Robust Tests against Weak Instruments for Limited Dependent Variable Models

Leandro Maschietto Magnusson
Department of Economics
Brown University
Leandro_Magnusson@brown.edu
http://www.econ.brown.edu/students/Leandro_Magnusson/

October 17, 2006

Abstract

This paper presents tests for the structural parameter of limited dependent variable models with endogenous explanatory variables that are robust against weak instruments. These tests are derived from the minimum distance objective function. The proposed method has two convenient properties: it is easy to implement and it avoids unnecessary assumptions about the identification of untested parameters. I compare the performances of robust and Wald tests for the sample selection model with an endogenous explanatory variable in a simulation experiment. Tests on female labor supply illustrate this analysis.

JEL classification: C12, C34, C35.
Keywords: Limited dependent variables models, weak instruments, minimum distance estimation and female labor supply.

1 Introduction

In this paper I present tests for limited dependent variable models with endogenous explanatory variables that are robust against the presence of weak instruments. These tests have two main advantages compared to the ones proposed by Kleibergen (2005b): they are easier to implement and they do not require the identification condition of untested parameters.

Consistent estimation requires the fulfillment of the identification conditions. It is hard to obtain general conditions for global identification in nonlinear models. In some cases they are taken as assumption; sometimes they are verified case by case. A necessary condition for global identification is that the expected value of the Jacobian of the moment condition must be a full rank matrix in a small neighborhood of the true parameter (see Newey and McFadden (1994)). The lack of identification can produce biased point estimators. Also, inferences delivered by Wald, Lagrange multiplier (LM) and Distance Metric (DM) tests are misleading (see Stock and Wright (2000)).

"Weak instruments" is the term that denotes the lack of identification of the structure parameter in linear instrumental variables models.[1] The pioneering test that is robust against the presence of weak instruments is the AR-test (see Anderson and Rubin (1949)). Kleibergen (2002) proposes the K-test, a Lagrange multiplier test based on the asymptotic independence between the moment and its Jacobian under the null hypothesis. Moreira (2003) derives the conditional Wald and the conditional likelihood ratio (LR_c) tests, which are not pivotal. His tests partition an invariant sufficient statistic into two independent statistics and condition one statistic on the other. Kleibergen (2005b) extends the K and the LR_c tests to the generalized method of moments models using the objective function of the continuous updating estimator (CUE - see Hansen et al. (1996)).

Kleibergen's tests are difficult to implement for limited dependent variable models with endogenous explanatory variables. Besides the cost of implementation, those tests demand that the untested parameters fulfill the identification condition under the null hypothesis for their consistent estimation.

I derive robust tests from a minimum distance objective function. I first define the unrestricted model, in which all parameters are identified, regardless of whether the structural parameter's identification is valid. Therefore, the unrestricted parameters and their covariance matrix are consistently estimated. Secondly, I define the restriction that maps the structural and unrestricted

[1] Some works that discuss the small sample properties of the IV estimators are Nelson and Startz (1990), Buse (1992) and Bound et al. (1995).

parameters to set up the minimum distance objective function.

This paper is divided into six more sections. In the following section I present the general limited dependent variable model with endogenous explanatory variables and discuss the identification condition of the structural parameter. The third section contains the derivation of the robust tests using the MD approach, followed, in the fourth section, by two applications of them: one for the selection model with an endogenous explanatory variable and another for the endogenous probit. The fifth section investigates the performance of the robust and the Wald tests by simulating their power curves for the selection model with an endogenous explanatory variable. In the sixth section I use the robust tests to construct confidence intervals for the female labor supply study presented by Mroz (1987). The last section is the conclusion. Proofs, mathematical passages, extra graphs and data description are in the appendices. The drawbacks of Kleibergen's tests for this class of models are also in the appendices.

2 Simultaneous Equation Limited Dependent Variable Models: Model and Identification

Limited dependent variable models with endogenous explanatory variables have been studied since the seventies (see Lee (1981) for a review; Newey (1987) and Blundell and Smith (1989); Lee (1996) provides a survey for semi- and nonparametric approach). They depart from the following latent limited information structural model:

$$\begin{cases} Y_i^* = X_i^*\beta + W_i\gamma + U_i \\ X_i^* = Z_i\Pi_z + W_i\Pi_w + V_i \end{cases} \tag{2.1}$$

where Y_i^* and U_i are scalars, X_i^* and V_i are $1 \times m$ vectors of endogenous variables and residuals, W_i is a $1 \times k_w$ vector of included instruments and Z_i is a $1 \times k_z$ vector of excluded instruments. For simplicity, I assume that X_i^* is always observed. Rather than observing Y_i^*, we observe:

$$Y_i = f(Y_i^*, \nu) \tag{2.2}$$

where f is a known function and ν is a $k_\nu \times 1$ vector of additional parameters. This representation is compatible with several limited dependent variables models. Let $\mathbb{1}(\cdot)$ be an indicator function. Under the normality assumption, $Y_i = \mathbb{1}(Y_i^* > 0)$ is the endogenous probit and $Y_i = Y_i^*\mathbb{1}(Y_i^* \in A)$ is the endogenous Tobit, where $A \subset \mathbb{R}$ (see Amemiya (1979), Smith and Blundell (1986) and Rivers and Vuong (1988)). Other examples included in this framework are the linear simultaneous equations, the ordered choice, the Box-Cox transformation and the combination of Box-Cox with limited dependent variable models (see Poirier (1978) and Lankford and Wyckoff (1991)).

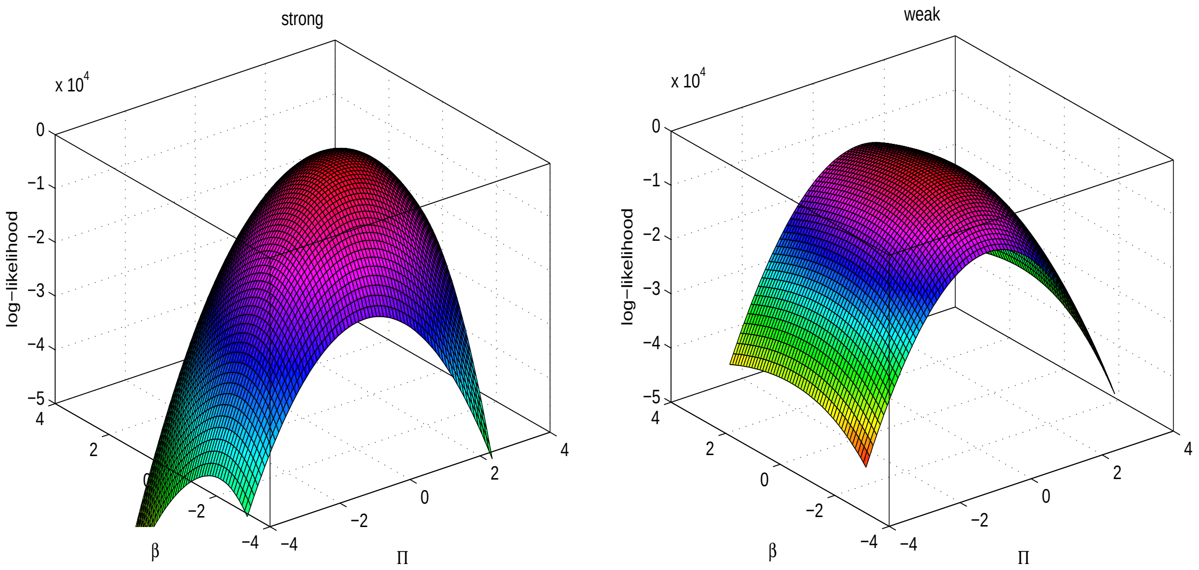
The regular identification condition for the structural parameter β requires that Π_z is a full column rank matrix, i.e, the excluded instruments should be "correlated" with the endogenous variables. Identification fails if there exists a non-full rank matrix in a small neighborhood of Π_z, including the case that Π_z is not itself full ranked. When β is not totally identified, the excluded exogenous variables are labeled as weak instruments. In their presence, estimators for the structural parameter β are inconsistent and the asymptotic properties of the classical tests are invalid since they depend on nuisance parameters (see Stock and Wright (2000)).

Figure 1 illustrates the shape of the log-likelihood function of the endogenous probit model when the instruments are strong and weak. The endogenous probit log-likelihood is:

$$L_n(\beta, \Pi_z, \rho, \sigma_v^2) = \sum_{i=1}^{n} \Bigg\{ y_i \ln\left(\Phi\left(\frac{x_i\beta + (x_i - z_i\Pi_z)}{1-\rho^2}\right)\right) + (1-y_i)\ln\left(1 - \Phi\left(\frac{x_i\beta + (x_i - z_i\Pi_z)}{1-\rho^2}\right)\right)$$
$$- 0.5\ln(2\pi\sigma_v^2) - 0.5\left(\frac{x_i - z_i\Pi}{\sigma_v}\right)^2 \Bigg\}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively. In this example there is no included exogenous variable and only one instrument, Z_i, which follows a standard normal distribution. I simulated 2000 observations. The true value of the structural parameter, β_0, is 0. I also set $\Pi_z = 1$ and $\Pi_z = 0.1$ to mimic strong and weak instruments. The log-likelihood functions are evaluated assuming that σ_v^2 equals 1 and the correlation coefficient ρ equals 0.5.

Fig. 1: Endogenous Probit log-likelihood functions for strong and weak instrument.



Clearly, when the instrument is strong, the log-likelihood is globally concave and can be uniquely maximized. In the case of weak instrument, the log-likelihood function resembles a negative semidefinite quadratic form. The smoothness along the line where $\Pi_z = 0$ indicates the lack of global identification of β.

To capture how weak instruments affect asymptotic behavior of statistical tests, I follow Staiger and Stock (1997) and model Π_z as local to zero.

Definition 1. *Let* C *be a full rank matrix.* Π_z *has the following asymptotic behavior in case of strong, weak and irrelevant instruments:*

i) $\Pi_z = \mathrm{C}$,

ii) $\Pi_z = \Pi_{z,n} = \frac{\mathrm{C}}{\sqrt{n}}$,

iii) $\Pi_z = 0$.

3 Weak Instruments Robust Tests for Limited Dependent Variable Models

I derive tests robust against the presence of weak instruments using a minimum distance (MD) objective function. They are modified versions of existing tests and denoted by the subscript M. This section contains the general case. In the following one, I illustrate the general case for two limited dependent variable models: the selection model with an endogenous variable and the endogenous probit model.

For the classical MD estimator there exists an auxiliary unrestricted parameter vector θ and a consistent estimator $\hat{\theta}_n$ such that:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Lambda_0)$$

where θ_0 is the true value of the parameter under the data generating process and Λ_0 is the asymptotic variance-covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. The estimation of the auxiliary parameter and its asymptotic variance-covariance matrix do not depend on the structural parameter nor on the quality of the instruments. The parameters θ and β are linked by a differentiable vector function $\bar{r} : \Theta \times \mathbb{B} \times \mathrm{Z} \to \mathbb{R}^k$ such that:

$$\bar{r}(\theta, \beta, \zeta) = \Big[r_1(\theta, \beta)' \;\; \vdots \;\; r_2(\theta, \beta)' + \zeta' \Big]'$$

where $r_1(\theta, \beta)$ is a $m \times 1$ vector and $r_2(\theta, \beta)$ is a $(k-m) \times 1$ vector. The parameter ζ indicates whether the overidentifying restriction holds: the unrestricted and the structural model are equivalent when $\zeta = 0$.

Define $r : \Theta \times \mathbb{B} \to \mathbb{R}^k$ where $r(\theta, \beta) = \bar{r}(\theta, \beta, 0)$. Under the true data degenerating process, we observe:

$$\bar{r}(\theta_0, \beta_0, 0) = r(\theta_0, \beta_0) = 0$$

Following Gorieroux and Monfort (1989), the objective function of the efficient minimum distance estimator (or the optimal asymptotic least square estimator) is:

$$S_n(\beta) = \frac{n}{2}\; r_n(\beta)' \left[R_n(\beta) \hat{\Lambda}_n R_n(\beta)' \right]^{-1} r_n(\beta) \tag{3.1}$$

where $r_n(\beta) = r(\hat{\theta}_n, \beta)$, $R_n(\beta) = \frac{\partial r(\theta, \beta)}{\partial \theta}\Big|_{\theta = \hat{\theta}_n}$ and $\hat{\Lambda}_n$ is a consistent estimator of Λ_0. Define

$$\hat{\Psi}_\beta = R_n(\beta) \hat{\Lambda}_n R_n(\beta)'$$

Under the null hypothesis $\mathrm{H}_0^S : \beta = \beta_0$ and $\zeta = 0$, the modified version of the S-test proposed by Stock and Wright (2000) is:

$$\begin{aligned} S_{\mathrm{M}}(\beta_0) &= \; n\; r_n(\beta_0)' \left[R_n(\beta_0) \hat{\Lambda}_n R_n(\beta_0)' \right]^{-1} r_n(\beta_0) \\ &= \; n\; r_n(\beta_0)'\; \hat{\Psi}_{\beta_0}^{-1}\; r_n(\beta_0) \end{aligned} \tag{3.2}$$

The S_{M}-test tests simultaneously the value of the structural parameter and the overidentifying restriction. Moreover, if $R_0 = \frac{\partial r(\theta_0, \beta_0)}{\partial \theta}$ is a full column rank matrix then the S_{M}-test converges asymptotically to a χ^2-distribution under H_0^S no matter if β is identified or not:

Theorem 3.1. *Assume that $\hat{\theta}_n$ and $\hat{\Lambda}_n$ are consistent estimators of θ_0 and Λ_0, respectively, which are not functions of β. Moreover, assume that $r(\theta, \beta)$ is a continuous differentiable real function with $r(\theta_0, \beta_0) = 0$ and R_0 is a full column rank matrix. Then under the null hypothesis $\mathrm{H}_0^S : \beta = \beta_0$ and $\zeta = 0$*

$$S_{\mathrm{M}}(\beta_0) \xrightarrow{d} \chi^2(k)$$

whether or not the structural parameter is identified.

