Tulane Economics Working Paper Series

Inference in Limited Dependent Variable Models Robust to Weak Identification

Leandro M. Magnusson
Department of Economics
206 Tilton Hall
Tulane University
New Orleans, LA 70118
lmagnuss@tulane.edu

Working Paper 0801
Original version: September 2008
This version: April 2009

Abstract

We propose tests for structural parameters in limited dependent variable models with endogenous explanatory variables. We use the generalized minimum distance principle. These tests are of the correct size regardless of whether the structural parameters are identified. Relating to the current tests, our tests is especially appropriate to models whose moment conditions are nonlinear in the parameters. Moreover, they are computationally simple, allowing them to be implemented in a large number of statistical software packages. We compare our tests to Wald tests by performing simulation experiments. We then use them to analyze both the female labor supply and the demand for cigarettes.

Keywords: weak identification, minimum chi-square estimation, hypothesis testing, limited dependent variable models
JEL codes: C12, C30, C34

Inference in Limited Dependent Variable Models Robust to Weak Identification *

Leandro M. Magnusson†
Tulane University

August 25, 2009

Abstract

We propose tests for structural parameters in limited dependent variable models with endogenous explanatory variables. These tests are based upon the generalized minimum distance principle. They are of the correct size regardless of whether the structural parameters are identified and are especially appropriate for models whose moment conditions are nonlinear in the parameters. Moreover, they are computationally simple, allowing them to be implemented using a large number of statistical software packages. We compare our tests to Wald tests in a simulation experiment and use them to analyze the female labor supply and the demand for cigarettes.

JEL classification: C12, C30, C34.
Keywords: Weak identification, minimum chi-square estimation, hypothesis testing, limited dependent variable models.

*I would like to thank Frank Kleibergen, Sophocles Mavroeidis, Sun-Jae Jun, Toru Kitagawa, Christian Bontemps and two anonymous referees for their helpful suggestions. I am also grateful for the suggestions of the participants of 2008 EC^2 conference in Rome.

†Department of Economics, Tulane University, 6823 Saint Charles Ave., 308 Tilton Hall, Tulane University, New Orleans LA 70118, USA, telephone: +1 (504) 862-8355, fax: +1 (504) 865-5869, e-mail: lmagnuss@tulane.edu, webpage: http://greenspace.tulane.edu/lmagnuss/.

1 Introduction

In this paper, we use the generalized minimum distance approach to derive tests for structural parameters in limited dependent variable models with endogenous explanatory variables. These tests are of the correct size even when the parameters are unidentified. The generalized minimum distance approach is especially convenient when the moment conditions are nonlinear in the parameters.

As shown by Staiger and Stock (1997) and Stock and Wright (2000), Wald, Lagrange multiplier (LM) and likelihood-ratio (LR) tests have nonstandard limiting distributions when the parameter is not identified, and, thus, inference based on these tests are unreliable. In the case of linear instrumental variable models, several tests are robust to parameter identification failure like the AR, see Anderson and Rubin (1949), the K, see Kleibergen (2002), the conditional likelihood-ratio (CLR), see Moreira (2003), among others. For nonlinear models, the extensions of these tests are based on the generalized method of moments (GMM). The starting point is the objective function of the continuous updating estimator (CUE). Stock and Wright (2000) formulate the S-test as an extension of the AR-test. Kleibergen (2005) proposes a new K-test which is the quadratic form of the score of the CUE's objective function. In the same paper, he derives the GMM extension of the CLR-test.

The generalized minimum distance approach is based on a link function which relates the structural and reduced form parameters. By avoiding moment conditions, this approach permits the construction of robust tests for a class of models where GMM would involve solving constrained nonlinear systems, or where the GMM is not feasible because the moments are not differentiable.

From an applied point of view, our proposed tests are useful because they are easier to compute. In many models, they can be computed using the built-in functions of regular statistical software packages. Moreover, confidence intervals based on these tests do not require the estimation of untested parameters under the null hypothesis at every hypothesized value of the parameter of interest.

However, the convenience of the proposed tests goes beyond their computational ease. Their asymptotic properties are derived from the asymptotic properties of the reduced form parameters, which do not depend on the structural parameters. Necessary conditions for the implementation are standard assumptions of minimum distance estimation. Under these assumptions, the reduced form parameters can be estimated either parametrically or semi-parametrically.

The paper is organized as follows. In the next section, we illustrate the use of the generalized minimum distance principle for estimating parameters in limited dependent variable models. The tests are derived in Section 3, and, in the subsequent section, we suggest an algorithm for the implementation of these tests in a class of limited dependent variable models. Next, in order to compare the performance of the proposed tests to Wald and robust GMM tests, we simulate endogenous Tobit and endogenous Poisson count data models. As an application of the tests, Section 6 considers the married female labor supply described by Blundell and Smith (1989) and Lee (1995) and the demand for cigarettes described by Mullahy (1997). The Appendix contains all proofs.

2 Minimum Distance Principle for Limited Dependent Variable Models

The minimum distance principle for estimation explores the underlying relation between structural parameters, denoted by β, and reduced form parameters, denoted by π. This relation is described by a system of implicity functions of the form $r(\pi, \beta) = 0$. In the literature, r is known as the *link function*. Let $\hat{\pi}$ be an estimator of π, such that $\hat{\pi} \xrightarrow{p} \pi$. We estimate β by forcing $||r(\hat{\pi}, \beta)|| = 0$ where $|| \cdot ||$ represents an appropriately weighted norm.

The next example illustrates the minimum distance method introduced by Amemiya (1979) for estimating parameters of limited depend variable models. For the use of this method in cross sectional models, see Madalla (1983), Blundell and Smith (1989), Lee (1995) and Blundell et al. (2007), and, in panel data models, see Arellano et al. (1999) and Jones and Labeaga (2003).

Example 1: Let $(\mathrm{y}^*, \mathrm{x}^*)$ be a vector of latent variables generated by a linear simultaneous system:

$$\begin{cases} \mathrm{y}^* = \mathrm{x}^*\beta_1 + \mathrm{w}\gamma + u \\ \mathrm{x}^* = \mathrm{z}\pi_{\mathrm{z}} + \mathrm{w}\pi_{\mathrm{w}} + v \end{cases} \tag{1}$$

where x^* is correlated with the stochastic disturbance u, and $\mathbf{z} = (\mathrm{z}, \mathrm{w})$ is a vector of exogenous variables. The reduced form representation of (1) is:

$$\begin{cases} \mathrm{y}^* = \mathbf{z}\delta_{\mathrm{y}} + \mathrm{e} \\ \mathrm{x}^* = \mathbf{z}\pi_{\mathrm{x}} + v. \end{cases} \tag{2}$$

Let L_{w} be a selection matrix, such that $\mathrm{w} = \mathbf{z}L_{\mathrm{w}}$. From systems (1) and (2), we have:

$$\mathbf{z}\delta_{\mathrm{y}} = \mathbf{z}\pi_{\mathrm{x}}\beta_1 + \mathbf{z}L_{\mathrm{w}}\gamma + \zeta,$$

where $\zeta = u + (v\beta_1 - e)$. Under the assumption that $\mathbb{E}(\zeta|\mathbf{z}) = 0$, it must be the case that

$$\underbrace{\delta_{\mathrm{y}} - \pi_{\mathrm{x}}\beta_1 - L_{\mathrm{w}}\gamma}_{r(\pi,\beta)} = 0.$$

The estimation of $\beta = (\beta_1, \gamma)$ is based on the minimization of the objective function:

$$\left(\hat{\delta}_{\mathrm{y}} - \hat{\pi}_{\mathrm{x}}\beta_1 - L_{\mathrm{w}}\gamma\right)' \hat{\Psi}^{-1} \left(\hat{\delta}_{\mathrm{y}} - \hat{\pi}_{\mathrm{x}}\beta_1 - L_{\mathrm{w}}\gamma\right), \tag{3}$$

where $\hat{\pi} = \mathrm{vec}[\hat{\delta}_{\mathrm{y}}, \hat{\pi}_{\mathrm{x}}]$ are estimates of the reduced form parameters and $\hat{\Psi}$ is a weighting matrix, i.e., we choose β that minimizes the distance of $r(\hat{\pi}, \beta)$ measured by the norm $||\cdot||_{\hat{\Psi}}$.

■

The reduced form vector π can be estimated parametrically or semiparametrically, according to the latent nature of $(\mathrm{y}^*, \mathrm{x}^*)$ and the distribution of (e, v). The consistent estimation of β depends on the identification condition, which is related to the rank of $\frac{\partial r(\pi,\beta)}{\partial \beta}$.[1] We use local concepts of identification as follows: β is identified if $\frac{\partial r(\pi,\beta)}{\partial \beta}$ is a full rank matrix; weakly identified if $\frac{\partial r(\pi,\beta)}{\partial \beta}$ is singular or $\frac{\partial r(\pi,\beta)}{\partial \beta} = \frac{C}{\sqrt{n}}$ where C is a full rank matrix and n is the sample size, and unidentified if $\frac{\partial r(\pi,\beta)}{\partial \beta}$ is a null matrix. In Example 1, the identification of β is given by the rank of π_{z}. If β is weakly identified or unidentified, then the minimum distance estimator is biased and the limiting normal asymptotic results do not hold. Consequently, the usual approach to inference which is based on the limiting distribution of the minimum distance estimator, for example the

[1] $\frac{\partial f(\theta)}{\partial \theta}$, the derivative of a $k_f \times 1$ vector function $f(\theta)$ by the $k_\theta \times 1$ vector θ, is a $k_f \times k_\theta$ is a matrix.

Wald and likelihood-ratio tests, is misleading.

Although the estimation of β is unreliable under weak and non identification cases, we can still perform hypothesis testing about the value of the structural parameter β. In the following section we present tests based on the minimum distance principle which have the null hypothesis of the form $H_0 : \beta = \beta_0$, against the alternative hypothesis $H_a : \beta \neq \beta_0$. The proposed tests are of the correct size even when identification is weak or absent.

3 Generalized Minimum Distance Robust Tests

The objective function (3) is an example of a broad class of minimum distance objective functions. Let π be a $k_\pi \times 1$ vector representing the reduced form parameters, and β be a $m \times 1$ vector of structural parameters. The values of π and β under the true data generating process are π_0 and β_0, respectively. The estimator of π is denoted by $\hat{\pi}$. Let r be a $q \times 1$ real vector value function defined as $r : \mathbb{R}^{k_\pi} \times \mathbb{R}^m \longrightarrow \mathbb{R}^q$, with $r(\pi, \beta)$ as a typical element. The *link function* r represents the distance between structural and reduced form parameters; and q, the number of restrictions imposed on the reduced form parameters, measures the dimension of this distance. These tests rely on the following regularity conditions:

Assumption 1 *(Regularity conditions)*

i. $\hat{\pi} \xrightarrow{p} \pi_0$, $\sqrt{n}(\hat{\pi} - \pi_0) \xrightarrow{d} \mathcal{N}(0, \Lambda_0)$ *where* Λ_0 *is a symmetric, positive definite covariance matrix;* $\hat{\Lambda} \xrightarrow{p} \Lambda_0$.

ii. $r(\pi, \beta)$ *is continuous on* $\mathbb{R}^{k_\pi} \times \mathbb{R}^m$, *differentiable in* π *on a neighborhood of* π_0 *and twice differentiable in* β. *Moreover,* $\operatorname{rank}\left[\frac{\partial r(\pi,\beta)}{\partial \pi}\right] = q$.

iii. $r(\pi_0, \beta_0) = 0$.

The above conditions are the same as the ones commonly adopted in minimum distance estimation - see Newey (1985) assumptions 1 and 2, Lee (1992), assumption 1, and Gourieroux and Monfort (1995), assumption 9.5. Assumption 1.i. states that the reduced form parameter is root-n consistent and asymptotically normal, and the its asymptotic variance matrix is consistently estimable. Newey and McFadden (1994) provide more primitive conditions if $\hat{\pi}$ is a maximum likelihood or a GMM estimator.

In a model combining censored and continuous endogenous variables, Newey (1985) presents conditions for estimating $\sqrt{n}$-consistent and asymptotically normal reduced form parameters which do not rely on the normality distribution of residuals. The definiteness of Λ_0 in 1.*i*., together with differentiability of $r(\pi, \beta)$ in π, and the full rank of $\frac{\partial r(\pi,\beta)}{\partial \pi}$ in 1.*ii*., are necessary for deriving the asymptotic distribution of $r(\hat{\pi}, \beta)$.

Assumption 1 deserves further explanation. First, we do not require that $\frac{\partial r(\pi,\beta)}{\partial \beta}$ is a full rank matrix, which is necessary for estimating β - see Newey (1985), assumption 1, and Lee (1992), assumption 2. Therefore, Assumption 1 holds independently of the structural parameter identification. Second, Kleibergen (2005) uses smoothness of the empirical moment condition for deriving weak identification robust tests. Some limited dependent variable models, such as the symmetric censored and the winsorized least squares discussed in Section 6, have non-differentiable moments. Unlike GMM tests, our tests rely on the differentiability of the link function (Assumption 1.*ii*), and, as a consequence, the reduced form parameters can be estimated semiparametrically. Finally, in binary choice models or in models with a selection equation, a scale normalization on the variance of the residuals is necessary for estimating π.

