

Article

# A GMM-Based Test for Normal Disturbances of the Heckman Sample Selection Model

Michael Pfaffermayr

<sup>1</sup> Department of Economics, University of Innsbruck, Universitaetsstrasse 15, Innsbruck 6020, Austria; E-Mail: Michael.Pfaffermayr@uibk.ac.at; Tel.: +43-512-507-7359

<sup>2</sup> Austrian Institute of Economic Research, P.O.-Box 91, Vienna A-1103, Austria

External Editor: Kerry Patterson

*Received: 31 July 2014; in revised form: 10 October 2014 / Accepted: 14 October 2014 /*

*Published: 23 October 2014*

---

**Abstract:** The Heckman sample selection model relies on the assumption of normal and homoskedastic disturbances. However, before considering more general, alternative semiparametric models that do not need the normality assumption, it seems useful to test this assumption. Following Meijer and Wansbeek (2007), the present contribution derives a GMM-based pseudo-score LM test on whether the third and fourth moments of the disturbances of the outcome equation of the Heckman model conform to those implied by the truncated normal distribution. The test is easy to calculate and in Monte Carlo simulations it shows good performance for sample sizes of 1000 or larger.

**Keywords:** sample selection model; GMM; normality; pseudo-score LM test

**JEL classifications:** C23; C21

---

## 1. Introduction

The assumption of bivariate normal and homoskedastic disturbances is a prerequisite for the consistency of the maximum likelihood estimator of the Heckman sample selection model. Moreover, some studies focus on the prediction of counterfactuals based on the Heckman sample selection model taking into account both changes in participation and outcome, which is often only feasible under the

assumption of bivariate normality.<sup>1</sup> Lastly, under the assumption of bivariate normality one can estimate the Heckman sample selection model by maximum likelihood methods that are less sensitive to weak exclusion restrictions.

Before employing alternative semiparametric estimators that do not need the normality assumption (see e.g., Newey, 2009 [3]), it seems useful to test the underlying normality assumption of sample selection models. So far, the literature offers several approaches to test this hypothesis.<sup>2</sup> Bera *et. al.*, (1984) [6] develop an LM test for normality of the disturbances in the general Pearson framework, which implies testing the moments up to order four. Lee (1984) [7] proposes Lagrangian multiplier tests within the bivariate Edgeworth series of distributions. Van der Klaauw and Koning (1993) [8] derive LR tests in a similar setting, while Montes-Rojas (2011) [9] proposes LM and  $C(\alpha)$  tests that are likewise based on bivariate Edgeworth series expansions, but robust to local misspecification in nuisance distributional parameters. In general, these approaches tend to lead to complicated test statistics that are sometimes difficult to implement in standard econometric software. More importantly, some of these tests for bivariate normality seem to exhibit unsatisfactory performance in Monte Carlo simulations and are rejected too often in small to medium samples sizes, especially if the parameter of the Mills' ratio is high in absolute value (see e.g., Montes-Rojas, 2011 [9], Table 1). This motivates Montes-Rojas (2011) [9] to focus on the assumptions of the two-step estimator that requires less restrictive assumptions, namely a normal marginal distribution of the disturbances of the selection equation and a linear conditional expectation of the disturbances of the outcome equation. He proposes to test for marginal normality and linearity of the conditional expectation of outcome model separately and shows that the corresponding locally size-robust test statistics based on the two-step estimator perform well in terms size and power .

In a possibly neglected, but very valuable paper, Meijer and Wansbeek (2007) [10] embed the two-step estimator of the Heckman sample selection model in a GMM-framework. In addition, they argue that within this framework it is easily possible to add moment conditions for designing Wald tests in order to check the assumption of bivariate normality and homoskedasticity of the disturbances. Their approach does not attempt to develop a most powerful test, rather they intended to design a relatively simple test for normality that can be used as an alternative to the existing tests. The test can be interpreted as a conditional moment test and checks whether the third and fourth moments of the disturbances of the outcome equation of the Heckman model conform to those implied by the truncated normal distribution. For  $H_0$  to hold, the test in addition requires normally distributed disturbances of the selection equation and the absence of heteroskedasticity in both the outcome and the selection equation.

Meijer and Wansbeek (2007) [10] do not explicitly derive the corresponding test statistic nor do they provide Monte Carlo simulations on its performances in finite samples. The present contribution takes up their approach arguing that a GMM based pseudo-score LM test is well suited to test the hypothesis of bivariate normality and is easy to calculate. The derived LM test is similar to the widely used Jarque

---

<sup>1</sup> An example is the estimation of gravity models of bilateral trade flows with missing and/or zero trade. Here, the assumption of bivariate normality turns out important for deriving comparative static results with respect changes in the external and internal margin of trade following Yen and Rosinski (2008) [1] and Staub (2014) [2].

<sup>2</sup> There is also work available that proposes normality tests for the Tobit model (see Skeels and Vella, 1999 [4] and Drukker, 2002 [5]).

and Bera LM test (1980) [11], and in the absence of sample selection reverts to their LM test statistic. Monte Carlo simulations show good performance of the proposed test for samples of sizes of 1000 or larger, especially if a powerful exclusion restriction is available.

## 2. The GMM Based Pseudo-Score LM Test for Normality

In a cross-section of  $n$  units the Heckman (1979) [12] sample selection model is given as

$$\begin{aligned} y_{1i}^* &= z_i' \gamma + u_{1i} \\ y_{2i}^* &= x_i' \beta + u_{2i} \\ d_i &= \begin{cases} 1 & \text{if } y_{1i}^* > 0 \\ 0 & \text{otherwise} \end{cases} \\ y_{2i} &= \begin{cases} y_{2i}^* & \text{if } d_i = 1 \\ \text{unobserved} & \text{if } d_i = 0 \end{cases} \end{aligned}$$

where  $y_{1i}^*$  and  $y_{2i}^*$  denote latent random variables. The outcome variable,  $y_{2i}^*$ , is observed if the latent variable  $y_{1i}^* > 0$  or, equivalently, if  $d_i = 1$ .  $z_i$  is a  $k_1 \times 1$  vector containing the exogenous variables of the selection equation and  $x_i$  is the  $k_2 \times 1$  vector of the exogenous variables of the outcome equation.  $z_i$  may include the variables in  $x_i$ , but also additional ones so that an exclusion restriction holds.  $\gamma$  and  $\beta$  denote the corresponding parameter vectors. Under  $H_0$  the disturbances are assumed to be distributed as bivariate normal, *i.e.*,

$$\left( \begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix} \middle| x_i, z_i \right) \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \tau \\ \tau & \sigma^2 \end{bmatrix} \right)$$

It is easy to show that under these assumptions

$$\begin{aligned} p_i &\equiv E[d_i] = \Phi(z_i' \gamma) \\ \lambda_i &\equiv E[u_{1i} | u_{1i} \geq -z_i' \gamma] = \frac{\phi(-z_i' \gamma)}{1 - \Phi(-z_i' \gamma)} = \frac{\phi(z_i' \gamma)}{\Phi(z_i' \gamma)} \end{aligned}$$

where  $\lambda_i$  denotes the inverse Mills' ratio. Under the normal assumption one can specify  $u_{2i} = \tau u_{1i} + \varepsilon_i$  so that  $\varepsilon_i \sim iid N(0, \sigma^2 - \tau^2)$ .  $\varepsilon_i$  is independent of  $u_{1i}$  as  $E[\varepsilon_i u_{1i}] = E[(u_{2i} - \tau u_{1i}) u_{1i}] = \tau - \tau = 0$ . Since  $E[u_{1i} | d_i = 1] = \lambda_i$ , it holds that  $E[\tau(u_{1i} - \lambda_i) + \varepsilon_i | d_i = 1] = 0$ . Therefore, the two-step Heckman sample selection model includes the estimated inverse Mills' ratio in the outcome equation as an additional regressor. For the observed outcome at  $d_i = 1$  the model can be written as

$$\begin{aligned} y_{2i}^* &= x_i' \beta + \tau \lambda_i + \tau (u_{1i} - \lambda_i) + \varepsilon_i \\ &\equiv w_i' \alpha + \tau v_i + \varepsilon_i \end{aligned} \tag{1}$$

where  $v_i = u_{1i} - \lambda_i$  and  $E[y_{2i}^* | d_i = 1] = x_i' \beta + \tau \lambda_i$ .

