Università Politecnica delle Marche
DIPARTIMENTO DI ECONOMIA

Conditional Moment Tests for Normality in Bivariate Limited Dependent Variable Models: a Monte Carlo Study

Riccardo (Jack) Lucchetti   Claudia Pigini

QUADERNI DI RICERCA n. 357

June 2011

Abstract

In this paper, we run a Monte Carlo analysis of the finite-sample performance of an Information Matrix Test put forward by Smith (1985) for bivariate censored models. We use the bivariate probit model and Heckman selection model as examples.

Approximating the finite-sample distribution of this test statistic by its asymptotic distribution can lead to very misleading results: its size is severely distorted even in samples that common practice would judge to be perfectly adequate for asymptotics. This is especially true when the correlation coefficient is far from zero.

Power properties of the test statistic are investigated by using bivariate $t_{(6)}$ and $\chi^2_{(1)}$ alternatives. The test has very low power against leptokurtosis, especially in the bivariate probit case, while power against asymmetry appears to be much more satisfactory.

In general, the performance of the Information Matrix test seems to be related to the amount of information on the latent variables which survives the censoring mechanism. A somewhat improved version of the test can be obtained, in some cases, by a careful choice of the moment conditions to employ.

JEL Class.: C12, C15, C24, C35

Keywords: Information Matrix test, Monte Carlo simulation, Bivariate Probit, Sample Selection Model

Indirizzo: Riccardo (Jack) Lucchetti: Dipartimento di Economia - Università Politecnica delle Marche – Ancona, Italy
r.lucchetti@univpm.it
Claudia Pigini: Dipartimento di Economia - Università Politecnica delle Marche – Ancona, Italy
c.pigini@univpm.it

Conditional Moment Tests for Normality in Bivariate Limited Dependent Variable Models: a Monte Carlo Study*

Riccardo (Jack) Lucchetti *Claudia Pigini*

1 Introduction

In microeconometric applications, the bivariate normality assumption is seldom, if ever, tested after estimating models with limited dependent variables, despite the fact that a mis-specified distribution makes the maximum likelihood estimator inconsistent (Robinson, 1982). This is not the case for univariate normality in limited dependent variable models, where tests were proposed first by Bera, Jarque, and Lee (1984) and then by Chesher and Irish (1987); the latter has become a well-established procedure and is interpreted as a conditional moment test for regressions with grouped data, Probit and Tobit models.

With continuous dependent variables, testing for multivariate normality has become common practice and several options are available. Tests based on measures of multivariate skewness and kurtosis have been first brought forward by Mardia (1970b), where $\sqrt{b_1}$ and b_2 have respectively a χ^2 and normal distribution. Bowman and Shenton (1975) derived a test based on approximating the distribution of $\sqrt{b_1}$ and b_2 by the Johnson System. Alternatively, Cox and Small (1978) (later reviewed by Cox and Wermuth (1994)) tested multivariate normality with repeated standard regression tests of non linearity. Finally, the *Omnibus Test* by Doornik and Hansen (2008) uses a conditional gamma distribution for kurtosis based on Shenton and Bowman (1977).

However, these tests can not be applied in models subject to truncation or censoring and there have been, to our knowledge, only few attempts to test

* *We would like to thank an anonymous reviewer and the participants to the* 2nd Gretl Conference, *Toruń, 2011 for useful comments and suggestions.*

bivariate normality, even though this is a key assumption for some widely popular limited dependent variable models that exploit the joint normality of the random error terms, such as Heckman's endogenous selection model and the bivariate (possibly ordered) probit model.

One way to test bivariate normality in this kind of models is to follow the strategy proposed by Smith, who derived the Information Matrix (IM) test statistic for misspecification in bivariate (Smith, 1985) and multivariate (Smith, 1987) limited dependent variable models. The Information Matrix test for bivariate normality proposed by Smith has several advantages. First of all, as argued in Smith (1985), it is applicable to all censoring schemes. Moreover, building on earlier work by Gourieroux, Monfort, Renault, and Trognon (1984), Smith introduced the definition of *Generalised Error Product of Order (r,s)* (GEP henceforth), which bypasses the analytical difficulties of computing the score and the Hessian matrix and yields a general rule for their derivation, thus making it possible to formalise the IM test for the most general case, that is a multivariate limited information simultaneous equation system in latent variables subject to an arbitrary censoring scheme. Second, the Information Matrix test may be interpreted, in the same way as in Chesher and Irish (1987), as a conditional moment test[1] (Newey, 1985; Tauchen, 1985). Therefore, it may make sense to focus on a subset of the moment conditions, that is those associated with skewness and kurtosis and to test them separately (Hall, 1987) when more appropriate, even though the IM test is originally a general test for mis-specification. Finally, an IM test is trivial to implement in software, when analytical expressions for the score and the Hessian are available.

Despite these qualities, however, Information Matrix tests are generally known to deliver a poor finite sample performance when applied to the linear and univariate limited dependent variable models: especially the OPG (Outer Proder of Gradient — see Davidson and MacKinnon (1993)) form of the IM test tends to over-reject the true null hypothesis in finite samples. Monte Carlo studies carried out by Taylor (1987) and by Kennan and Neumann (1988) reported the extremely poor approximation of the asymptotic χ^2 distribution to the finite sample distribution of the IM test statistic. Orme (1990) also argues that the Chesher-Lancaster OPG version leads to a high upward finite-sample size bias. He presents several variants of an IM test statistic for truncated normal regression and Probit model that all have an nR^2 interpretation, showing, by Monte Carlo simulations, that the size bias can be considerably reduced by using a version subject to some parameter

[1] CM tests were famously advocated, in a closely related context, by Pagan and Vella (1989).

restrictions. Moreover, the size bias increases with large samples the higher the number of indicators used to compute the IM test statistic and also, in probit and Tobit models, the power against leptokurtic alternatives is extremely sensitive to the degree of censoring of the dependent variable (Skeels and Vella, 1999).

These studies were followed by several attempts aimed to overcome the size bias problem. Chesher and Spady (1991) show that the poor χ^2 approximation ultimately depends on the method chosen to compute the test. They develop an approximation to the finite sample distribution based on a $O(n^{-1})$ Edgeworth expansion. Davidson and MacKinnon (1992) introduce a new form of the IM test based on double length artificial regression, which performs better in small samples than the OPG variant, at least in the univariate case. However, the DLR variant cannot be applied with limited dependent variable models. Finally, Horowitz (1994) and Dhaene and Hoorelbeke (2004) put forward solutions based on various versions of the bootstrap.

There is, however, no evidence about the finite-sample performance of the Information Matrix test in bivariate limited dependent variable models. The aim of this paper is therefore to explore the size and power properties of the Information Matrix test by means of a comprehensive Monte Carlo experiment for two very common limited dependent variable models, bivariate probit and Heckman's endogenous selection model.

The paper is organised as follows: section 2 describes in detail the derivation of the test statistic for the bivariate probit and Heckman selection model, while section 3 contains the results of the Monte Carlo experiments and draws a brief comparison with alternative approaches; section 4 concludes.

2 Moment Conditions for the Bivariate Probit and Heckman Selection Models

2.1 Conditional Moments tests

We will use the OPG variant of the Information Matrix test (Chesher, 1983; Lancaster, 1984) to test for bivariate normality in the bivariate probit and Heckman selection model.

As is well known, the IM test, introduced by White (1982), is based on the Information Matrix equality: under correct specification of the model, the variance of the score plus the expected value of the Hessian should be zero. This fact provides a set of moment conditions that can be used to test whether the model is correctly specified.

Let θ be the parameter vector. The Information Matrix test statistic is therefore a test for $E(C_i) = 0$, where

$$C_i = \text{vech}\left[\frac{\partial^2 \ell_i}{\partial\theta\partial\theta'} + G_i G_i'\right], \quad (1)$$

evaluated at $\theta = \hat{\theta}_{ML}$; ℓ_i is the contribution to the log-likelihood of the i-th observation and $G_i \equiv \frac{\partial \ell_i}{\partial \theta}$.

For the models we are considering here, deriving the score vector and the Hessian matrix directly from the log-likelihood for observables is not overly complex. However, in the general context of bivariate models with censoring, the general setup proposed by Smith (1985) (see section A in the Appendix) can be used to derive moment conditions for any censoring scheme. We choose the former approach as it leads to analytical expressions that lend themselves to a much easier translation into numerically efficient code[2].

The Information Matrix test can be computed by means of an OPG regression (see Davidson and MacKinnon (1993)): the test statistic equals nR^2 of the regression of an n-vector of ones on a matrix M, with typical row $M_i' = [G_i', C_i']$. It is important to note that, in general, M may not be of full column rank, since the asymptotic χ^2 distribution of the IM test statistic under the null hypothesis has a number of degrees of freedom given by rank$(M) - k$. The possibility of rank deficiency, however, can be easily handled via numerical methods and will be briefly discussed later in this section and more in detail in section B in the appendix.

In the next two subsections, we will compute these quantities explicitly for the Bivariate Probit and the Heckman Sample Selection models. Although these models are well known, we will expose them in detail so to establish our notation clearly.

