Will it always be necessary taking into account sample selection?

João Gomes[1], Tiago Oliveira

Departamento de Estatística e Investigação Operacional, Faculdade de Ciências da Universidade de Lisboa, 1749-016 Lisboa, Portugal

Instituto Superior de Estatística e Gestão Informação, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa , Portugal

jjgomes@fc.ul.pt, toliveira@isegi.unl.pt

Abstract

We compare the results obtained by sequential model (binary model for selection equation and the simple ordinary least squares (OLS) for selected equation) with the results obtained by Tobit II model estimation. We prove analytically and show by Monte Carlo simulation that the Tobit II only gives better results than OLS when the error correlation is different from zero, and when the index equation has more information than the structural equation. Therefore, we propose two new measures to quantify this lack of information.

Key-words: sample selection models; correlation errors; conditional expected values; Tobit II; hurdle; probit; OLS

JEL classification: C15; C24

[1] Work partially supported by Fundação para a Ciência e Tecnologia (FCT)-Plurianual 2009, under CMAF

1. Introduction

The capacity to estimate and test regression models over non randomly chosen subsamples is unquestionably one of the most significant innovations in regression models. Econometric analysis of the labour supply and wage function have frequently used a sample selection model named Tobit II. This model was a generalization of the standard Tobit I in econometrics, initially introduced by (Tobin, 1958).

In the sample selection model (Tobit II), the possibility of sample selection bias arises whenever one examines a subsample, and the unobservable factors determining inclusion in the subsample, are correlated with other unobservable factors that interfere on the variable of primary interest. For details of the model see (Amemiya, 1984, Heckman, 1978, Heckman, 1979, Newey, 1999, Vella, 1998).

There are a few recent studies that compare the Tobit I and II by Monte Carlo simulation (Flood and Gråsjö, 2001, Nawata, 2007). In this paper we compare the Tobit II with a sequential model (binary model for selection equation and simple ordinary least squares (OLS) for selected equation) usually referred as the hurdle model. Our work proves analytically that in OLS the inverse Mill's ratio is projected by hat matrix associated to model of primary interest and shows by Monte Carlo simulation that, when the binary equation does not have more information than the selected equation, the latter does not require information about the former. In this case the OLS and the Tobit II are similar, as shown by the estimates of marginal effects and root mean square error (RMSE).

The aim of this paper is to identify which situations require taking account the sample selection and to propose two measures to help upon this decision.

The remainder of this paper is organized as follows. The next section provides a basic framework. In section 3 we compare the conditional mean value of the two models and develop the two new measures referred above. In section 4 we use the Monte Carlo simulation to exemplify the theoretical results obtained. Finally, in section 5 we present our conclusions.

2. Basic framework

This model can be seen as a bivariate model. Initially, we have two latent variables $\mathbf{y}_1^* = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$ (structural equation) and $\mathbf{y}_2^* = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2$ (index equation).

$\mathbf{X_j}$ is the model matrix to $\mathbf{y_j^*}$, with dimension $n_j x k_j$, for j=1, 2 and $n_1<n_2$. We assume that the first column of $\mathbf{X_1}$ is **1** and $\mathbf{x_j}$ is a row of $\mathbf{X_j}$ such that to each observation, $y_j^* = \eta_j + \varepsilon_j$ with $\eta_j = \mathbf{x_j}\boldsymbol{\beta}_\mathbf{j}$. We further assume that $\varepsilon \equiv (\varepsilon_1, \varepsilon_2)$ is $\mathbf{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$.

A true observation is not (y_1^*, y_2^*) but rather (y_1, y_2), so that we have $y_2 = \begin{cases} 0 & y_2^* \le 0 \\ 1 & y_2^* > 0 \end{cases}$ and the following observation rule: $\begin{cases} y_1 = y_1^*, y_2 = 1 & if\ y_2^* > 0 \\ y_1 \quad \text{not observed}, y_2 = 0 & if\ y_2^* \le 0 \end{cases}$.

Therefore y_1 is censored if y_2=0, as this is a sample selection a bias is to be expected, meaning that the estimator of $E(y_1)$ to OLS based on the sub-sample, will be biased and should not be used.

Our main goal is to identify which sample selection situations are to be taken into account.

The Likelihood

The derivation of the likelihood function for the sample selection which accounts for the correlation errors model is presented in (Amemiya, 1984). In the literature this model is known as Tobit II.