Proof. See appendix A.1. □

Kleibergen's robust tests demand the estimation of untested parameters under the null hypothesis (see Kleibergen (2005b)). For limited dependent variable models the untested parameter's estimator comes from the solution of a nonlinear constraint optimization problem, which can be computationally unstable. Moreover, the consistent estimation requires the identification of the untested parameter under the null hypothesis which it is not easily verifiable.[2]

Differently from Kleibergen's tests, I redefine the parameter space, avoiding the constrained optimization problem and leaving the identification of θ independent of β. Such redefinition does not affect the consistency of the auxiliary unrestricted parameter's estimator as well as the consistency of its covariance matrix estimator.

The S_{M}-test's degrees of freedom are related to the dimension of $r(\theta, \beta)$, which may increase at the same pace as the number of instruments, decreasing the power of the S_{M}-test against the alternative hypothesis. In the GMM context, Kleibergen's solution comes from the asymptotic independence between the moments and their expected Jacobian (see Kleibergen (2005b)). I adopt the same idea, substituting the moments by the link function $r_n(\beta)$. For now, assume that $\zeta = 0$ and define $\mathrm{D}_n(\beta)$, a $k \times m$ matrix statistic for $\frac{\partial r(\theta,\beta)}{\partial \beta}$, as:

$$\mathrm{D}_n(\beta) = \begin{bmatrix} \mathrm{D}_{1,n}(\beta) & \dots & \mathrm{D}_{m,n}(\beta) \end{bmatrix}$$
$$\mathrm{D}_{j,n}(\beta) = \frac{\partial r_n(\beta)}{\partial \beta_j} - \frac{\partial R_n(\beta)}{\partial \beta_j} \hat{\Lambda}_n R_n(\beta)' \hat{\Psi}_\beta^{-1} r_n(\beta) \qquad \text{for} \;\; j = 1, \dots, m$$

where $\frac{\partial R_n(\beta)}{\partial \beta_j}$ is a matrix with the same dimension as $R_n(\beta)$ whose elements are derived with respect to β_j. Lemma 1 states that $\mathrm{D}_n(\beta)$ and $r_n(\beta)$ are asymptotically independent.

[2]See appendix B for a more detailed exposition of these two problems.

Lemma 1. *Assume that* $\frac{\partial}{\partial\theta}\mathrm{vec}\left[\frac{\partial r(\theta,\beta)}{\partial\beta}\right] = \left(\frac{\partial}{\partial\beta}\mathrm{vec}\left[\frac{\partial r(\theta,\beta)}{\partial\theta}\right]\right)'$. *Given* $\zeta = 0$, *the asymptotic joint distribution of* $\mathrm{D}_n(\beta)$ *and* $r_n(\beta)$ *under the null hypothesis* $\mathrm{H}_0^K : \beta = \beta_0$ *is:*

$$\sqrt{n}\begin{pmatrix} r_n(\beta_0) \\ \mathrm{vec}\left[\mathrm{D}_n(\beta_0) - \frac{\partial r(\theta_0,\beta_0)}{\partial\beta}\right]\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\Psi_0 & 0\\ 0 & \Xi_0\end{bmatrix}\right) \tag{3.3}$$

where:

$$\begin{aligned}
\Psi_0 &= R_0\Lambda_0 R_0' \\
R_0 &= \left.\frac{\partial r(\theta,\beta)}{\partial\theta}\right|_{\theta=\theta_0,\beta=\beta_0} \\
\Xi_0 &= \left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right)'\Lambda_0\left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right) - \left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right)'\Lambda_0 R_0'\Psi_0^{-1}R_0\Lambda_0\left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right) \\
\frac{\partial \mathrm{vec}[R_0]}{\partial\beta} &= \left.\frac{\partial}{\partial\beta}\mathrm{vec}\left[\frac{\partial r(\theta,\beta)}{\partial\theta}\right]\right|_{\theta=\theta_0,\beta=\beta_0}
\end{aligned}$$

Proof. See appendix A.2. □

The classical statistical tests, Wald, LM and DM, derived from the objective function (3.1), assumes that $\frac{\partial r(\theta_0,\beta_0)}{\partial\beta}$ has full rank (see Newey and McFadden (1994), section 9). If this assumption is not fulfilled, then these tests depend on the nuisance parameters (see Stock and Wright (2000)). Given $\zeta = 0$, $\mathrm{D}_n(\beta)$ and $r_n(\beta)$ are asymptotically independent under H_0^K. This independence is not affected by the rank of $\frac{\partial r(\theta_0,\beta_0)}{\partial\beta}$. I use this independence in order to derive the modified version of the K-test proposed by Kleibergen (2005b):

Theorem 3.2. *Define the* K_{M}*-test as:*

$$\begin{aligned}
K_{\mathrm{M}}(\beta_0) = &\ n\ r_n(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\mathrm{D}_n(\beta_0)\left[\mathrm{D}_n(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\mathrm{D}_n(\beta_0)\right]^{-1}\mathrm{D}_n(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}r_n(\beta_0) \\
= &\ n\ r_n(\beta_0)'\hat{\Psi}_{\beta_0}^{-\frac{1}{2}'}\ \hat{\mathrm{P}}_{\beta_0}\ \hat{\Psi}_{\beta_0}^{-\frac{1}{2}}r_n(\beta_0)
\end{aligned} \tag{3.4}$$

where:

$$\hat{\mathrm{P}}_{\beta_0} = \hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\mathrm{D}_n(\beta_0)\left[\mathrm{D}_n(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\mathrm{D}_n(\beta_0)\right]^{-1}\mathrm{D}_n(\beta_0)'\hat{\Psi}_{\beta_0}^{-\frac{1}{2}'}$$

Given $\zeta = 0$, *under assumptions of theorem 3.1, lemma 1 and* H_0^K, *we have:*

$$K_{\mathrm{M}}(\beta_0) \xrightarrow{d} \chi^2(m)$$

independently of the rank of $\frac{\partial r(\theta_0,\beta_0)}{\partial\beta}$.

Proof. See appendix A.3. □

As explained by Moreira (2003), since the Jacobian is estimated independently from the link function, the distribution of the K_{M}-test conditional on $\mathrm{D}_n(\beta_0)$ is free from the nuisance parameters. Moreover, its unconditional distribution is pivotal.

The computational cost to implement the K_{M}-test is lower compared to its GMM version. Solving the constrained optimization problem to estimate untested parameters and computing the variance-covariance matrix between the moments and the Jacobian are unnecessary. For some models, the K_{M}-test is even implementable using regular statistical softwares.[3]

The K_{M}-test is also obtained from the MD estimator objective function (3.1). The continuous updating estimator (CUE) solves the following first order condition:[4]

$$\frac{\partial S_n(\beta)}{\partial \beta} = n\ r_n(\beta)'\hat{\Psi}_{\beta}^{-1}\mathrm{D}_n(\beta) \tag{3.5}$$

Thence, the K_{M}-test is the quadratic form of the score of the continuous updating estimator evaluated at the hypothesized value of the parameter. However, differently from the original one where the CUE-GMM estimator attains the minimum value of the test, its minimum is attained at the CUE-MD estimator.

Inflexion and local minimum points also solve (3.5). At these points the K_{M}-test suffers spurious decline of power due to the failure of the overidentifying restriction, i.e, $\zeta \neq 0$. The $J_{K_{\mathrm{M}}}$-test is complementary to the K_{M}-test, which tests the hypothesis that $\mathrm{H}_0^J : \zeta = 0$, given $\beta = \beta_0$. Define $\hat{\mathrm{M}}_{\beta_0} = I_k - \hat{\mathrm{P}}_{\beta_0}$. Similarly to the J-test derived by Kleibergen (2004), the $J_{K_{\mathrm{M}}}$-test is:

$$\begin{aligned} J_{K_{\mathrm{M}}}(\beta_0) &= n\ r_n(\beta_0)'\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\ \hat{\mathrm{M}}_{\beta_0}\hat{\Psi}_{\beta_0}^{-\frac{1}{2}\,\prime}\, r_n(\beta_0) \\ J_{K_{\mathrm{M}}}(\beta_0) &\xrightarrow{d} \chi^2(k-m) \end{aligned} \tag{3.6}$$

Clearly, the K_{M} and the $J_{K_{\mathrm{M}}}$ tests are independent since they are an orthogonal decomposition of the S_{M}-test, i.e.,

$$S_{\mathrm{M}}(\beta_0) = K_{\mathrm{M}}(\beta_0) + J_{K_{\mathrm{M}}}(\beta_0)$$

At points where K_{M} suffers spurious decline of power, $J_{K_{\mathrm{M}}}$ takes the value of the S_{M}-test, which always has discriminatory power in those regions of the parameter space. So we may define a new test for the structural parameter combining both previous tests. Define $\tau_{K_{\mathrm{M}}}$ and $\tau_{J_{K_{\mathrm{M}}}}$, the levels of significance of K_{M} and $J_{K_{\mathrm{M}}}$ tests, respectively. The KJ_{M}-test, the combination test, has

[3]See Magnusson (2006) for the implementation of K_{M}-test in the endogenous Tobit model.

[4]See appendix A.4.

approximate significance level of $\tau = \tau_{K_\mathrm{M}} + \tau_{J_{K_\mathrm{M}}}$.[5] A suggested choice for τ_{K_M} is 0.04 and for $\tau_{J_{K_M}}$ is 0.01, since our main interest is to test the value of the structural parameter β.

Another robust test is the extension of the conditional likelihood ratio proposed by Moreira (2003) for the MD framework. In the present context the modified conditional likelihood-ratio test for one endogenous variable is:

$$LR_\mathrm{M}(\beta_0) = \frac{1}{2}\left\{S_\mathrm{M}(\beta_0) - \mathrm{rk}(\beta_0) + \sqrt{(S_\mathrm{M}(\beta_0) + \mathrm{rk}(\beta_0))^2 - 4J_{K_\mathrm{M}}(\beta_0)\mathrm{rk}(\beta_0)}\right\} \tag{3.7}$$

where:

$$\mathrm{rk}(\beta_0) = n\ \left\{\mathrm{D}_n(\beta_0)'\ \hat{\Xi}_{\beta_0}^{-1}\ \mathrm{D}_n(\beta_0)\right\}$$

and $\hat{\Xi}_{\beta_0}$ is a consistent estimator of Ξ_0. The asymptotic distribution of the LR_M test is not pivotal since it depends on the value of $\mathrm{rk}(\beta_0)$. However it is possible to simulate the critical values from the conditional distribution of the test by generating independent values of $\chi^2(1)$ and $\chi^2(k-1)$.[6]

Similarly to the first-stage F-test of linear IV models, rk can be interpreted as a statistic for the identification of the structural parameter. As the structural parameter becomes unidentifiable, the LR_M-test approximates to the S_M-test. On the other hand, as the structural parameter become fully identifiable, the LR_M-test approaches the K_M-test. Therefore, one may derive the limiting behavior of the LR_M-test as a function of rk:

$LR_\mathrm{M} \longrightarrow S_\mathrm{M}$ as $\mathrm{rk} \longrightarrow 0$ and $LR_\mathrm{M} \longrightarrow K_\mathrm{M}$ as $\mathrm{rk} \longrightarrow +\infty$.

Since the LR_M-test is a function of the K_M and J_{K_M} tests, we do not expected a spurious decline of power, differently from the K_M-test.

