# What about those zeroes?

# A Monte Carlo study of estimators for corner solution outcomes

**Bachelor's Thesis in Statistics, 15 HEC**

**Spring 2014**

**Author: Olof Bjerstaf**

**Supervisor: Xiangping Liu**

# What about those zeroes?

# A Monte Carlo study of estimators for corner solution outcomes

*By: Olof Bjerstaf*

## Abstract

*Partially continuous and limited dependent variables are commonly observed in empirical research. This thesis focuses on evaluating Tobit-type models, which are designed for these situations, with specific regard to small samples. By using a comparatively broader specification then has previously, model performance is evaluated for a range of scenarios. The results indicate that the simple Tobit has best small sample properties, although the bias diminishes above 100 observations for all models. Furthermore, all models are less biased for a log rather than level specification, suggesting that the former is preferable use in practice.*

Empirical researchers frequently encounters situation in which the dependent variable is substantially restricted, for example, bounded above zero. Models designed for these outcomes are generally referred to as limited dependent variable models, of which censored models is one among them. Depending on the cause for the limited range of the dependent variable, the censored models can be categorized into censored regression models and corner solution models. In a censored regression model, the (latent) dependent variable has a true value for everyone in the population. However, we only observe the values in a certain range due to data problems. While in a corner solution model, the restricted range of the outcome is the optimal decision of an individual and we observe the real outcome. The topic of thesis is to evaluate corner solution models.

Corner solution outcomes are typically observed for health expenditure. Whether an individual spends money on a doctor visit or not depends on the health status of the individual or perhaps the individual's perceived health status. Many persons in such a sample have zero expenditure, not because we cannot observe their visits to the doctor's office but because they feel that they are not sick. The large number of individuals with zero expenditure in the sample leads to a mass of zeros for the dependent variable. As the dependent variable for the corner solution outcome is restricted and hence only partially continuous, applying OLS will introduce a downward bias. While OLS is the best linear unbiased estimator for linear dependent variable according to the Gauss-Markov theorem, it fails to capture the non-linearity in the dependent variable. For a corner solution outcome, this renders inconsistent coefficient estimates (Woolridge, 2010:668).

One simple solution to the corner solution problem is to use a dichotomous-choice model, e.g. probit or logit model, to estimate whether the dependent variable is zero or positive. However, using a binary-choice model leads to an information loss because it ignores the variation in the positive range for the partially continuous dependent variable. Recognizing this issue, Tobin (1958) introduced the Tobit estimator as a solution. The Tobit model combines the discrete decision, i.e., to visit a doctor or not, with the continuous decision, i.e., how much health products to consume. More specifically, the Tobit model weights the underlying latent dependent variable by the probability that the latent variable will be above zero. The Tobit

estimator thus gives unbiased coefficient estimates for corner solution outcomes. However, an underlying assumption of the Tobit model is that the same mechanism determines the discrete and continuous decision, which is not necessarily true.

To address this problem with the Tobit model, Cragg (1971) proposed a two-part model (TPM). The underlying assumption of the TPM is that the decisions on whether to visit a doctor and how much medical service to consume are conditionally independent. Hence one estimates a probit model for the discrete choice, i.e., to buy health services or not, and an OLS model for the continuous decision, i.e. how much health services to consume. Although the TPM allows for different decision mechanisms between whether to buy and how much to consume, the conditional independence assumption of the error terms is not likely to hold (Woolridge 2010:697).

Heckman (1974, 1979) relaxed this assumption and allowed for correlation between the error terms in the discrete choice equation and continuous decision equation through an integrated two-stage model. The Heckman two-stage model can be estimated using either a limited information maximum likelihood estimator (hereafter referred to as LIML), or a full information maximum likelihood estimator (FIML). The FIML requires that we simultaneously fit the two stages using maximum likelihood estimation (MLE) and can be computationally intensive. In addition, the FIML can be highly sensitive to the normality assumption of the error terms. In order to resolve these issues, Heckman (1979) introduced LIML, which is a more robust, yet computationally simpler, two-stage estimator. The solution, termed as Heckman's correction, is to obtain an estimate of the inverse mills ratio from the first stage regression and insert it into the second stage regression as an additional independent variable. The inverse mills ratio captures the non-linear dependence between error terms. The Heckman model estimated by a LIML method is often referred to as "Heckit".

Although the four models are widely applied in empirical studies, there are relatively a few studies comparing them all and the existing literature often focuses on limited settings. For instance, Carson and Sun (2007) address the issue of a non-zero threshold. Dow and Norton (2003) focus on a selection criterion between LIML and TPM. Sigelman and Zeng (1999) examine scenarios in which data are not censored due to a data issue, but because of selection bias, i.e. individuals in the sample are self-selected to be zero, or the data are not censored at zero. Jonsson (2012) compares the performance (bias and variance) of LIML estimator for a set of model specifications. The author compares LIML with the classic Gauss-Markov (GM) model[1] and panel data models estimated by an Error Component Regression (ECR), and Random Coefficient Regression (RCR). Hay et al. (1987), Manning et al. (1987), and Leung and Yu (1996) all focus on TPM and LIML. They conclude that when there is "little variation" in the independent variables, the LIML suffers from collinearity issues and is less efficient than TPM. Flood and Gråsjö (2001) find that the TPM, with a known censoring point, is the most efficient model among Tobit, FIML, LIML and TPM. Their simulation

---

[1] The GM model has the well-known specification $y = x\beta + \varepsilon$, which thus constitutes the "baseline" for the linear part of all Tobit type models applied in the thesis.

results also indicate that simple Tobit model can perform, as good as or even better (i.e. has lower mean bias) than the more complex models FIML or LIML. The authors further examine effect of adjusting sample size from 500 to 1000 observations and the scenario of an omitted variable. The adjustments indicate no differences on the relative model performance. Although the author also reports marginal effects in addition coefficients, which marginal effects and whether these are calculated at the means is not disclosed. [23] Later evidence from Martin and Pham (2008) compare the Tobit, LIML and FIML models and indicate that a modified version of Tobit by Eaton-Tamura (1994) can be less biased than the LIML/FIML estimators, even if either of these latter is the true model.[4]

Besides the lack of a general all-encompassing study on respective performance of the four estimators, the existing studies focus on relatively large sample sizes, i.e., the asymptotical distribution of the estimators. The performance of these models for small samples is generally ignored. Almost all studies comparing the different models only focus on sample sizes ranging from 200 to 500 observations, or larger. An exception is Paarsch (1984). The author compares FIML with LIML for the samples size varying from 50 to 200 and finds that FIML outperforms LIML and introduce small bias.

Given well-known poor performance of maximum likelihood estimators (MLE) for small samples, the estimator procedure can induce bias in small samples.[5] This could be a particularly relevant issue when studying corner solution outcomes, as all the models are, at least partially, estimated by a maximum likelihood method. The results of Schoonbroodt (2004), show that MLE can induce a large biases for sample sizes below 200. Greene (2004) finds that whilst the Tobit estimator is unbiased for panel data with small sample size, probit/logit estimator can be biased. The findings can cast doubt on the relative performance of two-stage models vis-à-vis the simple Tobit model, as the two-stage models are estimated by a first stage probit estimator. The results from Jonsson (2012) indicates that the bias for the LIML estimator is negligible for a sample size above 100 and the bias diminishes almost completely as the sample size increase to and above 400. The author concludes this from using the set of specifications described above, with choice of sample sizes ranging from 100 to 1000 based on a preparatory study. To the best of my knowledge, there are few studies conducted on the importance of sample sizes, although this can have significant impact on estimator bias.

However, samples of relatively small size, i.e., 50 or 100, are not uncommon in several areas of applied research. Examples of small samples can be found in political science, psychology,

---

[2] Notably there are three types of marginal effects for Tobit models: Marginal effects on the latent variable, on the dependent variable conditional on being uncensored and the unconditional effects. See the theoretical section or Woolridge (2010:670) for a further discussion on different marginal effects and their implications.

[3] Depending on if marginal effects are computed at means (MEM) or on average (AME), the interpretation and results may vary substantially. This follows from the fact that each individual in the sample have a different marginal effect. In this thesis MEMs are the ones used and considered, as it the most commonly encountered in empirical research. For a discussion on issue and implementation in STATA, see Bartus (2005).

[4] The Tobit procedure proposed by Eaton-Tamura (1994) is a special case applied to deal with issues in trade flows, adding an extra constant to the threshold for the dependent variable.

