Selective donations: A non-technical discussion of the use of parametric models in the analysis of donation behavior

Maarten L. Buis, Department of Sociology, University of Tübingen

Pamala Wiepking, Department of Sociology & Erasmus Centre for Strategic Philanthropy (ECSP), Erasmus University Rotterdam, the Netherlands

November 8, 2011

Abstract

In philanthropic research, a much-neglected aspect is the methodology used for analyzing donating behaviour. This is a serious problem, as different methods lead to as many different conclusions. In this paper, we will discuss the limitations of using linear (OLS) regression analyses, as well as discuss the advantages and limitations of the two most often used parametric models for analyzing charitable donations: Tobit and Heckman Two-Stage regression analyses. Our main objectives in this paper are to explain Tobit and Heckman two-stage regression models in a non-technical way and to clarify the consequences for the findings derived with these models.

Keywords: Philanthropy, Giving, Methodology, Censoring, Sample selection model

Introduction

Analyzing charitable giving brings along a tricky methodological problem: How to treat people who have not donated any money? In recent literature different ways have been proposed for analyzing charitable giving, some giving solutions to the problem, others just ignore problematic aspects. One of the first methods used is linear or Ordinary Least Square (OLS) regression analysis on the amount of money donated (Boskin & Feldstein, 1977; E. Brown, 1987). However, linear regression analysis produces biased results, due to truncation or selection bias. Furthermore, it fits a straight line through the data and any straight line (except a horizontal line) will eventually become negative, leading to predicting negative donations (Rooney, Steinberg, & Schervish, 2001). Simply excluding the non-donors from the dataset is only a solution if one wants to make statements about the population of donors. It is very arguable that charitable donors are a non-random sample of the population, hence the results cannot be generalized to the entire population when non-donors are excluded (Breen, 1996; Rooney et al., 2001; Yen, 2002).

Another model often used when analysing charitable giving, is the Tobit model (Andreoni & Miller, 2002; Eleanor Brown, 2001; Rooney et al., 2001; Smith, Kehoe, & Cremer, 1995). Tobit is a form of truncated regression analysis, which can be used to censor the non-donors (left-censoring). Finally, there also is the Heckman Two-Stage analysis (Heckman, 1979), which estimates two models: one for the selection process whether or not to give (all cases), and one for the decision how much is given (donors only) (Rooney et al., 2001; Smith et al., 1995). In this paper we will discuss the mechanisms with respect to charitable giving implied by these different models, in order to help researches choose a model and interpret their results.

Why non-donors pose a problem

Theoretically, there are two mechanisms according to which people can make charitable donations. These two mechanisms are the *censoring-mechanism* and the *selection-mechanism.* We will start with describing the censoring-mechanism.

According to the censoring-mechanism, people first decide how much they are willing to give to a charitable organisation, let's call this amount y^*. There are reasons to believe that some these amounts are too low to actually donate, either because these amounts are socially undesirable, or because they are not worth the effort (such as wanting to donate €0.10 by credit slip). Therefore a person will only give y^* if y^* is more than some minimum. When y^* is less than that minimum, no donation will be made. According to this censoring-mechanism, there is an absolute cut-off point below which someone decides not to make a donation at all. Methodologically, the Tobit model is suitable to analyse charitable giving according to the censoring mechanism.

The censoring mechanism is rather strict, in the sense that the probability is directly related to the intended donation y^*. As a consequence, the effects of explanatory variables on the probability of donating are completely determined by the effects of these variables on the intended donations y^*. The selection-mechanism is less strict, it still allows for the possibility that persons with higher intended donations are more likely to donate, but it assumes that explanatory variables can have their own effect on the probability that people will make a donation. This selection-mechanism corresponds methodologically to the Heckman Two Stage regression model.

Both the censoring-mechanism, and the selection-mechanism imply that estimates obtained using standard regression techniques will be biased. One solution is to only make statements about donors, and hence exclude the zero-donations from the analysis. However, only including the sub-population of donors in the analyses will in some cases lead to biased results. The intuition behind this is displayed in figure 1.

Figure 1a shows a scatter plot of a hypothetical highly regular dataset. It shows that an explanatory variable *x* is positively related to the amount people donate. We represent this positive effect with the solid line. Figure 1a also shows that this relation is not perfect. Not every observation lies on the line. Some observations lie above the line and some below; these are 'errors'. We assume that on average these errors cancel each other out. This is true to an extreme extend in the hypothetical dataset displayed in figure 1: for each positive error (e.g. e_1) there is one negative error of exactly the same size (e_2). Each positive error can be thought of "trying to pull the estimated effect upwards". Similarly, each negative error "tries to pull the estimated effect downwards". However, the net error is zero, because the positive and negative errors are evenly matched.

