Inference in Limited Dependent Variable Models Robust to Weak Identification *

Leandro M. Magnusson†
Tulane University

September 29, 2008

Abstract

We propose tests for structural parameters in limited dependent variable models with endogenous explanatory variables using the classical minimum distance framework. These tests have the correct size whether the structural parameters are identified or not. Relating to the current tests, the application of ours is appropriate especially to models whose moment conditions are nonlinear in parameters. Moreover, the computation of ours tests is simple, allowing their implementation in a large number of statistical software packages. We compare our tests with Wald tests by performing simulation experiments. We use our tests to analyze the female labor supply and the demand for cigarette.

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 and Sun-Jae Jun for their helpful suggestions.

†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

1 Introduction

In this paper, we use the classical minimum distance approach to derive tests for structural parameters in limited dependent variable models with endogenous explanatory variables. These tests have the correct size even when these parameters are not identified. The minimum distance approach is specially convenient when the moment conditions are nonlinear in the parameters.

Lack of parameter identification invalidate Wald, Lagrange multiplier (LM) and likelihood-ratio (LR) tests, as shown by Staiger and Stock (1997) and Stock and Wright (2000). In the case of linear instrumental variable models, several tests are robust to parameter identification failure. We can mention the AR, in Anderson and Rubin (1949), the K, in Kleibergen (2002), the conditional likelihood-ratio, in Moreira (2003), the rank-type tests, in Andrews and Soares (2007), among others. For nonlinear models, the current tests are based on the GMM approach. They depart from the objective function of the continuous updating estimator. 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 continuous updating estimator. In the same paper he derives the CLR-test which is a function of a rank statistic and is not pivotal.[1]

The minimum distance approach is based on the definition of link functions, which relate structural and auxiliary parameters. By avoiding the direct use of moment conditions, the minimum distance permits the construction of robust tests for a class of models where the use of GMM tests would entail the solution of constrained nonlinear systems.

From the applied point of view, the tests that we propose are compelling mainly because of their ease of computation. In many models they can be carried out by using built-in functions within regular statistical software packages. Moreover, confidence intervals based on our tests do not require the estimation of untested parameters under the null hypothesis at every hypothesized value of the parameter of interest.

The convenience of the minimum distance approach, however, goes beyond its ease of computation. The asymptotic properties of our tests are derived from the asymptotic properties of the auxiliary parameters and do not depend on the structural parameters. Necessary conditions for the implementation of our tests are continuity and smoothness of the link function, $\sqrt{n}$-asymptotically consistency of the auxiliary parameters estimator

[1]A statistic is called pivotal if its sampling distribution does not depend on unknown parameters.

and consistency of their variance-covariance matrix estimator. Under these assumptions, auxiliary parameters can be estimated either parametrically or semiparametrically.

In the next section, we illustrate the application of our tests in the endogenous Tobit model. Since our tests can be applied to other models, we present their general version in Section 3. In a simulation experiment, we compare the performance of our tests to Wald tests. We perform simulations not only for the endogenous Tobit model, but also for the endogenous count data model. In Section 5 we consider the female labor supply described by Blundell and Smith (1989) and Lee (1995), and the demand for cigarettes described by Mullahy (1997) to illustrate the differences between inferences based on nonrobust tests and ours. We finish the paper with the conclusion. The Appendix contains all proofs and an algorithm describing the implementation of our tests for a class of limited dependent variable models.

2 Inference in the Endogenous Tobit Model

The classical minimum distance principle explores the underlying relation between structural and auxiliary parameters: the estimation of structural parameters is indirectly obtained from the auxiliary parameters estimates. This principle is widely used in Econometrics for estimation. Examples of models where the use of the minimum distance is appropriate are: limited dependent variable models with endogenous explanatory variables (see Newey (1987), Blundell and Smith (1989), Lee (1995), and Blundell et al. (2007)), linear panel models with unobserved heterogeneity (see Chamberlain (1984) and Jones and Labeaga (2003)), and autoregressive panel data models with sample selectivity (see Bover and Arellano (1997), and Arellano et al. (1999)). In macroeconomics, Sbordone (2005) and Li (2008) use the minimum distance estimation in dynamic forward looking models.

In this section we illustrate the application of our tests using the simple endogenous Tobit model discussed in Amemiya (1979) and Smith and Blundell (1986). Let h_i be the hours of work provided by individual i, x_i her unearned income and w_i other relevant variables such as individual characteristics (age, education, etc.). The labor supply is represented by:

$$h_i = \max\{0, x_i\beta + w_i\gamma + u_i\} \qquad \text{for} \quad i = 1, \ldots, n \tag{1}$$

where β is a scalar and γ is an $k_w \times 1$ vector. As Blundell et al. (2007) argue, x is an endogenous variable because non-observed preference for work may be correlated, for example, with asset income. Assume that there exists a vector of instruments z, with dimension $1 \times k_z$, such that:

$$x_i = z_i \Pi_z + w_i \Pi_w + v_i, \tag{2}$$

and z_i is uncorrelated with u_i. Our goal is to test the null hypothesis $H_0 : \beta = \beta_0$ without imposing any conditions about the identification of β, i.e, without assuming that Π_z is full-ranked. According to Stock and Wright (2000), if Π_z is not full-ranked, then the Lagrange multiplier, the likelihood-ratio and the Wald tests are not well approximated by their respective asymptotic distributions.

We propose tests that explore the representation of equation (1) in its reduced form:

$$h_i = \max\{0, z_i \pi_z + w_i \pi_w + e_i\} \tag{3}$$

If the reduced form is equivalent to the structural model, then $\pi_z = \Pi_z \beta$. Since z_i and w_i are exogenous in equations (2) and (3), under mild conditions the auxiliary parameters π_z and Π_z are consistently estimated regardless if β is identified or not. We derive tests for β using the restriction:

$$r(\pi_z, \Pi_z, \beta) = \pi_z - \Pi_z \beta$$

Our tests are based on the limiting distribution of $r(\hat{\pi}_z, \hat{\Pi}_z, \beta)$, where $\hat{\pi}_z$ and $\hat{\Pi}_z$ are, respectively, estimators of π_z and Π_z. The joint asymptotic distribution of $\hat{\pi}_z$ and $\hat{\Pi}_z$ is independent of β. Since the asymptotic properties of $r(\hat{\pi}_z, \hat{\Pi}_z, \beta)$ depends on the asymptotic properties of $\hat{\pi}_z$ and $\hat{\Pi}_z$, our tests are sized-correct independent of the rank of Π_z. Their formal derivation is in the next section.

Other identification-robust tests which can be applied to the endogenous Tobit model are proposed by Kleibergen (2005). These tests are based on the GMM framework. We argue that the implementation of our tests is more convenient in many situations, particularly in the endogenous Tobit model.

Kleibergen's tests require differentiability of the moment condition with respect to its parameters. A candidate for a smooth moment condition is the score of the likelihood function, which imposes restrictions on the joint distribution of the residuals. Assuming that u_i and v_i follow a bivariate normal distribution with zero mean and variance-covariance matrix $\Sigma = \begin{bmatrix}\sigma_u^2 & \Sigma_{\text{uv}} : \Sigma_{\text{vu}} & \Sigma_{\text{vv}}\end{bmatrix}$, the conditional likelihood function is well defined and twice differentiable. Define the pseudo-residuals:

$$e_i^{(1)} = d_i\left(\frac{y_i - \bar{w}_i\delta}{\sigma_e}\right) - (1-d_i)\frac{\phi_i}{1-\Phi_i}$$
$$e_i^{(2)} = d_i\left[\left(\frac{y_i - \bar{w}_i\delta}{\sigma_e}\right)^2 - 1\right] - (1-d_i)\left(\frac{\bar{w}_i\delta}{\sigma_e}\right)\frac{\phi_i}{1-\Phi_i}$$

where:

$$\begin{aligned}
d_i \quad &= \begin{cases} 1 & \text{if} \quad y_i > 0 \\ 0 & \text{otherwise} \end{cases} \\
v_i \quad &= x_i - z_i\Pi_z - w_i\Pi_w \\
\bar{w}_i\delta \quad &= x_i\beta + w_i\gamma + v_i\alpha \\
\alpha \quad &= \Sigma_{\text{vv}}^{-1}\Sigma_{\text{vu}} \\
\sigma_\varepsilon \quad &= \sigma_u^2 - \Sigma_{\text{uv}}\Sigma_{\text{vv}}^{-1}\Sigma_{\text{vu}}
\end{aligned}$$

and ϕ_i and Φ_i are, respectively, the normal density and distribution functions evaluated at $\frac{\bar{w}_i\delta}{\sigma_e}$. The score function of β, denoted by s_β, is:

$$s_\beta = \sum_i \frac{x_i e_i^{(1)}}{\sigma_\varepsilon} \tag{4}$$

Note from expression (4) that β is non-separable from γ, α, Π_z, Π_w, π_w and σ_e. Although these parameters are not being tested, they must be estimated under the null hypothesis and substituted in (4). Estimation of the untested parameters consists in solving $2(k_w+1)+k_z$ non-linear equations, derived from the remaining score fucntions:

$$\sum_i \begin{bmatrix} w_i & v_i \end{bmatrix}' \frac{e_i^{(1)}}{\sigma_e} = 0$$

$$\sum_i \frac{e_i^{(2)}}{2\sigma_\varepsilon^2} = 0$$

$$\sum_i \left(\frac{v_i' v_i}{n}\right)^{-1} \sum_i \begin{bmatrix} z_i & w_i \end{bmatrix}' v_i - \alpha \sum_i \frac{\begin{bmatrix} z_i & w_i \end{bmatrix}' e_i^{(1)}}{\sigma_\varepsilon} = 0$$

A second step is the estimation of the covariance between the Hessian and the score function. Our procedure represents a great simplification for testing the parameter β, since the estimation of auxiliary parameters does not require the solution of a constrained nonlinear equations system, neither the derivation of the asymptotic distribution of the untested parameters.

