Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lecr20

Robust Misspecification Tests for the Heckman's Two-Step Estimator

Gabriel V. Montes-Rojas [a]
[a] Department of Economics, City University London, London, UK

Available online: 24 Jan 2011

To cite this article: Gabriel V. Montes-Rojas (2011): Robust Misspecification Tests for the Heckman's Two-Step Estimator, Econometric Reviews, 30:2, 154-172

To link to this article: http://dx.doi.org/10.1080/07474938.2011.534035

Econometric Reviews, 30(2):154–172, 2011

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2011.534035

ROBUST MISSPECIFICATION TESTS FOR THE HECKMAN'S TWO-STEP ESTIMATOR

Gabriel V. Montes-Rojas
Department of Economics, City University London, London, UK

□ *This article constructs and evaluates Lagrange multiplier (LM) and Neyman's $C(\alpha)$ tests based on bivariate Edgeworth series expansions for the consistency of the Heckman's two-step estimator in sample selection models, that is, for 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 testing marginal normality and linearity of the conditional expectations separately have a better size performance than testing bivariate normality. Moreover, the robust variants of the tests have better empirical size than nonrobust tests, which determines that these tests can be successfully applied to detect specific departures from the null model of bivariate normality. Finally, the tests are applied to women's labor supply data.*

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

JEL Classification C12; C24.

1. INTRODUCTION

Sample selection models are widely used in applied econometrics (see Heckman, 1974, 1976, 1979). These models account for the interaction of the error terms in an outcome and a selection equation. An efficient maximum likelihood (ML) estimator can be constructed if the error terms follow a bivariate normal distribution, and a test for the validity of this estimator is based on testing for bivariate normality (see for instance Bera and John, 1983, Bera et al., 1984). However, the Heckman's two-step estimator requires weaker assumptions for consistency: normality of the error term in the selection equation and linearity of the conditional expectation of the outcome equation error term conditional

Address correspondence to Gabriel V. Montes-Rojas, Department of Economics, City University London, Northampton Square, London EC1V 0HB, United Kingdom; E-mail: Gabriel.Montes-Rojas.1@city.ac.uk

on the selection equation one (see Newey, 1999, for a general discussion). The goal of this article is to devise a testing procedure for these two conditions while allowing some local flexibility on the bivariate distribution of the error terms.

Lagrange multiplier (LM) misspecification tests using bivariate Edgeworth series expansions (BEE) were first proposed by Lee (1982, 1984). In BEE, the data generating process (DGP) can be approximated by a basis distribution function (bivariate normal) and additional information about skewness and kurtosis (or higher order moments). Lee (1984) constructs LM tests based on a subset of parameters related to marginal normality and linearity of the conditional expectation. However, these LM tests may have incorrect asymptotic size if the remaining parameters in the BEE are not considered. We provide Monte Carlo evidence pointing out to the severity of over-rejection of these tests. Therefore, the first contribution of the article is to construct adjusted LM statistics that are locally size-robust (as in Bera and Yoon, 1993), that is, that have correct asymptotic size in the presence of local distributional misspecification. The second contribution is to compute Neyman's $C(\alpha)$ test statistics, which are optimal (i.e., they have asymptotic power equal to LM tests) for any $\sqrt{n}$-consistent estimator (e.g., Heckman's two-step estimator), and to introduce a locally size-robust variant of these to accommodate to local misspecification. Monte Carlo evidence shows that locally size-robust statistics based on the two-step estimator perform better than those using ML.

Gabler et al. (1993) and van der Klaauw and Koning (1993) proposed an alternative methodology for testing normality in sample selection models using the flexible semi-nonparametric (SNP) model introduced by Gallant and Nychka (1987). BEE has two main advantages over SNP. First, it can distinguish among different sources of distributional misspecification, i.e., non-normality of one error term or non-linearity of the conditional expectation. SNP procedures reject bivariate normality without any constructive indication on what to do next. Second, if there is any departure from the null DGP in the SNP model, some parameters (ρ and σ in the model below) cannot be, in general, identified. Moreover, the LM procedures do not require estimation of additional parameters as in the van der Klaauw and Koning (1993) likelihood ratio tests.

This article is organized as follows. Section 2 presents a typical sample selection model with censoring. In Section 3, bivariate Edgeworth expansions are reviewed. The test statistics are described in 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. Conclusions and suggestions for future research on this topic are in Section 8.

2. SELECTION MODELS

Consider the following standard sample selection model with censoring. Let i index the observations in a random sample with $i = 1, 2, \ldots, n$, and let the outcome equation be

$$y_i = x_i\beta + u_i, \tag{1}$$

where y is an outcome of interest, x is a set of covariates, u is an error term (with mean 0 and finite variance σ^2), and β is the main parameter of interest.

Consider a censoring mechanism where y is observed only if a certain event (indexed by the binary variable c) occurs (i.e., $c = 1$). Assume that

$$c_i = \begin{cases} 1 & \text{if} \quad z_i\gamma + e_i > 0 \\ 0 & \text{if} \quad z_i\gamma + e_i \leq 0 \end{cases}, \tag{2}$$

where z is another set of covariates, not necessarily disjoint from x, but that satisfies the necessary exclusion restrictions, and e is another error term (assumed to have mean 0 and variance 1).

The inconsistency of the least squares estimator of β in Eq. (1) is the result of the nonindependence of u and e. Assuming bivariate normality the conditional mean function is

$$E(y \mid c = 1) = x\beta + \rho\sigma\lambda(-z\gamma), \tag{3}$$

where $\lambda(e) = \frac{\phi(e)}{1-\Phi(e)}$ is the inverse Mills ratio, and $\Phi(.)$ and $\phi(.)$ are the distribution and density functions of a standard normal random variable, respectively. While the maximum likelihood estimator (MLE) is sensitive to any kind of distributional misspecification, the Heckman's two-step estimator is robust to distributional misspecification if and only if the two following conditions are met:

C1: e has marginal normal distribution;
C2: $E(u|e) = \rho\sigma e$, i.e., the conditional expectation is linear.

C1 is the censoring mechanism and C2 specifies how the selection mechanism affects the outcome equation through the conditional expectation of the error terms. The joint distribution of (e, u) may still satisfy those conditions but may not be bivariate normal, and therefore, testing for C1 and C2 is less restrictive than testing for bivariate normality.[1]

[1]As an example, assume that $e \sim N(0, 1)$ and $u = \rho \cdot e + \psi$, where ψ and e are independent, and $E(\psi) = 0$. Then, both conditions are satisfied, but bivariate normality is not, if ψ follows a non-Gaussian distribution.