[5] This is commonly referred to as that the MLE exhibits poor small sample properties.

and in some areas of economics and finance (Dietrich, 2001; Hart and Clark, 1999). The examples of empirical works and practical situation with small samples are abundant. Gordy and Heitfeld (2002) examine default correlation thresholds among loans in a credit rating-based framework using small samples, due to the lack of historical data, i.e. when the time dimension (T) is small. The author considers T ranging from 20 to 160. Shadish et al (2008) study the practical implications of using non-randomized experiment samples compared to randomized samples. Their sample sizes used varies from 79 to 445 individuals. . Wang and Ray (1994) examine the effect of being the initiator of interstate wars on victory chance between Great Power for the years 1496-1814, with a sample consisting of 105 such wars. By using data on 160 Norwegian companies, Svendsen and Haugland (2011) investigate the effects transaction costs, strategic importance and institutional factors on cross-border investments. Giger (2009) examine political party representation with regard to electorate preferences using 121 to 271 observations. These are just a few examples among many empirical studies relying on relative small samples. Hence, it is important to understand the performance of these estimators when they are applied to small samples.

To summarize, earlier research have produced inconclusive evidence or limited insights on how to handle corner solution outcomes (Puhani (2000)). Each one of the four estimators is designed to fit an archetypical situation, wherefore their respective specific data generating processes all differ (DGP). However, in reality there is no clear cut on which model that fits what context, as the DGP is always unknown in practice. Thereby, misspecification can easily occur. This requires that we understand under which circumstances the model is appropriate beyond theoretical plausible scenarios and the bias of model coefficients when there is risk of misspecification. Besides, the issue of small samples persists in practice and so does the need to cope with it.[6] In this thesis a Monte Carlo study is implemented to compare bias of the four estimators described above.

I emphasize the importance of small samples in this study as this is only covered to a minor extent in previous literature. My goal is not necessarily to find a breakage point, but give an indication on the overall importance of sample size. In order to assess the bias of the four models, each of them is estimated with respect to the real DGPs of the four models. I allow for different truncation rates to examine the effects of the amount of zeroes in the sample on the estimates, given that early studies have shown this affects model performance. I also allow for various correlation structures for the error terms in the two stages to assess the impact of dependence on the level of bias. The remainder of the thesis is structured as follows: Section I outlines the theoretical background. Section II presents the Monte Carlo strategy. The results are reported in section III and section IV concludes.

## I. Theoretical background

---

[6] Small sample size is not necessarily the outcome of the availability of the data, and it can be balance in the tradeoff between accuracy and efficiency. If it is possible to draw robust conclusions out of a sample 50, why should a researcher should spend time and money to obtaining a sample of size 1000? See Dietrich (2001) for a short discussion in a practical context.

*A. Corner solution models*

*A.1 Tobit estimator*

Let $y^*$ be defined as a latent variable, observed only if $y^*$ is above the certain threshold $A$ and possible to estimated by a standard Tobit model, then:

$$y^* = Xb + u \tag{1}$$

$$u \sim N(0, \sigma^2)$$

$$y = y^* \ if \ y^* > A$$

$$y = A \ if \ y^* \leq A$$

The most common functional form in empirical research, which also applies to the analysis in this thesis, is to replace $A$ with 0. Here $X$ is a vector of independent variables, $b$ denotes a vector of regression coefficients, $u$ is the normally distributed error term with zero mean and unknown variance $\sigma^2$.

The Tobit model is estimated by using MLE. The log-likelihood function for a Tobit model with a zero threshold is:

$$Log \ L = \sum_{y=0} log\left(1 - \Phi\left(\frac{Xb}{\sigma}\right)\right) + \sum_{y>0} log\left(\frac{1}{\sigma}\phi(\frac{y-Xb}{\sigma})\right) \tag{2}$$

In line with conventional notations, $\Phi$ (.) corresponds to normal CDF and $\phi(.)$ to the normal PDF respectively. The log-likelihood function is then maximized with respect to the parameters $b$ and $\sigma$.[7]

The coefficients obtained by the MLE show estimated conditional expectation of $y^*$ given the independent variable $x$, i.e. $E(y^*|x)$. This only tell about how a change in $x$ effects the latent variable $y^*$, which is not of interest for corner solutions. In order to obtain an estimate of $E(y|x)$, the marginal (or partial) effects needs to computed (Dow and Norton (2003)). For the Tobit models there are three conditional means function, with three corresponding marginal effects:

- $E(y^*|x)$ – The effects on the latent variable $y^*$, equivalent to model coefficients of a linear specification.
- $E(y|x)$ – The "unconditional" effects of $x$ on $y$, including censored and uncensored observations.
- $E(y|y > 0, x)$ – The "conditional" effects of $x$ on $y$, restricted to uncensored observations.

For a corner solution outcome, we are interested in calculating the conditional or unconditional marginal effects. The choice depends on empirical context. Applying the model

---

[7] Something can be done with ease in many statistical packages, including STATA.

to health expenditure, we are interested in those who do not purchase any health too. For consumption of alcohol on the other hand, we might be better off by limiting the scope to actual drinkers. As $E(y|x)$ reflects the whole population, including both buyers and non-buyers, it is generally assumed to give a better view of actual utility choice of interest. Accordingly it is for most empirical situations the unconditional effects that are reported Dow and Norton (2003)..

$E(y|x)$ is a decomposition of the linear expectations conditional on being observed, $E(y|x, y > A)$, and not being observed, $E(y|x, y \leq A)$. Weighted by the probability of being above the threshold, $P(y > A)$ or not being above the threshold $P(y \leq A)$:

$$E(y|x) = P(y \leq A)E(y|x, y \leq A) + E(y|x, y > A)P(y > A)$$

$E(y|x, y \leq A) = 0$ for Tobit models, hence:

$$E(y|x) = E(y|x, y > A)P(y > A)$$

Where

$$P(y^* > A|x) = P(Xb + u > A|x)$$

$$= P(u > A - Xb|x)$$

$$= P\left(\frac{u}{\sigma} > \frac{A - Xb}{\sigma}\right) = 1 - \Phi\left(\frac{A - Xb}{\sigma}\right)$$

Set $A = 0$ and use the fact that the normal distribution is symmetric around it's mean:

$$1 - \Phi\left(\frac{-Xb}{\sigma}\right) = \Phi\left(\frac{Xb}{\sigma}\right)$$

In order to derive $E(y|x, y > 0)$, consider the density of the truncated standard normal variable $y$, with truncation at point $A$ and $\alpha = \frac{A - Xb}{\sigma}$:

$$f(y|y > A) = \frac{\phi(y)}{1 - \Phi(\alpha)} = \frac{\phi\left(\frac{1}{\sigma}\left(\frac{y - Xb}{\sigma}\right)\right)}{1 - \Phi\left(\frac{A - Xb}{\sigma}\right)}$$

It follows that the first moment of $y$ then is:

$$E(y|x, y > A) = Xb + \sigma\left(\frac{\phi\left(\frac{1}{\sigma}\left(\frac{A - Xb}{\sigma}\right)\right)}{1 - \Phi\left(\frac{A - Xb}{\sigma}\right)}\right)$$

For $A = 0$, it follows:

$$E(y|x, y > 0) = Xb + \sigma\lambda\left(\frac{Xb}{\sigma}\right)$$

In which $\lambda$ denotes the inverse mills ratio (IMR), which indicates the level of truncation in the sample:

$$\lambda\left(\frac{Xb}{\sigma}\right) = \frac{\phi(\frac{-Xb}{\sigma})}{1-\Phi\left(\frac{-Xb}{\sigma}\right)} = \frac{\phi(\frac{Xb}{\sigma})}{\Phi\left(\frac{Xb}{\sigma}\right)} \tag{3}$$

$$E(y|x) = \Phi(\tfrac{Xb}{\sigma})\,[Xb + \sigma\lambda\left(\tfrac{Xb}{\sigma}\right)]$$

$$\qquad = \Phi\left(\tfrac{Xb}{\sigma}\right)Xb + \sigma\phi(\tfrac{Xb}{\sigma}) \tag{4}$$

The unconditional marginal effects are then computed as the partial derivative of variable $x_j$:

$$\frac{\vartheta E(y|x)}{\vartheta x_j} = \frac{\vartheta P(y>0)}{\vartheta x_j}E(y|x, y>0) + \frac{\vartheta E(y|x,y>0)}{\vartheta x_j}P(y>0)$$

As the derivative of the CDF is the PDF, we have the following expression:

$$\frac{\vartheta E(y|x)}{\vartheta x_j} = \frac{b_j}{\sigma}\phi\left(\tfrac{Xb}{\sigma}\right)\left[Xb + \sigma\lambda(\tfrac{Xb}{\sigma})\right] + \Phi\left(\tfrac{Xb}{\sigma}\right)[b_j + b_j(-\lambda(\tfrac{Xb}{\sigma})[\tfrac{Xb}{\sigma} + \lambda(\tfrac{Xb}{\sigma})])$$

$$\qquad = (\tfrac{b_j}{\sigma})\phi\left(\tfrac{Xb}{\sigma}\right)\left[Xb + \sigma\lambda(\tfrac{Xb}{\sigma})\right] + \Phi\left(\tfrac{Xb}{\sigma}\right)(b_j(1 - \lambda\left(\tfrac{Xb}{\sigma}\right)\left[\tfrac{Xb}{\sigma} + \lambda\left(\tfrac{Xb}{\sigma}\right)\right])) \tag{5}$$