The construction of confidence intervals based on our tests is also simpler. The $1-\tau$ confidence set is formed by $\bar{\beta}$'s such that the null hypothesis $H_0 : \beta = \bar{\beta}$ is not rejected at the τ significance level. Using the GMM approach to build confidence intervals involves the re-estimation of untested parameters at each hypothesized value of $\bar{\beta}$. We do not estimate untested parameters, but only auxiliary parameters, which do not change with $\bar{\beta}$. Therefore, no re-estimation is necessary during our grid search.

Section 4 presents the simulations results for this model, comparing the performance of our tests and the Wald tests obtained from the maximum likelihood and two-step MD estimators. Since our tests are not restricted only to the endogenous Tobit model, we also perform simulations for the endogenous count data model discussed in Mullahy (1997).

3 Minimum Distance Robust Tests

In this section we present the general version of our robust tests. Consider θ as a $k_\theta \times 1$ vector representing the auxiliary parameters, and β as a $m \times 1$ vector of structural parameters. The estimator of θ is denoted by $\hat{\theta}$. The true values of θ and β under the data generating process are, respectively, θ_0 and β_0. The mapping $r : \Theta \times \mathbb{B} \rightarrow \Re^q$, with $m \leq q \leq k_\theta$, denoted as $r(\theta, \beta)$, represents the restrictions imposed on the auxiliary parameters. The minimum distance approach relies on the following regularity conditions:

Assumption 1. *(Regularity conditions)*

i. (Limiting distribution of the auxiliary parameters)

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Lambda_0) \tag{5}$$

where Λ_0 *is a symmetric positive definite covariance matrix.*

ii. (Existence of a consistent estimator for the variance-covariance matrix)

There exists $\hat{\Lambda}$ *such that*

$$\hat{\Lambda} \xrightarrow{p} \Lambda_0 \tag{6}$$

iii. (Continuity and differentiability of the link function)

$r(\theta, \beta)$ *is a twice continuous differentiable function on* $\mathbb{R}^{k_\theta} \times \mathbb{R}^m$*. Under the null hypothesis* $H_0 : \beta = \beta_0$*,* $r(\theta_0, \beta_0) = 0$*. Moreover,* $\text{rank}\left[\frac{\partial r(\theta,\beta)}{\partial \theta}\right] = q$*.*

Assumptions 1.i and 1.ii. assure that the auxiliary parameters estimator is root-N consistent and asymptotically normal. These assumptions hold independently of the structural parameter identification. Assumption 1.iii of continuity and differentiability is standard in minimum distance estimation and inference.

These assumptions deserve further comments:

1) We are not imposing the full rank condition on $\frac{\partial r(\theta,\beta)}{\partial \beta}$, a necessary assumption for estimating β by minimum distance.

2) In the GMM specification, smoothness of the empirical moment is necessary for the construction of robust tests. The minimum distance approach does not require such condition. Instead, it relies on the differentiability of the link function (Assumption 1.iii). As a consequence, auxiliary parameters can be estimated semiparametrically. For example, Lee (1995) uses a symmetry condition to estimate the auxiliary parameters of an endogenous Tobit model. Because symmetry implies a nondifferentiable moment restriction, the application of the GMM robust is not possible under this assumption.

By applying the delta-method, we derive the asymptotic distribution of $r(\hat{\theta}, \beta)$ which, under the null hypothesis $H_0 : \beta = \beta_0$, is:

$$\sqrt{n}\left(r(\hat{\theta}, \beta_0) - r(\theta_0, \beta_0)\right) \xrightarrow{d} \mathcal{N}\left(0, \Psi_{\beta_0}\right) \tag{7}$$

where:

$$\Psi_{\beta_0} = \left[\frac{\partial r(\theta_0, \beta_0)}{\partial \theta}\right] \Lambda_0 \left[\frac{\partial r(\theta_0, \beta_0)}{\partial \theta}\right]'$$

The asymptotic behavior of $r(\hat{\theta}, \beta_0)$ holds independently of the identification of β. Define $\mathrm{S_{MD}}(\beta)$ as the objective function of the optimal minimum distance estimator like in Gourieroux and Monfort (1989):

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

where:

$$\hat{\Psi}_\beta = \left[\frac{\partial r(\hat{\theta}, \beta)}{\partial \theta}\right] \hat{\Lambda} \left[\frac{\partial r(\hat{\theta}, \beta)}{\partial \theta}\right]'$$

From equation (7), $\mathrm{S_{MD}}(\beta)$ follows a chi-square distribution with q degrees of freedom under the null hypothesis:

$$\mathrm{S_{MD}}(\beta_0) \xrightarrow{d} \chi^2(q) \tag{9}$$

Remark 1. *The* $\mathrm{S_{MD}}$*-test is similar to the* S*-test proposed by Stock and Wright (2000) derived under the GMM framework. However it is important to emphasize the differences:*

1) The link function $r(\hat{\theta}, \beta)$ *represents the overidentification restrictions on the auxiliary parameters. It is not a sample average of empirical moments.*

2) Once $\hat{\theta}$ *and* $\hat{\Lambda}$ *are consistently estimated, the continuity of* $r(\theta, \beta)$ *guarantees that* $\hat{\Psi}_\beta$ *is consistent.*

3) The limiting distribution of the $\mathrm{S_{MD}}$*-test is solely derived from the asymptotic properties of the auxiliary parameter estimator* $\hat{\theta}$*.*

4) Usually the estimation of untested parameters can be avoided. In Section 2, for *example, we do not have to estimate γ, α, Π_z, Π_w, π_w nor σ_e to test β.*

The $\mathrm{S_{MD}}$-test tests simultaneously two hypotheses, which are the value of the structural parameter and the overidentification restrictions. As in Kleibergen (2007) we can decompose the $\mathrm{S_{MD}}$-test into two orthogonal statistics, namely, $\mathrm{K_{MD}}$ and $\mathrm{J_{MD}}$. The former statistic tests only the value of the structural parameter, while the latter tests only the overidentification restriction.

Theorem 1. *($\mathrm{K_{MD}}$- and $\mathrm{J_{MD}}$-tests)*

Define the $\mathrm{K_{MD}}$- and $\mathrm{J_{MD}}$-tests as:

$$\mathrm{K_{MD}}(\beta_0) = n \left[\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\ r(\hat{\theta},\beta_0)\right]' \ \hat{\mathrm{P}}_{\beta_0} \ \left[\hat{\Psi}_{\beta_0}^{-\frac{1}{2}} r(\hat{\theta},\beta_0)\right] \tag{10}$$

$$\mathrm{J_{MD}}(\beta_0) = n \left[\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\ r(\hat{\theta},\beta_0)\right]' \ \hat{\mathrm{M}}_{\beta_0} \ \left[\hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\ r(\hat{\theta},\beta_0)\right] \tag{11}$$

where:

$$\begin{aligned}
\hat{\mathrm{P}}_{\beta_0} &= \hat{\Psi}_{\beta_0}^{-\frac{1}{2}}\ \hat{\mathrm{D}}_{\beta_0} \left[\hat{\mathrm{D}}_{\beta_0}'\hat{\Psi}_{\beta_0}^{-1}\hat{\mathrm{D}}_{\beta_0}\right]^{-1} \hat{\mathrm{D}}_{\beta_0}'\hat{\Psi}_{\beta_0}^{-\frac{1}{2}\prime} \\
\hat{\mathrm{M}}_{\beta_0} &= I_q - \hat{\mathrm{P}}_{\beta_0} \\
\hat{\mathrm{D}}_{\beta_0} &= \left[\hat{\mathrm{D}}_1(\beta_0) \ \ \ldots \ \ \hat{\mathrm{D}}_m(\beta_0)\right] \\
\hat{\mathrm{D}}_j(\beta_0) &= \frac{\partial r(\hat{\theta},\beta_0)}{\partial \beta_j} - \left[\frac{\partial}{\partial\theta}\left(\frac{\partial r(\hat{\theta},\beta_0)}{\partial\beta_j}\right)\right]\hat{\Lambda}\left[\frac{\partial r(\hat{\theta},\beta_0)}{\partial\theta}\right]'\hat{\Psi}_{\beta_0}^{-1} r(\hat{\theta},\beta_0), \quad j = 1,\ldots,m
\end{aligned}$$

Under assumption 1 and $\mathrm{H}_0 : \beta = \beta_0$, we have:

$$\begin{aligned}
\mathrm{K_{MD}}(\beta_0) &\xrightarrow{d} \chi^2(m) \\
\mathrm{J_{MD}}(\beta_0) &\xrightarrow{d} \chi^2(q-m)
\end{aligned}$$

regardless if β is point identified or not. Also,

$$\mathrm{S_{MD}}(\beta_0) = \mathrm{K_{MD}}(\beta_0) + \mathrm{J_{MD}}(\beta_0) \tag{12}$$

Proof. See Appendix. □

The statistic $\hat{\mathrm{D}}_{\beta_0}$ is asymptotically independent of $r(\hat{\theta},\beta_0)$ under the null hypothesis. Given assumption C in Stock and Wright (2000), if β is identified, then $\hat{\mathrm{D}}_{\beta_0}$ converges in probability to $\frac{\partial r(\theta_0,\beta_0)}{\partial \beta}$. If not, 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 $r(\hat{\theta}, \beta_0)$, the distribution of the $\mathrm{K_{MD}}$-test, conditional on $\hat{\mathrm{D}}_{\beta_0}$, is free from nuisance parameters (see Moreira (2003)). Moreover, its unconditional distribution is pivotal.

