A Test for Bivariate Normality in Commonly Used Microeconometric Models

Riccardo (Jack) Lucchetti Claudia Pigini

Abstract

In this paper, we propose a test for bivariate normality in imperfectly observed models, based on the Information Matrix Test for bivariate censored models with bootstrap critical values. In order to evaluate its properties, we run a comprehensive Monte Carlo experiment, in which we use the bivariate probit model and Heckman sample selection model as examples.

We find that, while asymptotic critical values can be seriously misleading, the usage of bootstrap critical values results in a test that has excellent size and power properties even in small samples. Since this procedure is relatively inexpensive from a computational viewpoint and is easy to generalise to models with arbitrary censoring schemes, we recommend it as an important and valuable testing tool.

1 Introduction and motivation

Bivariate microeconometric models with censored, truncated or categorical response variables are widely employed in empirical applications. Examples range from familiar models like the bivariate (possibly ordered) probit model to more specialised tools.

The assumption of normality is almost always the mainstream choice. In some cases, there are alternatives: for example, a great deal of work has been done on non-parametric estimation of the sample selection model (see for example Das, Newey, and Vella, 2003). Maximum likelihood under normality, however, does have several advantages over alternative methods and therefore is often the practitioner’s choice; in some cases, such as for example the double-hurdle model (see Cragg, 1971) or the probit model with sample selection (see Van de Ven and Van Praag, 1981), it is the only available choice in pre-packaged software.

The distributional assumption, however, is seldom tested after estimating such models. This is rather unfortunate, since distributional misspecification is well known to make the maximum likelihood estimator inconsistent (Robinson, 1982; Smith, 1989), so that a QMLE approach is not applicable in this context. The reason may partly lie with the fact that, contrary to cases in which data are not censored, an established and computationally convenient procedure for normality testing does not exist.

A possible strategy to test the bivariate normality assumption in this kind of models has been put forward by Smith (1985), who derived the Information Matrix (IM henceforth) test statistic for misspecification in models subject to
imperfect observability. The generalisation of the IM test statistic to models with arbitrary censoring schemes is straightforward: it is derived from the Information Matrix equality, a general principle applicable to all models estimated by maximum likelihood; in addition, it does not require the specification of an alternative hypothesis. Moreover, it is of rather simple computation, especially in its Outer Product of the Gradient (OPG) form, which only requires expressions for the score and Hessian matrix.

IM tests are, however, generally known to deliver a poor finite sample performance: especially the OPG form of the IM test has been shown to reject the null hypothesis when true much too often in several other contexts. There is, however, no evidence yet about the finite-sample properties of the IM test statistic in bivariate limited dependent variable models.

However, the practice of bootstrapping critical values, as suggested by Horowitz (1994), is also known to improve the finite-sample performance of IM tests and can be easily adapted to bivariate limited dependent variable models. The aim of this paper is therefore to explore the size and power properties of the IM test both as originally proposed by Smith (1985) and with bootstrapped critical values to correct the size bias. We run a comprehensive Monte Carlo experiment for two very common microeconometric models: the sample selection model (Heckman, 1974) and the bivariate probit model.

Our experiments show two main results. The IM test as originally proposed by Smith has extremely disappointing finite-sample properties: the empirical size is severely biased even for samples which one would consider perfectly adequate for asymptotics. On the contrary, the test statistic has excellent size and power properties once critical values are bootstrapped. Interestingly, a trade-off seems to exist between the power of the test and the amount of information lost to censoring.

The paper is organised as follows: section 2 reviews some of the available test statistics for univariate and multivariate normality, with particular attention to the IM test and the approach proposed by Smith (1985); section 3 describes the test statistic and briefly illustrates how to bootstrap critical values; section 4 describes in detail the derivation of the test statistic for the two models and the experiment setup; section 5 contains the results of the Monte Carlo experiments for size and power; section 7 concludes.

2 Literature review

Testing bivariate and multivariate normality has become common practice in models with continuous dependent variables and several options are available. Tests based on measures of multivariate skewness ($\sqrt{b_1}$) and kurtosis ($b_2$) have been first brought forward by Mardia (1970), where $b_1$ and $b_2$ are $\chi^2$ and normally distributed, respectively. 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 cannot be applied in models subject to truncation or censoring, where normality of error terms is often the standard distributional
assumption.

In univariate limited dependent variable models, normality tests were proposed first by Bera, Jarque, and Lee (1984) and then by Chesher and Irish (1987): given a latent model

\[ y^*_i = x_i' \beta + \epsilon_i, \]

the test is based on the conditional moments \( E(\epsilon^3_i|y_i, x_i) \) and \( E(\epsilon^4_i|y_i, x_i) \), which can be calculated as functions of the estimates \( \hat{\beta} \). This has become a well-established procedure and is interpreted as a conditional moment test for regressions with grouped data, probit and Tobit models.

There have been, however, only few attempts to test bivariate normality, even though this is a key assumption for some widely popular models, such as the sample selection model (Heckman, 1974) and the bivariate (possibly ordered) probit model. A semi-parametric approach is followed by van der Klaauw and Koning (2003) who derived an LR test for bivariate normality in the sample selection model from the flexible likelihood function of Gallant and Nychka (1987); Lee (1984) proposes a LM test statistic based on the Edgeworth series expansion that serves as a suitable alternative distribution; its finite sample properties are analysed in Montes-Rojas (2011) and Murphy (2007) for the sample selection and bivariate probit model respectively.

The tests above all hinge on the specification of an alternative hypothesis which nests normality as a special case. A somewhat different approach, instead, was proposed by Richard Smith, who proposed to apply the Information Matrix test (White, 1982) to bivariate (Smith, 1985) and multivariate (Smith, 1987) limited dependent variable models. The use of the IM test statistic in this framework 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 IM 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 misspecification.

Finally, as emphasised in Chesher (1983) and Lancaster (1984), an IM test is trivial to implement in software by using the Outer Product of the Gradient (OPG) regression\(^2\); for most popular models, expressions for the score and the Hessian are usually available or, in other cases, their numerical computation is straightforward.

---

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

\(^2\)See for example Davidson and MacKinnon (2001).
Despite these qualities, however, IM tests are generally known to perform rather disappointingly in finite samples when applied to the linear and univariate limited dependent variable models: especially the OPG form of the IM test exhibits serious empirical size bias. Monte Carlo studies carried out by Taylor (1987) and by Kennan and Neumann (1988) reported the extremely poor approximation of the asymptotic $\chi^2$ distribution to the finite sample distribution of the IM test statistic. Orme (1990) also argues that the Chesher-Lancaster OPG version suffers from large 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, and shows by Monte Carlo simulations that the size bias can be considerably reduced by using a version subject to some parameter 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 $\chi^2$ 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 to limited dependent variable models.

In our experiment, we explore the usage of bootstrapped critical values instead of asymptotic ones to correct the IM test size bias, as originally suggested in Horowitz (1994). The practice of bootstrapping critical values has been known to dramatically improve the finite sample properties of asymptotic tests in a wide spectrum of situations; Horowitz (1994) shows that this is also the case in Tobit and probit models\footnote{For the latter, this result is also confirmed by Davidson and MacKinnon (1998)}. Differently from the double length regression, this method is easily generalisable to bivariate limited dependent variable models. In addition, it does not require cumbersome analytical derivations and recent advances in computing power commonly available make it practical and relatively inexpensive to implement in software.

3 Test for Bivariate Normality

3.1 Information Matrix and Conditional Moment Tests

As is well known, the IM test, introduced by White (1982), is based on the Information Matrix equality. Assume we have a sample of iid observations with log-likelihood for each observation $\ell_i(\theta)$, where $\theta$ is a $k$-vector of parameters. 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. The Information Matrix test statistic is therefore a test for $E(C_i) = 0$, where

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

(1)
and \( G_i \equiv \frac{\partial \ell_i}{\partial \theta} \); all quantities are evaluated at \( \theta = \hat{\theta}_{ML} \).