For each observation we have $L\left(y_1,y_2\right)=\left[P\left(y_2=y_2\right)\right]^{1-y_2}\left[P\left(y_2=y_2\mid y_1\right)f_{y_1}\left(y_1\right)\right]^{y_2}$

where $y_2 \in \{0,1\}$ *and* $y_1 \in \mathbb{R}$.

We know that

$$\varepsilon_2 \mid y_1 \sim N\left(0+\rho\frac{\sigma_1}{\sigma_2}\left(y_1-\eta_1\right),\sigma_1\sqrt{\left(1-\rho^2\right)}\right);$$

thus we have

$$P\left(y_2=1\mid y_1\right)=P\left(\varepsilon_2>-\eta_2\mid y_1\right)=\Phi\left[\frac{\eta_2+\rho\frac{\sigma_1}{\sigma_2}\left(y_1-\eta_1\right)}{\sigma_1\sqrt{\left(1-\rho^2\right)}}\right]$$

and

$$L\left(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,y_1,y_2,\sigma_1^2,\sigma_{12}\mid \mathbf{X}_1,\mathbf{X}_2\right)=\prod\left[1-\Phi\left(\eta_2\right)\right]^{1-y_2}\left[\Phi\left[\frac{\eta_2+\rho\frac{\sigma_1}{\sigma_2}\left(y_1-\eta_1\right)}{\sigma_1\sqrt{\left(1-\rho^2\right)}}\right]\frac{\phi\left(\frac{\left(y_1-\eta_1\right)}{\sigma_1}\right)}{\sigma_1}\right]^{y_2}$$

where ϕ and Φ are respectively the standard gaussian density and distribution functions.

Conditional mean in the sample selection model

The conditional mean of y_1 is given by:

$$E\left(y_1 \mid y_2=1\right)=E\left(\eta_1+\varepsilon_1 \mid y_2^*>0\right)=E\left(\eta_1+\varepsilon_1 \mid \eta_2+\varepsilon_2>0\right)$$
$$=\eta_1+E\left(\varepsilon_1 \mid \varepsilon_2>-\eta_2\right)=\eta_1+\rho\frac{\sigma_1}{\sigma_2}\lambda\left(\sigma_2^{-1}\eta_2\right)$$

where $\lambda(x)=\dfrac{\phi(x)}{\Phi(x)}$.

Note: the model of the index equation is a standard probit model, describing the choice y_2=1 or y_2=0. Therefore, a normalization restriction is required and one usually sets $\sigma_2^2=1$ such that $E\left(y_1 \mid y_2=1\right)=\eta_1+\rho\sigma_1\lambda\left(\eta_2\right)$ and $E\left(y_1\right)=\Phi\left(\eta_2\right)\eta_1+\rho\sigma_1\phi\left(\eta_2\right)$.

3. Comparing the Conditional Mean in OLS and Tobit II

Inverse Mills ratio and Hat Matrix

Let $\mathbf{H}=\mathbf{X}_1(\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1)^{-1}\mathbf{X}_1^{\mathrm{T}}$ be the hat matrix resulting from the sample selection process. This matrix is a projection matrix of $\mathbf{y}_1 \mid \mathbf{y}_2=\mathbf{1}$ in the subspace defined from $\mathbf{X}_1$. An interesting property of $\mathbf{H}$ is that, when $\boldsymbol{\eta}$ is a linear combination of the columns of $\mathbf{X}_1$, then $\mathbf{H}\boldsymbol{\eta}\equiv\boldsymbol{\eta}$.

The ratio $r(t)=\dfrac{\Phi(-t)}{\phi(t)}$ is known as Mill's ratio and its reciprocal $\lambda^*(t)=\dfrac{\phi(t)}{\Phi(-t)}$ is the so-called failure rate for the standard normal law. Here, our interest is in $\lambda(t)=\lambda^*(-t)=\dfrac{\phi(t)}{\Phi(t)}$.