Meijer and Wansbeek (2007) [10] embed the two-step Heckman sample selection estimator in a GMM framework and demonstrate that the estimation can be based on

$$\begin{aligned} h_{1,1i(k_1 \times 1)}(\theta_1) &\equiv \frac{(d_i - p_i) \phi_i}{p_i(1 - p_i)} z_i \\ h_{1,2i(k_2 \times 1)}(\theta_1) &\equiv d_i w_i (y_i - w_i' \alpha) = d_i w_i (\tau v_i + \varepsilon_i) \\ h_{1,3i(1 \times 1)}(\theta_1) &\equiv d_i [(y_i - w_i' \alpha)^2 - \varphi_{2,i}] = d_i [(\varepsilon_i + \tau v_i)^2 - \varphi_{2,i}] \end{aligned}$$

where  $\theta_1 = (\gamma', \beta', \tau, \sigma)'$  and  $\varphi_{k,i} = E[(\tau v_i + \varepsilon_i)^k | d_i = 1]$ ,  $k = 2, 3, 4$ . Note, there are as many parameters as moment conditions and the model is just-identified.

The first set of moment equations is based on  $\bar{h}_{1,1}(\theta_1) = \frac{1}{n} \sum_{i=1}^n h_{1,1i}(\theta_1)$  and refers to the score of the Probit model. Since these moment conditions do not include the parameters entering  $h_{1,2i}$  and  $h_{1,3i}$  (i.e.,  $\beta', \tau, \sigma$ ) and are exactly identified, estimation can proceed in steps: In the first step, one can solve  $\frac{1}{n} \sum_{i=1}^n h_{1,1i}(\hat{\gamma}) = 0$  and in the second step one solves the sample moment condition  $\bar{h}_{1,2}(\theta_1) = \frac{1}{n} \sum_{i=1}^n [h_{1,2i}(\hat{\gamma}', \hat{\beta}', \hat{\tau})] = 0$  using the estimated  $\hat{\gamma}$  derived in the first stage. This leads to the two-step Heckman estimator, which first estimates a Probit model, inserts the estimated Mills' ratio  $\hat{\lambda}_i$  as additional regressor in the outcome equation and applies OLS. Lastly, from  $\bar{h}_{1,3}(\theta_1) = \frac{1}{n} \sum h_{1,3i}(\hat{\gamma}', \hat{\beta}', \hat{\tau}, \hat{\sigma}) = 0$  one can obtain an estimator of  $\sigma^2$ .

As Meijer and Wansbeek (2007) [10] remark, a rough and simple test for normality can be based on two additional moment conditions that allow comparing the third and fourth moments of the estimated residuals of the outcome equation,  $y_i - w_i' \hat{\alpha}$ , with their theoretical counterparts based on the truncated normal distribution. These moment conditions use

$$\begin{aligned} h_{2,1i(1 \times 1)}(\theta_1, \theta_2) &\equiv d_i [(y_i - w_i' \alpha)^3 - \varphi_{3,i} - \xi] = d_i [(\tau v_i + \varepsilon_i)^3 - \varphi_{3,i} - \xi] \\ h_{2,2i(1 \times 1)}(\theta_1, \theta_2) &\equiv d_i [(y_i - w_i' \alpha)^4 - \varphi_{4,i} - \kappa] = d_i [(\tau v_i + \varepsilon_i)^4 - \varphi_{4,i} - \kappa] \end{aligned}$$

Thereby,  $\theta_2 = (\xi, \kappa)$  denotes additional parameters that are zero under normality. More importantly, under  $H_0$  the expectations  $\varphi_{k,i}$  can be derived recursively from the moments of the truncated normal distribution as shown in the Appendix (see also Meijer and Wansbeek, 2007, pp. 45–46) [10]. In general, these moments depend on the parameters  $\theta_1$  and, especially, on the inverse Mills' ratio  $\lambda_i$  and the parameter  $\tau$ .

To detect violations of the normality assumption, one can test  $H_0 : \xi = 0$  and  $\kappa = 0$  vs.  $H_0 : \xi \neq 0$  and/or  $\kappa \neq 0$ . Although this hypothesis checks the third and fourth moments of the disturbances of the two-step outcome Equation (1), it can only be true if  $\varphi_{3,i}$  and  $\varphi_{4,i}$  are the correct expected values. Therefore, the test additionally requires the moment conditions  $E[h_{1,1i}] = 0$  and  $E[h_{1,2i}] = 0$  to hold so that the parameters of both the selection equation and the outcome equation are consistently estimated. The present hypothesis is somewhat more restrictive than that tested, e.g., in Montes-Rojas (2011) [9], who emphasizes that the Heckman two-step estimator is robust to distributional misspecification if (i) the marginal distribution of  $u_{1i}$  is normal and (ii)  $E[u_{2i}|u_{1i}] = \tau u_{1i}$ , i.e., the conditional expectation is linear.<sup>3</sup>

In addition,  $H_0$  also requires the absence of heteroskedasticity ( see Meijer and Wansbeek, 2007, p. 46 ) [10]. To give an example, assume that  $u_{2i} = \tau u_{1i} + \varepsilon_i$  and  $u_{2i}$  are bivariate normal, but the variances of  $\varepsilon_i$  differ across  $i$  and are given as  $\sigma_i^2 - \tau^2$  (see also the DGP6 in the Monte Carlo set-up below and the excess kurtosis of DGP6 in Table 1 below). Then, it follows that  $\varphi_{k,i} = \sum_{j=0}^k \binom{k}{j} E[\varepsilon_i^{k-j}] \tau^j \psi_{j,i}$ , where  $\psi_{k,i} \equiv E[(v_i^k | d_i = 1)] = \sum_{j=0}^k (-1)^j \binom{k}{j} \mu_{j,i} \lambda_i^{k-j}$  and  $\mu_{j,i} = E[u_{1i}^k | u_{1i} > -z_i' \gamma]$  (see the

<sup>3</sup> Specifically, Montes-Rojas (2011)[9] mentions the case where  $u_{1i} \sim N(0, 1)$ ,  $u_{2i} = \tau u_{1i} + \varepsilon_i$  and  $u_{1i}$  and  $\varepsilon_i$  being independent, but  $\varepsilon_i$  does not follow a normal distribution.  $\varphi_{3,i}$  and  $\varphi_{4,i}$  the moments are  $E[\varepsilon_i^k]$  are left unrestricted and estimated from the residuals of the second-stage outcome equation.

Appendix). In this case, we have  $E[\varepsilon_i^2] = \sigma_i^2 - \tau^2$  and  $E[\varepsilon_i^4] = 3(\sigma_i^2 - \tau^2)^2$ , while the corresponding uneven moments are zero. Hence, under heteroskedasticity  $\varphi_{4,i}$  differs from that obtained under  $H_0$  which assumes  $E[\varepsilon_i^2] = \sigma^2 - \tau^2$  and  $E[\varepsilon_i^4] = 3(\sigma^2 - \tau^2)^2$  and the population moment condition  $E[h_{2,2i}(\theta_1, \theta_2)] = 0$  is violated. Hence, a test based on these moments should also be able to detect heteroskedasticity, although not in the most efficient way.

Applying a pseudo-score LM test (Newey and West, 1987 [13]; Hall, 2005 [14]), in this GMM-framework leads to a  $\chi^2(2)$ -test statistic that can be calculated easily. In order to derive the LM test statistic, define  $\bar{h}(\theta) = \frac{1}{n} \sum_{i=1}^n h_i(\theta)$  and  $\bar{\Psi}(\theta) = \frac{1}{n} \sum_{i=1}^n h_i(\theta)h_i(\theta)'$ , where  $h_i(\theta) \equiv (h'_{1,1i}, h'_{1,2i}, h_{1,3i}, h_{2,1i}, h_{2,2i})'$ . It is assumed that  $\Psi_0 = \text{plim}_{n \rightarrow \infty} \bar{\Psi}(\theta_0)$  exists, is positive definite and invertible. Under standard assumptions, it holds that

$$\begin{aligned} n^{1/2}\bar{h}(\theta_0) &\xrightarrow{d} N(0, \Psi_0) \\ n^{1/2}(\hat{\theta} - \theta_0) &\xrightarrow{d} N(0, A_0) \end{aligned}$$

where the subscript 0 indicates that  $H_0$  is assumed. Thereby,  $A_0 = G_0^{-1}\Psi_0(G_0^{-1})'$  and  $G_0$  is the probability limit of  $\bar{G}(\theta_0) = \frac{1}{n} \sum_{i=1}^n \frac{\partial h(\theta)}{\partial \theta} \Big|_{\theta=\theta_0}$ . Note,  $\bar{G}(\theta_0)$  is invertible as the model is just-identified.