Alternatively we could use a log-specification, which is common in the economic literature. Unconditional effects can then be interpreted as elasticity or semi-elasticity, depending on whether independent variables are in logs or not. For log-specification we thus estimate a model with the following functional form:

$$y^* = e^{(Xb+u)} \tag{6}$$

The log-specification retains the properties of the Tobit model described above. As the probability of being observed or not is not affected by logging the dependent variable, we simply use the first moment of the truncated lognormal distribution for obtaining $E(y|x, y \geq 0)$ and let $P(y^* > 0)$ remain as it is:

$$E(y|x) = e^{\left(Xb+\frac{\sigma^2}{2}\right)}\Phi\left(\frac{Xb+\sigma^2}{\sigma}\right) \tag{7}$$

The marginal effects are then the derivative:

$$\frac{\vartheta E(y|x)}{\vartheta x_j} = b_j e^{(Xb+\frac{\sigma^2}{2})}\phi\left(\frac{Xb+\sigma^2}{\sigma}\right) + b_j e^{(Xb+\frac{\sigma^2}{2})}\Phi(\frac{Xb+\sigma^2}{\sigma}) \tag{8}$$

*A.2 Two-part model (TPM)*

Now consider the TPM presented by Cragg (1971), allowing for different selection and outcome equations. Let's define a binary variable, $w$, which indicates whether a dependent variable is directly observed or not. The outcome equation is equal to Tobit, hence the TPM is:

$$w = XY + v \tag{9}$$

$$v \sim N(0,1)$$

$$y^* = Xb + u$$

$$Corr(v, u) = 0$$

$$y = wy^* = 1[X\Upsilon + v > 0](Xb + u) \tag{10}$$

Here $\Upsilon$ corresponds to the coefficient vector and $v$ the error term from the first stage probit model. The second is the same as for the Tobit model. The two error terms $u$ and $v$ are assumed to be bivariate normally distributed and uncorrelated.

Like in the Tobit case, the estimation of the TPM model involves specifying the log-likelihood function to be maximized:

$$Log\ L = \sum_{y=0} log\left(1 - \Phi\left(\frac{X\Upsilon}{\sigma}\right)\right) + \sum_{y>0}\left(log\Phi\left(\frac{X\Upsilon}{\sigma}\right) - log\Phi\left(\frac{Xb}{\sigma}\right) + log\phi\left(\frac{y-Xb}{\sigma}\right) - log(\frac{\sigma^2}{2})\right) \tag{11}$$

The first part of the log-likelihood function does not differ much from the Tobit model, with exception that by probit estimated coefficients now are used separately. The second part of the MLE looks a little bit more complicated, but corresponds to a truncated linear regression model. One can thereby use a probit for the first stage and use least squares for estimating the second, linear, part.[8]

As before there are three types of marginal effects that can be calculated. The unconditional expectations have the same functional form as Tobit:

$$E(y|x) = E(y|x, y > 0)P(y > 0)$$

By introducing $w$, the probability of being observed is estimated independently of the linear expectations. The conditional expectations thus differ slightly from Tobit:

$$E(y|x) = \Phi(X\Upsilon)[\left(Xb + \sigma\lambda\left(\frac{Xb}{\sigma}\right)\right)]$$

Marginal effects are given by (12):

$$\frac{\vartheta E(y|x)}{\vartheta x_j} = \Upsilon_j\phi(X\Upsilon)[\left(Xb + \sigma\lambda\left(\frac{Xb}{\sigma}\right)\right)] + b_j[1 - \lambda\left(\frac{Xb}{\sigma}\right)\left(Xb + \sigma\lambda\left(\frac{Xb}{\sigma}\right)\right)]\Phi(X\Upsilon) \tag{12}$$

The letter $\Upsilon_j$ corresponds to the coefficient of interest from the probit stage for variable, $j$. One can spot that the TPM is actually a nested Tobit, which reduces to latter if the coefficient estimates from the two stages are equal (Woolridge 2010: 697).

Alternatively we may wish to use a log-specification, for which the dependent variable of interest $y$ is given by:

$$y = 1[X\Upsilon + v > 0]e^{(Xb+u)} \tag{13}$$

Calculating expectations is quite straightforward as $P(y > 0)$ is the same in the truncated case:

---

[8] Alternatively in STATA one can use the "craggit" command by Burke (2009).

$$E(y|x, y > 0) = E(y^*|x, w = 1) = E(y^*|x) = e^{(Xb + \frac{\sigma^2}{2})}$$

$$E(y|x) = \Phi(XY) \times e^{(Xb + \frac{\sigma^2}{2})} \tag{14}$$

$$\frac{\vartheta E(y|x)}{\vartheta x_j} = Y_j e^{(Xb + \frac{\sigma^2}{2})} \phi(XY) + b_j e^{(Xb + \frac{\sigma^2}{2})} \Phi(XY) \tag{15}$$

*A.2 Heckman model: Full-information maximum likelihood (FIML) and Limited information maximum likelihood (LIML)*

The assumption of $Corr(v, u) = 0$ can be problematic, as the choice between "if to consume" and "how much to consume" is likely to be correlated. The violation of the zero-correlation assumption of the TPM renders the marginal effects estimates biased. Heckman (1974) presented a solution using a full maximum likelihood estimator, FIML, to account for the correlated error terms. The FIML estimator retains the basic TPM specification, but allows the two error terms to be correlated with parameter, $\rho$:

$$Corr(v, u) = \rho$$

The log-likelihood function is:

$$Log\, L = \sum_{y=0} log\left(1 - \Phi\left(\frac{XY}{\sigma}\right)\right) + \sum_{y>0}(log\, \phi\left(\frac{(y - Xb)}{\sigma}\right) + log\left(\Phi\left(\frac{XY + \frac{\rho}{\sigma}(y - Xb)}{(1 - \rho^2)}\right)\right) - log\Phi\left(\frac{Xb}{\sigma}\right) -$$

$$log(\frac{\sigma^2}{2})) \tag{16}$$

The log-likelihood function is maximized by simultaneously fitting the two parts, i.e. using *full information*. The method is however computationally intensive and might not even converge, especially when there is little information in the sample – i.e. small sample size or high truncation. Due to the computational issues, Heckman (1979) proposed a simpler alternative using a limited information maximum likelihood, LIML. The method is commonly referred to as Heckman's correction or "Heckit" in econometric literature. The procedure consists of two steps:

1. Estimate the selection equation by probit to obtain the linear prediction, $XY$. Compute $\lambda$ by (4) using the linear prediction from the probit, $XY$.

2. Include the $\lambda$ as an extra regressor in the outcome equation and estimate it by OLS. The non-linearity stemming from correlation between the two errors is then captured by $\lambda$.

Whether we use FIML or LIML makes no difference for how the unconditional expectations are computed. The calculation for marginal effects remains essentially the same as before, but we do adjust the expectation for the correlation parameter $\rho$:

$$E(y^*|x, w = 1) = E(y^*|v > -XY)$$

$$= E(Xb + u|v > -XY)$$

$$= Xb + E(u|v > -X\Upsilon) = Xb + \eta\lambda\left(\frac{X\Upsilon}{\sigma}\right)$$

Whereas $\eta = \sigma\rho$, the unconditional expectations for a LIML/FIML model is:

$$E(y|x) = \Phi(X\Upsilon)[Xb + \eta\lambda\left(\frac{X\Upsilon}{\sigma}\right)] \tag{17}$$

The marginal effects are:

$$\frac{\vartheta E(y|x)}{\vartheta x_j} = \Upsilon_j\phi(X\Upsilon)[Xb + \eta\lambda\left(\frac{X\Upsilon}{\sigma}\right)] + [b_j - \Upsilon_j\eta\lambda\left(\frac{X\Upsilon}{\sigma}\right)\left(X\Upsilon + \sigma\lambda\left(\frac{X\Upsilon}{\sigma}\right)\right)]\Phi(X\Upsilon) \tag{18}$$

The unconditional expectations can also be expressed for logged dependent variable, with respective specification and marginal effects of variable $x$:

$$E(y|x) = \Phi(X\Upsilon + \eta) \times e^{(Xb + \frac{\sigma^2}{2})} \tag{19}$$

$$\frac{\vartheta E(y|x)}{\vartheta x_j} = \Upsilon_j e^{(Xb + \frac{\sigma^2}{2})}\phi(X\Upsilon + \eta) + b_j e^{(Xb + \frac{\sigma^2}{2})}\Phi(X\Upsilon + \eta) \tag{20}$$

An added advantage of LIML over the FIML estimating procedure is that the former is robust to deviations from the normality assumption of the error terms (Puhani (2000)). As shown by Jonsson (2008) however, the dependent variable is still required to be normally distributed to obtain a correct estimate of $\lambda$. Another known issue with the LIML is the estimation of $\rho$. The correlation parameter is not necessarily bounded between -1 and 1, affecting the conditional expectations.[9] Moreover the LIML generally requires an "exclusion restriction" to suffice: an extra variable included in the selection but not in the outcome equation (Woolridge (2010:697-698)). Otherwise the LIML suffers from a collinearity issue as $\lambda$ estimated is from a set of variables and then included among them. The theoretical studies examining the importance of an exclusion restriction, e.g. Leung and Yu (1996) or Dow and Norton (2003), find it to be no more than a minor nuisance in large samples (n>1000).