For the models we are considering here, the score vector and the Hessian matrix can be computed analytically from the log-likelihood for observables; we take advantage of this possibility and use analytical derivatives throughout.\(^4\) However, this may become a problem in models with imperfect observability, that is models such as

\[
\begin{align*}
y^*_1 &= x'_1 \beta_1 + v_1 \\
y^*_2 &= x'_2 \beta_2 + v_2
\end{align*}
\]

in which one or both dependent variables are subject to some form of censoring. For example, in the bivariate probit model we do not observe the \( y^*_j \) variables, but only their sign. Fortunately, in the general context of bivariate models with imperfect observability, the setup proposed by Smith (1985) (see section A in the Appendix) can be used to derive moment conditions for any censoring scheme.

In this paper, however, we choose the derivative-based approach as it leads to analytical expressions that lend themselves to a much easier translation into numerically efficient code\(^5\).

The Information Matrix test can be computed by means of an OPG regression (see Davidson and MacKinnon (2001)): the test statistic equals \( nR^2 \) of the regression of an \( n \)-vector of ones on a matrix \( M \), with typical row \( M'_i = [G'_i, C'_i] \).

Under the null, the test statistic has an asymptotic \( \chi^2 \) distribution with degrees of freedom given by rank\((M) - k\).

It is important to note that, in general, \( M \) may not be of full column rank. One case 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 \( M \) are linear combinations of one another and, consequently, must be dropped from the OPG regression.

In the case of the bivariate probit model it can be proved that the degrees of freedom are bounded between zero and \( k(k + 1)/2 - 1 \). The upper bound for \( df \) the case of the sample selection model can also be proved 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 4.3 is given in section B, in the Appendix.

### 3.2 Bootstrapped Critical Values

The empirical distribution of the IM test statistic is known to be inadequately approximated by its asymptotic distribution. Considerable effort has been put into finding alternative forms of the test statistic or corrections to its size bias. We will follow the approach recommended by Horowitz (1994) of analysing the IM test finite-sample properties using bootstrapped critical values instead of asymptotic ones.

\(^4\)Numerical differentiation is clearly a possible, but more demanding computationally, alternative.

\(^5\)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.
In general, bootstrap methods are frequently advocated when exact tests are not available. Horowitz’s suggestion of using bootstrapped critical values hinges upon the difficulty of determining in which cases the IM test statistic is pivotal. Since in these situations bootstrapping corresponds to Monte Carlo simulation, we are not required to know whether the IM test is pivotal and we may bootstrap the critical values in any case.

In practice, bootstrapping critical values is straightforward to implement. In this section, we only lay out the basic computational steps. Of course, the theory and practice of bootstrap tests have been extensively discussed in more dedicated contributions: see for example Davidson and MacKinnon (1999b) and Davidson and MacKinnon (2006) for results on size and power of bootstrap tests, Davidson and MacKinnon (2000) for the optimal number of bootstrap replications, Davidson and MacKinnon (1999a) and Davidson and MacKinnon (2007) on optimisation algorithms for bootstrap testing. For a more detailed and extensive discussion on the application of this technique to the IM test, see Horowitz (1994).

A simple illustration can be given as:

1. estimate the model under examination using the original DGP; call $\hat{\theta}$ the parameter estimates and compute the test statistic $T$;
2. simulate a number $B$ of artificial datasets, using $\hat{\theta}$ as DGP parameters;
3. for each of these, compute the test on the simulated data $T_b$

By using the $B$ realisations of the test statistic $T_b$ for $b = 1, ..., B$, the bootstrapped critical values may be computed as the quantiles of this distribution; of particular interest in analysing finite-sample properties could be $T_b^{0.90}$, $T_b^{0.95}$ and $T_b^{0.99}$. Embedded in a Monte Carlo experiment, this procedure allows us to compare empirical size and power based on asymptotic and bootstrapped critical values: for a given nominal size $\alpha$, the empirical percentage of rejections are

$$\hat{\alpha} = (1/J) \sum_{j=1}^{J} I(T_j > c_{1-\alpha})$$ (2)

where $c_{1-\alpha}$ is the asymptotic critical value and

$$\hat{\alpha}^b = (1/J) \sum_{j=1}^{J} I(T_j > T_{b1-\alpha,j})$$ (3)

where $T_{b1-\alpha,j}$ is the bootstrapped critical value and $J$ is the number of Monte Carlo replications.

4 Experiment Design

In this section, we lay out all the necessary ingredients for our experiment. In the next two subsections, we will compute the moment conditions matrix explicitly for the bivariate probit and the sample selection models. Although these models are well known, we will present them in detail so to establish our notation clearly. Subsection 4.3 illustrates the setup of the Monte Carlo simulation.
4.1 The Sample Selection Model

The latent variable model is defined as

\[ y_i^* = x_i' \beta + v_{1i} \]
\[ d_i^* = w_i' \gamma + v_{2i} \]

where \( x_i \) and \( w_i \) are an \( m \)-vector and an \( h \)-vector of exogenous variables, respectively; the error terms \( v_{1i} \) and \( v_{2i} \) are assumed to be jointly normal with zero mean and covariance matrix

\[
V \begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho \sigma \\ \rho \sigma & 1 \end{pmatrix}
\]

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

\[
y_i = \begin{cases} 
  y_i^* & \text{if } d_i^* > 0 \\
  \text{NA} & \text{if } d_i^* \leq 0 
\end{cases}
\]

with \( d_i = I(d_i^* > 0) \). In order to have clear expressions, let us define \( u_i = \frac{y_i - x_i' \beta}{\sigma} \) and \( b_i = w_i' \gamma \). As customary in this kind of setting, we re-parametrise the bivariate normal density via hyperbolic functions so that instead of the correlation coefficient \( \rho \) we will be using \( \psi = \text{atanh}(\rho) \) and the associated quantities \( c_\psi = \cosh(\psi), s_\psi = \sinh(\psi) \) and \( t_\psi = \tanh(\psi) = s_\psi / c_\psi \); thus, the contribution to the log-likelihood for observation \( i \) can be written as

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

where \( a_i = c_\psi b_i + s_\psi w_i \) and \( \Phi \) is the standard normal distribution function. The parameter vector \( \theta' = (\beta', \gamma', \sigma, \psi) \) includes \( k = m + h + 2 \) parameters. By using the generalised residuals

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

the score elements for observation \( i \) can be written as:

\[ G_\beta^i = d_i \left( \frac{u_i - s_\psi \mu_i}{\sigma} \right) x_i' \]
\[ G_\gamma^i = [d_i \mu_i c_\psi - (1 - d_i) \mu_i] w_i' \]
\[ G_\sigma^i = d_i \left[ \mu_i (u_i - s_\psi \mu_i) - 1 \right] \]
\[ G_\psi^i = d_i \mu_i c_i \]

where \( c_i = c_\psi u_i + s_\psi b_i \); the moment conditions \( C_i \) are presented in Table 1.
Table 1: Moment Conditions for the Sample Selection Model