Under  $H_0$  the moment conditions  $E[h_{2,i}(\theta_1, \theta_2), \xi, \kappa]$ ,  $h_{2,i}(\theta_1, \theta_2) \equiv (h_{2,1i}, h_{2,2i})'$  referring to the third and fourth moments of the outcome equation are zero at  $\xi = 0$  and  $\kappa = 0$  and the separability result in Ahn and Schmidt (1995, Section 4) [15] can be applied. Denoting the restricted estimates under  $H_0$  by a tilde, using the invertibility of  $\bar{G}(\tilde{\theta})$  and the partitioned inverse of  $\Psi_n(\tilde{\theta}) = E[\bar{\Psi}(\tilde{\theta})]$ , the pseudo-score LM test statistic can be derived as (see the Appendix for details):

$$LM(\tilde{\theta}) = n\bar{h}'_2(\tilde{\theta}) \left( \Psi_{n,22}(\tilde{\theta}) - \Psi_{n,21}(\tilde{\theta})\Psi_{n,11}(\tilde{\theta})^{-1}\Psi_{n,12}(\tilde{\theta}) \right)^{-1} \bar{h}_2(\tilde{\theta})$$

Thereby,  $\bar{h}_2(\theta) = \frac{1}{n} \sum_{i=1}^n h_{2,i}(\theta)$  and we use  $\bar{h}'_1(\tilde{\theta}) = \frac{1}{n} \sum_{i=1}^n h_{1,i}(\tilde{\theta}) = 0$ , where  $h_{1,i}(\theta) = (h'_{1,1i}, h'_{1,2i}, h_{1,3i})'$ , as well as the partitioned inverse (see the Appendix)

$$\begin{aligned} \Psi_{n,11}(\theta) &= \frac{1}{n} \begin{bmatrix} Z'VZ & 0 & 0 \\ * & W'_1\Sigma_1W_1 & \sum_{d_i=1} w_i\varphi_{3,i} \\ * & * & \sum_{d_i=1} (\varphi_{4,i} - \varphi_{2,i}^2) \end{bmatrix} \\ \Psi_{n,22}(\theta) &= \frac{1}{n} \sum_{i=1} \begin{bmatrix} p_i(\varphi_{6,i} - \varphi_{3,i}^2) & p_i(\varphi_{7,i} - \varphi_{3,i}\varphi_{4,i}) \\ p_i(\varphi_{7,i} - \varphi_{3,i}\varphi_{4,i}) & p_i(\varphi_{8,i} - \varphi_{4,i}^2) \end{bmatrix} \\ \Psi_{n,12}(\theta) &= \frac{1}{n} \sum_{i=1} \begin{bmatrix} 0 & 0 \\ p_iw_i\varphi_{4,i} & p_iw_i\varphi_{5,i} \\ p_i(\varphi_{5,i} - \varphi_{2,i}\varphi_{3,i}) & p_i(\varphi_{6,i} - \varphi_{4,i}\varphi_{2,i}) \end{bmatrix} \end{aligned}$$

where  $V = \text{diag}\left(\frac{\phi_1^2}{p_1(1-p_1)}, \dots, \frac{\phi_n^2}{p_n(1-p_n)}\right)$ ,  $p_i = P(d_i = 1)$ ,  $Z_{n \times k_1} = (z_1, \dots, z_n)'$ ,  $W_{n \times k_2} = (w_1, \dots, w_n)'$ , and  $\Sigma = \text{diag}(\varphi_{2,1}, \dots, \varphi_{2,n})$ .  $\Sigma_1$  is obtained from  $\Sigma$  by deleting all rows and columns referring to  $d_i = 0$ , and similarly  $W_1$ .  $\Psi_n(\theta)$  can be consistently estimated by plugging in  $\tilde{\theta}$ . In addition, Meijer and Wansbeek (2007) [10] show that one can substitute  $d_i$  for  $p_i$  so that only information on the observed units is necessary. Note however, the summation runs over all observations (zero and ones in  $d_i$ ).

Under standard assumptions it follows that under  $H_0$  we have  $LM(\tilde{\theta}) \xrightarrow{d} \chi^2(2)$  (see Newey and West, 1987, pp. 781–782 [13] and Theorems 5.6 and 5.7 in Hall, 2005 [14]). In the absence of sample

selection ( $\tau = 0$ ) it holds that  $\varphi_{3,i} = \varphi_{5,i} = 0$ , while  $\varphi_{2,i} = \sigma^2$  and  $\varphi_{4,i} = 3\sigma^4$  and the LM test statistic reverts to that of Jarque and Bera (1980) [11].

### 3. Monte Carlo Simulation

Monte Carlo simulations may shed light on the performance of the proposed LM test in finite samples. It is based on a design that has been used previously by van der Klaauw and Koning (1993) [8] and Montes-Rojas (2011) [9], but includes a few modifications. The simulated model is specified as

$$\begin{aligned}y_{1i}^* &= -1z_{1i} + 1x_{2i} - 1 + u_{1i} \\ y_{2i}^* &= 0.5x_{1i} - 0.5x_{2i} + 1 + u_{2i}\end{aligned}$$

where for  $\rho \in \{-0.8, -0.4, 0.4, 0.8\}$  and  $\sigma^2 \in \{0.25, 1\}$ . The explanatory variables  $x_{1i}$ ,  $x_{2i}$ , and  $z_{1i}$  are generated as *iid*  $N(0, 3)$ ,  $N(0, 3)$  and  $U(-3, 3)$ , respectively. With respect to the disturbances,  $u_{1i}$ , and  $u_{2i}$  the following data generating processes are considered. Note DGP1-DGP3 imply  $Var[u_{1i}] = 1$  and  $Var[u_{2i}] = 0.25$ . In contrast, van der Klaauw and Koning (1993) [8] and Montes-Rojas (2011) [9] consider the case with  $Var[u_{2i}] = 5$  and thus receive less precise estimates of the slope parameters of the outcome equation.

#### 1. DGP1:

$$(u_{1i}, u_{2i}) \sim iid N\left(0, \begin{bmatrix} 1 & 0.5\rho \\ 0.5\rho & 0.25 \end{bmatrix}\right)$$

#### 2. DGP2:

$\varepsilon_{1i} \sim t(10)$ ,  $\varepsilon_{2i} \sim t(10)$ ,  $\varepsilon_{1i}$  and  $\varepsilon_{2i}$  being independent.

$$u_{1i} = \varepsilon_{1i} \left(\frac{10}{8}\right)^{-1/2}$$

$$u_{2i} = \sigma(1 + \rho^2)^{1/2} \left(\frac{10}{8}\right)^{-1/2} \varepsilon_{2i} + \rho\sigma u_{1i}$$

The degrees of freedom are set to 10 to guarantee that the moments up to order 4 exists.

#### 3. DGP3:

$\varepsilon_{1i} \sim \chi^2(20)$ ,  $\varepsilon_{2i} \sim \chi^2(30)$

$$u_{1i} = (\varepsilon_{1i} - 20) / \sqrt{40} - 20$$

$$u_{2i} = \sigma(1 + \rho^2)^{1/2} (\varepsilon_{2i} - 30) / \sqrt{60} + \rho\sigma u_{1i}$$

#### 4. DGP4:

$\varepsilon_{1i} \sim N(0, 1)$ ,  $\varepsilon_{2i} \sim \chi^2(30)$  and are independent.

$$u_{1i} = \varepsilon_{1i}$$

$$u_{2i} = \sigma(1 + \rho^2)^{1/2} (\varepsilon_{2i} - 30) / \sqrt{60} + \rho\sigma u_{1i}$$

#### 5. DGP5:

$\varepsilon_{1i} \sim \chi^2(20)$ ,  $\varepsilon_{3i} \sim N(0, 1)$  and are independent.

$$u_{1i} = (\varepsilon_{1i} - 20) / \sqrt{40} - 1$$

$$u_{2i} = \sigma(1 + \rho^2)^{1/2} \varepsilon_{2i} + \rho\sigma u_{1i}$$

#### 6. DGP6:

$\varepsilon_{1i} \sim N(0, 1)$ ,  $\varepsilon_{2i} \sim N(0, 0.25)$ ,  $\varepsilon_{1i}$  and  $\varepsilon_{2i}$  being independent.

$$c_i = 1 + e^{\frac{x_{1i}}{\sqrt{3}}} \left( e^{-\frac{1}{2}} - 1 \right) (e^1 - 1)^{-\frac{1}{2}}$$

$$u_{1i} = \varepsilon_{1i}$$

$$u_{2i} = \sigma(1 + \rho^2)^{1/2} \sqrt{c_i} \varepsilon_{2i} + 2\rho u_{1i}$$

#### 7. DGP7:

$\varepsilon_{1i} \sim N(0, 1)$ ,  $\varepsilon_{2i} \sim N(0, 1)$ ,  $\varepsilon_{1i}$  and  $\varepsilon_{2i}$  being independent.