If those conditions are not satisfied, alternative semiparametric estimators are available. If C1 is violated, γ can be estimated semi- or nonparametrically or by imposing other marginal distributions (Heckman et al., 2003). If C2 is violated, the outcome equation can be estimated by introducing additional terms (i.e., polynomials in the inverse Mill's ratio, as in Buchinsky, 1998) or by weighted least squares (see Newey et al., 1990). Other semiparametric procedures follow the SNP approach, where Hermite series expansions of the bivariate normal distribution allows for certain flexibility (van der Klaauw and Koning, 1993). However, semiparametric methods are less efficient and more computationally intensive than the Heckman's two-step method. Moreover, in general, only ML and Heckman's two-step estimators are available in software econometric packages. Therefore, testing C1 and C2 is important to avoid the potentially unnecessary cost of a semiparametric estimation.

3. BIVARIATE EDGEWORTH SERIES EXPANSIONS

Under some general conditions (see Chambers, 1967), the joint density function of (u, e), $g(u, e)$, can be expanded as a series of derivatives of the standard bivariate normal density function $\phi(u, e)$ (for simplicity, both u and e are assumed to have mean 0 and variance 1),

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

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 (BEE), which is a generalization of the univariate case, the Gram–Charlier and Edgeworth expansions. In this case, $\frac{\partial^{r+s}\phi(u,e)}{\partial u^r \partial e^s} = (-1)^{r+s} H_{rs}(u,e)\phi(u,e)$, where H_{rs} are the bivariate Hermite polynomials. These polynomials can be found in Mardia (1970), Johnson and Kotz (1972, Ch. 34) and Ord (1972, Appendix A).

Following Lee (1982, 1984), we only consider terms up to $r+s=4$ (i.e., assuming that the terms $r+s>4$ are zero), which is called a Type AA surface in Mardia (1970). For the type AA surface $A_{rs}=\kappa_{rs}$, where κ_{rs} denotes the cumulants of order (r, s). Note that κ_{rs} provides a measure of the DGP skewness and kurtosis. In particular, the parameters that satisfy $r+s=3$ determine skewness, and those with $r+s=4$ determine kurtosis.[2]

[2]For skewness: $\kappa_{30}=\mu_{30}$, $\kappa_{21}=\mu_{21}$, $\kappa_{12}=\mu_{12}$, and $\kappa_{03}=\mu_{03}$. For kurtosis: $\kappa_{40}=\mu_{40}-3$, $\kappa_{40}=\mu_{31}-3\rho$, $\kappa_{22}=\mu_{22}-2\rho^2-1$, $\kappa_{13}=\mu_{13}-3\rho$, and $\kappa_{04}=\mu_{04}-3$. Here we use the standard notation $\mu_{ij}=E(u^i e^j)$.

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

$$g(e) = \left[1 + \sum_{s=3}^{\infty} \frac{A_{0s}}{s!} H_{0s}(e)\right] \phi(e), \tag{5}$$

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

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

$$E(u|e) = \frac{\sum_{s=3}^{\infty}\left(\rho e + \frac{A_{0s}}{s!} H_{0,s+1}(e) + \frac{A_{1,s-1}}{(s-1)!} H_{0,s-1}(e)\right)}{1 + \sum_{s=3}^{\infty} \frac{A_{0s}}{s!} H_{0,s}(e)}, \tag{6}$$

where $\rho = corr(u, e)$. Under marginal normality of e,

$$E(u|e) = \rho e + \sum_{s=2}^{\infty}\left(\frac{A_{1,s}}{s!} H_{0,s}(e)\right). \tag{7}$$

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

Two major drawbacks of the BEE need to be mentioned. First, truncation of these series after a finite number of terms may produce negative values of the density function, and therefore, some restrictions on the multivariate series expansion are needed (Chambers, 1967; Jondeau and Rockinger, 2001). This is not, however, an issue for the LM tests developed here, as they only require estimation of the parameters under the null hypothesis DGP. Moreover, we will only consider local departures from the null DGP. Second, a feature of BEE is that fat tails and heavy skewness may not be captured appropriately.

4. STANDARD AND ROBUST LM AND NEYMAN'S $C(\alpha)$ TESTS

Lee (1984) derives LM tests for the bivariate normality of u and e, $H_0^{BN}: \kappa = 0$, where $\kappa = \{\kappa_{30}, \kappa_{21}, \kappa_{12}, \kappa_{03}, \kappa_{40}, \kappa_{31}, \kappa_{22}, \kappa_{13}, \kappa_{04}\}$, and for testing the marginal normality of e, $H_0^{C1}: \kappa_{03} = \kappa_{04} = 0$, testing conditional linearity, $H_0^{C2}: \kappa_{12} = \kappa_{13} = 0$, or both, $H_0^{C1C2}: \kappa_{03} = \kappa_{04} = \kappa_{12} = \kappa_{13} = 0$).[3]

[3]We consider the regularity conditions stated in Lee (1984, footnote 5, p. 854) to get the asymptotic distribution of the LM tests. Bera and Yoon (1993) statistics require the same conditions because they work under local departures from the joint null hypothesis.

For notational convenience define $\eta = \{\beta, \gamma, \sigma, \rho\}$; κ_0: cumulants of interest in either H_0^{BN}, H_0^{C1}, H_0^{C2} or H_0^{C1C2}; and κ_1: all κ except those in κ_0.

Note that, except for H_0^{BN}, the null hypotheses above do not completely specify the joint distribution of the error terms within the type AA surface. Moreover, the 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, 2001). Our strategy consists in testing either H_0^{C1}, H_0^{C2}, or H_0^{C1C2}, assuming that the DGP is bivariate normal (i.e., H_0^{BN}), because for this case it is straightforward to estimate η using MLE or the Heckman's two-step estimator.

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 nuisance parameters. Different tests can be obtained depending on: (i) what we assume about κ_1; and (ii) how η and κ_1 (if necessary) are estimated. 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.