In the literature there are many asymptotic results for Mill's ratio, and thus for $\lambda^*(t)$, when $t\to\infty$ (e.g. (Pinelis, 2002)). As the sample selection needs λ when t is in the neighbourhood of zero, such approaches are not interesting to us. Instead we can develop the following result:

Proposition 1 – If $\mathbf{H} = \mathbf{X}_1(\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T$ is the hat matrix of $\mathbf{y}_1 \mid \mathbf{y}_2 = 1$ and $\boldsymbol{\eta} = \mathbf{X}_1\boldsymbol{\beta}$ is a linear combination of the columns of $\mathbf{X}_1$ then using a Taylor expansion of the $\boldsymbol{\lambda}(\boldsymbol{\eta})$ around $\boldsymbol{\lambda}(\mathbf{0})$ with $\boldsymbol{\lambda}(\boldsymbol{\eta}) \equiv \boldsymbol{\lambda}$ and **M=I-H,** we obtain $\boldsymbol{\lambda}^T\mathbf{M}\boldsymbol{\lambda} \approx \frac{(4-\pi)^2}{2\pi^3}\boldsymbol{\eta}^{2^T}\mathbf{M}\boldsymbol{\eta}^2$, in the neighbourhood of zero ($\mathbf{f(x)}$ represents the vector whose components are $f(x)$).

As $\boldsymbol{\lambda}(\mathbf{0}) = \boldsymbol{\lambda}'(\mathbf{0}) = \mathbf{0}$, $\boldsymbol{\lambda} \approx \frac{\lambda''(0)}{2}\boldsymbol{\eta}^2$, then

$$\boldsymbol{\lambda} - \mathbf{H}\boldsymbol{\lambda} = \mathbf{M}\boldsymbol{\lambda} \approx \frac{\lambda''(0)}{2}\mathbf{M}\boldsymbol{\eta}^2 = \frac{4-\pi}{\sqrt{2\pi^3}}\mathbf{M}\boldsymbol{\eta}^2.$$

The mean values

As mentioned above, if $\mathbf{y} \equiv (\mathbf{y}_1 \mid \mathbf{y}_2 = \mathbf{1}, \mathbf{X}_1, \mathbf{X}_2)$, $\boldsymbol{\eta}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ and $\boldsymbol{\eta}_2 = \mathbf{X}_2\boldsymbol{\beta}_2$ for the rows of $\mathbf{X}_2$ where y_2=1, then

$$\mathbf{E}(\mathbf{y}) = \boldsymbol{\eta}_1 + \rho\sigma_1\boldsymbol{\lambda}(\boldsymbol{\eta}_2). \tag{1}$$

Moreover,

$$\mathbf{E}(\hat{\mathbf{y}}_{\mathbf{ols}}) = \mathbf{E}\left(\mathbf{X}_1\hat{\boldsymbol{\beta}}_{\mathbf{ols}}\right) = \mathbf{X}_1\mathbf{E}\left(\hat{\boldsymbol{\beta}}_{\mathbf{ols}}\right) = \mathbf{X}_1\left(\mathbf{X}_1^T\mathbf{X}_1\right)^{-1}\mathbf{X}_1^T\mathbf{E}(\mathbf{y}) = \boldsymbol{\eta}_1 + \rho\sigma_1\mathbf{H}\boldsymbol{\lambda}(\boldsymbol{\eta}_2). \tag{2}$$

What information can be lost when we replace $\boldsymbol{\lambda}(\boldsymbol{\eta}_2)$ by $\mathbf{H}\boldsymbol{\lambda}(\boldsymbol{\eta}_2)$?

Bias

$$\mathbf{B} = \mathbf{Bias}(\hat{\mathbf{y}}_{\mathbf{ols}}) = \mathbf{E}(\mathbf{y}) - \mathbf{E}(\hat{\mathbf{y}}_{\mathbf{ols}}) = \rho\sigma_1(\mathbf{I}-\mathbf{H})\boldsymbol{\lambda}(\boldsymbol{\eta}_2) = \rho\sigma_1\mathbf{M}\boldsymbol{\lambda}(\boldsymbol{\eta}_2).$$

Result

If the first column of $\mathbf{X}_1$ is **1** then $\mathbf{X}_1^T\mathbf{B} = \rho\sigma_1\mathbf{X}_1\mathbf{M}\boldsymbol{\lambda}(\boldsymbol{\eta}_2) = \mathbf{0}$ such that $\mathbf{1}^T\mathbf{B} = \mathbf{0}$ or

$\sum_{i=1}^{m} B_i = 0$.

