting distribution for the GMM over-identification statistic is obtained under the same conditions as for our novel statistic but has a larger degrees of freedom parameter. For some artificial datasets, we compute power curves and p-value plots that result from the different statistics. We apply the statistic to an IV regression of education on earnings from Card (1995).

1 Introduction

The quality of the instruments for the endogenous variables in Instrumental Variables (IV) regression is often not clear. When the quality of these instruments is poor, not only the estimates of the structural parameters are inprecise but also the standard limiting distributions do not apply to the estimators. Hence, we can not use the traditional likelihood based statistics, Wald, Likelihood Ratio and Lagrange Multiplier, to conduct tests on the structural parameters in the normal manner, see *e.g.* Staiger and Stock (1997), Dufour (1997) and Wang and Zivot (1998). Two statistics that can still be applied in these cases, as their limiting distributions do not depend on the quality of the instruments, are the Anderson-Rubin statistic, see Anderson and Rubin (1949), and the statistic that is recently proposed in Kleibergen (2000). Both statistics conduct joint tests on all structural parameters but differ in the degrees of freedom parameter of their χ^2 limiting distributions, that are equal to the number of

*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

instruments for the Anderson-Rubin statistic and the number of structural parameters for the statistic from Kleibergen (2000). In case that our hypothesis of interest only concerns a subset of the structural parameters, the degrees of freedom parameters of the limiting distributions of both statistics thus exceed the number of parameters in the hypothesis of interest.

Stock and Wright (2000) show, that in case the hypothesis of interest only concerns a subset of the structural parameters and the instruments are valid for the endogenous variables associated with the remaining structural parameters, that a Generalized Method of Moments (GMM) over-identification statistic has a χ^2 limiting distribution. This limiting distribution has a degrees of freedom parameter equal to the number of instruments minus the number of remaining structural parameters and is not affected by the quality of the instruments for the endogenous variables associated with the structural parameters in the hypothesis of interest. The GMM over-identication statistic is identical to Basmanns' (1960) over-identification statistic evaluated using the maximum likelihood estimator. We propose a novel statistic whose χ^2 limiting distribution applies under the same conditions as the GMM over-identification statistic but has a degrees of freedom parameter that is equal to the number of structural parameters in the hypothesis of interest. This degrees of freedom parameter is therefore less then or equal to the degrees of freedom parameter of the limiting distribution of the GMM over-identification statistic.

The outline of the paper is as follows. In the second section, we introduce the IV regression model. In section 3, we construct our novel statistic to test hypothezes on subsets of the structural form parameters. The statistic results from combining an over-identification statistic and the statistic from Kleibergen (2000). We therefore first briefly discuss these two statistics before we propose the novel one. Section 4 discusses the relationship between the novel statistic and the Lagrange Multiplier statistic. We illustrate the (in)sensitivity of the sampling distributions of different statistics to the quality of the instruments by generating datasets from Data Generating Processes (DGPs) with weak instruments and compute the resulting empirical distribution functions. Section 5 conducts a power comparison of different statistics to test hypothezes on subsets of the structural parameters in case of valid and weak instruments. Section 6 shows the shapes of (asymptotic) p-value plots that can result from these statistics. Section 7 applies the statistics to test the return on education in an IV regression of education on earnings from Card (1995). Next to (the length of) education and (log) wages, this IV regression model also contains two endogenous experience variables for which the instruments are valid. We can therefore apply the novel statistic to test the return on education. Finally, the eigth section concludes.

2 Instrumental Variable Regression Model

The Instrumental Variables (IV) 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

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

which 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

$$\Omega = var \begin{pmatrix} v_{1t} \\ V_{2t} \end{pmatrix} = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} e_1' \\ B \end{pmatrix}' \Sigma \begin{pmatrix} e_1' \\ B \end{pmatrix}, \tag{4}$$

where $e_1 : m \times 1$ is the first m dimensional unity vector. 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 $\Phi : k_2 \times m$. 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 Subset of Structural Parameters

The distributions of the likelihood-based statistics, Wald, Likelihood Ratio and Lagrange Multiplier, to test hypothezes on the structural parameters of the IV regression model depend on nuisance parameters even asymptotically, see *e.g.* Dufour (1997) and Staiger and Stock (1997). In Kleibergen (2000), a statistic for conducting joint tests on all the structural form parameters is constructed whose limiting distribution is pivotal such that it does not depend on nuisance parameters. The degrees of freedom parameter of this limiting distribution is also equal to the number of structural parameters which makes the statistic different from the Anderson-Rubin statistic whose pivotal χ^2 limiting distribution has a degrees of freedom parameter that is equal to the number of instruments. In the following, we construct a statistic to conduct tests on subsets of the structural form parameters. The statistic results from combining the statistic from Kleibergen (2000) with a statistic to test for over-identification. We therefore first briefly discuss the over-identification statistic and the statistic from Kleibergen (2000).

3.1 Over-Identification Statistic

Under the null-hypothesis of over-identification, the reduced form is equal to the restricted reduced form (3) while it is equal to the unrestricted reduced form (5) under the alternative hypothesis. We can therefore reflect the hypothezes as $H_0 : \Phi = \Pi B$ and $H_1 : \Phi \neq \Pi B$. A statistic that can be used to test this hypothesis reads

$$\begin{aligned}
F(H_0|H_1) &= tr\left[\left(S^{-1} - S^{-1}\hat{B}'(\hat{B}S^{-1}\hat{B}')^{-1}\hat{B}S^{-1}\right)Y'(M_{X\hat{\Pi}} - M_X)Y\right] \\
&= tr\left[\hat{B}_\perp'(\hat{B}_\perp S\hat{B}_\perp')^{-1}\hat{B}_\perp Y'(M_{X\hat{\Pi}} - M_X)Y\right] \\
&= \frac{1}{\frac{1}{T-k}\left(y_1 - Y_2\hat{\beta}\right)' M_X\left(y_1 - Y_2\hat{\beta}\right)}\left(y_1 - Y_2\hat{\beta}\right)'(M_{X\hat{\Pi}} - M_X)\left(y_1 - Y_2\hat{\beta}\right) \\
&= tr\left[(\hat{B}_\perp S\hat{B}_\perp')^{-1}\hat{B}_\perp Y'X(X'X)^{-1}\hat{\Pi}_\perp(\hat{\Pi}_\perp'(X'X)^{-1}\hat{\Pi}_\perp)^{-1}\hat{\Pi}_\perp'(X'X)^{-1}X'Y\hat{B}_\perp'\right] \\
&= tr\left[\hat{\lambda}\hat{\lambda}'\right] = \hat{\lambda}'\hat{\lambda},
\end{aligned} \tag{6}$$

where $M_V = I_T - V(V'V)^{-1}V'$ for any $T \times j$ dimensional full rank matrix V,

$$\begin{aligned}
\hat{\lambda} &= (\hat{\Pi}_\perp'(X'X)^{-1}\hat{\Pi}_\perp)^{-\frac{1}{2}}\hat{\Pi}_\perp'(X'X)^{-1}X'Y\hat{B}_\perp'(\hat{B}_\perp S\hat{B}_\perp')^{-\frac{1}{2}}, \\
S &= \tfrac{1}{T-k}Y'M_XY,
\end{aligned} \tag{7}$$

$\hat{\Pi}$, $\hat{B} = \begin{pmatrix} \hat{\beta} & I_{m-1} \end{pmatrix}$ are/contain the maximum likelihood estimators (mles) of β and Π.[1] $\hat{\Pi}_\perp : k \times (k-m+1)$ and $\hat{B}_\perp : 1 \times m$ are the normalized orthogonal complements of $\hat{\Pi}$ and $\hat{B}$ such that $\hat{\Pi}_\perp'\hat{\Pi} \equiv 0$, $\hat{\Pi}_\perp'\hat{\Pi}_\perp \equiv I_{k-m+1}$ and $\hat{B}\hat{B}_\perp' \equiv 0$, $\hat{B}_\perp\hat{B}_\perp' \equiv 1$. The third equation of (6) then results from specifying $\hat{B}_\perp$ as $(1 \; -\hat{\beta}')(1+\hat{\beta}'\hat{\beta})^{-\frac{1}{2}}$. To obtain (6) we have also used that, see *e.g.* Johansen (1991, lemma A.1),

$$\begin{aligned}
S^{-1} - S^{-1}\hat{B}'(\hat{B}S^{-1}\hat{B}')^{-1}\hat{B}S^{-1} &= \hat{B}_\perp'(\hat{B}_\perp S\hat{B}_\perp')^{-1}\hat{B}_\perp \\
Y'(M_{X\hat{\Pi}} - M_X)Y &= Y'X(X'X)^{-1}\hat{\Pi}_\perp(\hat{\Pi}_\perp'(X'X)^{-1}\hat{\Pi}_\perp)^{-1}\hat{\Pi}_\perp'(X'X)^{-1}X'Y.
\end{aligned} \tag{8}$$

[1]The maximum likelihood estimator of $\hat{\Pi}$ is obtained as $\hat{\Pi} = (X'X)^{-1}X'YS^{-1}\hat{B}'(\hat{B}S^{-1}\hat{B}')^{-1}$ and the maximum likelihood estimator of $\hat{\beta}$ is the limited information maximum likelihood estimator.

Under $H_0 : \Phi = \Pi B$ and the least squares estimator of Φ converges as

$$\sqrt{T}\left(\hat{\Phi} - \Pi B\right) \Rightarrow N(0, \Omega \otimes Q^{-1}), \tag{9}$$

where $\hat{\Phi} = (X'X)^{-1}X'Y$, $Q = p\lim_{T\to\infty}\frac{X'X}{T}$ and " $\Rightarrow$ " stands for weak convergence, see Billingsley (1986). To obtain the limiting behavior of (6), we assume that Π has a fixed full rank value such that

$$\sqrt{T}\hat{\Pi}'_{\perp}\hat{\Phi}\hat{B}'_{\perp} \Rightarrow N(0, B_{\perp}\Omega B'_{\perp} \otimes \Pi'_{\perp}Q^{-1}\Pi_{\perp}) \tag{10}$$

and

$$\hat{\lambda} \Rightarrow N(0, I_{k-m+1}), \tag{11}$$

which implies the limiting distribution of (6) under H_0 and a full rank value of Π

$$F(H_0|H_1) \Rightarrow \chi^2(k-m+1). \tag{12}$$

We explicitly construct $\hat{\lambda}$ in the derivation of the limiting distribution of (6) as we combine it with part of the statistic from Kleibergen (2000) that conducts a joint test on all structural parameters.

3.2 Joint Test on all Structural Parameters

In Kleibergen (2000), an asymptotically pivotal statistic for conducting joint tests on all of the structural parameters of the IV regression model is constructed. Before we adapt this statistic, by using the over-identification statistic discussed previously, for use to test hypothezes on subsets of the structural parameters, we first briefly discuss the statistic from Kleibergen (2000).

We consider the null-hypothesis $H_0 : \beta = 0$ under which the limiting distribution of the least squares estimator $\hat{\Phi}$ of the parameters of the unrestricted reduced form is normal

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

By either post-multiplying (13) by R, with

$$R = \begin{pmatrix} 1 & -\omega_{11}^{-1}\omega_{12} \\ 0 & I_{m-1} \end{pmatrix}, \tag{14}$$

or using the marginal and conditional limiting distributions of $\hat{\varphi}_1$ and $\hat{\Phi}_2$, with $\hat{\Phi} = \begin{pmatrix} \hat{\varphi}_1 & \hat{\Phi}_2 \end{pmatrix}$, $\hat{\varphi}_1 : k \times 1$, $\hat{\Phi}_2 : k \times (m-1)$, we obtain

$$\begin{array}{c} \sqrt{T}\left(\begin{pmatrix} \hat{\varphi}_1 & \hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12} \end{pmatrix} - \begin{pmatrix} 0 & \Pi \end{pmatrix}\right) \Rightarrow N(0, \begin{pmatrix} \omega_{11} & 0 \\ 0 & \Omega_{22.1} \end{pmatrix} \otimes Q^{-1}), \\ \Leftrightarrow \\ \left[\begin{array}{rl} \sqrt{T}\hat{\varphi}_1 & \Rightarrow N(0, \omega_{11} \otimes Q^{-1}) \\ \sqrt{T}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12} - \Pi\right) & \Rightarrow N(0, \Omega_{22.1} \otimes Q^{-1}) \end{array}\right], \end{array} \tag{15}$$

where $\Omega_{22.1} = \Omega_{22} - \omega_{21}\omega_{11}^{-1}\omega_{12}$, as

$$R'\Omega R = \begin{pmatrix} \omega_{11} & 0 \\ 0 & \Omega_{22.1} \end{pmatrix}. \tag{16}$$

Since $R'\Omega R$ is block diagonal, $\sqrt{T}\hat{\varphi}_1$ and $\sqrt{T}\left(\hat{\Phi}_2 - \hat{\varphi}_1\omega_{11}^{-1}\omega_{12}\right)$ are asymptotically stochastic independent.