In Example 1 we use a triangular specification, which results in a linear link function. However, the r function can be nonlinear in β, as in Blundell and Smith (1994). The next example presents a simplified version of their model.

Example 2: Consider the system:[2]

$$\begin{cases} \mathrm{y}^* = \mathrm{x}^*\beta_1 + \mathrm{w}\gamma + u \\ \mathrm{x}^* = \mathrm{y}\beta_2 + \mathrm{z}\pi_\mathrm{z} + v \end{cases} \tag{4}$$

where $\mathrm{y} = \max\{0, \mathrm{y}^*\}$ and x^* is continuously observed. The difference between (1) and (4) is that the observed y, not the latent y^*, determines x^*. When $\mathrm{y}^* > 0$, we derive the quasi reduced form system:

$$\begin{cases} \mathrm{y}^* = \mathrm{z}\delta_\mathrm{z} + \mathrm{w}\delta_\mathrm{w} + e \\ (\mathrm{x}^* - \beta_2\mathrm{y}) = \mathrm{z}\pi_\mathrm{z} + v \end{cases}$$

[2]The general model has a third equation which describes the mechanism in which the first equation is observed. We also impose a coherency statistical restriction by ignoring the observed y in the first equation.

The link function relating the reduced form and the structural parameter $\beta = (\beta_1, \beta_2)$ is:[3]

$$r(\pi, \beta) = \delta_{\mathrm{z}} - \frac{\pi_{\mathrm{z}}\beta_1}{1 - \beta_1\beta_2}, \tag{5}$$

and π_{z} is estimable for a given value of β_2.

■

By the delta method, we find the asymptotic distribution of $r(\hat{\pi}, \beta)$, which is:

$$\sqrt{n}\left(r(\hat{\pi}, \beta) - r(\pi_0, \beta)\right) \xrightarrow{d} \mathcal{N}\left(0, \Psi_\beta\right), \tag{6}$$

where:

$$\Psi_\beta = \left[\frac{\partial r(\pi_0, \beta)}{\partial \pi}\right] \Lambda_0 \left[\frac{\partial r(\pi_0, \beta)}{\partial \pi}\right]'. \tag{7}$$

The identification of β does not affect the derivation of the asymptotic distribution of $r(\hat{\pi}, \beta)$. Define the objective function of the optimal minimum distance estimator, $\mathrm{S}_{\mathrm{MD}}(\beta)$, as:

$$\mathrm{S}_{\mathrm{MD}}(\beta) = \ n \ \left[r(\hat{\pi}, \beta)\right]' \hat{\Psi}_\beta^{-1} \left[r(\hat{\pi}, \beta)\right], \tag{8}$$

where $\hat{\Psi}_\beta = \left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right] \hat{\Lambda} \left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right]'$. From equation (6), we show that $\mathrm{S}_{\mathrm{MD}}(\beta)$ follows a chi-squared distribution with q degrees of freedom under the null hypothesis.

Theorem 1 *(S_{MD}-test) Under Assumption 1 and the null hypothesis $H_0 : \beta = \beta_0$,*

$$\mathrm{S}_{\mathrm{MD}}(\beta_0) \xrightarrow{d} \chi^2(q)$$

independent of whether β is identified or not.

The S_{MD}-test is similar to the S-test proposed by Stock and Wright (2000) derived under the GMM framework. However, it is important to emphasize two differences. First, the link function is not a sample average of empirical moments. Under continuity and differentiability of $r(\pi, \beta)$, the limiting distribution of the S_{MD}-test is solely derived

[3]Lemieux et al. (1994) derive a similar link function, which is $(\delta_{\mathrm{z}} - \frac{\pi_{\mathrm{z}}\beta}{1-\beta},\ \delta_{\mathrm{w}} - \frac{\pi_{\mathrm{z}}}{1-\beta}) = 0$.

from the asymptotic properties of the reduced form parameter estimator $\hat{\pi}$ and its covariance $\hat{\Lambda}$. Second, in nonlinear models, because structural and nuisance parameters are nonseparable, testing a structural parameter involves the estimation of untested parameters under the null hypothesis. This is not the case for our tests. For the linear link function in Section 2, γ does not have to be estimated in order to test β_1. This property, illustrated in Sections 4 and 6, has important computational advantages, especially for the estimation of confidence intervals by inverting the statistical tests.

When the model is overidentified $(q > m)$, the $\mathrm{S_{MD}}$-test tests two hypotheses simultaneously: the value of the structural parameter vector and the $(q-m)$ overidentification restrictions. As a result, this test loses power under the alternative hypothesis along with an increasing number of overidentification restrictions.

As in Kleibergen (2007), we can decompose the $\mathrm{S_{MD}}$-test into two orthogonal statistics: $\mathrm{K_{MD}}$ and $\mathrm{J_{MD}}$. The first of these statistics tests only the value of the structural parameter, while the latter tests the overidentification restrictions.

Theorem 2 *($\mathrm{K_{MD}}$- and $\mathrm{J_{MD}}$-tests) Define the $\mathrm{K_{MD}}$- and $\mathrm{J_{MD}}$-tests as:*

$$\mathrm{K_{MD}}(\beta_0) = n \left[\hat{\Psi}_{\beta_0}^{-\frac{1}{2}} r(\hat{\pi},\beta_0)\right]' \mathrm{P}_{\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\hat{\mathrm{D}}_{\beta_0}} \left[\hat{\Psi}_{\beta_0}^{-\frac{1}{2}} r(\hat{\pi},\beta_0)\right],$$

$$\mathrm{J_{MD}}(\beta_0) = n \left[\hat{\Psi}_{\beta_0}^{-\frac{1}{2}} r(\hat{\pi},\beta_0)\right]' \mathrm{M}_{\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\hat{\mathrm{D}}_{\beta_0}} \left[\hat{\Psi}_{\beta_0}^{-\frac{1}{2}} r(\hat{\pi},\beta_0)\right],$$

where:

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

$$\mathrm{M}_{\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\hat{\mathrm{D}}_{\beta_0}} = \mathrm{I}_q - \mathrm{P}_{\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\hat{\mathrm{D}}_{\beta_0}},$$

$$\hat{\mathrm{D}}_{\beta_0} = \left[\hat{\mathrm{D}}_1(\beta_0) \quad \ldots \quad \hat{\mathrm{D}}_m(\beta_0)\right],$$

$$\hat{\mathrm{D}}_j(\beta_0) = \frac{\partial r(\hat{\pi},\beta_0)}{\partial \beta_j} - \left[\frac{\partial}{\partial \pi}\left(\frac{\partial r(\hat{\pi},\beta_0)}{\partial \beta_j}\right)\right] \hat{\Lambda} \left[\frac{\partial r(\hat{\pi},\beta_0)}{\partial \pi}\right]' \hat{\Psi}_{\beta_0}^{-1} r(\hat{\pi},\beta_0),$$

for $j = 1,\ldots,m$. Under assumption 1 and $\mathrm{H}_0 : \beta = \beta_0$, we have:

$$\mathrm{K_{MD}}(\beta_0) \xrightarrow{d} \chi^2(m) \quad \textit{and} \quad \mathrm{J_{MD}}(\beta_0) \xrightarrow{d} \chi^2(q-m),$$

regardless of whether β is point-identified. Also,

$$\mathrm{S}_{\mathrm{MD}}(\beta_0) = \mathrm{K}_{\mathrm{MD}}(\beta_0) + \mathrm{J}_{\mathrm{MD}}(\beta_0). \tag{9}$$

Unlike the S_{MD}, the K_{MD}-test is not affected by the number of overidentifying restrictions. The statistic $\hat{\mathrm{D}}_{\beta_0}$ is asymptotically independent of $r(\hat{\pi}, \beta_0)$ under the null hypothesis. If β is identified, then $\hat{\mathrm{D}}_{\beta_0}$ converges in probability to $\frac{\partial r(\pi_0,\beta_0)}{\partial \beta}$. If not, i.e. $\frac{\partial r}{\partial \beta}$ is close to a reduced rank value, then $\sqrt{n}\ \hat{\mathrm{D}}_{\beta_0}$ converges in distribution to a random variable. Because of the asymptotic independence between $\hat{\mathrm{D}}_{\beta_0}$ and $r(\hat{\pi}, \beta_0)$, the distribution of the K_{MD}-test, conditional on $\hat{\mathrm{D}}_{\beta_0}$, is free from nuisance parameters, see Kleibergen (2005). Moreover, its unconditional distribution is pivotal.

The derivative of the S_{MD} with respect to β, as shown in the Appendix, is:

$$-\frac{1}{2}\frac{\partial \mathrm{S}_{\mathrm{MD}}(\beta)}{\partial \beta} = \ n\ r(\hat{\pi}, \beta)' \hat{\Psi}_\beta^{-1} \hat{\mathrm{D}}_\beta. \tag{10}$$

The K_{MD}-test is the quadratic form of equation (10), weighted by its own variance. The minimum value of $\mathrm{S}_{\mathrm{MD}}(\beta)$ coincides with the point where the K_{MD}-test equals zero. This point is the generalized minimum distance continuous updating estimator (GMD-CUE), which is different from the GMM-CUE, the minimizer of the S-test.

The J_{MD}-test is related to the overidentification test proposed by Lee (1992). The latter results from substituting a minimum distance estimator for β in the S_{MD}-test. Therefore, it is not robust to identification failure.

Equation (10) shows that the K_{MD}-test will suffer from a spurious decline of power at inflection and local extrema points of the S_{MD}-test. Close to these points, the value of the J_{MD}-test approximates the value of the S_{MD}-test, which has discriminatory power - see equation (9). We define a new test for the structural parameter, the $\mathrm{KJ}_{\mathrm{MD}}$-test, by combining both the K_{MD}- and the J_{MD}-tests, see Kleibergen (2005). Let $\tau_{\mathrm{K}_{\mathrm{MD}}}$ and $\tau_{\mathrm{J}_{\mathrm{MD}}}$ be the significance levels of the K_{MD}- and J_{MD}-tests, respectively. The $\mathrm{KJ}_{\mathrm{MD}}$-test has an approximate significance level of $\tau = \tau_{\mathrm{K}_{\mathrm{MD}}} + \tau_{\mathrm{J}_{\mathrm{MD}}}$. Rejection of $\mathrm{KJ}_{\mathrm{MD}}$ occurs if K_{MD} rejects at $\tau_{\mathrm{K}_{\mathrm{MD}}}$ or if J_{MD} rejects at $\tau_{\mathrm{J}_{\mathrm{MD}}}$ significance levels. The J_{MD} component of the $\mathrm{KJ}_{\mathrm{MD}}$-test corrects the decline of power that affects the K_{MD}-test at the inflexion and local minima points.

An extension of Moreira's (2003) conditional likelihood-ratio (CLR) to the current framework, under the null hypothesis, is:

$$\mathrm{CLR}_{\mathrm{MD}}(\beta_0) = \frac{1}{2}\left\{ \mathrm{S}_{\mathrm{MD}}(\beta_0) - \mathrm{rk}(\beta_0) + \sqrt{\left[\mathrm{S}_{\mathrm{MD}}(\beta_0) + \mathrm{rk}(\beta_o)\right]^2 - 4\mathrm{J}_{\mathrm{MD}}(\beta_o)\mathrm{rk}(\beta_0)} \right\}, \quad (11)$$

where $\mathrm{rk}(\beta_0)$ is a statistic that tests the rank of $\hat{\mathrm{D}}_{\beta_0}$, see Kleibergen (2005). In case of one endogenous variable, $\mathrm{rk}(\beta_0) = \; n\left\{\hat{\mathrm{D}}_{\beta_0}'\hat{\Xi}_{\beta_0}^{-1}\hat{\mathrm{D}}_{\beta_0}\right\}$, where $\hat{\Xi}_{\beta_0}$ is the variance estimate of $\hat{\mathrm{D}}_{\beta_0}$, described by equation (21) in the Appendix. The presence of the $\mathrm{S}_{\mathrm{MD}}(\beta_0)$ statistic in (11) shows that the $\mathrm{CLR}_{\mathrm{MD}}$-test does not have the the spurious decline of power of the K_{MD}-test.

The asymptotic distribution of the $\mathrm{CLR}_{\mathrm{MD}}$ is not pivotal and depends on $\mathrm{rk}(\beta)$. The critical values of this test are calculated by simulating independent values of $\chi^2(m)$ and $\chi^2(q-m)$ random variables for a given value of $\mathrm{rk}(\beta)$.

Proposition 1 *($\mathrm{CLR}_{\mathrm{MD}}$-test) Under Assumption 1 and the null hypothesis we have:*

$$\mathrm{CLR}_{\mathrm{MD}}(\beta_0) \xrightarrow{p} \frac{1}{2}\left\{ \bar{\chi}_m + \bar{\chi}_{q-m} - \mathrm{rk}(\beta_0) + \sqrt{\left[\bar{\chi}_m + \bar{\chi}_{q-m} + \mathrm{rk}(\beta_0)\right]^2 - 4\bar{\chi}_{q-m}\mathrm{rk}(\beta_0)} \right\},$$

where $\bar{\chi}_m$ and $\bar{\chi}_{q-m}$ are independent chi-squared distributed random variables with m and $q-m$ degrees of freedom, respectively.