Conclusively there a several models applicable to corner solutions, all designed to cope with certain different problems experienced in empirical research. The subsequent section outlines the results obtained by earlier simulation studies on the relative performance of these models in terms of bias.

---

[9] For small samples in particular, absurdly large (or low) estimates are not uncommon for $\rho$.

*B. Earlier simulation studies*

Earlier studies have failed to present any present consensus regarding the relative model performance. Previously however, following simulations by Duan et al (1984), Hay et al (1987) and Manning et al (1987), TPM was considered the superior choice over LIML. The validity of these results was later challenged by Leung and Yu (1996). The authors showed that earlier conclusions were based on an erroneous use of insufficient variation in the independent variables, which induced a collinearity issue for LIML. After adjusting for the little variation using explanatory variables distributed over a larger range, the TPM was no least biased. Dow and Norton (2003) later reaffirmed these claims and concluded on the importance of using sufficient variation in the independent variables.

Papers examining non-normal errors are concerned with three alternatives, logistically, Cauchy or Laplace distributed errors. For logistically distributed error terms, TPM tends to be preferable over FIML/LIML. If the error terms are Cauchy all the models produce highly biased coefficient estimates. This is particularly augmented for the FIML estimator (Hay et al (1987)). In the case of Laplace distributed errors though, Paarsch (1984) concludes that FIML is less biased than LIML. An additional setback is that authors evaluating deviations from the normality assumption did not run more than 100 replications. With so few simulations one might also question the robustness of the findings.

There are inconclusive evidence of the implications of $\rho$ and level of truncation, besides a few main points: LIML is adversely affected by high correlation between error terms as well as high truncation rate. Neither truncation rate nor correlations has any significant impact on either FIML or TPM. The latter results are surprising as an underlying assumption of TPM is zero correlation between error terms. However, as shown by Duan et al (1984) the TPM is in practice generally unaffected by a violation of the conditional independence assumption.

There are few studies comparing the effect of estimating in levels or logs, despite evidence of sharp differences in bias from Martin and Pham (2008). Their results indicate that FIML/LIML models are comparatively more biased for a log-specification. The Tobit model however, is relatively less biased when estimated in logs instead of in levels. Dow and Norton (2003) evaluate both log and level specification for TPM and the Heckman models, finding little difference. Tsu and Liu (2008) assess the relative bias of the truncated TPM (10) with the logged TPM (13), concluding that (10) is less biased.

Despite the plethora of simulation studies, none evaluates model bias with respect to the three different situations the models were designed for. Thereby leaving a clear gap in understanding of model overall performance with respect to the DGPs. The strategy to narrow this gap is presented in subsequent section.

## II. Monte Carlo strategy

*A. Outline of background*

The models were compared using Monte Carlo simulations in order to evaluate performance under different settings. The method involves generating a large number of independent

samples, in which the models are estimated and compared to the true values. In accordance with the law of large numbers, the average bias in the samples will tend to the true model bias.

All models were examined with respect to the three DGPs, for sample size of $n = 50, 75, 100, 500$. The $n$ should be sufficiently low to cover the smallest sample size considerable in empirical research, for which 50 is the chosen bottom line. For the purpose of comparison and relative importance of sample size, $n$ is then increased up to a moderate sample size of 500 observations.

The models are estimated with respect to $y^*$, which renders coefficients that are of little interest for corner solutions. Instead of using raw coefficients, the marginal effects were calculated according to the formulas given in section I. As the marginal effects differ for each observation in the sample, one can either calculate average marginal effects or marginal effects at the means. Whilst both have respective advantages, the latter is the common choice in literature and interpretable as OLS coefficients[10]. Therefor marginal effects were calculated at the means.

Models were compared using average relative bias, which is calculated by:

$$Bias = \frac{\overline{ME_t} - \overline{ME_e}}{\overline{ME_t}} \qquad (21)$$

In which $\overline{ME_t}$ is the true average marginal effects and $\overline{ME_e}$ is the estimated average marginal effects from the sample in question, the bias can then be expressed as relative deviation from the true value in percent.

*B. Model Specifications*

The Tobit DGP was specified, using equation (1) and (6) in order to generate the data. The x-vector corresponded to three independent variables, $x_1, x_2, x_3$, all distributed $\sim U(0,10)$. Leung and Yu (1996) as well as Dow and Norton (2003) concluded that one need to include "sufficient" variation in $x$ for LIML to be an appropriate choice. The authors concluded that $U(0,10)$ is an adequate choice, hence it is used here as well. The slope parameters are all set to 1: $b_1 = b_2 = b_3 = 1$. The constant is retained for controlling the number of zeros in the sample, using a truncation rate of 25, 50 and 75 %. The standard deviation of the error term was set to $\sigma = 5$. The $\sigma$ is set higher than in most previous studies, in order to increase the level of variation in $u$. An exception is Stolzenberg and Relles (1990), which tested several different values of standard deviation, without getting any clear results on the implications. For the TPM/Heckman estimations respectively, the same set of regressors were used in the structural and index equation. The models' performances were evaluated using the average relative bias, calculated by (21).

The TPM was generated by (1), (9), (10) and (13). In contrast to Tobit, the index and structural equations do now differ. The second stage coefficients were set to equal values as

---

[10] See for instance Bartús (2005) for an elaborate discussion on the difference.

for Tobit DGP, the first stage coefficients were set accordingly: $\Upsilon_1 = 0.6, \Upsilon_2 = 0.7, \Upsilon_3 = 0.8$, reflecting differences between the respective mechanisms. The error term of the first stage is standard normal $v \sim (0,1)$, the second stage error term is as before $u \sim N(0, \sigma^2)$, with $\sigma = 5$. The intercepts are used to determine the expected number of zeros in the sample as for Tobit. Although an assumption of the TPM is that $\rho = 0$, this is relaxed to the test the validity in Duan et al (1984) conclusion that the TPM is unaffected by correlation. The correlation coefficient were set to range of strong positive to negative dependence: $\rho = -0.9, -0.5, 0, 0.5, 0.9$. When earlier papers, as Floodén and Gråsjö (2001) and Dow and Norton (2003), only assumed that $\rho = 0.5$ or $\rho = -0.5$ they left out importance of alternative correlation structures. Studies concerning specifically correlation structure are in turn generally limited by other specifications. Most researchers also restrict the values of $\rho$ to be bounded above zero, as it is empirically unlikely with negative correlation between the error terms (Puhani (2000)). However, as it still might occur in practice (e.g. Aristei and Peroni (2009)) and is therefore included in the thesis.

The FIML/LIML DGP was generated by equations (1), (9), (10) and (13). The setting was almost identical to that of the TPM. The difference is the inclusion of an exclusion restriction, as required by LIML. Consequently $x_3$ was dropped from the structural equation. Otherwise all coefficients retained the values specified above for the TPM DGP, including $\rho$ and $\sigma$.

Lastly deviations from the normality assumption were assessed. By using the inverse CDF of the Cauchy and Laplace distributions an initially normally distributed error term was transformed into the respective distribution of choice. These simulations are restricted to the Tobit DGP in logs to make it apprehensible. The choice follows from the fact that under this setting, all models performed well with only small deviations from the true value. Secondly, the other DGPs require a specified correlation structure between the error terms. The latter specification is problematic under this set-up for Cauchy or Laplace as this requires a non-linear transformation, for which Pearsson correlation measure is inappropriate (Embrechts et al 2003).[11] The parameters for the error term were set in similar fashion as before with location parameter 0 and scale parameter of 5. Previous research assessing distributional assumptions have returned inconclusive evidence, albeit Cauchy distributed errors tend to induce a bias for all models tested (Puhani (2000)).