Expression (15) contains the unobserved parameters ω_{11} and ω_{12} that have to be replaced by observable ones. We use the estimator S (7) for this purpose that we specify as $S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & S_{22} \end{pmatrix}$, $s_{11} : 1 \times 1$, $s_{12} = s_{21}' : 1 \times (m-1)$, $S_{22} : (m-1) \times (m-1)$. 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 therefore replace $\omega_{11}^{-1}\omega_{12}$ by $s_{11}^{-1}s_{12}$ in (15) and obtain that

$$\sqrt{T}\left(\hat{\Theta} - \Pi\right) \Rightarrow N(0, \Omega_{22.1} \otimes Q^{-1}), \tag{17}$$

where $\hat{\Theta} = \hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}$ and $\sqrt{T}\hat{\Theta}$ is asymptotically stochastic independent of $\sqrt{T}\hat{\varphi}_1$.[2] Consequently, since s_{11} is a consistent estimator of ω_{11}

$$\left(\hat{\Theta}'X'X\hat{\Theta}\right)^{-\frac{1}{2}} \hat{\Theta}'X'X\hat{\varphi}_1 s_{11}^{-\frac{1}{2}} \Rightarrow N(0, I_{m-1}) \tag{18}$$

and

$$\begin{aligned} F(H_0|H_1) &= \tfrac{1}{(m-1)s_{11}} y_1'X(X'X)^{-1}X'(Y_2 - y_1 s_{11}^{-1}s_{12})\left[(Y_2 - y_1 s_{11}^{-1}s_{12})'X(X'X)^{-1}X'\right. \\ &\quad \left.(Y_2 - y_1 s_{11}^{-1}s_{12})\right]^{-1}(Y_2 - y_1 s_{11}^{-1}s_{12})'X(X'X)^{-1}X'y_1 \\ &= \frac{\hat{\varphi}_1'X'X\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\right)\left(\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\right)'X'X\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\right)\right)^{-1}\left(\hat{\Phi}_2 - \hat{\varphi}_1 s_{11}^{-1}s_{12}\right)'X'X\hat{\varphi}_1}{(m-1)s_{11}} \\ &= \frac{\hat{\varphi}_1'X'X\hat{\Theta}\left(\hat{\Theta}'X'X\hat{\Theta}\right)^{-1}\hat{\Theta}'X'X\hat{\varphi}_1}{(m-1)s_{11}} \\ &\Rightarrow \frac{\chi^2(m-1)}{m-1}, \end{aligned} \tag{19}$$

which shows that the asymptotic distribution of (19) is completely characterized by the parameters under H_0 and does not depend on unobserved nuisance parameters. The difference between the limiting distribution in (19) and the limiting distributions of Likelihood Ratio, Wald and Lagrange Multiplier statistics, is that the limiting distribution (19) 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 (19) 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.

---

[2] (17) results as $\sqrt{T}\hat{\varphi}_1 \Rightarrow N(0, \omega_{11} \otimes Q^{-1})$ 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 $\sqrt{T}\hat{\varphi}_1 \Rightarrow N(0, \omega_{11} \otimes Q^{-1})$, $s_{11}^{-1}s_{12} \Rightarrow \omega_{11}^{-1}\omega_{12}$.

In the just-identified case k is equal to $m-1$ and $\hat{\Theta}$ in (19) is invertible. Statistic (19) and the Anderson-Rubin statistic are then identical. This shows that statistic (19) is a generalization of the Anderson-Rubin statistic from the just-identified to the over-identified case as a statistic to only test H_0. We note also that statistic (19) is invariant to transformations like $(y_1^* \ Y_2^*) = (y_1 \ Y_2)A$, with A a upper (block) triangular matrix, and $X^* = XC$, with C a square non-singular matrix.[3]

By using $y_1^* = y_1 - Y_2\beta_0$ instead of y_1 in all elements of (19) that contain y_1, (19) is an asymptotically pivotal statistic to test $H_0 : \beta = \beta_0$. In this manner (19) can be used to construct confidence sets, see Kleibergen (2000) for more details. In Kleibergen (2000) also the relationship between (19) and the likelihood ratio statistic is discussed and its' power is investigated.

3.3 Test on Subset of Structural Parameters

Statistic (19) can be used to conduct a joint test on all structural parameters and is asymptotically pivotal which implies that its' limiting distribution is not affected by weak instruments, see *e.g.* Staiger and Stock (1997) and Wang and Zivot (1998) for a discussion of the concept of a weak instrument. When we can determine a subset of the endogenous variables for which the instruments are known to be valid, we can also construct a statistic to test hypothezes on the structural parameters of the remaining subset of endogenous variables. The asymptotic distribution of the resulting statistic does not depend on the quality of the instruments for the latter subset of endogenous variables. To construct this statistic we consider the IV regression model

$$\begin{aligned} Y_1 &= X\Pi_1 + V_1, \\ Y_2 &= X\Pi_2 + V_2, \\ y_3 &= Y_1\beta_1 + Y_2\beta_2 + \varepsilon, \end{aligned} \tag{20}$$

where $y_3, \varepsilon : T \times 1$; $Y_2, V_2 : T \times m_2$; $Y_1, V_1 : T \times m_1$, $m = m_1 + m_2 + 1$, $\beta_1 : m_1 \times 1$, $\beta_2 : m_2 \times 1$, $\Pi_1 : k \times m_1$, $\Pi_2 : k \times m_2$, $Y = (Y_1 \ Y_2 \ y_3)$. We assume that the instruments for Y_2 are valid, which implies that Π_2 has a fixed full rank value, but make no assumptions about the quality of the instruments for Y_1, *i.e.* we make no assumptions about Π_1. We are interested in testing the hypothesis $H_0 : \beta_1 = 0$ against the alternative hypothesis $H_1 : \beta_1 \neq 0$. We specify the covariance matrix Ω and its estimator S (7) of the reduced form as,

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \ S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \tag{21}$$

where $\Omega_{11}, S_{11} : m_1 \times m_1$; $\Omega_{12}, \Omega_{21}', S_{12}, S_{21}' : m_1 \times (m_2+1)$; $\Omega_{22}, S_{22} : (m_2+1) \times (m_2+1)$, and assume that the order condition is satisfied such that $k \geq m_1 + m_2$.

The unrestricted reduced form of (20) can be specified as

$$Y = X\Phi + V, \tag{22}$$

[3] The invariance to transformations of X is straightforward to show. To show the invariance to the transformation $(y_1^* \ Y_2^*) = (y_1 \ Y_2)A$, consider that $A = \begin{pmatrix} a_{11} & a_{12} \\ 0 & A_{22} \end{pmatrix}$. This implies that $y_1^* = y_1 a_{11}$ and $Y_2^* = Y_2 A_{22} + y_1 a_{12}$ such that $s_{11}^* = a_{11} s_{11} a_{11}$ and $s_{12}^* = a_{11}(s_{12}A_{22} + s_{11}a_{12})$. As a consequence, $Y_2^* - y_1^* s_{11}^{*-1} s_{12}^* = Y_2 A_{22} + y_1 a_{12} - y_1 a_{11}(a_{11}s_{11}a_{11})^{-1} a_{11}(s_{12}A_{22} + s_{11}a_{12}) = (Y_2 - y_1 s_{11}^{-1} s_{12})A_{22}$. Both a_{11} and A_{22} cancel out in the expression of (19) which shows that (19) is invariant to these transformations.

where $V = (V_1\ V_2\ v_3)$, $v_3 = \varepsilon + V_1\beta_1 + V_2\beta_2$, $\Phi = (\Phi_1\ \Phi_2)$, $\Phi_1 : k \times m_1$, and $\Phi_2 : k \times (m_2+1)$.

Under $H_0 : \beta_1 = 0$, the limiting distribution of the least squares estimator $\hat{\Phi}$ reads

$$\sqrt{T}\left(\hat{\Phi} - \begin{pmatrix} \Pi_1 & \Pi_2 B_2 \end{pmatrix}\right) \Rightarrow N(0, \Omega \otimes Q^{-1}), \tag{23}$$

where $B_2 = \begin{pmatrix} I_{m_2} & \beta_2 \end{pmatrix}$. When we post-multiply (23) by

$$R = \begin{pmatrix} I_{m_1} & 0 \\ -\Omega_{22}^{-1}\Omega_{21} & I_{m_2+1} \end{pmatrix}, \tag{24}$$

the limiting expression becomes

$$\begin{array}{c} \sqrt{T}\left(\begin{pmatrix} \hat{\Phi}_1 - \hat{\Phi}_2\Omega_{22}^{-1}\Omega_{21} & \hat{\Phi}_2 \end{pmatrix} - \begin{pmatrix} \Pi_1 - \Pi_2 B_2 \Omega_{22}^{-1}\Omega_{21} & \Pi_2 B_2 \end{pmatrix}\right) \\ \Rightarrow N(0, \begin{pmatrix} \Omega_{11.2} & 0 \\ 0 & \Omega_{22} \end{pmatrix} \otimes Q^{-1}) \\ \Leftrightarrow \\ \left[\begin{array}{ll} \sqrt{T}\left(\left(\hat{\Phi}_1 - \hat{\Phi}_2\Omega_{22}^{-1}\Omega_{21}\right) - \left(\Pi_1 - \Pi_2 B_2\Omega_{22}^{-1}\Omega_{21}\right)\right) & \Rightarrow N(0, \Omega_{11.2} \otimes Q^{-1}) \\ \sqrt{T}\left(\hat{\Phi}_2 - \Pi_2 B_2\right) & \Rightarrow N(0, \Omega_{22} \otimes Q^{-1}) \end{array}\right], \end{array} \tag{25}$$

where $\hat{\Phi}_1 = (X'X)^{-1}X'Y_1$, $\hat{\Phi}_2 = (X'X)^{-1}X'(Y_2\ y_3)$, $\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$. Equation (25) shows that, under H_0, $\hat{\Phi}_1 - \hat{\Phi}_2\Omega_{22}^{-1}\Omega_{21}$ and $\hat{\Phi}_2$ are asymptotically stochastic independent. This asymptotic stochastic independence remains valid when we replace $\Omega_{22}^{-1}\Omega_{21}$ by $S_{22}^{-1}S_{21}$.[4]

Equation (25) shows that under H_0, because $\beta_1 = 0$, we can just consider the IV regression model that consists of the last $m_2 + 1$ equations of (20),

$$\begin{array}{ll} Y_2 & = X\Pi_2 + V_2, \\ y_3 & = Y_2\beta_2 + \varepsilon. \end{array} \tag{26}$$

We can test the hypothesis of over-identification in (26) by comparing the restricted and unrestricted reduced forms of (26). In order to do this, we use statistic $\hat{\lambda}$ (7)

$$\begin{array}{ll} \hat{\lambda} & = (\hat{\Pi}_{2\perp}'(X'X)^{-1}\hat{\Pi}_{2\perp})^{-\frac{1}{2}}\hat{\Pi}_{2\perp}'(X'X)^{-1}X'\begin{pmatrix} Y_2 & y_3 \end{pmatrix}\hat{B}_{2\perp}'(\hat{B}_{2\perp}S_{22}\hat{B}_{2\perp}')^{-\frac{1}{2}} \\ & = (\hat{\Pi}_{2\perp}'(X'X)^{-1}\hat{\Pi}_{2\perp})^{-\frac{1}{2}}\hat{\Pi}_{2\perp}'(X'X)^{-1}X'\left(y_3 - Y_2\hat{\beta}_2\right)\frac{1}{\left(\frac{1}{T-k}\left(y_3 - Y_2\hat{\beta}_2\right)' M_X \left(y_3 - Y_2\hat{\beta}_2\right)\right)^{\frac{1}{2}}}, \end{array} \tag{27}$$

where $\hat{\Pi}_2$ and $\hat{\beta}_2$ are the mles of Π_2 and β_2 under H_0, so in (26). Equation (11) shows that under H_0 and when Π_2 has full rank, such that also $\Phi_2 = \Pi_2 B_2$ has full rank,

$$\hat{\lambda} \Rightarrow N(0, I_{k-m_2}). \tag{28}$$

$\hat{\lambda}$ can be used to test the over-identification when we have imposed that $\beta_1 = 0$. The unrestricted reduced form equation of y_3 encompasses the restricted reduced form equation of y_3