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 $\hat{\eta}$ is the MLE estimator of η under bivariate normality. In that case, the LM test is

$$LM_{\kappa_0}(\hat{\theta}) = \frac{1}{n} d_{\kappa_0}(\hat{\theta})^\top \hat{J}_{\kappa_0 \cdot \eta}^{-1} d_{\kappa_0}(\hat{\theta}), \tag{8}$$

where for future reference

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

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

$$J_{a \cdot b}(\theta) = J_a - J_{ab} J_b^{-1} J_{ba} \text{ for } a, b = \eta, \kappa_0, \kappa_1,$$

and $\hat{J}$ is a consistent estimator of J, the information matrix (IM).

Under the corresponding null hypothesis, $H_0^{\kappa_0} : \kappa_0 = 0$, $LM_{\kappa_0} \Rightarrow \chi_p^2(0)$, where " $\Rightarrow$ " denotes asymptotic convergence in distribution and p is the number of parameters in the null hypothesis and the degrees of freedom of the chi-squared distribution. Under a sequence of local alternatives of the form $H_1^{\kappa_0}: \kappa_0 = \xi_0/\sqrt{n}$, $LM_{\kappa_0} \Rightarrow \chi_p^2(\lambda_0)$ where $\lambda_0 = \xi_0^\top J_{\kappa_0\cdot\eta}\xi_0$ denotes the non-centrality parameter of the chi-squared distribution. 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 $H_1^{\kappa_1}: \kappa_1 = \xi_1/\sqrt{n}$ and $LM_{\kappa_0} \Rightarrow \chi_p^2(\lambda_1)$ where $\lambda_1 = \xi_1^\top J_{\kappa_1\kappa_0\cdot\eta} J_{\kappa_0\cdot\eta}^{-1} J_{\kappa_0\kappa_1\cdot\eta}\xi_1$. 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 creates 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.[4]

If we restrict our attention to local misspecification of the type presented above, a "robust" LM test can be constructed as the berayoon93 adjusted LM tests or jaggiatrivedi94 conditional score tests

$$LM_{\kappa_0}^*(\hat{\theta}) = \frac{1}{n} d_{\kappa_0(\kappa_1)}(\hat{\theta})^\top \hat{J}_{\kappa_0\cdot\eta}^{-1} d_{\kappa_0(\kappa_1)}(\hat{\theta}), \tag{9}$$

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

In this case, under the null hypothesis $H_0^{\kappa_0}$ and a sequence of local alternatives in $H_1^{\kappa_1}$, $LM_{\kappa_0} \Rightarrow \chi_p^2(0)$, that is, the test statistic is locally size-robust under local misspecification of the unconsidered parameters. Under local alternatives of the form $H_1^{\kappa_0}$, $LM_{\kappa_0} \Rightarrow \chi_p^2(\lambda_2)$, where $\lambda_2 = \xi_0^{\mathrm{T}}(J_{\kappa_0\cdot\eta} - J_{\kappa_0\kappa_1\cdot\eta}J_{\kappa_1\cdot\eta}J_{\kappa_1\kappa_0\cdot\eta})\xi_0$. Since $\lambda_0 - \lambda_2 \geq 0$, the asymptotic power of this test will be lower than the LM test when there is no misspecification (this is the problem of *overtesting*). This test statistic is the robust (or adjusted) LM.

Let $\tilde{\theta} = (\tilde{\eta}, 0, 0)$ be a $\sqrt{n}$-consistent estimator under H_0^{C1C2}, different from the MLE. For our purposes, the Heckman's two-step estimator falls under this class of estimators. $LM_{\kappa_0}(\tilde{\theta})$ is no longer asymptotically chi-squared distributed, but it follows some other quadratic normal form, because the incorrect variance is used to normalize the score functions. Moreover, it will no longer be optimal, in the sense that it will have less power than $LM_{\kappa_0}(\hat{\theta})$. However, a simple adjustment of the scores given by the Neyman's $C(\alpha)$ tests have correct asymptotic size and power equivalent to that of $LM_{\kappa_0}(\hat{\theta})$[5]. The general form of this test is

$$C_{\kappa_0}(\tilde{\theta}) = \frac{1}{n} d_{\kappa_0\cdot\eta}(\tilde{\theta})^\top \hat{J}_{\kappa_0\cdot\eta}^{-1} d_{\kappa_0\cdot\eta}(\tilde{\theta}), \tag{10}$$

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

[5]See Bera and Bilias (2001) for an excellent discussion on Neyman's $C(\alpha)$ tests.

where $d_{\kappa_0\cdot\eta}(\theta) = d_{\kappa_0}(\theta) - \hat{J}_{\kappa_0\eta}\hat{J}_{\eta}^{-1}d_{\eta}(\theta)$ is the *effective score*. This test has the same asymptotic distribution as $LM_{\kappa_0}(\hat{\theta})$ under both $H_0^{\kappa_0}$ and $H_1^{\kappa_0}$.

Note the similarities with the robust LM test. In both cases, the nuisance parameters' influence has been taken out of the score functions of interest.[6] 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, i.e., $H_1^{\kappa_1}$. Unfortunately, additional assumptions are needed to obtain consistent estimates of the cumulants that do not belong to the null hypothesis.[7] For this reason, Neyman's $C(\alpha)$ tests may be futile to overcome the problem at hand of misspecification in κ_1. However, a simple extension of Bera and Yoon (1993) method provides a way of adjusting Neyman's $C(\alpha)$ tests for local misspecification of this type. Define

$$C^*_{\kappa_0}(\tilde{\theta}) = \frac{1}{n} d_{\kappa_0(\kappa_1)\cdot\eta}(\tilde{\theta})^\top \hat{J}^{-1}_{\kappa_0\cdot\eta} d_{\kappa_0(\kappa_1)\cdot\eta}(\tilde{\theta}), \tag{11}$$

where $d_{\kappa_0(\kappa_1)\cdot\eta}(\theta) = d_{\kappa_0\cdot\eta}(\theta) - \hat{J}_{\kappa_0\kappa_1\cdot\eta}\hat{J}^{-1}_{\kappa_1\cdot\eta}d_{\kappa_1\cdot\eta}(\theta)$ is the *adjusted effective score*. This test statistic is asymptotically equivalent to $LM^*_{\kappa_0}(\hat{\theta})$ under the null and alternative hypotheses; therefore, it is robust to the presence of $H_1^{\kappa_1}$.