Given this result, if we define $\overline{E(y)} = \frac{\sum E(y_1 \mid y_2 = 1)}{n_1}$ and $\overline{E(\hat{y}_{OLS})} = \frac{\sum E(\hat{y}_{OLS})}{n_1}$, we can then state that

$$\left(\mathbf{E(y)} - \overline{\mathbf{E(y)}}\right)^T \left(\mathbf{E(y)} - \overline{\mathbf{E(y)}}\right) = \left(\mathbf{E(\hat{y}_{OLS})} - \overline{\mathbf{E(\hat{y}_{OLS})}}\right)^T \left(\mathbf{E(\hat{y}_{OLS})} - \overline{\mathbf{E(\hat{y}_{OLS})}}\right) + \mathbf{B}^T\mathbf{B}.$$

Measure 1

We can measure the loss of information (**LI**) of OLS when compared with Tobit II by

$$\mathbf{LI} = \frac{\mathbf{B^T B}}{\mathbf{E}\left((\mathbf{y}) - \overline{\mathbf{E(y)}}\right)^T \left(\mathbf{E(y)} - \overline{\mathbf{E(y)}}\right)}.$$

Therefore 0< **LI**<1 where an **LI** near 0 means that there is no loss of information.

Remark

- When $\mathbf{X_2} \subset \mathbf{X_1}$ (the columns of $\mathbf{X_2}$ are in $\mathbf{X_1}$) then $\boldsymbol{\eta}_2$ is a linear combination of $\mathbf{X_1}$, and by proposition 1, we can write

$$\mathbf{LI} \approx \frac{\rho^2 \sigma_1^2 \frac{(4-\pi)^2}{2\pi^3} \boldsymbol{\eta}_2^{2T} \mathbf{M} \boldsymbol{\eta}_2^2}{\left(\mathbf{E(y)} - \overline{\mathbf{E(y)}}\right)^T \left(\mathbf{E(y)} - \overline{\mathbf{E(y)}}\right)}.$$

This value is close to zero and only has some impact when $\boldsymbol{\eta}_2$ is far from **zero**.

Measure 2

We further notice that $(\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}})^{\mathbf{T}}(\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}}) = \boldsymbol{\lambda}^{\mathbf{T}}\mathbf{M}\boldsymbol{\lambda} + (\mathbf{H}\boldsymbol{\lambda} - \overline{\mathbf{H}\boldsymbol{\lambda}})^{\mathbf{T}}(\mathbf{H}\boldsymbol{\lambda} - \overline{\mathbf{H}\boldsymbol{\lambda}})$ (with $\boldsymbol{\lambda}(\boldsymbol{\eta}_2) = \boldsymbol{\lambda}$).

Then

$$\mathbf{LI^*} = \frac{\boldsymbol{\lambda}^{\mathbf{T}}\mathbf{M}\boldsymbol{\lambda}}{(\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}})^{\mathbf{T}}(\boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}})}$$

is another measure of loss of information in OLS, but now referring only to λ. This new measure does not depend on ρ or σ_1, but it only makes sense to calculate it when $\rho \neq 0$. In summary, if we use the method of moments to estimate **LI** and **LI*,** we now have simple measures to quantify the loss of information when using OLS instead of Tobit II in the structural equation.

4. Monte Carlo Simulation

In order to evaluate the differences between these models a Monte Carlo simulation will be used. Our study applies a Tobit II for the data generation process (DGP). Furthermore, we have extensively used the software R in all our simulations, namely for developing the routines that implement the Tobit II. To our knowledge, such model has never been reported before in R. The experiments are based on a sample size of 2,000 observations; the number of trials is 500 in each case. The covariates, $\mathbf{x_1}$, $\mathbf{x_2}$ and $\mathbf{x_3}$, are random numbers with uniform distributions (0,b). In DGP we developed three different models to address the following questions.