2.2 Bivariate Probit Model

The latent variable model is defined as

$$y_1^* = x_1'\beta_1 + v_1 \quad (2)$$
$$y_2^* = x_2'\beta_2 + v_2 \quad (3)$$

where x_1 and x_2 are a k_1-vector and a k_2-vector of exogenous variables, respectively and the error terms v_1 and v_2 are assumed to be jointly normal

[2]On the other hand, Smith's derivation of the score and Hessian matrix elements as functions of the GEP(r, s) allows us to immediately recognise the moment conditions tested in the OPG regression. Such expressions are not reported in this paper. Section A contains a brief exposition of GEPs and their relationship to the derivatives of the log-likelihood.

with unit variances and correlation coefficient ρ. The observable random vector $y = (y_1, y_2)$ is related to latent random variable $y^* = (y_1^*, y_2^*)$ via $y_j = \mathbf{I}(y_j^* > 0)$ for $j = 1, 2$ and $\mathbf{I}(\cdot)$ is the indicator function. Let us consider for simplicity the case $y_1 = 0$ and $y_2 = 0$ and define

$$\begin{aligned} a_i &= -x_{1i}'\beta_1 \\ b_i &= -x_{2i}'\beta_2 \end{aligned}$$

As customary in this kind of setting, we reparametrise the bivariate normal density via hyperbolic functions so that instead of the correlation coefficient ρ we will be using $\alpha = \operatorname{atanh}(\rho)$ and the associated quantities $c_\alpha = \cosh(\alpha)$ and $s_\alpha = \sinh(\alpha)$; thus, the contribution to the log-likelihood for observation i can be written as

$$\ell_i = \ln \Phi_2(a_i, b_i, \alpha) = \ln P_i \tag{4}$$

where

$$\Phi_2(a_i, b_i, \alpha) = \int_{-\infty}^{a_i} \int_{-\infty}^{b_i} \varphi_2(y_1, y_2, \alpha) \mathrm{d}y_2 \mathrm{d}y_1$$

with Φ_2 and φ_2 being the bivariate standard normal cumulative distribution and density function, respectively. So we have a vector $\theta' = (\beta_1', \beta_2', \alpha)'$ of $k = k_1 + k_2 + 1$ parameters to be estimated by ML. Let us define the function

$$u_{b_i,a_i} = c_\alpha b_i - s_\alpha a_i$$

and write the score elements for observation i as

$$\begin{aligned} G_i^{\beta_1} = S_i^{a_i} x_{1i}' &= \frac{\varphi(a_i)\Phi(u_{b_i,a_i})}{P_i} x_{1i}' \\ G_i^{\beta_2} = S_i^{b_i} x_{2i}' &= \frac{\varphi(b_i)\Phi(u_{a_i,b_i})}{P_i} x_{2i}' \\ G_i^{\alpha} = S_i^{\alpha} &= \frac{\varphi(b_i)\varphi(u_{a_i,b_i})}{P_i c_\alpha} \end{aligned}$$

The moment conditions C_i, expressed as functions of the score elements, are shown in Table 1.

2.3 The Heckman Selection Model

The latent variable model is defined as

$$y_i^* = x_i'\beta + \varepsilon_i \tag{5}$$

$$d_i^* = w_i'\gamma + v_i \tag{6}$$

Table 1: Moment Conditions for the Bivariate Probit Model

	β_1	β_2	α
β_1	$-\left[a_i S_i^{a_i} + c_\alpha s_\alpha S_i^\alpha\right] x_{1i} x_{1i}'$	$c_\alpha^2 S_i^\alpha x_{1i} x_{2i}'$	$-u_{a_i,b_i} c_\alpha S_i^\alpha x_{1i}$
β_2		$-\left[b_i S_i^{b_i} + c_\alpha s_\alpha S_i^\alpha\right] x_{2i} x_{2i}'$	$-u_{b_i,a_i} c_\alpha S_i^\alpha x_{2i}$
α			$S_i^\alpha \left[u_{a_i,b_i} u_{b_i,a_i} - t_\alpha\right]$

where x_i and w_i are an m-vector and a h-vector of exogenous variables, respectively; the error terms ε_i and v_i are assumed to be jointly normal with zero mean and covariance matrix

$$V \begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}$$

The observable random vector (y, d) is related to latent random variables (y^*, d^*) as:

$$y_i = \begin{cases} y_i^* & \text{if} \quad d_i^* > 0 \\ \\ \text{NA} & \text{if} \quad d_i^* \leq 0 \end{cases} \tag{7}$$

with $d_i = \mathbf{I}(d_i^* > 0)$. The contribution of observation i to the log-likelihood is

$$\ell_i = (1 - d_i) \ln \Phi(-b_i) + d_i \ln \Phi(a_i) - d_i \left(\ln \sqrt{2\pi} + \ln \sigma + \frac{u_i^2}{2} \right) \tag{8}$$

where $a_i = c_\alpha b_i + s_\alpha u_i$, $b_i = w_i'\gamma$ and $u_i = \frac{y_i - x_i'\beta}{\sigma}$. The parameter vector $\theta' = (\beta', \gamma', \sigma, \alpha)$ includes $k = m+h+2$ parameters. By using the generalised residuals

$$\mu_i = d_i \frac{\varphi(a_i)}{\Phi(a_i)} + (1 - d_i) \frac{\varphi(b_i)}{\Phi(-b_i)}$$

the score elements for observation can be written as:

$$\begin{aligned} G_i^\beta &= d_i \left(\frac{u_i - s_\alpha \mu_i}{\sigma} \right) x_i' \\ G_i^\gamma &= \left[d_i \mu_i c_\alpha + (1 - d_i)(-\mu_i) \right] w_i' \\ G_i^\sigma &= d_i \left(\frac{u_i (u_i - s_\alpha \mu_i) - 1}{\sigma} \right) \\ G_i^\alpha &= d_i \mu_i c_i \end{aligned}$$

where $c_i = c_\alpha u_i + s_\alpha b_i$; the moment conditions C_i are presented in Table 2.

Table 2: Moment Conditions for the Heckman Selection Model

$C_i^{\beta,\beta}$	$d_i \frac{1}{\sigma} \left[G_i^\sigma - \frac{s_\alpha c_\alpha}{\sigma} G_i^\alpha \right] x_i x_i'$
$C_i^{\beta,\gamma}$	$d_i \frac{c_\alpha^2}{\sigma} G_i^\alpha x_i w_i'$
$C_i^{\beta,\sigma}$	$d_i \frac{1}{\sigma^2} u_i \left[s_\alpha \mu_i (1 - s_\alpha a_i - 2u_i) + u_i^2 - 2 \right] x_i' - \sigma G_i^\beta$
$C_i^{\beta,\alpha}$	$d_i \frac{1}{\sigma} \left[\mu_i (s_\alpha a_i c_i + u_i c_i - c_\alpha) \right] x_i'$
$C_i^{\gamma,\gamma}$	$d_i(-c_\alpha^2 \mu_i a_i) + (1 - d_i)(-\mu_i b_i) w_i w_i'$
$C_i^{\gamma,\sigma}$	$d_i \frac{1}{\sigma} \left[c_\alpha \mu_i (s_\alpha a_i u_i + u_i^2 - 1) \right] w_i'$
$C_i^{\gamma,\alpha}$	$d_i \left[-\mu_i (c_\alpha a_i c_i - s_\alpha) \right]$
$C_i^{\sigma,\sigma}$	$d_i \frac{1}{\sigma^2} \left[u_i^4 - 5u_i^2 + 2 + s_\alpha u_i \mu_i (4 - a_i s_\alpha u_i - 2u_i^2) \right]$
$C_i^{\sigma,\alpha}$	$d_i \frac{1}{\sigma} \mu_i \left[c_i u_i (s_\alpha a_i + u_i) - (c_\alpha u_i + c_i) \right]$
$C_i^{\alpha,\alpha}$	$d_i \mu_i a_i (1 - c_i^2)$

2.4 The Rank Deficiency Problem

As mentioned above, the possibility of the degrees of freedom of the IM test statistic being less than $k(k+1)/2$ needs further discussion. One case in which this happens is when the model includes constant terms or some of the regressors in the two equations are the same; in this case, the corresponding elements of M_i are linear combinations of one another and, consequently, must be dropped from the OPG regression. As shown in Hall (1987) in the context of the normal linear model, the test statistic can be asymptotically decomposed into the sum of three independent components each detecting heteroskedasticity, skewness and excess kurtosis. Therefore, dropping elements of M_i in the OPG regression due to redundancy may lead to failure in detecting mis-specification.

In the case of the bivariate probit model it can be proven that the degrees of freedom are bounded between zero and $k(k+1)/2-1$. The upper bound for *df* the case of the Heckman selection model can also be proven to equal $k(k+1)/2-1$ in the most favourable conditions, while a lower bound is more difficult to obtain, and depends on the precise setup of the model. A detailed analysis for the cases we will analyse in the Monte carlo experiment of Section 3 is given in section B, in the Appendix.

3 The Monte Carlo Experiment

3.1 The test statistics

As briefly explained in section 2.1, the OPG version of the Information Matrix test may be interpreted as a conditional moment test (Newey (1985) and Tauchen (1985)) and the moment conditions of interest may be tested separately (Hall, 1987).

We present the results of the Monte Carlo experiment for several variants of the the Information Matrix test. For both models, we first use all moment conditions, that is we include all non-redundant columns of the M matrix (see tables (1) and (2)) in the OPG regression resulting in a general test for misspecification. Next, we include in the OPG regression those columns of M that contain third and fourth moment conditions[3].

In the bivariate probit model, third moment conditions appear in $C_i^{\beta_1,\alpha}$ and $C_i^{\beta_2,\alpha}$, while the only condition on the fourth moment is $C_i^{\alpha,\alpha}$. It would be desirable to use a mix of moment conditions that considers third and fourth moments only. However, this becomes impossible if both equations (2) and (3) contain a constant term, due to collinearity in the moment conditions (see section B.1 in the Appendix for details) . This is of course a condition which will occur in every possible realistic setting, including the DGP used in our experiment.

In the sample selection model third moment conditions are contained in $C_i^{\beta,\sigma}$, $C_i^{\beta,\alpha}$, $C_i^{\gamma,\sigma}$, $C_i^{\gamma,\alpha}$ and fourth moment conditions are in $C_i^{\sigma,\sigma}$, $C_i^{\sigma,\alpha}$ and $C_i^{\alpha,\alpha}$. For this model we are able to test jointly third and fourth moment conditions since the only moment condition that gets dropped from the OPG regression due to collinearity is $C_i^{\alpha,\alpha}$ (see section B.2 in the Appendix).