Figure 1b shows the rationale of a Tobit process. It shows a scatter plot of the intended donations *y** and the actual donations *y* against the explanatory variable *x*. The relationship between *y** and *x* is the same as in figure 1a, i.e. there is a positive relation between *y** and *x* and each positive error is exactly matched by a negative error. However, not every person makes a donation. The censoring-mechanism, and with that, the Tobit process assumes that a person only donates if his intended donation (*y**) is above an absolute threshold. In this example it is assumed that people actually make a donation if their intended donation is higher than 4 euro. Using only information about donors would imply that the dataset consist only of the crosses. In that case the upward pull from the positive error a^+, b^+, c^+ and d^+, is no longer cancelled by the downward pull from the negative errors a^-, b^-, c^- and d^-. The upward pull from the positive errors causes the estimated effect to be rotated to a flatter position. In other words, the estimated effect of *x* on donations (dashed line) is weaker than the population effect (solid line).

<<Insert figure 1 about here>>

The Tobit process is rather strict: it assumes that all persons with an intended donation less than the threshold do not donate. The selection-mechanism and the Heckman two-stage process relax this assumption. However, only analysing the donors will still lead to biased results if the probability of donating is associated with intended donation. In particular one can expect that people with a lower intended donation are less likely to donate. As a consequence, people with positive errors are more likely to donate than people with negative errors. This is shown in figure 1c. This leads to a situation in which some positive errors no longer balance negative errors. The upward pull from a^+, b^+, c^+ and d^+ no longer cancel the downward pull from a^-, b^-, c^- and d^-. Again this leads to an estimated effect that is weaker (flatter) than the real (population) effect.

Why assumptions are the solution

The previous section showed that the use of normal regression analysis in analyzing charitable giving is appropriate as long as positive errors are (on average) balanced by negative errors. It also showed two plausible scenarios under which non-donors cause this assumption to fail. However, most regression textbooks will not only mention balanced errors, but also mention that the errors are assumed to be normally distributed around the regression line. Note that this is a stronger assumption. The errors in figure 1 balance each other out, but are not normally distributed around the regression line. The normality assumption is relevant for hypothesis testing. The correct regression line will be estimated if the errors are balanced but not normally distributed. However, the normality assumption will play a crucial role in estimating the regression line when one wants to correct for the bias caused by non-donors through Tobit or Heckman two-stage regression. First we explain how the normality assumption functions in normal regression. After that we show how Tobit and Heckman two-stage use the normality assumption to fit the correct regression line.

<<Insert figure 2 about here>>

In figure 2a there is a ‘cloud’ of data points centred on the regression line. The density of points around the line can be represented by a probability density function. The bell shape curve from the normal distribution is such a probability density function. One just needs two pieces of information in order to draw this curve: the mean and the variance. In linear regression the variance is considered to be constant, but the mean changes. For each value of x, the regression line at that point represents the mean. So there will be a different density curve for each value of x. Figure 2b shows three of these normal distributions. One can think of these curves as popping out of the paper, whereby the third dimension represents the density of points.

Figure 3 shows the distribution of donations according to a Tobit process. Again we assume that people only donate if their intended donation is higher than 4. The non-donors are represented by a vertical tick mark at zero.[1] This means that we do not know their intended donation (although it is below 4), but do know their value for x. The densities are the same as in figure 2b except that they are cut-off at 4. Figure 3b shows the density function when *x* equals 5. It is the second density curve in figure 3a rotated 90° clockwise. The left tail is cut off, so the positive errors (the right tail) are no longer balanced by the negative errors. If we can reconstruct the entire density function, than an estimate of the population regression line can be obtained. The data contains two pieces of information that are used to obtain this reconstruction, and these two pieces of information are derived from both the donors and the non-donors. The donors provide information on the part of the distribution represented by the solid line. This information can be used to extrapolate. The non-donors provide information on the proportion of non-donors. This proportion needs to equal the area under the dotted line. Tobit regression uses these two pieces of information to account for the missing information, and then estimates the regression using these densities. Notice that Tobit regression depends on two assumptions: First, the non-donors do not donate because they want to give less then the absolute cut-off value. This assumption cannot be tested, as it depends on a researchers theoretical view on donating behaviour, in this case the censoring-mechanism. Second, the intended donations are normally distributed around the regression line. This can be tested using Pagan and Vella's test for censored normality (1989).