An overview of the specifications used for the scenarios are presented below in table I:

**Table I. Overview of result tables and Monte Carlo specification used**

| Table: | DGP: | $b_1, b_2, b_3$ | $\Upsilon_1, \Upsilon_2, \Upsilon_3$ | $\rho$ | $x_1, x_2, x_3$ | $v$ | $u$ | $\sigma$ |
|---|---|---|---|---|---|---|---|---|
| II | Tobit | 1 | $N/A$ | $N/A$ | U(0,10) | $N/A$ | $N(0,\sigma^2)$ | 5 |
| III | Tobit(logs) | 1 | $N/A$ | $N/A$ | U(0,10) | $N/A$ | $N(0,\sigma^2)$ | 5 |
| IV | TPM | 1 | 0,6,0.7,0.8 | -0.9,0,0.9 | U(0,10) | $N(0,1)$ | $N(0,\sigma^2)$ | 5 |
| V | TPM (logs) | 1 | 0,6,0.7,0.8 | -0.9,0,0.9 | U(0,10) | $N(0,1)$ | $N(0,\sigma^2)$ | 5 |
| VI | Heckman | 1 | 0,6,0.7,0.9 | -0.9,0,0.9 | U(0,10) | $N(0,1)$ | $N(0,\sigma^2)$ | 5 |
| VII | Heckman (logs) | 1 | 0,6,0.7,0.9 | -0.9,0,0.9 | U(0,10) | $N(0,1)$ | $N(0,\sigma^2)$ | 5 |
| VIII | Tobit(logs) | 1 | $N/A$ | $N/A$ | U(0,10) | $N/A$ | $L(0,\sigma^2)$ | 5 |
| IX | Tobit(logs) | 1 | $N/A$ | $N/A$ | U(0,10) | $N/A$ | $C(0,\sigma^2)$ | 5 |

---

[11] Linear correlation is only retained through linear transformation, for non-linear transformations, as these ones, only "rank correlation" persist. Examples of measures of such are Kendall's Tau or Spearman's Rho, which is differs from commonly used Pearsson coefficient.

# III. Simulation results

The results presented are slimmed down to cut out the redundancies, i.e. when there were small variations between the specifications tested. There was little extra information obtained by setting $\rho=+/-0.5$, as for the intermediate sample sizes of 75/200 observations or using a 50 % truncation rate.[12] In the result section the relative bias of average marginal effects of the predictions are reported, along with standard errors. Following Dow and Norton (2003) the standard errors are estimated as the standard deviations of the replications of each statistic.

## A. *Level specification*

The tables are presented in the order the models were outlined above. The calculated bias corresponds to the mean of 500 replication ran for each of the different settings specified. The MLE estimators may experience problems of converging for small samples, resulting in implausible coefficient estimates and thus huge bias. Samples with extreme bias values were considered non-representable outliers and these were excluded.[13] The results from simulations are reported below, in table II, III and IV

**Table II. Relative bias of marginal effects in levels in %  DGP: Tobit level-specification**

| Sample: | | N=50 | | | | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Truncation: | Correlation: | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM |
| 25 % | *N/A* | 5.5 | 80.8 | 82.2 | 71.7 | 5.6 | 47.4 | 50.5 | 69.4 | 5.4 | 50.4 | 32 | 75.6 |
| | | (2.3) | (83.2) | (54.6) | (60.1) | (2.8) | (82.1) | (47.3) | (45.2) | (1.9) | (20.1) | (37.8) | (20.3) |
| 75 % | *N/A* | -7.2 | 82.5 | 92.9 | 70.3 | -13.4 | 39.2 | 53.6 | 43.6 | -13.3 | -47.2 | -35.6 | 69.4 |
| | | (8.3) | (410.5) | (146) | (240) | (7.3) | (130.1) | (116.3) | (121.2) | (3.5) | (120.2) | (104.3) | (103) |

Bias calculated as average over 500 replications. Standard errors are given within brackets.

**Table III. Relative bias of marginal effects in levels in %  DGP: TPM level-specification**

| Sample: | | N=50 | | | | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Truncation: | Correlation: | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM |
| | $\rho =-0.9$ | 2.7 | -25.8 | -26.7 | 18.5 | -1.4 | -0.1 | -0.3 | 2.2 | 0.9 | -3.5 | -3.5 | -0.1 |
| | | (15) | (32) | (28) | (20.2) | (11.3) | (25.8) | (24.2) | (13.2) | (4.9) | (10.6) | (10.7) | (7.1) |
| 25 % | $\rho =0$ | -20.2 | -19.8 | -13.2 | 21.6 | -1.6 | -5 | -5 | -0.9 | 4.4 | 0.9 | 0.4 | 2.2 |
| | | (19) | (38.3) | (39.3) | (28.3) | (14.8) | (24.9) | (25.3) | (18.2) | (5.8) | (11.3) | (9.6) | (8.3) |
| | $\rho =0.9$ | 10.5 | -23.6 | -5.3 | 47.2 | 0.9 | -1.2 | 1.1 | -0.8 | 0.9 | 4.7 | 0 | 3.4 |
| | | (19) | (31.1) | (29.3) | (29.8) | (15.1) | (20.3) | (22.8) | (18.6) | (6.1) | (13.4) | (9.1) | (10.3) |
| | $\rho =-0.9$ | -9.4 | -29.6 | -38.2 | -5.9 | 16.3 | 6.5 | 3.9 | -1 | 10.3 | 2.4 | -1.6 | 0.5 |
| | | (20) | (30) | (45) | (22) | (11.2) | (24.4) | (25) | (20.3) | (5.3) | (11.6) | (9.2) | (6.7) |
| 75 % | $\rho =0$ | 0.6 | -32.6 | -28 | 17.5 | 7.1 | -0.1 | 0.9 | -0.9 | 10.6 | 0 | -1.9 | -2.5 |
| | | (23) | (44) | (47.4) | (36) | (13.4) | (33.4) | (30.9) | (24.9) | (5.8) | (16.2) | (13.2) | (11.1) |
| | $\rho =0.9$ | 8.2 | -24.6 | -37 | 47.6 | -7.3 | -0.7 | 1 | -1.6 | 8.3 | 1.4 | 2.2 | 1.2 |
| | | (29) | (41.2) | (40.1) | (39.2) | (12.5) | (23.4) | (23.7) | (22.9) | (5.9) | (10.2) | (10.9) | (11.1) |

Bias calculated as average over 500 replications. Standard errors are given within brackets.

---

[12] Any results not reported can easily be obtained by the source code in STATA, available upon request from author. Besides, empirical studies rarely report the value of $\rho$, making it difficult to hypothesise concerning the general plausibility of different values. Papers reporting $\rho$ commonly find an insignificant value or close to zero, suggesting that correlation between error terms is only a minor concern in practice.

[13] It will be the case in which the MLE fails to converge by a long-shot, the model renders extreme coefficient estimates and is totally unusable. No sensible researcher could trust results as such in an empirical study. Neither would it make any sense for the purpose of this thesis to include such values.

**Table IV. Relative bias of marginal effects in levels in %  DGP: Heckman level-specification**

| Sample: | | N=50 | | | | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Truncation: | Correlation: | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM |
| | $\rho =-0.9$ | -45.4 | 2.7 | 3.2 | 7.5 | -54.1 | 5.1 | 0.5 | -22.9 | -44.6 | 12.2 | 4.1 | 2.8 |
| | | (5.4) | (18.2) | (13.7) | (35.2) | (3.1) | (12.4) | (12.4) | (37.1) | (3.3) | (17.1) | (16.3) | (17.1) |
| 25 % | $\rho =0$ | -31 | 13.2 | 17.4 | 48.8 | -29.4 | 12.3 | 13.5 | -35.3 | -33.2 | 16.2 | 16.9 | 41.5 |
| | | (9.2) | (30.6) | (37.2) | (38.1) | (7.4) | (22.1) | (15.3) | (55.7) | (3) | (10.2) | (7.8) | (20.2) |
| | $\rho =0.9$ | -22.4 | -22.8 | 33.8 | -86.2 | -17.2 | 16.3 | 16.5 | 45.3 | -17.5 | 16.1 | 16.4 | 50.5 |
| | | (9.2) | (30.6) | (19.5) | (59.3) | (7.1) | (14.5) | (15.6) | (61.7) | (3.4) | (4.4) | (5.1) | (24.3) |
| | $\rho =-0.9$ | -41.2 | 81.2 | 62.2 | 10.6 | -65.7 | 58.3 | 54.5 | -67.1 | -68.1 | 67.8 | 70.2 | -72.3 |
| | | (115.8) | (332.3) | (310.1) | (270.2) | (6.3) | (70.1) | (80.3) | (4.2) | (3.2) | (41.2) | (42.7) | (3.3) |
| 75 % | $\rho =0$ | -8.3 | 122.2 | 127.3 | 132.2 | -60.7 | 55.3 | 63.1 | -70.1 | -55.3 | 55.3 | 63.1 | -53.2 |
| | | (87.3) | (229.3) | (231.4) | (274.6) | (7.7) | (74.6) | (74.1) | (7.3) | (5.2) | (31.2) | (50.3) | (5.8) |
| | $\rho =0.9$ | -45.4 | 33.6 | 37.6 | 46.8 | -49.2 | 50.3 | 52.1 | -47.1 | -48.1 | 42.3 | 16.1 | -52.4 |
| | | (22.3) | (79.4) | (85.2) | (150.3) | (10.2) | (72.1) | (70.5) | (9.6) | (4.9) | (34.5) | (28.2) | (6.1) |

Bias calculated as average over 500 replications. Standard errors are given within brackets.