The richer structure of the Heckit model makes it possible to explore an alternative form of the test, obtained by selecting the two "diagonal" moment

[3]The presence of such moment conditions is not easy to spot inside the expressions given in section 2. However, it becomes obvious when rewriting those expressions using Smith's GEP formula.

conditions available in table (2), which contain, among others, fourth and cross-third moments, that is $C_i^{\sigma,\sigma}$ and $C_i^{\alpha,\alpha}$. We are unable to provide sound theoretical justification for our choice: a possible (vague) rationale that may be given is that, if a subset of the moment conditions has to be employed, those along the diagonal are more likely to carry useful information. We are so aware of the arbitrariness of our choice that we will refer to this form of the test as the result of a "cherry-picking" strategy.

All these variants of the IM test were analysed by running a Monte Carlo simulation based on 10000 replications and sample sizes of 1024, 4096 and 16384 observations.

3.2 Size Analysis

Tables (3), (4), (5) and (6) report the empirical size of the IM test statistic, at the 90%, 95% and 99% quantile of the appropriate χ^2 distribution. For the sample selection model results are reported for three different degrees of censoring: 10%, 50% and 75% of censored observations[4]. The correlation coefficient ρ between the random error terms varies from 0 to 0.90. Results are displayed only for $\rho = 0, 0.50, 0.90$ (results for $\rho = \pm 0.25$ and $\rho = 0.75$ are available upon request).

The DGP is described in detail at the end of Tables 3 and 4. The bold font indicates that the difference between the nominal and empirical size is not statistically significant at the 1% level.

The results of the Monte Carlo experiments for the bivariate probit model (Table 3) are consistent with the existing literature on its univariate counterpart (Skeels and Vella, 1999). As for the univariate probit model, the Information Matrix test using all moment conditions presents a severe size bias in the case of the bivariate extension as well. Although all tests appear to approach their nominal size in the limit, the null hypothesis is still highly over-rejected in samples as large as 16384 observations; the size distortion is especially severe for large values of ρ. In all three variants, a sample of 1024 observations is not enough to restrain the size bias. However, it seems that the size bias may be reduced if only the fourth moment condition is used.

The empirical size appears to be severely distorted in the Heckman selection model too. In all the variants of the test, the bias increases for non-zero values of the correlation coefficient. Also the stronger is the level of censoring considered, the higher is the size distortion (Tables 4, 5 and 6). The test based on all moment conditions abundantly over-rejects the true null hypothesis of correct specification even with a sample size of 16384 all over

[4] In practice, this amounted to setting the intercept in equation 6 to a suitable value.

the three degrees of censoring, while testing fewer moment conditions seems to make the test statistic approach faster its nominal size, especially when more information is available. The test with only fourth moment conditions exhibits a smaller size bias with a sample size of 4096, although it severely worsens when ρ is far from zero. The version with "cherry-picked" moment conditions exhibits the smallest distortion and behaves nicely with all sample sizes. However, even if less than the others, this variant's size is also sensitive to high values of the correlation coefficient and to the loss of information due to censoring.

It may be conjectured that the disappointing performance of the test statistics analysed here has to be ascribed to the usage of the OPG estimator: it has been known for a while that using the OPG matrix as an estimator of the information matrix leads to tests that have severe size distortions even in large samples (see for example Davidson and MacKinnon (1993, p. 477)). For this reason, we also experimented with a slight variant of the test which uses the analytical Hessian instead of the OPG estimator. The OPG-based statistic can be written as

$$W_{OPG} = c' \left[\bar{C}'\bar{C} - \bar{C}'G(G'G)^{-1}G'\bar{C} \right] c,$$

where $c = \bar{C}'\iota = \sum_{i=1}^{n} \bar{C}_i'$ and $\bar{C}$ is understood to be a suitable selection of the columns of C. By substituting $G'G$ with $-H$ (the negative analytical Hessian), we obtain the following form of the test

$$W_H = c' \left[\bar{C}'\bar{C} + \bar{C}'GH^{-1}G'\bar{C} \right] c,$$

which is asymptotically equivalent but should perform better than W_{OPG} if the conjecture is correct. However, we obtained strikingly similar results in all cases (these are not reported here, but of course are available upon request).

3.3 Power Analysis

We analyse the power of the Information Matrix test statistic against two alternatives running the experiments described above. For this purpose we use the same setup as in Gao and Lahiri (2000). Assume X_1 and X_2 are independent centred random variables. We use

$$v_1 = X_1 \tag{9}$$

$$v_2 = \rho X_2 + \sqrt{1-\rho^2} X_1 \tag{10}$$

as the two disturbance terms, which is equivalent to using a Gaussian copula. The marginal densities we will draw X_1 and X_2 from are a standardised $t_{(6)}$

for the first alternative and a standardised χ_1^2 for the second one. This choice enables us to focus on the consequences of leptokurtosis and asymmetry separately.

It can be said that the power properties of IM tests are generally quite unsatisfactory except in a few cases. Moreover, a number of unexpected results appear: for example, in some cases the power does not seem to be a monotone function of the sample size, putting the issue of test consistency into question. This is an aspect that needs to be further investigated and will form the object of future resaerch.

In the case of the bivariate probit model, the test statistic turns out to be generally very weak against the symmetric but leptokurtic alternative (table 7), unless an extremely large sample size is considered. While the power of the test is rather stable over different values of ρ when we consider all moment conditions, it remarkably worsens for values of the correlation coefficient higher than 0.5 when the variants based on third or fourth moments only are considered. On the other hand, the full and third-moments variants test statistic seems to be very powerful against skewness over all sample sizes (table 8)[5]. The same, unsurprisingly, cannot be said of the fourth-moment variant, which appears a rather poor choice.

Our results mirror closely those reported by Skeels and Vella (1999) for the univariate case. As we observe only the signs of the latent dependent variables, the information that can be captured by tail behaviour alone in latent variables with symmetric distributions is lost when the censoring takes place. That is, the observed signs remain unchanged whenever the mis-specification does not modify the proportions of zeros and ones in the dependent variables. As a result, the test fails to detect excess kurtosis. This also explains why the test statistic has good power against skewed alternatives, since the bivariate probit censoring scheme is able to keep the information about the change in the proportion of zeros and ones taken by observations produced by this particular form of mis-specification.

Results for the Heckman selection model are displayed in tables (9), (10) and (11). Only power for the bivariate $t_{(6)}$ is presented, since under a bivariate right and left skewed $\chi^2_{(1)}$ distribution the test *always* rejected the null in all our simulations.

All variants of the test exhibit power properties that are considerably more satisfactory than in the bivariate probit case: the censoring scheme is less strict than bivariate probit, so (partial) observability of y_t makes it

[5]Monte Carlo simulations have been run also to study the power against a left skewed bivariate χ_1^2. Results are practically the same as the ones reported against right skewness and are not reported in the paper.

possible, for example, to detect leptokurtosis. This is confirmed by the fact that the tests based on fourth moment conditions (with and without third moments) have more power against the $t_{(6)}$ and that power properties get worse as the degree of censoring grows large.

The test variant based on third moments only also shows good power properties, with the exception of the $\rho = 0$ case, in which the power seems to *decrease* with the sample size. However, it is worth noting that with our setup in the $\rho = 0$ case the two disturbances are effectively independent from one another, as opposed to just incorrelated; therefore, by symmetry of the marginals, all higher-order odd moments are zero or non-existent. The reverse is not necessarily true, so the power properties of this variant of the test in a more general case remain to be ascertained.

The "cherry-picked" variant of the test appears to have reasonably good power properties across all cases considered, taking of course into account the fact (common to all variants) that higher degrees of censoring lead to a noticeable power loss.

3.4 Comparison with Edgeworth-based Alternatives

Our Monte Carlo experiments reveal that the finite-sample performance of the Information Matrix test is rather poor, even in cases that can be reasonably considered good for asymptotics: results confirm the presence of a marked size bias and show a power loss due to censoring. In view of previous related results, such as those cited in section 1, this is hardly surprising. The only advantages IM tests seem to offer in this context are twofold: first, they are derived from a very general principle, which makes them easy to generalise to arbitrary censoring schemes; moreover, they are rather inexpensive to implement in software, which makes them suitable for numerically intensive procedures such as bootstrapping critical values to control for size.

The literature provides a parametric alternative to conditional moments tests for normality in bivariate limited dependent variable models. These are LM tests in which the distribution of the error terms is modelled via a bivariate truncated (type AA) Edgeworth expansion, such as the one proposed first by Murphy (2007) for the bivariate probit model and then by Montes-Rojas (2011) for the the Heckman selection model (building on earlier work by Lee (1984)). These tests, however, are not only more complex to compute; as Smith (1985) points out,

> the Type AA curve is not a proper p.d.f. and may yield negative values; ... [it] is used as an artefact to model locally non-normal behaviour of the errors and thus to generate test statistics with

> power against such local alternatives. Attempts to use the type AA curve to model non-normal density functions in practice have not been satisfactory ...

and a similar point was also raised by Mardia (1970a). Proponents of EE tests argue that this point is likely to be moot in a setting, such as the score test, in which only local properties of the approximate density are used, but the relevance of the problem clearly has to be established on a case-by-case basis.

However, since conditional moment tests also suffer from severe size distortions, the Edgeworth based tests may have a relatively better finite-sample performance. To the best of our knowledge, no comprehensive study on the comparison of these tests with conditional moments alternatives has been proposed.

The Montecarlo experiment run by Murphy (2007) is not directly comparable to ours because the reported critical values are bootstrapped. We are, however, able to run a Monte Carlo analysis on the same setup considered in Montes-Rojas (2011) (see end of table 12).[6] It would seem that the Edgeworth based test has in fact a more contained size bias in very small samples, while the are no major differences between the two alternatives for power against leptokurtosis (see table 13).

Another comparison we can draw between our IM test statistics and the Edgeworth-based ones is based on the famous dataset Mroz (1987) on female labour force participation; the results from the application on the Mroz dataset (see Montes-Rojas (2011) for details) displayed in table 14 are controversial. The IM variant that appears to deliver the best performance in terms of empirical size accepts the null hypothesis of bivariate normality in the wage equation while the Edgeworth based test's p-value is always zero, which is not entirely easy to reconcile with the idea that IM tests are more prone to mistakenly reject the null if true, while power properties are comparable.