In the linear instrumental variable model with homoskedastic residuals, Andrews et al. (2006) show that the CLR-test dominates the S- and K-tests in terms of power. However, this result is not yet extended to a more general class of models. The simulations reported in Section 5 have an example in which the $\mathrm{CLR}_{\mathrm{MD}}$-test does not dominate the K_{MD}-test in terms of power.

The proposed tests can be adapted in order to test only a subset of the structural parameter vector. The procedure consists of estimating the untested structural parameters under the null hypothesis by the GMD-CUE and replacing the estimated values in the original tests. If the estimated parameters are identified, the S_{MD}- and K_{MD}-tests remain chi-squared distributed with degrees of freedom reduced by the number of estimated parameters. If the estimated parameter is not identified, Kleibergen and Mavroeidis (2009) show that the limiting distributions of the tests are asymptotically bounded by the adjusted chi-squared distributions.

There is a correspondence between our tests and other robust tests for the linear instrumental variable model represented by the following system:

$$\begin{cases} \mathrm{y} = \mathrm{x}\beta_0 + u \\ \mathrm{x} = \mathrm{z}\pi_{\mathrm{z}} + v \end{cases}$$

We omit the included instruments, w, to simplify the exposition. The AR and K-tests are, respectively,

$$\mathrm{AR}(\beta_0) = \frac{(\mathrm{y} - \mathrm{x}\beta_0)' \, \mathrm{P_z} \, (\mathrm{y} - \mathrm{x}\beta_0)}{\hat{\sigma}^2_{\beta_0}} \quad \text{and} \quad \mathrm{K}(\beta_0) = \frac{(\mathrm{y} - \mathrm{x}\beta_0)' \, \mathrm{P}_{\mathrm{z}\tilde{\pi}_{\mathrm{z}}(\beta_0)} \, (\mathrm{y} - \mathrm{x}\beta_0)}{\hat{\sigma}^2_{\beta_0}},$$

where $\mathrm{P_a} = \mathrm{a}(\mathrm{a'a})^{-1}\mathrm{a}'$,

$$\hat{\sigma}^2_{\beta_0} = \frac{(\mathrm{y} - \mathrm{x}\beta_0)' \, \mathrm{M_z} \, (\mathrm{y} - \mathrm{x}\beta_0)}{n - k_{\mathrm{z}}} \quad \text{and} \quad \tilde{\pi}_{\mathrm{z}}(\beta_0) = (\mathrm{z'z})^{-1}\mathrm{z}' \left(\mathrm{x} - (\mathrm{y} - \mathrm{x}\beta_0) \frac{(\mathrm{y} - \mathrm{x}\beta_0)' \, \mathrm{M_z x}}{(n - k_{\mathrm{z}}) \; \hat{\sigma}^2_{\beta_0}} \right).$$

In the Appendix, we show that the AR- and K-tests are the same as the following $\mathrm{S_{MD}}$- and $\mathrm{K_{MD}}$-tests:

$$\begin{aligned} \mathrm{S_{MD}}(\beta_0) &= (\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0)' \hat{\Psi}^{-1}_{\beta_0} (\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0), \\ \mathrm{K_{MD}}(\beta_0) &= \left(\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0\right)' \hat{\Psi}^{-\frac{1}{2}'}_{\beta_0} \mathrm{P}_{\hat{\mathrm{D}}_{\beta_0}} \hat{\Psi}^{-\frac{1}{2}}_{\beta_0} \left(\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0\right), \end{aligned}$$

where $\hat{\delta}_{\mathrm{z}} = (\mathrm{z'z})^{-1}\mathrm{z'y}$, $\hat{\pi}_{\mathrm{z}} = (\mathrm{z'z})^{-1}\mathrm{z'x}$, $\hat{\Psi}_{\beta_0} = \hat{\sigma}^2_{\beta_0} \; (\mathrm{z'z})$ and $\hat{\mathrm{D}}_{\beta_0} = -\tilde{\pi}_{\mathrm{z}}(\beta_0)$.

In comparison with the tests proposed by Chernozhukov and Hansen (2008), the $\mathrm{S_{MD}}$ is the same as the Wald$-$S, but the $\mathrm{K_{MD}}$ differs from the Wald$-$K, which is based on testing $H_0 : \eta = 0$ in the following Gauss-Newton regression:

$$\mathrm{y} - \mathrm{x}\beta_0 = \mathrm{z}\tilde{\pi}_{\mathrm{z}}(\beta_0)\eta + \text{residuals}.$$

In the case of heteroskedastic or clustered residuals, these tests differ from the other tests in small samples. However, they are all asymptotically equivalent (see Appendix).

4 Implementation of Robust Tests

We provide an algorithm for the implementation of the proposed tests specific to the following class of limited dependent variable models:[4]

$$\begin{cases} \mathrm{y}^{*} = \mathrm{x}\beta + \mathrm{w}\gamma + u \\ \mathrm{x} = \mathrm{z}\pi_{\mathrm{z}} + \mathrm{w}\pi_{\mathrm{w}} + v \end{cases} \qquad \begin{cases} \mathrm{y}^{*} = \mathrm{z}\delta_{\mathrm{z}} + \mathrm{w}\delta_{\mathrm{w}} + e \\ \mathrm{x} = \mathrm{z}\pi_{\mathrm{z}} + \mathrm{w}\pi_{\mathrm{w}} + v \end{cases} \tag{12}$$

where y^* is latent and x is continuous and fully observed. The link function $r(\pi, \beta)$ derived from the system (12) is:

$$r(\pi, \beta) = \delta_{\mathrm{z}} - \pi_{\mathrm{z}}\beta, \tag{13}$$

where $\pi = \mathrm{vec}[\delta_{\mathrm{z}}, \pi_{\mathrm{z}}]$. In this model, $\frac{\partial r(\pi,\beta)}{\partial \pi} = [I_{k_{\mathrm{z}}} \; - \beta' \otimes I_{k_{\mathrm{z}}}]$ is of full rank, independent of the value of β. The variance of $r(\hat{\pi}, \beta)$, the $\hat{\mathrm{D}}_\beta$ statistic of the K_{MD}-test, and its variance matrix Ξ_β, which is necessary for computing the $\mathrm{CLR}_{\mathrm{MD}}$-test, are:

$$\Psi_\beta = \Lambda_{\delta_{\mathrm{z}}\delta_{\mathrm{z}}} - (\beta \otimes I_{k_{\mathrm{z}}})'\Lambda_{\pi_{\mathrm{z}}\delta_{\mathrm{z}}} - \Lambda_{\delta_{\mathrm{z}}\pi_{\mathrm{z}}}(\beta \otimes I_{k_{\mathrm{z}}}) + (\beta \otimes I_{k_{\mathrm{z}}})'\Lambda_{\pi_{\mathrm{z}}\pi_{\mathrm{z}}}(\beta \otimes I_{k_{\mathrm{z}}}), \tag{14a}$$

$$\mathrm{vec}\left[\hat{\mathrm{D}}_\beta\right] = -\mathrm{vec}[\hat{\pi}_{\mathrm{z}}] + [\Lambda_{\pi_{\mathrm{z}}\delta_{\mathrm{z}}} - \Lambda_{\pi_{\mathrm{z}}\pi_{\mathrm{z}}}(\beta \otimes I_{k_{\mathrm{z}}})]\,\Psi_\beta^{-1}\left(\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta\right), \tag{14b}$$

$$\Xi_\beta = \Lambda_{\pi_{\mathrm{z}}\pi_{\mathrm{z}}} - [\Lambda_{\pi_{\mathrm{z}}\delta_{\mathrm{z}}} - \Lambda_{\pi_{\mathrm{z}}\pi_{\mathrm{z}}}(\beta \otimes I_{k_{\mathrm{z}}})]\,\Psi_\beta^{-1}\left[\Lambda_{\delta_{\mathrm{z}}\pi_{\mathrm{z}}} - (\beta \otimes I_{k_{\mathrm{z}}})'\,\Lambda_{\pi_{\mathrm{z}}\pi_{\mathrm{z}}}\right]. \tag{14c}$$

where $\Lambda_{\delta_{\mathrm{z}}\delta_{\mathrm{z}}}$ and $\Lambda_{\pi_{\mathrm{z}}\pi_{\mathrm{z}}}$ are the asymptotic variances of $\sqrt{n}(\hat{\delta}_{\mathrm{z}} - \delta_{\mathrm{z}})$ and $\sqrt{n}[\mathrm{vec}(\hat{\pi}_{\mathrm{z}} - \pi_{\mathrm{z}})]$, respectively, and $\Lambda_{\pi_{\mathrm{z}}\delta_{\mathrm{z}}} = \Lambda'_{\delta_{\mathrm{z}}\pi_{\mathrm{z}}}$ is their asymptotic covariance.

We can also estimate the reduced form parameters by introducing a linear control function $v\alpha$. Then, the structural and reduced form equations become:

$$\begin{cases} \mathrm{y}^{*} = \mathrm{x}\beta + \mathrm{w}\gamma + v\alpha + \varepsilon \\ \mathrm{x} = \mathrm{z}\pi_{\mathrm{z}} + \mathrm{w}\pi_{\mathrm{w}} + v \end{cases} \qquad \begin{cases} \mathrm{y}^{*} = \mathrm{z}\delta_{\mathrm{z}} + \mathrm{w}\delta_{\mathrm{w}} + v\delta_v + \varepsilon \\ \mathrm{x} = \mathrm{z}\pi_{\mathrm{z}} + \mathrm{w}\pi_{\mathrm{w}} + v \end{cases} \tag{15}$$

where $\varepsilon = u - v\alpha$ and $\delta_v = \beta + \alpha$. We demonstrate that the use of the control function allows us to write $\Lambda_{\pi_{\mathrm{z}}\delta_{\mathrm{z}}} = \Lambda_{\pi_{\mathrm{z}}\pi_{\mathrm{z}}}[(\delta_v \otimes I_{k_{\mathrm{z}}})]$. The elements for computing the tests reduce

[4]Finlay and Magnusson (2009) have files available for implementing tests for the instrumental variable probit and Tobit models in *STATA*. These files are downloadable from http://greenspace.tulane.edu/kfinlay/research.html.

to:

$$\Psi_\beta = \Lambda_{\delta_z\delta_z} + [(\delta_v - \beta)' \otimes I_{k_z}]\, \Lambda_{\pi_z\pi_z}\, [(\delta_v - \beta) \otimes I_{k_z}]\,, \tag{16a}$$

$$\mathrm{vec}[\mathrm{D}_\beta] = -\mathrm{vec}[\pi_z] + \Lambda_{\pi_z\pi_z}\mathrm{vec}\big[\Psi_\beta^{-1}(\delta_z - \pi_z\beta)(\delta_v - \beta)'\big], \tag{16b}$$

$$\Xi_\beta = \Lambda_{\pi_z\pi_z} - \{\Lambda_{\pi_z\pi_z}\,[(\delta_v - \beta) \otimes I_{k_z}]\}\, \Psi_\beta^{-1}\, \{[(\delta_v - \beta)' \otimes I_{k_z}]\, \Lambda_{\pi_z\pi_z}\}\,. \tag{16c}$$

Further simplification is possible by assuming that v is homoskedastic (see the Appendix). The algorithm takes the following steps:

1. Estimate π_z and $\mathrm{Var}[\sqrt{n}(\hat{\pi}_z - \pi_z)]$ by OLS. Denote the estimated values by $\hat{\pi}_z$ and $\hat{\Lambda}_{\pi_z\pi_z}$. Compute $\hat{v}_i$, the OLS residuals.

2. Estimate δ_z, δ_w and δ_v from the following equation:

   $$\mathrm{y} = f\left(\mathrm{z}\delta_z + \mathrm{w}\delta_w + \hat{v}\delta_v + \tilde{\varepsilon}\right)$$

   where $f(\cdot)$ is a known function and $\tilde{\varepsilon} = \varepsilon - (\hat{v} - v)\delta_v$. Denote the estimates of δ_z and δ_v by, respectively, $\hat{\delta}_z$ and $\hat{\delta}_v$. We do not have to keep the estimate of δ_w because it is not part of the link function (13).

3. Save $\hat{\Lambda}_{\delta_z\delta_z}$, the output of the variance-covariance matrix estimate of $\hat{\delta}_z$.

4. Finally, substitute $\hat{\Lambda}_{\pi_z\pi_z}$, $\hat{\pi}_z$, and $\hat{\delta}_v$ into equations (16) with the hypothesized value of β.