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

$$-\frac{1}{2}\frac{\partial \mathrm{S_{MD}}(\beta)}{\partial \beta} = \ n\ r(\hat{\theta}, \beta)'\hat{\Psi}_{\beta}^{-1}\hat{\mathrm{D}}_{\beta} \tag{13}$$

The $\mathrm{K_{MD}}$-test is the quadratic form of equation (13), weighted by its own variance. The minimum value of $\mathrm{S_{MD}}(\beta)$ coincides with the point where the $\mathrm{K_{MD}}$-test equals zero. This point is the minimum distance continuous updating estimate (MD-CUE). The minimum value of the S-test is the GMM-CUE, which is often different from the MD-CUE.[2]

From (13), we have that the $\mathrm{K_{MD}}$-test suffers from a spurious decline of power at inflection and local minimum points of the $\mathrm{S_{MD}}$-test. Close to these points, the value of the $\mathrm{J_{MD}}$-test approximates to the value of the $\mathrm{S_{MD}}$-test, which has discriminatory power. Consequently, we may define a new test for the structural parameter, the $\mathrm{KJ_{MD}}$-test, by combining both the $\mathrm{K_{MD}}$- and the $\mathrm{J_{MD}}$-tests. Define $\tau_{\mathrm{K_{MD}}}$ and $\tau_{\mathrm{J_{MD}}}$ as the levels of significance of the $\mathrm{K_{MD}}$- and $\mathrm{J_{MD}}$-tests, respectively. The $\mathrm{KJ_{MD}}$-test has approximate significance level of $\tau = \tau_{\mathrm{K_{MD}}} + \tau_{\mathrm{J_{MD}}}$. Rejection occurs either if $\mathrm{K_{MD}}$ rejects at $\tau_{\mathrm{K_{MD}}}$ or if $\mathrm{J_{MD}}$ rejects at $\tau_{\mathrm{J_{MD}}}$.

The $\mathrm{S_{MD}}$-, $\mathrm{K_{MD}}$- and $\mathrm{J_{MD}}$-tests can be adapted in order to test only a subset of the structural parameter vector. The procedure consists in estimating the untested parameters under the null hypothesis by the minimum distance CUE estimator. The tests are calculated by replacing the estimated values into the original tests. However, if the untested parameters are not identified, the limiting distributions of the tests are not pivotal. It is possible to pretest the rank of the gradient of the link function with respect to the untested parameters. The size of the robust statistics is affected by this pre-test.

Recently, Monte Carlo studies conducted by Startz et al. (2004), in the context

[2]In the linear instrumental variable model under homoscedastic errors, the CUE-GMM is the same as the limited information maximum likelihood (LIML). Goldberger and Olkin (1971) show that LIML has a minimum distance interpretation. In this case, the minimum value of the S and $\mathrm{S_{MD}}$ are the same.

of linear instrumental variable model, and by Guggenberger and Smith (2005), in the context of generalized empirical likelihood (GEL), indicate that the size properties of the K- and the $\mathrm{LM}_{\mathrm{GEL}}$-tests for testing a subvector of structural parameters are not much affected by the identification of the untested parameters.

In many models, the implementation of our tests is feasible just using built-in functions of statistical software packages. Examples are some limited dependent variables models as the endogenous Tobit and the endogenous Probit. In the Appendix we describe an algorithm of our tests for the endogenous Tobit model.

4 Simulations

In order to analyze the performance of our tests, we carry out simulations not only for the Tobit but also for the endogenous count data model as in Mullahy (1997). In this type of model, the dependent variable, denoted as y, has a count outcome, $y \in \{0, 1, 2, \ldots\}$. Assume that there is one endogenous variable, x, and k_w exogenous variables, w. The expected conditional mean of y is:

$$\mathbb{E}\left[y|x, w, \eta; \beta, \gamma\right] = \exp\left(x\beta + w\gamma\right)\eta \tag{14}$$

where β and γ are the structural parameters of the model. The unobserved variable η is correlated with x and satisfies $\mathbb{E}\left[\eta\right] = 1$. The unitary mean of η is assumed without loss of generality since there is a constant term in w. Let z be a row vector of dimension k_z such that $\mathbb{E}\left[\eta|z, w\right]$ is a constant, and $\mathbb{E}\left[y|x, w, \eta, z\right] = \mathbb{E}\left[y|x, w, \eta\right]$. Assuming that:

$$x_i = z_i\Pi_z + w_i\Pi_w + v_i,$$

with $\mathbb{E}\left[v_i|z_i, w_i\right] = 0$, equation (14) can be written as:

$$\mathbb{E}\left[y|x, w, \eta; \beta, \gamma\right] = \exp\left(z_i\pi_z + w_i\pi_w\right)\xi_i \tag{15}$$

where $\xi_i = \exp(v_i)\eta_i$.

The general formulae of our tests, when applied to the endogenous Tobit and count data models, are:

$$\mathrm{S_{MD}}(\beta_0) = \left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)' \hat{\Psi}_{\beta_0}^{-1} \left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)$$
$$\mathrm{K_{MD}}(\beta_0) = \left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)' \hat{\Psi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0} \left(\hat{\hat{\Pi}}_{\beta_0}' \hat{\Psi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0}\right)^{-1} \hat{\hat{\Pi}}_{\beta_0}' \hat{\Psi}_{\beta_0}^{-1} \left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)$$

where:

$$\hat{\Psi}_{\beta_0} = \begin{bmatrix} I_{k_z} & -\beta_0 I_{k_z}\end{bmatrix} \hat{\Lambda} \begin{bmatrix} I_{k_z} & -\beta_0 I_{k_z}\end{bmatrix}'$$
$$\hat{\hat{\Pi}}_{\beta_0} = \hat{\Pi}_z - \begin{bmatrix} 0 & I_{k_z}\end{bmatrix} \hat{\Lambda} \begin{bmatrix} I_{k_z} & -\beta_0 I_{k_z}\end{bmatrix}' \hat{\Psi}_{\beta_0}^{-1} \left(\hat{\pi}_z - \hat{\Pi}_z\beta_0\right)$$

and I_{k_z} is the identity matrix of dimension k_z.

We estimate $\hat{\Pi}_z$ by ordinary least squares. For the endogenous Tobit model, we estimate π_z by the conditional generalized least squares (CGLS) proposed by Newey (1987), and by the symmetric censored least squares (SCLS) proposed by Powell (1986). For the endogenous count data model we estimate π_z by Poisson quasi-likelihood.

Besides $\mathrm{S_{MD}}$, $\mathrm{K_{MD}}$, $\mathrm{J_{MD}}$ and $\mathrm{KJ_{MD}}$, we also report the conditional likelihood-ratio (CLR) test derived by Moreira (2003). Assuming only one endogenous variable, the minimum distance version of the CLR, under the null hypothesis, is:

$$\mathrm{CLR_{MD}}(\beta_0) = \frac{1}{2}\left\{\mathrm{S_{MD}}(\beta_0) - \mathrm{rk}(\beta_0) + \sqrt{\left[\mathrm{S_{MD}}(\beta_0) + \mathrm{rk}(\beta_o)\right]^2 - 4\mathrm{J_{MD}}(\beta_o)\mathrm{rk}(\beta_0)}\right\} \quad (16)$$

where:

$$\mathrm{rk}(\beta_0) = \; n\left\{\hat{\hat{\Pi}}_{\beta_0}' \hat{\Xi}_{\beta_0}^{-1}\hat{\hat{\Pi}}_{\beta_0}\right\},$$
$$\hat{\Xi}_{\beta_0} = \hat{\Lambda}_{\Pi_z\Pi_z} - \left(\hat{\Lambda}_{\Pi_z\pi_z} - \beta_0\hat{\Lambda}_{\Pi_z\Pi_z}\right)\hat{\Psi}_{\beta_0}^{-1}\left(\hat{\Lambda}_{\pi_z\Pi_z} - \beta_0\hat{\Lambda}_{\Pi_z\Pi_z}\right),$$

$\hat{\Lambda}_{\Pi_z\Pi_z}$ is the variance estimate of $\hat{\Pi}_z$, and $\hat{\Lambda}_{\Pi_z\pi_z}$ is the covariance estimate between $\hat{\Pi}_z$ and $\hat{\pi}_z$. The asymptotic distribution of the $\mathrm{CLR_{MD}}$ is not pivotal and depends on $\mathrm{rk}(\beta)$. The critical values of this test are calculated by simulating independent values of $\chi^2(1)$ and $\chi^2(k_z - 1)$ for a given value of $\mathrm{rk}(\beta)$.