On this basis, the comparison is ambiguous and inconclusive. On the one hand, the complexity of computation and possible problems with values of the density function may divert from the use of these tests. On the other, such costs may be justified by a much better finite sample performace. The possible superiority of Edgeworth-based tests with respect to their IM counterparts has probably to be assessed by means of a comprehensive dedicated simulation experiment, which we leave for future research.

[6]In Tables 12 and 13, the values for the Edgeworth-based test are not computed by simulation, but simply copied from Montes-Rojas (2011) for ease of comparison. This is also the reason why the left bottom panel in Table 13 is empty.

4 Final Remarks

Approximating the finite-sample distribution of the IM test statistic by its asymptotic distribution can lead to very misleading results in both the bivariate probit and Heckman selection model: its size is severely biased even in samples that common practice would consider as perfectly adequate for asymptotics. This is especially true when the correlation coefficient is far from zero.

Power properties of the test statistic seem to be good (although somewhat unspectacular) against skewed alternatives, but rather dismal against leptokurtosis in the bivariate probit case. However, the latter result is, as shown by previous related literature, a necessary consequence of the particular censoring scheme we adopted and the possibility itself of testing for leptokurtosis in binary models appears to be in doubt.

A somehow related point can be made for the Heckman selection model: the greater the amount of information lost to censoring, the worse the finite-sample performance of the test statistic seems to be, either in terms of size bias and of power loss.

In conclusion: the IM test, in the form proposed by Smith (1985), does not provide a reliable tool to detect non-normality in bivariate limited dependent variable models unless in very large samples: the null is over-rejected when true and often accepted when false (especially for symmetric alternatives). However, if one considers variants based on a limited set of moment conditions, there may be cases where reasonable finite-sample properties and a rather simple computation may justify the choice of this test.

References

BERA, A. K., C. M. JARQUE, AND L.-F. LEE (1984): "Testing the Normality Assumption in Limited Dependent Variable Models," *International Economic Review*, 25(3), 563–78.

BOWMAN, K. O., AND L. R. SHENTON (1975): "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, 62(2), 243–250.

CHESHER, A. (1983): "The Information Matrix Test : Simplified Calculation via a Score Test Interpretation," *Economics Letters*, 13(1), 45–48.

CHESHER, A., AND M. IRISH (1987): "Residual Analysis in the Grouped and Censored Normal Linear Model," *Journal of Econometrics*, 34, 33–61.

Chesher, A., and R. Spady (1991): "Asymptotic Expansions of the Information Matrix Test Statistic," *Econometrica*, 59(3), 787–815.

Cox, D. R., and N. J. H. Small (1978): "Testing multivariate normality," *Biometrika*, 65(2), 263–272.

Cox, D. R., and N. Wermuth (1994): "Tests of linearity, multivariate normality and the adequacy of linear scores," *Applied Statistics*, 43, 347–355.

Davidson, R., and J. G. MacKinnon (1992): "A New Form of the Information Matrix Test," *Econometrica*, 60(1), 145–57.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and Inference in Econometrics*. Oxford University Press.

Dhaene, G., and D. Hoorelbeke (2004): "The Information Matrix Test with Bootstrap-Based Covariance Matrix Estimator," *Economics Letters*, 84, 341–347.

Doornik, J. A., and H. Hansen (2008): "An Omnibus Test for Univariate and Multivariate Normality," *Oxford Bulletin of Economics and Statistics*, 70(s1), 927–939.

Gao, C., and K. Lahiri (2000): "Further consequences of viewing LIML as an iterated Aitken estimator," *Journal of Econometrics*, 98(2), 187–202.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1984): "Résidus généralisé ou interprétations linéaires de l'économétrie non linéaire," Discussion Paper 8410, Document de travail, INSEE.

Hall, A. (1987): "The Information Matrix Test for the Linear Model," *Review of Economic Studies*, 54(2), 257–263.

Horowitz, J. L. (1994): "Bootstrap-Based Critical Values for the Information Matrix Test," *Journal of Econometrics*, 61, 365–411.

Kennan, I., and G. Neumann (1988): "Why does the information matrix test reject too often? A diagnosis of some Monte Carlo symptoms," Discussion Paper 88, Working papers in economics, Hoover Institution, Stanford University, Stanford, CA.

Lancaster, T. (1984): "The Covariance Matrix of the Information Matrix Test," *Econometrica*, 52, 1051–1053.

Lee, L.-F. (1984): “Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity,” *Econometrica*, 52(4), 843–863.

Mardia, K. V. (1970a): *Families of Bivariate Distributions*. London, Griffin.

——— (1970b): “Measures of multivariate skewness and kurtosis with applications,” *Biometrika*, 57(3), 519–530.

Montes-Rojas, G. V. (2011): “Robust misspecification tests for the Heckman’s two-step estimator,” *Econometric Reviews*, 30(2), 1–19.

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

Murphy, A. (2007): “Score tests of normality in bivariate probit models,” *Economics Letters*, 95(3), 374–379.

Newey, W. K. (1985): “Maximum Likelihood Specification Testing and Conditional Moment Tests,” *Econometrica*, 53, 1047–1070.

Orme, C. (1990): “The small-sample performance of the information-matrix test,” *Journal of Econometrics*, 46(3), 309–331.

Pagan, A., and F. Vella (1989): “Diagnostic Tests for Models Based on Individual Data: A Survey,” *Journal of Applied Econometrics*, 4(S), S29–59.

Robinson, P. M. (1982): “On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables,” *Econometrica*, 50(1), 27–41.

Shenton, L. R., and K. O. Bowman (1977): “A bivariate model for the distribution of $\sqrt{b_1}$ and b_2,” *Journal of the American Statistical Association*, 72, 206–211.

Skeels, C. L., and F. Vella (1999): “A Monte Carlo investigation of the sampling behavior of conditional moment tests in Tobit and Probit models,” *Journal of Econometrics*, 92(2), 275–294.

Smith, R. J. (1985): “Some Tests for Misspecification in Bivariate Limited Dependent Variables Models,” *Annales de l’INSÉÉ*, 59/60, 97–123.

——— (1987): "Testing the Normality Assumption in Multivariate Simultaneous Limited Dependent Variable Models," *Journal of Econometrics*, 34(1-2), 105–123.

Tauchen, G. E. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415–443.

Taylor, L. W. (1987): "The Size Bias of the Information Matrix Test," *Economics Letters*, 24, 63–67.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1), 1–25.

Table 3: Empirical Size, Bivariate Probit Model

	All Moment Conditions, df=9								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	46.98	56.32	87.31	38.14	47.91	82.83	24.11	33.17	72.67
4096	25.73	29.71	59.13	17.41	21.37	51.65	7.77	10.36	38.65
16384	14.72	17.18	31.83	8.62	10.61	24.00	2.65	4.02	13.57
	Third Moment Conditions, df=4								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	28.71	32.54	48.38	21.44	24.95	38.26	11.54	13.85	22.17
4096	16.90	18.79	28.63	10.99	12.51	20.93	4.32	5.23	10.62
16384	12.29	13.51	17.05	6.94	7.57	11.10	2.02	2.40	4.10
	Fourth Moment Conditions, df=1								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	11.85	16.47	30.12	6.07	10.78	22.18	1.44	4.09	11.70
4096	**10.40**	13.03	19.14	**5.34**	7.32	12.78	**1.01**	2.41	6.02
16384	**9.72**	11.73	13.10	**4.89**	6.41	7.96	**1.03**	1.49	3.06

Explanatory variables for eq. 1: constant and x_1; explanatory variables for eq. 2: constant and x_2; x_1 and x_2 are independent standard normal r. v. s. Parameters values are $\beta_{jr} = 1$ for $j = 0, 1$ and $r = 1, 2$. Monte Carlo results are based on 10000 replications.
For precise details on the costruction of the tests see 3.1

Table 4: Empirical Size, Heckman Selection Model, Censored Observations 10%

	All Moment Conditions, df=15								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	52.45	59.78	88.02	42.39	50.95	83.71	26.51	35.70	73.74
4096	28.10	32.59	58.74	19.77	24.27	50.51	9.27	12.22	35.62
16384	16.96	18.17	32.62	10.35	11.44	24.16	3.22	4.27	12.65
	Third Moment Conditions, df=7								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	20.72	27.59	53.37	13.04	18.45	44.28	4.65	8.13	28.90
4096	13.68	17.17	31.94	7.68	10.47	23.48	2.21	3.56	12.38
16384	10.87	12.37	18.94	5.71	6.88	11.89	**1.25**	1.94	4.59
	Fourth Moment Conditions, df=3								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	18.84	28.03	48.66	12.30	20.54	40.60	4.92	11.08	28.30
4096	12.62	18.74	29.38	7.53	12.55	22.79	2.24	5.19	13.38
16384	10.90	13.62	19.47	5.74	7.91	13.22	1.43	2.66	6.07
	Third and Fourth Moment Conditions, df=9								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	33.92	43.10	73.48	25.39	33.97	66.71	12.90	20.16	53.32
4096	18.65	25.13	44.50	11.86	17.32	36.43	4.55	8.00	23.61
16384	12.92	15.45	25.75	7.43	9.23	17.85	1.99	3.18	8.59
	Selected Moment Conditions, df=2								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	16.04	22.22	40.53	9.83	15.18	33.96	3.69	7.19	23.56
4096	11.91	16.38	25.13	6.54	10.20	18.91	1.90	3.74	11.21
16384	**10.44**	12.31	17.16	5.62	7.59	11.54	1.35	2.42	5.21

Explanatory variables for the main eq.: constant and x; explanatory variables for the selection eq. : constant and w; x and w are independent standard normal r. v. s. Parameters values are $\beta_r = 1$ for $j = 0, 1$, $\gamma_0 = \sqrt{2}\Phi^{-1}(p)$, where $p = 0.9$ is the percentage of uncensored observations, and $\gamma_1 = 1$. Monte Carlo results are based on 10000 replications.