---

[5] Let CR_KJ, CR_K and CR_J be the critical regions for KJ_M, K_M and J_{K_M} tests. Hence:

$$\begin{aligned}\Pr\left(KJ_\mathrm{M} \in CR_\mathrm{KJ}\right) &= \Pr\left(\{K_\mathrm{M} \in CR_\mathrm{K}\} \cap \{J_M \in CR_J\}\right) + \Pr\left(\{K_\mathrm{M} \in CR_\mathrm{K}\} \cap \{J_{K_\mathrm{M}} \notin CR_\mathrm{J}\}\right)\\ &\quad + \Pr\left(\{K_\mathrm{M} \notin CR_\mathrm{K}\} \cap \{J_{K_\mathrm{M}} \in CR_\mathrm{J}\}\right)\\ &= \tau_{K_\mathrm{M}}\tau_{J_{K_\mathrm{M}}} + \tau_{K_\mathrm{M}}(1-\tau_{J_{K_\mathrm{M}}}) + (1-\tau_{K_\mathrm{M}})(\tau_{J_{K_\mathrm{M}}})\\ &= \tau_{K_\mathrm{M}} + \tau_{J_{K_\mathrm{M}}} - \tau_{K_\mathrm{M}}\tau_{J_{K_\mathrm{M}}} \approx \tau.\end{aligned}$$

[6] See Kleibergen (2005a) for the extension of LR_M in the presence of more than one endogenous variable in the GMM context.

4 Robust Tests: two examples

I illustrate the application of the above robust tests for two limited dependent variable models: the selection model with an endogenous variable and the endogenous probit model.[7] Examples of both models are found in Mroz (1987), who uses the first model to investigate the labor supply of married women and in Evans et al. (1992), who use the last model to investigate the association between teenage pregnancy and high school dropouts.

4.1 The selection model with one endogenous variable

The selection model with one endogenous variable can be described as:

$$\begin{cases} Y_i = X_i\beta + W_t\gamma + U_i \\ X_i = Z_i\Pi_z + W_i\Pi_w + V_i \\ L_i = \mathbb{1}(Z_i f_z + W_i f_w + E_i > 0) \end{cases} \tag{4.1}$$

where Y_i is observed if $L_i = 1$ and X_i may be always observed or it may be observed when $L_i = 1$. Instead of providing the joint distribution of the residuals I depart from the following less restrictive assumptions:

i) $U_i = E_i\rho_u + \varepsilon_{u,t}$ and $V_i = E_i\rho_v + \varepsilon_{v,i}$;

ii) $\mathbb{E}\left[\varepsilon_{u,i}\right] = \mathbb{E}\left[\varepsilon_{v,i}\right] = 0$ and $\varepsilon_{u,i}, \varepsilon_{v,i} \perp E_i, Z_i, W_i$;

iii) $k_z \geq 2$ such that at least one element of the excluded instruments explains the selection equation;

iv) $\mathbb{E}[E_i|L_i = 1] = G(Z_i f_z + W_i f_w) = G(\bar{Z}_i f)$, where $\bar{Z}_i = \begin{bmatrix} Z_i & W_i \end{bmatrix}$ and $G(\cdot)$ is a known real function.

Under the above assumptions, the unrestricted model for the selected sample is:

$$\begin{cases} Y_i = Z_i\pi_z + W_i\pi_w + G(\bar{Z}_i f)\pi_g + e_{y,i} \\ X_i = Z_i\Pi_z + W_i\Pi_w + G(\bar{Z}_i f)\Pi_g + e_{x,i} \end{cases} \tag{4.2}$$

where $\mathbb{E}\left[e_{y,i}\right] = \mathbb{E}\left[e_{x,i}\right] = 0$ and $e_{y,i}, e_{x,i} \perp \bar{Z}_i$. The ordinary least squares is a consistent estimator for the unrestricted parameters after substituting $G(\bar{Z}_i f)$ by its estimate $G(\bar{Z}_i \hat{f})$, obtained from

[7] For a detailed explanation of both models see Wooldridge (2002), chapter 17, and Rivers and Vuong (1988), respectively.

the selection equation. Thence, if $G(\bar{Z}_i f)$ is consistently estimated, the identification condition for the unrestricted parameter is the same as for the ordinary least squares. The auxiliary parameter θ and the link mapping are, respectively:

$$\theta = \begin{bmatrix} \pi_z' & \pi_w' & \pi_g' & \Pi_z' & \text{vec}[\Pi_w]' & \Pi_g' \end{bmatrix}' \quad \text{and} \quad r(\theta, \beta) = \pi_z - \Pi_z \beta.$$

Let $\hat{\pi}_z$ and $\hat{\Pi}_z$ be the estimators of π and Π as described in the previous paragraph. The following are defined as the asymptotic variance-covariance matrix of $\sqrt{n}\begin{bmatrix}(\hat{\pi}_z - \pi_z^0)' & \text{vec}[\hat{\Pi}_z - \Pi_z^0]'\end{bmatrix}'$ and its consistent estimator:[8]

$$\mathrm{V}_0 = \begin{bmatrix} V_{\pi_z \pi_z} & V_{\pi_z \Pi_z} \\ V_{\Pi_z \pi_z} & V_{\Pi_z \Pi_z} \end{bmatrix} \quad \text{and} \quad \hat{\mathrm{V}} = \begin{bmatrix} \hat{V}_{\pi_z \pi_z} & \hat{V}_{\pi_z \Pi_z} \\ \hat{V}_{\Pi_z \pi_z} & \hat{V}_{\Pi_z \Pi_z} \end{bmatrix}.$$

The statistics for $\hat{\Psi}_{\beta_0}$ and $\mathrm{D}_n(\beta_0)$ are defined, respectively, as:[9]

$$\hat{\Psi}_{\beta_0} = \begin{bmatrix} I_{k_z} & -\beta_0 I_{k_z} \end{bmatrix} \begin{bmatrix} \hat{V}_{\pi_z \pi_z} & \hat{V}_{\pi_z \Pi_z} \\ \hat{V}_{\Pi_z \pi_z} & \hat{V}_{\Pi_z \Pi_z} \end{bmatrix} \begin{bmatrix} I_{k_z} \\ -\beta_0 I_{k_z} \end{bmatrix}$$

$$\mathrm{D}_n(\beta_0) = -n \left\{ \hat{\Pi}_z - (\hat{V}_{\Pi_z \pi_z} - \beta_0 \hat{V}_{\Pi_z \Pi_z}) \hat{\Psi}_{\beta_0}^{-1} (\hat{\pi}_z - \hat{\Pi}_z \beta_0) \right\}$$

The following corollary states the tests that are robust against weak instruments for the selection model with one endogenous variable.

Corollary 1. *Let* $\hat{\hat{\Pi}}_{\beta_0} = -\mathrm{D}_n(\beta_0)$*. The* S_{M}*,* K_{M} *and* $J_{K_{\mathrm{M}}}$ *tests for the selection model with one endogenous variable are:*

$$S_{\mathrm{M}}(\beta_0) = n\ (\hat{\pi}_z - \hat{\Pi}_z \beta_0)' \hat{\Psi}_{\beta_0}^{-1} (\hat{\pi}_z - \hat{\Pi}_z \beta_0) \tag{4.3a}$$

$$K_{\mathrm{M}}(\beta_0) = n\ (\hat{\pi}_z - \hat{\Pi}_z \beta_0)' \hat{\Psi}_{\beta_0}^{-1} \hat{\hat{\Pi}}_{\beta_0} \left(\hat{\hat{\Pi}}_{\beta_0}' \hat{\Psi}_{\beta_0}^{-1} \hat{\hat{\Pi}}_{\beta_0} \right)^{-1} \hat{\hat{\Pi}}_{\beta_0}' \hat{\Psi}_{\beta_0}^{-1} (\hat{\pi}_z - \hat{\Pi}_z \beta_0) \tag{4.3b}$$

$$J_{K_{\mathrm{M}}}(\beta_0) = S_{\mathrm{M}}(\beta_0) - K_{\mathrm{M}}(\beta_0) \tag{4.3c}$$

Under H_0^S*,* H_0^K *and* H_0^J*, as* $n \to +\infty$*:*

$$S_{\mathrm{M}}(\beta_0) \xrightarrow{d} \chi^2(k_z), \ K_{\mathrm{M}}(\beta_0) \xrightarrow{d} \chi^2(m) \text{ and } J_{K_{\mathrm{M}}}(\beta_0) \xrightarrow{d} \chi^2(k_z - m)$$

regardless of whether the instruments are strong, weak or irrelevant as in definition 1.

Proof. Directly from theorems 3.1 and 3.2. □

[8] See appendix A.5.1 equations (A.11), (A.12) and (A.13) for the detailed variance-covariance matrix formula.

[9] See appendix A.5 for the derivation of the statistics.

Estimates of Ξ_0 and rk are necessary to implement the LR_{M}-test. They are:

$$\hat{\hat{\Xi}}_{\beta_0} = \hat{V}_{\Pi_z\Pi_z} - (\hat{V}_{\Pi_z\pi_z} - \beta_0\hat{V}_{\Pi_z\Pi_z})\hat{\Psi}_{\beta_0}^{-1}(\hat{V}_{\pi_z\Pi_z} - \beta_0\hat{V}_{\Pi_z\Pi_z})$$
$$\mathrm{rk}(\beta_0) = n\ \hat{\hat{\Pi}}'_{\beta_0}\hat{\hat{\Xi}}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0}$$

After simulating the critical values, one may compute LR_{M}-test by substituting $S_{\mathrm{M}}(\beta_0)$, $J_{K_{\mathrm{M}}}(\beta_0)$ and $\mathrm{rk}(\beta_0)$ into (3.7).

The following subsection illustrates the application for another limited information model with endogenous explanatory variables: the endogenous probit model.

4.2 The endogenous probit model

The endogenous probit model is described as

$$\begin{cases} Y_i = \mathbb{1}(X_i\beta + W_i\gamma + U_i > 0) \\ X_i = Z_i\Pi_z + W_i\Pi_w + V_i \end{cases} \qquad (U_i, V_i)|Z_i, W_i \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_{\mathrm{u}}^2 & \Sigma_{\mathrm{uv}} \\ \Sigma_{\mathrm{vu}} & \Sigma_{\mathrm{vv}} \end{bmatrix}\right) \tag{4.4}$$

From the properties of the multivariate normal distribution we have:

$$U_i = V_i\alpha + \varepsilon_i, \qquad \varepsilon_i|Z_i, W_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

where $\sigma_\varepsilon^2 = \sigma_{\mathrm{u}}^2(1 - \rho'\rho)$ and $\rho = \frac{\Sigma_{\mathrm{vv}}^{-\frac{1}{2}}\Sigma_{\mathrm{vu}}}{\sigma_{\mathrm{u}}}$. The conditional unrestricted model becomes:

$$\begin{cases} Y_i = \mathbb{1}(Z_i\pi_z + W_i\pi_w + V_i\pi_v + \varepsilon_i > 0) \\ X_i = Z_i\Pi_z + W_i\Pi_w + V_i \end{cases} \tag{4.5}$$

I normalize $\sigma_\varepsilon = 1$. The unrestricted parameters can be estimated by the two-stage conditional maximum likelihood (TSCML) as proposed by Rivers and Vuong (1988). This method estimates Π_z and Π_w by ordinary least squares in the first stage. Its second stage is the application of the regular probit in the first equation with the substitutition of V_i by its estimate obtained from the first stage. The maximum likelihood and the TSCML deliver the same point estimates for this particular model (see Newey (1987), proposition 7). The identification conditions for the unrestricted parameters are the same as those for the ordinary least squares and for the classical probit models. Thence, the TSCML estimator for the unrestricted parameter is consistent under mild assumptions, regardless of the presence of weak instruments.