<<Insert figure 3 about here>>

Figure 4 shows a Heckman two-stage-process applied to predicting charitable giving. A Heckman two-stage process allows for the possibility that lower donations are associated with lower probabilities of donating. In that case the density curves are pushed down at lower values of donations and are kept intact at higher values. Figure 4a shows various density curves for different values of *x*. The curve for *x*=2 is associated with such low donations that the probability of donating is virtually zero everywhere. Consequently the curve reduces to a flat line. The curve for x=8 is associated with such high donations that the probability of donating is virtually 1 everywhere. Consequently the density curve remains unchanged. The curve for x=5 is an interesting intermediate case. The probability of donating changes considerably across its feasible range of donations. As a result the lower side of the curve is pushed down whereas the upper side of the curve remains unchanged. This is shown in more detail in figure 4b. The bottom panel shows how the probability of donating changes for different intended donations. The top panel shows the distribution of intended donations (dotted curve) and actual donations (solid curve). A person who intends to donate 4.5 euro (at the dashed vertical line) has a probability of donating of about .2. So even though the probability density of intended donations of 4.5 is .35 the probability density of actual donations is only .2*.35=.07. Once a person intends to donate more than 6 euro the probability of donating becomes virtually one and the curves of intended and actual donations overlap. If a person intends to donate less than 3 euro, the probability of donating becomes virtually zero and the probability density of the actual donations becomes zero. Again, the positive errors are no longer balanced by the negative errors. The upper panel in figure 4b also gives information about the proportion of non-donors: the area between the solid curve and the dotted curve represents this.

<<Insert figure 4 about here>>

Once the probability of donating and the probability density function of actual donations are known, then this information can be used to reconstruct the probability density function of intended donations. These reconstructed density curves can be used to estimate the regression parameter of x on the intended donations. The donors and non-donors give information about the probability of donating and the donors give information about the probability density function of actual donations. Heckman two-stage uses both sources of information to reconstruct the probability density functions of the intended donations and consequently the actual donations. In order to construct the probability density curve, Heckman two-stage assumes this curve is normally distributed. Therefore, applying the Heckman two-stage procedure depends on both correctly estimating the probabilities of donating, and on the assumption that intended donations are normally distributed. Bera et al. (1984) describe a test for the normality assumption in Heckman two-stage.

Within selection models the difficult part is to obtain information from the data about the relationship between intended donations and the probability of donating. One can identify this part of the model either by assuming that intended donations are normally distributed and the sub-model for the probability of donating is correctly specified, or by an exclusion restriction in the sub-model for intended donations (Bradley, Holden, & McClelland, 2005). The former option relies on strong assumptions that can easily be wrong. The latter option is less dependent on these assumptions, but it requires that the researcher uses at least one variable to estimate the probability of donating which is excluded when estimating the amount donated. It is often hard to find such a variable that influences the probability for donating but not the amount intended to donate. An example that has proven to be an useful selection variable is whether or not people are solicited to make a donation. For example, people who have been asked for a donation in church (for example in the two weeks prior to the survey) are very likely to have made a (religiously oriented) donation, but it is likely that being asked for this donation has not influenced the amount donated.

Interpretation of results

Once the parameter estimates are obtained, they have to be interpreted. The interpretation of the results from a Tobit or Heckman two-stage model are more complicated than in normal regression because they imply different types of effects of the explanatory variable:

- The effect of *x* on the intended donation.
- The effect of *x* on the probability of donating.
- The effect of *x* on the amount donors give when they give.
- The effect of *x* on all donations, including zero donations from non-donors.

The effect of the explanatory variable on the intended donation is easiest to compute: this is the ‘regression’ parameter for that variable. All other effects require some calculations. Not all statistical packages that allow you to do Tobit or Heckman two-stage have procedures to output all of these effects, so the formulas that can be used to compute them can be found in appendix A. Furthermore, these effects tend to change when the values of the explanatory variables change. A common solution is to calculate the effect for an individual who is average with respect to the explanatory variables: the effect of *x* when all explanatory variables are equal to their average. Alternatively, one can calculate the effect for an average individual: one computes the effect for each individual in the dataset and than one computes the average of those effects. These two often tend to be close but not the same. Conceptually, the latter gets closer to the idea of a summary measure of the effect. Finally, one can plot how the outcome of interest changes when one of the *x*’s change, while keeping the remaining *x*’s at their mean value.

Conclusion

The main goal of this paper was to give a non-technical overview of two of the most regularly used parametric models available to researchers for analyzing charitable giving. We started by showing that ordinary regression models can lead to predicting negative donations, and cannot adequately deal with the issue of non-donors. Next, we discussed Tobit and Heckman two-stage regression models as possible solutions to deal with censoring and sample selection in data on charitable giving. Finally, we gave some brief comments on how to interpret results of Tobit and Heckman two-stage regression analysis.

