Department of Economics
School of Social Sciences

Robust misspecification tests
for the Heckman's two-step estimator

Gabriel V. Montes-Rojas[1]
City University

Department of Economics

Discussion Paper Series

No. 08/01

[1] Department of Economics, City University, Northampton Square, London, EC1V 0HB, UK. *Gabriel.Montes-Rojas.1@city.ac.uk*

Robust misspecification tests for the Heckman's two-step estimator

Gabriel V. Montes-Rojas*
City University London

Abstract

We construct and evaluate LM and Neyman's $C(\alpha)$ tests based on bivariate Edgeworth expansions for the consistency of the Heckman's two-step estimator in selection models, that is, for the marginal normality and linearity of the conditional expectation of the error terms. The proposed tests are robust to local misspecification in nuisance distributional parameters. Monte Carlo results show that instead of testing bivariate normality, testing marginal normality and linearity of the conditional expectations separately have a better size performance. Moreover, the robust variants of the tests have better size and similar power to non-robust tests, which determines that these tests can be successfully applied to detect specific departures from the null model of bivariate normality. We apply the tests procedures to women's labor supply data.

Keywords: Heckman's two-step; LM tests; Neyman's $C(\alpha)$ tests

*City University London, Department of Economics, Northampton Square, London EC1V 0HB, United Kingdom, E-mail: Gabriel.Montes-Rojas.1@city.ac.uk, Tel: +44 (0)20 7040 8919

1 Introduction

Selection models are widely used in applied econometrics. Starting with the pioneering work of Heckman (1974, 1976, 1979), these models account for the interaction of the error terms in the outcome and selection equations. Two different misspecification problems are common in this setting: incorrect functional form (e.g. omitted variables or nonlinearity) and erroneous stochastic structure (distributional misspecification of the random variables). Through this paper we study the second type of misspecification for the most common estimation method, the Heckman's two-step estimator.

The assumption of bivariate normality provides efficient and easy-to-compute estimation procedures. Maximum likelihood estimates (MLE) and two-step procedures are available in all statistical packages. While the former relies on the complete specification of the error terms' distribution, the latter is more robust, in the sense that it requires less distributional assumptions. In general, these methods are not robust to misspecification of the data generating process (DGP). However, in practice, the normality assumption is used in a very large number of applications. A literature path concentrated on testing bivariate normality (Bera and John, 1983, Lee, 1982, 1984, Bera et al., 1984). However consistent estimates of the parameters of interest only requires normality of the error term in the selection equation, as well as linearity of the conditional expectation of the outcome equation error term conditional on the selection equation one. In this environment, bivariate normality is a much stronger assumption, which is in general rejected. Our goal is to devise a testing procedure that allows for separately testing for linearity of the conditional expectation and normality of marginal distributions, while allowing some local flexibility on the bivariate distribution of the random variables.

Misspecification tests using bivariate Edgeworth expansions (BEE) were first proposed by Lee (1982, 1984), which is the approach followed here. In BEE, the DGP can be approximated using a basis distribution function (bivariate normal) and information about the sample own and cross skewness and kurtosis is added to generate the set of prior admissible family of distribution functions. Our interest will be centered in a certain subset of parameters: those related to marginal normality and conditional linearity of the expectation terms.

Gabler et al. (1993) and van der Klaauw and Koning (1993) proposed a similar methodology for testing normality in sample selection models using flexible semi-nonparametric (SNP) specifications introduced by Gallant and Nychka (1987). Our procedure has two main advantages over SNP. First, it allows distinguishing among different sources of distributional misspecification, i.e. non-normality of any of the error terms or conditional non-linearity. SNP procedures reject bivariate normality without any constructive indication on what to do next. Second, in SNP, the distributional parameters (i.e. ρ and σ in the model below) cannot be, in general, identified if there is any departure from the null DGP. A major limitation of BEE is that it requires imposing restrictions on the parameters' estimates in order to satisfy the non-negativity of the density function (see for instance Jondeau and Rockinger, 1994). However, the Lagrange Multiplier (LM) procedures we follow do not require estimation of additional parameters as in the van der Klaauw and Koning (1993) likelihood ratio tests.

Unfortunately without considering the remaining parameters of the BEE (those not related to marginal normality and conditional linearity), LM tests would have incorrect asymptotic size. Therefore, the LM statistics are adjusted in a way such that the resulted statistic is orthogonal to the score functions of the unconsidered BEE parameters. This approach was first developed by Bera and Yoon (1993) to construct robust LM statistics. These tests are locally size-resistant, that is, they have correct asymptotic size and some optimal properties in the presence of local misspecification, although they may have less power if the alternative hypothesis is not misspecified. We also compute Neyman's $C(\alpha)$ test statistics, which are optimal for any $\sqrt{n}$-consistent estimator (*e.g.* Heckman's two-step estimator in selection models), and we introduce a locally size-robust variant of this test to accommodate to the same local misspecification type.

The paper is organized as follows. Section 2 presents a typical selection model, as well as the role of assuming bivariate normality in the estimation procedures. In section 3, bivariate Edgeworth expansions are analyzed, and the derivation of the test statistics are left for section 4. Section 5 contains details on the implementation and computation of the statistics, while section 6 presents Monte Carlo results. Section 7 applies the test procedures to Mroz (1987) women's labor supply dataset. Finally conclusions and suggestions for future research on this topic are in section 8.

2 Selection models

Consider the following standard selection model. Let i index the observations in a random sample sample with $i = 1, 2, ...n$. Assume that our interest is given by the following outcome equation:

$$(1) \quad y_i = x_i\beta + u_i$$

where y is an outcome of interest, x is a set of covariates, u is an error term and β is the main parameter of interest.

A selection mechanism is behind (1), and we only observe y if a certain event (indexed by the binary random variable c) occurs (i.e. $c = 1$). Assume the following selection process:

$$(2) \quad c_i = \left\{ \begin{array}{l} 1 \text{ if } z_i\gamma + e_i > 0 \\ 0 \text{ if } z_i\gamma + e_i \leq 0 \end{array} \right\}$$

where z is another set of covariates, not necessarily disjoint from x, and e is another error term. The inconsistency of the standard least squares estimates of equation (1) is in general the result of the non-independence of u and e, that is, the censoring mechanism depends on the bivariate distribution of these random variables. Let $G(u, e)$ denote the bivariate distribution function of the error terms, with corresponding density function $g(u, e)$. Bivariate normality of $g(.)$ results in a standard estimation method, which could be based in MLE or two-step estimation procedures. In this case the conditional mean function becomes

$$(3) \quad E(y|c = 1) = x\beta + \rho\sigma\lambda(-z\gamma)$$

where $\lambda(e) = \frac{\phi(e)}{1-\Phi(e)}$ is the inverse Mill's ratio. While MLE estimators are sensitive to any kind of distributional misspecification, the two-step procedure (known as the Heckman's two-step method) is robust to distributional misspecification if and only if the two following conditions are met:

Condition 1: e has marginal normal distribution.
Condition 2: $E(u|e) = \rho e$, i.e. the conditional expectation is linear.

The first condition specifies the form of the selection mechanism, while the second specifies how the selection mechanism affects the outcome equation though the conditional expectation of the error terms.

Several semi-parametric estimators have been proposed for these models if those conditions are not satisfied. If condition 1 is violated, there are alternative methods for estimating γ semi-parametrically or imposing other marginal distributions (Heckman and Navarro-Lozano, 2003). If condition 2 is violated, the outcome equation is estimated by introducing additional terms (i.e. polynomials in the inverse Mills ratio, see Buchinsky, 1998) or by weighted least squares (see Powell and Walker, 1990). Semi-parametric procedures follow Gallant and Nychka (1987) approach, where Hermite series expansions of the bivariate normal distribution allows for certain flexibility (van der Klaauw and Koning, 1993). However, semi-parametric methods are less efficient and more computationally intensive than the Heckman's two-step method. Furthermore, those models are not necessary robust to severe non-normality. Therefore, testing conditions 1 and 2 is important to avoid the potentially innecessary cost of a semi-parametric estimation and to detect potential inconsistencies.

3 Bivariate Edgeworth expansions

Under some general conditions (see Chambers, 1967), the joint density $g(u,e)$ can be expanded as a series of derivatives of the standard bivariate normal density function $\phi(u,e)$:

$$(4) \quad g(u,e) = \phi(u,e) + \sum_{r+s\geq 3}^{\infty} (-1)^{r+s} A_{rs} \frac{1}{r!s!} \frac{\partial^{r+s}\phi(u,e)}{\partial u^r \partial e^s}$$

where $A_{r,s}$ are functions of the cumulants (or semi-invariants) of u and e (see Mardia, 1970, Lee, 1982, 1984). This formulation is known as the bivariate Edgeworth series expansion[1], which is a generalization of the univariate case,

[1]As it is stated in Mardia (1970, p.12): "The work was initiated as early as 1896 by Edgeworth and he published further papers in 1905 and 1917. Van der Stok (1908),

the Gram-Charlier and Edgeworth expansions. It follows that the distribution function can also be expressed in this way as an expansion of bivariate distribution functions.

$$(5) \quad G(u,e) = \Phi(u,e) + \sum_{r+s\geq 3}^{\infty} (-1)^{r+s} A_{rs} \frac{1}{r!s!} \frac{\partial^{r+s-2}\phi(u,e)}{\partial u^{r-1}\partial e^{s-1}}$$

Moreover we have:

$$(6) \quad \frac{\partial^{r+s}\phi(u,e)}{\partial u^r \partial e^s} = (-1)^{r+s} H_{rs}(u,e)\phi(u,e)$$

where H_{rs} are the bivariate Hermite polynomials. The formal formulation of these polynomials can be found in Mardia (1970), Johnson and Kotz (1972, ch.34) and Ord (1972, Appendix A) and they are shown in the Appendix 1.