Table 5: Empirical Size, Heckman Selection Model, Censored Onservations 50%

All Moment Conditions, df=15									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	59.48	60.21	75.53	50.18	51.67	68.70	34.15	35.09	56.64
4096	30.96	31.50	42.29	22.74	22.69	32.74	11.09	11.07	19.74
16384	18.38	16.90	22.12	11.60	10.63	14.42	4.29	3.44	5.76
Third Moment Conditions, df=7									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	22.46	25.18	44.46	14.95	16.63	35.00	5.60	6.63	21.38
4096	15.00	15.65	24.25	8.63	9.04	16.80	2.41	2.91	7.50
16384	11.49	11.75	14.89	6.37	6.02	8.84	1.48	1.53	2.92
Fourth Moment Conditions, df=3									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	25.53	28.10	41.70	18.61	21.02	34.29	9.61	11.61	23.33
4096	15.99	18.32	24.95	10.15	11.87	18.23	3.73	5.08	9.80
16384	12.18	12.93	16.26	6.72	7.08	10.51	2.04	2.44	4.29
Third and Fourth Moment Conditions, df=9									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	40.45	42.85	60.66	31.73	33.69	51.99	18.56	19.81	38.52
4096	22.01	24.02	34.00	14.72	16.47	26.27	6.31	7.12	14.39
16384	13.85	14.77	19.49	8.27	8.68	12.74	2.52	2.92	5.00
Selected Moment Conditions, df=2									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	20.86	23.65	31.99	14.44	17.40	25.35	6.82	9.39	16.30
4096	13.86	16.05	19.35	8.25	10.37	13.44	2.84	4.38	6.73
16384	11.24	12.19	14.32	5.94	6.78	8.71	1.74	2.17	3.43

$p = 0.50$

Table 6: Empirical Size, Heckman Selection Model, Censored Observations 75%

All Moment Conditions, df=15

	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	75.23	73.46	83.96	67.78	65.93	78.56	52.44	50.82	66.80
4096	41.00	40.09	51.91	31.53	30.83	42.47	18.05	17.23	27.54
16384	21.01	21.20	28.76	13.70	13.56	20.78	5.56	5.32	10.02

Third Moment Conditions, df=7

	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	30.00	30.62	49.42	20.45	21.44	39.72	9.10	9.57	24.66
4096	17.19	17.33	26.68	10.04	10.51	18.53	3.39	3.46	8.68
16384	12.71	12.28	16.95	6.97	6.76	10.61	1.77	1.83	3.72

Fourth Moment Conditions, df=3

	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	32.87	33.11	47.73	25.52	25.88	35.19	15.25	15.62	23.80
4096	18.84	19.35	25.42	12.91	13.32	18.95	5.81	5.95	10.16
16384	12.51	13.36	16.82	7.82	7.70	11.12	2.65	2.59	4.40

Third and Fourth Moment Conditions, df=9

	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	55.26	54.95	67.45	46.36	45.79	59.74	31.84	31.25	45.91
4096	28.73	29.85	38.71	20.55	21.51	29.78	10.26	10.83	17.28
16384	17.02	16.72	22.07	10.42	10.52	15.32	3.74	3.75	6.56

Selected Moment Conditions, df=2

	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	24.77	27.77	33.33	18.26	20.94	26.18	9.64	12.01	17.01
4096	15.42	16.87	20.36	10.01	11.21	14.40	4.24	4.77	7.37
16384	11.68	12.46	14.56	6.29	7.19	8.91	2.00	2.21	3.23

$p = 0.25$

Table 7: Empirical Power, Bivariate $t_{(6)}$ Distribution, Bivariate Probit Model

	All Moment Conditions, df=15								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	38.16	48.06	86.01	28.74	38.68	80.41	15.45	23.92	67.13
4096	36.51	38.28	67.58	22.79	26.39	57.11	7.46	10.22	38.29
16384	97.23	96.74	98.60	94.14	92.85	96.41	78.65	75.92	85.77
	Third Moment Conditions, df=7								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	20.53	24.16	44.43	12.92	16.30	33.84	5.29	7.32	18.41
4096	25.41	21.38	26.51	13.37	11.00	16.74	2.92	2.36	6.15
16384	85.73	75.59	50.64	75.32	60.76	34.93	45.98	29.33	12.07
	Fourth Moment Conditions, df=3								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	13.02	15.40	27.82	6.93	7.98	18.76	1.51	2.22	8.37
4096	10.66	23.80	20.17	5.22	12.90	10.75	0.88	2.32	3.28
16384	10.23	67.47	45.88	5.23	52.00	28.63	0.95	22.51	7.48

Table 8: Empirical Power, Bivariate $\chi^2_{(1)}$ Distribution, Bivariate Probit Model

	All Moment Conditions, df=15								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	100	100	100	100	100	100	100	100	100
4096	100	100	100	100	100	100	100	100	100
16384	100	100	100	100	100	100	100	100	100
	Third Moment Conditions, df=7								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	100	100	99.98	100	100	99.92	100	100	99.74
4096	100	100	100	100	100	100	100	100	100
16384	100	100	100	100	100	100	100	100	100
	Fourth Moment Conditions, df=3								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	16.37	44.66	35.89	8.82	28.40	22.78	2.36	7.74	7.51
4096	12.11	90.83	74.71	6.17	80.98	58.67	1.26	50.77	26.66
16384	10.28	100	99.88	5.26	99.99	98.62	1.04	99.58	87.61

Table 9: Empirical Power, Bivariate $t_{(6)}$ Distribution, Heckman Selection Model, Censored Observations 10%

All Moment Conditions, df=15									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	92.16	91.42	97.00	86.24	86.06	94.79	69.44	69.07	87.85
4096	98.63	99.75	99.80	97.27	99.41	99.53	92.01	97.17	97.68
16384	99.59	100	100	99.14	100	100	97.66	100	100
Third Moment Conditions, df=7									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	43.17	75.03	80.66	33.28	64.99	72.14	18.27	43.34	52.62
4096	29.77	99.42	99.71	20.49	98.60	99.29	8.87	94.70	96.06
16384	18.65	100	100	11.14	100	100	3.83	100	100
Fourth Moment Conditions, df=3									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	96.84	95.80	94.45	91.73	89.94	87.57	71.59	67.79	67.91
4096	99.49	99.98	99.16	97.99	99.79	97.68	93.52	97.72	91.83
16384	99.80	100	99.99	99.27	100	99.91	97.13	99.99	99.19
Third and Fourth Moment Conditions, df=9									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	92.01	91.02	93.43	85.24	83.51	88.97	63.67	72.39	75.53
4096	99.39	99.90	99.91	98.36	99.77	99.63	94.03	98.74	97.86
16384	99.73	100	100	99.52	100	100	98.66	100	100
Selected Moment Conditions, df=2									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	98.84	98.01	96.12	94.39	92.54	90.08	74.79	71.83	68.80
4096	99.93	99.97	99.77	98.60	99.09	97.60	92.12	93.13	90.04
16384	100	100	99.99	99.29	99.92	98.76	96.04	98.87	94.86

Table 10: Empirical Power, Bivariate $t_{(6)}$ Distribution, Heckman Selection Model, Censored Observations 50%

	All Moment Conditions, df=15								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	85.13	88.02	95.15	77.66	81.99	92.21	59.61	66.55	84.18
4096	96.98	99.21	99.95	94.04	98.27	99.83	82.73	93.86	99.35
16384	99.29	100	100	98.78	100	100	96.28	100	100
	Third Moment Conditions, df=7								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	43.53	78.35	90.08	33.48	69.36	84.03	18.21	48.58	68.23
4096	27.29	99.31	99.99	18.54	98.33	99.95	7.43	93.64	99.25
16384	18.07	100	100	10.64	100	100	3.30	100	100
	Fourth Moment Conditions, df=3								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	87.85	88.10	87.38	76.95	77.58	77.85	48.54	50.56	54.10
4096	99.24	98.98	99.85	97.35	96.13	95.91	89.63	85.77	84.86
16384	99.62	99.89	99.80	98.93	98.87	98.41	96.28	93.96	93.37
	Third and Fourth Moment Conditions, df=9								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	80.37	83.86	92.26	69.30	75.61	87.79	45.74	55.81	75.40
4096	98.16	99.65	99.98	96.00	99.16	99.93	87.26	95.99	99.64
16384	99.67	100	100	99.31	100	100	97.80	100	100
	Selected Moment Conditions, df=2								
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	92.27	91.68	89.10	82.38	81.66	78.51	52.60	52.45	50.64
4096	99.80	99.94	99.88	97.68	98.43	98.12	88.36	88.53	87.88
16384	100	100	100	98.92	99.78	99.68	94.91	92.25	95.06

Table 11: Empirical Power, Bivariate $t_{(6)}$ Distribution, Heckman Selection Model, Censored Observations 75%

All Moment Conditions, df=15									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	84.39	92.18	97.02	77.23	88.01	95.47	61.33	77.14	89.51
4096	91.01	98.91	99.93	83.93	97.81	99.78	64.99	92.37	99.04
16384	98.46	100	100	97.26	100	100	92.57	100	100
Third Moment Conditions, df=7									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	53.63	81.76	90.81	34.40	74.09	85.74	26.20	56.64	73.22
4096	33.84	99.11	99.94	24.27	97.94	99.83	11.63	92.64	98.98
16384	21.07	100	100	13.31	100	100	4.54	100	100
Fourth Moment Conditions, df=3									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	67.89	83.71	84.43	54.23	74.40	75.09	29.59	52.35	54.02
4096	97.59	99.53	99.64	92.99	98.09	97.23	75.85	90.14	86.59
16384	99.51	100	100	98.47	99.90	99.66	94.41	99.04	95.77
Third and Fourth Moment Conditions, df=9									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	71.66	86.38	94.03	61.67	79.98	90.67	41.85	65.58	81.53
4096	93.08	99.42	99.95	86.48	98.49	99.83	65.33	93.79	99.32
16384	99.20	100	100	98.25	100	100	94.81	100	100
Selected Moment Conditions, df=2									
	90%			95%			99%		
n \ ρ	0.00	0.50	0.90	0.00	0.50	0.90	0.00	0.50	0.90
1024	73.36	82.96	83.32	58.89	71.20	72.71	30.78	44.94	48.06
4096	98.91	98.84	99.92	94.89	98.31	99.00	77.80	88.66	88.61
16384	99.91	100	100	98.42	100	99.99	92.74	98.72	96.44