We compare the performance of our tests with Wald tests which, according to our notation, are:

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

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

We simulate observations for both models and compute the rejection frequency of the tests at significance levels of 10%, 5% and 1%. The $\mathrm{KJ_{MD}}$-test uses a significance level of $\mathrm{K_{MD}}$ four times the significance level of $\mathrm{J_{MD}}$. The residuals u_i and v_i are the same for each simulation.

4.1 Endogenous Tobit

Consider the endogenous Tobit model discussed in Section 2:

$$\begin{cases} h_i = \max\{0, x_i\beta + w_i\gamma + u_i\} \\ x_i = z_i\Pi_z + w_i\Pi_w + v_i \end{cases}$$

We generate 10,000 random samples of 250 observations each, satisfying:

· w_i is a unitary constant;

· z_i is a 1×3 row vector drawn from independent standard normal distributions. It is the same in all simulations;

· u_i and v_i have correlation ρ_{uv}, $\rho_{\mathrm{uv}} \in \{0.1, 0.9\}$. The closer ρ_{uv} is to one, the more endogenous is the model;

· $\beta = 0.5$, $\gamma = 0.5$;

· Π_w assumes different values in order to guarantee that the expected number of censored observations is approximately 30% in each simulation;

· $\Pi_z = \begin{bmatrix} \Pi_{z_1} & 0 & 0 \end{bmatrix}'$ is a 3×1 column vector. The value of Π_{z_1} is set according to

$$\mu_z = \frac{\Pi_z' Z' Z \Pi_z}{k_z \sigma_v^2}$$

where $Z = \begin{bmatrix} z_1' & \dots & z_n' \end{bmatrix}'$, σ_v^2 is the variance of v_i, and μ_z is the concentration parameter divided by k_z. We choose μ_z equal to 10 and 1 to represent cases of strong and weak identification, respectively.[3]

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

The residuals u_i and v_i have symmetric joint distributions. We consider two scenarios. In the first one, the residuals follow a bivariate normal distribution with unitary variance and correlation ρ_{uv}, $\rho_{\text{uv}} \in \{0.1, 0.9\}$. The second scenario is different according to each estimator. In the case of the CGLS, we assume that $u_i = v_i\alpha + \varepsilon_i$, where v_i follows the t-distribution with 4 degrees of freedom and ε_i follows the standard normal distribution. The values of α are fixed such that ρ_{uv} is either 0.1 or 0.9. In the simulations with SCLS, we consider a bivariate t-distribution with 4 degrees of freedom and correlation ρ_{uv}.

We use three estimators for β: maximum likelihood, CGLS, and the two-stage SCLS (see Lee (1995)). In the SCLS case, we use the algorithm proposed by Silva (2001). This algorithm is a Newton-type method which checks the behavior of the objective function at each step of iteration. Silva (2001) shows that his algorithm converges faster and more frequently compared to the original algorithm in Powell (1986). However, Silva's algorithm does not eliminate convergence problems. Between 3% and 5% of the simulations did not converge. Our null hypothesis is $H_0 : \beta = 0.5$.

[Table 1 about here.]

[Table 2 about here.]

The Wald tests become size distorted when identification decreases. The distortion varies according to the degree of endogeneity. With $\mu_z = 1$, the Wald tests estimated by MLE or CGLS underreject the null hypothesis when $\rho = 0.1$ and overreject it when $\rho = 0.9$. The Wald-CGLS with $\rho = 0.9$ and normal distribution of the residuals is the most severe case of overrejection. The Wald-SCLS always underrejects. These results are related to the bias of the estimators for β: the lower the degree of identification and the higher the endogeneity, the more upward biased are the estimates.

Differently from the Wald tests, the robust identification minimum distance tests' performances are not affected by the level of identification nor by the degree of endogeneity.

4.2 Count Data Model

We simulate 10,000 samples with 250 observations each, according to the following rules:

$$\begin{cases} y_i \sim Poisson(\lambda_i) \\ \lambda_i = \exp\left(x_i\beta + w_i\gamma + u_i\right) \\ x_i = z_i\Pi_z + w_i\Pi_w + v_i \end{cases} \tag{17}$$

The variables and the parameters satisfy:

· w_i is a unitary constant and $\Pi_w = 0$;

· $\beta = 0.05$;

· z_i is a 1×3 row vector. Each element is drawn from the independent standard uniform distribution. They are the same in all simulations;

· $\Pi_z = \begin{bmatrix} \Pi_{z_1} & 0 & 0 \end{bmatrix}'$ is a 3×1 column vector. The value of Π_{z_1} is set according to

$$\mu_z = \frac{\Pi_z' Z' Z \Pi_z}{k_z \sigma_v^2}$$

where $Z = \begin{bmatrix} z_1' & \dots & z_n' \end{bmatrix}'$, σ_v^2 is the variance of v_i, and μ_z is the concentration parameter divided by k_z. We choose μ_z equal to 40 and 1 to represent strong and weak identification, respectively.

We want to evaluate the performances of the tests when there is overdispersion or not. In order to simulate a model without overdispersion, we drew u_i and v_i from a bivariate uniform distribution on the interval $[-0.5, 0.5]$ and set $\gamma = 0$. So, the mean of y_i was limited in the range $[1, 1.07]$ while its variance lied in $[1.13, 1.15]$. For the model with overdispersion, u_i and v_i were drawn from a standard bivariate normal distribution. In this case, $\gamma = -0.27$, so that the mean of y_i was in the range $[1.3, 1.4]$ and its variance was between 4 and 5.

To evaluate how differently the tests perform according to the degree of endogeneity, we consider two values for ρ_{uv}: 0.1 and 0.9.

We test the null hypothesis $H_0 : \beta = \beta_0$, where $\beta_0 = 0.05$. Results for $\mu_z = 40$ are in Table 3 and for $\mu_z = 1$, in Table 4. The nonrobust tests are listed according to the method used for estimating β: GMM or two-step MDE. Both estimators follow Mullahy (1997). The GMM estimator is based on the following empirical moment conditions:

$$f_n(\beta,\gamma) = \sum_i^n \begin{bmatrix} z_i' \\ w_i' \end{bmatrix} \left[\exp(-x_i\beta - w_i\gamma)y_i - 1\right], \tag{18}$$

while the minimum-distance estimator for the structural parameters is based on the following restrictions:

$$\begin{aligned} \pi_z - \Pi_z\beta &= 0 \\ \pi_w - \Pi_w\beta + \gamma &= 0 \end{aligned} \tag{19}$$

[Table 3 about here.]

[Table 4 about here.]

Changes both in the level of endogeneity and in the level of dispersion affect the behavior of the Wald tests. These changes are magnified when identification becomes weaker. In the presence of weak instruments the bias of both point estimators increases, affecting the size of the tests. In the overdispersion case with $\mu_z = 1$, for example, the rejection probability of the Wald-MDE test jumps from 1.31% when $\rho_{uv} = 0.1$ to 23.83% when $\rho_{uv} = 0.9$ while it is supposed to be 10%. Our tests' rejection probabilities are close to the expected asymptotic critical values, regardless of the level of endogeneity, degree of dispersion or strength of identification.

5 Two Applications

In this section we use the robust tests to construct confidence intervals and regions for the models of Section 4. The $1-\tau$ confidence interval or region consists in considering points of the parameter space which do not reject the null hypothesis $H_0 : \beta = \beta_0$ at significance level τ. For each example we compare our confidence intervals with those constructed by using the Wald tests.

5.1 Female Labor Supply

The married female labor supply model as in Blundell and Smith (1989) is represented by:

$$\begin{cases} h_i = \max\{0, x_i\beta + w_i\gamma + u_i\} \\ y_i = z_i\Pi_z + w_i\Pi_w + v_i \end{cases}$$

where h_i is weakly hours in paid work, 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 and its square, child dummy variables and a race dummy variable (1 if non-white, 0 otherwise). More details are in the Appendix.

The data set, originally obtained from the 1987 cross-section of the Michigan Panel Data Study of Income Dynamics, is the same as used by Lee (1995). The sample includes married couples with nonnegative total family income. The wife must be at 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 estimation methods discussed in the previous section, we consider the winsorized mean estimator (WME) suggested by Lee (1992) for estimating the auxiliary parameters. The WME is less restrictive than the Powell's SCLS estimator because the latter considers 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. As recommended by Lee (1995), our trimming parameter, denoted by w, is the point that minimizes the sum of the diagonal of the variances of the WME.

Table 5 displays the results, divided into two groups. In the first group we follow Mroz (1987) and consider functions of the included instruments as the excluded instruments: cubic terms of wife age and education. In the second group, we add 3 dummy variables related to the husband's occupation.[4] In the footnote, we report the exogeneity tests proposed by Smith and Blundell (1986) and the first-stage F-statistic.

[Table 5 about here.]

[4]see Table 8 in the Appendix.

The estimated values of β are negative, as expected, and increase after the inclusion of husband's occupation dummies. The difference between the first-stage F-statistics suggests that this increase is due to identification issues in the first model.

On Table 6, we present the 95% confidence intervals derived from the Wald and our tests.

[Table 6 about here.]

The Wald and our tests, except the S_{MD}-test,[5] have the same limiting distribution if β is identified. In the model with less instruments, the intervals derived from the Wald and our tests are different. In the SCLS case, the robust confidence intervals are larger than the nonrobust confidence interval. In the WME case, we observe the opposite. However, when other instruments are added, confidence intervals become identical. These results show that estimates in the first model are unreliable, even with the first-stage F-statistic above 10 and more than 3,000 observations. They suggest that, in the endogenous Tobit, the identification condition for estimating parameters might be even stronger than in the linear instrumental variables model.