Following Lee (1982, 1984) we can restrict our attention to terms up to $r + s = 4$ (i.e. assuming that the terms $r + s > 4$ are zero), which has been called a Type AA surface in Mardia (1970)[2]. For the type AA surface $A_{rs} = \kappa_{rs}$, where κ_{rs} denotes the cumulants of order (r, s).

A nice feature of BEE is that marginal distributions are univariate Edgeworth expansions, that is:

$$(7) \quad g(e) = \left[1 + \sum_{s=3}^{\infty} \frac{A_{0s}}{s!} H_{0s}(e)\right] \phi(e)$$

Charlier (1914), Jorgensen (1916). Wicksell (1917, a,b), K.Pearson (1925), an others contributed to the field in its developing stages. Pretorious (1930) gives an account of these development. Kendall (1949) extends this to bivariate distributions."

[2]Expanding the number of terms considered in the BEE do not affect the results presented below. We have some preliminar evidence on this matter.

where H_{0s} denotes the univariate Hermite polynomial of order s and ϕ is a standard univariate normal distribution. Therefore a test for the marginal normality of e can be based on $A_{0s} = 0, s = 3, 4, ...,$ or in the Type AA surface $\kappa_{03} = \kappa_{04} = 0$.

Moreover, the conditional expectation $E(u|e)$ is:

$$(8) \quad E(u|e) = \frac{\sum_{s=3}^{\infty} \left(\rho e + \frac{A_{0s}}{s!} H_{0,s+1}(e) + \frac{A_{1,s-1}}{(s-1)!} H_{0,s-1}(e) \right)}{1 + \sum_{s=3}^{\infty} \frac{A_{0s}}{s!} H_{0,s}(e)}$$

Under marginal normality of e,

$$(9) \quad E(u|e) = \rho e + \sum_{s=2}^{\infty} \left(\frac{A_{1,s}}{s!} H_{0,s}(e) \right)$$

Therefore a tests for the conditional linearity of the expectation can be based on $A_{1s} = 0, s = 2, 3, ...,$ or in the Type AA surface $\kappa_{12} = \kappa_{13} = 0$.

It is worth to notice that κ_{rs} provide measures of skewness and kurtosis of the DGP. In particular the parameters that satisfy $r + s = 3$ determine skewness, and the ones such that $r + s = 4$ determine kurtosis, that is,

Skewness	Kurtosis
$\kappa_{30} = \mu_{30}$	$\kappa_{40} = \mu_{40} - 3$
$\kappa_{21} = \mu_{21}$	$\kappa_{40} = \mu_{31} - 3\rho$
	$\kappa_{22} = \mu_{22} - 2\rho^2 - 1$
$\kappa_{12} = \mu_{12}$	$\kappa_{13} = \mu_{13} - 3\rho$
$\kappa_{03} = \mu_{03}$	$\kappa_{04} = \mu_{04} - 3$

(Here we use the standard notation $\mu_{ij} = E(u^i e^j)$, the raw moments of a standard bivariate normal distribution with correlation coefficient ρ.)

Two major drawbacks of the BEE need to be taken into account. First, truncation of these series after a finite number of terms may lend to negative values of the density function and second, some restrictions on the multivariate series expansion are needed (see Chambers, 1967). A feature of these expansions is that tail behavior is restricted. Fat tails and heavy skewness translate into multimodal densities. Whether this feature is a not desirable one has to be analyzed given the empirical problem at hand. In time series studies, fat tails are a distinctive feature that cannot be avoided. However, in labor and other empirical economics studies, outliers may be result of the existence of "different" subpopulations, each with a dissimilar probability distribution.

4 Different test procedures: standard and robust LM and Neyman's $C(\alpha)$ tests

Lee (1984) derives a LM test for the bivariate normality of u and e based on $H_0 : \kappa = 0$, where $\kappa = \{\kappa_{30}, \kappa_{21}, \kappa_{12}, \kappa_{03}, \kappa_{40}, \kappa_{31}, \kappa_{22}, \kappa_{13}, \kappa_{04}\}$. Note that three different subsets of interest naturally emerge: testing the marginal normality of e (condition 1, $H_0^{C1} : \kappa_{03} = \kappa_{04} = 0$), testing conditional linearity (condition 2, $H_0^{C1} : \kappa_{12} = \kappa_{13} = 0$) or both (conditions 1 and 2, $H_0^{C1C2} : \kappa_{03} = \kappa_{04} = \kappa_{12} = \kappa_{13} = 0$). Let p denote the number of cumulants that are tested in each case.

For notational convenience define:

$$\eta = \{\beta, \gamma, \sigma, \rho\}$$

$$\kappa_0 = \{\text{cumulants of interest in either } H_0, H_0^{C1}, H_0^{C2} \text{ or } H_0^{C1C2}\}$$

$$\kappa_1 = \{\text{all } \kappa \text{ except those in } \kappa_0\}$$

Note that those hypotheses considered above do not completely specify the joint distribution of the error terms. Moreover, maximum likelihood estimation of all the parameters involved in the model (i.e. η and κ) is a very difficult task (see for instance Jondeau and Rockinger, 1994). Our strategy consists on testing either H_0^{C1}, H_0^{C2} or H_0^{C1C2}, assuming that the DGP is bivariate normal (i.e. H_0), provided that for this case it is straightforward to estimate η using MLE or the Heckman's two-step estimator. Moreover, these estimators are available in any econometric package, while semi-parametric estimators are not. In sum, our strategy is to use LM-type tests which only require estimation under the joint null hypothesis, and to correct for the effect of potential local departures in κ_1.

Consider a test procedure for either H_0^{C1}, H_0^{C2} or H_0^{C1C2} and note that the remaining cumulants (i.e. in κ_1) that specify the distribution among the Type AA family are not our main interest. In other words, κ_1 should be considered as a nuisance parameter. Different tests can be obtained depending on: (i) what are we willing to assume about κ_1, and (ii) how η and κ_1 (if necessary) are estimated.

Let $L(\eta, \kappa_0, \kappa_1)$ denote the general log-likelihood function for the statistical model of interest. Denote $L_0(\eta)$ as the null model with alternatives $L_1(\eta, \kappa_0)$, $L_2(\eta, \kappa_1)$ or the full log-likelihood $L(\eta, \kappa_0, \kappa_1)$. Following the Bera and Yoon (1993) notation assume that $L_0(\eta) = L_1(\eta, \kappa_0 = 0) = L_2(\eta, \kappa_1 = 0)$; $L(\eta, \kappa_0, \kappa_1 = 0) = L_1(\eta, \kappa_0)$ and $L(\eta, \kappa_0 = 0, \kappa_1) = L_2(\eta, \kappa_1)$. Let us also denote $\theta \equiv (\eta, \kappa_0, \kappa_1)$ and $\hat{\theta} = (\hat{\eta}, 0, 0)$, where η is the MLE estimator of η under bivariate normality. In that case, the LM test is the preferred approach and it can be constructed as:

$$(10) \quad LM_{\kappa_0} = \frac{1}{n} d_{\kappa_0}(\hat{\theta})^\top J_{\kappa_0 \cdot \eta}(\hat{\theta})^{-1} d_{\kappa_0}(\hat{\theta})$$

where for future reference

$$d_a(\theta) \equiv \frac{\partial L(\theta)}{\partial a} = \sum_{i=1}^{n} \frac{\partial L_i(\theta)}{\partial a} \text{ for } a = \eta, \kappa_0, \kappa_1$$

$$J(\theta) = E\left[\frac{1}{n}\frac{\partial L(\theta)}{\partial \theta}\frac{\partial L(\theta)}{\partial \theta^\top}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial L_i(\theta)}{\partial \theta}\frac{\partial L_i(\theta)}{\partial \theta^\top}\right] = \begin{bmatrix} \mathrm{J}_\eta & \mathrm{J}_{\eta\kappa_0} & \mathrm{J}_{\eta\kappa_1} \\ \mathrm{J}_{\kappa_0\eta} & \mathrm{J}_{\kappa_0} & \mathrm{J}_{\kappa_0\kappa_1} \\ \mathrm{J}_{\kappa_1\eta} & \mathrm{J}_{\kappa_1\kappa_0} & \mathrm{J}_{\kappa_1} \end{bmatrix}$$

$$J_{a\cdot b}(\theta) = J_a(\theta) - J_{ab}(\theta)J_b(\theta)^{-1}J_{ba}(\theta)$$

Under the corresponding null hypothesis $LM_{\kappa_0} \Rightarrow \chi^2_p(0)$, where "$\Rightarrow$" denotes asymptotic convergence in distribution. Under a sequence of local alternatives of the form $H_1 : \kappa_0 = \xi_0/\sqrt{n}, LM_{\kappa_0} \Rightarrow \chi^2_p(\lambda_0)$ where $\lambda_0 = \xi_0^\top J_{\kappa_0\cdot\eta}\xi_0$ denotes the non-centrality parameter of the chi-squared distribution with p degrees of freedom. Now suppose that the true log-likelihood function is $L_2(\eta, \kappa_1)$, meaning that the alternative $L_1(\eta, \kappa_0)$ is now misspecified. In that case the sequence of local alternatives becomes $H_1 : \kappa_1 = \xi_1/\sqrt{n}$ and $LM_{\kappa_0} \Rightarrow \chi^2_p(\lambda_1)$ where $\lambda_1 = \xi_1^\top J_{\kappa_1\kappa_0\cdot\eta} J^{-1}_{\kappa_0\cdot\eta} J_{\kappa_0\kappa_1\cdot\eta}\xi_1$. Note that an effect of this misspecification is that, in general, the size of the test is not correct, even asymptotically. Therefore, not considering the presence of the nuisance parameters would create a problem of *undertesting*. It is worth to mention that this problem may not occur if $\xi_1 (\neq 0)$ belongs to the null space of $J_{\kappa_0\kappa_1\cdot\eta}$, or $J_{\kappa_0\kappa_1\cdot\eta}$ itself is zero[3].