5 Simulations

We simulate the endogenous Tobit and the endogenous Poisson count data models, which can be represented by the simultaneous system (12). In both cases, $\hat{\pi}_x = (\hat{\pi}_z, \hat{\pi}_w)$ is the ordinary least squares estimate. We estimate $\delta_y = (\delta_z, \delta_w)$ using Powell's (1986) symmetric censored least squares (SCLS) and Poisson quasi-likelihood method for, respectively, the endogenous Tobit model and the endogenous Poisson count data model. We compare the performance of the proposed robust tests with the Wald tests, defined as:

$$\mathrm{W}(\beta_0) = \left(\hat{\beta} - \beta_0\right) \hat{\mathrm{V}}_{\hat{\beta}}^{-1} \left(\hat{\beta} - \beta_0\right),$$

where $\hat{\beta}$ is an estimate of β, and $\hat{\mathrm{V}}_{\hat{\beta}}$ is the estimated variance of $\hat{\beta}$ evaluated at $\hat{\beta}$.

We compute the rejection frequency of the tests at 10%, 5% and 1% significance levels. For the $\mathrm{KJ_{MD}}$-test, the significance level of $\mathrm{K_{MD}}$ is four times the significance level of $\mathrm{J_{MD}}$. For both simulations, we generate 10,000 random samples of 200 observations each, satisfying the following conditions: $\beta_0 = 0$; w_i is an unitary constant; z_i is a 1×3 row vector drawn from independent uniform distributions and kept fixed in all simulations; and $\pi_{\mathrm{z}} = (\pi_{\mathrm{z}_1}, 0, 0)'$ is a 3×1 column vector. The value of π_{z_1} is set according to μ, the concentration parameter divided by k_{z}:

$$\mu = \frac{1}{k_{\mathrm{z}}}\left(\frac{\pi_{\mathrm{z}_1}'\mathrm{z}_1'\mathrm{z}_1\pi_{\mathrm{z}_1}}{\sigma_v^2}\right),$$

where z_1 is the $n \times 1$ vector of the first instrument and σ_v^2 is the variance of v_i. We chose μ to be 30 and 3, representing strong and weak identification, respectively.[5]

5.1 Endogenous Tobit

The endogenous Tobit model is represented by:

$$\begin{cases} \mathrm{y}_i = \max\{0, \mathrm{x}_i\beta + \mathrm{w}_i\gamma + u_i\} \\ \mathrm{x}_i = \mathrm{z}_i\pi_{\mathrm{z}} + \mathrm{w}_i\pi_{\mathrm{w}} + v_i \end{cases} \tag{17}$$

We consider two cases for the joint distribution of (u_i, v_i): a bivariate Laplace distribution with zero mean and unit variance, and a bivariate t-distribution with three degrees of freedom. In both cases, we first generate bivariate uniform distributed random variables with correlation coefficient ρ. Then, we generate the residuals $\{(u_i, v_i)\}_{i=1}^n$ using the inverse of the cumulative distribution functions; ρ is either 0.2 or 0.9, and $\pi_{\mathrm{w}} = 0.2$. The parameter γ takes on values of 0.7267 and 0.4901 for the t-Student and the Laplace residuals, respectively. We calibrate γ such that, on average, 25% of the observations are left censored. In computing the robust tests, we assumed that residuals are heteroskedastic of unknown form. The statistics are based on the elements defined in (14).

For estimating $\delta_{\mathrm{y}} = (\delta_{\mathrm{z}}, \delta_{\mathrm{w}})$ by SCLS we use the algorithm proposed by Silva (2001). This algorithm converges faster and more frequently compared to the original algorithm

[5] In linear instrumental variable models, Staiger and Stock (1997) suggest that values of μ below 10 indicate that the instruments are weak.

in Powell (1986). However, between 0.5% and 1% of the simulations did not converge. The Wald test is derived from the two-step minimum distance estimator in Lee (1995).

Table 1 reveals that the sizes of the Wald tests become distorted when identification decreases. The distortion varies according to the degree of endogeneity: the tests under-reject the null hypothesis when $\rho = 0.2$ and over-reject it when $\rho = 0.9$. These results are related to the bias of the minimum distance estimator of β: the lower the degree of identification and the higher the endogeneity, the more upwardly biased the two-step estimates are. Different from the Wald tests, the performance of our tests are neither affected by the level of identification nor by the degree of endogeneity.

Table 1: Size Comparison (in percentage) $H_0 : \beta = 0$, Endogenous Censored Model[a]

	$\mu = 30$						$\mu = 3$					
	$\rho = 0.2$			$\rho = 0.9$			$\rho = 0.2$			$\rho = 0.9$		
ACV[b]	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Laplace residuals												
Wald_{MD}	9.23	4.39	0.68	12.24	7.45	2.41	3.68	1.13	0.05	22.34	13.45	3.54
S_{MD}	12.07	6.87	2.08	12.24	6.41	1.85	12.07	6.87	2.08	12.24	6.41	1.85
K_{MD}	11.12	5.88	1.40	11.43	6.05	1.28	10.94	5.83	1.40	12.84	7.30	1.76
J_{MD}	11.54	6.66	1.86	11.43	6.33	1.57	11.70	6.81	1.89	10.45	5.43	1.41
KJ_{MD}	11.48	6.42	1.64	11.82	6.26	1.53	11.50	6.53	1.61	12.73	7.21	1.81
CLR_{MD}	11.12	6.43	1.32	11.64	6.69	1.31	11.68	6.68	1.97	13.36	7.34	1.87
t-Student residuals												
Wald_{MD}	9.93	4.91	0.88	12.60	7.46	2.37	4.18	1.30	0.07	20.25	12.03	3.28
S_{MD}	13.19	7.84	2.55	13.31	7.34	2.33	13.19	7.84	2.55	13.31	7.34	2.33
K_{MD}	11.89	6.47	1.68	12.07	6.67	1.57	11.77	6.51	1.69	13.40	7.47	1.91
J_{MD}	12.38	7.32	2.19	12.09	6.99	1.81	12.45	7.41	2.20	11.53	6.20	1.73
KJ_{MD}	12.64	7.35	2.09	12.98	7.13	1.98	12.85	7.42	2.11	13.57	7.75	2.19
CLR_{MD}	11.82	7.12	1.76	12.30	7.39	1.60	12.63	7.70	2.37	13.88	8.18	2.24

[a] 10,000 simulations; 200 observations per simulation.
[b] ACV: asymptotic critical value.

5.2 Poisson Count Data Model

The following system is a representation of the endogenous Poisson count data model:

$$\begin{cases} \mathrm{y}_i \sim Poisson(\lambda_i) \\ \lambda_i = \exp\left(\mathrm{x}_i\beta + \mathrm{w}_i\gamma + u_i\right) \\ \mathrm{x}_i = \mathrm{z_i}\pi_\mathrm{z} + \mathrm{w_i}\pi_\mathrm{w} + v_i \end{cases} \tag{18}$$

where λ_i is the mean of the Poisson distribution. We analyze the performance of the tests for over, equally and underdispersed data. Because the results are similar, we only report the equally dispersed case.[6] We also analyze tests that consider the presence of a linear control function, as described in equation (15).

Realizations from the system (18) are generated according to the following steps. First, we sample the random variables (ν_1, ν_2) from a bivariate uniform distribution, with correlation coefficient $\rho = \{0.2, 0.9\}$. We set $v = \nu_1$; y results from the inverse Poisson distribution, evaluated at ν_2. By fixing $\pi_\mathrm{w} = -0.5$ and $\gamma = \log(2.0)$, we obtain $\mathbb{E}[\mathrm{w}\pi_\mathrm{w} + v] = 0$ and $\mathbb{E}[\mathrm{y}] = \mathrm{Var}[\mathrm{y}] = 2$.

We also investigate the performance of the GMM robust tests in Stock and Wright (2000) and Kleibergen (2005), which are derived from the following empirical moment condition proposed by Mullahy (1997):

$$\frac{1}{n}\sum_{i}^{n}\begin{bmatrix}\mathrm{z}_i' \\ \mathrm{w}_i'\end{bmatrix}\left[\exp(-\mathrm{x}_i\beta - \mathrm{w}_i\gamma)\mathrm{y}_i - 1\right]. \tag{19}$$

The estimation of γ under the null hypothesis is necessary for computing the robust GMM tests. The non-robust tests are listed according to the method used for estimating β: two-step GMM or two-step minimum distance.

The results in Table 2 show that changes in the level of identification affect the behavior of the Wald tests. Similar to the endogenous Tobit model, they under-reject the null hypothesis when $\rho = 0.2$ and over-reject it when $\rho = 0.9$. For example, the rejection probability of the Wald two-step minimum distance test jumps from 3.78% when $\rho = 0.2$ to 43.08% when $\rho = 0.9$, while it is supposed to be 10%.

[6] The remaining results are available on the author's web site.

Table 2: Size Comparison (in percentage) $H_0 : \beta = 0$, Endogenous Poisson Count Data Model, equidispersion case[a]

	$\mu = 30$						$\mu = 3$					
	$\rho = 0.2$			$\rho = 0.9$			$\rho = 0.2$			$\rho = 0.9$		
ACV[b]	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Wald_{GMM}	8.86	3.84	0.48	14.07	9.14	3.74	4.12	1.23	0.05	30.84	22.49	10.59
Wald_{MD}	8.41	3.55	0.51	15.04	10.33	4.71	3.78	1.22	0.07	43.08	35.95	23.83
$\text{Wald}^{\text{c}}_{\text{MD}}$	9.96	4.47	0.75	15.87	11.39	5.40	4.64	1.61	0.08	46.42	39.55	27.79
S_{MD}	9.25	4.71	1.11	9.01	4.44	0.96	9.25	4.71	1.11	9.01	4.44	0.96
K_{MD}	10.22	4.73	0.95	9.62	4.80	0.73	10.12	5.10	1.04	9.41	4.74	0.86
J_{MD}	9.29	4.51	1.07	8.55	4.56	0.87	9.33	4.58	1.05	9.11	4.40	0.97
KJ_{MD}	9.79	4.79	0.94	9.30	4.70	0.81	9.96	5.19	0.89	9.35	4.76	0.81
CLR_{MD}	10.08	5.07	0.92	9.62	5.11	0.72	9.90	5.23	1.03	9.19	4.94	0.78
$\text{S}^{\text{c}}_{\text{MD}}$	12.44	6.60	1.76	10.09	5.22	0.89	12.44	6.60	1.76	10.09	5.22	0.89
$\text{K}^{\text{c}}_{\text{MD}}$	12.06	6.03	1.50	9.94	5.01	0.90	11.95	6.52	1.52	9.75	5.07	0.93
$\text{J}^{\text{c}}_{\text{MD}}$	11.51	6.06	1.51	9.92	5.03	1.10	11.65	6.24	1.52	10.19	5.21	1.09
$\text{KJ}^{\text{c}}_{\text{MD}}$	12.01	6.23	1.50	10.13	5.02	0.91	12.18	6.67	1.58	9.90	5.01	0.96
$\text{CLR}^{\text{c}}_{\text{MD}}$	12.08	6.73	1.38	9.98	5.39	0.88	12.55	6.88	1.45	9.89	5.22	0.88
S_{GMM}	10.23	4.88	0.97	9.56	4.65	0.68	10.23	4.88	0.97	9.56	4.65	0.68
K_{GMM}	9.64	4.40	0.80	8.97	4.42	0.66	9.73	4.80	0.81	9.09	4.36	0.76
J_{GMM}	10.84	5.53	1.11	10.53	5.07	0.91	11.00	5.30	1.02	10.27	4.89	0.88
KJ_{GMM}	9.37	4.42	0.81	9.20	4.37	0.66	9.69	4.78	0.76	9.03	4.34	0.74
CLR_{GMM}	9.66	4.74	0.77	9.09	4.76	0.55	9.82	4.78	0.77	9.39	4.53	0.63

[a] 10,000 simulations; 200 observations per simulation.
[b] ACV: asymptotic critical value.
[c] Uses the control function $v\alpha$.

The proposed and the GMM robust tests' rejection probabilities are close to the expected asymptotic critical values, regardless of the level of endogeneity, the degree of dispersion or the identification strength. The introduction of a control function has ambiguous results. It makes the nominal sizes closer to the asymptotic sizes when $\rho = 0.9$, and the opposite when $\rho = 0.2$.

5.3 Power - Endogenous Count Data Model

We investigate the power of the proposed tests for the endogenous count data model using the same data generating process of Subsection 5.2. We only report the results of the tests which do not incorporate a control function and therefore are less efficient.

Figure 1 compares the robust tests and the two-step GMM test according to their degree of endogeneity and identification “strength”. First, the size of the robust tests remain correct in all graphs, while the Wald test is biased even in cases in which the identification is relatively strong. Second, the S_{MD}-test has less discriminatory power than the K_{MD}-test, which is explained by the number of overidentification restrictions. Third, in three out of the four cases, the $\mathrm{CLR}_{\mathrm{MD}}$-test dominates the remaining robust tests. Finally, we also note that the K_{MD} suffers a decline of power at values differing from the hypothesized value.

Figure 2 illustrates the power of our tests and of robust GMM. We observe that the $\mathrm{KJ}_{\mathrm{MD}}$ and $\mathrm{CLR}_{\mathrm{MD}}$ tests dominate the $\mathrm{KJ}_{\mathrm{GMM}}$ and $\mathrm{CLR}_{\mathrm{GMM}}$, respectively, in all cases. The $\mathrm{S}_{\mathrm{GMM}}$ dominates the S_{MD}-test for positive values of β. The results are even more favorable when using a control function.