Question1: what happens if $\mathbf{X_2}$ has at least one more variable that $\mathbf{X_1}$? To address this question we have estimated <u>model 1</u>, where data were generated for the structural and index equations with $\mathbf{X_1} = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2]$, $\mathbf{X_2} = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$, $\boldsymbol{\beta}_1^T = [0.1, \text{-}0.2, 2]$, $\sigma_1 = 2$ and $\boldsymbol{\beta}_2^T = [-4.5, \text{-}0.1, 0.5, 1]$. We have also set b=20 for $\mathbf{x_1}$ and b=5 for $\mathbf{x_2}$ and $\mathbf{x_3}$, leading to the results that are in Table 1 and 2. In Table 1, we can see that the percentage of bias and root mean square error (RMSE) are higher for the hurdle model in the structural equation, especially when $|\rho|$ tends to 1. This difference is higher because $\mathbf{x}_3$ is only in the index equation, and the hurdle model doesn´t consider correlation errors stemming from sample selection. The **LI** value shows that the OLS gives rise to a normalized bias of 5% and 9% in **E(y1|y2=1)** when ρ=-0.8 or 0.8, respectively. **LI*** = 0.72 indicates that the use of **H**$\boldsymbol{\lambda}$ instead of $\boldsymbol{\lambda}$ yields a normalized impact of 72% concerning $\boldsymbol{\lambda}$.

Table 1 TO INSERT HERE

In Table 2 we can see that when $|\rho|$ tends to 1, the bias in estimated marginal effects increases. For ρ=0.8 the bias in the hurdle estimated marginal effects for $\mathbf{x}_1$ is approximately 31%. This reveals that for model 1 it is imperative to take into account sample selection.

Table 2 TO INSERT HERE

Question 2: what happens if $\mathbf{X}_2$ is equal to $\mathbf{X}_1$? <u>Model 2</u> was estimated in such a situation. The model for the DGP was $\mathbf{X}_1 = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2]$, $\mathbf{X}_2 = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2]$, $\boldsymbol{\beta}_1^T = [0.1, \text{-}0.2, 2]$, $\sigma_1 = 2$ and $\boldsymbol{\beta}_2^T = [-3.5, \text{-}0.1, 1]$. We also set b=20 for $\mathbf{x}_1$ and b=5 for $\mathbf{x}_2$, leading to the results in Table 3 and 4. The Tobit II and the hurdle models, have both a poor performance estimating marginal effects and RMSE. Furthermore, **LI** and **LI*** reveal that it is not important to take into account a sample selection.

Table 3 TO INSERT HERE

Table 4 TO INSERT HERE

Question 3: what happens if $\mathbf{X}_2$ has fewer variables than $\mathbf{X}_1$? To address this question we have <u>model 3</u>. For this model the DGP is $\mathbf{X}_1 = [\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2]$, $\mathbf{X}_2 = [\mathbf{1}, \mathbf{x}_1]$, $\boldsymbol{\beta}_1^T = [0.1, \text{-}0.2, 2]$,

$\sigma_1 = 2$ and $\boldsymbol{\beta}_2^T = [1, -0.3]$. We have set b=20 for $\mathbf{x_1}$ and $\mathbf{x_2}$. The results, in table 5 and 6, show that the marginal effect and the RMSE for $\mathbf{x_1}$ were not accurately estimated. Again, our measures, **LI** and **LI***, reveals accounting for sample selection is not helpful.

Table 5 TO INSERT HERE

Table 6 TO INSERT HERE

As mentioned in section 3 above, Figure 1 reveals that model 1 is the only one where "sample selection" is to be taken into account.

Figure 1 TO INSERT HERE

5. Conclusions

The sample selection process is not random only when the null hypothesis (H_0: $\rho=0$) is rejected. When $\rho \neq 0$ the OLS mean value estimator of $y_1|y_2=1$ is given by (2) instead (1). Thus, the loss of information is limited to the change of λ by $H\lambda$.

This non-random sample selection has significant bias if explanatory variables of the index equation ($\mathbf{X_2}$) contain additional information concerning the explained variable ($\mathbf{y_1}$), beyond the one in the explanatory variables of the structural equation ($\mathbf{X_1}$). We have developed two measures to quantify this additional information obtained from a Tobit II, in instead of a sequential model. One depends on ρ and σ, and the other depends only on additional information that $\mathbf{X_2}$ contains for $\mathbf{y_1}$. We have conducted several Monte Carlo simulations that corroborate these theoretical results.

References

Amemiya, T., 1984, Tobit Models - a Survey, Journal of Econometrics 24, 3-61.

Flood, L. and U. Gråsjö, 2001, A Monte Carlo simulation study of Tobit models, Applied Economics Letters 8, 581-584.

Heckman, J. J., 1978, Dummy Endogenous Variables in a Simultaneous Equation System, Econometrica 46, 931-959.

Heckman, J. J., 1979, Sample Selection Bias as a Specification Error, Econometrica 47, 153-161.