[4] $\hat{\Phi}$ is asymptotic stochastic independent from S. $S_{22}^{-1}S_{21}$ is therefore also asymptotic stochastic independent from $\hat{\Phi}$. As a consequence, $\sqrt{T}\left(\left(\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1}S_{21}\right) - \left(\Pi_1 - \Pi_2 B_2\Omega_{22}^{-1}\Omega_{21}\right)\right) = \sqrt{T}\left(\left(\hat{\Phi}_1 - \hat{\Phi}_2\Omega_{22}^{-1}\Omega_{21}\right) - \left(\Pi_1 - \Pi_2 B_2\Omega_{22}^{-1}\Omega_{21}\right)\right) - \sqrt{T}\hat{\Phi}_2\left(S_{22}^{-1}S_{21} - \Omega_{22}^{-1}\Omega_{21}\right) \Rightarrow \sqrt{T}\left(\left(\hat{\Phi}_1 - \hat{\Phi}_2\Omega_{22}^{-1}\Omega_{21}\right) - \left(\Pi_1 - \Pi_2 B_2\Omega_{22}^{-1}\Omega_{21}\right)\right)$ as $S_{22}^{-1}S_{21} \Rightarrow \Omega_{22}^{-1}\Omega_{21}$, $\sqrt{T}\left(\hat{\Phi}_2 - \Pi_2 B_2\right) \Rightarrow N(0, \Omega_{22} \otimes Q^{-1})$ and $S_{22}^{-1}S_{21}$ and $\sqrt{T}\hat{\Phi}_2$ are asymptotically stochastic independent.

that results from (20) which again encompasses the restricted reduced form equation of y_3 that results from (26). The over-identification statistic that reflects the difference between the unrestricted reduced form and (26) therefore also contains the difference between the restricted reduced form equations of y_3 that result from (20) and (26). To determine this latter difference, we use (25) to project $\hat{\lambda}$ onto an estimate of Π_1 in the direction that is not explained by Π_2. Since we estimate Π_2 beforehand in (26), only the part of Π_1 that does not lie in the direction of Π_2 is identified and can be used to identify β_1. The estimate that we then use is under H_0 asymptotically stochastic independent of $\hat{\lambda}$ and results from $\hat{\Pi}_2$ and $\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1} S_{21}$, which are also under H_0 asymptotically stochastic independent from one another. To conduct the projection, we pre-multiply $\hat{\lambda}$ (27) by

$$\left(\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1} S_{21}\right)' \hat{\Pi}_{2\perp} (\hat{\Pi}_{2\perp}'(X'X)^{-1}\hat{\Pi}_{2\perp})^{-\frac{1}{2}} \tag{29}$$

and obtain the statistic

$$\begin{aligned} \hat{\kappa} = & \left(\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1} S_{21}\right)' \hat{\Pi}_{2\perp} (\hat{\Pi}_{2\perp}'(X'X)^{-1}\hat{\Pi}_{2\perp})^{-1} \hat{\Pi}_{2\perp}' \hat{\Phi}_2 \hat{B}_{2\perp}' (\hat{B}_{2\perp} S_{22} \hat{B}_{2\perp}')^{-\frac{1}{2}} \\ = & \left(Y_1 - \begin{pmatrix} Y_2 & y_3 \end{pmatrix} S_{22}^{-1} S_{21}\right)' \left[M_{X\hat{\Pi}_2} - M_X\right] \begin{pmatrix} Y_2 & y_3 \end{pmatrix} \hat{B}_{2\perp}' (\hat{B}_{2\perp} S_{22} \hat{B}_{2\perp}')^{-\frac{1}{2}}. \end{aligned} \tag{30}$$

Equation (25) shows that, under H_0, $\hat{\Phi}_2$ is a consistent estimator of $\Pi_2 B_2$. Furthermore, we have assumed that Π_2 has full rank such that $\Pi_{2\perp}$ is properly defined. Hence, the limiting distribution of $\hat{\Pi}_{2\perp}' \left(\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1} S_{21}\right)$ can be characterized as

$$\sqrt{T}\left(\hat{\Pi}_{2\perp}'\left(\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1} S_{21}\right) - \Pi_{2\perp}'\Pi_1\right) \Rightarrow N(0, \Omega_{11.2} \otimes \Pi_{2\perp}' Q^{-1} \Pi_{2\perp}), \tag{31}$$

which results as $\hat{\Pi}_{2\perp}$ is a consistent estimator of $\Pi_{2\perp}$, such that $\Pi_2 B_2 \Omega_{22}^{-1} \Omega_{21}$ cancels out of the limiting expression, and since $\hat{\Pi}_{2\perp}$, that results from $\hat{\Phi}_2$, is asymptotically stochastic independent of $\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1} S_{21}$. Equation (31) shows that $\hat{\kappa}$ (30) is indeed a projection of $\hat{\lambda}$ onto an estimate of Π_1 in the direction that is not spanned by Π_2, *i.e.* $\Pi_{2\perp}'\Pi_1$.

We further normalize $\hat{\kappa}$ and then obtain the statistic

$$\begin{aligned} \hat{\mu} = & \left[\left(\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1} S_{21}\right)' \hat{\Pi}_{2\perp} (\hat{\Pi}_{2\perp}'(X'X)^{-1}\hat{\Pi}_{2\perp})^{-1} \hat{\Pi}_{2\perp}' \left(\hat{\Phi}_1 - \hat{\Phi}_2 S_{22}^{-1} S_{21}\right)\right]^{-\frac{1}{2}} \hat{\kappa} \\ = & \left[\left(Y_1 - \begin{pmatrix} Y_2 & y_3 \end{pmatrix} S_{22}^{-1} S_{21}\right)' \left[M_{X\hat{\Pi}_2} - M_X\right] \left(Y_1 - \begin{pmatrix} Y_2 & y_3 \end{pmatrix} S_{22}^{-1} S_{21}\right)\right]^{-\frac{1}{2}} \\ & \left(Y_1 - \begin{pmatrix} Y_2 & y_3 \end{pmatrix} S_{22}^{-1} S_{21}\right)' \left[M_{X\hat{\Pi}_2} - M_X\right] \left(y_3 - Y_2\hat{\beta}_2\right) \frac{1}{\left(\frac{1}{T-k}\left(y_3 - Y_2\hat{\beta}_2\right)' M_X \left(y_3 - Y_2\hat{\beta}_2\right)\right)^{\frac{1}{2}}}, \end{aligned} \tag{32}$$

which has, under H_0, a limiting distribution that is characterized by

$$\hat{\mu} \Rightarrow N(0, I_{m_1}). \tag{33}$$

The limiting distribution in (33) results as (28) and (31) are asymptotically stochastic independent and the order condition is satisfied in (20) such that $k - m_2 \geq m_1$. The statistic for testing $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ then results as

$$\begin{aligned} F(H_0|H_1) = & \frac{1}{m_1}\hat{\mu}'\hat{\mu} \\ = & \frac{1}{m_1} \frac{1}{\frac{1}{T-k}(y_3 - Y_2\hat{\beta}_2)' M_X (y_3 - Y_2\hat{\beta}_2)} \left(y_3 - Y_2\hat{\beta}_2\right)' \left[M_{X\hat{\Pi}_2} - M_X\right] \left(Y_1 - \begin{pmatrix} Y_2 & y_3 \end{pmatrix} S_{22}^{-1} S_{21}\right) \\ & \left[\left(Y_1 - \begin{pmatrix} Y_2 & y_3 \end{pmatrix} S_{22}^{-1} S_{21}\right)' \left[M_{X\hat{\Pi}_2} - M_X\right] \left(Y_1 - \begin{pmatrix} Y_2 & y_3 \end{pmatrix} S_{22}^{-1} S_{21}\right)\right]^{-1} \\ & \left(Y_1 - \begin{pmatrix} Y_2 & y_3 \end{pmatrix} S_{22}^{-1} S_{21}\right)' \left[M_{X\hat{\Pi}_2} - M_X\right] \left(y_3 - Y_2\hat{\beta}_2\right), \end{aligned} \tag{34}$$

and its' limiting behavior is under H_0 and the valid instrument assumption for Y_2 characterized by

$$F(H_0|H_1) \Rightarrow \frac{\chi^2(m_1)}{m_1}. \tag{35}$$

The limiting distribution (35) does not depend on the quality of the instruments for Y_1 and also its' degrees of freedom parameter is equal to the number of elements of β_1. This is because the limiting distribution (35) results from projecting $\hat{\lambda}$ onto the estimate of Π_1 in the direction of $\Pi_{2\perp}$, under H_0, (29). As (27) is asymptotically stochastic independent from (29), the limiting distribution (35) is not different when the instruments are weak or even invalid for Y_1. Statistic (34) is invariant to transformations of the variables that do offend H_0. Examples of these kind of transformations are $Y^* = (Y_1\ (Y_2\ y_3))A$, with A a lower (block) triangular $m \times m$ matrix, and $X^* = XC$, with C a full rank $k \times k$ matrix.[5]

Statistic (34) results from decomposing the joint statistic (19) which can be used to conduct a test on all structural parameters. The limiting distribution of the joint statistic (19) does not depend on the quality of the instruments. Statistic (34) conducts a conditional test on β_1 given β_2. The relationship on which we condition therefore has to be a valid one such that the instruments for Y_2 have to be relevant. This result is not surprising since when we decompose (19), the irrelavency of the quality of the instruments for the limiting distribution can only be present in one of the parts where we decompose (19) into. In our case this implies that the instruments are allowed to be weak for Y_1 but have to be valid for Y_2.

4 Relationship with Lagrange Multiplier Statistic

Although statistic (34) looks complicated, it is straightforward to compute. By using $y_3^* = y_3 - Y_1\beta_{10}$ instead of y_3 in all elements of (34) that contain y_3, (34) can also be used to test hypothezes for other non-zero values of β_1, $H_0^* : \beta_1 = \beta_{10}$. The expression for (34) then becomes

$$\begin{aligned} F(H_0^*|H_1) = \ & \tfrac{1}{m_1}\tfrac{1}{\frac{1}{T-k}(y_3^*-Y_2\hat{\beta}_2^*)'M_X(y_3^*-Y_2\hat{\beta}_2^*)}\left(y_3^* - Y_2\hat{\beta}_2^*\right)'\left[M_{X\hat{\Pi}_2^*} - M_X\right]\left(Y_1 - \begin{pmatrix} Y_2 & y_3^* \end{pmatrix} S_{22}^{*-1}S_{21}^*\right) \\ & \left[\left(Y_1 - \begin{pmatrix} Y_2 & y_3^* \end{pmatrix} S_{22}^{*-1}S_{21}^*\right)'\left[M_{X\hat{\Pi}_2^*} - M_X\right]\left(Y_1 - \begin{pmatrix} Y_2 & y_3^* \end{pmatrix} S_{22}^{*-1}S_{21}^*\right)\right]^{-1} \\ & \left(Y_1 - \begin{pmatrix} Y_2 & y_3^* \end{pmatrix} S_{22}^{*-1}S_{21}^*\right)'\left[M_{X\hat{\Pi}_2^*} - M_X\right]\left(y_3^* - Y_2\hat{\beta}_2^*\right), \end{aligned} \tag{36}$$

where $\hat{\beta}_2^*$ and $\hat{\Pi}_2^*$ are the mles of β_2 and Π_2 in (14) with y_3 replaced by $y_3^* = y_3 - Y_1\beta_{10}$, $S_{22}^* = \frac{1}{T-k}\begin{pmatrix} Y_2 & y_3 - Y_1\beta_{10} \end{pmatrix}'M_X\begin{pmatrix} Y_2 & y_3 - Y_1\beta_{10} \end{pmatrix}$, $S_{21}^* = \frac{1}{T-k}\begin{pmatrix} Y_2 & y_3 - Y_1\beta_{10} \end{pmatrix}'M_XY_1$. By specifying a grid of values of β_{10}, we can use statistic (36) to construct a (asymptotic pivotal) confidence set for β_1. Note that the (asymptotic pivotal) confidence sets of the parameters of