<table>
<thead>
<tr>
<th>Term</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{i}^{\beta,\beta}$</td>
<td>$d_i \frac{1}{\sigma} \left[ G_i^\sigma - \frac{s_{\psi}e_{\psi}}{\sigma} G_i^0 \right] x_i x'_i $</td>
</tr>
<tr>
<td>$C_{i}^{\beta,\gamma}$</td>
<td>$d_i \frac{\epsilon_{\gamma}^2}{\sigma} G_i^\gamma x_i w'_i $</td>
</tr>
<tr>
<td>$C_{i}^{\beta,\sigma}$</td>
<td>$d_i \frac{1}{\sigma} u_i \left[ s_{\psi} \mu_i \left( 1 - s_{\psi} a_i - 2 u_i \right) + u_i^2 - 2 \right] x'_i - \sigma G_i^\beta $</td>
</tr>
<tr>
<td>$C_{i}^{\beta,\psi}$</td>
<td>$d_i \frac{1}{\sigma} \left[ \mu_i (s_{\psi} a_i c_i + u_i c_i - c_{\psi}) \right] x'_i $</td>
</tr>
<tr>
<td>$C_{i}^{\gamma,\gamma}$</td>
<td>$d_i \left[ (c_{\psi} \mu_i - (1 - d_i) \mu_i b_i) w_i w'_i \right] $</td>
</tr>
<tr>
<td>$C_{i}^{\gamma,\sigma}$</td>
<td>$d_i \frac{1}{\sigma} \left[ c_{\psi} \mu_i (s_{\psi} a_i u_i + u_i^2 - 1) \right] w'_i $</td>
</tr>
<tr>
<td>$C_{i}^{\gamma,\psi}$</td>
<td>$d_i \left[ -\mu_i (c_{\psi} a_i c_i - s_{\psi}) \right] $</td>
</tr>
<tr>
<td>$C_{i}^{\sigma,\sigma}$</td>
<td>$d_i \frac{1}{\sigma} \left[ u_i^4 - 5 u_i^2 + 2 + s_{\psi} u_i \mu_i (4 - a_i s_{\psi} u_i - 2 u_i^2) \right] $</td>
</tr>
<tr>
<td>$C_{i}^{\sigma,\psi}$</td>
<td>$d_i \frac{1}{\sigma} \mu_i \left[ c_i u_i (s_{\psi} a_i + u_i) - (c_{\psi} u_i + c_i) \right] $</td>
</tr>
<tr>
<td>$C_{i}^{\psi,\psi}$</td>
<td>$d_i \mu_i a_i (1 - c_i^2) $</td>
</tr>
</tbody>
</table>
4.2 The Bivariate Probit Model

The latent variable model is defined as (dropping the observation index \(i\) for clarity)

\[
y_1^* = x_1' \beta_1 + v_1
\]

\[
y_2^* = x_2' \beta_2 + v_2
\]

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 with unit variances and correlation coefficient \(\rho\). The observable random vector \(y = (y_1, y_2)\) is related to latent random variable \(y^* = (y_1^*, y_2^*)\) via \(y_j = I(y_j^* > 0)\) for \(j = 1,2\) where \(I(\cdot)\) is the indicator function. Let us define \(a_i = x_1' \beta_1\) and \(b_i = x_2' \beta_2\). Then, the contribution to the log-likelihood for observation \(i\) can be written as \(\ell_i = \ln P_i\), where

\[
P_i \equiv y_1 y_2 \Phi_2(a_i, b_i, \psi) + y_1 (1 - y_2) \Phi_2(a_i, -b_i, -\psi) + \Phi_2(-a_i, b_i, -\psi) + (1 - y_1)(1 - y_2) \Phi_2(-a_i, -b_i, \psi)
\]

and \(\Phi_2\) is the bivariate standard normal distribution function. So we have a vector \(\theta' = (\beta_1', \beta_2', \psi)\) of \(k = k_1 + k_2 + 1\) parameters to be estimated by ML. Let us define the function

\[
u_{ba} = c_{\psi} b_i - s_{\psi} a_i
\]

and write the score elements for observation \(i\) as

\[
G_{\beta_1} = \frac{\varphi(c_{\psi} b_i - s_{\psi} a_i)}{P_i} x_1' c_{\psi} S_{\psi} i
\]

\[
G_{\beta_2} = \frac{\varphi(c_{\psi} b_i - s_{\psi} a_i)}{P_i} x_2' c_{\psi} S_{\psi} i
\]

\[
G_{\psi} = \frac{\varphi(c_{\psi} b_i - s_{\psi} a_i)}{P_i c_{\psi}}
\]

The moment conditions \(C_i\), expressed as functions of the score elements, are shown in Table 2.

Table 2: Moment Conditions for the Bivariate Probit Model

<table>
<thead>
<tr>
<th>(\beta_1)</th>
<th>(\beta_2)</th>
<th>(\psi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-a_i S_{\psi} i + c_{\psi} s_{\psi} S_{i} ) (x_1 x_1')</td>
<td>(c_{\psi} S_{\psi} i x_1 x_2')</td>
<td>(-u_{ba} c_{\psi} S_{\psi} i x_1)</td>
</tr>
<tr>
<td>(-b_i S_{\psi} i + c_{\psi} s_{\psi} S_{i} ) (x_2 x_2')</td>
<td>(-u_{ba} c_{\psi} S_{\psi} i x_2)</td>
<td>(S_{\psi} i [u_{ba} u_{ba} - t_{\psi}])</td>
</tr>
</tbody>
</table>

4.3 Test Statistics and Monte Carlo Setup

We run the Monte Carlo simulation for the sample selection and bivariate probit model. For the sample selection model results are reported for two different
degrees of censoring: 25% and 75% of censored observations. The correlation coefficient \( \rho \) between the random error terms in both models is set to \( \rho = 0.50, 0.75 \). For a detailed description of the DGP see tables 5 and 7.

For the power analysis, we use a standardised bivariate \( \chi^2 \) distribution, generated via a Gaussian copula. In practice, we generate the disturbances \( v_1 \) and \( v_2 \) as

\[
\begin{align*}
v_1 &= \frac{F^{-1}_\nu[\Phi(x_1)] - \nu}{\sqrt{2\nu}} \\
v_2 &= \frac{F^{-1}_\nu[\Phi(x_2)] - \nu}{\sqrt{2\nu}}
\end{align*}
\]

where \( F^{-1}_\nu(\cdot) \) is the inverse cumulative distribution function of a \( \chi^2 \) rv with \( \nu \) degrees of freedom and \( (x_1, x_2) \) are a bivariate standard Gaussian rv with the desired correlation coefficient. Clearly, as \( \nu \to \infty \) the departure from normality vanishes. By setting \( \nu \) to different values we are able to study the behaviour of the test statistic when departures from bivariate normality are more or less severe. The values of \( \nu \) that we use are \( \nu = 2, 4, 6, 8, 20 \).

In the size analysis, we run the experiment with 10000 Monte Carlo replications and sample sizes are 256, 1024 and 4096. To analyse power we run instead 1024 Monte Carlo replications with samples of 256 and 1024 observations. In both cases, the number of bootstrap replications is 400. Horowitz (1994) suggests that 100 of bootstrap replications are sufficient; however, when this number is too small a power loss can occur (see Davidson and MacKinnon (2000)).

As briefly explained in section 3.1, 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). This opens the possibility of performing the normality test by focusing on variously defined subsets of the available moment conditions.

One obvious choice is to use all moment conditions, that is that we including all non-redundant columns of the \( M \) matrix (see tables (2) and (1)) in the OPG regression resulting in a general test for misspecification. This leads to the test we call the “all-moments” test, which is computed for both the models we analyse.

Next, we include in the OPG regression those columns of \( M \) that contain third and fourth moment conditions separately\(^7\). These two choices lead to test variants called “third-moments” and “fourth-moments” tests, respectively.\(^8\) In the sample selection model third moment conditions are contained in \( C_i^{\beta,\sigma} \), \( C_i^{\gamma,\sigma} \), \( C_i^{\beta,\psi} \), \( C_i^{\gamma,\psi} \) and fourth moment conditions are in \( C_i^{\sigma,\sigma} \), \( C_i^{\sigma,\psi} \), \( C_i^{\psi,\psi} \), \( C_i^{\beta,\psi} \) and \( C_i^{\gamma,\psi} \). In the bivariate probit model, instead, third moment conditions appear in \( C_i^{\beta_1,\psi} \) and \( C_i^{\beta_2,\psi} \) and the only condition on the fourth moment is \( C_i^{\psi,\psi} \).