[3]This condition may be exploited in certain occasions to show that standard LM tests can be used without having to control for misspecification in some parameters. In our case, some covariance terms among the Hermite polynomials are zero under the assumption of

If we restrict our attention to local misspecification of the type presented above, a "robust" LM tests can be constructed as the Bera and Yoon (1993) adjusted LM tests or Jaggia and Trivedi (1994) conditional score tests.

$$(11) \quad LM^*_{\kappa_0} = \frac{1}{n} d_{\kappa_0(\kappa_1)}(\hat{\theta})^\top J_{\kappa_0 \cdot \eta}(\hat{\theta})^{-1} d_{\kappa_0(\kappa_1)}(\hat{\theta})$$

where $d_{\kappa_0(\kappa_1)}(\theta) = d_{\kappa_0}(\theta) - J_{\kappa_0\kappa_1\cdot\eta}(\theta) J_{\kappa_1\cdot\eta}(\theta)^{-1} d_{\kappa_1}(\theta)$ is the *adjusted score.*

In this case, under the null hypothesis and a sequence of local alternatives in κ_1, $LM_{\kappa_0} \Rightarrow \chi^2_p(0)$, that is, the test statistic is robust under local misspecification of the unconsidered parameters. Under local alternatives of the form $H_1 : \kappa_0 = \xi_0/\sqrt{n}$, $LM_{\kappa_0} \Rightarrow \chi^2_p(\lambda_2)$ where $\lambda_2 = \xi_0^\top (J_{\kappa_0\cdot\eta} - J_{\kappa_0\kappa_1\cdot\eta} J_{\kappa_1\cdot\eta} J_{\kappa_1\kappa_0\cdot\eta})\xi_0$. Since $\lambda_0 - \lambda_2 \geq 0$, the asymptotic power of this test will be less than when there is no misspecification (this is the problem of *overtesting*). This test statistic is the robust LM.

If $\sqrt{n}$-consistent estimates are available, Neyman's $C(\alpha)$ tests are optimal. The Heckman's two-step estimator falls under this class if conditions 1 and 2 (H_0^{C1C2} above) are satisfied. Let $\tilde{\theta} = (\tilde{\eta}, 0, 0)$ be that estimator under H_0. Unfortunately additional assumptions are needed to obtain consistent estimates of the cumulants that do not belong to the null hypothesis[4]. For this reason, Neyman's $C(\alpha)$ tests may be futile to overcome the problem at hand. The general form of this test is:

$$(12) \quad C_{\kappa_0} = \frac{1}{n} d_{\kappa_0\cdot\eta}(\tilde{\theta})^\top J_{\kappa_0\cdot\eta}(\tilde{\theta})^{-1} d_{\kappa_0\cdot\eta}(\tilde{\theta})$$

where $d_{\kappa_0\cdot\eta}(\theta) = d_{\kappa_0}(\theta) - J_{\kappa_0\eta}(\theta) J_\eta(\theta)^{-1} d_{\eta\kappa_1}(\theta)$ is the *effecttive score* and $\tilde{\theta}$

bivariate normality. This determines that some off-diagonal terms of the Jacobian matrix may be asymptotically negligible. We do not exploit this line of research.

[4]Only κ_{30} and κ_{40} can be consistently estimated.

denotes a $\sqrt{n}$-consistent estimator of θ. This test has the same asymptotic distribution as LM_{κ_0} under the null and alternative hypothesis.

Note the similarities with the LM robust test. In both cases, the nuisance parameters' influence has been taken out of the score functions of interest[5]. Also note that κ_1 does not enter in the computation of the Neyman's $C(\alpha)$ statistic, given that it is assumed to be equal to zero, and therefore it is not robust to local deviations in κ_1 . A simple extension of Bera and Yoon (1993) provides a way of adjusting Neyman's $C(\alpha)$ tests for local misspecification of this type. In this case we have

$$(13)\quad C^*_{\kappa_0} = \frac{1}{n} d_{\kappa_0(\kappa_1)\cdot\eta}(\tilde{\theta})^\top J_{\kappa_0\cdot\eta}(\tilde{\theta})^{-1} d_{\kappa_0(\kappa_1)\cdot\eta}(\tilde{\theta})$$

where $d_{\kappa_0(\kappa_1)\cdot\eta}(\theta) = d_{\kappa_0\cdot\eta}(\theta) - J_{\kappa_0\kappa_1\cdot\eta}(\theta) J_{\kappa_1\cdot\eta}(\theta)^{-1} d_{\kappa_1\cdot\eta}(\theta)$ is the *adjusted effective score.*

This test statistics is asymptotically equivalent to $LM^*_{\kappa_0}$ under the null and alternative hypotheses, therefore it is robust in the sense explained above and optimal when a $\sqrt{n}$-consistent estimator of θ is used.

In the following sections we fully explore the properties of the LM and Neyman's $C(\alpha)$ tests in their standard and robust variants.

5 Computation of the test statistics

5.1 MLE and score functions

The general log-likelihood function in this case is:

[5]As stated in Jaggia and Trivedi (1994), Neyman's $C(\alpha)$ tests are special cases of their conditional score tests.

(14)

$$L(\beta,\gamma,\sigma,\rho,\kappa) = \sum_{i=1}^{n} \mathbf{1}[c_i = 1]\left(ln(\Phi(\nu_i) + K_i) - \frac{u_i^2}{2} - \ln\sigma \right) + \mathbf{1}[c_i = 0]\left(ln(\Phi(-z_i\gamma) + \bar{K}_i) \right)$$

$$u_i = \frac{y_i - x_i\beta}{\sigma}; \nu_i = \frac{z_i\gamma + \rho u_i}{\sqrt{1-\rho^2}};$$

$$K_i = \sum_{3 \le r+s \le 4} \frac{\kappa_{rs}(-1)^{r+s}}{r!s!} \int_{-\infty}^{v_i} H_{rs}(u_i, e)\phi(e|u_i)de$$

$$\bar{K}_i = \sum_{3 \le r+s \le 4} \frac{\kappa_{rs}(-1)^{r+s}}{r!s!} \int_{-\infty}^{\infty}\int_{-\infty}^{v_i} H_{rs}(u, e)\phi(e|u)de$$

Note that the above formulation allows us to restrict our attention to univariate normal densities, which in turn simplifies the algebra considerably. The score functions evaluated at H_0 are easily obtained as:

$$(15) \quad \frac{\partial L}{\partial \beta}_{H_0} = \sum_{i=1}^{n} \mathbf{1}[c_i = 1] \frac{x_i}{\sigma(1-\rho^2)} \left(-\frac{\phi(\nu_i)}{\Phi(\nu_i)} \frac{\rho}{\sqrt{1-\rho^2}} + u_i \right)$$

$$(16) \quad \frac{\partial L}{\partial \gamma}_{H_0} = \sum_{i=1}^{n} \mathbf{1}[c_i = 1] z_i \left(\frac{\phi(\nu_i)}{\Phi(\nu_i)} \frac{\rho}{\sqrt{1-\rho^2}} \right) - \mathbf{1}[c_i = 0] z_i \left(\frac{\phi(\nu_i)}{\Phi(\nu_i)} \right)$$

$$(17) \quad \frac{\partial L}{\partial \sigma}_{H_0} = \sum_{i=1}^{n} \mathbf{1}[c_i = 1] \left[\frac{u_i}{\sigma} \left(-\frac{\phi(\nu_i)}{\Phi(\nu_i)} \frac{\rho}{\sqrt{1-\rho^2}} + u_i \right) - 1/\sigma \right]$$

$$(18) \quad \frac{\partial L}{\partial \rho}_{H_0} = \sum_{i=1}^{n} \mathbf{1}[c_i = 1] \frac{\phi(\nu_i)}{\sqrt{1-\rho^2}\Phi(\nu_i)} \left(u_i + \frac{\rho \nu_i}{1-\rho^2} \right)$$

$$(19) \quad \frac{\partial L}{\partial \kappa_{rs}}_{H_0} = \sum_{i=1}^{n} \sum_{3 \leq r+s \leq 4} \frac{(-1)^{r+s}}{r!s!} \mathbf{1}[c_i = 1] \int_{-\infty}^{v_i} H_{rs}(u_i, e) \phi(e|u_i) de$$
$$+ \frac{(-1)^{r+s}}{r!s!} \mathbf{1}[c_i = 0] \int_{-\infty}^{\infty} \int_{-\infty}^{v_i} H_{rs}(u, e) \phi(e|u) de$$

Appendix 2 derives explicit expressions for the conditional expectation of the Hermite polynomials. Following Lee (1984, ,p849–850) the score functions of the κ parameters are asymptotically equivalent to the difference between the estimated sample truncated moments of orders (r, s) with $r + s = 3, 4$ and the theoretical truncated moments of the bivariate normal distribution.