5. COMPUTATION OF THE TEST STATISTICS

5.1. MLE and Score Functions

The log-likelihood function is

$$\begin{aligned} L(\beta, \gamma, \sigma, \rho, \kappa) = \sum_{i=1}^{n} 1[c_i = 1]\left(ln(\Phi(\nu_i) + K_i) - \frac{u_i^2}{2} - \ln\sigma \right) \\ + \mathbf{1}[c_i = 0](ln(\Phi(-z_i\gamma) + \bar{K}_i)), \end{aligned} \tag{12}$$

where $u_i = \frac{y_i - x_i\beta}{\sigma}$, $\nu_i = \frac{z_i\gamma + \rho u_i}{\sqrt{1-\rho^2}}$, $K_i = \sum_{3\le r+s\le 4} \frac{\kappa_{rs}(-1)^{r+s}}{r!s!}\int_{-\infty}^{v_i} H_{rs}(u_i, e)\phi(e|u_i)de$, and $\bar{K}_i = \sum_{3\le r+s\le 4}\frac{\kappa_{rs}(-1)^{r+s}}{r!s!}\int_{-\infty}^{\infty}\int_{-\infty}^{v_i} H_{rs}(u, e)\phi(e|u)de$.

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^{BN} are

$$\frac{\partial L}{\partial \beta}|_{H_0^{BN}} = \sum_{i=1}^{n}\mathbf{1}[c_i = 1]\frac{x_i}{\sigma(1-\rho^2)}\left(-\frac{\phi(\nu_i)}{\Phi(\nu_i)}\frac{\rho}{\sqrt{1-\rho^2}} + u_i\right), \tag{13}$$

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

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

$$\frac{\partial L}{\partial \gamma}|_{H_0^{BN}} = \sum_{i=1}^{n} \mathbf{1}[c_i = 1] z_i \left(\frac{\phi(\nu_i)}{\Phi(\nu_i)} \frac{\rho}{\sqrt{1-\rho^2}} \right) - \mathbf{1}[c_i = 0] z_i \left(\frac{\phi(\nu_i)}{\Phi(\nu_i)} \right), \tag{14}$$

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

$$\frac{\partial L}{\partial \rho}|_{H_0^{BN}} = \sum_{i=1}^{n} \mathbf{1}[c_i = 1] \frac{\phi(\nu_i)}{\sqrt{1-\rho^2}\Phi(\nu_i)} \left(u_i + \frac{\rho \nu_i}{1-\rho^2} \right), \tag{16}$$

$$\begin{aligned} \frac{\partial L}{\partial \kappa_{rs}}|_{H_0^{BN}} = \sum_{i=1}^{n} \sum_{3 \le r+s \le 4} \frac{(-1)^{r+s}}{r!s!} \mathbf{1}[c_i = 1] \int_{-\infty}^{v_i} H_{rs}(u_i, e) \phi(e|u_i) de \\ + \frac{(-1)^{r+s}}{r!s!} \mathbf{1}[c_i = 0] \int_{-\infty}^{\infty} \int_{-\infty}^{v_i} H_{rs}(u, e) \phi(e|u) de. \end{aligned} \tag{17}$$

Following Lee (1984, pp. 849–850) 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

The small sample performance of LM tests depends on how the IM is estimated. Consistent estimates can be obtained by the outer product gradient (OPG) method, that is, $\hat{J}_{ab}^{OPG} = \frac{1}{n}\sum_{i=1}^{n} d_{ia}^{\top} d_{ib}$, for $a, b = \eta, \kappa_0, \kappa_1$. However, several Monte Carlo studies showed that models that use the expectation of $\hat{J}_{ab}^{OPG}$ or evaluate the second derivative of the log-likelihood function (Hessian) under the null have a much better performance in terms of empirical size (see Godfrey and Orme, 2001, and the references therein).

In our case, the best results, both in terms of size and power, are obtained using the simulated expectation of the elements in J using a bootstrap procedure. We consider m random draws, indexed by (j), of a bivariate normal distribution $(u_i^{(j)}, e_i^{(j)})$, $i = 1, 2 \ldots n$, with parameters given by the estimates of ρ and σ (either ML or Heckman's two-step). In turn, this is used to generate $y_i^{(j)}$ and $c_i^{(j)}$ conditional on X and Z and based on the estimates of β and γ. That is, $y_i^{(j)} = \beta x_i + u_i^{(j)}$ and $c_i^{(j)} = \mathbf{1}[\gamma z_i + e_i^{(j)}]$. Then, for each (j), re-estimate the parameters in η (either ML or Heckman's two-step, $\hat{\eta}^{(j)}$) and compute

$$\hat{J}_{ab}^{(j)} = \frac{1}{n} \sum_{i=1}^{n} d_{ia}(\hat{\theta}^{(j)}) d_{ib}(\hat{\theta}^{(j)})^{\top}, \quad \text{for } a, b = \eta, \kappa_0, \kappa_1, \tag{18}$$

where $\hat{\theta}^{(j)} = (\hat{\eta}^{(j)}, 0, 0)$. Finally, the estimate of J is given by

$$\hat{J}_{ab}^{SIM} = \frac{1}{m} \sum_{j=1}^{m} \hat{J}_{ab}^{(j)}, \text{for } a, b = \eta, \kappa_0, \kappa_1. \tag{19}$$

6. MONTE CARLO RESULTS

6.1. Baseline Model

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

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i, z_i^* = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 x_{i2} + e_i, c_i = \mathbf{1}[z_i^* > 0],$$

where $x_{i1}, x_{i2} \sim iid\, N(0, 3)$, $z_{i1} \sim iid\, Uniform(-3, 3)$, $\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = -0.5$, $\gamma_0 = 1$, $\gamma_1 = -1$, $\gamma_2 = 1$, $Var(u) = 4$, $Var(e) = 1$, and $Corr(u, e) = \rho$.

Monte Carlo experiments will be based on different sample sizes of $n = 200$, 500, 1000 and correlation parameter values $\rho = 0.2$, 0.4, 0.6, 0.8. To simulate the IM, we use $m = 500$. Rejection rates are based on 5% and 10% theoretical sizes.

We consider standard and robust LM and Neyman's C(α) tests for the following hypothesis: H_0^{C1}: tests for marginal normality of e; H_0^{C2}: tests for linearity of the conditional expectation of u given e; H_0^{C1C2}: tests for the validity of both conditions, i.e., the assumptions needed for the Heckman's two-step estimator; and H_0^{BN}: tests for bivariate normality. The estimation of the parameters is done in STATA 10.1 (heckman command). For numerical stability, the MLE estimator computes arctan(ρ) instead of ρ, and log(σ) instead of σ.