[5] To show the invariance of (34) to the transformation $Y^* = (Y_1\ (Y_2\ y_3))A$ with A lower block triangular, consider that $A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$. This implies that $Y_1^* = Y_1A_{11} + (Y_2\ y_3)A_{21}$, $(Y_2^*\ y_3^*) = (Y_2\ y_3)A_{22}$ and therefore $S_{21}^* = A_{22}'(S_{22}A_{21}+S_{21}A_{11})$, $S_{22}^* = A_{22}'S_{22}A_{22}$. As a consequence, $Y_1^*-(Y_2^*\ y_3^*)S_{22}^{*-1}S_{21}^* = Y_1A_{11}+(Y_2\ y_3)A_{21} - (Y_2\ y_3)A_{22}(A_{22}'S_{22}A_{22})^{-1}A_{22}'(S_{22}A_{21} + S_{21}A_{11}) = (Y_1 - (Y_2\ y_3)S_{22}^{-1}S_{21})A_{11}$, which is the result that is needed to have invariance of (34) to the transformation $Y^* = (Y_1\ (Y_2\ y_3))A$. The invariance to the transformation $X^* = XC$ can be shown by considering that $\hat{\Pi}_2^* = C^{-1}\hat{\Pi}_2$. As a consequence, $M_{X^*\hat{\Pi}_2^*} - M_{X^*} = M_{X\hat{\Pi}_2} - M_X$.

the IV regression model can be unbounded, discontinuous or empty, see *e.g.* Dufour (1997) and Zivot *et. al.* (1998).

In order to understand statistic (36), we compare it with the Lagrange Multiplier statistic to test the hypothesis $H_0^*: \beta_1 = \beta_{10}$ which has, when both the instruments for Y_1 (!) and Y_2 are valid, a χ^2 limiting distribution with m_1 degrees of freedom. The Lagrange Multiplier Statistic, see *e.g.* Engle (1984), for testing $H_0^*: \beta_1 = \beta_{10}$ against $H_1: \beta_1 \neq \beta_{10}$ results from an auxiliary regression of $X\hat{\Phi}_1$ on the estimated residuals for the equation of y_3 in (26) with y_3 replaced by $y_3^* = y_3 - Y_1\beta_{10}$,

$$\begin{aligned} LM(H_0^*|H_1) &= \tfrac{1}{\frac{1}{T}(y_3^*-Y_2\hat{\beta}_2^*)'(y_3^*-Y_2\hat{\beta}_2^*)}(y_3^* - Y_2\hat{\beta}_2^*)'X\hat{\Phi}_1\left(\hat{\Phi}_1'X'X\hat{\Phi}_1\right)^{-1}\hat{\Phi}_1'X(y_3^*-Y_2\hat{\beta}_2^*) \\ &= \tfrac{1}{\frac{1}{T}(y_3^*-Y_2\hat{\beta}_2^*)'(y_3^*-Y_2\hat{\beta}_2^*)}(y_3^* - Y_2\hat{\beta}_2^*)'X(X'X)^{-1}X'Y_1\,(Y_1'X(X'X)^{-1}X'Y_1)^{-1} \\ &\quad Y_1'X(X'X)^{-1}X(y_3^*-Y_2\hat{\beta}_2^*). \end{aligned} \tag{37}$$

There are two important differences between (37) and (36). The first difference results from (8) which shows, as

$$M_{X\hat{\Pi}_2^*} - M_X \quad = X(X'X)^{-1}\hat{\Pi}_{2\perp}^*(\hat{\Pi}_{2\perp}^{*\prime}(X'X)^{-1}\hat{\Pi}_{2\perp}^*)^{-1}\hat{\Pi}_{2\perp}^{*\prime}(X'X)^{-1}X', \tag{38}$$

that (36) only contains the elements of $((X'X)^{-1}X')\ (y_3^* - Y_2\hat{\beta}_2^*)$ and $((X'X)^{-1}X')\ (Y_1 - \begin{pmatrix} Y_2 & y_3^* \end{pmatrix} S_{22}^{*-1}S_{21}^*)$ that lie in the space orthogonal to $\hat{\Pi}_2^*$. The Lagrange Multiplier statistic (37) uses all components of $((X'X)^{-1}X')\ (y_3^* - Y_2\hat{\beta}_2^*)$ and $((X'X)^{-1}X')\ Y_1$. The second difference is that (36) involves $\left(Y_1 - \begin{pmatrix} Y_2 & y_3^* \end{pmatrix} S_{22}^{*-1}S_{21}^*\right)$ while (37) only involves Y_1. These two differences imply that (36) uses the estimate of $(\Pi_{2\perp}')\,\Pi_1$, $\hat{\Pi}_{2\perp}^{*\prime}(X'X)^{-1}X'\left(Y_1 - \begin{pmatrix} Y_2 & y_3^* \end{pmatrix} S_{22}^{*-1}S_{21}^*\right)$, that is under H_0 asymptotically stochastic independent of $(X'X)^{-1}X'(y_3^* - Y_2\hat{\beta}_2^*)$. The Lagrange Multiplier statistic uses the estimate of Π_1, $(X'X)^{-1}X'Y_1$ which is under H_0 not asymptotically stochastic independent of $(X'X)^{-1}X'(y_3^* - Y_2\hat{\beta}_2^*)$. This implies that the limiting distribution of the Lagrange Multiplier statistic depends, like the other likelihood-based test statistics, Likelihood Ratio and Wald, on nuisance parameters and is not pivotal.

4.1 Empirical Distribution Function

To illustrate the properties of the limiting distribution of (36) compared to the limiting distributions of the Lagrange Multiplier statistic (37) and the likelihood ratio statistic, we computed the empirical distribution functions of these statistics for an IV regression model with strong endogeneity. We therefore simulated data from the model

$$\begin{aligned} y_1 &= x_1\pi_1 + v_1, \\ y_2 &= x_2\pi_2 + v_2, \\ y_3 &= y_1\beta_1 + y_2\beta_2 + \varepsilon, \end{aligned} \tag{39}$$

where y_1, y_2, y_3, x_1, x_2, v_1, v_2, $\varepsilon : T \times 1$; $T = 100$, $\beta_1 = 0$, $\beta_2 = 1$, $\pi_2 = 1$, and x_1 and x_2 are independently generated from a $N(0, I_T)$ distribution and are fixed throughout the simulation experiment. Both m_1 and m_2 are thus equal to one. We generate disturbances $(v_1\ v_2\ \varepsilon)$ from a $N(0, \Sigma \otimes I_T)$ distribution with

$$\Sigma = \begin{pmatrix} 1 & 0.8 & 0.9 \\ 0.8 & 1 & 0.6 \\ 0.9 & 0.6 & 1 \end{pmatrix}. \tag{40}$$

The covariance matrix Σ implies strong endogeneity between y_1, y_2, y_3.

Our first Data Generating Process (DGP) has a value of π_1 equal to one which implies a valid instrument for y_1. Instead of estimating the just-identified model (39), we add three additional nonsense instruments that are independently generated from a $N(0, I_k)$ distribution. We thus estimate a model like

$$\begin{aligned} y_1 &= X\psi_1 + v_1, \\ y_2 &= X\psi_2 + v_2, \\ y_3 &= y_1\beta_1 + y_2\beta_2 + \varepsilon, \end{aligned} \tag{41}$$

with $X = (x_1\ x_2\ X_3)$, $X_3 : T \times (k - m + 1)$; ψ_1, $\psi_2 : k \times 1$, $k = 5$, which has a degree of over-identification equal to three, while the model from which we generate the data is (39). We simulated 5000 datasets from DGP (39) with π_1 equal to 1 and for each simulation we estimated the over-identified model (41). Figure 1 contains the empirical distribution functions based on these simulations for statistic (36), the Lagrange Multiplier statistic (37) and the likelihood ratio statistic that test the hypothesis $H_0 : \beta_1 = 0$. As the instrument for y_1 is a valid one, all empirical distribution functions coincide with the limiting distribution which is also shown.

For figure 2 we have again simulated from DGP (39) but now with a value of π_1 equal to 0.1 which implies that the instrument for y_1 is weak. For the estimation model (41) we have added 18 nonsense instruments, such that $k = 20$, that are independently generated from $N(0, I_k)$ distributions. Figure 2 contains the empirical distribution functions based on these simulations for statistic (36), the Lagrange Multiplier statistic (37) and the likelihood ratio statistic that test the hypothesis $H_0 : \beta_1 = 0$. As the instrument is weak and the degree of over-identification is substantial, both the distributions of the Lagrange Multiplier and the likelihood ratio statistic lie relatively far from their $\chi^2(1)$ limiting distribution. Since the limiting distribution of statistic (36) is pivotal, its' empirical distribution function still coincides with the limiting distribution.

5 Power Comparison

To further investigate statistic (36), we conducted a power comparison of it with a few other statistics. Next to the Lagrange Multiplier statistic (37), we also compare statistic (36) with the likelihood ratio statistic and the GMM objective function statistic that is analyzed in Stock and Wright (2000). Stock and Wright show that the GMM objective function can be used as a statistic to conduct tests on subsets of the structural form parameters and is closely related to the Anderson-Rubin statistic, see Anderson and Rubin (1949). They show that, under the same assumptions that we make to obtain the limiting distribution of (36) of which the valid instrument assumption for Y_2 is the most important one, the GMM objective function as a statistic to conduct tests on β_1 has a limiting distribution that does not depend on the validity of the instruments for Y_1. The difference between the limiting distributions of the GMM objective function statistic and statistic (36) therefore only concerns the degrees of freedom parameter which is equal to the number of elements of β_1 for (36) and the number of elements of β_1 plus the degree of over-identification for the GMM objective function statistic, see Stock and Wright (2000, theorem 3). In our setting, a way to represent the GMM objective function as a statistic to test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ is by means of the over-identification

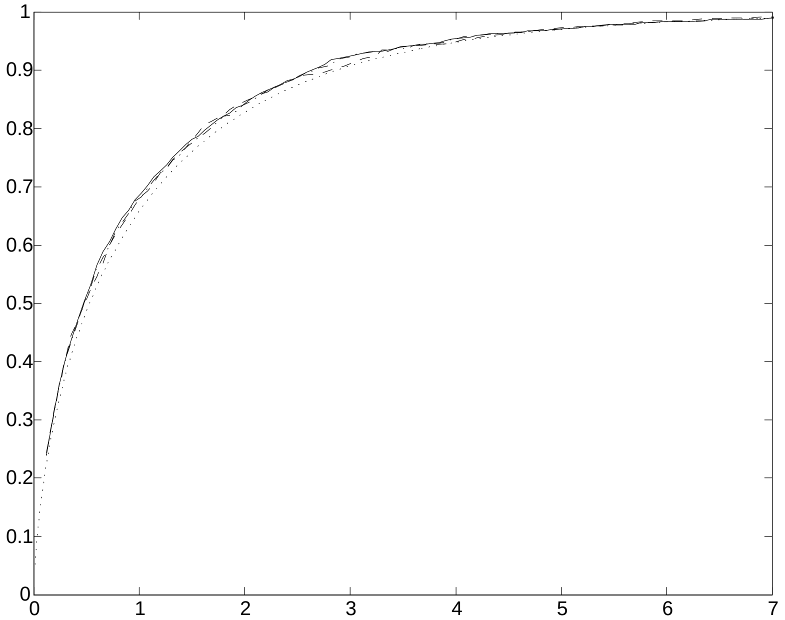

Figure 1: Empirical distribution functions of statistic (36) (-), Lagrange Multiplier statistic (37) (- -), likelihood ratio statistic (-.-), and $\chi^2(1)$ limiting distribution (..) for DGP (39) with $\pi_1 = 1$ while the estimated model is (41) with $k = 5$, $k - m + 1 = 3$.