Finally, we analyse a variant of the test which considers third and fourth moments jointly. Curiously, this turns out to be unfeasible for the bivariate probit

---

6This was done by setting set the intercept parameter in equation (5) to a suitable value.

7The presence of such moment conditions is not easy to spot inside the expressions given in sections 4.1 and 4.2. However, it becomes obvious when rewriting those expressions using Smith’s GEP formula.

8Labelling these tests as “skewness” and “kurtosis” tests would be tempting, but misleading, since these conditions contain cross-moment conditions which are not readily interpretable in terms of the shape of the joint density.
model if both equations (8) and (9) contain a constant term, due to collinearity in the moment conditions (see section B.2 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. A similar phenomenon also happens for the Heckit model (the condition $C_1^{\psi,\psi}$ becomes collinear — see section B.1 in the Appendix) but the problem is less serious and the test is computable. We call this last test the “third-and-fourth-moments” test.

5 Experiment Results

Table 3: Rejection frequency at nominal size = 5%, $\rho = 0.5$, 1024 observations, sample selection and bivariate probit model

<table>
<thead>
<tr>
<th></th>
<th>Heckit 25% censoring</th>
<th>Heckit 75% censoring</th>
<th>Bivariate probit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic $\chi^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>59.25</td>
<td>67.10</td>
<td>48.75</td>
</tr>
<tr>
<td>$3^{rd}$</td>
<td>16.08</td>
<td>21.83</td>
<td>20.32</td>
</tr>
<tr>
<td>$4^{th}$</td>
<td>20.26</td>
<td>26.02</td>
<td>15.88</td>
</tr>
<tr>
<td>$3^{rd}, 4^{th}$</td>
<td>31.43</td>
<td>46.81</td>
<td></td>
</tr>
<tr>
<td>Bootstrap</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>4.21</td>
<td>4.97</td>
<td>4.75</td>
</tr>
<tr>
<td>$3^{rd}$</td>
<td>4.83</td>
<td>4.87</td>
<td>4.41</td>
</tr>
<tr>
<td>$4^{th}$</td>
<td>4.75</td>
<td>4.71</td>
<td>4.75</td>
</tr>
<tr>
<td>$3^{rd}, 4^{th}$</td>
<td>4.45</td>
<td>5.22</td>
<td></td>
</tr>
</tbody>
</table>

The bold font indicates that the empirical size is within a 95% confidence interval centred on the nominal size.

We ran all combinations of experiments listed in subsection 4.3, which required a sizeable computational effort. The output is likewise quite enormous and would take several dozen pages to be presented (even in graphical form). An example is displayed in Table 3, which shows the performance in terms of size of the four variants of the IM test for our three models of choice, limited to the case $\rho = 0.5$, $n = 1024$ and $\alpha = 0.05$.

The most striking feature that is apparent in Table 3 is that, while the empirical size of the asymptotic version of the test is very far from its nominal level and over-rejections are substantial, the bootstrap variant performs very well, with only a slight tendency to under-reject and, in most cases, the empirical rejection frequency is not significantly different from the nominal size. This is true for all variants of the test considered, although it must be said that the size bias for the asymptotic version differs quite substantially across versions of the test, which implies that if one had to rely on asymptotic critical values, then the choice of which moment conditions to employ would be of critical importance.

These results basically hold unchanged across all the DGPs we used in the Monte Carlo experiment: we got qualitatively similar results for all values of $\rho$, for all values of $\alpha$ and the effect of the sample size on the test statistic was in all cases the expected one (that is, the larger the sample the smaller the size bias).
However, it is not easy to present these results in tabular form: the output for the whole experiment would include 27 tables like Table 3 for the size analysis alone. Repeating the above for the DGP s generated under non-normality for the power analysis would result in 135 more tables. Of course, all output is available upon request, but we deemed it more informative for the reader to concentrate on those variants which appeared to display the best performance in terms of size and power and, therefore, to be of interest to the practitioner.

From now on, therefore, what we call “the” test is the “third-and-fourth-moments” variant for the sample selection model and the “all-moments” variant for the bivariate probit model. By even a cursory analysis of the complete set of results, the choice was quite obvious. However, we also include the results for the experiments on the “all-moments” variant for the sample selection models in Tables 5 and 6 in the appendix.

5.1 Size

Tables 5, 6 and 7 report the empirical size of the IM test statistic:

Apart from marginal differences, there are several result that may be considered consistent across all experiments:

- As expected, size bias declines with sample size; however, while the slight tendency to under-reject that the bootstrap variant exhibits is confined to small samples and quickly vanishes with 1000 observations or more, the situation is markedly different for the test built on asymptotic critical values: in no case reported on the tables the asymptotic version of the test yields a frequency of rejections that was not statistically different from the nominal one. Smaller experiments (not reported here, but available on request) with 16384 observations also gave disappointing results in most cases. We consider it as a very noteworthy result that over 16000 iid observations may not be considered, under some circumstances, a sample ‘large enough’ to trust in asymptotics.

- The quantity of information that gets lost due to censoring affects negatively the size properties of the test. Again, however, this tendency is more pronounced for the asymptotic-based test, while the bootstrap-based version seems largely unaffected.

- There seems to be a slight tendency for the size properties of the tests to worsen as $\rho$ moves away from 0. However, this is by no means general and, again, is much less of a problem for the bootstrap version.

These considerations become rather evident when considering Figures 1–3, which depict empirical size graphically (to enhance readability, curves were smoothed via Bézier splines and the y axis is in log scale).

5.2 Power

As anticipated in section 4.3, we analyse the power of the Information Matrix test statistic by employing the standardised bivariate $\chi^2$ distribution, for several

\[ \chi^2 \]

9Tables 5 and 6 for the sample selection models also include the “all-moment” variant of the test to let the reader compare it with the preferred one.
Figure 1: Size Analysis, Sample Selection Model, 25% of Censored Obs.
Figure 2: Size Analysis, Sample Selection Model, 75% of Censored Obs.
Figure 3: Size Analysis, Bivariate Probit Model

- Rej. freq. at nominal size = 10% (log scale)
- Rej. freq. at nominal size = 5% (log scale)
- Rej. freq. at nominal size = 1% (log scale)
- Sample size
- Asymptotic
- Bootstrap
- \( \rho = 0.0 \)
- \( \rho = 0.5 \)
- \( \rho = 0.75 \)
degrees of freedom, as the true DGP. This choice was dictated by the following consideration: in ordinary circumstances, one may want to explore the power of a normality test by considering alternative distributions with different degrees of asymmetry and/or kurtosis. In our context, however, we felt it more important to focus on asymmetry rather than kurtosis because in most censoring schemes (such as, for example, in the bivariate probit model) what is observed is simply the sign of the disturbance term. Although it is likely that the power of a normality test should be rather low under a symmetric alternative,\(^\text{10}\) it is also likely that the shape of the distribution tails should have little impact on the actual distribution of the ML estimator. Therefore, even if the normality-based ML estimator is technically inconsistent\(^\text{11}\) when the true DGP is symmetric but non-normal, this could be considered as a lesser drawback, in the eyes of a practitioner, than those potentially arising from asymmetry. The \(\chi^2\) distribution thus provides a convenient way to measure departure from normality via the degrees-of-freedom parameter as depicted in Figure 4.

In our analysis of the power of the test, we made several simplifying choices: first, although we did compute the percentage of rejections for the asymptotic version of the IM test, we will only comment on the power of the bootstrap version of the test: it makes little sense, in our eyes, to analyse the power properties of a test that has abysmal size properties such as the asymptotic version (see section 5.1).