6.2. Empirical Size and Power

6.2.1. Empirical Size

Tables 1 and 2 present empirical size simulations at the 5% and 10% theoretical sizes, respectively. LM tests have increasing over-rejection as the correlation parameter approaches to one. These are particularly severe for testing conditional linearity (H_0^{C2}), even for a large sample size (n = 1000). Neyman's C(α) tests show better size performance for large correlation parameter values and small sample sizes, although rejection rates in this case are also above the theoretical level. In all cases, the Bera and Yoon (1993) robustification procedure reduces the over-rejection significantly. Moreover, the latter test statistics show that the χ^2 approximation in small samples is relatively accurate for both the 5% and 10% significance levels.

Overall, 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^{C1C2} and H_0^{BN}. In all cases, tests for H_0^{BN} show clear over–rejection. Moreover, joint tests for the validity of the Heckman's two step estimator show rejections rates slightly above those of the tests for H_0^{C2}. In this case, as before, robust tests have a better

TABLE 1 Monte Carlo simulations: Empirical size. DGP: bivariate normal. Theoretical size 5%

		n			n			n			n		
		200	500	1000	200	500	1000	200	500	1000	200	500	1000
	ρ	$LM_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1}	0.2	0.072	0.059	0.052	0.062	0.054	0.052	0.068	0.046	0.046	0.055	0.052	0.049
	0.4	0.098	0.088	0.087	0.065	0.056	0.055	0.076	0.054	0.051	0.065	0.052	0.052
	0.6	0.121	0.110	0.099	0.068	0.059	0.056	0.078	0.072	0.068	0.067	0.057	0.051
	0.8	0.133	0.128	0.112	0.078	0.062	0.061	0.083	0.075	0.072	0.070	0.067	0.056
	ρ	$LM_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$		
H_0^{C2}	0.2	0.098	0.072	0.062	0.072	0.065	0.055	0.077	0.068	0.056	0.042	0.040	0.039
	0.4	0.121	0.096	0.077	0.084	0.076	0.057	0.085	0.074	0.065	0.051	0.052	0.049
	0.6	0.175	0.145	0.120	0.099	0.087	0.065	0.099	0.092	0.083	0.060	0.053	0.052
	0.8	0.268	0.230	0.201	0.113	0.090	0.070	0.111	0.102	0.089	0.066	0.057	0.055
	ρ	$LM_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1C2}	0.2	0.091	0.059	0.052	0.069	0.060	0.059	0.080	0.069	0.060	0.055	0.050	0.048
	0.4	0.128	0.088	0.087	0.074	0.069	0.061	0.091	0.087	0.076	0.061	0.059	0.053
	0.6	0.151	0.135	0.120	0.102	0.093	0.066	0.106	0.098	0.080	0.068	0.059	0.052
	0.8	0.250	0.226	0.210	0.123	0.092	0.078	0.135	0.116	0.117	0.076	0.067	0.060
	ρ	$LM_{\kappa}(\hat{\theta})$						$C_{\kappa}(\tilde{\theta})$					
H_0^{BN}	0.2	0.242	0.150	0.135				0.162	0.128	0.106			
	0.4	0.366	0.251	0.220				0.250	0.187	0.150			
	0.6	0.405	0.332	0.297				0.369	0.292	0.262			
	0.8	0.449	0.370	0.354				0.399	0.351	0.317			

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

performance. These results suggest that increasing the number of parameters being tested increases the size unnecessarily.

6.2.2. Power and Size Under Nonlinear Conditional Expectation

An important feature of these tests is to detect departures from linearity in the conditional expectation (H_0^{C2}). In order to observe the performance under nonlinear conditional expectation, we evaluate an ad-hoc DGP with a nonlinear relationship between the error terms, but still satisfying the marginal normality of the selection equation error

$$e = w_1,\ u = 2 * \rho\ e + (e^2 - 1) + \sqrt{4 - 2 - (2 * \rho)^2}\, w_2,$$

where w_1 and w_2 are independent standard normal random variables.[8] Table 3 reports these simulations.

The tests for marginal normality are affected by the nonlinearity in the conditional expectation. Note that increasing the sample size, n, does not solve the over–rejection problem. The simulations show that using Bera

[8]Note that we can only evaluate $\rho = 0.2,\ 0.4,\ 0.6$.

TABLE 2 Monte Carlo simulations: Empirical size. DGP: bivariate normal. Theoretical size 10%

		n			n			n			n		
		200	500	1000	200	500	1000	200	500	1000	200	500	1000
	ρ	$LM_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1}	0.2	0.145	0.131	0.121	0.129	0.113	0.109	0.136	0.130	0.119	0.105	0.098	0.094
	0.4	0.185	0.156	0.087	0.129	0.120	0.110	0.150	0.130	0.121	0.130	0.125	0.109
	0.6	0.201	0.160	0.123	0.138	0.124	0.114	0.156	0.136	0.126	0.136	0.057	0.110
	0.8	0.245	0.187	0.156	0.156	0.134	0.129	0.162	0.141	0.133	0.142	0.132	0.124
	ρ	$LM_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$		
H_0^{C2}	0.2	0.182	0.165	0.148	0.144	0.131	0.120	0.157	0.132	0.102	0.101	0.094	0.088
	0.4	0.225	0.187	0.077	0.165	0.146	0.125	0.166	0.153	0.116	0.105	0.098	0.092
	0.6	0.278	0.215	0.192	0.199	0.169	0.130	0.178	0.163	0.132	0.111	0.102	0.099
	0.8	0.341	0.300	0.261	0.215	0.154	0.133	0.223	0.182	0.149	0.130	0.114	0.106
	ρ	$LM_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1C2}	0.2	0.195	0.172	0.150	0.153	0.136	0.128	0.170	0.151	0.126	0.113	0.109	0.102
	0.4	0.236	0.199	0.087	0.174	0.165	0.129	0.184	0.167	0.136	0.119	0.108	0.105
	0.6	0.295	0.241	0.120	0.209	0.178	0.135	0.200	0.178	0.154	0.121	0.109	0.105
	0.8	0.370	0.312	0.290	0.223	0.189	0.150	0.245	0.197	0.164	0.142	0.133	0.125
	ρ	$LM_{\kappa}(\hat{\theta})$						$C_{\kappa}(\tilde{\theta})$					
H_0^{BN}	0.2	0.306	0.247	0.203				0.251	0.189	0.150			
	0.4	0.366	0.302	0.249				0.306	0.257	0.205			
	0.6	0.475	0.395	0.302				0.458	0.391	0.275			
	0.8	0.549	0.432	0.379				0.511	0.450	0.353			

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

and Yoon (1993) procedure and Neyman's C(α) tests together with the Heckman's two step estimator in particular, that is $C^*_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$, robustify the statistic against misspecification in the cumulants not being tested.