In this model, the auxiliary parameter and the link function are, respectively:

$$\theta = \begin{bmatrix} \pi_z' & \pi_w' & \pi_v' & \mathrm{vec}[\Pi_z]' & \mathrm{vec}[\Pi_w]' \end{bmatrix}'$$
$$r(\theta, \beta) = \pi_z - \Pi_z\beta$$

The asymptotic covariance matrices of $r(\hat{\theta}_n, \beta_0)$ and $\mathrm{D}_n(\beta_0)$ are:[10]

$$\Psi_0 = \Gamma^{zz} + (\pi_v^0 - \beta_0)'\Sigma_{\mathrm{vv}}(\pi_v^0 - \beta)\mathbb{E}[Z_i^{\perp\prime} Z_i^{\perp}]^{-1}$$
$$\Xi_0 = \Sigma_{\mathrm{vv}} \otimes \mathbb{E}[Z_i^{\perp\prime} Z_i^{\perp}]^{-1} - \Sigma_{\mathrm{vv}}(\pi_v^0 - \beta_0)(\pi_v^0 - \beta_0)'\Sigma_{\mathrm{vv}} \otimes \left(\mathbb{E}[Z_i^{\perp\prime} Z_i^{\perp}]\Psi_0\mathbb{E}[Z_i^{\perp\prime} Z_i^{\perp}]\right)^{-1}$$

where Γ^{zz} is the block of the inverse Fisher information matrix relative to the parameter π_z and $Z_i^{\perp}$ represents Z_i projected out from the space spanned by W_i. Let $\hat{\pi}_v$ and $\hat{\pi}_z$ be the TSCML estimators of π_v and π_z, respectively, $\hat{\Pi}_z$ be the ordinary least squares from the first stage, $\hat{\Sigma}_{\mathrm{vv}} = \frac{X^{\perp\prime} M_{Z^{\perp}} X^{\perp}}{n-(k_z+k_w)}$ and $\hat{\Gamma}^{zz}$ be the estimator of Γ^{zz}. The statistics for $\hat{\Psi}_{\beta_0}$ and $\mathrm{D}_n(\beta_0)$ are defined as:

$$\hat{\Psi}_{\beta_0} = \hat{\Gamma}^{zz} + (\hat{\pi}_v - \beta_0)'\hat{\Sigma}_{\mathrm{vv}}(\hat{\pi}_v - \beta_0)\left(\frac{Z^{\perp\prime} Z^{\perp}}{n}\right)^{-1}$$
$$\mathrm{D}_n(\beta_0) = -\left(\hat{\Pi}_z - \left(\frac{Z^{\perp\prime} Z^{\perp}}{n}\right)^{-1}\hat{\Psi}_{\beta_0}^{-1}(\hat{\pi}_z - \hat{\Pi}_z\beta_0)(\hat{\pi}_v - \beta_0)'\hat{\Sigma}_{\mathrm{vv}}\right)$$

The following corollary states the S_{M}, K_{M} and $J_{K_{\mathrm{M}}}$ tests for the endogenous probit model:

Corollary 2. *Let $\hat{\hat{\Pi}}_{\beta_0} = -\mathrm{D}_n(\beta_0)$. The S_{M}, K_{M} and $J_{K_{\mathrm{M}}}$ tests for the endogenous probit model are:*

$$S_{\mathrm{M}}(\beta_0) = n\ (\hat{\pi}_z - \hat{\Pi}_z\beta_0)'\hat{\Psi}_{\beta_0}^{-1}(\hat{\pi}_z - \hat{\Pi}_z\beta_0) \tag{4.6a}$$
$$K_{\mathrm{M}}(\beta_0) = n\ (\hat{\pi}_z - \hat{\Pi}_z\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0}\left(\hat{\hat{\Pi}}'_{\beta_0}\hat{\Psi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0}\right)^{-1}\hat{\hat{\Pi}}'_{\beta_0}\hat{\Psi}_{\beta_0}^{-1}(\hat{\pi}_z - \hat{\Pi}_z\beta_0) \tag{4.6b}$$
$$J_{K_{\mathrm{M}}}(\beta_0) = S_{\mathrm{M}}(\beta_0) - K_{\mathrm{M}}(\beta_0) \tag{4.6c}$$

Under H_0^S, H_0^K and H_0^J, as $n \to +\infty$:

$$S_{\mathrm{M}}(\beta_0) \xrightarrow{d} \chi^2(k_z),\ K_{\mathrm{M}}(\beta_0) \xrightarrow{d} \chi^2(m) \text{ and } J_{K_{\mathrm{M}}}(\beta_0) \xrightarrow{d} \chi^2(k_z - m)$$

regardless of whether the instruments are strong, weak or irrelevant as in definition 1.

Proof. Directly from theorems 3.1 and 3.2. □

Finally, in order to implement the LR_{M}-test, I define $\hat{\hat{\Xi}}_{\beta_0}$, the consistent estimator for Ξ_0, and for $\mathrm{rk}(\beta_0)$ as:

$$\hat{\hat{\Xi}}_{\beta_0} = \hat{\Sigma}_{\mathrm{vv}} \otimes \left(\frac{Z^{\perp\prime} Z^{\perp}}{n}\right)^{-1} - \hat{\Sigma}_{\mathrm{vv}}(\hat{\pi}_v - \beta_0)(\hat{\pi}_v - \beta_0)'\hat{\Sigma}_{\mathrm{vv}} \otimes \left(\frac{Z^{\perp\prime} Z^{\perp}}{n}\hat{\Psi}_{\beta_0}\frac{Z^{\perp\prime} Z^{\perp}}{n}\right)^{-1} \tag{4.7a}$$
$$\mathrm{rk}(\beta_0) = n\ \hat{\hat{\Pi}}'_{\beta_0}\hat{\hat{\Xi}}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0} \tag{4.7b}$$

Substituting $S_{\mathrm{M}}(\beta_0)$, $J_{K_{\mathrm{M}}}(\beta_0)$ and $\mathrm{rk}(\beta_0)$ into (3.7) one may compute the $LR_{\mathrm{M}}(\beta_0)$-test after simulating its critical value.

[10] The derivation of both covariance matrices are in appendix A.5.2.

5 Power Investigation

I investigate the power of the tests that are robust against weak instrument by testing the hypothesis $H_0 : \beta = 0$ at the 5% significance level. The model under investigation is the sample selection with one endogenous variable described in subsection 4.1.[11] The same model is used for the empirical application described in the next section.

The design of the simulations uses the following latent model:

$$\begin{cases} Y_i^* = X_i^* \beta + U_i \\ X_i^* = Z_i \Pi_z + V_i \\ L_i^* = W_i f_w + E_i \end{cases} \tag{5.1}$$

where E_i comes from a standard logistic distribution, $U_i = \alpha_u E_i + e_{yi}$, $V_i = \alpha_v E_i + e_{xi}$ with e_{yi} and e_{yi} resulting from a bivariate normal distribution with mean zero, variance one and correlation coefficient ρ_{yx}. There is only one explanatory endogenous variable X_i^* and one instrument, Z_i, correlated to it. L_i^* is a latent variable for the selection equation which is explained by the instrument W_i. I added three irrelevant instruments in the model so the total number of instruments equals five. All instruments come from independent normally distributed random normals with zero mean and unitary variance and are kept constant for all simulations.

Stock and Yogo (2004) developed two tests to capture the presence of weak instruments for linear IV regression models. When there is only one endogenous variable, the tests consist of comparing the first-stage F-statistic to its respective critical values, which lie in the range of $(10, 27)$.[12] I set $\lambda_z = \frac{\Pi_z' Z' Z \Pi_z}{k \Sigma_{\mathrm{vv}}}$, also known as the concentration parameter, equal to 20, 10, 3, 1 and 0.01 in order to mimic very strong, strong, medium, weak and inept instruments. I also set λ_w, the selection equation concentration parameter, equal to 10 for all simulations.[13] The number of latent observations for each simulation is 500. Table 1 summarizes the simulation designs.

[11] The simulations from the endogenous probit model give similar results as the one reported here. They are available upon request.

[12] See Stock and Yogo (2004) tables 1 and 2, second and third columns. The critical values are no less than 10, the rule of thumb to detect the presence of weak instruments suggested by Staiger and Stock (1997).

[13] Simulations with different values for λ_w were performed. The results are very similar to the ones presented here and therefore they are not reported.

Table 1: Simulation designs[a]

$\lambda_w = \frac{f_w' W' W f_w}{k\sigma_E^2}$	α_u	α_v	ρ_{yx}	$\lambda_z = \frac{\Pi_z' Z' Z \Pi_z}{k\Sigma_{\mathrm{vv}}}$
				20
			0.0	10
10	0.8	0.9	0.5	3
			0.9	1
				0.01

[a] Number of latent observations = 500.

For each value of β, 2500 samples were generated from the latent model in (5.1). The structural model is described in (4.1). I assume that Y_i and X_i are observed if $L_i = 1$ and that Z_i, W_i and the three irrelevant instruments are always observed. I test the null hypothesis $H_0 : \beta = 0$ for each simulation and compute the proportion of rejected tests in order to build the power curves at the 5% significance level. Here only the results for the case in which $\rho_{yx} = 0.5$ are reported, while the remaining are in the appendix C. I also analyze the performance of the Wald test, which is defined as:

$$\mathrm{W}(\beta_0) = n\ (\hat{\beta} - \beta_0)' \hat{V}_\beta^{-1} (\hat{\beta} - \beta_0) \tag{5.2}$$

where $\hat{\beta}$ is the multistage estimator and $\hat{V}_\beta$ is the variance estimator of $\hat{\beta}$. The multistage estimator is the TSLS estimator applied to:

$$\begin{cases} Y_i = X_i\beta + W_i\gamma + G(\bar{Z}_i\hat{f})\delta + \hat{\varepsilon}_{yi} \\ X_i = Z_i\Pi_z + W_i\Pi_w + G(\bar{Z}_i\hat{f})\Pi_g + \hat{\varepsilon}_{xi} \end{cases} \tag{5.3}$$

where $\hat{f}$ is the estimator from the logit selection equation.[14]

Graphs of figure 2 show the power curves for the five robust tests described in section 3 and for the Wald test, using different values of λ_z. In case of robust tests, the rejection probability attains approximately 5% when β equals 0 for all graphs, reflecting that they reach the correct level under regular and weak instruments asymptotics. As expected, the S_{M}-test has poor power compared to the others because of the large number of instruments compared to the number of endogenous variables. The graphs also illustrate the spurious behavior of the K_{M} at the left side of the x-axis. In this region, the combination test does not suffer from spurious decline of power. The power curves approximate to the 5% horizontal line as the instruments become weaker. This suggests that confidence intervals derived by inverting the robust tests may be unbounded.

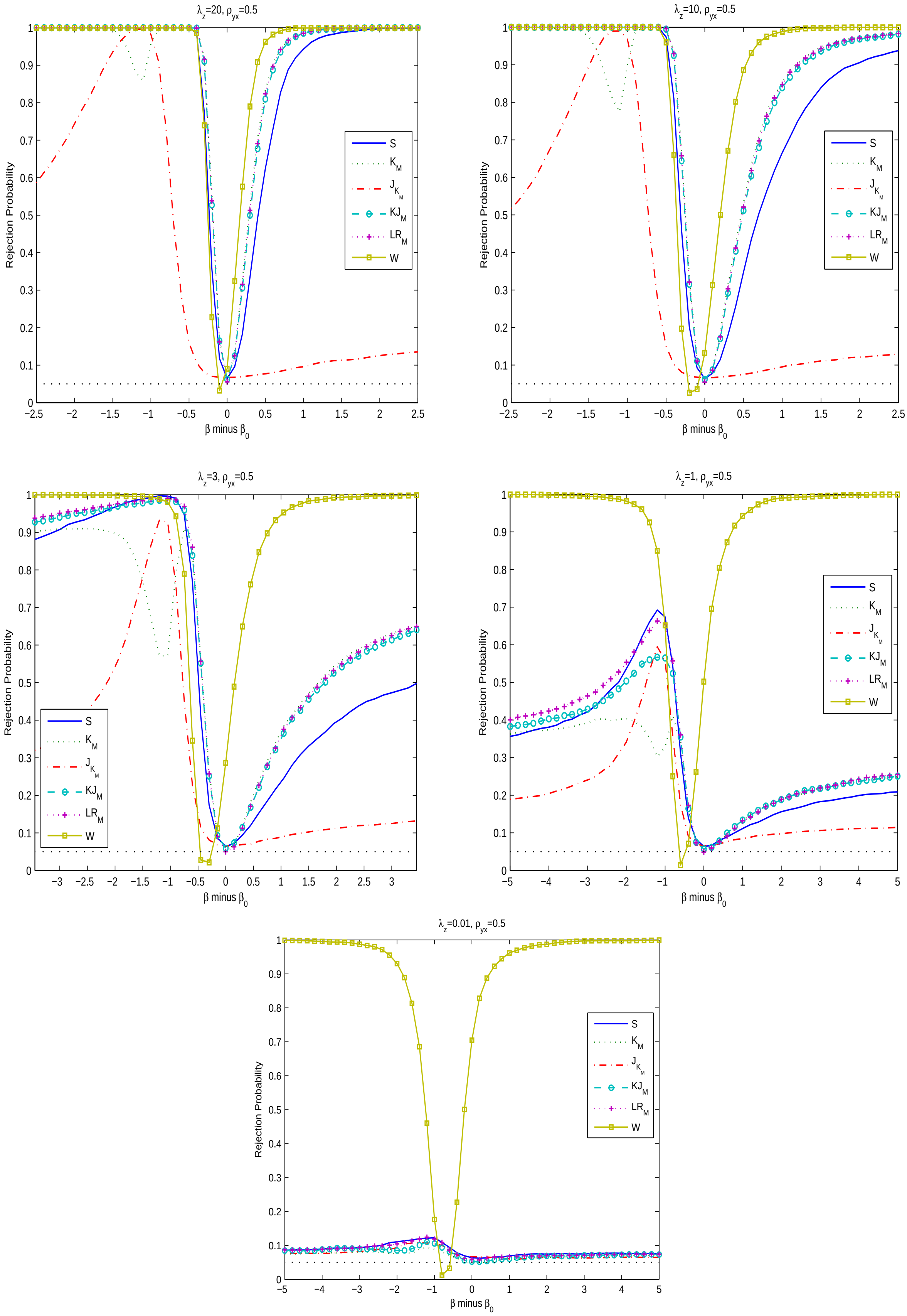
[14] Simulations with probit selection show the same results and they are not reported.