Second, to keep the computational effort within reasonable limits we only considered samples of 256 and 1024 observations and we used a smaller number of Monte Carlo simulations (1024). Moreover, it is not necessary, like in the size analysis, to perform a formal test of equality of the percentage of rejections to a pre-specified nominal value so the smaller number of Monte Carlo replications should pose no particular problems.

Third, we decided that presenting our results in tabular form would have

---

\(^{10}\) For example, in a univariate context, Skeels and Vella (1999) find that the Chesher-Irish normality test has better power for the Tobit model than for the probit model.

\(^{11}\) See Robinson (1982) and Smith (1989).
Figure 5: Power analysis, Sample Selection Model, 25% of Censored Obs.
Figure 6: Power analysis, Sample Selection Model, 75% of Censored Obs.
Figure 7: Power analysis, Bivariate Probit Model
been far too cumbersome, so our main results are summarised in Figures 5–7, which are organised as follows: since we are mainly interested in the power properties of the test for several degrees of deviations from normality, the x axis represents the number of the degrees of freedom of the bivariate $\chi^2$ random variate used in the experiment.

As expected, the power of the test statistic decreases in the number of degrees of freedom $\nu$ of the $\chi^2$ distribution and increases in the sample size: in some cases the percentage of rejections is very close to 100.

The correlation coefficient $\rho$ between the two disturbance terms seems to have a mixed effect: in Heckman’s sample selection model, power seems to decrease as $\rho$ moves away from 0; on the contrary, different values of $\rho$ in the bivariate probit model seem to leave the power unaffected.

Finally, the comparison between the two versions of the Heckit model seems to reinforce the idea that power is decreasing in the degree of censoring: for a sample as large as 1024, power is extremely good when the degree of censoring is 25%, but is much less satisfactory when the degree of censoring equals 75%.

6 A real data example

In order to provide a real-life example of the IM test statistic, we chose the empirical model by Martins (2001). The author estimates a wage equation on a sample of 2339 married women below the age of 60 interviewed in 1991. As only 60% of the women considered is employed, a selection equation describing labor force participation is needed. As customary in this kind of applications, a sample selection model is estimated by maximum likelihood under normality. The author, however, after applying the normality test by Horowitz and Härdle (1994), finds that the probit model is inadequate for the selection equation. The solution proposed in the paper is to estimate the sample selection model semi-parametrically by using the two-step series estimator, as per Newey (1999).

Table 4 replicates Martins’ ML estimates and shows the IM test for joint normality. As can be seen from the table, the presence of a non-normality issue is strongly confirmed, since the IM test statistic takes a value of 406.43. The asymptotic distribution of the test would be in this case $\chi^2$ with 25 degrees of freedom, which would lead, of course to reject the normality hypothesis.

The same conclusion would be reached by considering bootstrap critical values: out of 400 bootstrap replications, not a single one even approaches the value of the IM test obtained for the real data, so it makes little sense even to speak of a p-value. However, the descriptive statistics on the 400 bootstrap replications show that the divergence of the finite-sample distribution of the IM test from the asymptotic one is striking: a $\chi^2_{25}$ distributions has mean 25 and standard deviation 7.07, which are very far from the bootstrap statistics (39.2 and 14.16, respectively).

Interestingly, by applying univariate normality tests to both equations, it turns out that the misspecification of the marginal distributions seems far more severe for the main equation than for the selection one. Other than being the
Table 4: Sample selection model and test of bivariate normality

<table>
<thead>
<tr>
<th></th>
<th>Main equation</th>
<th>Selection equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>4.484***</td>
<td>-0.575</td>
</tr>
<tr>
<td>edu</td>
<td>0.114***</td>
<td>0.150***</td>
</tr>
<tr>
<td>pexp</td>
<td>0.132*</td>
<td></td>
</tr>
<tr>
<td>pexp2</td>
<td>-0.003</td>
<td></td>
</tr>
<tr>
<td>pexpchd</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>pexpchd2</td>
<td>-0.011*</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td></td>
<td>0.807***</td>
</tr>
<tr>
<td>age2</td>
<td></td>
<td>-0.123***</td>
</tr>
<tr>
<td>child</td>
<td></td>
<td>-0.118***</td>
</tr>
<tr>
<td>ychild</td>
<td></td>
<td>-0.090</td>
</tr>
<tr>
<td>hw</td>
<td></td>
<td>-0.103</td>
</tr>
<tr>
<td>( \hat{\rho} )</td>
<td>0.35*</td>
<td>Log-lik.</td>
</tr>
<tr>
<td>Total obs.</td>
<td>2339</td>
<td>Cens. obs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>939 (40.0%)</td>
</tr>
</tbody>
</table>

Univariate normality tests

Main eq.: 362.853 p-value = 1.612e-79
Sel eq.: 11.2899 p-value = 0.00353

Bivariate normality test

IM Test statistic = 406.4274 (asymptotic p-value = 1.49906e-70)

Summary statistics on bootstrap distribution

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Skewness</th>
<th>Ex. kurtosis</th>
<th>95% perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>39.234</td>
<td>0.84616</td>
<td>0.59760</td>
<td>66.931</td>
</tr>
<tr>
<td>Median</td>
<td>37.469</td>
<td>0.59760</td>
<td>20.357</td>
<td>66.931</td>
</tr>
<tr>
<td>Minimum</td>
<td>13.128</td>
<td>0.59760</td>
<td>5% perc.</td>
<td>20.357</td>
</tr>
<tr>
<td>Maximum</td>
<td>87.732</td>
<td>0.59760</td>
<td>95% perc.</td>
<td>66.931</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>14.165</td>
<td>IQ range</td>
<td>18.566</td>
<td></td>
</tr>
</tbody>
</table>

Correct statistical choice, a test of bivariate normality in this kind of models is able to capture the joint distributional misspecification.

7 Conclusions

The IM test, in the form originally proposed by Smith (1985), despite being very attractive for its portability, its ease of generalisation to all possible censoring schemes and its simplicity of computation, does not provide a reliable tool to detect non-normality in bivariate limited dependent variable models unless in very large samples. For commonly used models, the IM test suffers form considerable size bias even in samples with several thousands observations, which common practice would consider as perfectly adequate for asymptotics. Even after selecting carefully crafted subsets of the available moment conditions, the situation only improves marginally. This finding is consistent with previously
available results on the finite-sample performance of the Information Matrix test.

However, a simple bootstrap-based computation of the appropriate critical values seems to be extremely effective: size properties of the test are generally very good even in rather small samples, while it retains good power, especially against those alternatives which are more likely to represent a problem in practical cases.

In conclusion, the bootstrap variant of the Information Matrix test provides a simple and only mildly computationally intensive way to perform a joint normality test in bivariate models with some form of censoring. Since non-normality is a potentially very disruptive cause of misspecification, usage of this test procedure is highly recommended.