The tables above provide some interesting insights about the model performance. First, sample size is important for bias, invariant to the DGP specified. The exception is Tobit, whereas the results indicate that the model works well regardless of sample size issues. For all other models sample size have strong effect on model bias. As illustrated in Table III, with TPM as DGP used. At a sample size of 100, the bias practically diminishes for all models in Table III. The standard errors remain high though, until the sample size is 500. In a clear majority of cases the TPM render least biased predictions, although Tobit standard errors always are the lowest for all scenarios. Interestingly the bias of the Tobit coefficient increases with sample size, following lower standard errors and higher level of certainty regarding bias. A similar pattern is evident all throughout table IV for the Heckman and TPM as well. As predicted a shift from smaller samples to larger have major impact on the relative bias. The contrast is the arguably much larger between 50 and 100, than going from 100 to 500. Thus the results indicate that a sample size of 100 should be sufficient to obtain relatively unbiased estimates and further improvements are less pronounced. One should take notice that the overall decrease in bias from a larger sample is smaller for the less truncated case. If there are few censored observations in the sample, there is also more information and all models are to a lesser degree affected by sample size.

Considering impact of DGP, all models tends to perform the best at their "own", as expected. A notable observation is that Tobit can perform as good as or even better than TPM for the TPM DGP. Whilst the TPM is overall the least biased in Table III, the well-known overall superiority found in previous literature is absent (Dow and Norton (2003)). In Table II the TPM is evidently even more biased than FIML/LIML and the most biased in Table IV. For Tobit DGP, the Tobit model is clearly preferable over alternatives. One reasonable explanation of this, lies within the perceived sensitivity of Probit to the standard normally distributed error term as found by Greene (2004). On the contrary in table IV the Tobit is the most biased of all models. Arguably the exclusion restriction renders a specification error for Tobit, leading to a large underestimation in expected marginal effects. Strangely though, similar results are observed for the TPM estimator under this set-up. Suggestively the reason for the sharp downward bias could be another than simply the exclusion restriction, which should only affect Tobit.

There is no evidence that the TPM provides better estimates for $\rho = 0$. The importance of uncorrelated error terms, as stressed in econometric literature (Woolridge (2010:691)), is not supported by the results. The TPM is shown to be relatively unaffected by correlated error terms. The results thus reaffirm earlier findings on the insignificance of this assumption. If there is any impact of correlation on the TPM, it is that the model is less biased for strong negative correlation, $\rho = -0.9$. Scenarios with $\rho = 0.9$ are on the other hand slightly more biased for the TPM. LIML/FIML both are largely unaffected by correlation structure, although a little less biased for $\rho = 0.9$. An unforeseen results as if there is a non-linear relation between the structural and index equation, the TPM should be biased. The findings do not support the results of Manning et al (1987) or Flooden and Gråsjö (2001), on the overall supremacy of TPM over FIML/LIML. An explanation can be as suggested Leung and Yu (1996) and Dow and Norton (2003), that LIML depend crucially on variation in the independent variables to avoid a collinearity issue. The Tobit model fares generally the best for $\rho = 0.9$ and worse for $\rho = -0.9$. Recalling that the Tobit model is a special case of TPM, in which the Probit and linear coefficients are equal, a plausible explanation is that Tobit relies on a high degree of similarity between the equations. Hence for scenarios with high level of correlation between the error terms less bias should be observed for Tobit, which also is the case.

Truncation rate is an important source for bias in all models, regardless of DGP considered. A higher level of zeros in the sample leads to an information loss and thus reduced accuracy of the models. The small diversity in highly truncated samples further leads to estimation problems, comparable to those of the small samples. The link between number of zeros and estimator performance is however unclear for all models except Tobit, which is always adversely affected by an increase in truncation. Otherwise for TPM DGP as shown in Table III there is little or no effect on the other models. In Table II/IV on the contrary there are sharp differences for FIML/LIML/TPM between the both truncation levels. The standard errors are particularly larger for when there are more zeros in the sample. Thereby one can conclude that the importance of truncation is related to the underlying DGP, which in practice always is unknown.

*B. Log specification*

Proceeding with the same logged specification, results for the various scenarios are displayed in table V-VII:

**Table V. Relative bias of marginal effects in levels in %  DGP: Tobit**

| Sample: | | N=50 | | | | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Truncation: | Correlation: | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM |
| 25 % | N/A | 5.6 | 4.8 | 4.3 | 6.8 | 5.7 | 4.6 | 7.8 | 5.7 | 6.2 | 4.8 | 4.7 | 6.9 |
| | | (2.7) | (13.7) | (9.3) | (5.2) | (1.6) | (6) | (6.2) | (4.5) | (0.7) | (3.5) | (2.9) | (4.1) |
| 75 % | N/A | -3.7 | -6.7 | 4.5 | -9.1 | -2.4 | -5.8 | -2.9 | -9.5 | -1.4 | -9.8 | -6.8 | -7.3 |
| | | (11.7) | (63.2) | (17.3) | (15.2) | (6.5) | (48.32) | (13.1) | (10.2) | (3.1) | (11.3) | (6.5) | (7.1) |

Bias calculated as average over 500 simulations. Standard errors are given within brackets.

**Table VI. Relative bias of marginal effects in levels in %  DGP: TPM log-specification**

| Sample: | | N=50 | | | | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Truncation: | Correlation: | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM |
| | $\rho$ =-0.9 | 3.8 | -1.5 | -3.9 | -4.7 | 2.8 | -2.5 | -0.8 | -2.8 | 3.2 | 2.2 | 5 | 0.8 |
| | | (2.9) | (7.7) | (8.7) | (4.8) | (2.8) | (6.3) | (6.5) | (3.2) | (1) | (2.6) | (2.2) | (1.5) |
| 25 % | $\rho$ =0 | 4.1 | -31.1 | -7.4 | -5.8 | 3.5 | -2.5 | 0.6 | 0.1 | 4.3 | -2.1 | 2.5 | 2.1 |
| | | (3,7) | (7,2) | (7,1) | (5,2) | (3,2) | (6,8) | (6,2) | (2,5) | (2,2) | (4,2) | (4) | (1,2) |
| | $\rho$ =0.9 | 4 | -31.5 | -9.1 | -6.6 | 3.7 | -3.4 | -0.2 | 2.5 | 5.7 | -2.7 | 0.8 | 3.8 |
| | | (2.6) | (6.3) | (5.8) | (4.9) | (2.2) | (5.8) | (5.6) | (4.3) | (2) | (4.1) | (4.3) | (1.3) |
| | $\rho$ =-0.9 | -3.8 | 9.9 | -15.1 | -10.3 | -7 | -6.7 | -10.2 | -12.8 | -5 | -4.6 | -2.8 | -9.4 |
| | | (6.9) | (17.7) | (9.7) | (7.3) | (3.7) | (7.3) | (6.3) | (5.2) | (1.9) | (4.2) | (4) | (2.3) |
| 75 % | $\rho$ =0 | 4 | -3.7 | -14 | -8.7 | 2.4 | -3.7 | -2.5 | -2.7 | 3.2 | -1.9 | 0.4 | -0.3 |
| | | (6.7) | (31.2) | (13) | (7.3) | (6) | (17.2) | (12) | (4.5) | (2) | (4.5) | (4.3) | (2) |
| | $\rho$ =0.9 | 5.4 | 35.5 | -12.8 | -10.6 | 5 | -7.1 | -3.9 | 1.4 | 5.4 | -6 | -2.8 | 3.9 |
| | | (6.4) | (12.2) | (8.2) | (7.7) | (4.9) | (7.9) | (7.84) | (7.21) | (1.9) | (3) | (3.2) | (2.5) |

Bias calculated as average over 500 simulations. Standard errors are given within brackets.