6 Two Applications

We illustrate the use of the robust tests by constructing confidence intervals and regions for the two models of Section 5: endogenous Tobit and Poisson count data. The former model is illustrated by the married female labor supply, see Blundell and Smith (1989) and Lee (1995), while the latter is exemplified by the demand for cigarettes, see Mullahy (1997). The $1-\tau$ confidence set is formed by the points of the parameter space which do not reject the null hypothesis at significance level τ.

6.1 Female Labor Supply

Consider the married female labor supply model of Blundell and Smith (1989). In equation (17), y_i represents weekly hours in paid work, and x_i is other household income measured in US$1,000.00, which includes unearned income and savings. Besides a constant term, w_i includes demographic variables: female age and its square, education

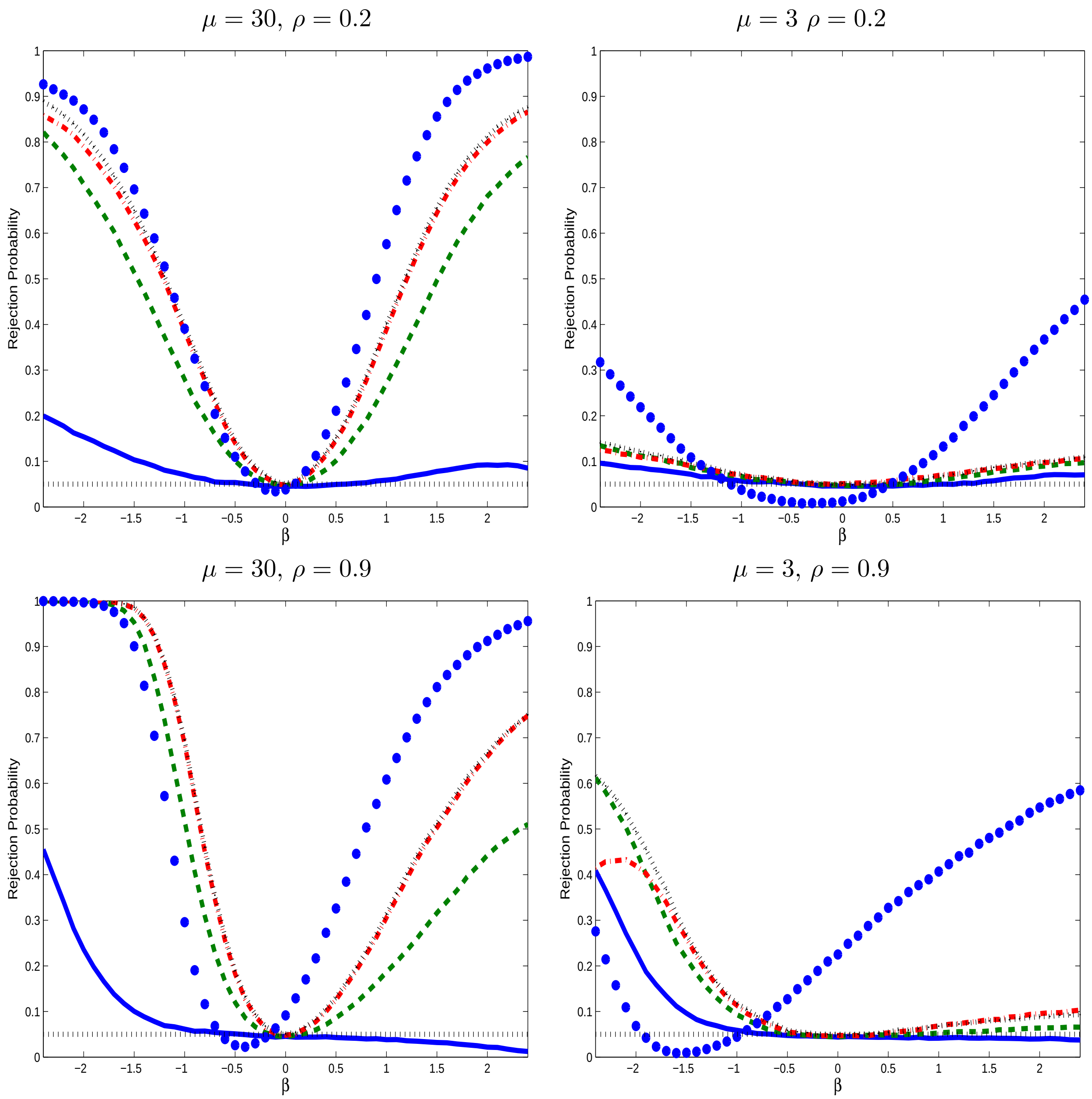

Figure 1: Power Curves of S_{MD} (dashed line), J_{MD} (solid line), K_{MD} (dash-dot line), $\mathrm{CLR}_{\mathrm{MD}}$ (dotted line), and $\mathrm{Wald}_{\mathrm{GMM}}$ ('circle') for testing $H_0 : \beta = 0$ at 5% significance level in the endogenous count data model. Sample size: 200 observations. Monte Carlo simulations: 10,000 replications.

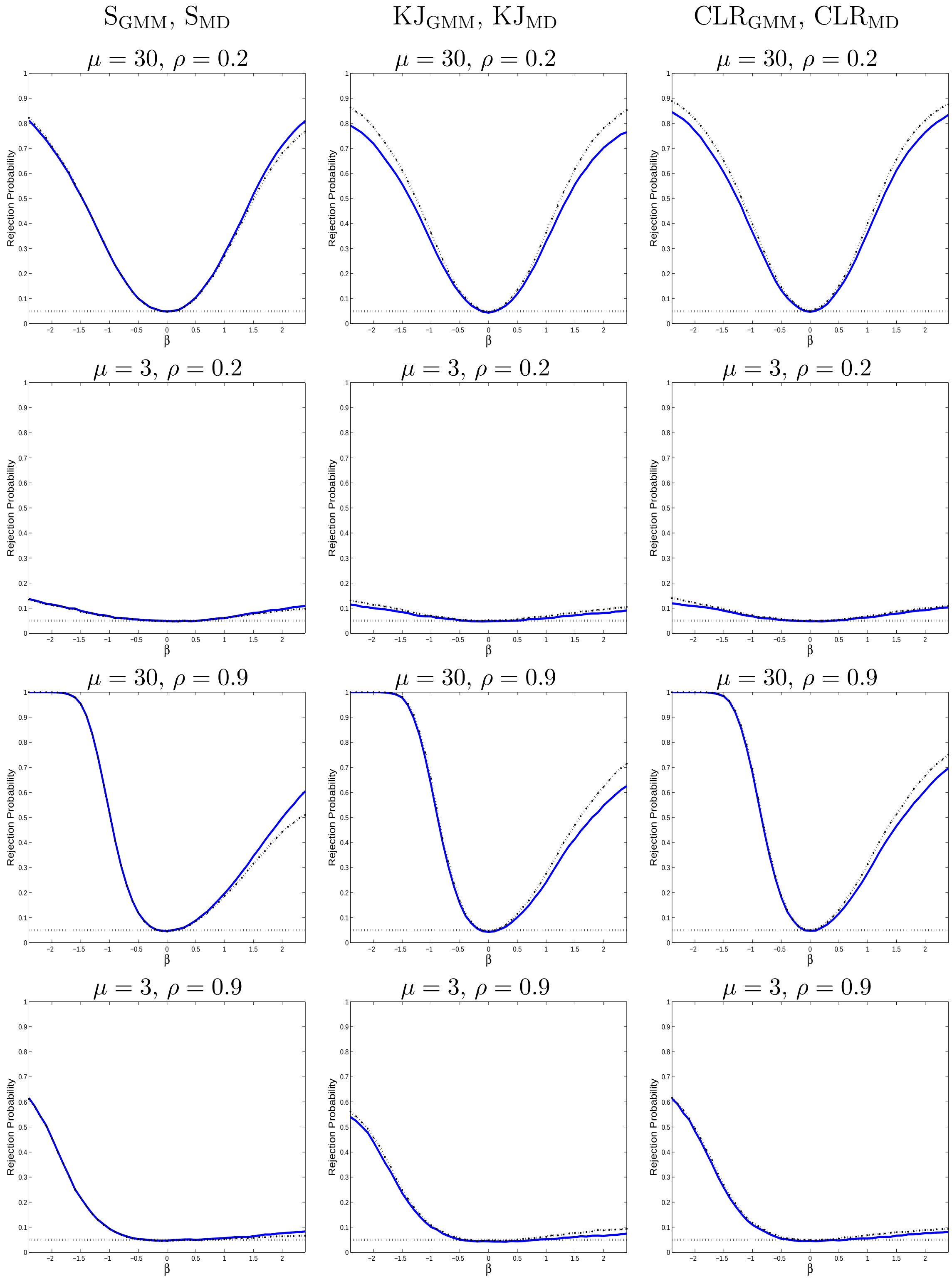

Figure 2: Power Curves of S_{GMM}, S_{MD}, KJ_{GMM}, KJ_{MD}, CLR_{GMM}, and CLR_{MD}, for testing $H_0 : \beta = 0$ at 5% significance level in the endogenous count data model. Keys: GMM-tests - solid line, MD-tests - dotted line

and its square, three child dummy variables and a race dummy variable.[7]

The dataset was originally obtained from the 1987 cross-section of the Michigan Panel Data Study of Income Dynamics and is the same as used by Lee (1995). The sample includes married couples with nonnegative total family income. The female household member must be of working age (18-64) and not self-employed. From the 3,382 married females, 895 were not working, which is, approximately, 26.4% of the total number of observations.

Besides the SCLS, we consider the winsorized mean estimator (WME) suggested by Lee (1995) to estimate the reduced form parameters. The WME is less restrictive than Powell's SCLS estimator because the latter considers a symmetric distribution of the residuals, while the former assumes only local symmetry. On the other hand, the WME demands the definition of a trimming parameter that imposes local symmetry. Our trimming parameter, denoted by w, is the point that minimizes the sum of the diagonal of the variances of the WME.

Table 3 presents the 95% confidence intervals derived from the two-step Wald estimator and our tests. The results are divided in two groups. In the first group, we follow Mroz (1987) and consider functions of the included instruments as the excluded instruments: cubic terms of the female age and education. In the second group, we add three dummy variables related to the male's occupation. In the footnote, we report the exogeneity tests proposed by Smith and Blundell (1986) and the first stage F-statistic.

For the model in column (a), the intervals derived from the Wald differ from those derived from the robust tests. In the SCLS case, the robust confidence intervals are larger than the non-robust confidence interval. In the WME case, the opposite occurs. However, when dummies are added for husband's occupations, the confidence intervals become identical, except for the S_{MD}. These results show that the Wald confidence intervals in column (a) are unreliable, even with a first-stage F-statistic above 10 and more than 3,000 observations.

6.2 Cigarette Demand Function

Mullahy (1997) suggests a Poisson-type regression to investigate the impact of smoking habits on cigarette consumption. The dataset consists of $6,160$ responses of males

[7] See Table 3 footnote.

Table 3: 95% Confidence Interval - Other Household Income

Estimation Method		Instruments[a] (a)[b]	(a)+(b)[c]
SCLS	Wald[c]	$[-0.36,\ 0.14]$	$[-0.20,\ 0.03]$
	S_{MD}	$[-0.59,\ 0.23]$	$[-0.23,\ 0.06]$
	KJ_{MD}	$[-0.79,\ 0.14]$	$[-0.20,\ 0.03]$
	CLR_{MD}	$[-0.60,\ 0.13]$	$[-0.20,\ 0.03]$
WME	Wald	$[-0.55,\ 0.20]$	$[-0.23,\ -0.01]$
	S_{MD}	$[-0.26,\ 0.09]$	$[-0.30,\ 0.05]$
	KJ_{MD}	$[-0.57,\ 0.15]$	$[-0.23,\ -0.01]$
	CLR_{MD}	$[-0.38,\ 0.12]$	$[-0.23,\ -0.01]$

Included Instruments: $age = \frac{\text{Age-40}}{10}$, where Age is female's age in years, age^2, $educ =$ (Education $-$ 8), where Education is females's education in years, $educ^2$, three child dummy variables ($C1$: child of age 0-5, $C2$: child of age 6-13, $C3$: child of age 14-17), *Race* (1 if non-white and 0 otherwise).
Excluded Instruments: three male occupation dummies ($O1$: manager or professional, $O2$: sales worker or clerical or craftsman, $O3$: farm-related worker).
Number of observations: 3,382.
[a] (a) $age \times education$, age^3, $education^3$, $age^2 \times education$, and $age \times education^2$.
(b) $O1$, $O2$ and $O3$.
[b] Exogeneity t-test: -0.56. First-stage F-statistic: 15.08.
[c] Exogeneity t-test: -3.29. First-stage F-statistic: 32.15.

to the Smoking Supplement of the 1979 National Health Interview Survey. In equation (18), y_i represents the number of cigarettes consumed, measured in packs per day. The endogenous explanatory variable x_i is the smoking habit stock measure K210.[8] The vector of instruments w_i includes: the state-level average per-pack cigarette in 1979; the individual's age in years; his years of education and its square; his family income in US$ 1,000.00; a race dummy variable (white equals one, zero otherwise), and a constant. As excluded instruments, we use: an interaction term between age and education; the state-level average price per-pack of cigarettes in 1978; and the number of years the state's restaurant smoke restrictions had been in place in 1979.