The power for detecting nonlinearity, H_0^{C2}, increases monotonically as both n and ρ increase. Taken together with the empirical size under bivariate normality, and in particular with the over–rejection cases for the LM and Neyman's C(α), the robust versions $LM^*_{\kappa_{12}\kappa_{13}}(\hat{\theta})$ and $C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$ show good power properties. Finally, tests for both conditions, H_0^{C1C2}, show similar power to the individual tests for H_0^{C2}.

6.2.3. Power and Size Under Nonmarginal Normality

As discussed above, the validity of the Heckman's two-step estimator also depends on the marginal normality of the error term in the selection equation (H_0^{C1}). Note that tests for this condition can be constructed on the selection equation separately, as for instance, tests for distributional misspecification of the *probit* model. Nevertheless, for completeness, we also evaluate the validity of H_0^{C1} using the BEE approach proposed in this article. Following van der Klaauw and Koning (1993), we consider two bivariate processes that satisfy H_0^{C2} (see that article for details on the construction of these DGPs): a bivariate t-Student distribution process with

TABLE 3 Monte Carlo simulations: Empirical size and power. DGP: bivariate distribution with non-linearity in conditional expectation. Theoretical size 5%

		n			n			n			n		
		200	500	1000	200	500	1000	200	500	1000	200	500	1000
	ρ	$LM_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1}	0.2	0.272	0.204	0.153	0.108	0.100	0.086	0.253	0.191	0.130	0.054	0.048	0.037
	0.4	0.307	0.260	0.199	0.145	0.120	0.099	0.318	0.274	0.205	0.084	0.068	0.042
	0.6	0.383	0.297	0.252	0.169	0.147	0.105	0.374	0.285	0.245	0.102	0.077	0.056
	ρ	$LM_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$		
H_0^{C2}	0.2	0.316	0.395	0.485	0.121	0.301	0.453	0.295	0.446	0.500	0.094	0.370	0.464
	0.4	0.522	0.768	0.932	0.211	0.559	0.906	0.491	0.637	0.808	0.125	0.477	0.788
	0.6	0.759	0.985	1.000	0.286	0.815	0.994	0.470	0.792	0.941	0.116	0.521	0.888
	ρ	$LM_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1C2}	0.2	0.345	0.421	0.416	0.225	0.406	0.416	0.270	0.425	0.471	0.150	0.404	0.471
	0.4	0.523	0.754	0.920	0.321	0.725	0.920	0.366	0.609	0.787	0.187	0.570	0.785
	0.6	0.761	0.983	1.000	0.434	0.938	0.999	0.450	0.770	0.943	0.227	0.723	0.942
	ρ	$LM_{\kappa}(\hat{\theta})$						$C_{\kappa}(\tilde{\theta})$					
H_0^{BN}	0.2	0.892	0.955	1.000				0.780	0.881	0.978			
	0.4	0.998	0.992	1.000				0.863	0.899	0.993			
	0.6	0.998	1.000	1.000				0.985	0.998	1.000			

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

TABLE 4 Monte Carlo simulations: Empirical size. DGP: bivariate t-Student. Theoretical size 5%

		n			n			n			n		
		200	500	1000	200	500	1000	200	500	1000	200	500	1000
	ρ	$LM_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1}	0.2	0.556	0.700	0.842	0.405	0.650	0.833	0.543	0.646	0.846	0.465	0.622	0.800
	0.4	0.598	0.782	0.893	0.579	0.777	0.856	0.603	0.688	0.790	0.569	0.652	0.769
	0.6	0.640	0.942	0.998	0.631	0.935	0.980	0.658	0.785	0.945	0.616	0.618	0.782
	0.8	0.850	0.999	1.000	0.846	0.999	1.000	0.852	0.998	1.000	0.816	0.947	0.999
	ρ	$LM_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$		
H_0^{C2}	0.2	0.198	0.175	0.156	0.107	0.095	0.072	0.077	0.068	0.056	0.042	0.040	0.039
	0.4	0.251	0.190	0.160	0.119	0.103	0.082	0.085	0.074	0.065	0.051	0.052	0.049
	0.6	0.306	0.255	0.225	0.132	0.107	0.099	0.099	0.092	0.083	0.060	0.053	0.052
	0.8	0.369	0.332	0.295	0.154	0.126	0.107	0.111	0.102	0.089	0.066	0.057	0.055
	ρ	$LM_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1C2}	0.2	0.602	0.799	0.956	0.405	0.650	0.833	0.743	0.902	0.990	0.662	0.890	0.903
	0.4	0.706	0.868	0.999	0.579	0.777	0.856	0.850	0.988	0.997	0.801	0.952	0.965
	0.6	0.887	0.975	1.000	0.631	0.935	0.980	0.953	1.000	1.000	0.616	0.618	0.782
	0.8	0.952	0.999	1.000	0.846	0.999	1.000	0.995	1.000	1.000	0.816	0.947	0.999
	ρ	$LM_{\kappa}(\hat{\theta})$						$C_{\kappa}(\tilde{\theta})$					
H_0^{BN}	0.2	0.906	0.967	0.971				0.853	0.899	0.959			
	0.4	0.977	0.982	1.000				0.886	0.931	0.981			
	0.6	0.999	0.998	1.000				0.902	0.982	0.999			
	0.8	1.000	1.000	1.000				0.963	0.995	1.000			