The confidence intervals of the S_{MD}-CGLS and the $\mathrm{KJ}_{\mathrm{MD}}$-CGLS tests are empty, suggesting that the models that assume normality distribution of the residuals are misspecified. We cannot reject specifications that rely on the overall and/or local symmetry according to the same tests.

5.2 Cigarette Demand Function

Mullahy (1997) suggests a Poisson-type regression to investigate the impact of smoking habit on the consumption of cigarette. The data set consists of $6,160$ answers of males to the Smoking Supplement of the 1979 National Health Interview Survey. To describe the model, we repeat equation (14):

$$\mathbb{E}\left(y_i | x_i, w_i, \beta, \eta_i; \alpha, \gamma\right) = \exp\left(x_i \beta + w_i \gamma\right) \eta_i$$

where y is the number of smoked cigarettes measured in packs per day. The endogenous explanatory variable x_i is the smoking habit stock measure K210.[6] The vector of

[5] The confidence set derived from the S_{MD}-test is larger because its limiting χ^2-distribution has more degrees of freedom than the remaining tests.

[6] K210 is an index of the habit-forming effects of prior cigarette consumption.

included instruments w_i contains: the state-level average per-pack cigarette in 1979; the individual's age in years and its square; his years of education and its square; his family income in U$ 1,000.00; a race dummy variable (white equals one, zero otherwise), and a constant. 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 habit and unobserved smoking characteristics are also correlated. The instruments used by Mullahy (1997) are: cubic terms of age and education; an interaction term between age and education; stage-level average price per-pack cigarette in 1978; and number of years the state's restaurant smoke restrictions had been placed in 1979.

The estimates of smoking habit stock and cigarette price are displayed on Table 7.

[Table 7 about here.]

The results in columns (3) and (4) are very close to those reported by Mullahy (1997). The sign of the estimates are positive for habit stock and negative for cigarette price and restaurant restriction, as expected. Columns (1)'s and (2)'s first-stage F-statistic is very low, suggesting that stage-level average price in 1978 and restaurant restrictions are weak instruments for the smoking habit index. So, the identification of the structural parameters comes from the remaining instruments, which are functions of the included exogenous variables.

Figure 1 shows the one minus p-value function derived from the robust and non-robust tests for the models (3) and (4) on Table 7. The intersection between the one minus p-value function and the 0.95 horizontal line delimits the 95% confidence interval for the smoke stock variable.

[Figure 1 about here.]

From equation (13) we know that the CUE-MDE minimizes the S_{MD}-test and, at this estimate, the K_{MD}-test is zero. The GMM estimate is the minimum value of the GMM one minus p-value function, and the analogous is valid for the MD estimate. We observe that the GMM and the CUE-MD estimates are close one to another, but they are different from the two-step MDE. According to Mullahy, the different values of GMM and MD estimates are due to the misspecification in the first-stage, which assumes a linear relation between endogenous variable and instruments. However, the J_{MD} overidentification test does not reject the model with a linear first-stage. We conclude that the difference between the two-step MD and the GMM estimates is due to the weak identification problem and not to a model missspecification. This is clearly captured by the confidence intervals: the confidence interval generated by Wald-GMM test is just a subset contained in the robust confidence intervals.

In the next figure we illustrate the impact of weak instruments on the inference of exogenous variables. We compute 95% confidence regions for the smoking habit stock and cigarette price parameters using Wald-MDE and Wald-GMM as the nonrobust tests, and S_{MD} and $\mathrm{KJ}_{\mathrm{MD}}$ as the robust ones.

[Figure 2 about here.]

Differently from the confidence regions of the Wald-tests, the robust tests' regions are not elliptic. The Wald-MDE gives a smaller confidence region relatively to the robust confidence regions. The Wald-GMM confidence region does not cover the same parameter space as the robust tests' confidence regions. The panels suggest that the presence of weak instruments affects the inference of the other exogenous variables in the model.

A policy question is whether changes in cigarettes prices and adoption of restaurant restriction affect the demand for cigarettes. We jointly test if cigarette price and restaurant restriction are statistically significant, considering habit stock in the range between 0 and 0.012. The Wald-GMM test did not reject the hypothesis that the coefficients are insignificant at the 5% significance level. In the same range, the robust tests S_{MD} and K_{MD} reject the same hypothesis at the 1% significance level.

6 Conclusion

We extend weak identification robust tests to models in which nonlinearities in the moment conditions turn the current GMM tests untractable. Our tests, which are based on the classical minimum distance principle, avoid non-linearity problems because they do not require direct inference about the structural parameters. The crucial assumptions, instead, are about the relation between structural and auxiliary parameters and the asymptotic behavior of the auxiliary parameters estimator. The simplicity of this approach extends to its computational implementation, which can be made by regular statistical software packages. Simulations show that our tests perform well in case of weak identification and different degrees of endogeneity.

A Proofs

A.1 Proof of Theorem 1

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

$$\sqrt{n}\begin{pmatrix} r(\hat{\theta},\beta_0)-r(\theta_0,\beta_0) \\ \text{vec}\left[\frac{\partial r(\hat{\theta},\beta_0)}{\partial\beta}-\frac{\partial r(\theta_0,\beta_0)}{\partial\beta}\right]\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix} \Psi_{\beta_0} & \left[\frac{\partial r(\theta_0,\beta_0)}{\partial\theta}\right]\Lambda_0 F_0' \\ F_0\Lambda_0\left[\frac{\partial r(\theta_0,\beta_0)}{\partial\theta}\right]' & F_0\Lambda_0 F_0'\end{bmatrix}\right) \tag{20}$$

where:

$$\begin{aligned}\Psi_{\beta_0} &= \left[\frac{\partial r(\theta_0,\beta_0)}{\partial\theta}\right]\Lambda_0\left[\frac{\partial r(\theta_0,\beta_0)}{\partial\theta}\right]' \\ F_0 &= \frac{\partial}{\partial\theta}\left[\text{vec}\left(\frac{\partial r(\theta_0,\beta_0)}{\partial\beta}\right)\right]\end{aligned}$$

Define the following lower-block triangular matrix:

$$\begin{bmatrix} I_q & 0_{q\times mq} \\ -\hat{F}_0\hat{\Lambda}\left[\frac{\partial r(\hat{\theta},\beta_0)}{\partial\theta}\right]'\hat{\Psi}_{\beta_0}^{-1} & I_q\otimes I_m\end{bmatrix} \tag{21}$$

where $\hat{F}_0=\frac{\partial}{\partial\theta}\left[\text{vec}\left(\frac{\partial r(\hat{\theta},\beta_0)}{\partial\beta}\right)\right]$. The pre-multiplication of (20) by (21) give us:

$$\sqrt{n}\begin{pmatrix} r(\hat{\theta},\beta_0) \\ \text{vec}\left[\hat{\mathrm{D}}(\beta_0)-\frac{\partial r(\theta_0,\beta_0)}{\partial\beta}\right]\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\Psi_{\beta_0} & 0 \\ 0 & \Xi_{\beta_0}\end{bmatrix}\right)$$

where:

$$\Xi_{\beta_0} = F_0\Lambda_0F_0' \;-\; F_0\Lambda_0\frac{\partial r(\theta_0,\beta_0)}{\partial\theta}'\Psi_{\beta_0}^{-1}\frac{\partial r(\theta_0,\beta_0)}{\partial\theta}\Lambda_0F_0$$

The statistic $\hat{\mathrm{D}}(\beta_0)$ is asymptotically independent of $r(\hat{\theta},\beta_0)$ under the null hypothesis, independently of the rank$\left[\frac{\partial r(\theta_0,\beta_0)}{\partial\beta}\right]$. In the case of full rank condition, we have:

$$\hat{\mathrm{D}}(\beta_0) \xrightarrow{p} \mathrm{C}$$

$$\sqrt{n}\ \left[\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\hat{\mathrm{D}}(\beta_0)\right]^{-\frac{1}{2}}\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}r(\hat{\theta},\beta_0) \xrightarrow{d} \left(\mathrm{C}'\Psi_{\beta_0}^{-1}\mathrm{C}\right)^{-\frac{1}{2}}\mathrm{C}'\Psi_{\beta_0}^{-1}\psi_r \equiv \mathcal{N}\left(0, I_m\right)$$

where $\mathrm{C} = \frac{\partial r(\theta_0,\beta_0)}{\partial\beta}$ and ψ_r is a multivariate standard normal distribution with dimension q. As in Staiger and Stock (1997), if $\frac{\partial r(\theta_0,\beta_0)}{\partial\beta}$ is not full ranked, then, we assume:

$$\sqrt{n}\ \mathrm{vec}\left[\hat{\mathrm{D}}(\beta_0)\right] \xrightarrow{d} \psi_D$$

where ψ_D is a normal distribution with variance Ξ_{β_0}. Therefore:

$$n\ \hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}r(\hat{\theta},\beta_0) \xrightarrow{d} \psi_D'\Psi_{\beta_0}^{-1}\psi_r$$

Conditioning on ψ_D, the limit density function of the above expression is

$$\psi_D'\Psi_{\beta_0}^{-1}\psi_r\big|\,\psi_D \equiv \mathcal{N}\left(0, \psi_D'\Psi_{\beta_0}^{-1}\psi_D\right)$$

Since the ψ_D and ψ_r are independent, we have

$$\left(\psi_D'\ \Psi_{\beta_0}^{-1}\psi_D\right)^{-\frac{1}{2}}\psi_D'\Psi_{\beta_0}^{-1}\psi_r \equiv \mathcal{N}(0, I_m)$$

and

$$\left[\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}\hat{\mathrm{D}}(\beta_0)\right]^{-\frac{1}{2}}\hat{\mathrm{D}}(\beta_0)'\hat{\Psi}_{\beta_0}^{-1}r(\hat{\theta},\beta_0) \xrightarrow{d} \mathcal{N}\left(0, I_m\right)$$

unconditionally.