In general, the power curves of the robust tests resemble the ones obtained by Kleibergen (2005a) for the case of linear limited information models. Determining from the graphs which robust test dominates the other is difficult, although the LR_{M} seems to have better power properties compared to the remaining robust tests when the instruments are strong.

On the other hand, the Wald test is biased and has the wrong size at the null hypothesis in all cases. When β equals 0, i.e, when the size and the power should be the same, the rejection probability varies from 7% (strong instruments) to 80% (inept instruments). This behavior is explained by the poor small sample properties of the point estimator.[15] Also, the shape of the power curves does not change when the instruments become weaker, remaining steep even when identification of the structural parameter fails. Thence the Wald test may produce bounded confidence intervals with wrong coverage probability.

[15] For the cases considered here, the point estimator is upward biased even when the instruments are strong. The bias becomes negligible when the sample increases above 800 observations for both cases of strong instruments.

Fig. 2: Power curves for Robust and Wald tests.



6 Empirical Application: Labor Supply of Married Women

I apply the robust tests to construct confidence intervals for the structural parameter of the labor supply of married women. Mroz (1987) describes the structural model, which takes as exogenous husband's behavior, as:

$$\begin{cases} Y_i = X_i\beta + W_i\gamma + U_i \\ X_i = Z_i\Pi_z + W_i\Pi_z + V_i \\ L_i = 1(Z_i f_z + W_i f_w + E_i > 0) \end{cases} \tag{6.1}$$

where Y_i is the hours of work during a given year, X_i is a measure of female wage rate, W_i is a set of control variables which includes, besides a constant term, a measure of other income received by the household and demographic variables (age, number of children less than six, number of children between six and nineteen). The set of instruments Z_i includes background variables, quadratic and cubic terms of age and education and nonwife income. L_i is the labor force participation indicator. More details about the variables are in appendix D. The data set used in that work comes from the Michigan Panel Study of Income Dynamics. The sample consists of 753 white women between the ages of 30 and 60 in 1975, with 428 working at some time during the year. X_i, the wage, is her average hourly of earnings, which is obtained as the division of the total yearly earnings by hours of work. Thence, by construction, X_i is an endogenous variable.

Even though the information about hours of work and exogenous variables is always available, the wage is observed only for the women who participate in the labor force ($L_i = 1$). There is a combination of sample selection and endogeneity in the model. Mroz (1987) estimates the structural using the multistage procedure described in the previous section.[16]

Points of the parameter space which do not reject the hypotheses $H_0 : \beta = \beta_0$ at 5% belong to the 95% confidence interval for the structural parameter. The plots of one minus p-value for the robust tests are shown by figures 3 and 4. In figure 3 the selection equation is probit while in figure 4 the selection equation is logit. The intersection between the one minus p-value and the 95% horizontal lines delimits the confidence interval. I also report the confidence interval obtained by Mroz (1987) using the Wald test of equation (5.2).

[16] See Heckman (1974) for the maximum likelihood estimation. There are two advantages to the multistage method compared to the likelihood: it relaxes the distributional assumption of the residuals and is computationally easy to implement.

Fig. 3: Confidence Intervals using statistical tests: probit selection equation

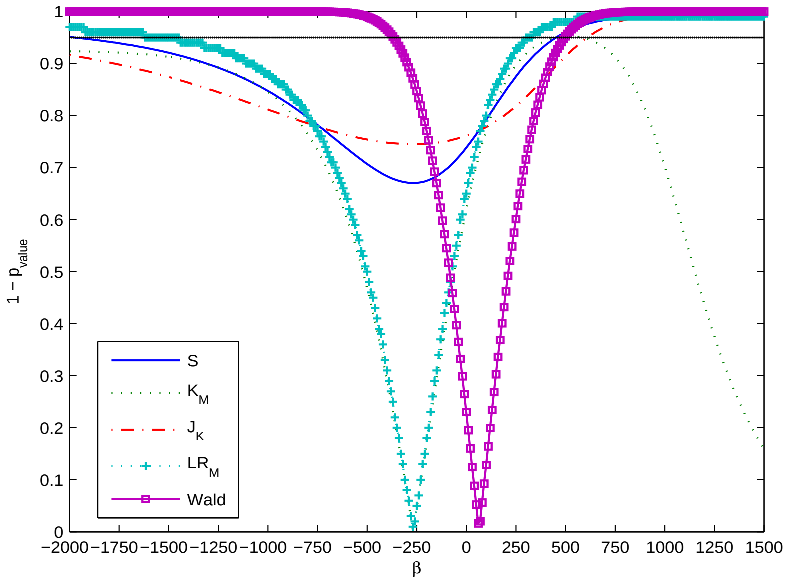
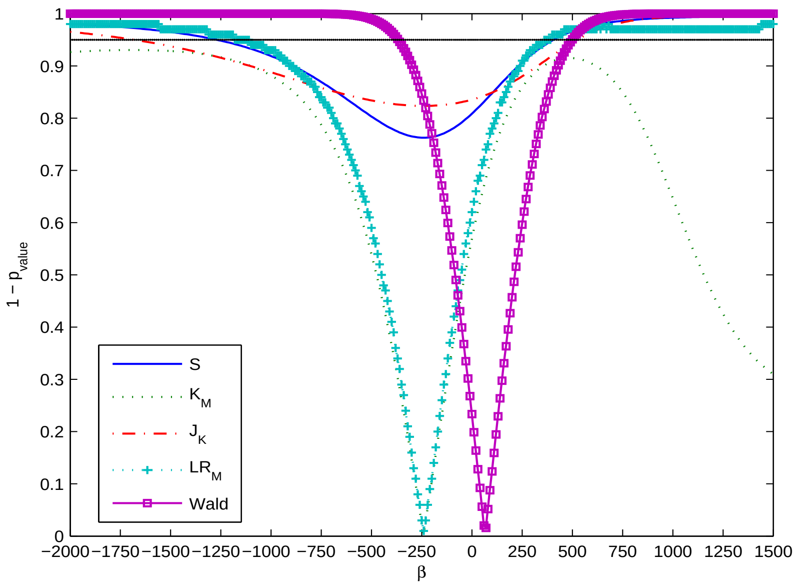


Fig. 4: Confidence Intervals using statistical tests: logit selection equation



The confidence intervals are similar for both figures. The LR_{M} produces smaller confidence sets compared to the remaining robust statistics. The K_{M}-test behaves spuriously on the positive part of the real line resulting in large and even non-convex confidence intervals. It is also clear that K_{M} is not minimized at the multistage estimator.

The range of the Wald statistic confidence intervals is nearly contained in the robust tests confidence intervals. Given the results of the previous section, the point estimative may be upward biased leading to wrong inference about the structural parameter. The difference in the ranges of the confidence intervals produced by robust and Wald tests reinforces the possible presence of weak instruments.

7 Conclusion

In this paper I show how to obtain tests that are robust against weak instruments for limited information models with endogenous explanatory variables. I derive those tests using a minimum distance objective function. I illustrate the method with two examples: the selection model with an endogenous explanatory variable and the endogenous probit. Simulations were conducted for comparing robust and Wald tests. The former perform according to the local asymptotics as introduced by Staiger and Stock (1997), while the latter overrejects the null hypothesis at the true value. An empirical application of the robust tests is conducted in order to build confidence intervals for the structural parameter of the labor supply of married women.

A Proofs

A.1 Proof of Theorem 3.1

By the Slustky and the continuous mapping theorems

$$R_n(\beta_0)\hat{\Lambda}_n R_n(\beta_0)' \xrightarrow{p} R_0\Lambda_0 R_0'$$

Thence, by the central limited theorem combined again with the Slutsky theorem one may find

$$\left(R_n(\beta_0)\hat{\Lambda}_n R_n(\beta_0)'\right)^{-\frac{1}{2}} \sqrt{n}\ r_n(\beta_0) \xrightarrow{d} \mathcal{N}\left(0, I_k\right)$$

and $S_{\mathrm{M}}(\beta_0) \xrightarrow{d} \chi^2(k)$, as $n \to +\infty$.

A.2 Proof of Lemma 1

First, I introduce another lemma:

Lemma 2. *Given that $\zeta = 0$, under the null hypothesis $\mathrm{H}_0^K : \beta = \beta_0$, the asymptotic join distribution between $r_n(\beta_0)$ and $\frac{\partial r_n(\beta_0)}{\partial \beta}$ is:*

$$\sqrt{n}\begin{pmatrix} r_n(\beta_0) \\ \mathrm{vec}\left[\frac{\partial r_n(\beta_0)}{\partial\beta} - \frac{\partial r(\theta_0,\beta_0)}{\partial\beta}\right]\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\Psi_0 & R_0\Lambda_0\left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right) \\ \left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right)'\Lambda_0 R_0' & \left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right)'\Lambda_0\left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right)\end{bmatrix}\right) \quad \text{(A.1)}$$

Proof.

The assumption that $\frac{\partial}{\partial\theta}\mathrm{vec}\left[\frac{\partial r(\theta,\beta)}{\partial\beta}\right] = \left(\frac{\partial}{\partial\beta}\mathrm{vec}\left[\frac{\partial r(\theta,\beta)}{\partial\theta}\right]\right)'$ implies that

$$\frac{\partial^2 r(\theta,\beta)}{\partial\theta_h\partial\beta_j} = \frac{\partial^2 r(\theta,\beta)}{\partial\beta_j\partial\theta_h} \quad \text{for} \quad j = 1,\ldots,m \text{ and } h = 1,\ldots,b.$$

By Taylor expansion:

$$\sqrt{n}\begin{pmatrix} r_n(\beta_0) \\ \mathrm{vec}\left[\frac{\partial r_n(\beta_0)}{\partial\beta} - \frac{\partial r(\theta_0,\beta_0)}{\partial\beta}\right]\end{pmatrix} = \begin{bmatrix}\frac{\partial r(\theta^*,\beta_0)}{\partial\theta} \\ \frac{\partial}{\partial\theta}\mathrm{vec}\left[\frac{\partial r(\theta^*,\beta_0)}{\partial\beta}\right]\end{bmatrix}\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$$

where θ^* lies between $\hat{\theta}_n$ and θ_0. By the Slutsky theorem combined with the central limit theorem one may find that, as $n \to +\infty$:

$$\sqrt{n}\begin{pmatrix} r_n(\beta_0) \\ \mathrm{vec}\left[\frac{\partial r_n(\beta_0)}{\partial\beta} - \frac{\partial r(\theta_0,\beta_0)}{\partial\beta}\right]\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}R_0\Lambda_0 R_0' & R_0\Lambda_0\left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right) \\ \left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right)'\Lambda_0 R_0' & \left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right)'\Lambda_0\left(\frac{\partial \mathrm{vec}[R_0]}{\partial\beta}\right)\end{bmatrix}\right)$$

□

Consider the following lower-block triangular matrix:

$$\begin{bmatrix} I_k & 0 \\ -\left(\frac{\partial \text{vec}[R_0]}{\partial \beta}\right)' \Lambda_0 R_0' \Psi_0^{-1} & I_k \otimes I_m \end{bmatrix} \tag{A.2}$$

Pre-multiplying (A.1) by (A.2) one may get

$$\sqrt{n} \begin{pmatrix} r_n(\beta_0) \\ \text{vec}\left[\text{D}_n(\beta_0) - \frac{\partial r(\theta_0, \beta_0)}{\partial \beta}\right] \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Psi_0 & 0 \\ 0 & \Xi_0 \end{bmatrix} \right) \tag{A.3}$$

where:

$$\Xi_0 = \left(\frac{\partial \text{vec}[R_0]}{\partial \beta}\right)' \Lambda_0 \left(\frac{\partial \text{vec}[R_0]}{\partial \beta}\right) - \left(\frac{\partial \text{vec}[R_0]}{\partial \beta}\right)' \Lambda_0 R_0' \Psi_0^{-1} R_0 \Lambda_0 \left(\frac{\partial \text{vec}[R_0]}{\partial \beta}\right) \tag{A.4}$$