Nawata, K., 2007, A monte carlo analysis of the type II tobit maximum likelihood estimator when the true model is the type I tobit model, Economics Bulletin 3, 1-10.

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

Pinelis, I., 2002, Monotonicity properties of the relative error of a padé aproximation for Mill's ratio, Journal of Inequalities in Pure and Applied Mathematics 3, 1-8.

Tobin, J., 1958, Estimation of relationships for limited dependent variables, Econometrica 26, 24--36.

Vella, F., 1998, Estimating models with sample selection bias: A survey, Journal of Human Resources 33, 127-169.

Table 1 – Monte Carlo simulation, 2,000 observations for model 1

Para-meters	True value	$\rho = -0.8$				$\rho = 0$				$\rho = 0.8$			
		Tobit II		**Hurdle**		**Tobit II**		**Hurdle**		**Tobit II**		**Hurdle**	
		% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse
β_{10}	0.1	-1.20%	0.291	1653.30%	1.669	11.30%	0.366	-3.00%	0.268	-2.50%	0.300	-1652.40%	1.671
β_{11}	-0.2	-0.45%	0.017	-27.75%	0.058	-0.30%	0.020	-0.15%	0.019	0.55%	0.017	27.90%	0.058
β_{12}	2	-0.02%	0.071	-14.37%	0.295	-0.11%	0.085	0.02%	0.075	0.06%	0.077	14.42%	0.296
β_{20}	-4.5	-0.90%	0.242	-0.89%	0.249	-0.48%	0.240	-0.48%	0.249	-0.54%	0.224	-0.55%	0.234
β_{21}	-0.1	-0.30%	0.009	-0.40%	0.009	-0.80%	0.009	-0.80%	0.009	-0.90%	0.009	-1.00%	0.009
β_{22}	0.5	-0.74%	0.035	-0.80%	0.036	-0.76%	0.038	-0.76%	0.038	-0.20%	0.037	-0.24%	0.038
β_{23}	1	-0.83%	0.055	-0.84%	0.057	-0.45%	0.054	-0.45%	0.054	-0.79%	0.051	-0.79%	0.054
ρ		-0.02%	0.044	-	-	-	0.109	-	-	-0.30%	0.040	-	-
σ	2	0.36%	0.090	11.09%	0.231	0.20%	0.075	-0.01%	0.074	0.48%	0.082	11.29%	0.233
LI		4.96%				0.00%				8.91%			
LI*		71.59%				-				71.52%			

Table 2 – Bias in estimated marginal effects for structural equation in model 1

Variable	$\rho = -0.8$		$\rho = 0$		$\rho = 0.8$	
	Tobit II	**Hurdle**	**Tobit II**	**Hurdle**	**Tobit II**	**Hurdle**
x1	-0.93%	-11.81%	-0.40%	0.55%	2.18%	30.85%
x2	-0.40%	-6.58%	-0.12%	0.17%	0.37%	10.36%

Table 3 – Monte Carlo simulation, 2,000 observations for model 2

Parameters	True value	$\rho = -0.8$				$\rho = 0$				$\rho = 0.8$			
		Tobit II		**Hurdle**		**Tobit II**		**Hurdle**		**Tobit II**		**Hurdle**	
		% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse
β_{10}	0.1	622.10%	2.776	5142.00%	5.168	-0.10%	3.123	-64.90%	0.701	-643.40%	2.901	-5139.50%	5.156
β_{11}	-0.2	-6.80%	0.060	-51.00%	0.104	-1.25%	0.067	-0.70%	0.023	7.40%	0.059	50.85%	0.104
β_{12}	2	-6.65%	0.587	-53.48%	1.077	0.09%	0.656	0.81%	0.171	6.89%	0.609	53.34%	1.072
β_{20}	-3.5	-0.75%	0.213	-0.62%	0.214	-0.05%	0.220	-0.02%	0.216	-0.56%	0.213	-0.43%	0.212
β_{21}	-0.1	-0.70%	0.009	-0.40%	0.009	-0.60%	0.009	-0.60%	0.009	-0.30%	0.009	-0.10%	0.009
β_{22}	1	-0.77%	0.057	-0.59%	0.058	-0.18%	0.062	-0.13%	0.060	-0.63%	0.059	-0.46%	0.059
ρ		11.34%	0.427	-	-	-	0.435	-	-	12.08%	0.438	-	-
σ	2	2.06%	0.189	21.93%	0.443	-5.77%	0.185	0.08%	0.086	2.19%	0.199	22.16%	0.448
LI		0.21%				0.00%				2.17%			
LI*		1.77%				-				1.78%			