A.2 Derivation of equation (13)

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

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

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

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

The second term of equation (22) simplifies to:

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

Using the fact that

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

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

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

Since both terms are scalars, (23) simplifies to

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

and (22) becomes:

$$-\frac{1}{2}\frac{\partial \mathrm{S}_{\mathrm{MD}}(\beta)}{\partial \beta} = \; n\; r(\hat{\theta},\beta)'\hat{\Psi}_{\beta}^{-1}\hat{\mathrm{D}}(\beta) \tag{24}$$

where:

$$\begin{aligned}
\hat{\mathrm{D}}(\beta) &= \begin{bmatrix} \hat{\mathrm{D}}_1(\beta) & \dots & \hat{\mathrm{D}}_m(\beta) \end{bmatrix} \\
\hat{\mathrm{D}}_j(\beta_0) &= \frac{\partial r(\hat{\theta},\beta_0)}{\partial \beta_j} - \; \left[\frac{\partial}{\partial \theta}\left(\frac{\partial r(\hat{\theta},\beta_0)}{\partial \beta_j}\right)\right]\hat{\Lambda}\left[\frac{\partial r(\hat{\theta},\beta_0)}{\partial \theta}\right]'\hat{\Psi}_{\beta_0}^{-1} r(\hat{\theta},\beta_0), \quad j = 1,\dots,m
\end{aligned}$$

B Implementation of Robust Tests for a Class of Limited Dependent Variable Models

The appendix shows an generic algorithm for implementing our tests. This algorithm is specific for the following class of models:

$$\begin{cases} y^* = x\beta + w\gamma + u \\ x = z\Pi_z + w\Pi_w + v \end{cases}$$

where y^* is latent and x is continuously observed. We assume that u and v is jointly normally distributed with mean zero and variance-covariance $\Sigma = \begin{bmatrix} \sigma_{uu} & \Sigma_{uv} & \vdots & \Sigma_{vu} & \Sigma_{vv} \end{bmatrix}$. Rather then observing y^*, we observe:

$$y = g(y^*, \nu)$$

This representation is compatible with several limited dependent variable models. To illustrate the implementation of our robust tests we consider the endogenous Tobit model as example. The algorithm can be extended to other limited dependent variable models straight forwardly. The endogenous Tobit model, as presented by Smith and Blundell (1986), is:

$$\begin{cases} y_i = \max\{0, x_i\beta + w_i\gamma + u_i\} \\ x_i = z_i\Pi_z + w_i\Pi_w + v_i \end{cases}$$

The above model has the following reduced-form representation:

$$\begin{cases} y_i = \max\{0, z_i\pi_z + w_i\pi_w + v_i\pi_v + e_i\} \\ x_i = z_i\Pi_z + w_i\Pi_w + v_i \end{cases}$$

where $e_i = u_i - v_i\kappa$, and $\kappa = \Sigma_{vv}^{-1}\Sigma_{vu}$. Define the matrix $\bar{Q}$ as

$$\bar{Q} = \begin{bmatrix} Q_{zz} & Q_{zw} & \vdots & Q_{wz} & Q_{ww} \end{bmatrix} \tag{25}$$

the quadratic matrix, where $Q_{zz} = \sum_i z_i' z_i$, $Q_{zw} = Q_{wz}' = \sum_i z_i' w_i$ and $Q_{ww} = \sum_i w_i' w_i$. Define also $Q_{zz.w} = Q_{zz} - Q_{zw}Q_{ww}^{-1}Q_{wz}$. We have that $Q_{zz.w}^{-1}$ the upper right block of

$\bar{Q}^{-1}$.

The algorithm test takes the following steps:

1. Estimate Π_z and Σ_{vv} by OLS. Denote the estimated values as $\hat{\Pi}_z$ and $\hat{\Sigma}_{vv}$. Compute also $\hat{v}_i$, the OLS estimated residuals.

2. Estimate π_z and π_w using the following Tobit equation:

$$y_i = \max\{0, z_i\pi_z + w_i\pi_w + \hat{v}_i\pi_v + \tilde{e}_i\}$$

   Denote the estimated values of π_z and π_w as, respectively, $\hat{\pi}_z$ and $\hat{\pi}_w$.

3. Save $\hat{\Gamma}_{\pi_z\pi_z}$, the output of the variance-covariance matrix estimate of $\hat{\pi}_z$. Note that this is not the "correct" variance-covariance of $\hat{\pi}_z$ since we are not adjusting for the presence of $\hat{v}_i$.

Now we have all the elements to compute the minimum distance robust tests. Using the same notation as in the body of the text we have:

$$\begin{aligned} r(\hat{\pi}_z, \hat{\Pi}_z, \beta) &= \hat{\pi}_z - \hat{\Pi}_z\beta \\ \hat{\Psi}_\beta &= \hat{\Gamma}_{\pi_z\pi_z} + (\hat{\pi}_v - \beta)'\hat{\Sigma}_{vv}(\hat{\pi}_v - \beta)Q_{zz.w}^{-1} \\ \hat{\hat{\Pi}}_\beta &= \hat{\Pi}_z - Q_{zz.w}^{-1}\hat{\Psi}_\beta^{-1}(\hat{\pi}_v - \beta)(\hat{\pi}_v - \beta)'\hat{\Sigma}_{vv} \end{aligned}$$

In this example, the robust tests has the following configuration:

$$\begin{aligned} \mathrm{S}_{\mathrm{MD}}(\beta) &= \left(\hat{\pi}_z - \hat{\Pi}_z\beta\right)'\hat{\Psi}_\beta^{-1}\left(\hat{\pi}_z - \hat{\Pi}_z\beta\right) \\ \mathrm{K}_{\mathrm{MD}}(\beta) &= \left(\hat{\pi}_z - \hat{\Pi}_z\beta\right)'\hat{\Psi}_\beta^{-1}\hat{\hat{\Pi}}_\beta\left(\hat{\hat{\Pi}}_\beta'\hat{\Psi}_\beta^{-1}\hat{\hat{\Pi}}_\beta\right)^{-1}\hat{\hat{\Pi}}_\beta'\hat{\Psi}_\beta^{-1}\left(\hat{\pi}_z - \hat{\Pi}_z\beta\right) \\ \mathrm{J}_{\mathrm{MD}}(\beta) &= \mathrm{S}_{\mathrm{MD}}(\beta) - \mathrm{K}_{\mathrm{MD}}(\beta) \end{aligned}$$

C Data Appendix - Married Female Labor Supply

The data set was extracted from 1987 wave of Michigan Panel Study of Income Dynamics PSID. We rescale the variables in order to match the definition used by Blundell and Smith (1989).

[Table 8 about here.]

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. and G. Soares (2007). Rank tests for instrumental variables regression with weak instruments. *Econometric Theory 23*(6), 1033–1082.

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.

Bover, O. and M. Arellano (1997). Estimating dynamic limited dependent variable models from panel data. *Investigaciones Economicas 21*(2), 141–165.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, Volume II, Chapter 22, pp. 1248–1318. Elsevier.

Goldberger, A. and I. Olkin (1971). A minimum-distance interpretation of limited-information estimation. *Econometrica 39*(3), 635–639.

Gourieroux, C. and A. Monfort (1989). A general framework for testing a null hypothesis in a "mixed form". *Econometric Theory 5*(1), 63–82.

Guggenberger, P. and R. J. Smith (2005). Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory 21*(4), 667–709.

Jones, A. M. and J. M. Labeaga (2003). Individual heterogeneity and censoring in panel data estimates of tobacco expenditure. *Journal of Applied Ecometrics 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.

Lee, M.-J. (1992). Winsorized mean estimator for censored regression. *Econometric Theory 8*, 368–382.

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.

Li, H. (2008). Estimation and testing of euler equation models with time-varying reduced form coefficients. *Journal of Econometrics* (forthcoming).

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. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics 36*(3), 231–250.

Powell, J. L. (1986). Symmetrically trimmed least squares estimation for tobit models. *Econometrica 54*(6), 1435–1460.

Sbordone, A. M. (2005). Do expected future marginal costs drive inflation dynamics. *Journal of Monetary Economics 52*, 1183–1197.

Silva, J. M. C. S. (2001). Influence diagnostics and estimation algorithms for powell's scls. *Journal of Business and 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.

Startz, R., E. Zivot, and C. R. Nelson (2004). Improved inference in weakly identified instrumental variable regression. Working Paper, University of Washington.

Stock, J. H. and J. Wright (2000). Gmm with weak identification. *Econometrica 68*(5), 1055–1096.