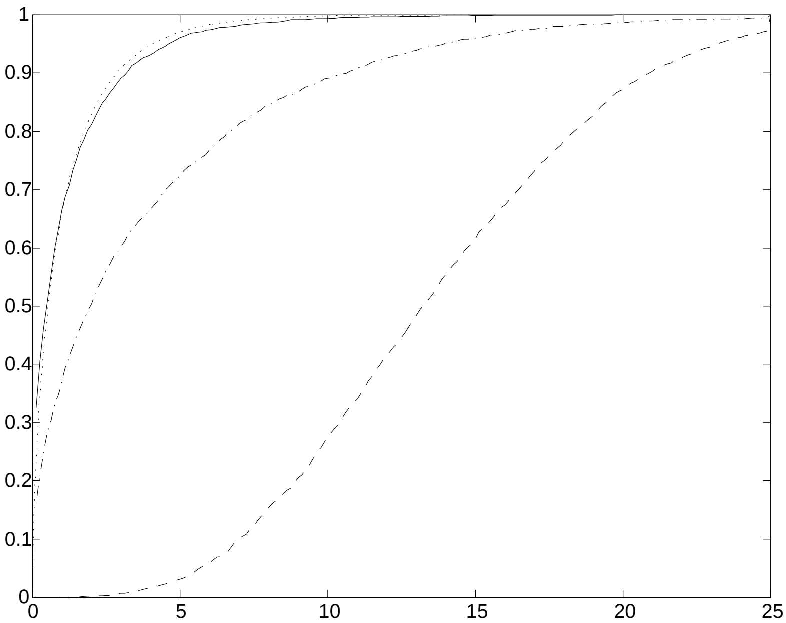

Figure 2: Empirical distribution functions of statistic (36) (-), Lagrange Multiplier statistic (37) (- -), likelihood ratio statistic (-.-), and $\chi^2(1)$ limiting distribution (..) for DGP (39) with $\pi_1 = 0.1$ while the estimated model is (41) with $k = 20$, $k - m + 1 = 18$.

statistic, see Basmann (1960),

$$S(H_0|H_1) = \frac{1}{k-m_2}\frac{1}{\frac{1}{T-k}(y_3 - Y_2\hat{\beta}_2)'M_X(y_3 - Y_2\hat{\beta}_2)}(y_3 - Y_2\hat{\beta}_2)'X(X'X)^{-1}X'(y_3 - Y_2\hat{\beta}_2). \tag{42}$$

Under the assumptions made in section 2 and that Π_2 has full rank, such that the instruments are valid for Y_2, the limiting distribution of $S(H_0|H_1)$ evaluated using the maximum likelihood estimator $\hat{\beta}_2$ is under $H_0 : \beta_1 = 0$ characterized by, see *e.g.* Stock and Wright (2000),

$$S(H_0|H_1) \Rightarrow \frac{\chi^2(l)}{l}, \tag{43}$$

where l is equal to the number of elements of β_1 plus the degree of over-identification, $l = m_1 + (k - m - 1) = m_1 + (k - m_1 - m_2) = k - m_2$. Statistic (42) is identical to the over-identification statistic $\hat{\lambda}'\hat{\lambda}/l$ (6), where $\hat{\lambda}$ results from (27), which can therefore in the same manner as (42) be used to test hypothezes on β_1.[6]

We use DGP (39) with all of its' parameter settings to construct power curves for tests of the hypothesis $H_0 : \beta_1 = 0$ for various values of the true underlying value of β_1. We varied the quality of the instrument(s) for Y_1 in DGP (39) by using values of π_1 equal to 0.1, weak instrument, and 1, valid instrument. In a similar way as in section 4.1, (41) is the estimated model and contains additional nonsense instruments. By adding nonsense instruments, we can analyze the sensitivity of the power curves of the different statistics to the degree of over-identification.

5.1 Valid Instrument, $\pi_1 = 1$

Figures 3 to 5 contain power curves of statistic (36), the over-identification statistic (42), the Lagrange Multiplier statistic (37) and the likelihood ratio statistic that test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ with 5% (asymptotic) significance in case of a valid instrument for Y_1, *i.e.* $\pi_1 = 1$. The estimated model (41) is exactly identified in figure 3 ($k = 2$) while it is over-identified with a degree of over-identification equal to three ($k = 5$) in figure 4 and

[6] The maximum likelihood estimator of β_2 is obtained from an eigenvector of the polynomial equation $|\eta(Y_2\ y_3)'(Y_2\ y_3) - (Y_2\ y_3)'X(X'X)^{-1}X'(Y_2\ y_3)| = 0$, see *e.g.* Hausman (1983). This eigenvector is therefore also an eigenvector of the polynomial equation $|\mu(Y_2\ y_3)'(Y_2\ y_3) - (T-k)(Y_2\ y_3)'S_{22}(Y_2\ y_3)| = 0$ as $S_{22} = \frac{1}{T-k}(Y_2\ y_3)'M_X(Y_2\ y_3)$. Hence, $(Y_2\ y_3)'(Y_2\ y_3)\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix} = \frac{T-k}{\mu}S_{22}\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix}$, for some value of μ. When we construct the maximum likelihood estimator of $\hat{\Pi}_2$ as, in footnote 1 such that, $\hat{\Pi}_2 = (X'X)^{-1}X'(Y_2\ y_3)S_{22}^{-1}\hat{B}_2'(\hat{B}_2S_{22}\hat{B}_2')^{-1}$, it results that $\hat{\Pi}_2'X'(y_3 - Y_2\hat{\beta}_2) = (\hat{B}_2S_{22}\hat{B}_2')^{-1}\hat{B}_2S_{22}^{-1}(Y_2\ y_3)'X(X'X)^{-1}X'(Y_2\ y_3)\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix}$. We can now use that $\hat{B}_2\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix} = 0$, as $\hat{B}_2 = (I_{m_2}\ \hat{\beta}_2)$, and the expression for S_{22}, to obtain $\hat{\Pi}_2'X'(y_3 - Y_2\hat{\beta}_2) = (\hat{B}_2S_{22}\hat{B}_2')^{-1}\hat{B}_2S_{22}^{-1}(Y_2\ y_3)'(Y_2\ y_3)\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix}$. We note that $\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix}$ is such that $(Y_2\ y_3)'(Y_2\ y_3)\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix} = \frac{(T-k)}{\mu}S_{22}\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix}$ and therefore $\hat{\Pi}_2'X'(y_3 - Y_2\hat{\beta}_2) = \frac{(T-k)}{\mu}(\hat{B}_2S_{22}\hat{B}_2')^{-1}\hat{B}_2S_{22}^{-1}S_{22}\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix} = \frac{(T-k)}{\mu}(\hat{B}_2S_{22}\hat{B}_2')^{-1}\hat{B}_2\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix} = 0$ as $\hat{B}_2\begin{pmatrix} -\hat{\beta}_2 \\ 1 \end{pmatrix} = 0$. This implies that $M_{X\hat{\Pi}}(y_3 - Y_2\hat{\beta}_2) = (y_3 - Y_2\hat{\beta}_2)$ and that statistics (42) and (6) coincide. This shows the importance of using the maximum likelihood estimator when we construct the statistics as the equality of (42) and (6) only holds when they are evaluated using the maximum likelihood estimator.

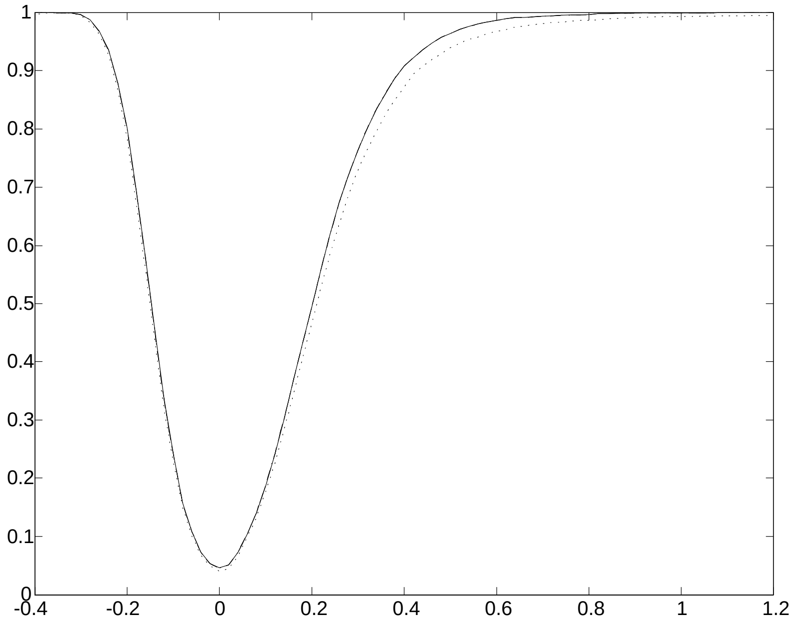

Figure 3: Power curves of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ for various values of the true β_1 in DGP (39) with $\pi_1 = 1$, $k = 2$.

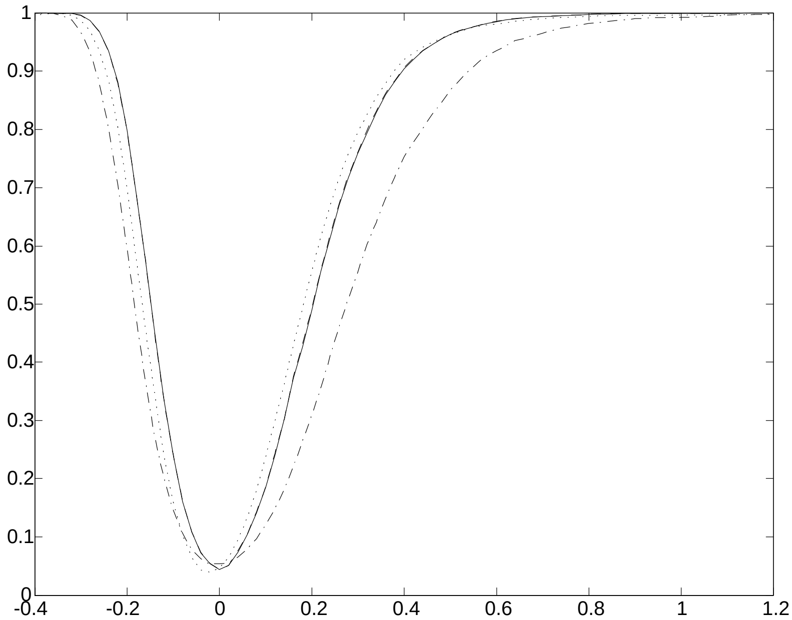

Figure 4: Power curves of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ for various values of the true β_1 in DGP (39) with $\pi_1 = 1$, $k = 5$.

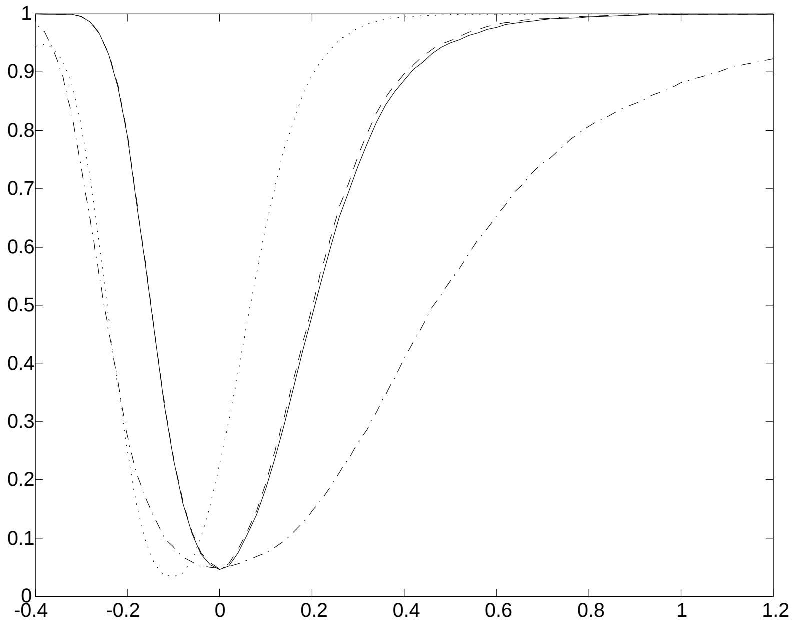

Figure 5: Power curves of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ for various values of the true β_1 in DGP (39) with $\pi_1 = 1$, $k = 20$.

eighteen ($k = 20$) in figure 5. For every value of β_1, we computed the power for testing H_0 using 5000 simulations.