5.2 Information matrix estimates

A very important feature of the LM tests is how the information matrix (IM) is estimated. Consistent estimates can be obtained by the outer product gradient (OPG) method:

$$(20) \quad \hat{J}_{ab}^{OPG} = \frac{1}{n} \sum_{i=1}^{n} d_{ia}^{\top} d_{ib}, \ \ a, b = \eta, \kappa$$

A major drawback of this approach is that several Monte Carlo studies showed that models that use the expectation of (20) or evaluate the second derivative of the log-likelihood function (Hessian) under the null have a much better performance in terms of empirical size. We found that that the best results, both in terms of size and power, are obtained using the simulated expectation of J_{ab} where b random draws of size m are generated and the score functions are created for each (j) simulation (Godfrey and Orme, 2001). That is:

$$(21) \quad \hat{J}_{ab}^{(j)} = \frac{1}{m} \sum_{i=1}^{m} d_{ia}^{\top} d_{ib}, \ \ a, b = \eta, \kappa$$

$$(22) \quad \hat{J}_{ab}^{SIM} = \frac{1}{b} \sum_{j=1}^{b} \hat{J_{ab}^{(j)}}, \ \ a, b = \eta, \kappa$$

We consider random draws for $(u_i, e_i), i = 1, 2...m$, conditional on X and Z and based on our prior estimates of η. In turn, this is used to generate $y_i^{(j)}$ and $c_i^{(j)}$ from which (22) is generated.

6 Monte Carlo results

6.1 Baseline model

Our baseline model is similar to that of van der Klaauw and Koning (1993).

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

$$z_i^* = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 x_{i2} + e_i$$

$$c_i = \mathbf{1}[z_i^* > 0]$$

where

$$x_{i1}, x_{i2} \sim iid\ N(0, 3);$$

$$z_{i1} \sim iid\ Uniform(-3, 3);$$

$$\beta_0 = 1, \beta_1 = 0.5, \beta_2 = -0.5, \gamma_0 = 1, \gamma_1 = -1, \gamma_2 = 1;$$

$$VAR(u) = 4, VAR(e) = 1, CORR(u, e) = 0.2$$

Monte Carlo experiments for estimating the empirical size and power will be based on a sample size of $n = 1000$ (other sample sizes are available from the Author upon request). To simulate the information matrix we use $b = 100$, $m = n$. All Monte Carlo experiments use 1000 random sample generations. Rejection rates are based on a theoretical size of 5%.

We test that all the cumulants are zero (bivariate normality, H_0 above), that both marginal normality and conditional linearity are satisfied (H_0^{C1C2}), and these two hypotheses individually (H_0^{C1} and H_0^{C2} respectively). Robust tests are presented for the last three cases. We compute all tests under both MLE and two-step estimation procedures.

The estimation of the parameters is done in `STATA 9.1` by MLE and two-step procedures (`heckman` command). For numerical stability, the MLE estimator estimates $\arctan(\rho)$ instead of ρ and $\log(\sigma)$ instead of σ .

6.2 Empirical size and power

6.2.1 Empirical size

Table 1 presents the empirical size (at the 5% level) for standard and robust LM and $C(\alpha)$ tests. Using simulation for estimating the IM gets better size results than alternative methods (in particular OPG and Hessian methods, reported in an earlier version of the paper, and available from the Author upon request). The results show that testing for H_0^{C1} and H_0^{C2} separately results in a better size performance that the joint tests H_0 or H_0^{C1C2}. Moreover, only the robust variants of the tests show correct asymptotic size. Although not reported, similar results are observed for other sample sizes.

6.2.2 Power and size under local perturbations on selected parameters

The purpose of robust LM tests is to have similar power to non-robust variants, but smaller size on local departures from bivariate normality. In order to verify the properties of the tests, we generate observations from different BEE. In each case, a perturbation is applied to a selected parameter in κ. We report the empirical size and power obtained from samples generated using non-zero values of selected κ's, with domain on the set $\{0.2, 0.4, ...3\}$.

The samples were generated using the rejection method. As before the sample size is set to $n = 1000$, and for each parameter value we generate 1000 random samples. We report LM and $C(\alpha)$ tests in their standard and robust variants, where the first are applied to MLE and the second to two-step estimators.

In the first case, Figures 1 and 2, we apply some skewness on e (affecting κ_{03}) which should be detected by the H_0^{C1}-tests (tests for marginal normality of e, denoted by H_0' in the figure) statistics, but it should not affect H_0^{C2}-tests (tests for linearity of the conditional expectation, denoted by H_0'' in the figure). For all cases we observe that rejection rates increase with the size of the perturbation, but H_0^{C1}-tests do it considerably faster than H_0^{C2}-tests. Moreover, the robust variants have smaller rejection rates for both LM and $C(\alpha)$ tests, which determines that they have less power but better size. LM tests show higher power (although similar size) than the Neyman's $C(\alpha)$ tests.

In Figures 3 and 4, the DGP is constructed using different values of κ_{04}. As in the last case, we observe that rejection rates increase at higher speed for H_0^{C1}-tests than for H_0^{C2}-tests. However, only in the robust LM test we observe that the difference in the rejection rates between both types is enough to show that the test is successful in detecting the correct departure from bivariate normality.

When the samples are generated using non-zero values of κ_{12} (Figures 5 and 6), tests for H_0^{C2} should reject, while H_0^{C1}-tests should not. Again, all the test statistics are responsive to perturbations in κ_{12}, and only the robust variants show good size performance for H_0^{C1}-tests. In terms of power, LM tests overshadow $C(\alpha)$ tests for H_0^{C2}. However when perturbations in κ_{13} are applied (Figures 7 and 8) the tests do not reject even for high values of the cumulant parameter. This determines that the conditional linearity may only be detected in the direction of κ_{12}.

In the next figures, we apply some perturbation in those parameters that are not of interest either for H_0^{C1} nor H_0^{C2} tests. In particular we consider perturbation in κ_{30} (Figures 9 and 10) and κ_{21} (Figures 11 and 12)[6]. As expected all the tests are non-responsive to these DGPs, and $C(\alpha)$ tests show the best size performances, provided that the two-step estimation procedures are consistent even in the presence of perturbations in the nuisance parameters. Note that H_0^{C2}-tests are responsive to perturbations in κ_{21}, although the rejection rates are of no significant concern in the robust variant. In this case, the robust $C(\alpha)$ test has the best size performance.

6.2.3 Power and size under non-linear conditional expectation

An important feature of these tests, which cannot be found in other tests, is to detect departures from linearity in the conditional expectation. In order to observe the performance under non-linear conditional expectation, we evaluate an *ad-hoc* DGP with non-linear relationships between the error terms, satisfying the marginal normality of the selection equation error:

$$e = w_1$$

$$u = (e^2 - 1) + w_2$$

where w_1 and w_2 are independent standard normal random variables. Table 2 reports empirical results for this case. As in Table 1, we only observe a correct size for robust $C(\alpha)$ tests for marginal normality while every other variant over rejects considerably. However, the tests for non-linearity are successful in detecting it.

7 Empirical application

Lee (1984) studies the effect of being in a labor union on the workers

[6]Similar results are obtained for the remaining cumulants; available from the Author upon request.

wage. In this case, there is a strong selectivity for individuals who are in a union vs. those that are not affiliated. Using the OPG method to estimate the IM this author found strong evidence to reject H_0, that is the hypothesis that all cumulants are zero, although he cannot reject H_0^{C1}. Given our Monte Carlo results we may conclude that rejection of H_0 occurs too often and it may not be used as evidence to reject the use of two-step estimation methods.

In order to evaluate the performance of our testing procedures we use the well known Mroz (1987) database for studying the labor supply of married women. This database was widely used as a reference for many non-parametric and semi-parametric applications. For instance, Powell and Walker (1990) apply their semi-parametric procedure to this database. Moreover, the topic constitutes the most important application of selection models, since the original developments made by Heckman were intended to be applied here.

The sample consists of 753 women of whom 428 where working at the time of the study (the details on the construction of this database are in the Mroz paper). The dependent variable h is the annual hours of work and the regressors X included the logarithm of the wage rate (lw, assumed endogenous), family income less wife's labor income, indicators for young and older children in the family and the wife's number of years of age and education. The conditioning variables Z included the exogenous variables in X, plus years of labor experience, other background variables and various interaction terms (specified in Mroz, 1987, Table IX). Additionally actual experience was included as a exogenous variable in his Table X.

This model consists of three stages. First, a participation equation is estimated using the Z exogenous variables. Second, a wage-equation is estimated for working women only. Finally, the predicted wage is used in the hours -equation instead of lw. The paper compares two-stage least squares in the working women only sub-sample with selection models with different assumptions about the distributional properties of participation equation (using normal, logit and log-normal). Mroz concluded that failure to control by self-selection yields biased results if actual experience is included as an additional exogenous regressor, however there is not any evidence of biases if it is not included. In both cases, similar estimates are found across selection models with different distributional specifications.

The test statistics proposed in this paper are calculated for both the wage- and hours-equations. For the latter, we include the predicted value of the log-wage from the wage-equation. Test statistics appear in Tables 3 and 4- part A excludes actual experience, while part B has this variable as an additional exogenous regressor.