Appendix

The formulas that are presented here can be used to turn raw output from your estimation command into the desired marginal effects or predicted values. The raw output will generally give you (using the notation defined below) β and σ for the Tobit regression and β, α, σ and ρ for Heckman two-stage regression. When computing the marginal effects one would choose a value for each x_j (and w_j), usually the mean, and compute $\boldsymbol{x}_i\boldsymbol{\beta}$, δ_i, and λ_i, and plug these into the relevant equation from the table below. The predicted values are generally used to graph how the predicted intended donation, donation, donation by donors only, or probability of donating, changes when one x (or w) changes. So, one calculates for a lot values of the variable of interest the appropriate predicted value, while keeping the value for all other variable fixed, usually at the mean.

		Tobit	Heckman two-stage
Intended donations	Predicted value	$E(y^*)=x_i\beta$	$E(y^*)=x_i\beta$
	Marginal effect	$\dfrac{\partial E(y^*)}{\partial x_j}=\beta_j$	$\dfrac{\partial E(y^*)}{\partial x_j}=\beta_j$
Donations if donating	Predicted value	$E(y \mid y^*>c)=x_i\beta+\sigma\lambda_i+c$	$E(y \mid d)=x_i\beta+\rho\sigma\lambda_i$
	Marginal effect	$\dfrac{\partial E(y \mid y^*>c)}{\partial x_j}=\beta_j(1-\lambda_i^2-\delta_i\lambda_i)$	$\dfrac{\partial E(y \mid d)}{\partial x_j}=\beta_j-\alpha_j\rho\sigma(\delta_i\lambda_i-\lambda_i^2)$
Donations (including zero donations)	Predicted value	$E(y)=\Phi(\delta_i)(x_i\beta+\sigma\lambda_i)+(1-\Phi(\delta_i))$	$E(y)=E(y \mid d)\Pr(d)$
	Marginal effect	$\dfrac{\partial E(y)}{\partial x_j}=\Phi(\delta_i)\beta_j+(c-c_y)\phi(\delta_i)\dfrac{\beta_j}{\sigma}$	$\dfrac{\partial E(y)}{\partial x_j}=E(y \mid d)\dfrac{\partial \Pr(d)}{\partial w_j}+\dfrac{\partial E(y \mid d)}{\partial x_j}\Pr(d)$
Probability	Predicted value	$\Pr(y^*>c)=\Phi(\delta_i)$	$\Pr(d)=\Phi(\delta_i)$
	Marginal effect	$\dfrac{\partial \Pr(y^*>c)}{\partial x_j}=\phi(\delta_i)\dfrac{\beta_j}{\sigma}$	$\dfrac{\partial \Pr(d)}{\partial w_j}=\phi(\delta_i)\alpha_j$

Notation Tobit

β_j Regression coefficient for the j^{th} variable.

σ Standard error or the standard deviation of the normal distributions around the regression line.
c Minimum acceptable donation, usually but not necessarily zero.
c_y Amount someone gives when he wants to give less than c, usually but not necessarily zero.
δ_i $\left(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_J x_{Ji} - c\right)/\sigma$
$\boldsymbol{x}_i\boldsymbol{\beta}$ $\beta_0 + \beta_1 x_{1i} + \cdots + \beta_J x_{Ji}$
ϕ Probability density function of the standard normal distribution.
$\mathbf{\Phi}$ Cumulative density function of the standard normal distribution.
λ_i $\phi(\delta_i)/\Phi(\delta_i)$

Notation Heckman two-stage
β_j Regression coefficient explaining intended donations for the j^{th} variable.
x_j j^{th} variable explaining intended donations.
α_j Regression coefficient explaining probability of donating for the j^{th} variable.
w_j j^{th} variable explaining probability of donating.
σ Standard error or the standard deviation of the normal distributions around the regression line.
ρ The strength of the relation between probability of donating and intended donation.
δ_i $\alpha_0 + \alpha_1 w_{1i} + \cdots + \alpha_J w_{Ji}$
$\boldsymbol{x}_i\boldsymbol{\beta}$ $\beta_0 + \beta_1 x_{1i} + \cdots + \beta_J x_{Ji}$
ϕ Probability density function of the standard normal distribution.
$\mathbf{\Phi}$ Cumulative density function of the standard normal distribution.
λ_i $\phi(\delta_i)/\Phi(\delta_i)$
$(y|d)$ The amount someone donations if he donates
$\Pr(d)$ Probability someone donates

Endnote

[1] Imagine that one tries to fit a line through all data points (both donors and non-donors). In this case the line would be too steep, i.e. all the zero donations would draw the line downwards. The effect of *x* would now be overestimated. If the non-donors are included during a regression analysis, the regression line can be either over- or underestimated, depending upon what value the non-donors are given (Sigelman & Zeng, 1999).