In case of exact identification, $k = 2$, figure 3 shows that the power curves of statistics (36), (42) and the likelihood ratio statistic coincide. The power curve of the Lagrange Multiplier statistic (37) slighty differs from the other power curves. When $k = 5$ such that the degree of over-identification is equal to three, figure 4 shows that the power curves of statistic (36) and the likelihood ratio statistic still coincide. Because of the larger degrees of freedom parameter of its' limiting distribution ($k - 1 = 4$), the power curve of statistic (42) lies below the power curves of the likelihood ratio statistic and statistic (36). Figure 4 also shows that the difference between the power curves of the Lagrange Multiplier statistic (37) and the likelihood ratio statistic and statistic (36) has increased compared to figure 3. Figure 5 shows the power curves in case of a large degree of (nonsense) over-identification ($k - m + 1 = 18$). The power curves of the likelihood ratio statistic and statistic (36) still coincide but the difference with the power curve of statistic (42) has further increased compared to figure 4. This is because of the substantial difference in the degrees of freedom parameters of the limiting distribution of statistic (42), for which it is equal to $k - 1 = 19$, and the likelihood ratio statistic and statistic (36), for which it is equal to 1. Also the difference with the Lagrange Multiplier statistic (37) has increased and it now has a substantial size distortion.

Since the instrument for Y_1 is valid, the likelihood ratio statistic has a χ^2 limiting distribution and the limiting distribution is a good approximation of the finite sample distribution. This explains why the likelihood ratio statistic has the correct (asymptotic) size for all values of k. To analyze the potential size distortion of the likelihood ratio statistic, we therefore in the next simulation experiment use a weak instrument for Y_1.

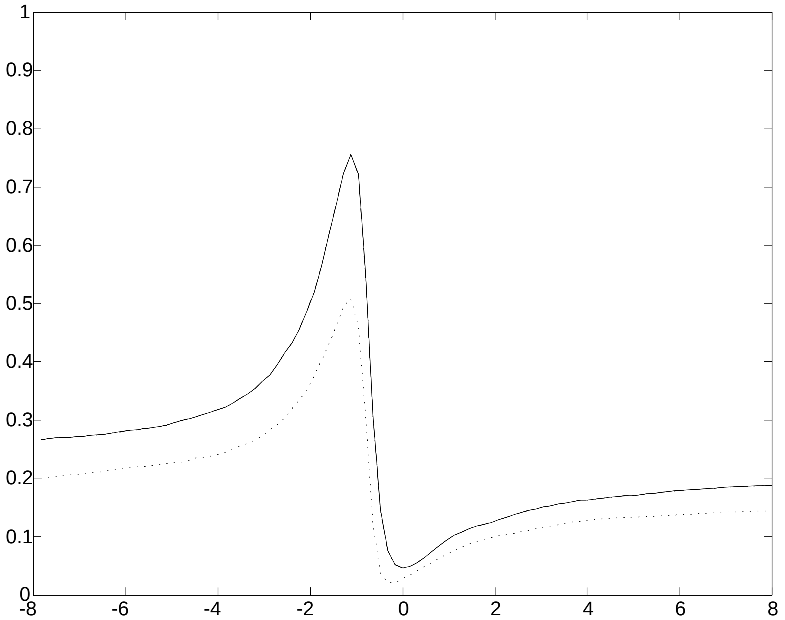

Figure 6: Power curves of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ for various values of the true β_1 in DGP (39) with $\pi_1 = 0.1$, $k = 2$.

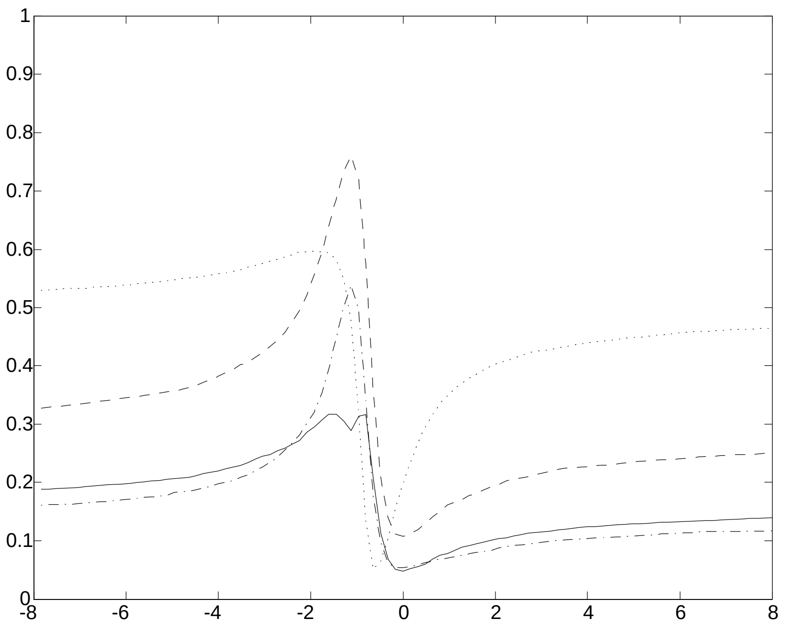

Figure 7: Power curves of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ for various values of the true β_1 in DGP (39) with $\pi_1 = 0.1$, $k = 5$.

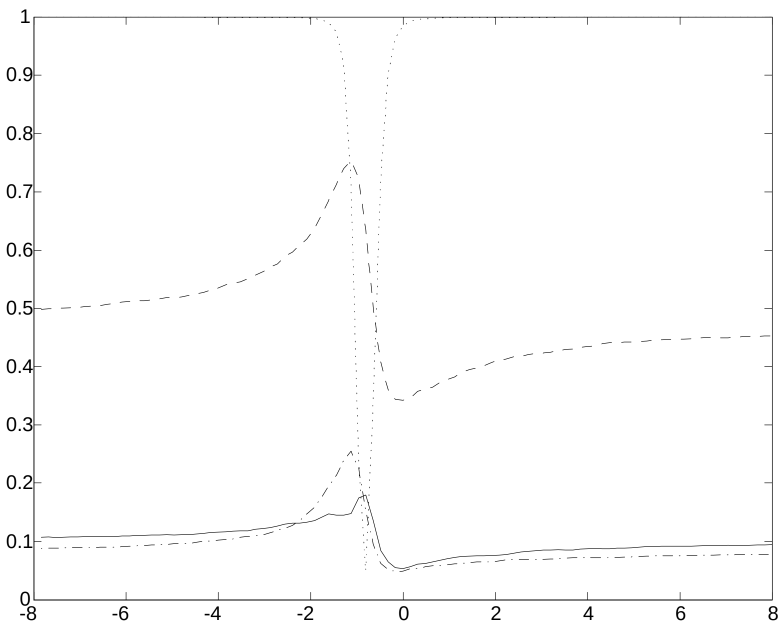

Figure 8: Power curves of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ for various values of the true β_1 in DGP (39) with $\pi_1 = 0.1$, $k = 20$.

5.2 Weak Instrument, $\pi_1 = 0.1$

Figures 6 to 8 contain power curves of statistic (36), the over-identification statistic (42), the Lagrange Multiplier statistic (37) and the likelihood ratio statistic that test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ with 5% (asymptotic) significance in case of a weak instrument for Y_1, *i.e.* $\pi_1 = 0.1$. The estimated model (41) is exactly identified in figure 6 ($k = 2$) while it is over-identified with a degree of over-identification equal to three ($k = 5$) in figure 7 and eighteen ($k = 20$) in figure 8. For every value of β_1, we computed the power for testing H_0 using 5000 simulations.

Figures 6 to 8 show that all test statistics have low power in case of a weak instrument. This already low power further decreases when we add the nonsense instruments. In case of just identification ($k = 2$), the likelihood ratio statistic has the correct (5%) asymptotic size and its' power curve coincides with the power curves of statistics (36) and (42). The size of the likelihood ratio statistic gets distorted when we add nonsense instruments. This size distortion also occurs for the Lagrange Multiplier statistic but not for statistics (36) and (42). The latter two statistics do not get size distorted as they have pivotal limiting distributions. This illustrates the importance of using statistics with pivotal limiting distributions. Note also that in case $k = 20$, figure 8, the Lagrange Multiplier statistic has a lot of spurious power while the likelihood ratio statistic is severely size (larger than 30%) distorted.

6 Confidence Sets

By specifying a grid of values of β_{10}, we can use statistic (36) to construct a confidence set for β_1. The resulting confidence set does (asymptotically) not depend on the value of the other parameters as we use a statistic with a pivotal limiting distribution. These confidence sets

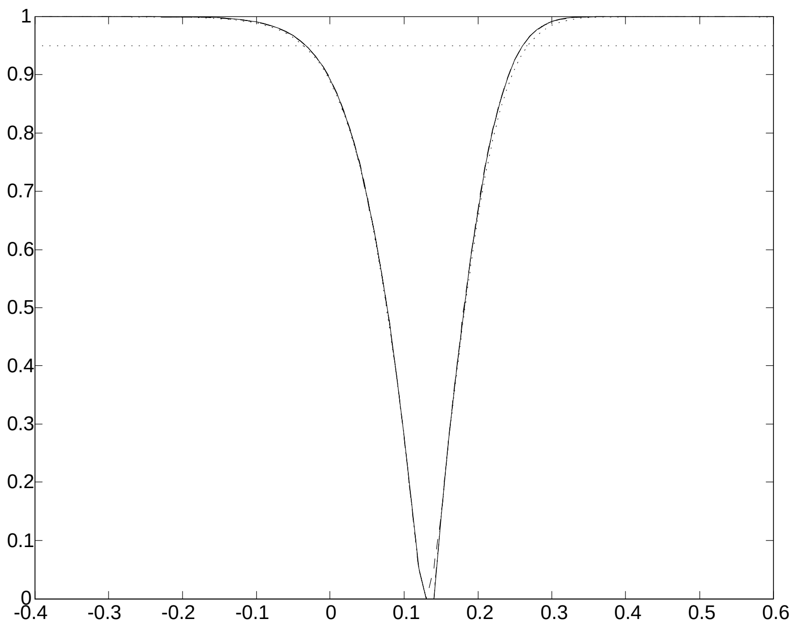

Figure 9: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of the true β_{10} for dataset that is simulated from DGP (39) with $\beta_1 = 0$, $\pi_1 = 1$, $k = 2$.

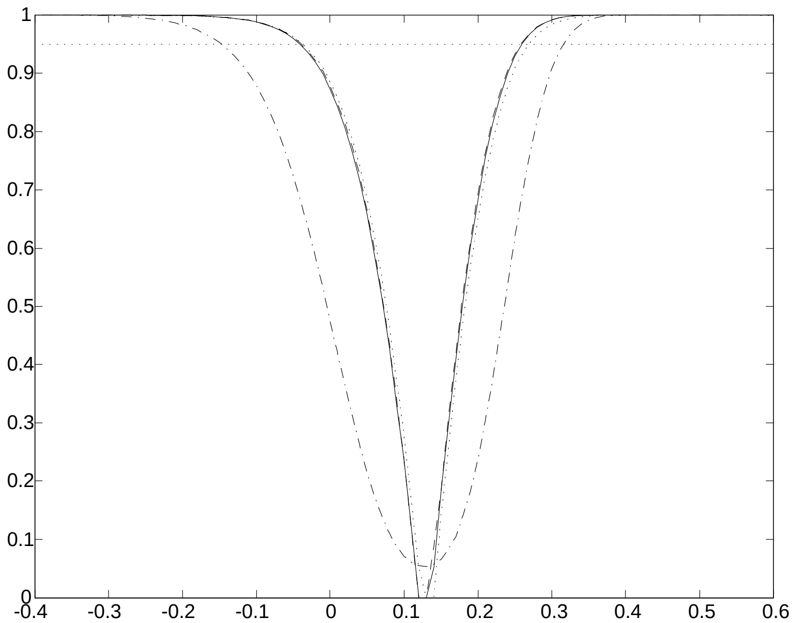

Figure 10: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of the true β_{10} for dataset that is simulated from DGP (39) with $\beta_1 = 0$, $\pi_1 = 1$, $k = 5$.