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

TABLE 5 Monte Carlo simulations: Empirical size. DGP: bivariate χ^2. Theoretical size 5%

		n			n			n			n		
		200	500	1000	200	500	1000	200	500	1000	200	500	1000
	ρ	$LM_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1}	0.2	0.781	0.877	0.992	0.603	0.741	0.852	0.667	0.774	0.841	0.467	0.555	0.621
	0.4	0.882	0.969	0.999	0.764	0.851	0.932	0.754	0.837	0.906	0.522	0.600	0.663
	0.6	0.953	0.992	1.000	0.840	0.935	0.990	0.823	0.893	0.947	0.697	0.759	0.767
	0.8	1.000	1.000	1.000	0.899	0.955	0.995	0.909	0.995	1.000	0.874	0.902	0.911
	ρ	$LM_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$		
H_0^{C2}	0.2	0.201	0.212	0.230	0.169	0.162	0.154	0.251	0.225	0.201	0.162	0.150	0.131
	0.4	0.252	0.249	0.256	0.184	0.179	0.180	0.260	0.231	0.210	0.159	0.143	0.141
	0.6	0.261	0.270	0.271	0.232	0.206	0.199	0.285	0.281	0.274	0.189	0.163	0.146
	0.8	0.333	0.299	0.291	0.246	0.221	0.195	0.320	0.299	0.271	0.202	0.192	0.185
	ρ	$LM_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$LM^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta})$			$C_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$			$C^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta})$		
H_0^{C1C2}	0.2	0.972	0.961	0.991	0.771	0.782	0.807	0.884	0.906	0.899	0.661	0.720	0.709
	0.4	0.980	0.976	0.989	0.847	0.850	0.863	0.913	0.933	0.940	0.762	0.803	0.822
	0.6	0.999	1.000	1.000	0.901	0.935	0.990	0.953	0.974	1.000	0.883	0.875	0.935
	0.8	1.000	1.000	1.000	1.000	1.000	1.000	0.972	0.969	1.000	0.901	0.938	0.968
	ρ	$LM_{\kappa}(\hat{\theta})$						$C_{\kappa}(\tilde{\theta})$					
H_0^{BN}	0.2	1.000	1.000	1.000				1.000	1.000	1.000			
	0.4	1.000	1.000	1.000				1.000	1.000	1.000			
	0.6	1.000	1.000	1.000				1.000	1.000	1.000			
	0.8	1.000	1.000	1.000				1.000	1.000	1.000			

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

3 degrees of freedom (see Table 4) and a bivariate χ^2 with 1 degree of freedom (see Table 5).

Tests for marginal normality have considerable power for detecting non-Gaussian DGPs. As before, the rejection rates increase both with n and ρ. Moreover, robust LM variants of the tests for conditional linearity show rejection rates closer to the theoretical size than its nonrobust variants. However, significant over–rejection is observed in the χ^2 DGP. The best size performance is observed for the $C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta})$ test statistic.

7. EMPIRICAL APPLICATION

In order to evaluate the performance of the proposed tests, we use the well-known Mroz (1987) database for studying the labor supply of married women.[9] This database was widely used as a reference for many

[9]Lee (1984) studies the effect of being in a labor union on the workers 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^{BN} and H_0^{C2}, 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 rates in standard LM tests occur too often and therefore, they may not be used as evidence to reject the validity of two-step estimation methods.

nonparametric and semiparametric applications. For instance, Newey et al. (1990) apply their semiparametric procedure to this database. Moreover, the topic constitutes the most important application of sample selection models, since the original developments made by Heckman were intended to be applied here.

The sample consists of 753 women, of whom 428 were working at the time of the study. 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 (nwifeinc), indicators for young and old children in the family (kidslt6: number of kids less than 6 years old; kidsge6: number of kids with at least 6 years old) and the wife's years of age and education. The conditioning variables *Z* included the exogenous variables in *X*, plus years of labor experience.

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*. Mroz's article compares different selection models with different assumptions about the distributional properties of participation equation (using normal, logit, and log-normal distributions). He concluded that failure to control for self-selection yields biased results.

The test statistics proposed in this article are calculated for both wage and hours equation. Regression results and the test statistics developed in this article appear in Tables 6 and 7 for the wage and hours equation, respectively.

The wage equation results in Table 6 show that the estimated correlation coefficient is very low (0.028 in MLE; 0.065 in the Heckman's two-step estimator). As a result, the selection equation has no significant effect in the wage-outcome equation, and therefore, both MLE and Heckman's two-step estimator are similar to OLS. Moreover, the LM statistics developed in this article cannot reject the validity of the normality and conditional linearity assumptions.

The hours equation, however, show very different results for ML and Heckman's two-step estimators. The MLE of the correlation parameter is -0.28, which contrasts sharply with the two-step estimate of -0.80. As a result, regression coefficients differ as well. Of particular importance is the estimated coefficient of the effect of wages on hours worked. MLE estimates suggest that increasing wages by 1% would increase hours worked by 11 hours, which is statistically significant. However, Heckman's two-step procedure shows that the effect is just 2.5 hours, but not statistically significant. The test for marginal normality, H_0^{C1}, cannot reject the normality of the error term in the selection equation. Note that in this case, the nonrobust variant of the C(α) test calculated with the two-step estimator has a very high value, certainly due to the high value of the estimated correlation

TABLE 6 Empirical application: Wage equation

	OLS	MLE	Heckman's two step
		Outcome equation	
educ	0.107 (0.014)	0.108 (0.015)	0.110 (0.016)
age	0.0003 (0.0049)	−0.0001 (0.0053)	−0.0007 (0.0060)
exper	0.042 (0.013)	0.043 (0.015)	0.045 (0.018)
expersq	−0.0008 (0.0004)	−0.0008 (0.0004)	−0.0009 (0.0004)
		Selection equation	
educ		0.131 (0.025)	0.130 (0.025)
age		−0.053 (0.008)	−0.051 (0.006)
exper		0.123 (0.019)	0.123 (0.019)
expersq		−0.0019 (0.0006)	−0.0019 (0.0006)
kidslt6		−0.867 (0.119)	−0.867 (0.119)
kidsge6		0.036 (0.043)	0.036 (0.043)
nwifeinc		−0.012 (0.005)	−0.012 (0.005)
ρ		0.028 (0.163)	0.065
σ		0.663 (0.023)	0.664
mills ratio^{-1}			0.043 (0.168)
H_0^{C1}		$LM_{\kappa_{03}\kappa_{04}}(\hat{\theta}) = 3.69$	$C_{\kappa_{03}\kappa_{04}}(\tilde{\theta}) = 3.42$
		$LM^*_{\kappa_{03}\kappa_{04}}(\hat{\theta}) = 3.05$	$C^*_{\kappa_{03}\kappa_{04}}(\tilde{\theta}) = 3.35$
H_0^{C2}		$LM_{\kappa_{12}\kappa_{13}}(\hat{\theta}) = 1.56$	$C_{\kappa_{12}\kappa_{13}}(\tilde{\theta}) = 0.81$
		$LM^*_{\kappa_{12}\kappa_{13}}(\hat{\theta}) = 0.25$	$C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta}) = 0.56$
H_0^{C1C2}		$LM_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta}) = 4.72$	$C_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta}) = 4.19$
		$LM^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta}) = 4.05$	$C^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta}) = 3.98$
H_0^{BN}		$LM_{\kappa}(\hat{\theta}) = 383.3$	$C_{\kappa}(\tilde{\theta}) = 342.9$