For both equations we cannot reject the marginal normality hypothesis except for the two-step estimaton procedures that produce extremely high values of the correlation parameter, which translate into very high values of the test statistics. In general, the robust variants of the tests show smaller values than the non-robust test statistics, and Neyman's $C(\alpha)$ tests show more consistent results across estimators than LM tests. These results are in line with Mroz (1987) and Powell and Walker (1990) assertions that normality in the participation equation cannot be rejected. Moreover, the tests for conditional linearity show that this hypothesis cannot be rejected. These test results are in line with Powell and Walker (1990) who find that semi-parametric estimators give similar estimates to that of the Heckman's two-step method.

8 Conclusions and suggestions for future research

We derived robust variants (in the sense that they have correct asymptotic size under local misspecification of the alternative hypothesis) of the Lee (1984) LM tests for distributional misspecification in sample selection models. Our purpose was to test separately the two conditions needed for applying two-step Heckman selection models. We also adjusted LM tests for the case where the two-step estimator is used instead of MLE, using Neyman's $C(\alpha)$ tests statistics.

Monte Carlo results show that bivariate normality is rejected too often, and therefore, testing fewer restrictions may provide better empirical size. Robust LM and Neyman's $C(\alpha)$ statistics show the best size performance for testing marginal normality of the selection equation error term and conditional linearity of the error terms.

We explore size and power properties of the tests when local perturbations are applied to selected parameters. In general robust Neyman's $C(\alpha)$ tests show good empirical size, but LM tests have better power performances. When perturbations on the skewness of the distribution of e are used, robust LM and $C(\alpha)$ tests statistics show good size performances, but LM procedures have better power properties. When the conditional linearity assumption is affected only the robust variant of the $C(\alpha)$ test have a good size performance. In this case, the test statistics are responsive only in one of the two directions considered in the paper.

We apply the tests procedures to the well-known Mroz (1987) database for women's labor supply decisions. Our results show that, in general, the selection equation's marginal normality and linearity of the conditional expectation of the error terms hypotheses cannot be rejected.

Robust LM tests provide a satisfactory procedure for testing distributional misspecification when the alternative hypothesis is not completely specified. Additional research is needed for studying the covariance structure of the Hermite polynomials in multivariate distributions, which may be useful for LM tests robustness under misspecified alternatives. On the other hand, more research on the estimation of BEE, other than SNP, is necessary to construct likelihood ratio and Wald tests, and to provide an efficient estimation procedure for this semi-parametric approach.

Table 1: Monte Carlo simulations - Empirical size

Test statistic	*Standard*				*Robust*		
Hypothesis	H_0	H_0^{C1C2}	H_0^{C1}	H_0^{C2}	H_0^{C1C2}	H_0^{C1}	H_0^{C2}
	MLE and LM tests						
ρ=-0.2	0.110	0.082	0.068	0.092	0.099	0.074	0.087
ρ=0.2	0.118	0.107	0.093	0.093	0.102	0.084	0.080
	Heckman's two-step and LM tests						
ρ=-0.2	0.137	0.121	0.078	0.125	0.100	0.068	0.105
ρ=0.2	0.148	0.141	0.097	0.13	0.131	0.070	0.108
	MLE and $C(\alpha)$ tests						
ρ=-0.2	0.088	0.091	0.034	0.103	0.103	0.036	0.076
ρ=0.2	0.085	0.094	0.062	0.103	0.195	0.053	0.099
	Heckman's two-step and $C(\alpha)$ tests						
ρ=-0.2	0.095	0.111	0.042	0.129	0.075	0.041	0.068
ρ=0.2	0.110	0.114	0.058	0.128	0.070	0.055	0.066

Notes: Theoretical size 0.05. Rejection rates based on 1000 replications. See text for details.

Table 2: Monte Carlo simulations - Empirical size and power against a bivariate distribution with non-linearity in the conditional expectation

Test statistic		*Standard*				*Robust*	
Hypothesis	H_0	H_0^{C1C2}	H_0^{C1}	H_0^{C2}	H_0^{C1C2}	H_0^{C1}	H_0^{C2}
				LM tests			
MLE	1.000	0.997	0.999	0.991	0.928	0.525	0.923
Heckman's two-step	1.000	0.912	0.913	0.903	0.901	0.402	0.896
				$C(\alpha)$ tests			
MLE	1.000	1.000	0.991	1.000	0.922	0.053	0.743
Heckman's two-step	1.000	0.960	0.923	0.967	0.938	0.052	0.731

Notes: Theoretical size 0.05. Rejection rates based on 1000 replications. See text for details.

Table 3: Empirical application - Tests for marginal normality

A-Without experience as exogenous covariate					
Estimator		LM		$C(\alpha)$	
		Standard	Robust	Standard	Robust
Wage Equation					
MLE	$\widehat{\rho} = 0.129$	3.66	4.87	3.12	3.30
Two-step	$\widehat{\rho} = 0.887$	440.3	246.8	284.6	19.96
Hours Equation					
MLE	$\widehat{\rho} = -0.161$	4.55	6.21	3.78	4.38
Two-step	$\widehat{\rho} = -0.275$	10.57	4.11	6.77	2.50
B-With experience as exogenous covariate					
Estimator		LM		$C(\alpha)$	
		Standard	Robust	Standard	Robust
Wage Equation					
MLE	$\widehat{\rho} = 0.091$	0.50	1.36	0.37	0.54
Two-step	$\widehat{\rho} = 0.731$	117.5	4.85	260.4	2.00
Hours Equation					
MLE	$\widehat{\rho} = -0.461$	9.60	9.26	13.49	3.00
Two-step	$\widehat{\rho} = -0.811$	431.1	24.5	401.9	4.67

Notes: Critical values for $\chi_2(2)$: 4.61 (10%), 5.99 (5%) and 9.21 (1%).

Table 4: Empirical application - Tests for conditional linearity

Estimator		LM		$C(\alpha)$	
		Standard	Robust	Standard	Robust
A-Without experience as exogenous covariate					
Wage Equation					
MLE	$\widehat{\rho} = 0.129$	0.34	4.56	0.39	0.78
Two-step	$\widehat{\rho} = 0.887$	469.9	2.46	301.6	2.38
Hours Equation					
MLE	$\widehat{\rho} = -0.161$	4.90	0.99	2.80	1.17
Two-step	$\widehat{\rho} = -0.275$	24.2	4.84	20.7	5.87
B-With experience as exogenous covariate					
Estimator		LM		$C(\alpha)$	
		Standard	Robust	Standard	Robust
Wage Equation					
MLE	$\widehat{\rho} = 0.091$	2.32	6.36	1.79	4.71
Two-step	$\widehat{\rho} = 0.731$	160.0	3.17	298.1	2.66
Hours Equation					
MLE	$\widehat{\rho} = -0.461$	21.2	3.28	103.5	6.80
Two-step	$\widehat{\rho} = -0.811$	522.6	4.60	473.7	4.62

Notes: Critical values for $\chi_2(2)$: 4.61 (10%), 5.99 (5%) and 9.21 (1%).