**Table VII. Relative bias of marginal effects in levels in %  DGP: Heckman log-specification**

| Sample: | | N=50 | | | | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Truncation: | Correlation: | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM |
| | $\rho$ =-0.9 | -14.4 | 24.4 | 20.0 | -11.7 | -20.5 | 17.6 | 15.3 | -19.8 | -12.6 | 9.2 | 15.6 | -13.5 |
| | | (5.7) | (46.3) | (26.4) | (7.8) | (3.3) | (17.1) | (19.3) | (4.8) | (3.3) | (17.1) | (16.3) | (4.1) |
| 25 % | $\rho$ =0 | -15.9 | 13.4 | 15.9 | -13.7 | -12.9 | 3.5 | 8.6 | -8.3 | -13.2 | 4.8 | 9.6 | -9.5 |
| | | (4.3) | (23.3) | (26.8) | (6.2) | (2.6) | (14.8) | (13.2) | (4.2) | (1.3) | (8.1) | (7.8) | (2.1) |
| | $\rho$ =0.9 | -9.9 | 3.1 | 6.6 | -5.4 | -8 | 2.1 | 4.2 | -3.4 | -9.2 | 3.3 | 2.6 | -4.5 |
| | | (4.3) | (22.2) | (15.5) | (6.2) | (2.3) | (12.4) | (9.3) | (3.9) | (0.9) | (4.4) | (4.8) | (1.5) |
| | $\rho$ =-0.9 | -48.9 | 33.5 | 25.6 | -45.5 | -65.7 | 58.3 | 54.5 | -67.1 | -58.1 | 47.8 | 40.2 | -52.3 |
| | | (6.3) | (81.4) | (87.3) | (9.3) | (6.3) | (70.1) | (80.3) | (4.2) | (3.2) | (41.2) | (42.7) | (3.3) |
| 75 % | $\rho$ =0 | -52.3 | 35.5 | 30.6 | -51.5 | -60.7 | 55.3 | 63.1 | -70.1 | -55.3 | 55.3 | 63.1 | -53.2 |
| | | (7.1) | (94.4) | (60.5) | (8.9) | (7.7) | (74.6) | (74.1) | (7.3) | (5.2) | (31.2) | (50.3) | (5.8) |
| | $\rho$ =0.9 | -49.5 | 61.4 | 77.5 | -57.4 | -49.2 | 50.3 | 52.1 | -47.1 | -48.1 | 42.3 | 16.1 | -52.4 |
| | | (7.1) | (94.4) | (60.5) | (8.9) | (8.3) | (68.4) | (60.3) | (9.4) | (5.2) | (22.1) | (30.2) | (5.8) |

Bias calculated as average over 500 simulations. Standard errors are given within brackets.

Previously pronounced issues related to the other models using Tobit DGP with level-specification are now non-existent. All models applied in general have low bias throughout Table VI-VII. The exception is shown in Table VII for the high truncation case, for which all models are severely biased. Otherwise the Tobit tends to be less biased and have lower standard errors compared to the other models, regardless of sample size or DGP. In contrast to the case with level-specification, the TPM is often on pair with or just slightly worse off than Tobit. Thereby in line with the findings of e.g. Flood and Gråsjö (2001), being superior the Heckman alternatives. The FIML and LIML are on overall the most biased models and have highest standard errors, even for the Heckman DGP. There is no difference observed between the FIML or LIML estimator here either. In contrast to the findings of Dow and Norton (2003) one can thus state that FIML holds preference of LIML for a log-specification.

Sample size tends to have a less importance for logs than for levels. In contrast to table II-IV, there is no clear pattern of systematic discrepancies for the three DGPs. In Table V and VI, the link between sample size and bias is hardly any. The results in Table VII show that there are some improvements occurring in the shift from 50 to 100 observations, for the less truncated case. The case with the 75 % indicates no shifts as such and the models all remain equally biased. The standard deviations decline by the number of observations, but not the bias of the coefficients itself. Thereby reducing the uncertainty of the estimates but still giving biased such.

The absence of implications from the correlation structure in Table II-IV, further holds for log-specification. There are well-defined, yet contradictory, patterns for the Heckman DGP with respect to correlation. In less-truncated scenarios negative correlation is associated with higher bias, for high-truncated scenarios the opposite is true. The underlying reason for these results is unclear and difficult to hypothesize about, especially as it is non-apparent for the TPM case.

The effect of truncation level does not differ substantially from the previous section. For neither of the DGPs are there clear discrepancies between the level and the log specifications with regard to bias. All estimators a slightly more biased for highly truncated cases. The standard errors for all estimators are mostly lower than in the previous section, irrespective of the set-up. The relatively higher standard errors for the high-level truncation case remains and all models are less biased for the low level truncation case. The results are not at all affirmative to the findings of Stewart (2009), stating that Tobit suffers from an increase in zeros whilst TPM remains equally biased A more reasonable conclusion must be as for the level-specification, if one model is affected by more zeros then all are.

### C. Alternative error specification

So far the analysis has concerned sensitivity towards sample size, correlation structure and level of truncation for the different DGPs. Cauchy and Laplace are both fat-tailed distributions, contrasting to the thin tails of a normal distribution. Beyond the established fact that Cauchy tends to pose a severe problem, there is inconclusive evidence on the implications of the deviations from normality. Most importantly there is preciously little done on comparing all the models, not just FIML and LIML. In this section alternative error specifications are analyzed. A single error term and Tobit DGP is, as in previous studies, used. The results from the simulations are shown below in Table VIII-IX:

**Table VIII. Relative bias of marginal effects in levels in %  DGP: Tobit log specification – Cauchy errors**

| Sample: | | N=50 | | | | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Truncation: | Correlation: | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM |
| 25 % | N/A | -17.2 | -0.4 | -15 | -28.7 | -33.4 | -22.3 | -34.8 | -45.3 | -44.5 | -66.3 | -53.7 | -66.7 |
| | | (14.3) | (116.5) | (27.3) | (21.5) | (12.3) | (95.1) | (17.7) | (23.3) | (10.2) | (13.8) | (10.9) | (17.3) |
| 75 % | N/A | -22.4 | 33.1 | -15.2 | -23.3 | -38.5 | 33.2 | -37.9 | -66.4 | -49.8 | -65.7 | -59.4 | -95.2 |
| | | (9.2) | (203.5) | (40.2) | (130.3) | (7.9) | (215.3) | (25.4) | (78.2) | (8.3) | (68.9) | (15.7) | (12.95) |

Bias calculated as average over 500 simulations. Standard errors are given within brackets.

**Table IX. Relative bias of marginal effects in logs in %  DGP: : Tobit log specification – Laplace errors**

| Sample: | | N=50 | | | | N=100 | | | | N=500 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Truncation: | Correlation: | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM | Tobit | LIML | FIML | TPM |
| 25 % | N/A | 4.3 | 1.9 | 6.3 | 6.1 | 3.7 | -1.1 | 4 | 7 | 4.9 | -1.1 | 0.1 | -7.3 |
| | | (2.8) | (23.7) | (11.9) | (7.2) | (1.9) | (15.3) | (8.7) | (5.4) | (1.1) | (5.8) | (4.1) | (4) |
| 75 % | N/A | -4.1 | 18.5 | -2.8 | -11.7 | -6.2 | -13.8 | -8.3 | -13.5 | -2.8 | -27.1 | -13.5 | -21.4 |
| | | (9.6) | (135.4) | (22.5) | (16.7) | (7.2) | (71.5) | (3.8) | (2.2) | (0.3) | (19.4) | (6) | (10.5) |

Bias calculated as average over 500 simulations. Standard errors are given within brackets.

Following the properties of the Cauchy with a variance increasing with observations, we should expect that the level of bias increase with observations. The results in Table VIII are in line with this argument. For all models the fit worsens substantial with an increase in the number of observations, as previously found. The bias is more prominent in the highly

truncated case, although the standard errors are almost the same. As in the tables above, Tobit has lowest standard errors and LIML the highest.

A Laplace distributed error term has less impact on the coefficient estimates as illustrated in Table IX. Similar to the scenarios with normally distributed error terms and Tobit DGP above, all models perform well. The level of bias is unaffected by the sample size for the 25 % scenario. With a truncation rate of 75 % however, bias increase with sample size for all models but Tobit. Thus reassembling the characteristics of Cauchy distributed errors, without the extremes though. Conclusively deviations from the normality assumption of the error terms might well influence model performance, regardless of model

## IV. Conclusion

In many disciplines there are optimal outcomes in which the agent's utility maximizing choice is bounded in an interval, so called corner solutions. The most common approach to corner solutions in practice is by using a Tobit-type model. The three common Tobit-type models, Tobit, TPM and FIML/LIML all are pledged with their own constraints and issues. Evaluating the three different models with respect to ideal situation through a various set of scenarios the relative efficiency is studied in this thesis. The results clearly implicate that regardless of DGP specified, the "true model" always ends up on top. Interestingly the simple Tobit model is shown to a good general choice, have relatively little bias for most scenarios considered. The exception is the Heckman DGP, in which the exclusion restricts takes shape of an omitted variable bias for Tobit. Another surprising result is that FIML/LIML tends to be less biased for a TPM DGP, than for the Heckman DGP. On the contrary to earlier findings, general superiority of TPM over the other models cannot be established. Neither can previous conclusions on relative importance of correlation/error structure be supported.