A.3 Proof of Theorem 3.2

Independently of the rank of $\frac{\partial r(\theta_0, \beta_0)}{\partial \beta}$, as $n \to +\infty$

$$R_n(\beta_0) \hat{\Lambda}_n R_n(\beta_0)' \xrightarrow{p} R_0 \Lambda_0 R_0' \equiv \Psi_0$$

where Ψ_0 is a positive definite matrix. Define ψ_r, the limiting distribution of $\sqrt{n} r_n(\beta_0)$, and C, a full rank matrix. I separate the proof in three cases: (i) $\frac{\partial r(\theta_0, \beta_0)}{\partial \beta} = \text{C}$, (ii) $\frac{\partial r(\theta_0, \beta_0)}{\partial \beta} = \frac{\text{C}}{n}$ and (iii) $\frac{\partial r(\theta_0, \beta_0)}{\partial \beta} = 0$.

i $\text{D}_n(\beta_0) \xrightarrow{p} \text{C}$. Therefore

$$\sqrt{n} \left\{ \text{D}_n(\beta_0)' \left[R_n(\beta_0) \hat{\Lambda}_n R_n(\beta_0)'\right]^{-1} \text{D}_n(\beta_0) \right\}^{-\frac{1}{2}} \text{D}_n(\beta_0)' \left[R_n(\beta_0) \hat{\Lambda}_n R_n(\beta_0)'\right]^{-1} r_n(\beta_0) \xrightarrow{d}$$
$$\left(\text{C}' \Psi_0^{-1} \text{C}\right)^{-\frac{1}{2}} \text{C}' \Psi_0^{-1} \psi_r \equiv \mathcal{N}\left(0, I_m\right)$$

ii $\sqrt{n}\, \text{vec}\left[\text{D}_n(\beta_0)\right] \xrightarrow{d} \mathcal{N}(\text{vec}\left[\text{C}\right], \Xi_0) \equiv \psi_D$. Therefore

$$n \text{D}_n(\beta_0)' \left[R_n(\beta_0) \hat{\Lambda}_n R_n(\beta_0)'\right]^{-1} r_n(\beta_0) \xrightarrow{d} \psi_D' \Psi_0^{-1} \psi_r$$

The conditional distribution of the above expression is

$$\psi_D' \Psi_0^{-1} \psi_r \big| \psi_D \equiv \mathcal{N}\left(0, \psi_D' \Psi_0^{-1} \psi_D\right)$$

Since the ψ_D and ψ_r are independent, we have

$$\left(\psi_D' \Psi_0^{-1} \psi_D\right)^{-\frac{1}{2}} \psi_D' \Psi_0^{-1} \psi_r \equiv \mathcal{N}(0, I_m)$$

and

$$n \left\{ \text{D}_n(\beta_0)' \left[R_n(\beta_0) \hat{\Lambda}_n R_n(\beta_0)'\right]^{-1} \text{D}_n(\beta_0) \right\}^{-\frac{1}{2}} \text{D}_n(\beta_0)' \left[R_n(\beta_0) \hat{\Lambda}_n R_n(\beta_0)'\right]^{-1} r_n(\beta_0) \xrightarrow{d} \mathcal{N}\left(0, I_m\right)$$

unconditionally.

iii $\sqrt{n}\, \text{vec}\left[\text{D}_n(\beta_0)\right] \xrightarrow{d} \mathcal{N}(0, \Xi_0) \equiv \psi_D$. The proof is the same as case (ii).

A.4 Derivation of equation (3.5)

The first order condition of the continuous updating estimator is:

$$\begin{aligned}\frac{\partial S_n(\beta)}{\partial \beta} &= n\ r_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}\frac{\partial r_n(\beta)}{\partial \beta} \\ &\quad + \frac{n}{2}\left(r_n(\beta)'\otimes r_n(\beta)'\right)\frac{\partial}{\partial \beta}\mathrm{vec}\left[\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}\right]\end{aligned} \tag{A.5}$$

The second term of the right hand side is:

$$-\left\{\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}\otimes\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}\right\}\frac{\partial}{\partial \beta}\mathrm{vec}\left[R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right]$$

The derivative term of the above equation is:

$$\left(R_n(\beta)\hat{\Lambda}_n\otimes I\right)\frac{\partial}{\partial \beta}\mathrm{vec}[R_n(\beta)] + \left(I\otimes R_n(\beta)\hat{\Lambda}_n\right)\frac{\partial}{\partial \beta}\mathrm{vec}[R_n(\beta)']$$

Thence the second term of (A.1) simplifies to:

$$\begin{aligned}&\left\{r_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}R_n(\beta)\hat{\Lambda}_n\otimes r_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}\right\}\frac{\partial}{\partial \beta}\mathrm{vec}[R_n(\beta)] \\ &+\left\{r_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}\otimes r_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}R_n(\beta)\hat{\Lambda}_n\right\}\frac{\partial}{\partial \beta}\mathrm{vec}[R_n(\beta)']\end{aligned} \tag{A.6}$$

Using the fact that

$$\frac{\partial}{\partial \beta_j}\mathrm{vec}[R_n(\beta)] = \mathrm{vec}\left[\frac{\partial R_n(\beta)}{\partial \beta_j}\right] \quad \text{and} \quad \frac{\partial}{\partial \beta_i}\mathrm{vec}[R_n(\beta)'] = \mathrm{vec}\left[\frac{\partial R_n(\beta)'}{\partial \beta_j}\right],$$

the i^{th} column of (A.6) is:[17]

$$\begin{aligned}&r_n(\theta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}\frac{\partial R_n(\beta)}{\partial \beta_j}\ \hat{\Lambda}_n R_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}r_n(\beta) \\ &+\ r_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}R_n(\beta)\hat{\Lambda}_n\frac{\partial R_n(\beta)'}{\partial \beta_j}\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}r_n(\beta)\end{aligned}$$

Since both terms are scalars, (A.6) simplifies to

$$\begin{aligned}&2n\ r_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n^{-1}R_n(\beta)'\right)^{-1}\times \\ &\quad\left[\tfrac{\partial R_n(\beta)}{\partial \beta_1}\quad\dots\quad\tfrac{\partial R_n(\beta)}{\partial \beta_m}\right]\hat{\Lambda}_n R_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}r_n(\beta)\end{aligned}$$

and (A.5) becomes:

$$\frac{\partial S_n(\beta)}{\partial \beta} = n\ r_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}\mathrm{D}_n(\beta) \tag{A.7}$$

where:

$$\begin{aligned}\mathrm{D}_n(\beta) &= \left[\mathrm{D}_{1,n}(\beta)\quad\dots\quad\mathrm{D}_{m,n}(\beta)\right] \\ \mathrm{D}_{j,n}(\beta) &= \frac{\partial r_n(\beta)}{\partial \beta_j} - \frac{\partial R_n(\beta)}{\partial \beta_j}\ \hat{\Lambda}_n R_n(\beta)'\left(R_n(\beta)\hat{\Lambda}_n R_n(\beta)'\right)^{-1}r_n(\beta)\quad \text{for}\ \ j = 1,\dots,m\end{aligned}$$

---

[17] $\frac{\partial R_n(\beta)}{\partial \beta_j}$ has the same dimension of $R_n(\beta)$ and each element is derived with respect to β_j.

A.5 The statistics for the robust tests

Before presenting the statistics of the robust tests, I introduce some notations which are suitable for both examples in section 3. The definition of the link functions are:

$$\begin{aligned} r(\theta_0, \beta_0) &= \pi_z^0 - \Pi_z^0 \beta_0 \\ r_n(\beta_0) &= \hat{\pi}_z - \hat{\Pi}_z \beta_0 \end{aligned}$$

From the above equations we derive:

$$R_n(\beta_0) = R_0 = \begin{bmatrix} I_{k_z} & -\beta_0' \otimes I_{k_z} \end{bmatrix} \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \tag{A.8a}$$

$$\frac{\partial r_n(\beta_0)}{\partial \beta} = -\hat{\Pi}_z \tag{A.8b}$$

$$\frac{\partial \mathrm{vec}\,[R_0]}{\partial \beta} = -H_2' \tag{A.8c}$$

where H_1 and H_2 are selection matrices: H_1 selects the k_z rows related to the parameter π_z while H_2 selects the $k_z m$ rows related to $\mathrm{vec}\,[\Pi_z]$. Note that $R_n(\beta_0)$ is a constant matrix. Also, for simplifying notation, define:

$$\hat{\Lambda}_n^H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \hat{\Lambda}_n \begin{bmatrix} H_1' & H_2' \end{bmatrix} = \begin{bmatrix} \hat{\Lambda}_{11} & \hat{\Lambda}_{12} \\ \hat{\Lambda}_{21} & \hat{\Lambda}_{22} \end{bmatrix} \tag{A.9}$$

From (A.8a)-(A.8c) and (A.9), $\hat{\Psi}_{\beta_0}$, $\mathrm{D}_n(\beta_0)$ and $\hat{\Xi}_{\beta_0}$ are rewritten as:

$$\hat{\Psi}_{\beta_0} = R_0 \hat{\Lambda}_n R_0' = \begin{bmatrix} I_{k_z} & -\beta_0' \otimes I_{k_z} \end{bmatrix} \begin{bmatrix} \hat{\Lambda}_{11} & \hat{\Lambda}_{12} \\ \hat{\Lambda}_{21} & \hat{\Lambda}_{22} \end{bmatrix} \begin{bmatrix} I_{k_z} \\ -\beta_0 \otimes I_{k_z} \end{bmatrix} \tag{A.10a}$$

$$\mathrm{D}_n(\beta_0) = -\left\{ \hat{\Pi}_z - \begin{bmatrix} \hat{\Lambda}_{21} & \hat{\Lambda}_{22} \end{bmatrix} \begin{bmatrix} I_{k_z} \\ -\beta_0 \otimes I_{k_z} \end{bmatrix} \hat{\Psi}_{\beta_0}^{-1} \left(\hat{\pi}_z - \hat{\Pi}_z \beta_0 \right) \right\} \tag{A.10b}$$

$$\hat{\Xi}_{\beta_0} = \hat{\Lambda}_{22} - \begin{bmatrix} \hat{\Lambda}_{21} & \hat{\Lambda}_{22} \end{bmatrix} \begin{bmatrix} I_{k_z} \\ -\beta_0 \otimes I_{k_z} \end{bmatrix} \hat{\Psi}_{\beta_0}^{-1} \begin{bmatrix} I_{k_z} & -\beta_0' \otimes I_{k_z} \end{bmatrix} \begin{bmatrix} \hat{\Lambda}_{12} \\ \hat{\Lambda}_{22} \end{bmatrix} \tag{A.10c}$$

A.5.1 The robust tests' statistics: selection model with one endogenous regressor

Define $\bar{Z}_i = \begin{bmatrix} Z_i & W_i \end{bmatrix}$ and $\tilde{Z}_i = \begin{bmatrix} Z_i & W_i & G(\bar{Z}_i f) \end{bmatrix}$. Assuming heteroscedasticity of unknown form, Lee (1982) provides the following sandwich asymptotic variance-covariance matrix for the unrestricted parameters:

$$\Lambda_0 = B^{-1} \Delta_0 B^{-1\prime} \tag{A.11}$$

where:

$$B = \mathbb{E}\begin{bmatrix} \tilde{Z}_i'\tilde{Z}_i & 0 \\ 0 & \tilde{Z}_i'\tilde{Z}_i \end{bmatrix} \tag{A.12}$$

$$\Delta_0 = \mathbb{E}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \tilde{Z}_i'\tilde{Z}_i e_{yi}^2 & \tilde{Z}_i'\tilde{Z}_i e_{yi}e_{xi} \\ \tilde{Z}_i'\tilde{Z}_i e_{yi}e_{xi} & \tilde{Z}_i'\tilde{Z}_i e_{xi}^2 \end{pmatrix}\right] + \\ \mathbb{E}\begin{bmatrix} \pi_g^0\tilde{Z}_i' \, \nabla_f G(\bar{Z}_i f_0) \\ \Pi_g^0\tilde{Z}_i' \, \nabla_f G(\bar{Z}_i f_0) \end{bmatrix} V_{as}(\hat{f}) \, \mathbb{E}\begin{bmatrix} \pi_g^0\tilde{Z}_i' \, \nabla_f G(\bar{Z}_i f_0) \\ \Pi_g^0\tilde{Z}_i' \, \nabla_f G(\bar{Z}_i f_0) \end{bmatrix}' \tag{A.13}$$

and $V_{as}(\hat{f})$ is the asymptotic variance-covariance of $\sqrt{n}\,(\hat{f} - f_0)$ which is obtained from the selection equation. The $G(\bar{Z}_i f)$ function derived from the probit selection equation is the inverse Mills ratio, i.e:

$$G(\bar{Z}_i f) = \frac{\phi\left(\bar{Z}_i f\right)}{\Phi\left(\bar{Z}_i f\right)}$$