References


Table 5: Sample Selection Model, 25% of Censored Observations: Empirical Size

<table>
<thead>
<tr>
<th>n</th>
<th>ρ</th>
<th>Asymptotic χ²</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>81.59</td>
<td>8.24</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>85.30</td>
<td>7.68</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>91.97</td>
<td>6.15</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>59.31</td>
<td>8.24</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>68.08</td>
<td>8.91</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>79.76</td>
<td>9.98</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>46.95</td>
<td>10.35</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>59.52</td>
<td>10.16</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>70.65</td>
<td>9.99</td>
</tr>
</tbody>
</table>

third-and-fourth-moment, nominal size = 10%

<table>
<thead>
<tr>
<th>n</th>
<th>ρ</th>
<th>Asymptotic χ²</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>63.27</td>
<td>8.81</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>67.32</td>
<td>8.32</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>78.38</td>
<td>7.08</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>54.07</td>
<td>10.08</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>58.98</td>
<td>9.38</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>71.77</td>
<td>9.60</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>47.24</td>
<td>10.41</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>50.53</td>
<td>10.01</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>63.32</td>
<td>10.07</td>
</tr>
</tbody>
</table>

all-moments, nominal size = 5%

<table>
<thead>
<tr>
<th>n</th>
<th>ρ</th>
<th>Asymptotic χ²</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>74.39</td>
<td>3.99</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>79.19</td>
<td>3.60</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>88.11</td>
<td>2.43</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>49.37</td>
<td>4.70</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>59.25</td>
<td>4.21</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>72.75</td>
<td>4.21</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>37.24</td>
<td>5.13</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>50.53</td>
<td>4.98</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>63.32</td>
<td>4.80</td>
</tr>
</tbody>
</table>

third-and-fourth-moments, nominal size = 5%

<table>
<thead>
<tr>
<th>n</th>
<th>ρ</th>
<th>Asymptotic χ²</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>84.07</td>
<td>4.11</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>58.98</td>
<td>3.85</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>71.77</td>
<td>3.02</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>26.07</td>
<td>4.88</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>31.43</td>
<td>4.45</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>41.53</td>
<td>4.58</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>12.83</td>
<td>5.42</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>15.73</td>
<td>4.61</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>21.38</td>
<td>5.16</td>
</tr>
</tbody>
</table>

all-moments, nominal size at 1%

<table>
<thead>
<tr>
<th>n</th>
<th>ρ</th>
<th>Asymptotic χ²</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>57.68</td>
<td>0.60</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>64.16</td>
<td>0.55</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>77.95</td>
<td>0.29</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>32.17</td>
<td>0.91</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>41.79</td>
<td>0.77</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>58.06</td>
<td>0.5</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>22.52</td>
<td>1.02</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>34.28</td>
<td>0.97</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>48.84</td>
<td>0.88</td>
</tr>
</tbody>
</table>

third-and-fourth-moments, nominal size = 1%

<table>
<thead>
<tr>
<th>n</th>
<th>ρ</th>
<th>Asymptotic χ²</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>57.68</td>
<td>0.60</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>64.16</td>
<td>0.55</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>77.95</td>
<td>0.29</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>32.17</td>
<td>0.91</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>41.79</td>
<td>0.77</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>58.06</td>
<td>0.5</td>
</tr>
<tr>
<td>256</td>
<td>0.00</td>
<td>22.52</td>
<td>1.02</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>34.28</td>
<td>0.97</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>48.84</td>
<td>0.88</td>
</tr>
</tbody>
</table>

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 β_m = 1 for m = 0, 1, γ_0 = √3Φ⁻¹(p), where p is the percentage of uncensored observations, and γ_1 = 1. The standard deviation σ = 1.
Table 6: Sample Selection Model, 75% of Censored Observations: Empirical Size

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\rho$</th>
<th>Asymptotic $\chi^2_{15}$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>98.52 98.45 98.48</td>
<td>8.87 8.98 7.77</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>80.58 75.04 76.73</td>
<td>10.22 10.06 9.70</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>45.94 41.14 49.53</td>
<td>10.22 10.15 10.54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\rho$</th>
<th>Asymptotic $\chi^2_9$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>256.00 256.00 256.00</td>
<td>10.18 10.16 9.90</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>1024.00 1024.00 1024.00</td>
<td>9.90 9.87 9.67</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>4096.00 4096.00 4096.00</td>
<td>9.67 9.67 9.67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\rho$</th>
<th>Asymptotic $\chi^2_{15}$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>97.94 97.64 97.59</td>
<td>9.57 9.23 8.12</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>73.97 67.10 69.97</td>
<td>10.18 10.06 9.90</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>36.50 31.88 40.43</td>
<td>10.45 10.07 10.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\rho$</th>
<th>Asymptotic $\chi^2_9$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>90.29 90.16 90.38</td>
<td>9.57 9.23 8.12</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>57.32 56.28 58.63</td>
<td>10.18 10.06 9.90</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>19.20 30.30 33.18</td>
<td>10.45 10.07 10.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\rho$</th>
<th>Asymptotic $\chi^2_{15}$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>104.02 104.35 104.35</td>
<td>0.75 0.64 3.71</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>59.41 52.01 54.79</td>
<td>0.95 0.84 0.92</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>21.50 18.05 25.92</td>
<td>0.95 1.15 1.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\rho$</th>
<th>Asymptotic $\chi^2_9$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>0.00</td>
<td>76.48 76.26 77.07</td>
<td>0.90 0.75 0.64</td>
</tr>
<tr>
<td>1024</td>
<td>0.50</td>
<td>32.93 31.91 35.44</td>
<td>0.99 0.96 0.88</td>
</tr>
<tr>
<td>4096</td>
<td>0.75</td>
<td>10.77 10.78 13.25</td>
<td>1.09 1.23 1.03</td>
</tr>
</tbody>
</table>

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 $\beta_m = 1$ for $m = 0, 1$, $\gamma_0 = \sqrt{3p^{-1}(p)}$, where $p$ is the percentage of uncensored observations, and $\gamma_1 = 1$. The standard deviation $\sigma = 1$. 

27
Table 7: Bivariate Probit Model: Empirical Size

<table>
<thead>
<tr>
<th></th>
<th>Asymptotic $\chi^2_\rho$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>256</td>
<td>78.48</td>
<td>82.29</td>
</tr>
<tr>
<td>1024</td>
<td>48.77</td>
<td>57.65</td>
</tr>
<tr>
<td>4096</td>
<td>26.56</td>
<td>30.55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Asymptotic $\chi^2_\rho$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>256</td>
<td>71.74</td>
<td>76.15</td>
</tr>
<tr>
<td>1024</td>
<td>40.14</td>
<td>48.75</td>
</tr>
<tr>
<td>4096</td>
<td>19.26</td>
<td>22.71</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Asymptotic $\chi^2_\rho$</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>256</td>
<td>57.89</td>
<td>62.38</td>
</tr>
<tr>
<td>1024</td>
<td>26.46</td>
<td>32.27</td>
</tr>
<tr>
<td>4096</td>
<td>9.17</td>
<td>11.53</td>
</tr>
</tbody>
</table>

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 $r = 1, 2$.  

28
A Information Matrix tests in bivariate limited dependent variable models

Consider the latent variable model:

\[ y_1^* = x_1' \beta_1 + v_1 \]  
\[ y_2^* = x_2' \beta_2 + v_2 \]  

where \( v_1 \) and \( v_2 \) have the following bivariate normal distribution:

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

where \( x_1 \) and \( x_2 \) are vectors of exogenous variables and \( \beta_1, \beta_2 \) 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:

\[ y_1^* = x_1' \beta_1 + \rho v_2 + u \]  
\[ y_2^* = x_2' \beta_2 + v_2 \]

with \( u|x_1, v_2 \sim N(0, \omega_{11}^2) \) where \( \omega_{11}^2 = \omega_1^2 - \omega_{12}^2/\omega_2^2 \).