Tables

Table 1: Size Comparison (in percentage) $H_0 : \beta = 0.5$, Endogenous Tobit Model, $\mu_z = 10$[a]

	$\rho = 0.1$						$\rho = 0.9$					
	normal			t			normal			t		
ACV[b]	10	5	1	10	5	1	10	5	1	10	5	1
Wald-MLE	9.66	4.26	0.55	·	·	·	10.74	6.75	2.64	·	·	·
Wald-CGLS	8.59	3.58	0.42	8.87	3.96	0.46	14.50	10.07	4.57	14.48	10.06	4.17
Wald-SCLS	10.24	5.13	1.08	10.79	5.83	1.36	8.52	4.68	1.26	7.73	4.01	0.96
S_{MD}-CGLS	10.24	5.42	1.06	10.36	5.15	1.05	10.40	5.20	0.98	10.27	5.24	1.09
K_{MD}-CGLS	10.21	5.05	0.97	10.18	5.31	1.10	10.21	5.12	0.98	9.92	4.75	0.81
J_{MD}-CGLS	10.31	5.27	1.16	10.29	5.31	1.16	10.02	5.32	1.06	10.93	5.46	1.08
KJ_{MD}-CGLS	10.35	4.98	1.04	10.21	5.32	1.13	10.07	5.25	0.92	9.71	4.78	0.89
CLR_{MD}-CGLS	10.38	5.40	0.89	10.08	5.66	1.02	10.18	5.41	0.95	9.97	5.04	0.83
S_{MD}-SCLS	12.36	7.50	2.74	12.05	7.07	2.43	10.23	6.06	2.16	8.41	4.72	1.85
K_{MD}-SCLS	11.27	6.10	1.89	10.90	6.15	1.81	9.22	5.16	1.50	8.34	4.50	1.21
J_{MD}-SCLS	11.65	6.78	2.02	11.19	6.63	2.05	9.11	5.36	1.61	8.11	4.72	1.42
KJ_{MD}-SCLS	11.88	7.02	2.23	11.55	6.88	2.23	9.92	5.93	1.90	8.47	4.94	1.54
CLR_{MD}-SCLS	11.41	6.81	1.95	11.10	6.54	1.89	9.52	5.64	1.61	8.46	4.87	1.26

[a] 10,000 simulations; 250 observations per simulation.
[b] ACV: asymptotic critical value in percentage.

Table 2: Size Comparison (in percentage) $H_0 : \beta = 0.5$ - Endogenous Tobit Model- $\mu_z = 1$[a]

	$\rho = 0.1$						$\rho = 0.9$					
	normal			t			normal			t		
ACV[b]	10	5	1	10	5	1	10	5	1	10	5	1
Wald-MLE	2.81	0.73	0.07	·	·	·	16.61	13.11	8.30	·	·	·
Wald-CGLS	3.06	0.76	0.03	3.21	0.77	0.01	43.53	36.68	25.31	43.39	36.87	25.21
Wald-SCLS	3.48	1.04	0.04	4.82	1.75	0.14	8.50	3.58	.39	5.54	2.27	0.20
S_{MD}-CGLS	10.40	5.47	1.11	10.35	5.16	1.15	10.31	5.22	1.01	10.35	5.16	1.15
K_{MD}-CGLS	10.34	5.33	1.10	10.31	4.85	0.87	9.81	5.22	0.87	9.73	4.74	0.84
J_{MD}-CGLS	10.57	5.42	1.06	10.40	5.35	1.18	10.29	5.13	0.96	10.65	5.31	1.09
KJ_{MD}-CGLS	10.34	5.13	1.00	9.86	5.03	1.06	9.97	5.13	1.00	9.77	4.85	0.95
CLR_{MD}-CGLS	10.34	5.53	1.06	10.09	5.37	1.06	9.88	5.58	0.88	9.98	5.38	0.83
S_{MD}-SCLS	11.35	6.73	2.22	10.22	6.15	1.82	9.17	5.51	1.90	8.04	4.60	1.39
K_{MD}-SCLS	11.09	6.01	1.77	10.12	5.28	1.39	11.06	6.37	1.93	9.73	5.40	1.57
J_{MD}-SCLS	10.23	5.53	1.50	9.41	5.36	1.48	7.21	3.74	1.17	6.13	3.47	0.97
KJ_{MD}-SCLS	11.10	6.42	1.95	10.18	5.75	1.74	10.63	6.34	2.00	9.27	5.40	1.58
CLR_{MD}-SCLS	11.20	7.01	2.16	10.31	6.28	1.69	11.06	6.88	2.14	9.34	5.68	1.64

[a] 10,000 simulations; 250 observations in each simulation.
[b] ACV: asymptotic critical value in percentage.

Table 3: Sizes (in percentage) for testing $H_0 : \beta = \beta_0$ at 10%, 5% and 1% significance levels - $\mu_z = 40$ [a]

	$\rho_{uv} = 0.1$						$\rho_{uv} = 0.9$					
	regular dispersion			overdispersion			regular dispersion			overdispersion		
ACV[b]	10	5	1	10	5	1	10	5	1	10	5	1
Wald-GMM	7.86	3.27	0.33	10.76	5.45	1.20	9.04	3.85	0.47	17.72	11.15	4.32
Wald-MDE	8.75	3.95	0.41	9.77	4.18	0.60	9.60	4.41	0.66	12.79	7.65	2.88
S_{MD}	10.34	5.62	1.26	11.41	5.87	1.17	10.47	5.69	1.25	12.34	6.46	1.64
K_{MD}	10.69	5.38	1.21	11.59	6.19	1.38	10.66	5.73	1.15	12.50	6.72	1.67
J_{MD}	9.98	5.19	1.09	10.56	5.61	1.07	10.03	5.36	1.16	11.23	5.84	1.30
KJ_{MD}	10.49	5.42	1.17	11.67	5.99	1.30	10.41	5.71	1.03	12.37	6.72	1.78
CLR_{MD}	10.84	5.81	1.17	11.69	6.59	1.32	10.69	6.17	1.15	12.45	7.09	1.66

[a] 10,000 simulations; 250 observations in each simulation.
[b] ACV: asymptotic critical value in percentage.

Table 4: Sizes (in percentage) for testing $H_0 : \beta = \beta_0$ at 10%, 5% and 1% significance levels - $\mu_z = 1$[a]

	$\rho_{uv} = 0.1$						$\rho_{uv} = 0.9$					
	regular dispersion			overdispersion			regular dispersion			overdispersion		
ACV[b]	10	5	1	10	5	1	10	5	1	10	5	1
Wald-GMM	2.35	0.56	0.01	13.79	7.79	1.93	4.2	1.28	0.09	28.89	19.04	6.49
Wald-MDE	1.23	0.30	0.00	1.31	0.29	0.01	3.94	1.49	0.11	23.83	14.68	4.23
S_{MD}	10.78	5.82	1.35	11.76	6.30	1.60	10.26	5.19	0.96	12.41	6.32	1.33
K_{MD}	10.77	5.92	1.22	11.68	5.78	1.44	10.43	5.39	1.24	12.23	6.50	1.44
J_{MD}	10.04	5.52	1.30	11.20	5.69	1.35	9.84	5.06	0.90	11.40	5.92	1.25
KJ_{MD}	11.16	5.94	1.31	11.55	6.14	1.44	10.11	4.99	1.18	12.49	6.47	1.43
CLR_{MD}	11.06	5.74	1.37	11.77	6.36	1.53	10.09	5.50	1.18	12.46	6.62	1.39

[a] 10,000 simulations; 250 observations in each simulation.
[b] ACV: asymptotic critical value in percentage.

Table 5: Female Labor Supply - Weekly hours in paid work [a]

Estimation Method	β	$\|t\|$-value	overidentification p-value	instruments [b]
MLE [c]	-0.057	(1.78)	·	(a)
CGLS	-0.077	(0.82)	0.001	(a)
SCLS	-0.111	(0.88)	0.35	(a)
WME ($\mathsf{w} = 11$)	-0.173	(0.91)	0.05	(a)
MLE [d]	-0.037	(1.01)	·	(a) + (b)
CGLS	-0.044	(0.85)	0.003	(a) + (b)
SCLS	-0.084	(1.45)	0.20	(a) + (b)
WME ($\mathsf{w} = 15$)	-0.119	(2.08)	0.45	(a) + (b)

[a] Number of observations: 3,382. The absolute value of the t-statistics are in parentheses.

[b] (a) $age \times education$, age^3, $education^3$, $age^2 \times education$, and $age \times education^2$. (b) 3 male occupation dummies.

[c] Exogeneity t-test: -0.56. First-stage F-statistic: 15.08.

[d] Exogeneity t-test: -3.29. First-stage F-statistic: 32.15.

Table 6: 95% Confidence Interval - Other Household Income

Estimation Method		Instruments	
		(a)	(a)+(b)
CGLS	Wald	$[-0.26,\ 0.11]$	$[-0.15,\ 0.06]$
	S_{MD}	$\emptyset$	$\emptyset$
	$\mathrm{KJ}_{\mathrm{MD}}$	$\emptyset$	$\emptyset$
	$\mathrm{CLR}_{\mathrm{MD}}$	$[-0.28,\ 0.16]$	$[-0.14,\ 0.08]$
SCLS	Wald[c]	$[-0.36,\ 0.14]$	$[-0.20,\ 0.03]$
	S_{MD}	$[-0.59,\ 0.23]$	$[-0.23,\ 0.06]$
	$\mathrm{KJ}_{\mathrm{MD}}$	$[-0.79,\ 0.14]$	$[-0.20,\ 0.03]$
	$\mathrm{CLR}_{\mathrm{MD}}$	$[-0.60,\ 0.13]$	$[-0.20,\ 0.03]$
WME	Wald	$[-0.55,\ 0.20]$	$[-0.23,\ -0.01]$
	S_{MD}	$[-0.26,\ 0.09]$	$[-0.30,\ 0.05]$
	$\mathrm{KJ}_{\mathrm{MD}}$	$[-0.57,\ 0.15]$	$[-0.23,\ -0.01]$
	$\mathrm{CLR}_{\mathrm{MD}}$	$[-0.38,\ 0.12]$	$[-0.23,\ -0.01]$

[a] Number of observations: 3,382.