$$c_i = 1 + e^{\frac{x_{2i}}{\sqrt{3}}} \left( e^{-\frac{1}{2}} - 1 \right) (e^1 - 1)^{-\frac{1}{2}}$$

$$u_{1i} = \varepsilon_{1i} \sqrt{c_i}$$

$$u_{2i} = \sigma(1 + \rho^2)^{1/2} \varepsilon_{2i} + \sigma\rho u_{1i}$$

DGP1 serves as a reference to assess the size of the pseudo-score LM test. The second DGP deviates from the bivariate normal in terms of a higher kurtosis, while DGP3 exhibits both higher skewness and kurtosis than the normal. DGP4 allows for deviation from normality in the outcome equation, while keeping the normality assumption in the selection equation. DGP5 reverses this pattern. The disturbances of the outcome equation are normal and those of the selection equation are not. DGP6 and DGP7 introduce heteroskedasticity in either the outcome or the selection equation, respectively. In case of the latter two, the variances of  $u_{i1}$  and  $u_{i2}$  is normalized to an average of 1 and 0.25, respectively. Note, the explanatory variables are held fixed in repeated samples.

Overall, for these DGPs four experiments are considered. In the baseline Experiment 1 (first row of the figures of graphs) 37% of the data remain unobserved and in the absence of sample selection the implied  $R^2$  amounts to  $1 - \frac{0.25}{1.75} = 0.86$  using  $Var(u_{2i}) = 0.25$  and  $Var(y_{2i}^*) = 1.75$ . Experiment 2 (second row of the figures of graphs) analyzed the performance of the Heckman two-step estimator under a weaker exclusion restriction, assuming  $z_{1i} \sim iid U(-1, 1)$  so that  $Var(z_{1i}) = 1/3$ : Experiment 3 (third row of the figures of graphs) sets the constant of the outcome equation to zero so that 49% instead of 37% units are unobserved. Lastly, Experiments 4 (fourth row of the figures of graphs) considers a weaker fit in the outcome equation setting  $Var(u_{2i}) = 1$  so that in the absence of sample selection we have  $R^2 = 0.43$ .

Table 1 summarizes the average variance, skewness and kurtosis of the generated disturbances  $u_{1i}$  and  $u_{2i}$  under Experiment 1. In DGP2-DGP7, depending on  $\rho$ , the average kurtosis of  $u_{2i}$  varies between 3.00 and 5.68, while the kurtosis of  $u_{1i}$  lies between 3.07 and 3.58 in DGP5. In the other ones the kurtosis of  $u_{1i}$  is held constant taking values 2.99 (DGPs 1,4 and 6), 3.97 (DGP2), 3.58 (DGP3) and 5.71 (DGP7), respectively. The skewness coefficient of the generated disturbances is zero for all DGPs except for DGP3 with corresponding values of 0.63 ( $u_{1i}$ ) and  $-0.21$  to  $0.43$  ( $u_{2i}$ ) and DGP5 where the skewness of  $u_{1i}$  varies between 0.14 and 0.63.

**Table 1.** Variance, Skewness and Kurtosis of the simulated disturbances.

DGP	$\rho$	$u_1$			$u_2$		
		Variance	Skewness	Kurtosis	Variance	Skewness	Kurtosis
1	all	1.00	0.00	2.99	0.25	0.00	2.99
2	-0.8	1.00	0.00	3.97	0.25	0.00	3.52
2	-0.4	1.00	0.00	3.97	0.25	0.00	3.70
2	0.0	1.00	0.00	3.97	0.25	0.00	3.96
2	0.4	1.00	0.00	3.97	0.25	0.00	3.70
2	0.8	1.00	0.00	3.97	0.25	0.00	3.52
3	-0.8	1.00	0.63	3.58	0.25	-0.21	3.29
3	-0.4	1.00	0.63	3.58	0.25	0.35	3.28
3	0.0	1.00	0.63	3.58	0.25	0.51	3.38
3	0.4	1.00	0.63	3.58	0.25	0.43	3.28
3	0.8	1.00	0.63	3.58	0.25	0.43	3.29
4	-0.8	1.00	0.00	2.99	0.25	0.11	3.05
4	-0.4	1.00	0.00	2.99	0.25	0.39	3.27
4	0.0	1.00	0.00	2.99	0.25	0.51	3.38
4	0.4	1.00	0.00	2.99	0.25	0.39	3.27
4	0.8	1.00	0.00	2.99	0.25	0.11	3.04
5	-0.8	1.00	0.63	3.58	0.25	0.00	2.99
5	-0.4	1.00	0.63	3.58	0.25	0.00	2.99
5	0.0	1.00	0.63	3.58	0.25	0.00	2.99
5	0.4	1.00	0.63	3.58	0.25	0.00	2.99
5	0.8	1.00	0.63	3.58	0.25	0.00	2.99
6	-0.8	1.00	0.00	2.99	0.25	0.00	3.34
6	-0.4	1.00	0.00	2.99	0.25	0.00	4.89
6	0.0	1.00	0.00	2.99	0.25	0.00	5.68
6	0.4	1.00	0.00	2.99	0.25	0.00	4.89
6	0.8	1.00	0.00	2.99	0.25	0.00	3.35
7	-0.8	0.99	0.00	5.71	0.25	0.00	4.11
7	-0.4	0.99	0.00	5.71	0.25	0.00	3.06
7	0.0	0.99	0.00	5.71	0.25	0.00	2.99
7	0.4	0.99	0.00	5.71	0.25	0.00	3.06
7	0.8	0.99	0.00	5.71	0.25	0.00	4.11

Following Davidson and MacKinnon (1998) [16] the size and power is analyzed in terms of size-discrepancy and power-size curves. The former is based on the empirical cumulative distribution function of the  $p$ -values,  $p_r$ , defined as  $F(q) = \frac{1}{R} \sum_{r=1}^R I(p_r \leq q)$ , where  $R$  is the number of Monte Carlo replications. The size-discrepancy curves are defined as plots of  $F(q) - q$  against  $q$  under the assumption that  $H_0$  holds and DGP1 is the correct one. In addition, one can use a Kolmogorov and Smirnov test to see whether  $F(q) - q$  differs significantly from 0 (see Davidson and MacKinnon 1998, p. 11) [16]. The size-power curves plot power against size, *i.e.*,  $F_{H_1}(q)$  against  $F_{H_0}(q)$ . In both plots  $q \in [0, 0.15]$  and step size is 0.001. An important feature of this procedure is that it avoids size adjustments of the power curves if the tests reject too often under  $H_0$ .

Figure 1 exhibits the size-discrepancy plots for Experiments 1–4 and sample sizes  $n = 500, 1000, 2000$ . The plots show that the pseudo-score LM test is properly sized for  $\rho = -0.4$  and  $\rho = 0.4$



in all experiments, while it slightly over-rejects at  $\rho = -0.8$  and  $\rho = 0.8$ , especially at a small sample size ( $n = 500$ ). For example, at a nominal test size of 0.05 and a sample size of 1000 the size of LM test is too high by 0.012 percentage points at  $|\rho| = 0.8$ . For  $\rho = -0.4$  and  $\rho = 0.4$  the size-discrepancy is within the Kolmogorov and Smirnov 5% confidence of bound  $p \pm 0.0096$  for  $p$ -values smaller than 0.1. A similar result has also been mentioned in Montes-Rojas (2011) [9] in case of robust LM and  $C(\alpha)$  tests. A weaker exclusion restriction, setting  $Var(z_{1i}) = 1/3$  in Experiment 2, increases the size-discrepancy at high absolute values of  $\rho$  (Experiment 2, row 2 of Figure 1), but hardly affects the size of the test at  $|\rho| = 0.4$ . The size-discrepancy remains in the confidence bounds at medium values of  $\rho$ . Increasing the share of unobserved values to 0.49 (Experiment 3, row 3 of Figure 1) hardly affects the size-discrepancy. Lastly, Experiment 4 (last row of Figure 1) shows that a weaker fit ( $Var(u_{2,i}) = 1$ ) does not result in a larger size distortion as compared to the baseline in the first row of Figure 1. As one would expect, a larger number of observations generally enhances the performance of the LM test (see the last column in Figure 1). However, the large sample approximation improves relatively slowly with sample size under a weak exclusion restriction at high absolute values of  $\rho$  (confer the second row of graphs in Figure 1).