The log-likelihood for the latent variables \( \ell^* \) can be split into conditional \( \ell^*_{12} \) and marginal \( \ell^* \) log-likelihoods so that:

\[ \ell^*(y_1^*, y_2^*; \theta) = \ell^*_{12}(y_1^*|y_2^*; \theta) + \ell^*_2(y_2^*; \theta_2) \]  

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:

\[ \frac{\partial \ell}{\partial \theta} = E \left[ \frac{\partial \ell^*}{\partial \theta} \mid y \right] \]  

and the Hessian matrix as:

\[ \frac{\partial^2 \ell}{\partial \theta \partial \theta'} = E \left[ \frac{\partial^2 \ell^*}{\partial \theta \partial \theta'} \mid y \right] + V \left[ \frac{\partial \ell^*}{\partial \theta} \mid y \right] \]  

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:

\[ \varepsilon^r \xi^s = E(\varepsilon^r \xi^s|y) - E(\varepsilon^r) \]  

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

\[ \text{A rigorous proof is given in Gourieroux, Monfort, Renault, and Trognon (1984).} \]
where $\varepsilon = u/\omega_{1,2}$ 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 sample selection model

As mentioned in 3.1, in the case of the sample selection model we are only able to prove that the upper bound for $df$ is equal $k(k+1)/2 - 1$ since the last moment condition $C_{\psi, \psi}$ is always dropped in the OPG regression. 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 $x_i$s with $s = 1, \ldots, m$ and $w_{ir}$ with $r = 1, \ldots, h$ 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 (1)). $C_{\psi, \psi}$ can be written as a linear combination of other columns of matrix $M$. For this purpose, the following expressions are developed. Consider first the generic $r$-condition $C^{r, \sigma}_{i}$ also as a function of $u_i$ and $b_i$:

$$C^{r, \sigma}_{i} = \frac{1}{\sigma} \mu_i \left[ s_\psi c_\psi^2 u_i b_i + s_\psi^2 c_\psi u_i^2 + c_\psi u_i^2 - c_\psi b_i \right] w_{ir}$$

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

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

which gives

$$\sigma \sum_{r=1}^{h} \gamma_r C^{r, \sigma}_{i} = \mu_i \left[ s_\psi c_\psi^2 u_i b_i + s_\psi^2 c_\psi u_i^2 + c_\psi u_i^2 - c_\psi b_i \right]$$

(20)

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

$$t_\psi^2 \sigma \sum_{r=1}^{h} \gamma_r C^{r, \sigma}_{i} = \mu_i \left[ s_\psi^3 u_i b_i^2 + s_\psi^2 c_\psi u_i^2 b_i - \frac{s_\psi^2}{c_\psi} b_i \right]$$

(21)
Similar transformations applied to \( C_i^{\gamma} \) yield

\[
t_{\psi} \sum_{r=1}^{h} \gamma_r C_i^{\gamma} = \mu_i \left[ -s_{\psi}^2 c_{\psi} b_1^2 - s_{\psi} c_{\psi} u_i b_2^2 - s_{\psi}^2 u_i b_2^2 - s_{\psi}^2 c_{\psi} u_i^2 b_1 + \frac{s_{\psi}^2}{c_{\psi}} b_i \right]
\]  

(22)

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

\[
\sigma t_{\psi} C_i^{\sigma} = \mu_i \left[ -2s_{\psi} u_i - \frac{s_{\psi}^2}{c_{\psi}} b_i + s_{\psi} c_{\psi} u_i^2 b_1 + 2s_{\psi}^2 c_{\psi} u_i^2 b_i \right]
\]  

(23)

and, finally, we need

\[
t_{\psi} G_i^{\psi} = \mu_i \left[ \frac{s_{\psi}^2}{c_{\psi}} b_i + s_{\psi} u_i \right]
\]  

(24)

Since \( C_i^{\psi} \), written as a function of \( u_i \) and \( b_i \) is

\[
C_i^{\psi} = \mu_i \left[ a_i (1 - c_i^2) \right] = 
\mu_i \left[ -s_{\psi} c_{\psi}^2 u_i^3 - s_{\psi} c_{\psi} b_1^3 - c_{\psi}^2 u_i b_1 - s_{\psi} u_i b_2 - 2s_{\psi} c_{\psi} u_i^2 b_2 - 2s_{\psi}^2 c_{\psi} u_i^2 b_1 + c_{\psi} b_1 + s_{\psi} u_i \right]
\]  

(25)

it can now be expressed as a linear combination of (20), (21), (22), (23) and (24)

\[
C_i^{\psi} = -\sigma (1 - t_{\psi}^2) \sum_{r=1}^{h} \gamma_r C_i^{\gamma,r} + t_{\psi} \sum_{r=1}^{h} \gamma_r C_i^{\gamma,r} - \sigma t_{\psi} C_i^{\sigma} - t_{\psi} G_i^{\psi}
\]

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 \((\beta_0, \gamma_0, \gamma_1, \sigma, \psi)\). Six of the fifteen moment conditions are dropped in the OPG regression. Naturally

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

\[
C_i^{\beta_0, \gamma_0, \gamma_0} = d_i \frac{t_{\psi}^2}{\sigma} G_i^{\psi}
\]

are dropped as they are linear combinations of the score elements. Also \( C_i^{\gamma_0, \gamma_0} \) is a linear combination of \( G_i^\sigma, G_i^{\gamma_1}, G_i^{\psi} \). First notice that in this specific setup
\[ b_i = \gamma_0 + w_i \gamma_1 \]

and

\[ a_i = c_\psi \gamma_0 + c_\psi w_i \gamma_1 + s_\psi u_i; \quad c_i = s_\psi \gamma_0 + s_\psi w_i \gamma_1 + c_\psi u_i \]

then write

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

Given the following transformations, only for uncensored observations,

\[ -c_\psi^3 \gamma_0 G_{i}^{\gamma_0} = -c_\psi^3 \gamma_0 \mu_i; \quad -c_\psi^2 \gamma_1 G_{i}^{\gamma_1} = -c_\psi^2 w_i \gamma_1 \mu_i \]

and

\[ -s_\psi c_\psi G_{i}^{\psi} = -\mu_i \left[-c_\psi^3 \gamma_0 - c_\psi^2 w_i \gamma_1 - c_\psi^2 s_\psi u_i\right], \]

\[ C_{i}^{\gamma_0, \gamma_1} \text{ can be written as} \]

\[ C_{i}^{\gamma_0, \gamma_1} = -d_i \left[c_\psi s_\psi G_{i}^{\psi} + \gamma_0 G_{i}^{\gamma_0} + \gamma_1 G_{i}^{\gamma_1}\right] - (1 - d_i) \left[\gamma_0 G_{i}^{\gamma_0} + \gamma_1 G_{i}^{\gamma_1}\right]. \]

\[ C_{i}^{\gamma_0, \sigma} \text{ is also a linear combination of } G_{i}^{\psi} \text{ and of the moment conditions } C_{i}^{\beta_0, \gamma_1} \text{ and } C_{i}^{\beta_0, \psi}. \]

Rearranging algebraically \( C_{i}^{\gamma_0, \sigma} \) and \( C_{i}^{\beta_0, \psi} \) we get

\[ C_{i}^{\gamma_0, \sigma} = C_{i}^{\beta_0, \psi} - \mu_i \frac{1}{\sigma} \left[s_\psi^2 c_\psi b_i + s_\psi c_\psi^2 u_i\right] (\gamma_0 + \gamma_1 w_i). \] (26)

Applying the following transformations

\[ \frac{s_\psi c_\psi}{\sigma} G_{i}^{\gamma_0} = \frac{\mu_i}{\sigma} \left[s_\psi^2 c_\psi b_i + s_\psi c_\psi^2 u_i\right] \gamma_0 \]

\[ \frac{s_\psi c_\psi}{c_\psi} C_{i}^{\beta_0, \gamma_1} = \frac{\mu_i}{\sigma} \left[s_\psi^2 c_\psi b_i + s_\psi c_\psi^2 u_i\right] \gamma_1 w_i \]

we can rewrite (26) as

\[ C_{i}^{\gamma_0, \sigma} = C_{i}^{\beta_0, \psi} - \frac{s_\psi c_\psi}{\sigma} G_{i}^{\psi} - \frac{s_\psi}{c_\psi} \gamma_1 C_{i}^{\beta_0, \gamma_1}. \] (27)

The OPG regression also drops

\[ C_{i}^{\gamma_0, \psi} = -\mu_i (c_\psi a_i c_1 - s_\psi). \] (28)

Considering

\[ -\sigma \frac{s_\psi c_\psi}{s_\psi} C_{i}^{\beta_0, \psi} = -\mu_i \left(c_\psi a_i c_1 + \frac{c_\psi}{s_\psi} u_i c_1 - \frac{c_\psi^2}{s_\psi}\right) \]

and

\[ \frac{\sigma}{s_\psi c_\psi} C_{i}^{\gamma_0, \sigma} = \mu_i \left(c_\psi b_i u_i + \frac{c_\psi^2 u_i^2}{s_\psi} - \frac{1}{s_\psi}\right) \]

we can rewrite (28) as

\[ C_{i}^{\gamma_0, \psi} = -\sigma \frac{c_\psi}{s_\psi} C_{i}^{\beta_0, \psi} + \frac{\sigma}{s_\psi c_\psi} C_{i}^{\gamma_0, \sigma} \] (29)
and substituting (27) in (29) we finally get
\[ C_{i}^{\gamma_{0},\psi} = -\sigma_{s}^{\psi} C_{i}^{\gamma_{0},\psi} - \gamma_{0} G_{i}^{\psi} - \sigma_{s}^{\psi} C_{i}^{\gamma_{1},\psi} \]

Finally the last column to be dropped is \( C^{\psi,\psi}_{i} \), 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.