We compute the 90% and the 95% confidence regions for the smoking habit stock (β), and cigarette price (γ) using S_{GMM}, and S_{MD}, KJ_{GMM} and KJ_{MD}. They are illustrated

[8]K210 is an index of the habit-forming effects of prior cigarette consumption. The author argues that the smoking habit and past unobserved determinants of smoking are correlated. Since the latter is also correlated with the present unobserved determinants of smoking, it turns out that smoking is correlated with its unobserved characteristics.

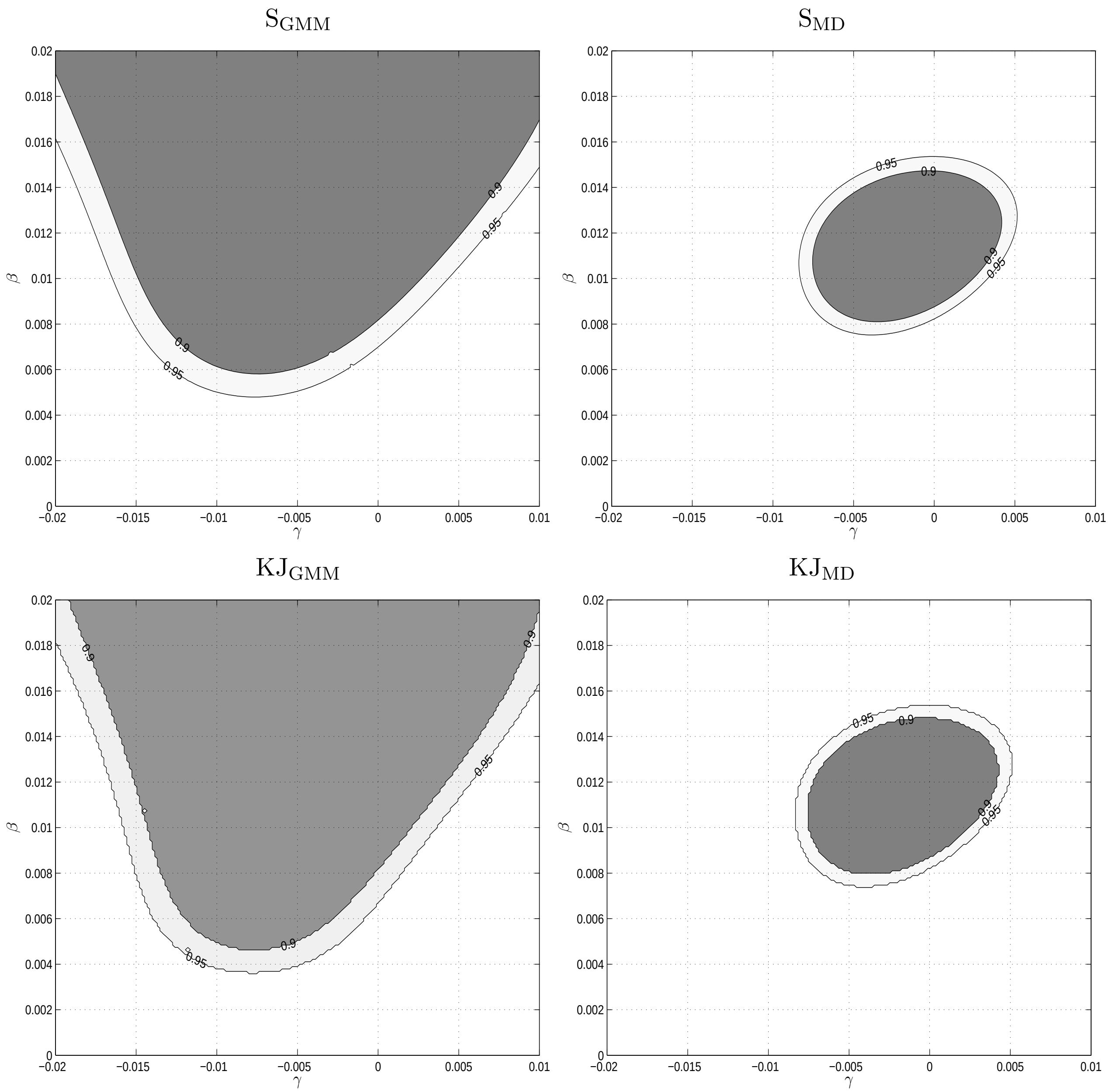

Figure 3: 90% and 95% Confidence Regions derived from S_{GMM}, S_{MD}, KJ_{GMM}, KJ_{MD} and derived from endogenous count data model. Number of observations: 6160. First-stage F-statistic: 9.78.

in Figure 3.

In this example, the GMM-tests require estimation of the eight parameters of included instruments for each hypothesized value in the grid search. This procedure involves optimizing a system of nonlinear functions. As a consequence, in parts of the parameter space, the optimization algorithm is unstable and may not converge. Additionally, the computational time for estimating the confidence region increases significantly. The S_{MD}- and $\mathrm{KJ}_{\mathrm{MD}}$-tests do not involve the estimation of the untested parameters. They require solving only one optimization problem, which is the reduced form parameters estimation.[9] Moreover, their confidence regions are smaller.

7 Conclusion

We develop tests robust to weak identification in the context of models where nonlinearities in the moment conditions make conventional GMM-procedures intractable. These tests, based on the generalized minimum distance principle, avoid nonlinearity problems because they do not require direct inference about the structural parameters. Instead, the crucial assumptions concern the relationship between the structural and reduced form parameters and the asymptotic behavior of the reduced form parameter estimator. The simplicity of this approach extends to its computational implementation, which can be conducted using regular statistical software packages. Simulations show that these tests perform well in case of weak identification and different degrees of endogeneity.

[9] In this example the grid is of 30,000 points. Using the same computer, the S_{MD}- and $\mathrm{KJ}_{\mathrm{MD}}$-tests take 8.64 seconds to calculate the confidence region, while the GMM tests last more than 26 hours.

A Proofs

A.1 Proof of Theorem 1

Under the null hypothesis, $r(\pi_0, \beta_0) = 0$ and, by continuity of the link function in π and the Cramer theorem, the asymptotic behavior of $r(\hat{\pi}, \beta_0)$ is:

$$\sqrt{n}\ r(\hat{\pi}, \beta_0) = \left[\frac{\partial r(\pi_0, \beta_0)}{\partial \pi}\right] \sqrt{n}\,(\hat{\pi} - \pi_0) + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, \Psi_{\beta_0}\right),$$

where $\Psi_{\beta_0} = \left[\frac{\partial r(\pi_0,\beta_0)}{\partial \pi}\right] \Lambda_0 \left[\frac{\partial r(\pi_0,\beta_0)}{\partial \pi}\right]'$. Since $\hat{\Lambda} \xrightarrow{p} \Lambda_0$ and $\frac{\partial r(\hat{\pi},\beta_0)}{\partial \pi} \xrightarrow{p} \frac{\partial r(\pi_0,\beta_0)}{\partial \pi}$, $\mathrm{S}_{\mathrm{MD}}(\beta_0) \xrightarrow{d} \chi^2(q)$ by the Slutzky theorem.

A.2 Proof of Theorem 2

From assumption 1 and a Taylor expansion, the asymptotic joint distribution between $r(\hat{\pi}, \beta)$ and $\frac{\partial r(\hat{\pi},\beta)}{\partial \beta}$, under the null hypothesis, is:

$$\sqrt{n}\begin{pmatrix} r(\hat{\pi}, \beta_0) - r(\pi_0, \beta_0) \\ \mathrm{vec}\left[\frac{\partial r(\hat{\pi},\beta_0)}{\partial \beta} - \frac{\partial r(\pi_0,\beta_0)}{\partial \beta}\right]\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Psi_{\beta_0} & \left[\frac{\partial r(\pi_0,\beta_0)}{\partial \pi}\right] \Lambda_0 F_0' \\ F_0 \Lambda_0 \left[\frac{\partial r(\pi_0,\beta_0)}{\partial \pi}\right]' & F_0 \Lambda_0 F_0' \end{bmatrix}\right) \quad (20)$$

where Ψ_{β_0} is defined in equation (7) and $F_0 = \frac{\partial}{\partial \pi}\left[\mathrm{vec}\left(\frac{\partial r(\pi_0,\beta_0)}{\partial \beta}\right)\right]$. The pre-multiplication of (20) by the lower-block triangular matrix $[I_q, 0_{q\times m} : \hat{F}_0, I_{qm}]$, where $\hat{F}_0 = \frac{\partial}{\partial \pi}\left[\mathrm{vec}\left(\frac{\partial r(\hat{\pi},\beta_0)}{\partial \beta}\right)\right]$, results in:

$$\sqrt{n}\begin{pmatrix} r(\hat{\pi}, \beta_0) \\ \mathrm{vec}\left[\hat{\mathrm{D}}(\beta_0) - \frac{\partial r(\pi_0,\beta_0)}{\partial \beta}\right]\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Psi_{\beta_0} & 0 \\ 0 & \Xi_{\beta_0} \end{bmatrix}\right),$$

where:

$$\Xi_{\beta_0} = F_0 \Lambda_0 F_0' \ - \ F_0 \Lambda_0 \frac{\partial r(\pi_0, \beta_0)}{\partial \pi}' \Psi_{\beta_0}^{-1} \frac{\partial r(\pi_0, \beta_0)}{\partial \pi} \Lambda_0 F_0 \quad (21)$$

Thus, $\hat{\mathrm{D}}(\beta_0)$ and $r(\hat{\pi}, \beta_0)$ are asymptotically independent, regardless the rank of C, where $C = \frac{\partial r(\pi_0,\beta_0)}{\partial \beta}$.

Let ψ_r be the limiting distribution of $\sqrt{n}\,[r(\hat{\pi}, \beta_0)]$. If C has full rank, then $\hat{\mathrm{D}}(\beta_0) \xrightarrow{p}$

C and $\sqrt{n}\ \left[\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\hat{\mathrm{D}}(\beta_0)\right]^{-\frac{1}{2}}\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}r(\hat{\pi},\beta_0) \xrightarrow{d} \left(\mathrm{C}'\Psi_{\beta_0}^{-1}\mathrm{C}\right)^{-\frac{1}{2}}\mathrm{C}'\Psi_{\beta_0}^{-1}\psi_r$. The last term is $\mathcal{N}\left(0, I_m\right)$.

If C is singular, then, as in Kleibergen (2002, 2005), $\sqrt{n}\ \mathrm{vec}\left[\hat{\mathrm{D}}(\beta_0)\right] \xrightarrow{d} \psi_D$, where ψ_D is a $qm \times 1$ multivariate normal distribution with variance Ξ_{β_0}. In this case, $\left[\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\hat{\mathrm{D}}(\beta_0)\right]^{-\frac{1}{2}}\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}r(\hat{\pi},\beta_0) \xrightarrow{d} \left(\psi_D'\ \Psi_{\beta_0}^{-1}\psi_D\right)^{-\frac{1}{2}}\psi_D'\Psi_{\beta_0}^{-1}\psi_r$. The conditional distribution $\psi_D'\Psi_{\beta_0}^{-1}\psi_r|\psi_D$ follows a multivariate normal with mean zero and variance $\psi_D'\Psi_{\beta_0}^{-1}\psi_D$. Since ψ_r and ψ_D are independent, the marginal and conditional distributions are the same. This implies that $\left(\psi_D'\ \Psi_{\beta_0}^{-1}\psi_D\right)^{-\frac{1}{2}}\psi_D'\Psi_{\beta_0}^{-1}\psi_r \equiv \mathcal{N}(0, I_m)$, and $\left[\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\hat{\mathrm{D}}(\beta_0)\right]^{-\frac{1}{2}}\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}r(\hat{\pi},\beta_0) \xrightarrow{d} \mathcal{N}\left(0, I_m\right)$, unconditionally.