Figures 2–4 present the power-size plots of the pseudo-score LM test for the DGPs 2–3, 4–5 and 6–7, respectively. In general and in line with the literature, for all DGPs referring to the alternative hypothesis we observe lower power of the pseudo-score LM test at high absolute values of  $\rho$ , but especially so at  $\rho = -0.8$ . If the distribution of the disturbances of the outcome equation exhibits both skewness and excess kurtosis (DGP3) the simulated power of the pseudo-score LM test is higher than that of a symmetric distribution with fatter tails than the normal distribution except for ( $\rho = -0.8$ ). Furthermore, for DGP3 the power is generally lower at  $\rho = -0.8$  as compared to large positive values ( $\rho = 0.8$ ), which reflects differences in the skewness of the distribution of  $u_{2i}$  with respect to  $\rho$  (confer Table 1).

Figure 3 illustrates the power of the pseudo-score LM test under non-normality in either the outcome (DGP4) or the selection equation (DGP5) but not in both. Under DGP4 the pseudo-score LM test exhibits high power at intermediate absolute values of  $\rho$ , while at high absolute values of  $\rho$  the power tends to be lower as the weight of  $u_{1i}$  (that is assumed to be normal) is higher in the disturbances of the outcome equation. In case of DGP5 we see the reversed pattern. Deviations from normality are only detected in case of high absolute values of  $\rho$ . Actually, under DGP5 the test has no power at all at  $\rho = 0$ , since in this case there is no effect of the truncation of  $u_{i1}$  and disturbances of the outcome equation are normal. This results can be found in all four considered Experiments.

Figure 4 presents the size-power plot of DGP6 and DGP7 and refers to heteroskedasticity. DGP6 allows for heteroskedasticity in the outcome equation and DGP7 in the selection equation. The power-size curves indicate that the pseudo-score LM test is also able to detect this type of deviation from the model assumptions as heteroskedasticity translates into pronounced excess kurtosis of the disturbances of the outcome equation. For DGP6 this is the case at medium to low values of  $|\rho|$ . DGP7 introduces heteroskedasticity in Probit selection model. In this case, the LM test exhibits power at high absolute values of  $\rho$ , but has virtual no power at  $\rho = -0.4$  and  $0.4$ . The reason is that the nominal kurtosis of  $u_{2i}$  is hardly affected (amounting to 3.06, see Table 1) and the bias of the Mills' ratio and the estimated coefficients of the outcome equation, especially that of the Mills' ratio turn out small in comparison.

Figure 1. Size-discrepancy plot.

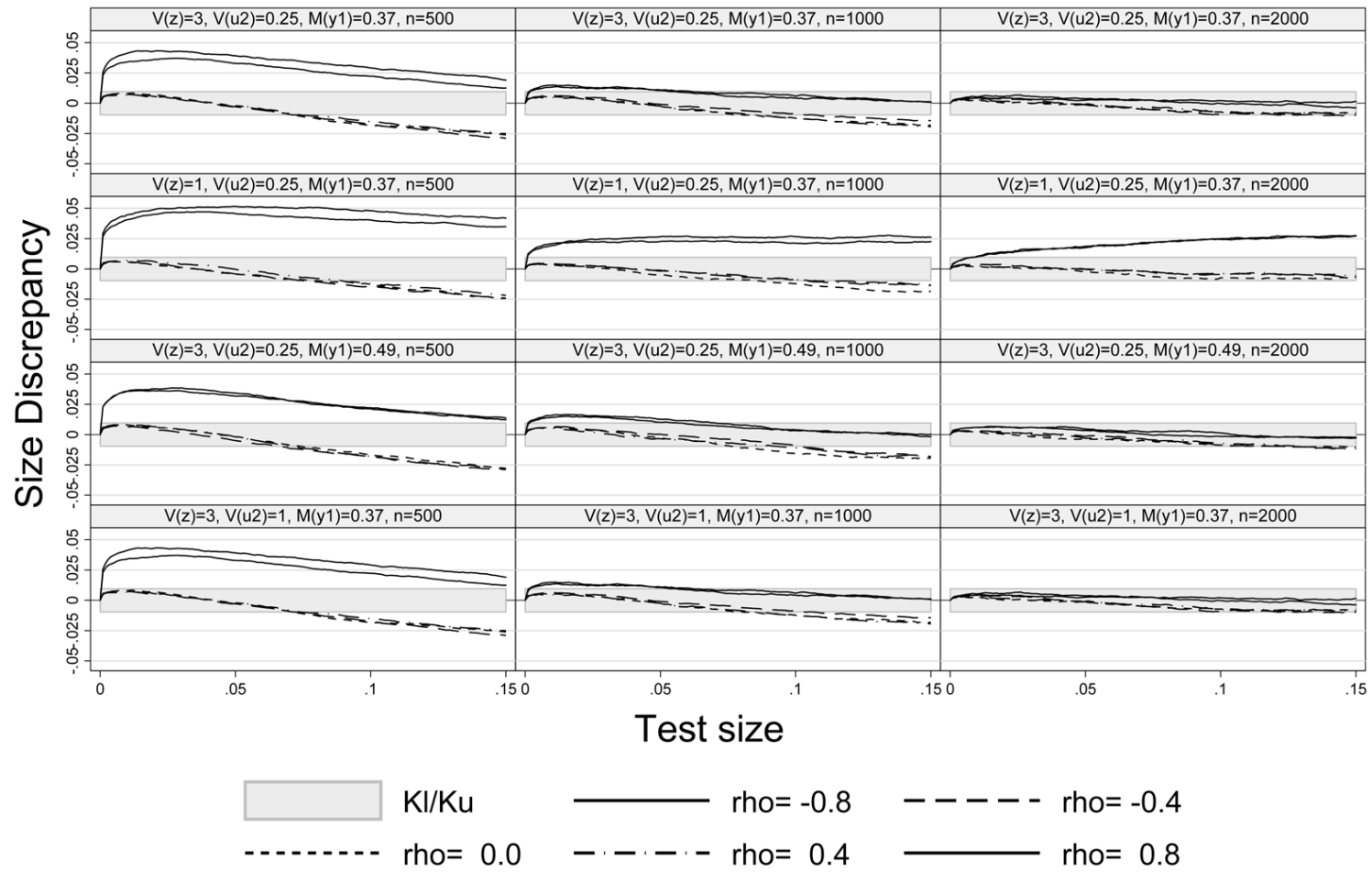


Figure 2. Size power plot, DGP1-DGP3,  $n = 1000$ .