### B.2 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,
\[ -x_{1i}\beta_{1} = a_{i} = \bar{a} \quad -x_{2i}\beta_{2} = b_{i} = \bar{b} \quad P_{i} = \bar{P} \]

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
\[ G_{i}^{a} = \bar{S}^{a} \quad G_{i}^{b} = \bar{S}^{b} \quad G_{i}^{\psi} = \bar{S}^{\psi} \]

This makes every moment condition \( C_{i} \) a linear combination of the score elements (compare Tables 2 and 8), which means that all these conditions are collinear to the score matrix and, as a consequence, do not contribute to the rank of \( M \), as defined in Section 3.1.

Table 8: Moment Conditions for the Bivariate Probit Model with only two constant terms as regressors

<table>
<thead>
<tr>
<th></th>
<th>( \beta_{1} )</th>
<th>( \beta_{2} )</th>
<th>( \psi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{1} )</td>
<td>(-\bar{a}S^{a} - c_{\psi}s_{\psi}\bar{S}^{\psi})</td>
<td>(c_{\psi}S^{\psi})</td>
<td>(-u_{a,b,c_{\psi}}S^{\psi})</td>
</tr>
<tr>
<td>( \beta_{2} )</td>
<td>(-\bar{b}S^{b} - c_{\psi}s_{\psi}\bar{S}^{\psi})</td>
<td>(-u_{b,a,c_{\psi}}S^{\psi})</td>
<td>(\bar{S}^{\psi}(u_{a,b}u_{b,a} - t_{\psi}))</td>
</tr>
</tbody>
</table>

Let us turn to the opposite extreme case, with non-overlapping sets of regressors\(^{17}\). It is possible to prove that the last moment condition, \( C^{\psi,\psi}_{i} \) 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
\[ a = \sum_{r=1}^{k_{1}} x_{1r}\beta_{1r} \quad b = \sum_{s=1}^{k_{2}} x_{2s}\beta_{2s} \]

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

\(^{17}\) 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.
Now write the moment condition associated with the cross-derivative of $\beta_1t$ and $\psi$ as:

$$C^{\beta_1r, \beta_2s} = c_\psi S^\psi x_{1r} x_{2s}$$

(30)

By using (30), the previous expression becomes

$$C^{\beta_1r, \psi} = \left[ c_\psi \left( \sum_{r=1}^{k_1} x_{1r} \beta_{1r} \right) x_{1t} - s_\psi \left( \sum_{s=1}^{k_2} x_{2s} \beta_{2s} \right) x_{1t} \right] c_\psi S^\psi$$

By using (30), the previous expression becomes

$$C^{\beta_1r, \psi} = -c_\psi^2 S^\psi \sum_{r=1}^{k_1} x_{1t} x_{1r} \beta_{1r} + t_\psi \sum_{s=1}^{k_2} C^{\beta_1r, \beta_2s} \beta_{2s}$$

(31)

By symmetry, the moment condition associated with the cross-derivative of $\beta_{2m}$ and $\psi$ can be written as

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

(32)

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

$$C^{\psi, \psi} = S^\psi \left[ c_\psi^2 a b + s_\psi^2 a b - s_\psi c_\psi b^2 - t_\psi \right] =

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

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

Note that

$$c_\psi^2 S^\psi \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} x_{1r} x_{2s} \beta_{1r} \beta_{2s} = \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} C^{\beta_{1r}, \beta_{2s}} \beta_{1r} \beta_{2s}$$

(33)

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

(34)

and that by multiplying (31) by $\beta_{1t} t_\psi$ one gets

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

(35)

Similarly, (32) may be multiplied by $\beta_{2m} t_\psi$ to obtain

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

(36)
Finally, after rearranging (33), (34), (35) and (36), we can rewrite $C^{\psi, \psi}$ as

\[
C^{\psi, \psi} = (1 - t_\psi^2) \sum_{r=1}^{k_1} \sum_{s=1}^{k_2} C^{\beta_1 r, \beta_2 s} \beta_1 r \beta_2 s + t_\psi \sum_{r=1}^{k_1} C^{\beta_1 r, \psi} \beta_1 r + t_\psi \sum_{s=1}^{k_2} C^{\beta_2 s, \psi} \beta_2 s - t_\psi S^{\psi, \psi},
\]

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 $C^{\psi, \psi}$ 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

\[
a = \sum_{r=1}^{q} x_r \beta_1 r, \quad b = \sum_{r=1}^{q} x_r \beta_2 r.
\]

Similarly, $G^{\beta_j}$ has $q$ elements

\[
[G^{\beta_{1j}}, G^{\beta_{2j}}, \ldots, G^{\beta_{qj}}]
\]

for $j = 1, 2$, such that

\[
G^{\beta_{1r}} = S^a x_r, \quad G^{\beta_{2r}} = S^b x_r
\]

for $r = 1, \ldots, q$ (see also section 4.2).

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

\[
C^{\beta_{11}, \beta_{11}} = - \left[ \sum_{r=1}^{q} x_r \beta_1 r \right] S^a + s_\psi c_\psi S^{\psi} = - \left[ \sum_{r=1}^{q} G^{\beta_{1r}} \beta_1 r + s_\psi c_\psi S^{\psi} \right]
\]

\[
C^{\beta_{21}, \beta_{21}} = - \left[ \sum_{r=1}^{q} x_r \beta_2 r \right] S^b + s_\psi c_\psi S^{\psi} = - \left[ \sum_{r=1}^{q} G^{\beta_{2r}} \beta_2 r + s_\psi c_\psi S^{\psi} \right]
\]

Consider now the $q^2$ elements associated with cross derivatives of $\beta_1 r, \beta_2 s$

\[
C^{\beta_{1s}, \beta_{2s}} = c_\psi^2 S^{\psi} x_r x_s
\]

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

\[
C^{\beta_{11}, \beta_{21}} = c_\psi^2 S^{\psi}.
\]
The are also 2q elements, collinear to other columns of $M$, associated with the cross derivatives of regressors with $\psi$ since

$$C^{\beta_1, \psi} = -\left[ c_\psi \left( \sum_{r=1}^{q} x_r \beta_{1r} \right) - s_\psi \left( \sum_{r=1}^{q} x_r \beta_{2r} \right) \right] c_\psi S^\psi x_t =$$

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

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

and as well

$$C^{\beta_2, \psi} = -\left[ c_\psi \left( \sum_{r=1}^{q} x_r \beta_{2r} \right) - s_\psi \left( \sum_{r=1}^{q} x_r \beta_{1r} \right) \right] c_\psi S^\psi x_t =$$

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

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

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

$$df = k(k+1)/2 - 2 - q^2 + q(q+1)/2 - 1 - 2q - 1$$

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

$$df = 3 \left[ \frac{q(q+1)}{2} - 1 \right]$$.