Table 12: Edgeworth-based test vs conditional moment test for the Heckit model, bivariate normal distribution, empirical size at 90% and 95%

	Selected Moment Conditions, df=2							
	90%				95%			
n \ ρ	0.20	0.40	0.60	0.80	0.20	0.40	0.60	0.80
200	56.40	61.85	70.95	73.50	49.43	54.64	64.84	68.44
500	39.43	46.20	56.48	67.01	33.03	39.40	48.44	60.68
1000	31.00	35.70	41.10	53.15	23.70	30.40	34.10	47.45

	EE, df= 9							
	90%				95%			
n \ ρ	0.20	0.40	0.60	0.80	0.20	0.40	0.60	0.80
200	30.60	36.60	47.50	54.90	24.20	36.60	40.50	44.90
500	24.70	30.20	39.50	43.20	15.00	25.10	33.20	37.00
1000	20.30	24.90	30.20	37.90	13.50	22.00	29.70	35.40

Main eq.: $y_i^* = 1 + 0.5x_1 - 0.5x_2 + \varepsilon$ where $x_1, x_2 \sim N(0, 3)$ and $\varepsilon_i \sim N(0; 4)$. Selection eq.: $d_i^* = 1 - w_1 + x_2 + v$ where $w_i \sim U[-3; 3]$ and $v \sim N(0, 1)$; Censored observations are about 50%. Monte Carlo results are based on 1000 replications.

Table 13: Edgeworth-based test vs conditional moment test for the Heckit model, bivariate t_3 distribution, empirical size at 90% and 95%

	Selected Moment Conditions, df=2							
	90%				95%			
n \ ρ	0.20	0.40	0.60	0.80	0.20	0.40	0.60	0.80
200	91.98	89.78	88.95	89.75.	87.24	84.09	84.75	85.73
500	99.30	99.20	98.38	98.03	97.90	97.20	96.35	94.81
1000	100	100	100	99.70	99.60	99.60	100	99.70
	EE, df= 9							
	90%				95%			
n \ ρ	0.20	0.40	0.60	0.80	0.20	0.40	0.60	0.80
200					90.60	97.70	99.90	100
500					96.70	98.20	99.80	100
1000					100	100	100	100

The values for the Edgeworth-based test are simply copied from Montes-Rojas (2011). In his paper, the results at the 90% level for the case here considered are not reported, so the left bottom panel is left blank.

Table 14: Empirical Application: Test Statistics

	Wage Equation		Hours Equation	
All	$\chi^2_{93} = 209.32$	p-val.=0.000	$\chi^2_{106} = 270.96$	p-val.=0.000
Third	$\chi^2_{21} = 74.02$	p-val.=0.000	$\chi^2_{23} = 87.10$	p-val.=0.000
Fourth	$\chi^2_{3} = 34.03$	p-val.=0.000	$\chi^2_{3} = 25.79$	p-val.=0.000
Third & Fourth	$\chi^2_{25} = 97.65$	p-val.=0.000	$\chi^2_{23} = 76.18$	p-val.=0.000
Cherry-Picking	$\chi^2_{2} = 7.80$	p-val.=0.020	$\chi^2_{2} = 9.33$	p-val.=0.009
Edgw.-based	$\chi^2_{9} = 383.3$	p-val.=0.000	$\chi^2_{9} = 67.1$	p-val.= 0.000

A Information Matrix Tests in Bivariate Limited Dependent Variable Models

Consider the latent variable model[7]:

$$
\begin{aligned}
y_1^* &= x_1'\beta_1 + v_1 \qquad (11)\\
y_2^* &= x_2'\beta_2 + v_2 \qquad (12)
\end{aligned}
$$

where v_1 and v_2 have the following bivariate normal distribution:

$$
\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix}\right] \qquad (13)
$$

where x_1 and x_2 are vectors of exogenous variables and β_1, β_2 are the parameter vectors. Since v_1 can be written as $v_1 = \rho v_2 + u$, where $\rho = \omega_{12}/\omega_2^2$, the model for $y_1^*|y_2^*$ is:

$$
\begin{aligned}
y_1^* &= x_1'\beta_1 + \rho v_2 + u \qquad (14)\\
y_2^* &= x_2'\beta_2 + v_2 \qquad (15)
\end{aligned}
$$

with $u|x_1, v_2 \sim N(0, \omega_{11.2}^2)$ where $\omega_{11.2}^2 = \omega_1^2 - \omega_{12}/\omega_2^2$.

The log-likelihood for the latent variables ℓ^* can be split into conditional ℓ_{12}^* and marginal ℓ_2^* log-likelihoods so that:

$$
\ell^*(y_1^*, y_2^*; \theta) = \ell_{12}^*(y_1^*|y_2^*; \theta) + \ell_2^*(y_2^*; \theta_2) \qquad (16)
$$

where $\theta = (\theta_1', \theta_2')'$, $\theta_1 = (\beta_1', \rho)$ and $\theta_2 = (\beta_2', \omega_2^2)$. For an iid sample (y_i, x_i), the observational rules for y_1^* and y_2^* are assumed to be independent from the parameters.

The following are the key results on the likelihood function (Gourieroux, Monfort, Renault, and Trognon, 1984) crucial to the presentation of the test statistic for a model subject to an arbitrary censoring scheme. The score and Hessian matrix elements for observables can be derived quite easily from the score and Hessian matrix for the unobservables as follows[8]:

$$
\frac{\partial \ell}{\partial \theta} = E\left[\left.\frac{\partial \ell^*}{\partial \theta}\right| y\right] \qquad (17)
$$

[7]Smith's original paper is more general than what presented here, as the simultaneous-equations case is considered; however, this is not necessary here so we may skip the resulting complications.

[8]A rigorous proof is given in Gourieroux, Monfort, Renault, and Trognon (1984).

and the Hessian matrix as:

$$\frac{\partial^2 \ell}{\partial\theta\partial\theta'} = E\left[\frac{\partial^2 \ell^*}{\partial\theta\partial\theta'}\middle| y\right] + V\left[\frac{\partial \ell^*}{\partial\theta}\middle| y\right] \tag{18}$$

These quantities can be shown to be functions of the *Generalised Error Product of Order (r,s)* GEP(r, s), introduced by Smith, defined as:

$$\overline{\varepsilon^r \xi^s} = E(\varepsilon^r \xi^s | y) - E(\varepsilon^r \xi^s) \tag{19}$$

where $\varepsilon = u/\omega_{11.2}$ and $\xi = v_2/\omega_2$. $E(\varepsilon^r \xi^s|y)$ is the expectation conditional on the censoring scheme, that is the relevant region of integration. In (Smith, 1985, section 1) some examples are given. The sample counterpart of the GEP(r, s) is the *Generalised Residual Product of Order (r,s)*, GRP(r, s), which is the GEP(r, s) evaluated at $\hat{\theta}_{ML}$ the ML estimator of model (14)-(15).

The Information Matrix test statistic is based on the moment conditions

$$C_i = \text{vech}\left[\frac{\partial^2 \ell_i}{\partial\theta\partial\theta'} + G_i G_i'\right].$$

evaluated at $\theta = \hat{\theta}_{ML}$. The contributions to the Hessian matrix and to the outer product of the gradient are derived using (17) and (18) and are linear functions of the GRP(r, s), $r + s \leq 4$. Expressions for the model (14)-(15) are given in (Smith, 1985, Appendix 1 and 2), for both unobservables and observables.

B Rank Analysis

B.1 Rank Analysis for the Bivariate Probit Model

Let us start from the extreme case in which in each equation the only regressor is a constant; then,

$$-x_{1i}\beta_1 = a_i = \bar{a} \qquad -x_{2i}\beta_2 = b_i = \bar{b} \qquad P_i = \bar{P}$$

are also constant across observations with differences depending only on the observational rule. Therefore the three score elements G_i are also constant across observations

$$G_i^{a_i} = \bar{S}^{\bar{a}} \qquad G_i^{b_i} = \bar{S}^{\bar{b}} \qquad G_i^{\alpha} = \bar{S}^{\alpha}$$

This makes every moment condition C_i a linear combination of the score elements (compare Tables 1 and 15), which means that all these conditions

Table 15: Moment Conditions for the Bivariate Probit Model with only two constant terms as regressors

	β_1	β_2	α
β_1	$-\bar{a}\bar{S}^{\bar{a}} - c_\alpha s_\alpha \bar{S}^\alpha$	$c_\alpha^2 \bar{S}^\alpha$	$-u_{\bar{a},\bar{b}} c_\alpha \bar{S}^\alpha$
β_2		$-\bar{b}\bar{S}^{\bar{b}} - c_\alpha s_\alpha \bar{S}^\alpha$	$-u_{\bar{b},\bar{a}} c_\alpha \bar{S}^\alpha$
α			$\bar{S}^\alpha (u_{\bar{a},\bar{b}} u_{\bar{b},\bar{a}} - t_\alpha)$

are collinear to the score matrix and, as a consequence, do not contribute to the rank of M, as defined in Section 2.1.

Let us turn to the opposite extreme case, with non-overlapping sets of regressors[9]. It is possible to prove that the last moment condition, $C_i^{\alpha,\alpha}$ is always collinear to the rest of the columns of M, even though no suspicious redundancy is apparent.