Notes: Standard errors in parenthesis. Critical values for χ^2_2: 4.61 (10%), 5.99 (5%), 9.21 (1%); χ^2_4: 7.78 (10%), 9.49 (5%), 13.28 (1%); χ^2_9: 14.68 (10%), 16.92 (5%), 21.67 (1%).

parameter, which translates into a high value of the test statistics. However, the robust versions, LM* and C^*, cannot reject the null hypothesis. As expected, these results are similar to those found in the wage equation test statistics, because the same selection equation was used in both cases. In fact, these results are also in line with Mroz (1987) and Newey et al. (1990) assertions that normality in the participation equation cannot be rejected. Moreover, we cannot reject the conditional linearity hypothesis, H_0^{C2}. These test results are in line with those in Newey et al. (1990), who find that semiparametric estimators have similar estimates to that of the Heckman's two-step method. Overall, the tests suggest that the Heckman's two-step procedure is valid. The joint test H_0^{C1C2} provides mixed evidence on this. However, as argued above, the individual tests should be given more weight.

The significant differences between the two estimators in the hours equation suggest that the bivariate distribution of the error terms is not bivariate normal. In fact, this is suggested by the high values of the test statistics for H_0^{BN}, despite the fact that the Monte Carlo simulations of Lee's (1982, 1984) statistics for testing bivariate normality were well above the empirical size.

TABLE 7 Empirical application: Hours equation

	OLS	MLE	Heckman's two step
		Outcome equation	
educ	−195.0 (25.2)	−159.9 (40.4)	−96.1 (66.5)
age	−25.8 (4.21)	−10.13 (6.71)	1.58 (12.2)
kidslt6	−431.6 (59.2)	−200 (121.5)	20.5 (231.6)
kidsge6	−38.6 (23.3)	−79.2 (30.5)	−87.7 (34.3)
nwifeinc	−4.55 (2.54)	0.78 (3.77)	4.25 (5.06)
lwage	2067.4 (178.0)	1135.9 (3.95.6)	250.9 (830.3)
		Selection equation	
educ		0.132 (0.025)	0.130 (0.025)
age		−0.053 (0.009)	−0.051 (0.006)
exper		0.120 (0.018)	0.123 (0.019)
expersq		−0.0019 (0.0006)	−0.0019 (0.0006)
kidslt6		−0.867 (0.119)	−0.867 (0.119)
kidsge6		0.036 (0.043)	0.036 (0.043)
nwifeinc		−0.012 (0.005)	−0.012 (0.005)
ρ		−0.286 (0.212)	−0.807
σ		736.7 (34.1)	867.5
mills ratio^{-1}			−699.7 (452.5)
H_0^{C1}		$LM_{\kappa_{03}\kappa_{04}}(\hat{\theta}) = 4.41$	$C_{\kappa_{03}\kappa_{04}}(\tilde{\theta}) = 9.06$
		$LM^*_{\kappa_{03}\kappa_{04}}(\hat{\theta}) = 3.32$	$C^*_{\kappa_{03}\kappa_{04}}(\tilde{\theta}) = 4.23$
H_0^{C2}		$LM_{\kappa_{12}\kappa_{13}}(\hat{\theta}) = 7.82$	$C_{\kappa_{12}\kappa_{13}}(\tilde{\theta}) = 5.75$
		$LM^*_{\kappa_{12}\kappa_{13}}(\hat{\theta}) = 2.17$	$C^*_{\kappa_{12}\kappa_{13}}(\tilde{\theta}) = 4.15$
H_0^{C1C2}		$LM_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta}) = 10.03$	$C_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta}) = 15.2$
		$LM^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\hat{\theta}) = 5.06$	$C^*_{\kappa_{12}\kappa_{13}\kappa_{03}\kappa_{04}}(\tilde{\theta}) = 10.9$
H_0^{BN}		$LM_{\kappa}(\hat{\theta}) = 67.1$	$C_{\kappa}(\tilde{\theta}) = 560.5$

Notes: Standard errors in parenthesis. Critical values for χ^2_2: 4.61 (10%), 5.99 (5%), 9.21 (1%); χ^2_4: 7.78 (10%), 9.49 (5%), 13.28 (1%); χ^2_9: 14.68 (10%), 16.92 (5%), 21.67 (1%).

8. CONCLUSIONS AND FUTURE RESEARCH

The article 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. It also constructed adjusted LM tests for the case where the Heckman's 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.

The tests procedures are applied to the well-known Mroz (1987) database for women's labor supply. The 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. Moreover, 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 semiparametric approach.

ACKNOWLEDGMENTS

I am in debt to Anil Bera, Roger Koenker, Todd Elder, Zhonjgun Qu, Thomas Mroz, Esfandiar Maasoumi, an Associate Editor, and two anonymous referees for helpful comments and suggestions. A previous version of this article appeared as a chapter in my Ph.D. in Economics Dissertation at the University of Illinois Urbana-Champaign. All remaining errors are mine.

REFERENCES

Bera, A. K., Bilias, Y. (2001). Rao's score, Neyman C(α) and Silvey's LM tests: an essay on historical developments and some new results. *Journal of Statistical Planning and Inference* 97:9–44.

Bera, A. K., Jarque, C. M., Lee, L.-F. (1984). Testing the normality assumption in limited dependent variable models. *International Economic Review* 25:563–578.

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

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

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

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

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

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

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

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

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* 5:475–492.

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

Heckman, J. J., Tobias, J. L., Vytlacil, E. (2003). Simple estimators for treatment parameters in a latent-variable framework. *Review of Economics and Statistics* 85:748–755.

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

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

Jondeau, E., Rockinger, M. (2001). Gram-Charlier densities. *Journal of Economic Dynamics and Control* 25:1457–1483.

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

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

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* 55:765–799.

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

Newey, W. K. (1999). Consistency of two-step sample selection estimators despite misspecification of distribution. *Economics Letters* 63:129–132.

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

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