Table 4 – Bias in estimated marginal effects for model 2

Variable	$\rho = -0.8$		$\rho = 0$		$\rho = 0.8$	
	Tobit II	**Hurdle**	**Tobit II**	**Hurdle**	**Tobit II**	**Hurdle**
x1	-4.98%	3.71%	-1.26%	1.31%	10.81%	-9.98%
x2	-5.08%	5.41%	0.16%	-0.20%	11.45%	-14.56%

Table 5 – Monte Carlo simulation, 2,000 observations for model 3

Para-meters	True value	ρ = –0.8				ρ = 0				ρ = 0.8			
		Tobit II		**Hurdle**		**Tobit II**		**Hurdle**		**Tobit II**		**Hurdle**	
		% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse	% Bias	Rmse
β_{10}	0.1	37.40%	0.243	296.10%	0.350	-0.40%	0.304	-7.90%	0.257	-50.10%	0.247	-297.70%	0.353
β_{11}	-0.2	-19.90%	0.169	-158.05%	0.318	-1.20%	0.208	-0.20%	0.044	23.70%	0.182	159.00%	0.320
β_{12}	2	0.03%	0.014	0.05%	0.014	0.05%	0.017	0.05%	0.019	-0.04%	0.013	-0.06%	0.014
β_{20}	1	-0.54%	0.083	-0.30%	0.084	-0.02%	0.082	0.08%	0.085	-0.31%	0.080	-0.05%	0.081
β_{21}	-0.3	-0.27%	0.015	-0.13%	0.015	0.03%	0.014	0.10%	0.015	-0.30%	0.015	-0.13%	0.016
ρ		12.28%	0.428	-	-		0.459	-	-	14.36%	0.456	-	-
σ	2	1.91%	0.176	20.70%	0.418	-5.21%	0.187	-0.31%	0.077	2.09%	0.180	20.80%	0.420
LI		0.01%				0.00%				0.01%			
LI*		1.39%				-				1.38%			

Table 6 – Bias in estimated marginal effects for model 3

	$\rho = -0.8$		$\rho = 0$		$\rho = 0.8$	
Variable	**Tobit II**	**Hurdle**	**Tobit II**	**Hurdle**	**Tobit II**	**Hurdle**
x1	-8.92%	12.88%	-1.57%	1.42%	-94.93%	130.02%
x2	0.03%	-0.05%	0.05%	-0.05%	-0.04%	0.06%

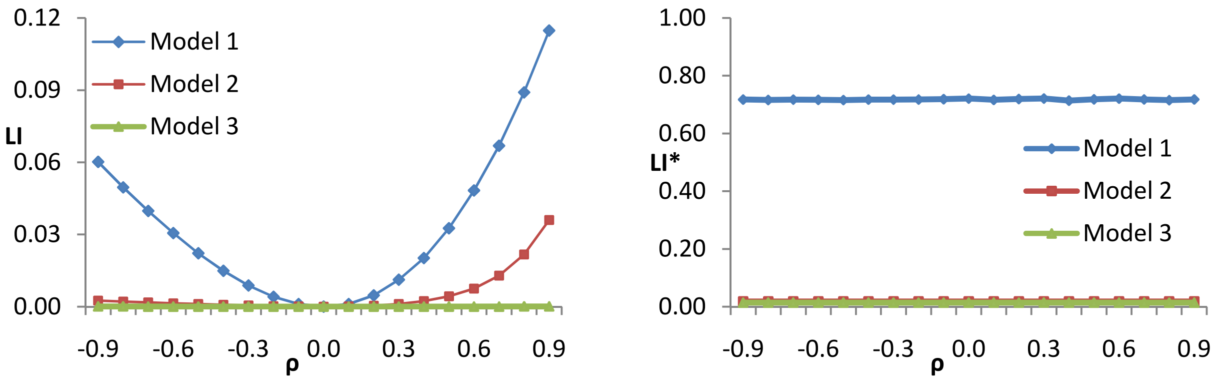

Figure 1 – Measures of **LI** and **LI*** to ρ goes from zero to one in three models