If the selection equation is logistic, then we have:

$$G(\bar{Z}_i f) = \big[1 + \exp\big(-\bar{Z}_i f\big)\big] \times \ln\big[1 + \exp\big(-\bar{Z}_i f\big)\big] + \bar{Z}_i f \exp\big(-\bar{Z}_i f\big)$$

Consistent estimators for B and Δ_0 are:

$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \hat{\tilde{Z}}_i'\hat{\tilde{Z}}_i & 0 \\ 0 & \hat{\tilde{Z}}_i'\hat{\tilde{Z}}_i \end{bmatrix} \tag{A.14a}$$

$$\hat{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \hat{\tilde{Z}}_i'\hat{\tilde{Z}}_i\hat{e}_{yi}^2 & \hat{\tilde{Z}}_i'\hat{\tilde{Z}}_i\hat{e}_{yi}\hat{e}_{xi} \\ \hat{\tilde{Z}}_i'\hat{\tilde{Z}}_i\hat{e}_{yi}\hat{e}_{xi} & \hat{\tilde{Z}}_i'\hat{\tilde{Z}}_i\hat{e}_{xi}^2 \end{bmatrix} + \\ \left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \hat{\pi}_g\hat{\tilde{Z}}_i' \, \nabla_f G(\bar{Z}_i\hat{f}) \\ \hat{\Pi}_g\hat{\tilde{Z}}_i' \, \nabla_f G(\bar{Z}_i\hat{f}) \end{bmatrix}\right) \hat{V}_{as}(\hat{f}) \left(\frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \hat{\pi}_g\hat{\tilde{Z}}_i' \, \nabla_f G(\bar{Z}_i\hat{f}) \\ \hat{\Pi}_g\hat{\tilde{Z}}_i' \, \nabla_f G(\bar{Z}_i\hat{f}) \end{bmatrix}\right)' \tag{A.14b}$$

where:

$$\hat{\tilde{Z}}_i = \begin{bmatrix} Z_i & W_i & G(\bar{Z}_i\hat{f}) \end{bmatrix}$$

$$\hat{e}_{yi} = Y_i - Z_i\hat{\pi}_z - W_i\hat{\pi}_w - G(\bar{Z}_i\hat{f})\hat{\pi}_g$$

$$\hat{e}_{xi} = X_i - Z_i\hat{\Pi}_z - W_i\hat{\Pi}_w - G(\bar{Z}_i\hat{f})\hat{\Pi}_g$$

and $\hat{V}_{as}(\hat{f})$ is a consistent estimator of the Fisher information matrix derived from the selection equation.

A.5.2 The robust tests's statistics: endogenous probit model

Define $\bar{\pi} = \begin{bmatrix} \pi_z & \pi_w \end{bmatrix}$. The asymptotic variance-covariance matrix for the unrestricted parameters is given by Proposition 5 of Newey (1987):

$$\Lambda_0 = \begin{bmatrix} \begin{bmatrix} \Gamma_{\bar{\pi}\bar{\pi}} & \Gamma_{\bar{\pi}\pi_v} \\ \Gamma_{\pi_v\bar{\pi}} & \Gamma_{\pi_v\pi_v} \end{bmatrix}^{-1} + (\pi_v^0)'\Sigma_{\mathrm{vv}}\pi_v^0 \begin{bmatrix} (\mathbb{E}\left[\bar{Z}_i'\bar{Z}_i\right])^{-1} & 0 \\ 0 & 0 \end{bmatrix} & (\pi_v^0)'\Sigma_{\mathrm{vv}} \otimes \begin{bmatrix} (\mathbb{E}\left[\bar{Z}_i'\bar{Z}_i\right])^{-1} \\ 0 \end{bmatrix} \\ \\ \Sigma_{\mathrm{vv}}\pi_v^0 \otimes \begin{bmatrix} (\mathbb{E}\left[\bar{Z}_i'\bar{Z}_i\right])^{-1} & 0 \end{bmatrix} & \Sigma_{\mathrm{vv}} \otimes (\mathbb{E}\left[\bar{Z}_i'\bar{Z}_i\right])^{-1} \end{bmatrix} \tag{A.15}$$

where $\Gamma = \begin{bmatrix} \Gamma_{\bar{\pi}\bar{\pi}} & \Gamma_{\bar{\pi}\pi_v} \\ \Gamma_{\pi_v\bar{\pi}} & \Gamma_{\pi_v\pi_v} \end{bmatrix}$ is the probit's Fisher information matrix. Let Γ^{zz} denote the $k_z \times k_z$ upper left-hand block of the matrix of Γ^{-1} and $\hat{\Gamma}^{zz}$ its consistent estimator. From (A.9) we have:

$$\begin{aligned} \hat{\Lambda}_{11} &= \hat{\Gamma}^{zz} + \hat{\pi}_v'\hat{\Sigma}_{\mathrm{vv}}\hat{\pi}_v \left(\tfrac{Z^{\perp\prime}Z^{\perp}}{n}\right)^{-1} \\ \hat{\Lambda}_{12} = \hat{\Lambda}_{21}' &= \hat{\pi}_v'\hat{\Sigma}_{\mathrm{vv}} \otimes \left(\tfrac{Z^{\perp\prime}Z^{\perp}}{n}\right)^{-1} \\ \hat{\Lambda}_{22} &= \hat{\Sigma}_{\mathrm{vv}} \otimes \left(\tfrac{Z^{\perp\prime}Z^{\perp}}{n}\right)^{-1} \end{aligned}$$

Substituting the above relations into (A.10a)-(A.10c) we find:

$$\begin{aligned} \hat{\Psi}_{\beta_0} &= \hat{\Gamma}^{zz} + (\hat{\pi}_v - \beta_0)'\hat{\Sigma}_{\mathrm{vv}}(\hat{\pi}_v - \beta_0) \left(\tfrac{Z^{\perp\prime}Z^{\perp}}{n}\right)^{-1} \\ \mathrm{D}_n(\beta_0) &= -\left\{\hat{\Pi}_z - \left(\tfrac{Z^{\perp\prime}Z^{\perp}}{n}\right)^{-1} \hat{\Psi}_{\beta_0}^{-1}(\hat{\pi}_z - \hat{\Pi}_z\beta_0)(\hat{\pi}_v - \beta_0)'\hat{\Sigma}_{\mathrm{vv}}\right\} \\ \hat{\Xi}_{\beta_0} &= \hat{\Sigma}_{\mathrm{vv}} \otimes \left(\tfrac{Z^{\perp\prime}Z^{\perp}}{n}\right)^{-1} - \hat{\Sigma}_{\mathrm{vv}}(\hat{\pi}_v - \beta_0)(\hat{\pi}_v - \beta_0)'\hat{\Sigma}_{\mathrm{vv}} \otimes \left(\tfrac{Z^{\perp\prime}Z^{\perp}}{n}\hat{\Psi}_{\beta_0}\tfrac{Z^{\perp\prime}Z^{\perp}}{n}\right)^{-1} \end{aligned}$$

B Kleibergen's GMM Robust Tests for Limited Dependent Variable Models

Kleibergen's robust tests use the moment conditions for the structural parameter which are valid under the null hypothesis. Finding moments as a function of the structural parameter is only difficult in limited dependent variable models. Therefore, his tests demands first the estimation of the untested parameters.

The moment equation restriction is defined as

$$\mathbb{E}[g(\Upsilon_i, \eta, \beta)] = g_0(\eta, \beta) \quad \text{with} \quad g_0(\eta_0, \beta_0) = 0$$

where $\Upsilon_i = \{Y_i, X_i, W_i, Z_i\}$, η is a $p \times 1$ vector of additional parameters that are not being tested and $g : \mathbb{E} \times \mathbb{B} \rightarrow \mathbb{R}^k$ is a $k_g \times 1$ real continuously differentiable function with $k_g \geq m + p$, where m is the dimension of the structural parameter. The GMM objective function is:

$$\begin{aligned} \hat{Q}_n(\eta, \beta) &= \frac{1}{2n} \left(\sum_{i=1}^{n} g(\Upsilon_i, \eta, \beta) \right)' \left[\hat{\Omega}_n(\eta, \beta) \right]^{-1} \left(\sum_{i=1}^{n} g(\Upsilon_i, \eta, \beta) \right) \\ &= \frac{1}{2n} \; \hat{g}_n(\eta, \beta)' \left[\hat{\Omega}_n(\eta, \beta) \right]^{-1} \hat{g}_n(\eta, \beta) \end{aligned} \tag{B.1}$$

where $\hat{\Omega}_n(\eta, \beta)$ is a consistent estimator of $\Omega(\eta, \beta)$, the asymptotic variance-covariance matrix of $\frac{1}{\sqrt{n}} \; \hat{g}_n(\eta, \beta)$. Under the null hypothesis $\mathrm{H}_0 : \beta = \beta_0$, Kleibergen's estimator of the untested parameters, $\hat{\eta}_{\beta_0}$, solves:

$$\frac{1}{n} \; \hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0)' \left[\hat{\Omega}_n(\hat{\eta}_{\beta_0}, \beta_0) \right]^{-1} \hat{\mathrm{J}}_{\eta,n}(\hat{\eta}_{\beta_0}, \beta_0) = 0$$

where:

$$\begin{aligned} \hat{\mathrm{J}}_{\eta,n}(\hat{\eta}_{\beta_0}, \beta_0) &= \left[\hat{\mathrm{J}}_{\eta_1,n}(\hat{\eta}_{\beta_0}, \beta_0,) \quad \ldots \quad \hat{\mathrm{J}}_{\eta_p,n}(\hat{\eta}_{\beta_0}, \beta_0) \right] \\ \hat{\mathrm{J}}_{\eta_i,n}(\hat{\eta}_{\beta_0}, \beta_0) &= \frac{\partial \hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0)}{\partial \eta_i} \; - \; \hat{\mathrm{G}}_{\eta_i,n}(\hat{\eta}_{\beta_0}, \beta_0) \left[\hat{\Omega}_n(\hat{\eta}_{\beta_0}, \beta_0) \right]^{-1} \hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0) \quad \text{for} \quad i = 1, \ldots, p \end{aligned}$$

and $\hat{\mathrm{G}}_{\eta_i,n}(\hat{\eta}_{\beta_0}, \beta_0)$ is a consistent estimator of the covariance matrix between $\hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0)$ and $\frac{\partial \hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0)}{\partial \eta_i}$.

Let $\mathrm{J}_{\eta_0} = \mathbb{E}\left[\frac{\partial g(\upsilon, \eta_0, \beta_0)}{\partial \eta} \right]$. The consistency of $\hat{\eta}_{\beta_0}$ requires the identification of η under the null hypothesis. If $\Omega(\eta_0, \beta_0)$ is positive semidefinite, then the necessary and sufficient condition for local identification is that J_{η_0} is a full column rank matrix.[18] Due to nonlinearities of the

[18] See Kleibergen (2005b), assumption 3.

limited dependent variable models, the first difficulty of Kleibergen's tests is to verify whether the identification of η holds under the null hypothesis.

Once $\hat{\eta}_{\beta_0}$ is obtained, the second stage of Kleibergen's robust tests is the computation of $\hat{\mathrm{J}}_{\beta,n}(\hat{\eta}_{\beta_0}, \beta_0)$, a statistic for the Jacobian of the moments, where:

$$\hat{\mathrm{J}}_{\beta,n}(\hat{\eta}_{\beta_0}, \beta_0) = \left[\hat{\mathrm{J}}_{\beta_1,n}(\hat{\eta}_{\beta_0}, \beta_0) \quad \ldots \quad \hat{\mathrm{J}}_{\beta_m,n}(\hat{\eta}_{\beta_0}, \beta_0)\right]$$
$$\hat{\mathrm{J}}_{\beta_j,n}(\hat{\eta}_{\beta_0}, \beta_0) = \frac{\partial \hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0)}{\partial \beta_j} \; - \; \hat{\mathrm{G}}_{\beta_j,n}(\hat{\eta}_{\beta_0}, \beta_0)\left[\hat{\Omega}_n(\hat{\eta}_{\beta_0}, \beta_0)\right]^{-1}\hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0) \quad \text{for} \quad j = 1, \ldots, m$$

and $\hat{\mathrm{G}}_{\beta_j,n}(\hat{\eta}_{\beta_0}, \beta_0)$ is a consistent estimator of the covariance matrix between $\hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0)$ and $\frac{\partial \hat{g}_n(\hat{\eta}_{\beta_0}, \beta_0)}{\partial \beta_j}$.