Figure 1: BEE with arbitrary values of κ_{03}-LM tests

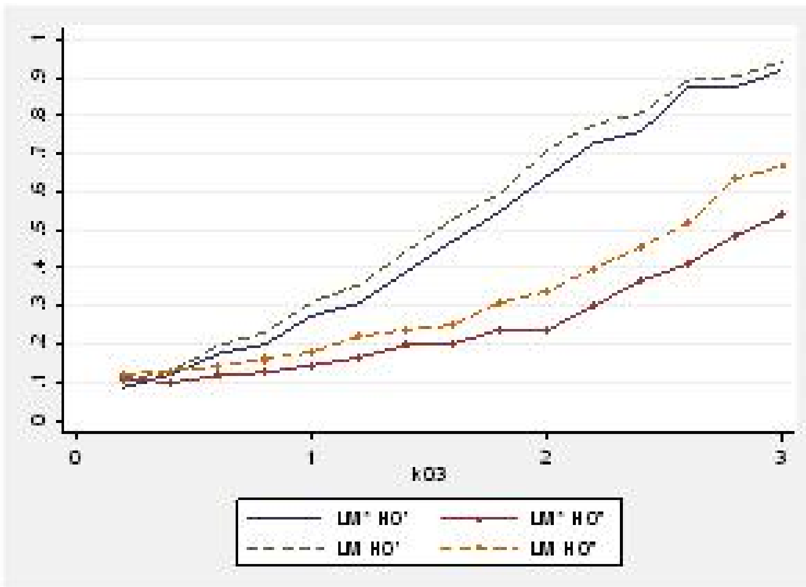


Figure 2: BEE with arbitrary values of κ_{03}-$C(\alpha)$ tests

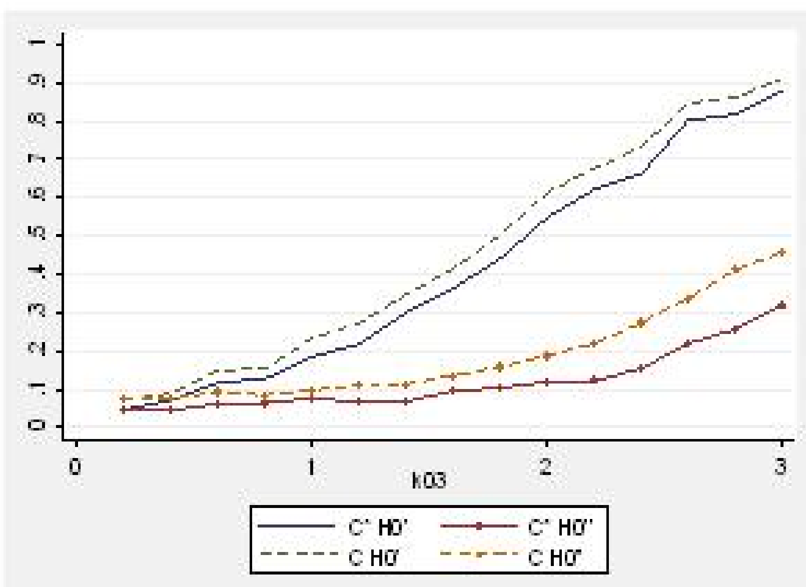


Figure 3: BEE with arbitrary values of κ_{04}-LM tests

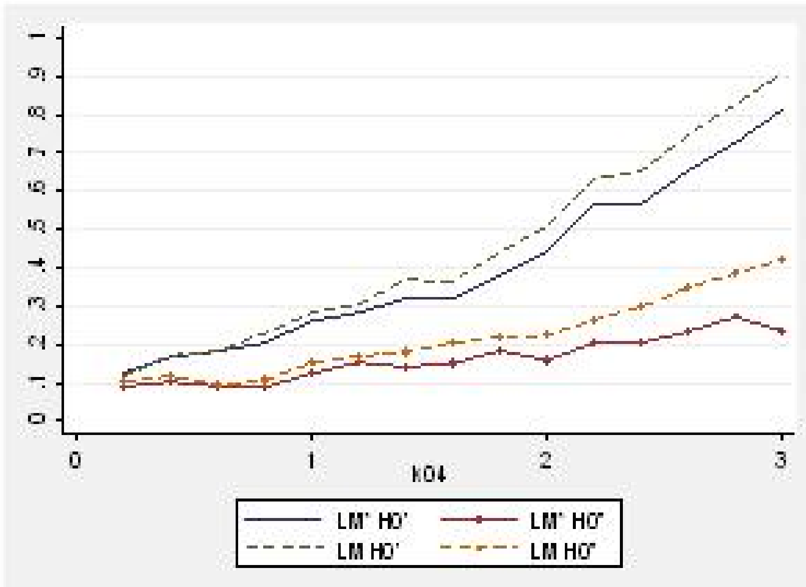


Figure 4: BEE with arbitrary values of κ_{04}-$C(\alpha)$ tests

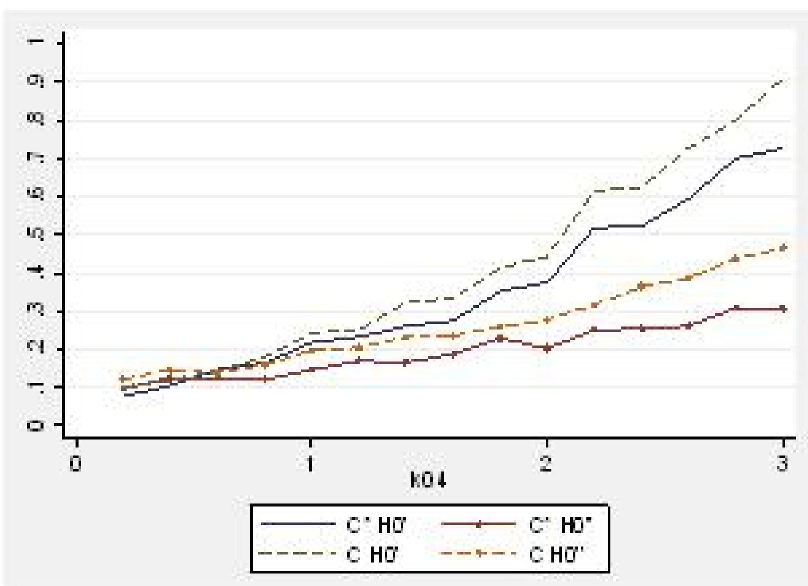


Figure 5: BEE with arbitrary values of κ_{12}-LM tests

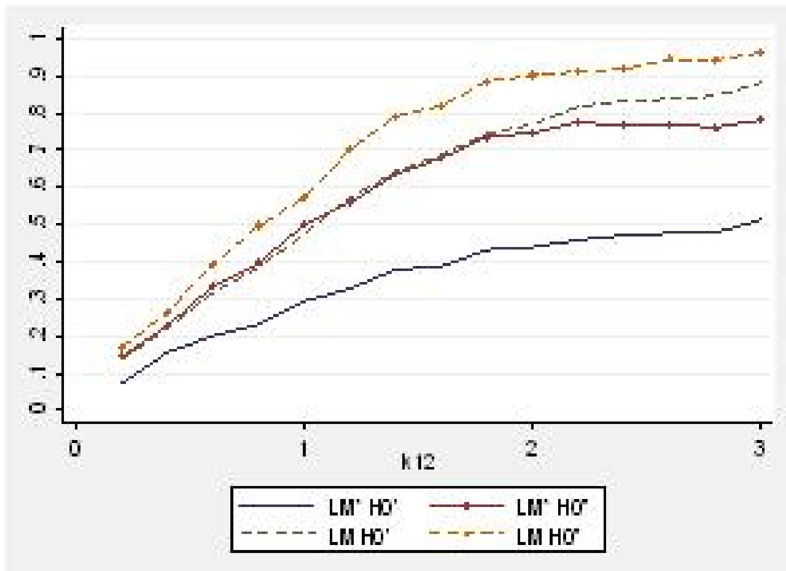


Figure 6: BEE with arbitrary values of κ_{12}-$C(\alpha)$ tests

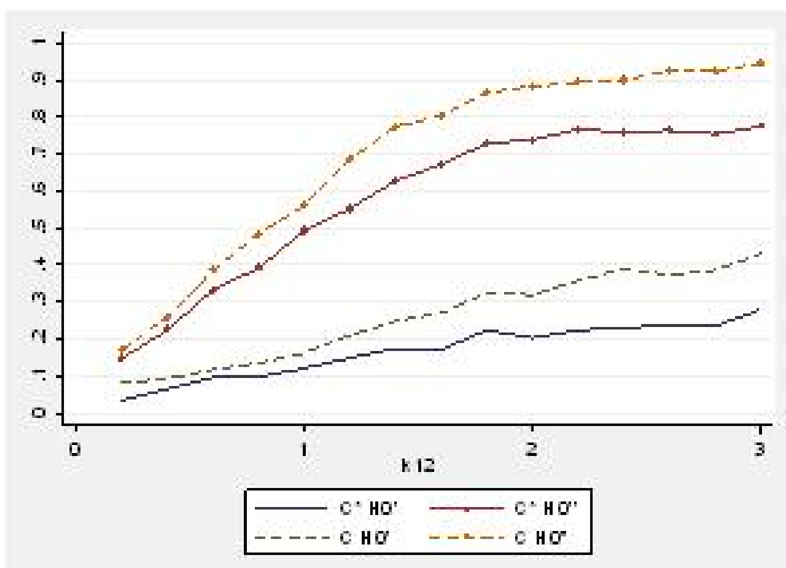


Figure 7: BEE with arbitrary values of κ_{13}-LM tests

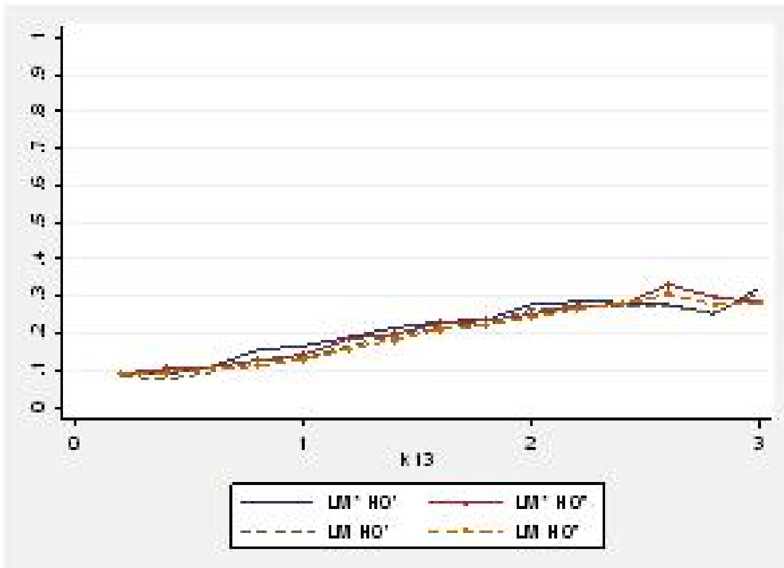


Figure 8: BEE with arbitrary values of κ_{13}-$C(\alpha)$ tests

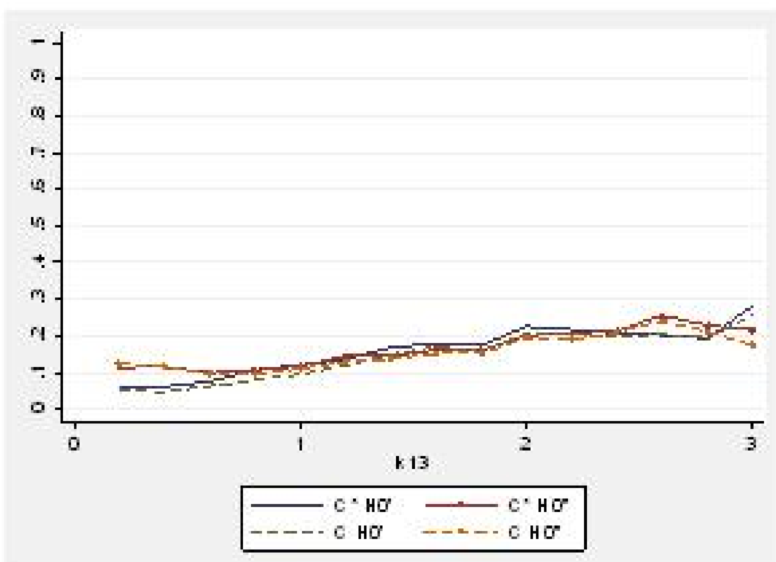


Figure 9: BEE with arbitrary values of κ_{30}-LM tests

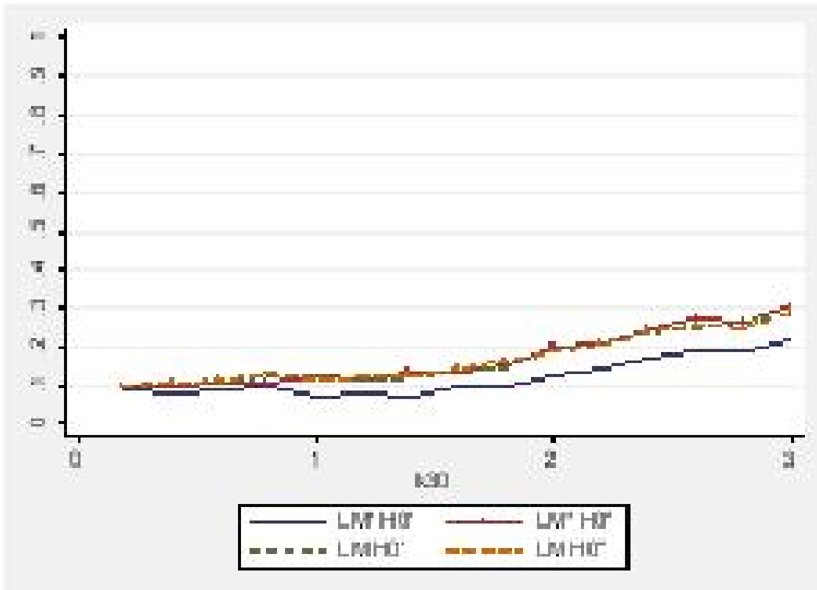


Figure 10: BEE with arbitrary values of κ_{30}-$C(\alpha)$ tests

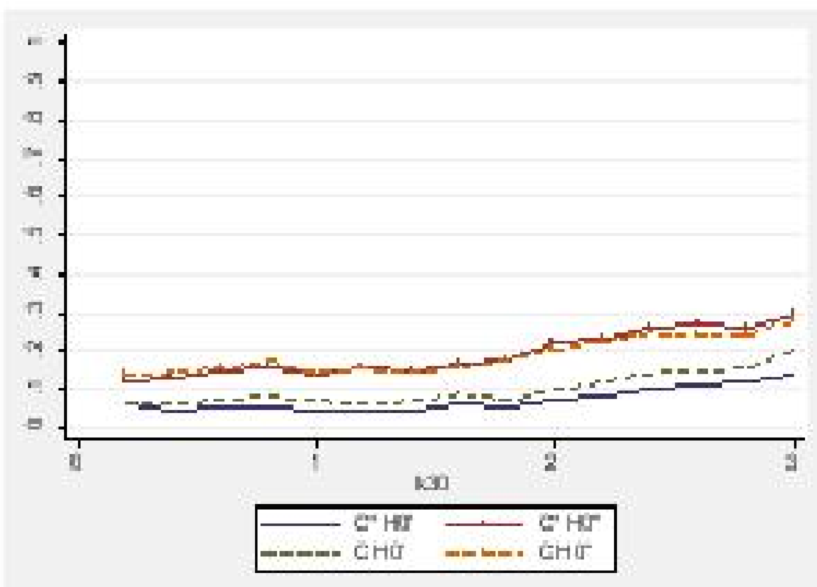


Figure 11: BEE with arbitrary values of κ_{21}-LM tests

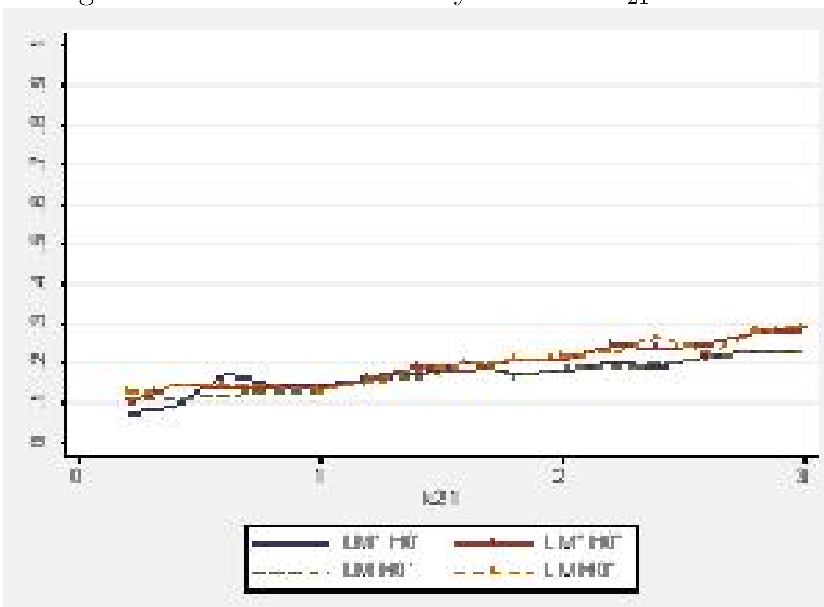
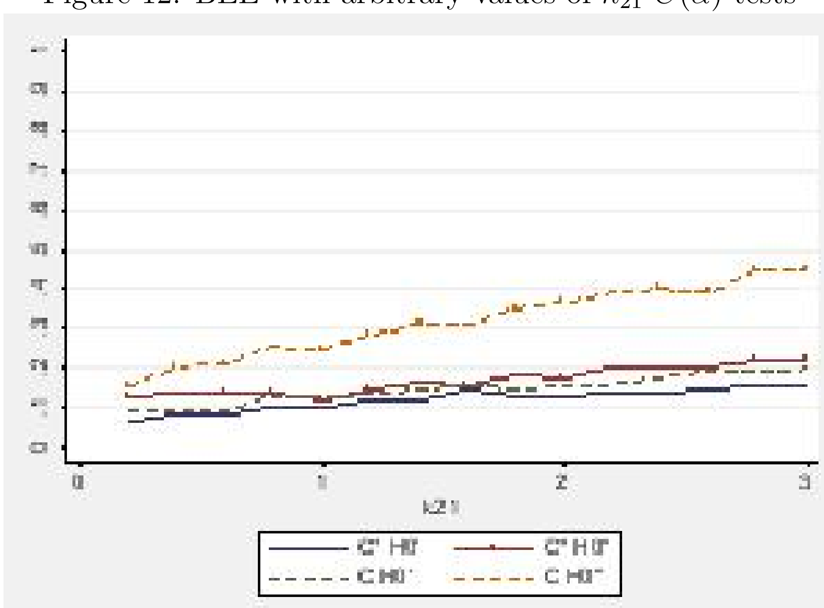


Figure 12: BEE with arbitrary values of κ_{21}-$C(\alpha)$ tests



Appendix 1: Bivariate Hermite polynomials

In this Appendix we derive the bivariate Hermite polynomials that correspond to $r + s \leq 4$. Let

$$H_{10}(u, e; \rho) = (-1)(u - \rho e)/(1 - \rho^2)$$