[b] (a) $age \times education$, age^3, $education^3$, $age^2 \times education$, and $age \times education^2$.
(b) 3 male occupation dummies.

Table 7: Cigarette Demand Function Estimates[a]

Dependent Variable: smoked cigarettes per day (in packs)

Variable	Estimation Method			
	MDE (1)	GMM (2)	MDE (3)	GMM (4)
Habit Stock	0.0117	0.0035	0.0061	0.0040
	(1.06)	(0.93)	(2.60)	(1.66)
Cigarette Price	-0.0026	-0.0096	-0.0044	-0.0084
	(0.49)	(1.98)	(1.61)	(1.91)
Restaurant Restriction	-0.0174	-0.0476	-0.0684	-0.059
	(0.16)	(0.90)	(1.80)	(1.15)
Instrumental variables[b]	(a)	(a)	(a)+(b)	(a)+(b)
Overidentification test p-value	0.18	0.24	0.16	0.05

[a] Number of observations: 6160.
The absolute value of the t-statistics are in parentheses. First-stage F-statistic on the excluded instruments are 0.21 for columns 1 and 2, and 5.90 for columns 3 and 4.

[b] (a) - stage-level average price per-pack cigarette in 1978, number of years the state's restaurant smoking restrictions had been placed.
(b) - age^3, $education^3$, and $age \times education$.

Table 8: Definition of the variables, 3382 observations, 895 left-censored, 1987 US PSID

Variable	Definition
h	wife's working hours per weak
x	other household's income in \$1000
age	$\frac{\text{Age-40}}{10}$, where Age is wife's age in years
age^2	$\frac{(\text{Age}-40)^2}{100}$
$educ$	(Education − 8), where Education is wife's education in years
$educ^2$	$(\text{Education} - 8)^2$
C1	1 for any child between ages 0 to 5 and 0 otherwise
C2	1 for any child between ages 6 to 13 and 0 otherwise
C3	1 for any child between ages 14 to 17 and 0 otherwise
Race	1 if non-white and 0 otherwise
Husband occ 1	1 if husband is manager or professional and 0 otherwise
Husband occ 2	1 if husband is sales worker or clerical or craftsman and 0 otherwise
Husband occ 3	1 if husband is farm-related worker and 0 otherwise

Figures

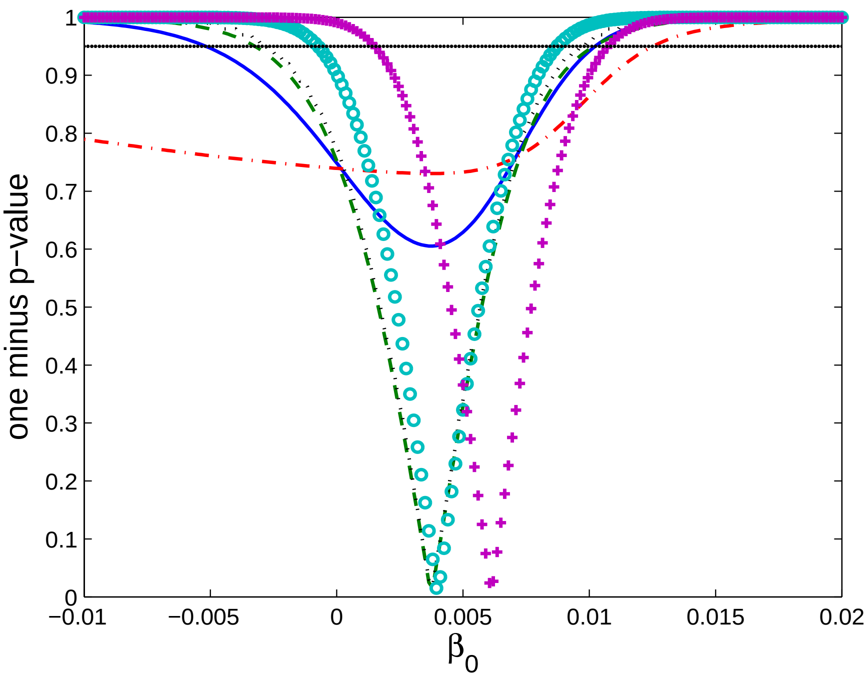

Figure 1: One minus p-value for statistics that test the value of habit stock parameter at different values of the segment line. Key: S_{MD}, solid line; K_{MD}, dashed line; J_{MD}, dashed-dotted line; $\mathrm{CRL}_{\mathrm{MD}}$, dotted line; GMM, circle; MDE, plus.

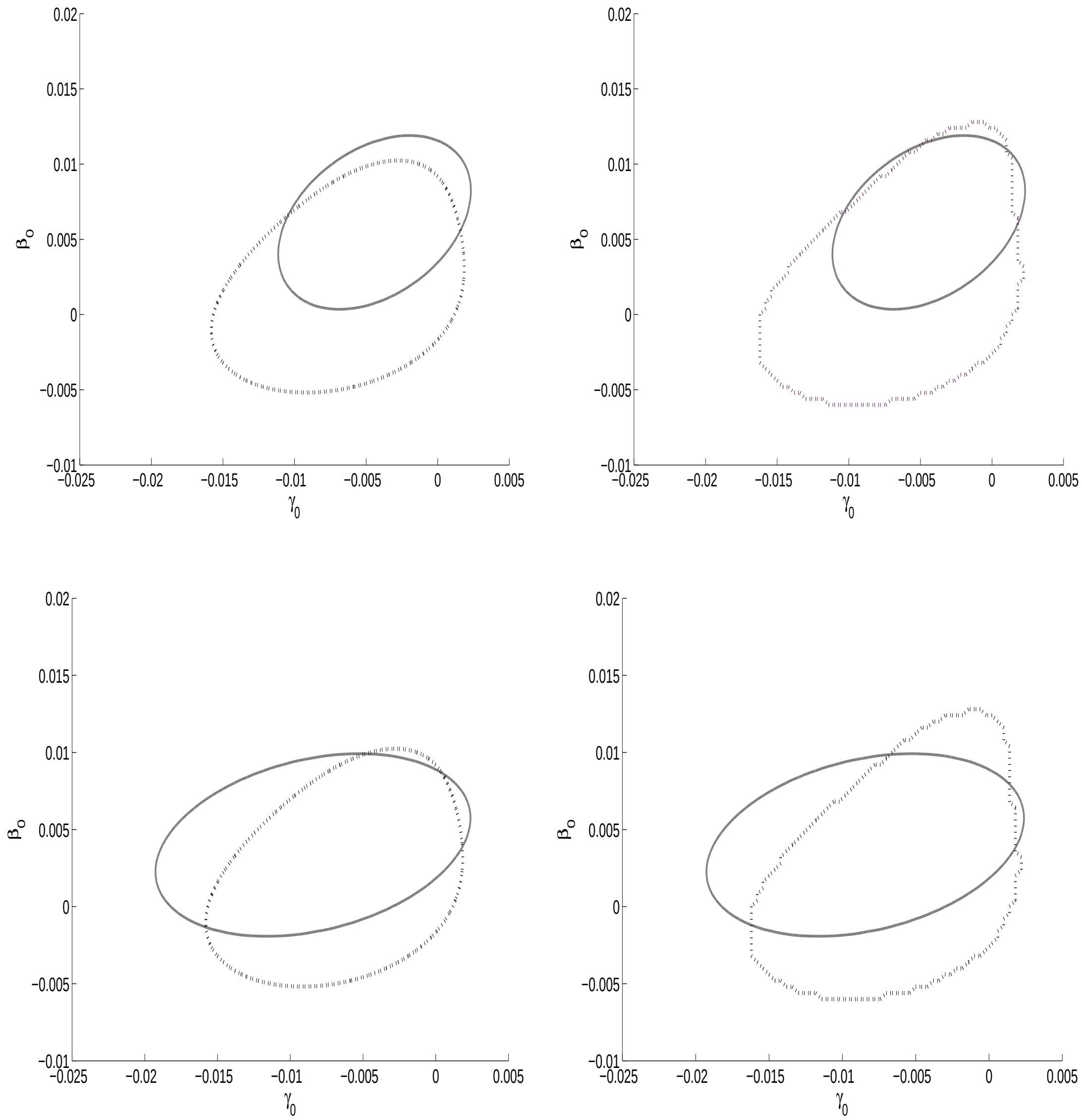

Figure 2: 95% confidence interval for habit stock (β) and cigarette price (γ). Key: Top left: MDE, solid line and S_{MD}, dotted. Top right: MDE, solid line and KJ_{MD}, dotted. Bottom left: GMM, solid line and S_{MD}, dotted. Bottom right: GMM, solid line and KJ_{MD}, dotted.