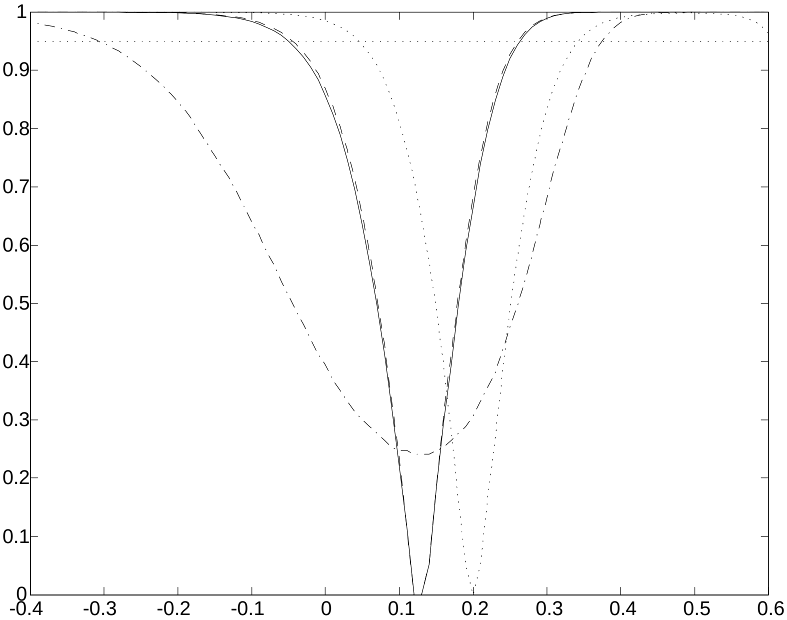

Figure 11: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of the true β_{10} for dataset that is simulated from DGP (39) with $\beta_1 = 0$, $\pi_1 = 1$, $k = 20$.

can have peculiar shapes and can, for example, be infinite, discontinuous or empty, see *e.g.* Dufour (1997).

We use DGP (39) with $\beta_1 = 0$ and the other parameters at their initial settings, such that $\pi_2 = 1$, to generate two datasets. One of these datasets has a valid instrument for Y_1, *i.e.* $\pi_1 = 1$, and one has a weak instrument for Y_1, *i.e.* $\pi_1 = 0.1$. We then add nonsense instruments to the model that we estimate and analyze the sensitivity of the confidence sets to the quality of the instruments and the addition of nonsense instruments.

6.1 Valid Instrument, $\pi_1 = 1$

Figures 9 to 11 contain (asymptotic) p-value plots of statistic (36), the over-identification statistic (42), the Lagrange Multiplier statistic (37) and the likelihood ratio statistic that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of β_{10} in case of a valid instrument for Y_1, *i.e.* $\pi_1 = 1$. The estimated model (41) is just identified in figure 9 ($k = 2$) while it is over-identified with a degree of over-identification equal to three ($k = 5$) in figure 10 and eighteen ($k = 20$) in figure 11. The figures also contain a straight line at 0.95 that enables us to construct the 95% confidence set in a straightforward way.

The p-value plots almost completely coincide in case that $k = 2$, figure 9. When $k = 5$, figure 10 shows that the confidence sets that result from statistic (42) are distinctly larger than the confidence sets that result from the likelihood ratio statistic and statistic (36). This results because of the larger degrees of freedom parameter of the limiting distribution of statistic (42), 4, compared to the degrees of freedom parameter of the limiting distributions of statistic (36) and the likelihood ratio statistic, 1. The difference becomes even more pronounced when $k = 20$. Figure 11 shows that the p-value plot that results from the Lagrange multiplier statistic (37) especially becomes different from the p-value plots from the likelihood ratio

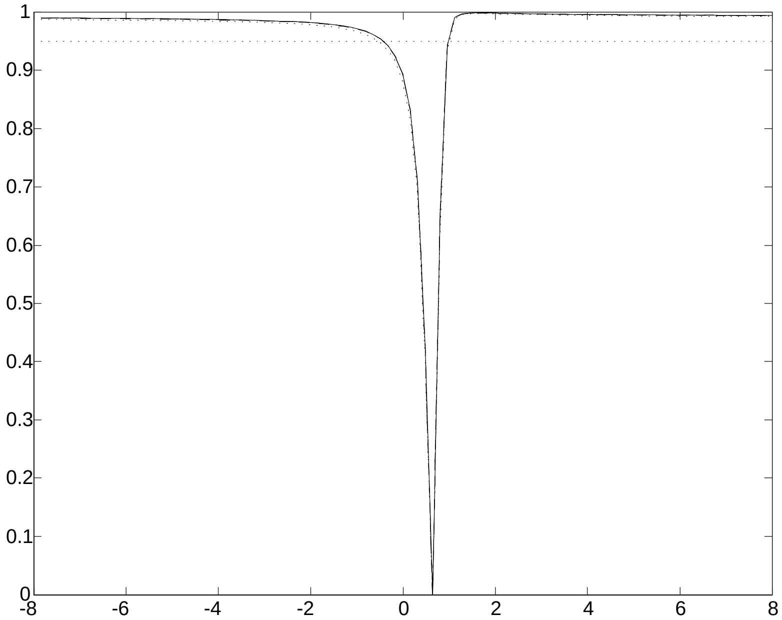

Figure 12: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of the true β_{10} for dataset that is simulated from DGP (39) with $\beta_1 = 0$, $\pi_1 = 0.1$, $k = 2$.

statistic and statistic (36) when the degree of (nonsense) over-identification is large.

Because the instrument for Y_1 is valid, the p-value plots of the likelihood ratio statistic and statistic (36) almost completely coincide for all values of k. A reason for this is that the limiting distributions of (36) and the likelihood ratio statistic coincide in case of valid instruments. This was also found in section 5.1 where the power curves of these statistics almost completely coincided for all values of k in case of a valid instrument for Y_1. We therefore also simulate a dataset from DGP (39) with a weak instrument for Y_1 such that the similarity of the limiting distribution does not apply.

6.2 Weak Instrument, $\pi_1 = 0.1$

Figures 12 to 14 contain (asymptotic) p-value plots of statistic (36), the over-identification statistic (42), the Lagrange Multiplier statistic (37) and the likelihood ratio statistic that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of β_{10} in case of a weak instrument for Y_1, *i.e.* $\pi_1 = 0.1$. The estimated model (41) is just identified in figure 12 ($k = 2$) while it is over-identified with a degree of over-identification equal to three ($k = 5$) in figure 13 and eighteen ($k = 20$) in figure 14. The figures also contain a straight line at 0.95 that enables us to construct the 95% confidence set in a straightforward way.

In case of just-identification, $k = 2$, figure 12 shows that all p-value plots coincide. When $k = 5$, figure 13 shows that the p-value plots become substantially different but still a close relationship between statistic (36) and the likelihood ratio statistic exists. Statistic (36) leads to larger confidence sets than the likelihood ratio statistic and its' p-value plot also has a sudden decrease at $\beta_{10} \approx 1.5$. This decrease becomes even more pronounced when we add more nonsense instruments as shown in figure 14. The p-value plots of the likelihood ratio statistic and statistic (36) in figure 14 are quite different. This results from the large size

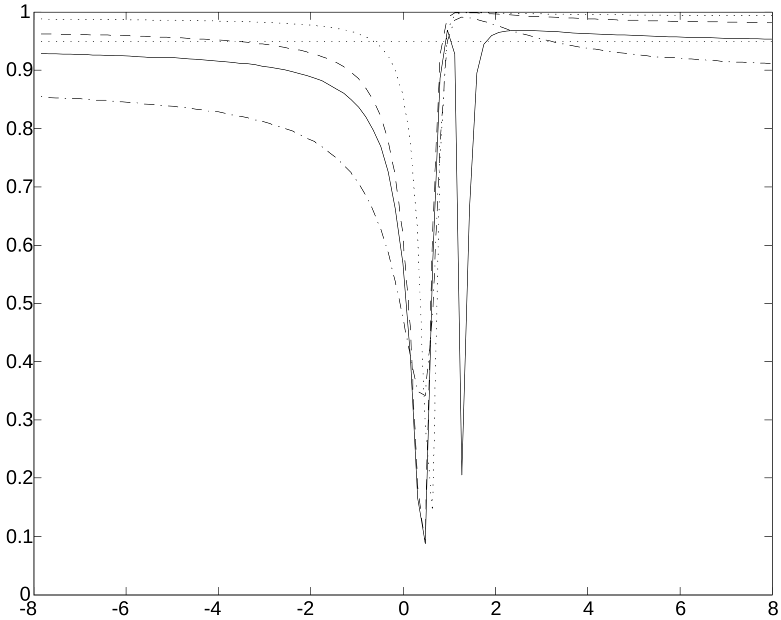

Figure 13: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of the true β_{10} for dataset that is simulated from DGP (39) with $\beta_1 = 0$, $\pi_1 = 0.1$, $k = 5$.

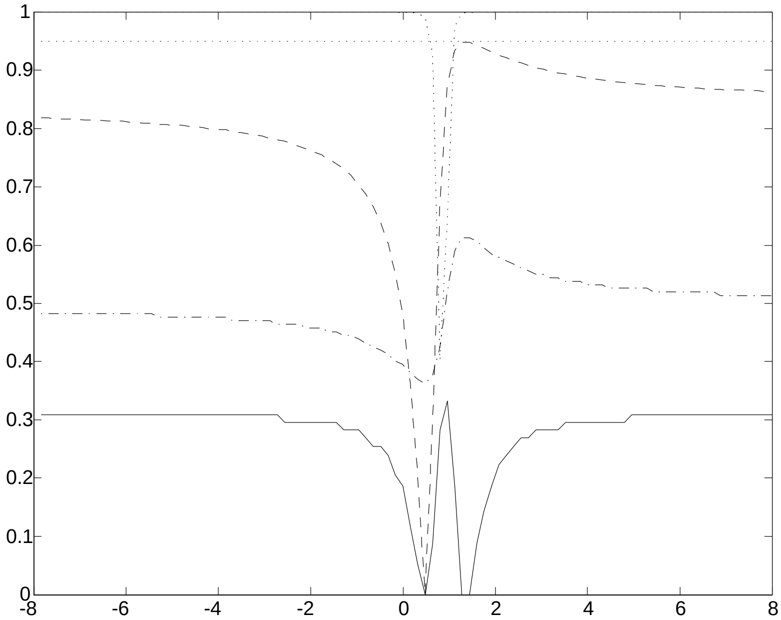

Figure 14: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of the true β_{10} for dataset that is simulated from DGP (39) with $\beta_1 = 0$, $\pi_1 = 0.1$, $k = 20$.

distortion of the likelihood ratio statistic shown in figure 8 for this DGP. The p-value plot of statistic (36) shows that one can not make any statement about the value of β_1 in this DGP which accords with the flat power curve shown in figure 8. The p-value plot of the likelihood ratio statistic gives the false impression that something can be said about the value of β_1. This is even more so for the Lagrange multiplier statistic which even gives the (very mistaken) impression that one can make an accurate statement about the value of β_1. Figure 14 shows the importance of using statistics with pivotal limiting distributions as both statistics (42) and (36) show that no sensible statement about the value of β_1 can be made.