The focus point of the study was to examine and evaluate the specific bias entailed to small samples for corner solutions. The results show that the number of observations may have a significant impact on coefficient bias. In line with the findings of Jonsson (2008), the effect of sample size is pronounced up to 100 observations and almost fully gone at 500. Unexpectedly the Tobit model is shown to be remarkably robust towards issues with small samples and is only affected to a minor extent, in contrast to the more advanced TPM and Heckman models. The most apparent drawback from smaller samples is however convergence problems for the MLE, in particular for the FIML. The lack of information might further make it problematic to estimate the inverse mills ratio and thus make LIML inappropriate. TPM/Tobit are only affected by these issues to a smaller extent. A related information issue is the number of zeros in the samples, whereas more zeros should imply more biased coefficients and larger standard errors. The evidence from simulations indicates that this is not necessarily the case. Although the bias tends to be only slightly higher for the more truncated case, the magnitude of the standard errors is always larger.

Unlike the findings of Martin and Pham (2008), there is no evidence of a systematic dissimilarity levels and logs for relative model bias. Except for the Tobit DGP whereas logging the dependent variable substantially improve the fit for the other models. All the models are on overall less biased when estimated in logs than levels in absolute terms.

Particularly for the small samples with 50 observations this difference can be substantial for the TPM/Heckman DGPs. Thereby suggesting that working with variables in logs instead of levels may have a positive impact on model performance, especially if the DGP is Tobit.

The implications of a non-normal distribution of the error term are highly dependent on the alternative distribution. For a Cauchy distributed error term all estimates are highly biased and no model can be preferential over another. On the other hand the non-normal distribution is Laplace, there is little deviation from the results of a normally distributed error term. Thus the outcome of the simulations indicates that any issues related to violations of the normality assumption, are likely to be bounded to that specific distribution rather than deviation from normality. It should however be stressed that as sample size increase for the highly truncated Laplace case, the bias approaches that associated with a Cauchy distributed error term.

The way to adequately approach corners solutions will remain a pressing concern in the empirical literature. It is essential to recall the reason for the truncated/censored sample, let it plausible outcome caused by the agent's action and not a data issue which is often assumed. The researcher must therefore from start be aware of whether it is a censored, latent or truncated dependent variable that is of interest. In the latter case my results indicates that the basic Tobit model provide a good general fit to the data, unless there is risk of an omitted variable in the selection equation. As the TPM is found to be equally sensitive to this issue only the LIML/FIML gives relatively unbiased estimates for such scenarios. Additionally, Tobit is by shown to be by far more robust to small samples, in line with findings of Greene (2004), than the other models. Thereby a good starting point for empirical research with small samples would be to use the basic Tobit and possibly compare it with Heckit. Future theoretical papers should seek to expand on the generality of the results, more closely scrutinizing the deviation from the established truth of TPM's otherwise superior performance.

## References

Aristei, D., and Pieroni, L., 2009, Addiction, social interactions and gender differences in cigarette consumption, *Empirica*, Springer, vol. 36(3), 245-272

Bartus, T., 2005, Estimation of marginal effects using margeff, *The Stata Journal*, 5(3), 309-329

Burke, W., 2009, Fitting and interpreting Cragg's tobit alternative using Stata, *Stata Journal*, 9(4), 584-592

Carson, R.T., and Sun, Y., 2007, The Tobit model with a non-zero threshold, *Econometrics Journal*, volume 10(3), 488-502

Cragg, J., 1971, Some Statictical models for limited dependent variables with application to the demand for durable goods, *Econometrica*, 39(5), 829-844.

Dietrich, J., 2001, The Effects of Choice-based Sampling and Small-sample Bias on Past Fair Lending Exams, *OCC Economics Working Paper,* 2001-2

Dow, W. H., and Norton, E. C., 2003, Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions, *Health Services and Outcomes Research methodology*, March 2003, Volume 4, issue 1, 5-18

Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P. (1984) Choosing Between the Sample-Selection Model and the Multi-Part Model. *Journal of Business & Economic Statistics*, 2, 3, 283-289.

Eaton, J., and Tamura, A. 1994, Bilateralism and regionalism in Japanese and U.S. trade and direct foreign investment patterns, *Journal of Japanese and International economies*, 8, 478-510

Embrechts, P., Lindskog, F., McNeil, A., 2003, Modelling dependence with copulas and applications to Risk Management, In: *Handbook of Heavy Tailed Distributions in Finance* Ed: S. Rachev, Elsevier, Chapter 8, 329-384.

Giger, N., 2009, Party Policy representation over time and across nations, *Paper prepared for presentation at the ECPR General Conference, Potsdam,* September 10-12

Gordy, M., and E. Heitfield., 2002, Estimating default correlations from short panels of credit rating performance data, *Working Paper*, Federal Reserve Board

Greene, W., 2004, The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects, *Econometrics Journal*, volume 7, 98–119

Hart, R. A. Jr., and Clark, D.H., 1999, Does Size Matter? Exploring the Small Sample Properties of Maximum Likelihood Estimation, *Paper prepared for presentation at the Annual Meeting of the Midwest Political Science Association*, March, Chicago, IL.

Hay, J.W., Leu, R., and Fohrer, P., 1987, Ordinary least squares and sample-selection models of health-care demand, *Journal of Business and Economic Statistics,* 5, 499–506

Heckman, J., 1974, "Shadow Prices, Market Wages and Labor Supply," *Econometrica*, volume 42(4), 679-94

Heckman, J., 1979, Sample selection bias as a specification error, *Econometrica,* Volume 47 (1): 153–61

Hsu, A. C, and Liu, S-C., 2008, The hurdle models choice between truncated normal and lognormal, *Applied Economics*, Volume 40(2), 201-207

Hurd, M., 1979, Estimation in truncated samples when there is heteroscedasticity. *Journal of Econometrics*, 11(2-3), 247-258

Jonsson, R., 2012, When does Heckman's two-step procedure for censored data work and when does it not? *Statistical Papers*, February 2012, Volume 53(1), 33-49

Leung, S.F., and Yu, S., 1996, On the choice between sample selection and two-part models, *Journal of Econometrics,* 72, 197–229

Manning,W.G., Duan, N., and Rogers,W.H., 1987, Monte Carlo evidence on the choice between sample selection and two-part models, *Journal of Econometrics,* 35, 59–82

Martin, W., and Pham, C. S., 2008, Estimating the gravity equation when zero trade ows are frequent, Technical report.

Paarsch, Harry J., 1984, A Monte Carlo Comparison of Estimators for Censored Regression Models, *Journal of Econometrics*, 24, 197-213.

Puhani, P,A., 2000, The Heckman correction for sample selection and its critique. *Journal of Economic Surveys,* volume 14(1), 53-68

Radcliff, B., 1992, The Welfare State, Turnout, and the Economy: A Comparative Analysis, *American Political Science Review* 86(2): 444–454.

Relles, D., and Stolzenberg, R.M., 1990, Theory Testing in a World of Constrained Research Design: The Significance of Heckman's Censored Sampling Bias Correction for Nonexperimental, *Research Sociological Methods and Research*, 18, 395-415

Schoonbroodt, A., 2004, Small Sample Bias Using Maximum Likelihood versus Moments: The Case of a Simple Search Model of the Labor Market. *Working Paper, University of Minnesota*, 1-29.

Shadish, W. R., Clark, M.H., and Steiner, P.M., 2008. Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random to non-random assignment, *Journal of American Statistical association*, 103, 1334-1356.

Sigelman, L., and Zeng, L., 1999, Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models, *Political Analysis,* 8, 167–182.

Smith, K. B., and Meier, K., 1995, Politics and the Quality of Education: Improving Student Performance, *Political Research Quarterly,* 48(2).

Stewart, J., 2009, Tobit or Not Tobit?, *IZA Discussion Paper*, No. 4588

Svendsen, M.F., and Haugland, S.A., 2011, Host country institutional pressures and cross-border relationship governance, *International Business Review*, 20(3), 324-337

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

Wang, K., and Ray, J.L., 1994, Beginners and Winners: The Fate of Initiators of Interstate Wars Involving Great Powers Since 1495, *International Studies Quarterly* 38(1):139–154.

Wooldridge, J.M., 2010, *Econometric analysis of cross section and panel data*, MIT Press second edition, Cambridge, MA,