Dropping the i index for clarity, consider $x_{1r} \neq x_{2s}$ for every $r = 1, ..., k_1$ and $s = 1, ..., k_2$; then

$$a = \sum_{r=1}^{k_1} x_{1r}\beta_{1r} \qquad b = \sum_{s=1}^{k_2} x_{2s}\beta_{2s}$$

The generic condition associated with cross-derivatives of β_{1r} and β_{2s} (see also table (1)), may be written as:

$$C^{\beta_{1r},\beta_{2s}} = c_\alpha^2 S^\alpha x_{1r} x_{2s} \tag{20}$$

Now write the moment condition associated with the cross-derivative of β_{1t} and α as:

$$C^{\beta_{1t},\alpha} = -\left[c_\alpha\left(\sum_{r=1}^{k_1} x_{1r}\beta_{1r}\right)x_{1t} - s_\alpha\left(\sum_{s=1}^{k_2} x_{2s}\beta_{2s}\right)x_{1t}\right]c_\alpha S^\alpha$$

By using (20), the previous expression becomes

$$C^{\beta_{1t},\alpha} = -c_\alpha^2 S^\alpha \sum_{r=1}^{k_1} x_{1t} x_{1r}\beta_{1r} + t_\alpha \sum_{s=1}^{k_2} C^{\beta_{1t},\beta_{2s}}\beta_{2s} \tag{21}$$

[9]More formally: we are assuming that the space spanned by the two sets of regressors have no elements in common. Note that this excludes the presence of constant terms in both equations.

By symmetry, the moment condition associated with the cross-derivative of β_{2m} and α can be written as

$$C^{\beta_{2m},\alpha} = -c_\alpha^2 S^\alpha \sum_{s=1}^{k_2} x_{2m} x_{2s} \beta_{2s} + t_\alpha \sum_{r=1}^{k_1} C^{\beta_{1r},\beta_{2m}} \beta_{1r} \tag{22}$$

Let us now rewrite $C^{\alpha,\alpha}$ as follows:

$$C^{\alpha,\alpha} = S^\alpha [c_\alpha^2 ab + s_\alpha^2 ab - s_\alpha c_\alpha a^2 - s_\alpha c_\alpha b^2 - t_\alpha] =$$

$$= c_\alpha^2 S^\alpha \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} x_{1r} x_{2s} \beta_{1r} \beta_{2s} + s_\alpha^2 S^\alpha \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} x_{1r} x_{2s} \beta_{1r} \beta_{2s} +$$

$$-s_\alpha c_\alpha S^\alpha \sum_{r=1}^{k_1} \sum_{t=1}^{k_1} x_{1r} x_{1t} \beta_{1r} \beta_{1t} - s_\alpha c_\alpha S^\alpha \sum_{s=1}^{k_2} \sum_{m=1}^{k_2} x_{2s} x_{2m} \beta_{2s} \beta_{2m} - S^\alpha t_\alpha.$$

Note that

$$c_\alpha^2 S^\alpha \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} x_{1r} x_{2s} \beta_{1r} \beta_{2s} = \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} C^{\beta_{1r}\beta_{2s}} \beta_{1r} \beta_{2s} \tag{23}$$

$$s_\alpha^2 S^\alpha \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} x_{1r} x_{2s} \beta_{1r} \beta_{2s} = t_\alpha^2 \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} C^{\beta_{1r}\beta_{2s}} \beta_{1r} \beta_{2s}. \tag{24}$$

and that by multiplying (21) by $\beta_{1t} t_\alpha$ one gets

$$-s_\alpha c_\alpha S^\alpha \sum_{r=1}^{k_1} \sum_{t=1}^{k_1} x_{1r} x_{1t} \beta_{1r} \beta_{1t} = t_\alpha \sum_{t=1}^{k_1} C^{\beta_{1t},\alpha} \beta_{1t} - t_\alpha^2 \sum_{t=1}^{k_1} \sum_{s=1}^{k_2} C^{\beta_{1t},\beta_{2s}} \beta_{1t}, \beta_{2s}; \tag{25}$$

similarly, (22) may be multiplied by $\beta_{2m} t_\alpha$ to obtain

$$-s_\alpha c_\alpha S^\alpha \sum_{s=1}^{k_2} \sum_{m=1}^{k_2} x_{2s} x_{2m} \beta_{2s} \beta_{2m} = t_\alpha \sum_{m=1}^{k_2} C^{\beta_{2m},\alpha} \beta_{2m} - t_\alpha^2 \sum_{r=1}^{k_1} \sum_{m=1}^{k_2} C^{\beta_{1r},\beta_{2m}} \beta_{1r}, \beta_{2m}. \tag{26}$$

Finally, after rearranging (23), (24), (25) and (26), we can rewrite $C^{\alpha,\alpha}$ as

$$C^{\alpha,\alpha} = (1 - t_\alpha^2) \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} C^{\beta_{1r},\beta_{2s}} \beta_{1r} \beta_{2s} +$$

$$t_\alpha \sum_{r=1}^{k_1} C^{\beta_{1r},\alpha} \beta_{1r} + t_\alpha \sum_{s=1}^{k_2} C^{\beta_{2s},\alpha} \beta_{2s} - t_\alpha S^\alpha,$$

that is, a linear combination of elements of other columns of M_i.

Different combinations of constant and duplicated regressors across equations lead to intermediate cases. We are particularly interested in studying the case (which often occurs in practice) in which we have the same set of regressors for both equations including constant terms. Other than the $C^{\alpha,\alpha}$ element, we now prove that in this particular case other moment conditions are always collinear in the OPG regression. Moreover, in this special case an explicit formula to determine *a priori* the rank of M can be obtained.

Consider $x_1 = x_2 = x$ and $k_1 = k_2 = q$ so (again, the i index is dropped) $x' = (1, x_2, \ldots, x_q)$ and

$$a = \sum_{r=1}^{q} x_r \beta_{1r} \qquad b = \sum_{r=1}^{q} x_r \beta_{2r}.$$

Similarly, G^{β_j} has q elements

$$[G^{\beta_{j1}}, G^{\beta_{j2}}, \ldots, G^{\beta_{jq}}]$$

for $j = 1, 2$, such that

$$G^{\beta_{1r}} = S^a x_r \qquad G^{\beta_{2r}} = S^b x_r$$

for $r = 1, ..., q$ (see also section (2)).

For a start, the three moment conditions associated to the two constant terms get dropped as

$$C^{\beta_{11},\beta_{11}} = -\Big[\Big(\sum_{r=1}^{q} x_r \beta_{1r}\Big) S^a + s_\alpha c_\alpha S^\alpha\Big] = -\Big[\sum_{r=1}^{q} G^{\beta_{1r}} \beta_{1r} + s_\alpha c_\alpha S^\alpha\Big]$$

$$C^{\beta_{21},\beta_{21}} = -\Big[\Big(\sum_{r=1}^{q} x_r \beta_{2r}\Big) S^b + s_\alpha c_\alpha S^\alpha\Big] = -\Big[\sum_{r=1}^{q} G^{\beta_{2r}} \beta_{2r} + s_\alpha c_\alpha S^\alpha\Big]$$

Consider now the q^2 elements associated with cross derivatives of β_{1r}, β_{2s}

$$C^{\beta_{1r},\beta_{2s}} = c_\alpha^2 S^\alpha x_r x_s$$

with $s = 1, ..., q$. Since the sets of regressors are the same, the number of elements dropped due to collinearity will be $q^2 - q(q+1)/2$ plus the condition associated with the cross derivative of the constant terms

$$C^{\beta_{11},\beta_{21}} = c_\alpha^2 S^\alpha.$$

The are also $2q$ elements, collinear to other comlumns of M, associated with the cross derivatives of regressors with α since

$$C^{\beta_{1t},\alpha} = -\Big[c_\alpha\Big(\sum_{r=1}^{q} x_r \beta_{1r}\Big) - s_\alpha\Big(\sum_{r=1}^{q} x_r \beta_{2r}\Big)\Big] c_\alpha S^\alpha x_t =$$

$$-c_\alpha^2 S^\alpha \sum_{r=1}^{q} x_t x_r \beta_{1r} + s_\alpha c_\alpha S^\alpha \sum_{r=1}^{q} x_t x_r \beta_{2r} =$$

$$-\sum_{r=1}^{q} C^{\beta_{1r},\beta_{2t}} \beta_{1t} + t_\alpha \sum_{r=1}^{q} C^{\beta_{1r},\beta_{2t}} \beta_{2r}$$

and as well

$$C^{\beta_{2t},\alpha} = -\Big[c_\alpha\Big(\sum_{r=1}^{q} x_r\beta_{2r}\Big) - s_\alpha\Big(\sum_{r=1}^{q} x_r\beta_{1r}\Big)\Big] c_\alpha S^\alpha x_t =$$

$$-\sum_{r=1}^{q} C^{\beta_{1r},\beta_{2t}} \beta_{2r} + t_\alpha \sum_{r=1}^{q} C^{\beta_{1r},\beta_{2t}} \beta_{1r}$$

Finally, as shown earlier in this section, $C^{\alpha,\alpha}$ is always a linear combination of other comlumns of M. So, in this setup, the number of degrees of freedom amounts to

$$df = \frac{k(k+1)}{2} - 2 - q^2 + \frac{q(q+1)}{2} - 1 - 2q - 1$$

and since $k = 2q + 1$ we have

$$df = 3\Big(\frac{q(q+1)}{2} - 1\Big).$$

B.2 Rank Analysis for the Heckman Selection Model

In the the case of the Heckman selection model we are only able to prove that the upper bound for df is also equal $k(k+1)/2-1$ since the last moment condition $C_i^{\alpha,\alpha}$ is always dropped in the OPG regression as in the bivariate probit model. This happens when different sets of regressors without constant terms for the two equations are considered. The lower bound is more difficult to determine, and will be derived only for a specific setup as an example.