The second difficulty of Kleibergen's tests is are their implementation. $\hat{\mathrm{G}}_{\eta_1,n}(\hat{\eta}_{\beta_0}, \beta_0)$,..., $\hat{\mathrm{G}}_{\eta_p,n}(\hat{\eta}_{\beta_0}, \beta_0)$, $\hat{\mathrm{J}}_{\eta,n}(\hat{\eta}_{\beta_0}, \beta_0)$ and $\hat{\eta}_{\beta_0}$ in the first stage and $\hat{\mathrm{G}}_{\beta_1,n}(\hat{\beta}_{\beta_0}, \beta_0,)$,..., $\hat{\mathrm{G}}_{\beta_m,n}(\hat{\beta}_{\beta_0}, \beta_0)$, $\hat{\mathrm{J}}_{\beta,n}(\hat{\eta}_{\beta_0}, \beta_0)$ in the second stage must be estimated.[19] When the dimension of η is high, a usual characteristic of limited dependent models, the computation of the estimates becomes very unstable.

The difficulties illustrated above show that Kleibergen's tests for limited dependent variable models are inappropriate and inconvenient to use. Instead of using the GMM objective function, I derive tests robust against the presence of weak instruments using a minimum distance objective function. This approach is easy to implement and avoids the constrained maximization behind Kleibergen's method.

[19]In the likelihood context, $\hat{\mathrm{G}}_{\eta_i,n}(\hat{\eta}_{\beta_0}, \beta_0)$ and $\hat{\mathrm{G}}_{\beta_j,n}(\hat{\eta}_{\beta_0}, \beta_0)$ are estimators for the covariance between the score function and the Hessian.

C Power Curves

Fig. 5: Power curves for Robust and Wald tests - selection model.

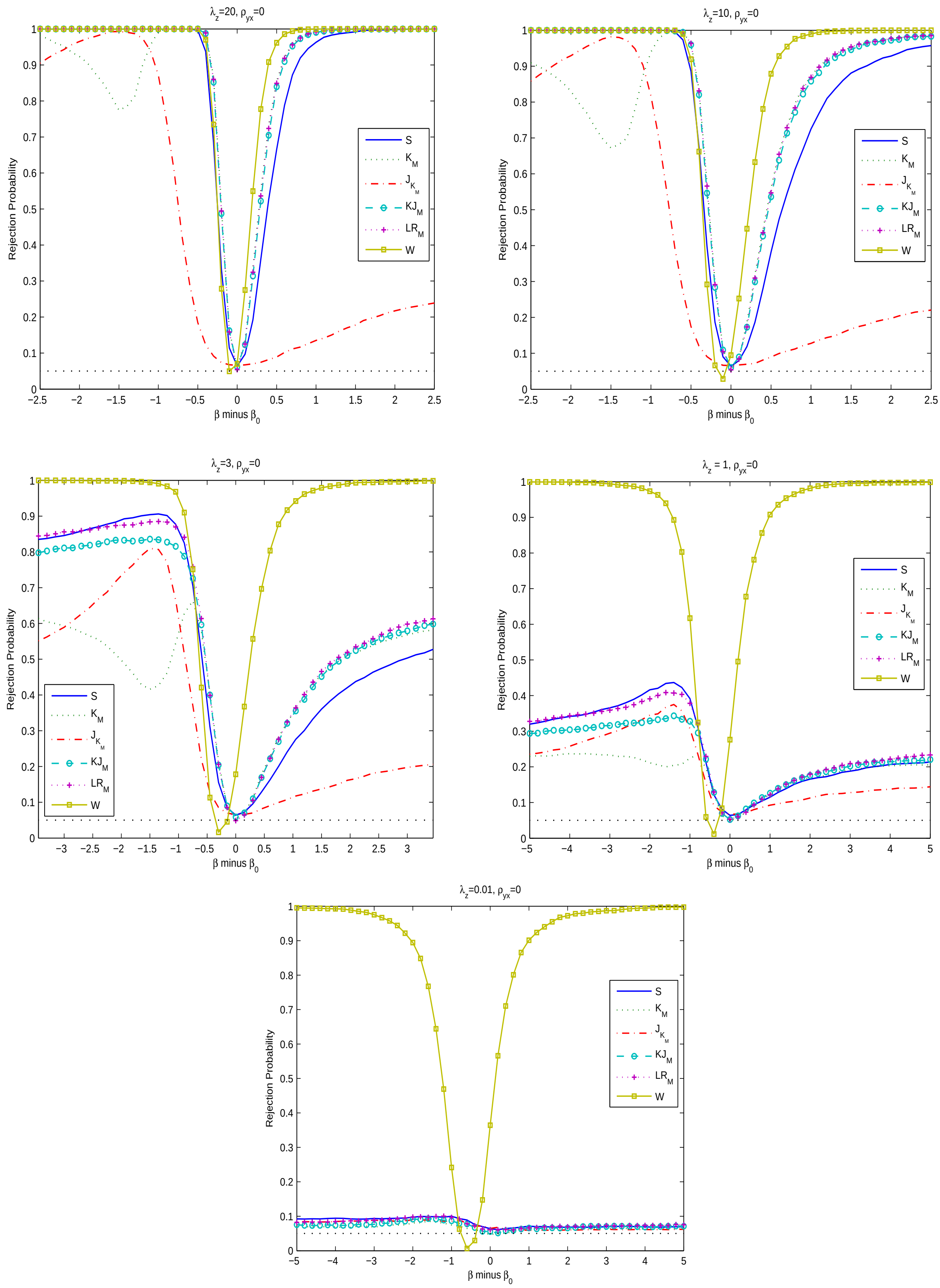
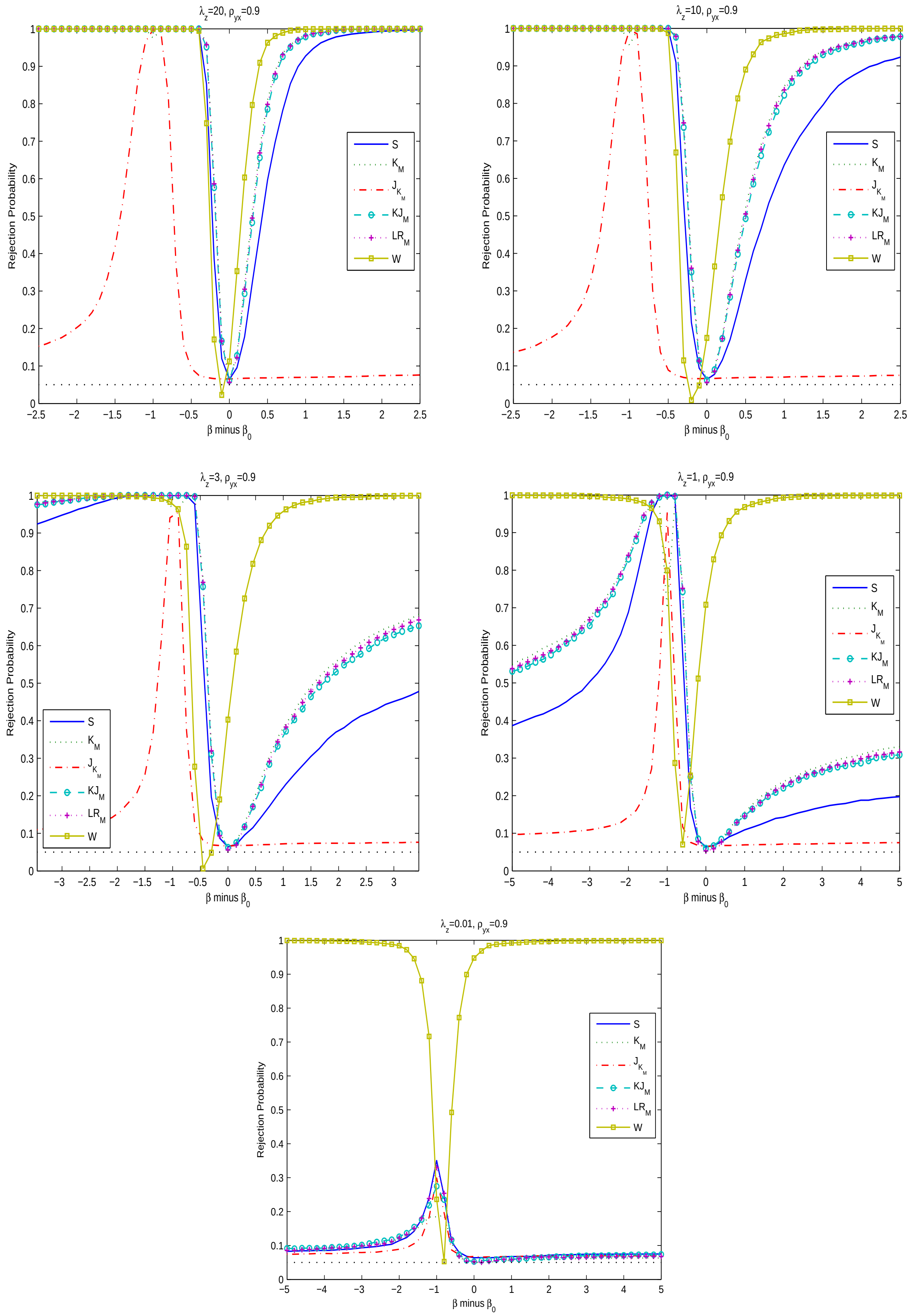


Fig. 6: Power curves for Robust and Wald tests - selection model.



D Data Description

The data set comes from the University of Michigan Panel Study of Income Dynamics PSID for the year 1975 (interviewed 1976). It consists of 753 married white women between ages 30 and 60, with 428 working at some time during the year. The data is available in package 'micEcon' of R statistical software.

Table 2: Definition of the variables, 753 married white women, University of Michigan - PSID - 1975

Variable		definition
Y	h	wife yearly hours of work
X	w_f	average hourly wage
W	age	wife's age in years
	$k1$	number of kids with 5 years old or younger
	$k2$	number of kids between 6 and 18 years old
	w_o	other family income[a]
	ed	years of schooling
Z	ed_m	mother's education in years
	ed_f	father's education in years
	un	unemployment rate in the county of residence
	dc	dummy variable: $dc = 1$ if live in large city (SMSA)
	ter	quadratic and cubic terms of education and age [b]

[a] $w_o = \frac{inc_h - h \times w_f}{1000}$, where inc_h is the household income.

[b] The terms are: age^2, ed^2, age^3, ed^3, $age \times ed$, $age^2 \times ed$ and $age \times ed^2$.

References

Amemiya, T. (1979). The estimation of simultaneous-equation tobit model. *International Economic Review 20*(1), 169–181.

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics 20*(1), 46–63.

Blundell, R. W. and R. J. Smith (1989). Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies 56*(1), 37–58.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*(430), 443–450.

Buse, A. (1992). The bias of instrumental variable estimators. *Econometrica 1*(1), 173–180.

Evans, W. N., W. Oates, and R. Schwab (1992). Measuring peer effects: a study of teenage behavior. *Journal of Political Economy 100*(5), 966–991.

Gorieroux, C. and A. Monfort (1989). A general framework for testing a null hypothesis in a "mixed form". *Econometric Theory 5*, 63–82.

Hansen, L. P., J. Heaton, and A. Yaron (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business and Economics Statistics 14*(3), 262–280.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica 42*(4), 679–694.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica 70*(5), 1781–1803.

Kleibergen, F. (2004). Testing subsets of structural parameters in the instrumental variables regression model. *The Review of Economics and Statistics 86*(1), 418–423.

Kleibergen, F. (2005a). Generalizing weak instrument robust iv statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics* (forthcoming).

Kleibergen, F. (2005b). Testing parameters in gmm without assuming that they are identified. *Econometrica 73*(4), 1103–1124.

Lankford, R. H. and J. H. Wyckoff (1991). Modeling charitable giving using box-cox standard tobit model. *The Review of Economics and Statistics 73*(3), 460–470.

Lee, L.-F. (1981). Simultaneoous equations models with discrete and censored dependent variables. In C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 346–364. MIT Press.

Lee, L.-F. (1982). Some approaches to the correction of selectivity bias. *Review of Economic Studies 49*(3), 335–372.

Lee, M.-J. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models* (first ed.). Springer.

Magnusson, L. M. (2006). Weak instruments robust tests for censored models. Working Paper, Brown University.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*(4), 1027–1048.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica 55*(4), 765–799.

Nelson, C. R. and R. Startz (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica 58*(4), 967–976.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume IV, Chapter 36, pp. 2111–2245. Elsevier.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics 36*, 231–250.

Poirier, D. J. (1978). The use of the box-cox transformation in limeted dependent variable models. *Journal of the American Statistical Association 73*(362), 284–287.

Rivers, D. and Q. H. Vuong (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics 39*(3), 347–366.

Smith, R. J. and R. W. Blundell (1986). An exogeneity test for simultaneuos equation tobit model with an application to labor supply. *Econometrica 54*(3), 679–685.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H. and J. Wright (2000). Gmm with weak identification. *Econometrica 68*, 1055–1096.

Stock, J. H. and M. Yogo (2004). Testing for weak instruments in linear iv regression. Working Paper, Department of Economics, Harvard University.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.