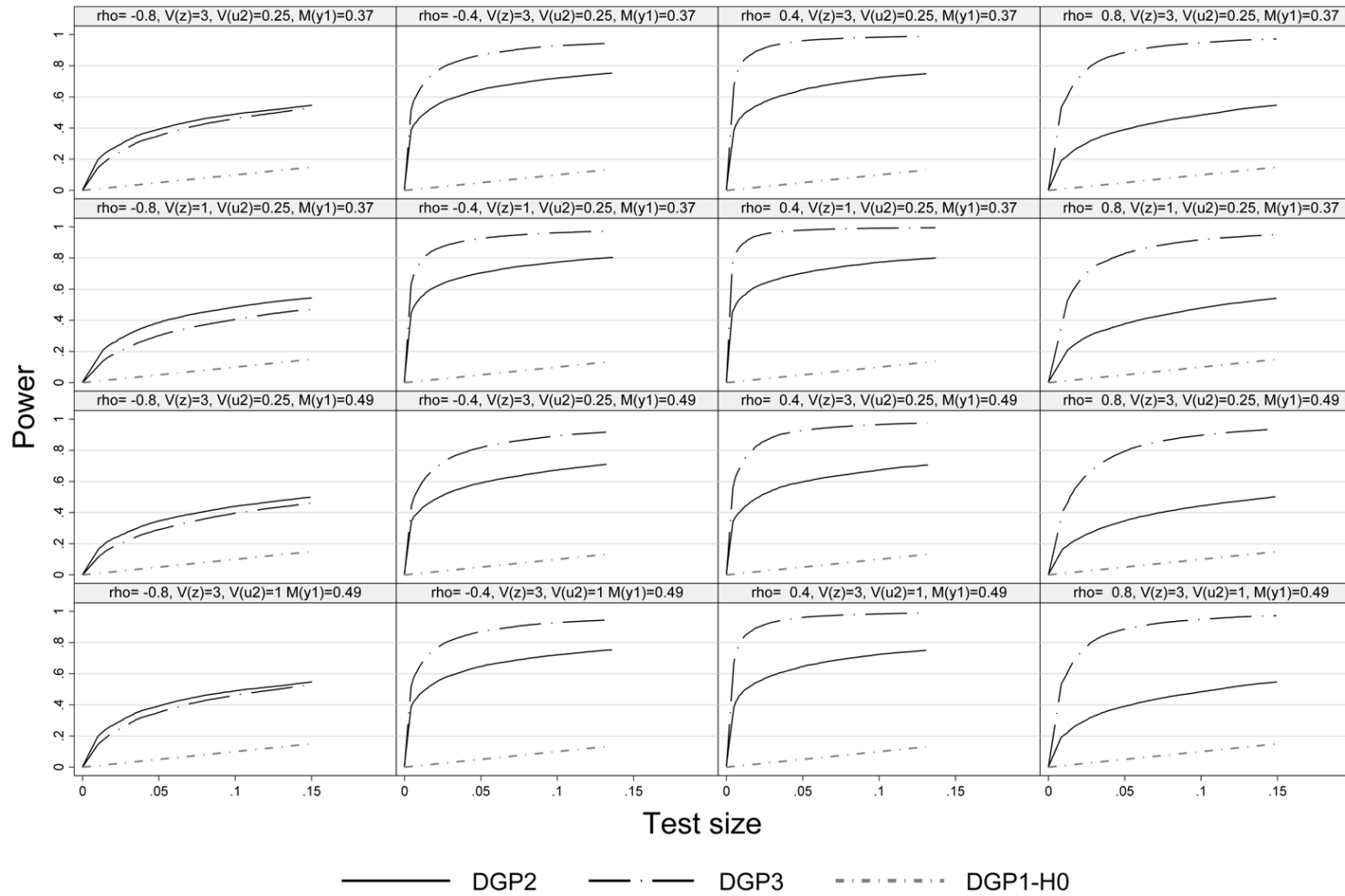


Figure 3. Size power plot, DGP1, DGP4 and DGP 5,  $n = 1000$ .

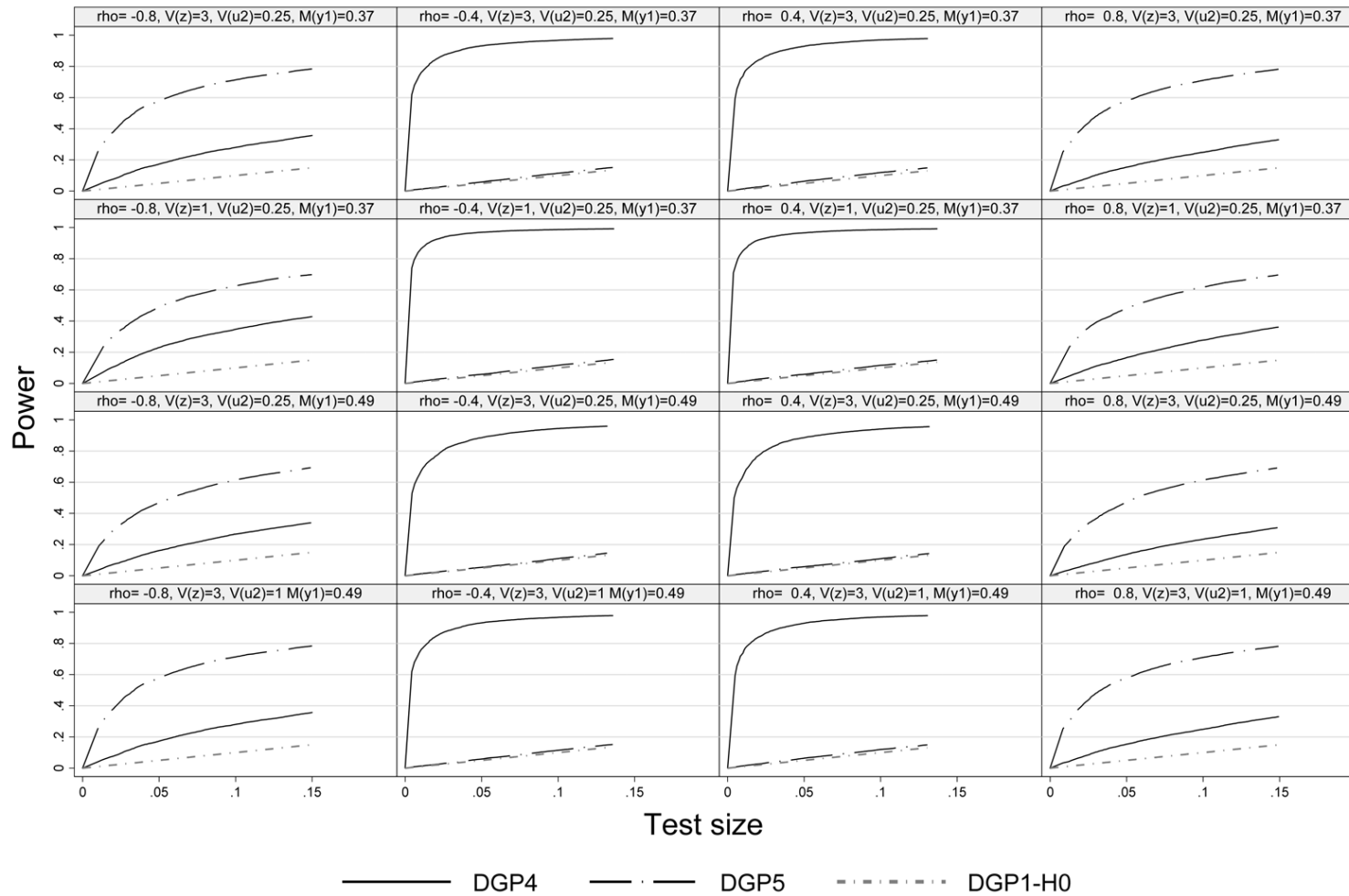
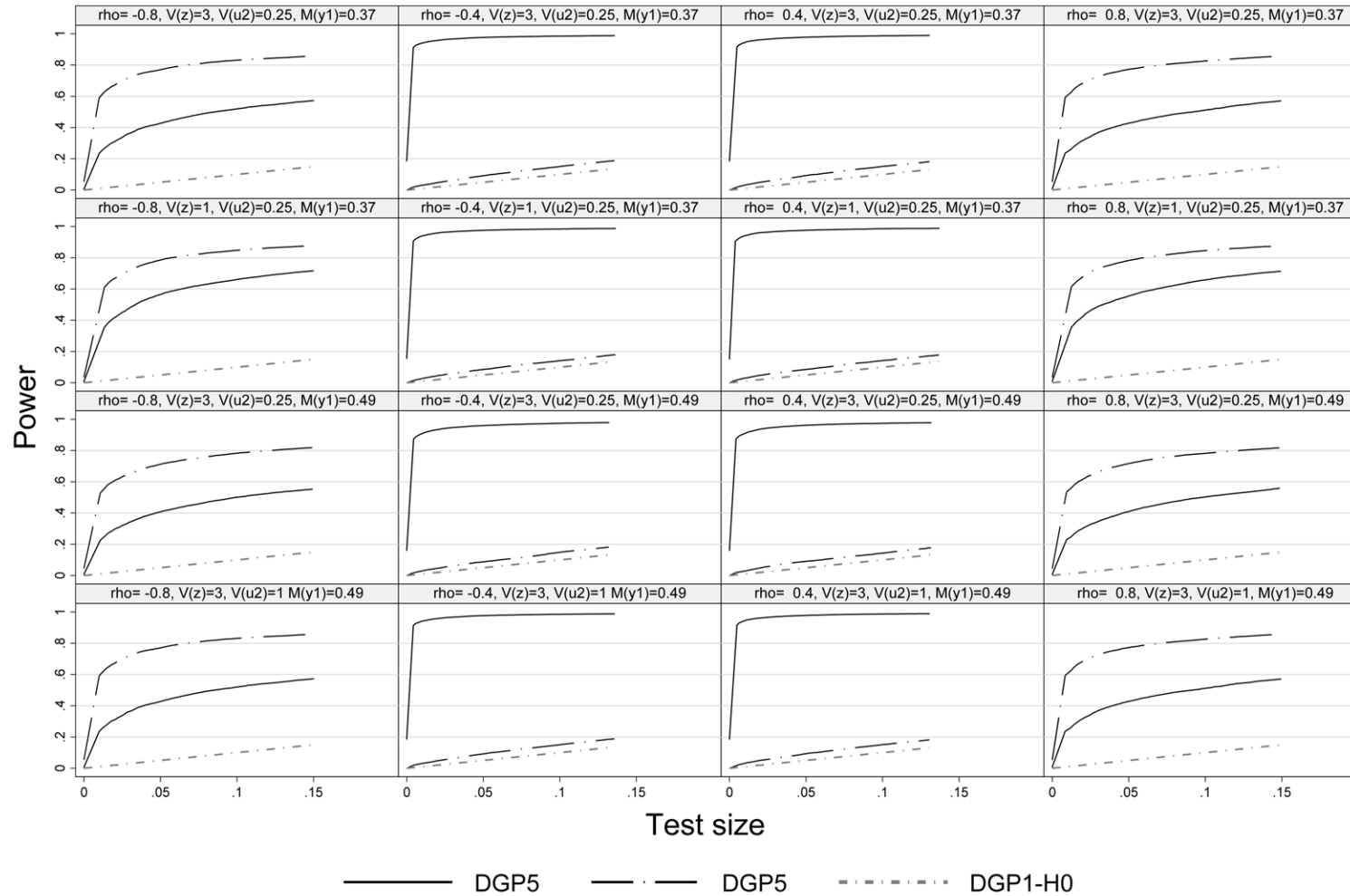