Let us now consider two sets of completely different regressors x_{is} with $s = 1, ..., m$ and w_{ir} with $r = 1, ..., h$ without constant terms. All the moment conditions that are going to be considered are non-zero only for uncensored observations, so the d_i index will be dropped to simplify the notation (see also table (2)). $C_i^{\alpha,\alpha}$ can be written as a linear combination of other columns of matrix M. For this purpose, the following expressions are devloped. Consider first the generic r-condition $C_i^{\gamma_r,\sigma}$ also as a function of u_i and b_i:

$$C_i^{\gamma_r,\sigma} = \frac{1}{\sigma}\mu_i \left[s_\alpha c_\alpha^2 u_i b_i + s_\alpha^2 c_\alpha u_i^2 + c_\alpha u_i^2 - c_\alpha\right] w_{ir}$$

We now multiply $C_i^{\gamma_r,\sigma}$ by $\sigma\gamma_r$ and then sum across the r moment conditions obtaining

$$\sigma \sum_{r=1}^{h} \gamma_r C_i^{\gamma_r,\sigma} = \mu_i \left[s_\alpha c_\alpha^2 u_i b_i + s_\alpha^2 c_\alpha u_i^2 + c_\alpha u_i^2 - c_\alpha \right] \sum_{r=1}^{h} \gamma_r w_{ir}$$

which gives

$$\sigma \sum_{r=1}^{h} \gamma_r C_i^{\gamma_r,\sigma} = \mu_i \left[s_\alpha c_\alpha^2 u_i b_i^2 + c_\alpha^3 u_i^2 b_i - c_\alpha b_i \right] \tag{27}$$

since $\sum_{r=1}^{h} \gamma_r w_{ir} = b_i$ and $c_\alpha s_\alpha^2 + c_\alpha = c_\alpha^3$. We will later need also

$$t_\alpha^2 \sigma \sum_{r=1}^{h} \gamma_r C_i^{\gamma_r,\sigma} = \mu_i \left[s_\alpha^3 u_i b_i^2 + s_\alpha^2 c_\alpha u_i^2 b_i - \frac{s_\alpha^2}{c_\alpha} b_i \right] \tag{28}$$

Similar transformations applied to $C_i^{\gamma_r,\alpha}$ yield

$$t_\alpha \sum_{r=1}^{h} \gamma_r C_i^{\gamma_r,\alpha} = \mu_i \left[-s_\alpha^2 c_\alpha b_i^3 - s_\alpha c_\alpha^2 u_i b_i^2 - s_\alpha^3 u_i b_i^2 - s_\alpha^2 c_\alpha u_i^2 b_i + \frac{s_\alpha^2}{c_\alpha} b_i \right] \tag{29}$$

Let us now write $C_i^{\sigma,\alpha}$ as a function of u_i and b_i. We get

$$\sigma t_\alpha C_i^{\sigma,\alpha} = \mu_i \left[-2 s_\alpha u_i - \frac{s_\alpha^2}{c_\alpha} b_i + s_\alpha c_\alpha^2 u_i^3 + s_\alpha^3 u_i b_i^2 + 2 s_\alpha^2 c_\alpha u_i^2 b_i \right] \tag{30}$$

and, finally, we need

$$t_\alpha G_i^\alpha = \mu_i \left[\frac{s_\alpha^2}{c_\alpha} b_i + s_\alpha u_i \right] \tag{31}$$

Since $C_i^{\alpha,\alpha}$, written as a function of u_i and b_i is

$$C_i^{\alpha,\alpha} = \mu_i \left[a_i (1 - c_i^2) \right] =$$

$$\mu_i \left[-s_\alpha c_\alpha^2 u_i^3 - s_\alpha^2 c_\alpha b_i^3 - c_\alpha^3 u_i^2 b_i - s_\alpha^3 u_i b_i^2 - 2 s_\alpha^2 c_\alpha u_i^2 b_i - 2 s_\alpha c_\alpha^2 u_i b_i^2 + c_\alpha b_i + s_\alpha u_i \right] \tag{32}$$

it can now be expressed as a linear combination of (27), (28), (29), (30) and (31)

$$C_i^{\alpha,\alpha} = -\sigma(1 - t_\alpha^2) \sum_{r=1}^{h} \gamma_r C_i^{\gamma_r,\sigma} + t_\alpha \sum_{r=1}^{h} \gamma_r C_i^{\gamma_r,\alpha} - \sigma t_\alpha C_i^{\sigma,\alpha} - t_\alpha G_i^\alpha$$

While it makes sense to consider the case of the same sets of regressors for the rank analysis in the bivariate probit model and therefore to choose the limiting case of only two constants to study *df*'s lower bound, the choice of the case study for the Heckman selection model needs further discussion. First of all, it is not possible to consider only two constants since the model would not be identified. Secondly, it is quite common to see applications with two at least slightly different sets of regressors. Therefore we believe that the simplest form of a reasonable setup is one containing a constant term and a continuous regressor w in the selection equation, and only a constant term in the main equation. The vector of parameters of the model just described is $(\beta_0, \gamma_0, \gamma_1, \sigma, \alpha)$. Six of the fifteen moment conditions are dropped in the OPG regression. Naturally

$$C_i^{\beta_0,\beta_0} = d_i \frac{1}{\sigma}\left[G_i^{\sigma} - \frac{s_\alpha c_\alpha}{\sigma} G_i^{\alpha}\right]$$

$$C_i^{\beta_0,\gamma_0} = d_i \frac{c_\alpha^2}{\sigma} G_i^{\alpha}$$

are dropped as they are linear combinations of the score elements. Also $C_i^{\gamma_0,\gamma_0}$ is a linear combination of $G_i^{\gamma_0}, G_i^{\gamma_1}, G_i^{\alpha}$. First notice that in this specific setup

$$b_i = \gamma_0 + w_i\gamma_1$$

and

$$a_i = c_\alpha\gamma_0 + c_\alpha w_i\gamma_1 + s_\alpha u_i; \quad c_i = s_\alpha\gamma_0 + s_\alpha w_i\gamma_1 + c_\alpha u_i$$

then write

$$C_i^{\gamma_0,\gamma_0} = d_i\mu_i\left[-c_\alpha^3\gamma_0 - c_\alpha^3 w_i\gamma_1 - s_\alpha c_\alpha^2 u_i\right] - (1-d_i)\mu_i\left[\gamma_0 + w_i\gamma_1\right]$$

Given the following transformations, only for uncensored observations,

$$-c_\alpha^2\gamma_0 G_i^{\gamma_0} = -c_\alpha^3\gamma_0\mu_i; \quad -c_\alpha^2\gamma_1 G_i^{\gamma_1} = -c_\alpha^3 w_i\gamma_1\mu_i$$

and

$$-s_\alpha c_\alpha G_i^{\alpha} = -\mu_i\left[-c_\alpha^3\gamma_0 - c_\alpha^3 w_i\gamma_1 - c_\alpha^2 s_\alpha u_i\right],$$

$C_i^{\gamma_0,\gamma_0}$ can be written as

$$C_i^{\gamma_0,\gamma_0} = -d_i\left[c_\alpha s_\alpha G_i^{\alpha} + \gamma_0 G_i^{\gamma_0} + \gamma_1 G_i^{\gamma_1}\right] - (1-d_i)\left[\gamma_0 G_i^{\gamma_0} + \gamma_1 G_i^{\gamma_1}\right].$$

$C_i^{\gamma_0,\sigma}$ is also a linear combination of G_i^{α} and of the moment conditions $C_i^{\beta_0,\gamma_1}$ and $C_i^{\beta_0,\alpha}$. Rearranging algebraically $C_i^{\gamma_0,\sigma}$ and $C_i^{\beta_0,\alpha}$ we get

$$C_i^{\gamma_0,\sigma} = C_i^{\beta_0,\alpha} - \mu_i \frac{1}{\sigma}\left[s_\alpha^2 c_\alpha b_i + s_\alpha c_\alpha^2 u_i\right](\gamma_0 + \gamma_1 w_i). \tag{33}$$

Applying the following transformations

$$\frac{s_\alpha c_\alpha \gamma_0}{\sigma} G_i^\alpha = \frac{\mu_i}{\sigma}\left[s_\alpha^2 c_\alpha b_i + s_\alpha c_\alpha^2 u_i\right]\gamma_0$$

$$\frac{s_\alpha}{c_\alpha}\gamma_i C_i^{\beta_0,\gamma_1} = \frac{\mu_i}{\sigma}\left[s_\alpha^2 c_\alpha b_i + s_\alpha c_\alpha^2 u_i\right]\gamma_1 w_i$$

we can rewrite (33) as

$$C_i^{\gamma_0,\sigma} = C_i^{\beta_0,\alpha} - \frac{s_\alpha c_\alpha \gamma_0}{\sigma} G_i^\alpha - \frac{s_\alpha}{c_\alpha}\gamma_1 C_i^{\beta_0,\gamma_1}. \tag{34}$$

The OPG regression also drops

$$C_i^{\gamma_0,\alpha} = -\mu_i(c_\alpha a_i c_i - s_\alpha). \tag{35}$$

Considering

$$-\sigma\frac{c_\alpha}{s_\alpha} C_i^{\beta_0,\alpha} = -\mu_i\left(c_\alpha a_i c_i + \frac{c_\alpha}{s_\alpha} u_i c_i - \frac{c_\alpha^2}{s_\alpha}\right)$$

and

$$\frac{\sigma}{s_\alpha c_\alpha} C_i^{\gamma_0,\sigma} = \mu_i\left(c_\alpha b_i u_i + \frac{c_\alpha^2}{s_\alpha} u_i^2 - \frac{1}{s_\alpha}\right)$$

we can rewrite (35) as

$$C_i^{\gamma_0,\alpha} = -\sigma\frac{c_\alpha}{s_\alpha} C_i^{\beta_0,\alpha} + \frac{\sigma}{s_\alpha c_\alpha} C_i^{\gamma_0,\sigma} \tag{36}$$

and substituting (34) in (36) we finally get

$$C_i^{\gamma_0,\alpha} = -\sigma\frac{s_\alpha}{c_\alpha} C_i^{\beta_0,\alpha} - \gamma_0 G_i^\alpha - \frac{\sigma}{c_\alpha^2}\gamma_1 C^{\beta_0,\gamma_1}$$

Finally the last column to be dropped is $C_i^{\alpha,\alpha}$ as we discussed earlier. So in this simple setup the number of moment conditions are $k(k+1)/2 = 15$ of which only 9 are kept for testing.