A.3 Derivation of equation (10)

The first order condition of $\mathrm{S}_{\mathrm{MD}}(\beta)$ with respect to β is:

$$-\frac{1}{2}\frac{\partial \mathrm{S}_{\mathrm{MD}}(\beta)}{\partial \beta} = \ n\ r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1}\frac{\partial r(\hat{\pi},\beta)}{\partial \beta} + \frac{n}{2}\left(r(\hat{\pi},\beta)' \otimes r(\hat{\pi},\beta)'\right)\frac{\partial \mathrm{vec}\left[\hat{\Psi}_{\beta}^{-1}\right]}{\partial \beta} \tag{22}$$

The partial derivative of $\hat{\Psi}_{\beta}^{-1}$ with respect to β is:

$$\begin{aligned}
&-\left\{\hat{\Psi}_{\beta}^{-1} \otimes \hat{\Psi}_{\beta}^{-1}\right\}\frac{\partial \mathrm{vec}\left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\hat{\Lambda}\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}'\right]}{\partial \beta} \\
&-\left\{\hat{\Psi}_{\beta}^{-1} \otimes \hat{\Psi}_{\beta}^{-1}\right\}\left\{\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\hat{\Lambda} \otimes I_q\right)\frac{\partial \mathrm{vec}\left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right]}{\partial \beta} + \left(I_q \otimes \frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\hat{\Lambda}\right)\frac{\partial \mathrm{vec}\left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}'\right]}{\partial \beta}\right\}
\end{aligned}$$

The second term of equation (22) simplifies to:

$$\begin{aligned}
&\left(r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1}\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\hat{\Lambda} \otimes r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1}\right)\frac{\partial \mathrm{vec}\left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right]}{\partial \beta} \\
&\quad + \left(r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1} \otimes r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1}\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\hat{\Lambda}\right)\frac{\partial \mathrm{vec}\left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right]'}{\partial \beta}
\end{aligned} \tag{23}$$

Using the fact that

$$\frac{\partial \text{vec}\left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right]}{\partial \beta_j} = \text{vec}\left[\frac{\partial}{\partial \beta_j}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)\right] \quad \text{and} \quad \frac{\partial \text{vec}\left[\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right]'}{\partial \beta_j} = \text{vec}\left[\frac{\partial}{\partial \beta_j}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)'\right],$$

the j^{th} column of (23) is:

$$\begin{aligned} & r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1}\left[\frac{\partial}{\partial \beta_j}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)\right]\hat{\Lambda}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)'\hat{\Psi}_{\beta}^{-1}r(\hat{\pi},\beta) \\ & \quad + \; r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)\hat{\Lambda}\left[\frac{\partial}{\partial \beta_j}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)'\right]\hat{\Psi}_{\beta}^{-1}r(\hat{\pi},\beta) \end{aligned}$$

Since both terms are scalars, (23) simplifies to

$$2\; r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1}\left[\left(\frac{\partial}{\partial \beta_1}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)\right) \quad \ldots \quad \left(\frac{\partial}{\partial \beta_m}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)\right)\right]\hat{\Lambda}\left(\frac{\partial r(\hat{\pi},\beta)}{\partial \pi}\right)'\hat{\Psi}_{\beta}^{-1}r(\hat{\pi},\beta)$$

and (22) becomes:

$$-\frac{1}{2}\frac{\partial \mathrm{S}_{\mathrm{MD}}(\beta)}{\partial \beta} = \; n\; r(\hat{\pi},\beta)'\hat{\Psi}_{\beta}^{-1}\hat{\mathrm{D}}(\beta),$$

where $\hat{\mathrm{D}}(\beta) = \left[\hat{\mathrm{D}}_1(\beta) \quad \ldots \quad \hat{\mathrm{D}}_m(\beta)\right]$ and, for $j = 1, \ldots, m$,

$$\hat{\mathrm{D}}_j(\beta_0) = \frac{\partial r(\hat{\pi},\beta_0)}{\partial \beta_j} - \; \left[\frac{\partial}{\partial \pi}\left(\frac{\partial r(\hat{\pi},\beta_0)}{\partial \beta_j}\right)\right]\hat{\Lambda}\left[\frac{\partial r(\hat{\pi},\beta_0)}{\partial \pi}\right]'\hat{\Psi}_{\beta_0}^{-1}r(\hat{\pi},\beta_0).$$

B Robust Tests in Linear Instrumental Variable Models

The linear limited information model and its unrestricted reduced form is:

$$\begin{cases} \mathrm{y} = \mathrm{x}\beta + u \\ \mathrm{x} = \mathrm{z}\pi_{\mathrm{z}} + v \end{cases} \qquad \begin{cases} \mathrm{y} = \mathrm{z}\delta_{\mathrm{z}} + e \\ \mathrm{x} = \mathrm{z}\pi_{\mathrm{z}} + v \end{cases} \tag{24}$$

The included exogenous regressors w is omitted for exposition clarity. The OLS estimators of δ_{z} and π_{z} are $\hat{\delta}_{\mathrm{z}} = (\mathrm{z}'\mathrm{z})^{-1}\mathrm{z}'\mathrm{y}$ and $\hat{\pi}_{\mathrm{z}} = (\mathrm{z}'\mathrm{z})^{-1}\mathrm{z}'\mathrm{x}$, respectively. Thus, the link function $(\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0)$ can be rewritten as $(\mathrm{z}'\mathrm{z})^{-1}\mathrm{z}'(\mathrm{y} - \mathrm{x}\beta_0)$.

Note that $\frac{\partial r(\pi,\beta)}{\partial \pi} = \left[\begin{smallmatrix} 1 & -\beta_0' \end{smallmatrix}\right] \otimes I_{k_{\mathrm{z}}}$. If the residuals are homoskedastic, the asymptotic variance of $\mathrm{vec}[\delta_{\mathrm{z}}, \pi_{\mathrm{z}}]$ is $\Omega \otimes \left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)$, where Ω is $\mathrm{Var}(e, v)$. Thus, by definiton, Ψ_{β_0} is:

$$\left[\begin{pmatrix} 1 & -\beta_0' \end{pmatrix} \otimes I_{k_{\mathrm{z}}}\right] \left\{\Omega \otimes \left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1}\right\} \left[\begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} \otimes I_{k_{\mathrm{z}}}\right]$$

Let $\hat{\Omega} = \left[\begin{smallmatrix} \mathrm{y}' \\ \mathrm{x}' \end{smallmatrix}\right] \mathrm{M}_{\mathrm{z}} \left[\begin{smallmatrix} \mathrm{y} & \mathrm{x} \end{smallmatrix}\right] / (n - k_{\mathrm{z}})$ be the estimator of Ω. After substituting $\hat{\Omega}$ into Ψ_{β_0}, we find that $\hat{\Psi}_{\beta_0} = \hat{\sigma}^2_{\beta_0}\left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1}$ where $\hat{\sigma}^2_{\beta_0} = \frac{(\mathrm{y}-\mathrm{x}\beta_0)'M_{\mathrm{z}}(\mathrm{y}-\mathrm{x}\beta_0)}{n-k_{\mathrm{z}}}$. Thus, we have $\mathrm{S}_{\mathrm{MD}} = \mathrm{AR}$.

The matrix Ω can be partition as $\Omega = [\Omega_{ee}, \Omega_{ev} : \Omega_{ve}, \Omega_{vv}]$. Since $\frac{\partial r(\hat{\pi},\beta_0)}{\partial \beta} = -\hat{\pi}_{\mathrm{z}}$ and $\frac{\partial}{\partial \pi}\left(\frac{\partial r(\hat{\pi},\beta_0)}{\partial \beta}\right) = [0 \ I_{k_{\mathrm{z}}m}]$, by definition, $\mathrm{vec}[\hat{\mathrm{D}}(\beta)]$ is:

$$\begin{aligned} &= -\,\mathrm{vec}[\hat{\pi}_{\mathrm{z}}] + \left\{\begin{bmatrix} \hat{\Omega}_{ve} & \hat{\Omega}_{vv} \end{bmatrix} \otimes \left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1}\right\} \left[\begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} \otimes \mathrm{I}_{\mathrm{k_z}}\right] \hat{\Psi}^{-1}_{\beta_0}(\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0) \\ &= -\,\mathrm{vec}[\hat{\pi}_z] + \mathrm{vec}\left[\left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1} \hat{\Psi}^{-1}_{\beta_0}(\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0) \begin{bmatrix} \hat{\Omega}_{ev} \\ -\beta_0'\hat{\Omega}_{vv} \end{bmatrix}\right] \end{aligned}$$

Since $\frac{(y-x\beta_0)'M_z x}{(n-k)} = \left[\begin{smallmatrix} 1 & -\beta_0' \end{smallmatrix}\right] \left[\begin{smallmatrix} \hat{\Omega}_{ev} \\ \hat{\Omega}_{vv} \end{smallmatrix}\right]$, we have that $\hat{\mathrm{D}}(\beta) = -\tilde{\pi}_{\mathrm{z}}(\beta_0)$.

In the case of heteroskedastic residuals, the White estimator of the asymptotic covariance matrix of $\sqrt{n}\Big((\hat{\delta}_{\mathrm{z}} - \delta_{\mathrm{z}})', \mathrm{vec}(\hat{\pi}_{\mathrm{z}} - \pi_{\mathrm{z}})'\Big)'$ can be written as:

$$\hat{\Lambda} = \left[I_{m+1} \otimes \left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1} \mathrm{z}'\right] \begin{bmatrix} \mathrm{diag}(\hat{e}_i^2) & \mathrm{diag}\left[\mathrm{vec}(\hat{v}_i\hat{e}_i)'\right] \\ \mathrm{diag}\left[\mathrm{vec}(\hat{v}_i\hat{e}_i)\right] & \mathrm{diag}\left[\mathrm{vec}(\hat{v}_i\hat{v}_i')\right] \end{bmatrix} \left[I_{m+1} \otimes \mathrm{z}\left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1}\right]$$

where $\operatorname{diag}(t_i)$ represents a diagonal matrix whose typical element in the main diagonal i is t_i. Pre-multiplying by $\left[\begin{pmatrix} 1 & -\beta_0' \end{pmatrix} \otimes I_{k_z}\right]$ and pos-multiplying by $\left[\begin{pmatrix} 1 \\ -\beta_0 \end{pmatrix} \otimes I_{k_z}\right]$ results in $\hat{\Psi}_{\beta_0}$, which is:

$$= \left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1} n^{-1}\mathrm{z}'\left\{\operatorname{diag}(\hat{e}_i^2) - (\beta_0' \otimes I_n)\operatorname{diag}\left[\operatorname{vec}(\hat{v}_i\hat{e}_i)\right] - \operatorname{diag}\left[\operatorname{vec}(\hat{v}_i\hat{e}_i)'\right](\beta_0 \otimes I_n)\right.$$

$$\left. + (\beta_0' \otimes I_n)\operatorname{diag}\left(\operatorname{vec}(\hat{v}_i\hat{v}_i')\right](\beta_0 \otimes I_n)\right\} \mathrm{z}\left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1}, \quad \text{or}$$

$$= \left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1} n^{-1}\mathrm{z}'\left\{\operatorname{diag}\left[(\hat{e}_i - \hat{v}_i\beta_0)^2\right]\right\} \mathrm{z}\left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)^{-1}$$

Then, the White covariance matrix with ordinary least square estimate for the reduced form parameters give us the following result:

$$\sqrt{n}\,(\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0)'\left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)\left(n^{-1}\mathrm{z}'\left\{\operatorname{diag}\left[(\hat{e}_i - \hat{v}_i\beta_0)^2\right]\right\}\mathrm{z}\right)^{-1}\left(\frac{\mathrm{z}'\mathrm{z}}{n}\right)\sqrt{n}\,((\hat{\delta}_{\mathrm{z}} - \hat{\pi}_{\mathrm{z}}\beta_0)),$$

$$\frac{1}{\sqrt{n}}\,(\mathrm{y} - \mathrm{x}\beta_0)'\mathrm{z}\left(n^{-1}\mathrm{z}'\left\{\operatorname{diag}\left[(\hat{u}_i(\beta_0))^2\right]\right\}\mathrm{z}\right)^{-1}\frac{1}{\sqrt{n}}\,\mathrm{z}'(\mathrm{y} - \mathrm{x}\beta_0)$$

where $\hat{u}(\beta_0) = (\hat{e} - \hat{v}\beta_0) = \mathrm{M}_{\mathrm{z}}\,(\mathrm{y} - \mathrm{x}\beta_0)$. Kleibergen's (2007) S- and K-tests use $u(\beta_0) = \mathrm{y} - \mathrm{x}\beta_0$ in the last equation for estimating the variance covariance matrix. The same proof extends for the case of clustered and autocorrelated residuals.

Since

$$\frac{1}{n}\left\{\mathrm{z}'\left\{\operatorname{diag}\left[(\hat{u}_i(\beta_0))^2\right]\right\}\mathrm{z} - \mathrm{z}'\left\{\operatorname{diag}\left[(u_i(\beta_0))^2\right]\right\}\mathrm{z}\right\} = o_p(1),$$

the S- and the K-tests are asymptotically equivalent to the S_{MD}- and the K_{MD}-tests, respectively. Chernozhukov and Hansen (2008) demonstrate the asymptotic equivalent between the Wald $-$ S and Wald $-$ K and S and K.