Figure 4. Size power plot, DGP1, DGP6 and DGP7,  $n = 1000$ .



Comparing the first and second row of graphs in Figures 2–4 indicates that there is not much power lost with the weaker exclusion restriction. A higher share of unobserved units tends to slightly reduce the power of the LM test as one would expect (see the graphs in row 3 vs. 1 in Figures 2–4). Comparing the first and the last row in Figures 2–4 indicates that a weaker fit (*i.e.*,  $Var(u_{2i})$  is increased from 0.25 to 1) does not result in a significant loss of power. Lastly, as expected a larger sample size improves the power of the pseudo-score LM test across the board.<sup>4</sup>

#### 4. Conclusions

Using Meijer and Wansbeek's (2007) [10] GMM-approach for two-step estimators of the Heckman sample selection model, this paper introduces a pseudo-score LM test to check the assumption of normality and homoskedasticity of the disturbances, a prerequisite for the consistency of this estimator. The GMM-based pseudo-score LM test is easy to calculate and similar to the widely used Jarque and Bera (1980) [11] LM test. Indeed, in the absence of sample selection it reverts to their LM test statistic. In particular, the test checks whether the third and fourth moments of the disturbances of the outcome equation of the Heckman model conform to those implied by the truncated normal distribution. Under  $H_0$  normal disturbances of the selection equation and the absence of heteroskedasticity in both the outcome and the selection equation are additionally required.

Monte Carlo simulations show good performance of the pseudo-score LM test for samples of size 1000 or larger and a powerful exclusion restriction. However, in line with other tests of the normality assumption of the Heckman sample selection model proposed in the literature the pseudo-score LM test tends to be oversized, although only slightly, if the correlation of the disturbances of the selection and the outcome equation is high in absolute value or if the exclusion restrictions are weak. Hence, this test can be recommended for sample sizes of 1000 or larger.

#### Acknowledgments

I am very grateful to Tom Wansbeek and two anonymous referees for detailed and constructive comments on an earlier draft. A Stata ado-file for this test is available at: <http://homepage.uibk.ac.at/c43236/publications.html>.

#### Appendix

**Deriving**  $E[(\tau v_i + \varepsilon_i)^k | d = 1]$  :

Let  $Z \sim N(0, 1)$  and consider  $\mu_k(a_i) = E[Z^k | Z > a_i]$ ,  $k = 1, 2, \dots$ . The derivation the moments of  $Z^k | Z > a_i$  uses the following recursive formula (Meijer and Wansbeek, 2007, p. 45) [10]:

$$\begin{aligned}\mu_0(a_i) &= 1 \\ \mu_1(a_i) &= \lambda_i \\ \mu_k(a_i) &= (k - 1)\mu_{k-2}(a_i) + a_i^{k-1}\lambda_i, \quad k \geq 2\end{aligned}$$

<sup>4</sup> The corresponding figures for a larger sample size of  $n = 2000$  are available upon request from the author.

Setting  $a_i = -z_i'\gamma$  and abbreviating  $\mu_k(a_i) = \mu_{k,i}$ , one obtains

$$\psi_{k,i} \equiv E[v_i^k | d_i = 1] = \sum_{j=0}^k (-1)^j \binom{k}{j} \mu_{j,i} \lambda_i^{k-j}$$

and based on these results one can calculate the moments of  $(\tau v_i + \varepsilon_i)^k$  as

$$\begin{aligned} \varphi_{k,i} &\equiv E\left[(\tau v_i + \varepsilon_i)^k | d = 1\right] = E\left[\sum_{j=0}^k \binom{k}{j} \varepsilon_i^{k-j} (\tau v_i)^j | d = 1\right] \\ &= \sum_{j=0}^k \binom{k}{j} \mu_{\varepsilon,k-j} \tau^j \psi_{j,i} \end{aligned}$$

where  $\mu_{\varepsilon,k-j} \equiv E[\varepsilon_i^k]$ .

**Pseudo-score LM test:**

Denoting the GMM-estimates under  $H_0$  by  $\tilde{\theta}$ , the pseudo-score LM test can be written as (see Hayashi, 2000, p.491–493 [17], Newey and West, 1987, p. 780 [13] and Hall, 2005, p. 162 [14]) : <sup>5</sup>

$$LM = n\bar{h}'(\tilde{\theta})\Psi_n(\tilde{\theta})^{-1}\bar{G}'(\tilde{\theta})(\bar{G}'(\tilde{\theta})\Psi_n(\tilde{\theta})^{-1}\bar{G}(\tilde{\theta}))^{-1}\bar{G}(\tilde{\theta})\Psi_n(\tilde{\theta})^{-1}\bar{h}(\tilde{\theta})$$

where  $\Psi_n(\tilde{\theta}) = E[\bar{\Psi}(\tilde{\theta})]$  is a consistent estimator of  $\Psi_0$  under  $H_0$ . Using the fact that  $\bar{G}(\tilde{\theta})$  is invertible yields the LM test statistic as

$$LM = n\bar{h}'(\tilde{\theta})\Psi_n(\tilde{\theta})^{-1}\bar{h}(\tilde{\theta})$$

which can be further simplified using the partitioned inverse

$$LM = n\bar{h}_2(\tilde{\theta})' \left( \Psi_{n,22}(\tilde{\theta}) - \Psi_{n,21}(\tilde{\theta})\Psi_{n,11}^{-1}(\tilde{\theta})\Psi_{n,12}(\tilde{\theta}) \right)^{-1} \bar{h}_2(\tilde{\theta})$$

since  $\bar{h}_1(\tilde{\theta}) = 0$ .

**Variance of moments:**

Under fairly general conditions (see Amemiya, 1985, Section 3.4) [18],  $\lim_{n \rightarrow \infty} E[\bar{\Psi}(\theta)] = plim_{n \rightarrow \infty} \bar{\Psi}(\theta)$  and in the formulas for the asymptotic covariance matrix, one can replace  $\bar{\Psi}(\theta)$  by its expectation. Note  $\Psi_n(\theta_0) = E[\bar{\Psi}(\theta_0)]$  can be estimated consistently in the usual way by  $\Psi_n(\tilde{\theta})$ . To obtain the estimate  $\Psi_n(\tilde{\theta})$ , we partition  $\Psi_n(\theta)$  in accordance to  $\bar{h}(\theta) = (\bar{h}_1(\theta)', \bar{h}_2(\theta)')'$  as

$$\Psi_n(\theta) = \begin{bmatrix} \Psi_{n,11}(\theta) & \Psi_{n,12}(\theta) \\ \Psi_{n,12}(\theta)' & \Psi_{n,22}(\theta) \end{bmatrix}$$

<sup>5</sup> Newey and West (1987) [13] propose to use the unrestricted estimator  $\bar{\Psi}(\hat{\theta})$ , a route that is not followed here.