7 Application to Return on Education

Card (1995) analyzes the return of education on earnings. He uses proximity to college as an instrument in an IV regression of (length of) education on (the log) wage. The proximity influences the costs of college education and is therefore directly related to the (length of) education but only indirectly (through the education) to earnings. We use the dataset from Card (1995) and the statistics discussed previously to construct confidence sets for the return on education. The dataset of Card (1995) consists of data obtained from the National Longitudinal Survey of Young Men. This survey started in 1966 and continued untill 1981. We use the 1976 subsample which is a cross-sectional dataset that consists of 3010 observations. Other variables that are contained in the dataset are besides four variables indicating the proximity to college, the length of education, the log-wages, experience, IQ score, age and racial, metropolitan, family and regional indicators. For more details on the dataset we refer to Card (1995).

The model that is used by Card is identical to (20) and reads for individual i

$$\begin{array}{lllll} e_i & = x_i\pi_1 + & & z_i\gamma_1 & +v_{1i} \\ Y_i & = x_i\Pi_2 + & & z_i\Gamma_2 & +V_{2i} \\ w_i & = & e_i\beta_1 + Y_i\beta_2 + z_i\delta & & +\varepsilon_i \end{array} \tag{44}$$

where e_i is the length of education of individual i, $Y_i = (\text{exp}_i\ \text{exp}_i^2)'$ contains the experience (exp) and experienced squared of individual i, $z_i = (1\ \text{race}_i\ \text{smsa}_i\ \text{south}_i)'$ consists of a constant and indicator variables of the race, residence in a metropolitan area and residence in the South part of the United States, w_i is the (logarithm of the) wage of individual i. z_i contains the included exogenous variables and x_i contains the instruments. The instruments in x_i consist of the age and squared age of individual i and a selection of at least one of the four proximity to college indicators. v_{1i}, v_{2i} and ε_i are the disturbances of the model. The IQ score is lateron added to z_i as an additional included exogenous variable.

Estimates of the unrestricted reduced form equations of the endogenous variables included in the equation of w_i in (44) reveal that the R^2 is high for the equations for the experience variables exp_i, and exp_i^2 ($\text{R}^2 \approx 61\%$) and low for the education equation ($\text{R}^2 \approx 12\%$). This results as experience is partly constructed from age (exp_i=age_i-6-e_i). The R^2's show that the instruments are valid for the experience variables and perhaps weak for education. This implies that the assumptions under which statistics (36) and (42), when applied to conduct tests on β_1, have pivotal χ^2 limiting distribution are satisfied since the instruments are valid for the endogenous experience variables.

Table 1 contains estimates of the return on education β_1 for three different specifications of model (44) and three estimation methods. Figures 15 to 17 show p-value plots of statistics that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ for various values of β_{10} for the three different specifications

Instrument\Estimation Method	**OLS**	**2SLS**	**LIML**
age, age^2, indicator for prox to 4 year college	0.074 0.0035	0.133 0.051	0.133 0.051
age, age^2, indicators for prox to 2, 4 and 4 year public college	0.074 0.0035	0.162 0.041	0.18 0.048
age, age^2, indicators for prox to 2, 4 and 4 year public college + IQ score is incorporated as included exogenous variable	0.072 0.0036	0.172 0.054	0.21 0.073

Table 1: Estimates of return on education β_1 (standard error is listed below).

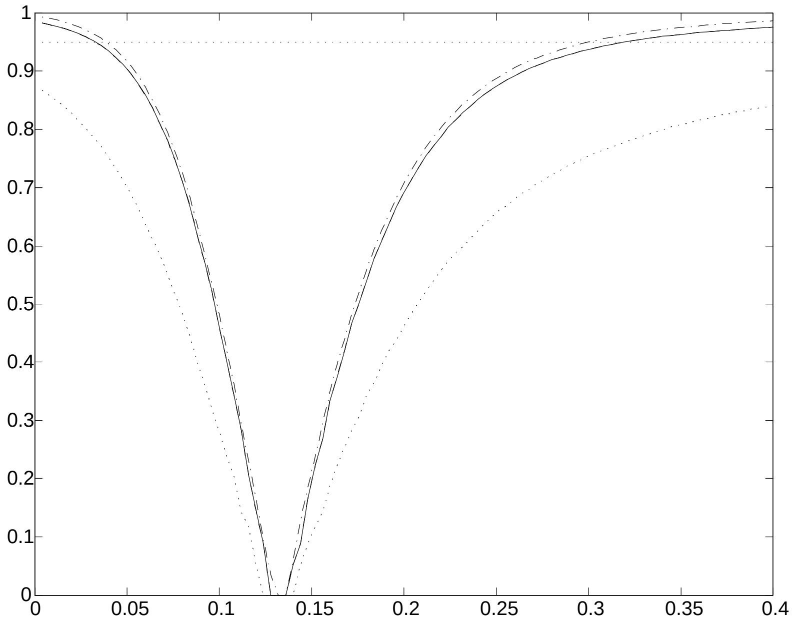

Figure 15: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of β_{10} for return on education dataset from Card (1995). Instruments are age, age^2 and proximity to four year college.

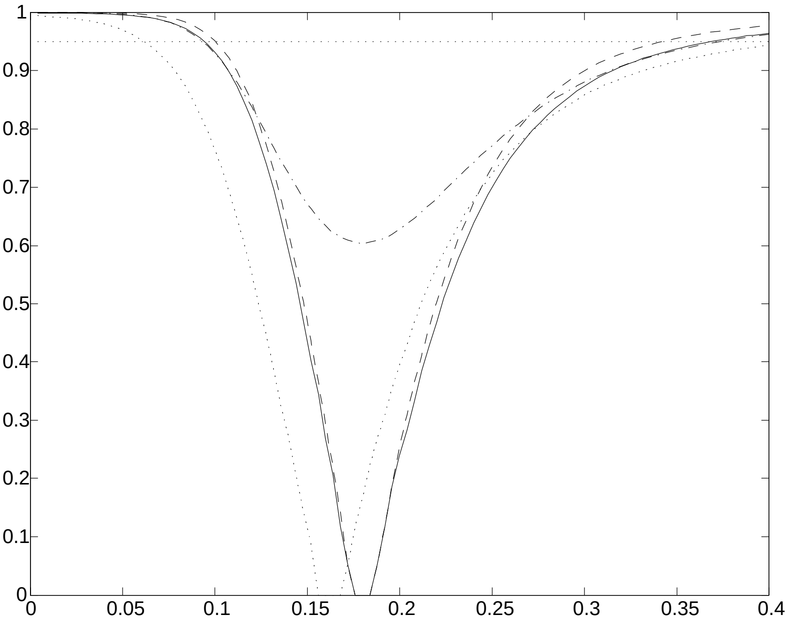

Figure 16: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of β_{10} for return on education dataset from Card (1995). Instruments are age, age^2 and indicators for proximity to two, four and four year public colleges.

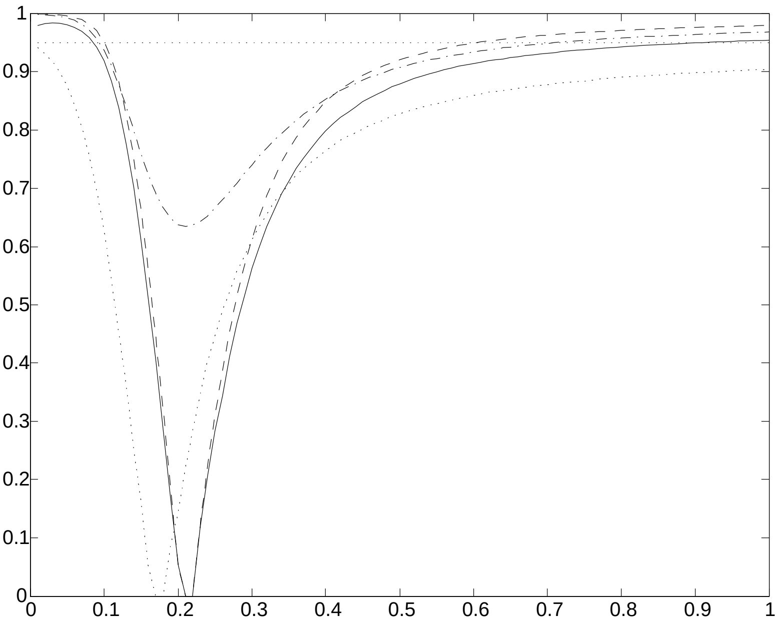

Figure 17: p-value plots of statistics (36) (-), (42) (-.), LM (37) (..) and LR (- -) that test the hypothesis $H_0 : \beta_1 = \beta_{10}$ against $H_1 : \beta_1 \neq \beta_{10}$ for various values of β_{10} for return on education dataset from Card (1995). Instruments are age, age^2 and indicators for proximity to two, four and four year public colleges. The IQ score is incorporated as an exogenous variable in the earnings equation.

of model (44). The first specification in table 1 and figure 15 is just identified as the number of instruments in x_i, 3, is equal to the number of endogenous variables included in the earnings equation. Because of the just identification, the 2SLS and LIML estimates of β_1 and the p-value plots that result from the over-identification statistic (42) and statistic (36) coincide. The p-value plots of the latter two statistics are, because of this just identification, almost identical to the p-value plot of the likelihood ratio statistic.

The OLS estimates of β_1 in the first and second specification of (44) are also identical which results as the models only differ with respect to the instruments and these are not used to construct the OLS estimates. The degree of over-identification in this second specification of (44) is equal to two. Hence, the LIML and 2SLS estimates of β_1 and the p-value plots in figure 16 that result from the over-identification statistic (42) and statistic (36) no longer coincide. Also the difference between the p-value plot of statistic (36) and the likelihood ratio statistic has slightly increased compared to figure 15.

The third specification of (44) includes the IQ score as an exogenous variable in the equation for w_i, so in z_i. With respect to the estimates of β_1, it primarily leads to a change of the LIML estimate but hardly of the OLS and 2SLS estimates. It also leads to an increase of the standard errors. When we evaluate the p-value plots in figure 17, however, we notice a dramatic change. The asymmetry that was already present in the p-value plots in figure 15 and 16 has strongly increased. Also the difference between the p-value plots that result from the Likelihood Ratio statistic and statistic (36) has increased especially with respect to the 95% confidence set that results from them. These 95% confidence sets are also much larger than the ones that result from the standard errors. This shows that the 95% confidence sets that are based on the standard errors will have a too small asymptotic coverage probability as the standard errors under-estimate the uncertainty associated with β_1. This shows that we should use the (pivotal) p-value plots to construct confidence sets. We also note the similarity in figures 15-17 between the p-value plots that result from statistic (36) and the Likelihood Ratio statistic and the substantial difference between these and the confidence sets that are based on the 2SLS t-values that are primarily used in the literature, see *e.g.* Card (1995).

8 Conclusions

We developed a novel statistic to test hypothezes on subsets of the structural parameters in an IV regression model. The statistic has a χ^2 limiting distribution with a degrees of freedom parameter that is equal to the number of tested parameters. The χ^2 limiting distribution is not affected by the validity of the instruments for the endogenous variables that are associated with the structural parameters in the hypothesis of interest. A key assumption is, however, that the instruments are valid for the remaining endogenous variables. Applications of the statistic show that it is closely related to the likelihood ratio statistic in case of valid instruments but can be quite different in case of weak instruments. This results as the standard limiting distribution does then not apply to the likelihood ratio statistic. The limiting distribution of the novel statistic is not affected by the validity of the instruments.

In future work, we will analyze the possiblity of generalizations of the statistic for joint tests on all structural parameters from Kleibergen (2000) and the statistic for subsets introduced in this paper that are applicable in a GMM setting.

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] Basmann, R.L. On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics. *Journal of the American Statistical Association*, **55**:650–659, 1960.

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

[4] Card, D. Using geographic variation in college proximity to estimate the return to schooling. In L.N. Christofides, E.K. Grant and R. Swidinsky, editor, *Aspects of Labour Market Behaviour: essays in honor of John Vanderkamp*, pages 201–222. University of Toronto Press, Toronto, Canada, 1995. (NBER Working Paper 4483 (1993)).

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

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

[7] Engle, R.F. Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, Volume 2.* Elsevier Science (Amsterdam), 1984.

[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] Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, **59**:1551–1580, (1991).

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

[12] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. Tinbergen Institute Discussion Paper TI 2000-055/4, 2000b.

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

[14] 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.

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

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

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

[18] Stock, J.H. and J.H. Wright. GMM with Weak Identification. *Econometrica*, **68**:1055–1096, 2000.

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

[20] 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.