$$H_{01}(u, e; \rho) = (-1)(e - \rho u)/(1 - \rho^2)$$

be obtained by taking the first and second derivative of a bivariate normal distribution with respect to u and e. In a similar fashion:

Define

$$a = \frac{-1}{1 - \rho^2}$$
$$b = \frac{\rho}{1 - \rho^2}$$

Then the Hermite polynomials can be obtained by obtaining higher order derivatives.

$$H_{20}(u, e; \rho) = H_{10}(u, e; \rho)^2 + a$$
$$H_{02}(u, e; \rho) = H_{01}(u, e; \rho)^2 + a$$
$$H_{30}(u, e; \rho) = H_{10}(u, e; \rho)^3 + 3aH_{10}(u, e; \rho) + 3a^2$$
$$H_{03}(u, e; \rho) = H_{01}(u, e; \rho)^3 + 3aH_{01}(u, e; \rho) + 3a^2$$
$$H_{40}(u, e; \rho) = H_{10}(u, e; \rho)^4 + 6aH_{10}(u, e; \rho)^2 + 3a^2$$
$$H_{04}(u, e; \rho) = H_{01}(u, e; \rho)^4 + 6aH_{01}(u, e; \rho)^2 + 3a^2$$
$$H_{21}(u, e; \rho) = H_{01}(u, e; \rho)H_{10}(u, e; \rho)^2 + aH_{01}(u, e; \rho) + 2bH_{10}(u, e; \rho)$$
$$H_{12}(u, e; \rho) = H_{10}(u, e; \rho)H_{01}(u, e; \rho)^2 + a * H_{10}(u, e; \rho) + 2bH_{01}(u, e; \rho)$$
$$H_{31}(u, e; \rho) = H_{01}(u, e; \rho)H_{10}(u, e; \rho)^3 + 3aH_{01}(u, e; \rho)H_{10}(u, e; \rho) + 3bH_{10}(u, e; \rho)^2 + 3ab$$
$$H_{13}(u, e; \rho) = H_{10}(u, e; \rho)H_{01}(u, e; \rho)^3 + 3aH_{10}(u, e; \rho)H_{01}(u, e; \rho) + 3bH_{01}(u, e; \rho)^2 + 3ab$$

$$H_{22}(u,e;\rho) = H_{01}(u,e;\rho)^2 H_{10}(u,e;\rho)^2 + aH_{01}(u,e;\rho)^2 + aH_{10}(u,e;\rho)^2 + 3bH_{01}(u,e;\rho)H_{10}(u,e;\rho) + 2b^2 + a^2$$