Using

$$h_{1,i}(\theta)h_{1,i}(\theta)' = \begin{bmatrix} (d_i - p_i) \frac{\phi_i z_i}{p_i(1-p_i)} \\ d_i w_i (\tau v_i + \varepsilon_i) \\ d_i [(\tau v_i + \varepsilon_i)^2 - \varphi_{2,i}] \end{bmatrix} \begin{bmatrix} \frac{(d_i - p_i)\phi_i z_i}{p_i(1-p_i)} & d_i w_i' (\tau v_i + \varepsilon_i) & d_i [(\tau v_i + \varepsilon_i)^2 - \varphi_{2,i}] \end{bmatrix} = \begin{bmatrix} (d_i - p_i)^2 \left(\frac{\phi_i}{p_i(1-p_i)}\right)^2 z_i z_i' & \frac{d_i(d_i - p_i)\phi_i}{p_i(1-p_i)} z_i w_i' (\tau v_i + \varepsilon_i) & \frac{(d_i - p_i)\phi_i}{p_i(1-p_i)} d_i [(\tau v_i + \varepsilon_i)^2 - \varphi_{2,i}] z_i \\ * & d_i w_i w_i' (\tau v_i + \varepsilon_i)^2 & d_i [(\tau v_i + \varepsilon_i)^3 - \varphi_{2,i}(\tau v_i + \varepsilon_i)] w_i \\ * & * & d_i [(\tau v_i + \varepsilon_i)^2 - \varphi_{2,i}]^2 \end{bmatrix}$$

one obtains for the off-diagonal elements:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n E[(d_i - p_i) \frac{\phi_i}{p_i(1-p_i)} d_i (\tau v_i + \varepsilon_i) z_i w_i'] &= 0 \\ \frac{1}{n} \sum_{i=1}^n E[\frac{(d_i - p_i)\phi_i}{p_i(1-p_i)} d_i [(\tau v_i + \varepsilon_i)^2 - \varphi_{2,i}] z_i] &= 0 \\ \frac{1}{n} \sum_{i=1}^n E[d_i [(\varepsilon_i + \tau v_i)^3 - \varphi_{2,i}(\tau v_i + \varepsilon_i)] w_i] &= \frac{1}{n} \sum_{i=1}^n p_i \varphi_{3,i} w_i \end{aligned}$$

Some of the explanatory variables summarized in  $w_i$  may not be observed at  $d_i = 0$ . However, one can use the reasoning in Meijer and Wansbeek (2007) [10] and establish

$$\begin{aligned} plim_{n \rightarrow \infty} \frac{1}{n} (W_1' W_1) - \lim_{n \rightarrow \infty} \frac{1}{n} W_1' \Pi W_1 &= plim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n d_i w_i w_i' - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n p_i w_i w_i' = 0 \\ plim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (d_i - p_i) \varphi_{k,i} w_i &= 0, \quad k = 1, \dots, 8 \\ plim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (d_i - p_i) \varphi_{k,i} \varphi_{l,i} &= 0, \quad k, l = 1, \dots, 4. \end{aligned}$$

Here,  $\Pi = diag(p_1, \dots, p_n)$  and  $W_1$  is derived from  $W$  by skipping all rows with  $d_i = 0$ . Hence, one can use

$$\Psi_{n,11}(\theta) = \frac{1}{n} \begin{bmatrix} Z' V Z & 0 & 0 \\ * & W_1' \Sigma_1 W_1 & \sum_{d_i=1} w_i \varphi_{3,i} \\ * & * & \sum_{d_i=1} (\varphi_{4,i} - \varphi_{2,i}^2) \end{bmatrix}$$

where  $V = diag(\frac{\phi_1^2}{p_1(1-p_1)}, \dots, \frac{\phi_n^2}{p_n(1-p_n)})$ ,  $W_{n \times k_2} = (w_1, \dots, w_n)'$ , and  $\Sigma = diag(\varphi_{2,1}, \dots, \varphi_{2,n})$ .  $\Sigma_1$  is obtained from  $\Sigma$  by deleting all rows and columns referring to  $d_i = 0$ , and similarly  $W_1$ . Similar arguments yield at  $\xi = \kappa = 0$

$$\Psi_{n,22}(\theta) = \frac{1}{n} \sum_{i=1}^N \begin{bmatrix} p_i (\varphi_{6,i} - \varphi_{3,i}^2) & p_i (\varphi_{7,i} - \varphi_{3,i} \varphi_{4,i}) \\ p_i (\varphi_{7,i} - \varphi_{3,i} \varphi_{4,i}) & p_i (\varphi_{8,i} - \varphi_{4,i}^2) \end{bmatrix}$$

and

$$\Psi_{n,12}(\theta) = \frac{1}{n} \sum_{i=1}^n \begin{bmatrix} 0 & 0 \\ p_i w_i \varphi_{4,i} & p_i w_i \varphi_{5,i} \\ p_i (\varphi_{5,i} - \varphi_{2,i} \varphi_{3,i}) & p_i (\varphi_{6,i} - \varphi_{4,i} \varphi_{2,i}) \end{bmatrix}$$



Again, we can insert  $d_i$  for  $p_i$ . Applying the formula for the partitioned inverse yields the simplification of the pseudo-score LM test statistic:

$$\begin{aligned} LM &= n\bar{h}'(\tilde{\theta})\Psi_n(\tilde{\theta})^{-1}\bar{h}'(\tilde{\theta}) \\ &= n\left[0, \bar{h}_2(\tilde{\theta})'\right] \begin{bmatrix} \Psi_{n,11}(\tilde{\theta}) & \Psi_{n,12}(\tilde{\theta}) \\ \Psi_{n,21}(\tilde{\theta}) & \Psi_{n,22}(\tilde{\theta}) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \bar{h}_2(\tilde{\theta}) \end{bmatrix} \\ &= n\bar{h}_2(\tilde{\theta})' \left( \Psi_{n,22}(\tilde{\theta}) - \Psi_{n,21}(\tilde{\theta})\Psi_{n,11}(\tilde{\theta})^{-1}\Psi_{n,12}(\tilde{\theta}) \right)^{-1} \bar{h}_2(\tilde{\theta}) \end{aligned}$$

which is asymptotically distributed as  $\chi^2(2)$  under  $H_0$ .

### Conflicts of Interest

The author declares no conflict of interest.

### References

1. Yen, S.T.; Rosinski, J. On the marginal effects of variables in the log-transformed sample selection models. *Econ. Lett.* **2008**, *100*, 4–8.
2. Staub, K.E. A causal interpretation of extensive and intensive margin effects in generalized Tobit models. *Rev. Econ. Stat.* **2014**, *96*, 371–375.
3. Newey, W.K. Two-step series estimation of sample selection models. *Econom. J.* **2009**, *12* (Suppl. 1), 217–229.
4. Skeels, C.L.; Vella, F. A Monte Carlo investigation of the sampling behavior of conditional moment tests in Tobit and Probit models. *J. Econom.* **1999**, *92*, 275–294.
5. Drukker, D.M. Bootstrapping a conditional moments test for normality after Tobit estimation. *Stata J.* **2002**, *2*, 125–139.
6. Bera, A.K.; Jarque, C.M.; Lee, L.-F. Testing the normality assumption in limited dependent variable models. *Int. Econ. Rev.* **1984**, *25*, 563–578.
7. Lee, L.-F. Tests for the bivariate normal distribution in econometric models with selectivity. *Econometrica* **1984**, *52*, 843–863.
8. Van der Klaauw, B.; Koning, R.H. Testing the normality assumption in the sample selection model with and application to travel demand. *J. Bus. Econ. Stat.* **1993**, *21*, 31–42.
9. Montes-Rojas, G.V. Robust misspecification tests for the Heckman's two-step estimator. *Econom. Rev.* **2011**, *30*, 154–172.
10. Meijer, E.; Wansbeek, T. The sample selection model from a method of moments perspective. *Econom. Rev.* **2007**, *26*, 25–51.
11. Jarque, C.; Bera, A. Efficient tests for normality, homoskedasticity and serial independence of regression residuals. *Econ. Lett.* **1980**, *6*, 255–259.
12. Heckman, J.J. Sample selection bias as a specification error. *Econometrica* **1979**, *47*, 153–161.
13. Newey, W.K.; West, K.D. Hypothesis testing with efficient method of moments estimation. *Int. Econ. Rev.* **1987**, *28*, 777–787.
14. Hall, A.R. *Generalized Methods of Moments*; Oxford University Press: Oxford, UK, 2005.

15. Ahn, S.C.; Schmidt, P. A separability result for GMM estimation, with applications to GLS prediction and conditional Moment Tests. *Econom. Rev.* **1995**, *14*, 19–34.
16. Davidson, R.; MacKinnon, J.G. Graphical methods for investigating the size and power of hypothesis tests. *Manch. Sch.* **1998**, *66*, 1–26.
17. Hayashi, F. *Econometrics*; Princeton University Press: Princeton, NJ, USA; Oxford, UK, 2000.
18. Amemiya, T. *Advanced Econometrics*; Harvard University Press: Harvard, UK, 1985.

© 2014 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).