C The algorithm for computing the robust tests using built-in functions

We derived tests for the following class of limited dependent variable models with unrestricted reduced form representation:

$$\begin{cases} \mathrm{y}^{*}= \mathrm{x}\beta + \mathrm{w}\gamma + u \\ \mathrm{x}= \mathrm{z}\pi_{\mathrm{z}} + \mathrm{w}\pi_{\mathrm{w}} + v \end{cases} \qquad \begin{cases} \mathrm{y}^{*} = \mathrm{z}\delta_{\mathrm{z}} + \mathrm{w}\delta_{\mathrm{w}} + v\delta_{v} + \varepsilon \\ \mathrm{x} = \mathrm{z}\pi_{\mathrm{z}} + \mathrm{w}\pi_{\mathrm{w}} + v \end{cases},$$

where $u = v\alpha + \varepsilon$ and $\delta_v = \alpha + \beta$. Define $\mathbf{z} = (\mathrm{z}, \mathrm{w})$, $\delta = (\delta_{\mathrm{z}}', \delta_{\mathrm{w}}', \delta_v')'$, $\pi_{\mathrm{x}} = (\pi_{\mathrm{z}}; \pi_{\mathrm{w}})$, and $h = \mathrm{z}\delta_{\mathrm{z}} + \mathrm{w}\delta_{\mathrm{w}} + v\delta_v$. We consider estimators of δ and π_{x} which have influence function on the following form $\sum_{i=1}^{n} \mathrm{g}_i(\delta, \pi_{\mathrm{x}}) = \mathrm{g}_n(\delta, \pi_{\mathrm{x}}) = \begin{pmatrix} \mathrm{g}_{1,n}(\delta, \pi_{\mathrm{x}})' & \mathrm{g}_{2,n}(\pi_{\mathrm{x}})' \end{pmatrix}'$, where:

$$\mathrm{g}_{1,n}(\delta, \pi_{\mathrm{x}}) = \begin{pmatrix} \mathbf{z}'\tilde{\varepsilon}(h) \\ v'\tilde{\varepsilon}(h) \end{pmatrix}, \quad \text{and} \quad \mathrm{g}_{2,n}(\pi_{\mathrm{x}}) = \mathrm{vec}(\mathbf{z}'v),$$

where $\tilde{\varepsilon}$ is an $n \times 1$ vector of generalized residuals, i.e, $\mathbb{E}\left[\tilde{\varepsilon}_i | v_i, \mathbf{z}_i; \delta, \pi_{\mathrm{x}}\right] = 0$ for $i = 1, ..., n$. This vector is a function of the reduced parameters through h. This representation accommodates several limited dependent variable models, including the endogenous probit, Tobit and count data models. In case of maximum or quasi-maximum likelihood, $\mathrm{g}_{1,n}(\delta, \pi_{\mathrm{x}})$ is the score function.[10]

Define $\sigma_i = \frac{\partial \tilde{\varepsilon}_i}{\partial h_i}$. The Jacobian of $\mathrm{g}_n(\delta, \pi_{\mathrm{x}})$ has the following form

$$\mathrm{H}_n = \begin{bmatrix} \mathrm{H}_{\delta\delta,n} & \mathrm{H}_{\delta\pi_{\mathrm{x}},n} \\ 0 & I_m \otimes (\mathbf{z}'\mathbf{z}) \end{bmatrix} = \begin{bmatrix} \mathbf{z}'\Sigma\ \mathbf{z} & \mathbf{z}'\Sigma\ v & -\delta_v' \otimes \mathbf{z}'\Sigma\ \mathbf{z} \\ v'\Sigma\ \mathbf{z} & v'\Sigma\ v & -\delta_v' \otimes v'\Sigma\ \mathbf{z} \\ 0 & 0 & I_m \otimes (\mathbf{z}'\mathbf{z}) \end{bmatrix},$$

where $\Sigma = \mathrm{diag}(\sigma_i)$. We use that $\frac{\partial \mathrm{g}_{1,n}(\delta, \pi_{\mathrm{x}})}{\partial \mathrm{vec}\pi_{\mathrm{x}}} = \mathbf{z}' \frac{\partial \tilde{\varepsilon}(h)}{\partial \mathrm{vec}(h)} \frac{\partial \mathrm{vec}(h)}{\partial \mathrm{vec}(\pi_{\mathrm{x}})}$ and $\frac{\partial \mathrm{vec}(h)}{\partial \mathrm{vec}(\pi_{\mathrm{x}})} = -\delta_v' \otimes \mathbf{z}$ in the derivation of the last block column. Because of the conditional independence we

[10] In the Tobit model, σ_{ε}^2, the variance of ε, is a nuisance parameter in the reduced form model. In this case, $\mathrm{g}_n^{(1)}(\delta, \pi_{\mathrm{x}})$ is the 'effective' score, which is obtained as the residual of regressing $\mathrm{g}_n^{(1)}(\delta, \pi_{\mathrm{x}})$ on the score of σ_{ε}^2.

have

$$\frac{1}{\sqrt{n}}\left[\begin{array}{c} \mathrm{g}_{1,n}(\delta,\pi_{\mathrm{x}}) \\ \mathrm{g}_{2,n}(\pi_{\mathrm{x}}) \end{array}\right] \to N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right],\left[\begin{array}{cc} \mathrm{G} & 0 \\ 0 & \Upsilon \end{array}\right]\right),$$

where $\mathrm{G} = \mathbb{E}[\mathrm{g}_i(\delta,\pi_{\mathrm{x}})\mathrm{g}_i(\delta,\bar{\pi})'|\mathbf{z}_i, v_i]$ and $\Upsilon = \lim_{n\to+\infty}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(I_m\otimes\mathbf{z}_i')\,v_i'v_i\,(I_m\otimes\mathbf{z}_i)\right]$. The asymptotic distribution of the reduced form parameter estimator is:

$$\sqrt{n}\left(\begin{array}{c} \hat{\delta}-\delta \\ \mathrm{vec}(\hat{\pi}_{\mathrm{x}}-\pi_{\mathrm{x}}) \end{array}\right) \overset{a}{=} -\left(\frac{1}{n}\left[\begin{array}{cc} \mathrm{H}_{\delta\delta,n} & \mathrm{H}_{\delta\pi_{\mathrm{x}},n} \\ 0 & I_m\otimes(\mathbf{z}'\mathbf{z}) \end{array}\right]\right)^{-1}\frac{1}{\sqrt{n}}\left(\begin{array}{c} \mathrm{g}_{1,n}(\delta,\pi_{\mathrm{x}}) \\ \mathrm{g}_{2,n}(\pi_{\mathrm{x}}) \end{array}\right).$$

After some simplification, we find that the asymptotic variance is:

$$\left[\begin{array}{cc} \mathrm{H}_{\delta\delta}^{-1}\mathrm{G}\mathrm{H}_{\delta\delta}^{-1}+(\delta_v'\otimes I_k)\,\Lambda_{\pi_{\mathrm{x}}\pi_{\mathrm{x}}}\,(\delta_v\otimes I_k) & (\delta_v'\otimes I_k)\,\Lambda_{\pi_{\mathrm{x}}\pi_{\mathrm{x}}} \\ \Lambda_{\pi_{\mathrm{x}}\pi_{\mathrm{x}}}\,(\delta_v\otimes I_k) & \Lambda_{\pi_{\mathrm{x}}\pi_{\mathrm{x}}} \end{array}\right]$$

where $n^{-1}\mathrm{H}_{\delta\delta,n} \xrightarrow{p} \mathrm{H}_{\delta\delta}$, $\Lambda_{\pi_{\mathrm{x}}\pi_{\mathrm{x}}} = (\mathrm{I}\otimes\mathrm{Q})^{-1}\Upsilon(\mathrm{I}\otimes\mathrm{Q})^{-1}$ and $n^{-1}\mathbf{z}'\mathbf{z} \xrightarrow{p} \mathrm{Q}$. The term $\mathrm{H}_{\delta\delta}^{-1}\mathrm{G}\mathrm{H}_{\delta\delta}^{-1}$ is the variance-covariance of the quasi-maximum likelihood estimator. Pre multiplying by $\left[\left(1\ -\beta_0'\right)\otimes\left(I_{k_z}\ \ 0\right)\right]$ and post multiplying by $\left[\left(\begin{smallmatrix} 1 \\ -\beta_0 \end{smallmatrix}\right)\otimes\left(\begin{smallmatrix} I_{k_z} \\ 0 \end{smallmatrix}\right)\right]$ results in Ψ_{β_0}, which is

$$\left(\mathrm{H}_{\delta\delta}^{-1}\mathrm{G}\mathrm{H}_{\delta\delta}^{-1}\right)_{\delta_{\mathrm{z}}\delta_{\mathrm{z}}}+((\delta_v-\beta_0)'\otimes\mathrm{I}_{\mathrm{k_z}})\,\Lambda_{\pi_{\mathrm{z}}\pi_{\mathrm{z}}}\,((\delta_v-\beta_0)\otimes\mathrm{I}_{\mathrm{k_z}}) \tag{25}$$

where $\left(\mathrm{H}_{\delta\delta}^{-1}\mathrm{G}\mathrm{H}_{\delta\delta}^{-1}\right)_{\delta_{\mathrm{z}}\delta_{\mathrm{z}}}$ is the $\mathrm{k_z}\times\mathrm{k_z}$ variance of $\hat{\delta}_{\mathrm{z}}$. In the simulations for the endogenous count data model, the computation of $\hat{\Psi}_{\beta}$ is based on equation (25). If the likelihood is correctly specified, $\mathrm{H}_{\delta\delta}^{-1}\mathrm{G}\mathrm{H}_{\delta\delta}^{-1} = \mathrm{H}_{\delta\delta}^{-1}$ by the matrix information equality. If residuals are homoskedastic, the covariance matrix estimator derived from ordinary least squares can be simplified to $\Lambda_{\pi_{\mathrm{x}}\pi_{\mathrm{x}}} = (\Sigma_{vv}\otimes\mathrm{Q}^{-1})$. The matrix Q can be partitioned as $\mathrm{Q} = [\mathrm{Q_{zz}},\mathrm{Q_{zw}} : \mathrm{Q_{wz}},\mathrm{Q_{ww}}]$. Then, equation (25) simplifies to

$$\left(\mathrm{H}_{\delta\delta}^{-1}\right)_{\delta_{\mathrm{z}}\delta_{\mathrm{z}}}+(\delta_{\mathrm{v}}-\beta_0)'\Sigma_{\mathrm{vv}}(\delta_{\mathrm{v}}-\beta_0)\ \mathrm{Q}_{\mathrm{zz.w}}^{-1}$$

where $\mathrm{Q_{zz.w}} = \mathrm{Q_{zz}}-\mathrm{Q_{zw}}\mathrm{Q_{ww}^{-1}}\mathrm{Q_{wz}}$. This is the variance equation used for computing the robust tests for endogenous probit and Tobit models in Finlay and Magnusson (2009).

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.

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

Arellano, M., O. Bover, and J. M. Labeaga (1999). Autoregressive models with sample selectivity for panel data. In C. Hsiao, K. Lahiri, L.-F. Lee, and M. H. Pesaran (Eds.), *Analysis of Panels and Limited Dependent Variable Models*. Cambridge University Press.

Blundell, R., T. MaCurdy, and C. Meghir (2007). Labor supply models: Unoserved heterogeneity, nonparticipation and dynamics. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6A, Chapter 69, pp. 4668–4775. Elsevier.

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.

Blundell, R. W. and R. J. Smith (1994). Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables. *Journal of Econometrics 64*, 355–373.

Chernozhukov, V. and C. Hansen (2008). The Reduced Form: A Simple Approach to Inference with Weak Instruments. *Economic Letters 100*(1), 68–71.

Finlay, K. and L. M. Magnusson (2009). Implementing Tests for a General Class of Instrumental Variable Models that Are Robust to Weak Instruments. *The Stata Journal* (forthcoming).

Gourieroux, C. and A. Monfort (1995). *Statistics and Econometrics Models*, Volume 1. Cambridge University Press.

Jones, A. M. and J. M. Labeaga (2003). Individual Heterogeneity and Censoring in Panel Data Estimates of Tobacco Expenditure. *Journal of Applied Econometrics 18*(2), 157–177.

Kleibergen, F. (2002). Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica 70*(5), 1781–1803.

Kleibergen, F. (2005). Testing Parameters in GMM without Assuming that they are Identified. *Econometrica 73*(4), 1103–1123.

Kleibergen, F. (2007). Generalizing Weak Instrument Robust IV Statistics Towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics. *Journal of Econometrics 139*(1), 181–216.

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

Lee, L.-F. (1992). Amemiya's Generalized Least Squares and Tests of the Overidentification in Simultaneous Equation Models with Qualitative or Limited Dependent Variables. *Econometric Reviews 11*(3), 319–328.

Lee, M.-J. (1995). Semi-parametric Estimation of Simultaneous Equations with Limited Dependent Variables: a Case Study of Female Labor Supply. *Journal of Applied Econometrics 10*(2), 187–200.

Lemieux, T., B. Fortrin, and P. Fréchette (1994). The Effect of Taxes on Labor Supply in the Underground Economy. *American Economic Review 84*(1), 231–254.

Madalla, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Econometrica Society Monographs. Cambridge University Press.

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 Empircal Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica 55*(4), 765–799.

Mullahy, J. (1997). Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior. *Review of Economics and Statistics 79*(4), 586–593.

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. (1985). Semiparametric Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables. *Annales de L'Insee 59-60*, 219–236.

Powell, J. L. (1986). Symmetrically Trimmed Least Squares Estimation for Tobit Models. *Econometrica 54*(6), 1435–1460.

Silva, J. M. C. S. (2001). Influence Diagnostics and Estimation Algorithms for Powell's SCLS. *Journal of Business & Economic Statistics 19*(1), 55–62.

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*(5), 1055–1096.