Some useful algebra:

$$(u-\rho e)^2 = u^2 - 2\rho ue + \rho^2 e$$
$$(u-\rho e)^3 = u^3 - 3\rho u^2 e + 3\rho^2 ue^2 - \rho^3 e^3$$
$$(u-\rho e)^4 = u^4 - 4\rho u^3 e + 6\rho^2 u^2 e^2 - 4\rho^3 ue^3 + \rho^4 e^4$$
$$(e-\rho u)^2 = e^2 - 2\rho eu + \rho^2 u$$
$$(e-\rho u)^3 = e^3 - 3\rho e^2 u + 3\rho^2 eu^2 - \rho^3 u^3$$
$$(e-\rho u)^4 = e^4 - 4\rho e^3 u + 6\rho^2 e^2 u^2 - 4\rho^3 eu^3 + \rho^4 u^4$$
$$(u-\rho e)(e-\rho u) = -\rho u^2 + (1+\rho^2)ue - \rho e^2$$
$$(u-\rho e)^2(e-\rho u) = -\rho u^3 + (1+2\rho^2)u^2 e - \rho(2+\rho^2)e^2 u + \rho^2 e^3$$
$$(e-\rho u)^2(u-\rho e) = -\rho e^3 + (1+2\rho^2)e^2 u - \rho(2+\rho^2)u^2 e + \rho^2 u^3$$
$$(u-\rho e)^3(e-\rho u) = -\rho u^4 + (1+3\rho^2)u^3 e - 3\rho(1+\rho^2)u^2 e^2 + \rho^2(3+\rho^2)ue^3 - \rho^3 e^4$$
$$(e-\rho u)^3(u-\rho e) = -\rho e^4 + (1+3\rho^2)e^3 u - 3\rho(1+\rho^2)e^2 u^2 + \rho^2(3+\rho^2)eu^3 - \rho^3 u^4$$
$$(u-\rho e)^2(e-\rho u)^2 = \rho^2 u^4 - 2\rho^2(1+\rho^2)u^3 e + (1+4\rho^2+\rho^4)u^2 e^2 - 2\rho^2(1+\rho^2)ue^3 + \rho^2 e^4$$

Appendix 2: Truncated conditional moments

Denote $\phi(e|u) \equiv N(\rho u, 1-\rho^2)$. Then we have:

$$\frac{\partial \phi(e|u)}{\partial e} = -\frac{e-\rho u}{1-\rho^2}\phi(e|u) \tag{23}$$

By multiplying by $(1-\rho^2)e^{j-1}$, integrating by parts parts and rearranging terms we obtain the following recursive formulas:

$$\int_{-z\gamma}^{\infty} e^j \phi(e|u)de = \rho u \int_{-z\gamma}^{\infty} e^{j-1}\phi(e|u)de + (1-\rho^2)(j-1)\int_{-z\gamma}^{\infty} e^{j-2}\phi(e|u)de + (1-\rho^2)(-z\gamma)^{j-1}\phi(-z\gamma|u) \tag{24}$$

(25)

$$\int_{-\infty}^{-z\gamma} e^j \phi(e|u)de = \rho u \int_{-\infty}^{-z\gamma} e^{j-1}\phi(e|u)de + (1-\rho^2)(j-1)\int_{-\infty}^{-z\gamma} e^{j-2}\phi(e|u)de - (1-\rho^2)(-z\gamma)^{j-1}\phi(-z\gamma|u)$$

(24) corresponds to the truncation in the case of $c = 1$, and (25) to $c = 0$. Define:

$$(26) \quad \lambda_{u_i} = \frac{\phi(\nu_i)}{\Phi(\nu_i)\sqrt{1-\rho^2}}$$

where ν_i is defined as in (14). Therefore we have the following expressions for the truncated expectations:

$$E_i^*(e^j|u_i) \equiv E(e^j|u_i, e > -z_i\gamma), j = 1, 2...$$

$$E_i^*(e|u_i) = \rho u_i + (1-\rho^2)\lambda_{u_i}$$

$$E_i^*(e^2|u_i) = \rho u_i E_i^*(e|u_i) + (1-\rho^2) + (-z_i\gamma)\lambda_{u_i}$$

$$E_i^*(e^3|u_i) = \rho u_i E_i^*(e^2|u_i) + 2(1-\rho^2)E_i^*(e|u_i) + (1-\rho^2)(-z_i\gamma)^2\lambda_{u_i}$$

$$E_i^*(e^4|u_i) = \rho u_i E_i^*(e^3|u_i) + 3(1-\rho^2)E_i^*(e|u_i) + (1-\rho^2)(-z_i\gamma)^3\lambda_{u_i}$$

To evaluate the censored model we also need the double integration of the Hermite polynomials. Define:

$$(27) \quad \lambda_i = \frac{\phi(-z_i\gamma)}{\Phi(-z_i\gamma)}$$

Therefore we have:

$$E_{i*}(e^j u^h) \equiv E(e^j u^h|e < -z_i\gamma), j, h = 0, 1, 2...$$

$$E_{i*}(e) = \lambda_i$$

$$E_{i*}(e^2) = E_{i*}(e) + (-z_i\gamma)\lambda_i$$

$$E_{i*}(e^3) = 2E_{i*}(e) + (-z_i\gamma)^2\lambda_i$$

$$E_{i*}(e^4) = 3E_{i*}(e) + (-z_i\gamma)^3\lambda_i$$

$$E_{i*}(u) = \rho E_{i*}(e)$$

$$E_{i*}(u^2) = \rho^2 E_{i*}(e^2) + (1-\rho^2)$$

$$E_{i*}(u^3) = \rho^3 E_{i*}(e^3) + 3\rho(1-\rho^2)E_{i*}(e^3)$$

$$E_{i*}(u^4) = \rho^4 E_{i*}(e^4) + 6\rho^2(1-\rho^2)E_{i*}(e^2) + 3(1-\rho^2)^2$$

$$E_{i*}(eu) = \rho E_{i*}(e^2)$$

$$E_{i*}(ue^2) = \rho E_{i*}(e^3)$$

$$E_{i*}(ue^3) = \rho E_{i*}(e^4)$$

$$E_{i*}(u^2 e) = \rho^2 E_{i*}(e^3) + (1-\rho^2)E_{i*}(e)$$

$$E_{i*}(u^2 e^2) = \rho^2 E_{i*}(e^4) + (1-\rho^2)E_{i*}(e^2)$$

$$E_{i*}(u^3 e) = \rho^3 E_{i*}(e^4) + 3\rho(1-\rho^2)E_{i*}(e^2)$$

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), 563–578.

Bera, A. K. and S. John (1983). Tests for multivariate normality with Pearson alternatives. *Communications in Statistics: Theory and Methods 12*(1), 103–117.

Bera, A. K. and M. J. Yoon (1993). Specification testing with locally misspecified alternatives. *Econometric Theory 9*(4), 649–658.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the usa: a quantile regression approach. *Journal of Applied Econometrics 13*(1), 1–30.

Chambers, J. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika 54*(3-4), 367–383.

Gabler, S., F. Laisney, and M. Lechner (1993). Seminonparametric estimation of binary-choice models with an application to labor-force participation. *Journal of Business and Economic Statistics 11*(1), 61–81.

Gallant, A. R. and D. W. Nychka (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica 55*(2), 363–390.

Godfrey, L. and C. Orme (2001). On improving the robustness and reliability of Rao's score test. *Journal of Statistical Planning and Inference 97*(1), 153–176.

Heckman, J. J. (1974). Shadow wages, market wages and labor supply. *Econometrica 42*(4), 679–693.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and simple estimator for such models. *Annals of Economic and Social Measurement 5*(4), 475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Heckman, J. J. and S. Navarro-Lozano (2003). Simple estimators for treatment parameters in a latent-variable framework. *Review of Economics and Statistics 85*(3), 748–755.

Jaggia, S. and P. K. Trivedi (1994). Joint and separate score tests for state dependence and unobserved heterogeneity. *Journal of Econometrics 60*(1-2), 273–291.

Johnson, N. and S. Kotz (1972). *Continuous multivariate distributions.* New York: John Wiley and Sons.

Jondeau, E. and M. Rockinger (1994). Gram-Charlier densities. *Journal of Economic Dynamics and Control 25*(10), 1457–1483.

Lee, L.-F. (1982). Some approaches to the correction of selectivity bias. *Review of Economic Studies 49*(3), 355–372.

Lee, L.-F. (1984). Tests for the bivariate normal distribution in econometric models with selectivity. *Econometrica 52*(4), 843–863.

Mardia, K. V. (1970). *Families of Bivariate Distributions.* Darien, Connecticut: Hafner Publishing Company.

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

Ord, J. K. (1972). *Families of Frequency Distributions.* London: Griffin.

Powell, W. K. N. J. and J. Walker (1990). Semiparametric estimation of selection models: Some empirical results. *American Economic Review 80*(2), 324–328.

van der Klaauw, B. and R. H. Koning (1993). Testing the normality assumption in the sample selection model with and application to travel demand. *Journal of Business and Economic Statistics 21*(1), 31–42.