References

Andreoni, J., & Miller, J. (2002). Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica, 70*(2), 737-753.

Bera, A. K., Jarque, C. M., & Lee, L.-F. (1984). Testing the Normality Assumption in Limited Dependent Variable Models. *International Economic Review, 25*(3), 563-578.

Boskin, M. J., & Feldstein, M. (1977). Effects of the Charitable Deduction of Contribution by Low and Middle Income Households: New Evidence of the National Survey of Philanthropy. *The Review of Economics and Statistics, 59*, 351-354.

Bradley, R., Holden, S., & McClelland, R. (2005). A Robust Estimation of the Effects of Taxation on Charitable Contributions. *Contemporary Economic Policy, 23*(4), 545-554.

Breen, R. (1996). *Regression Models: Cencored, Sample Selected, or Truncated Data*. Thousand Oaks: Sage.

Brown, E. (1987). Tax Incentives and Charitable Giving: Evidence from New Survey Data. *Public Finance Quarterly, 15*, 386-396.

Brown, E. (2001). *Making Philanthropy Work: Social Capital and Human Capital as Predictors of Household Giving*. Claremont, CA.: Claremont Colleges Working Papers in Economics, Claremont University Consortium.

Heckman, J. (1979). Sample Selection Bias as Specification Error. *Econometrica, 47*, 153-162.

Pagan, A., & Vella, F. (1989). Diagnostic Tests for Models Based on Individual Data: A Survey. *Journal of Applied Econometrics, 4*, S29-S59.

Rooney, P., Steinberg, K., & Schervish, P. (2001). A Methodological Comparison of Giving Surveys: Indiana as a Test Case. *Nonprofit and Voluntary Sector Quarterly, 30*(3), 551-568.

Sigelman, L., & Zeng, L. (1999). Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models. *Political Analysis, 8*(2), 167-182.

Smith, V. H., Kehoe, M. R., & Cremer, M. E. (1995). The Private Provision of Public Goods: Altruism and Voluntary Giving. *Journal of Public Economics, 58*, 107-126.

Yen, S. T. (2002). An Econometric Analysis of Household Donations in the USA. *Applied Economics Letters, 9*, 837-841.

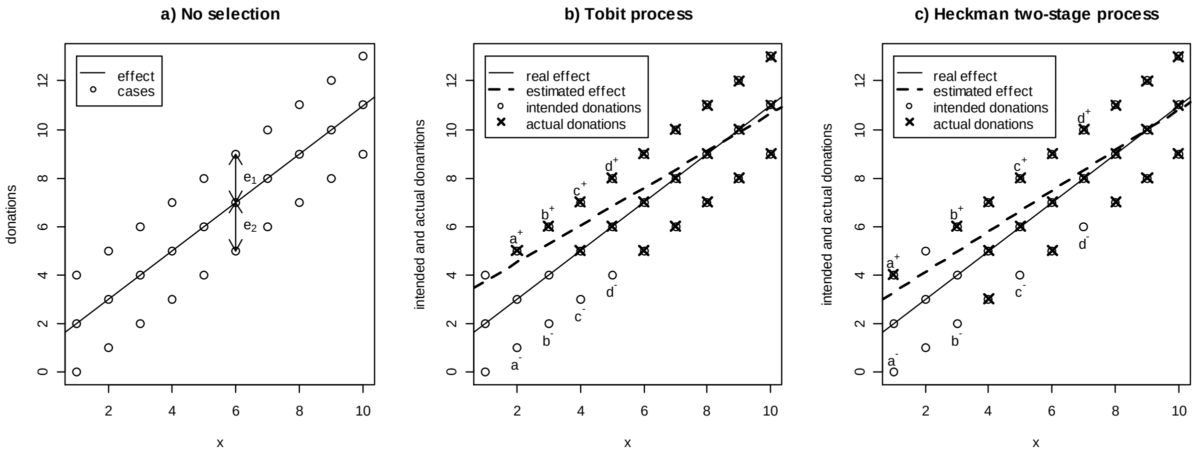

Figure 1 Rationale behind (a) normal; (b) Tobit; and (c) Heckman two-stage regression analyses

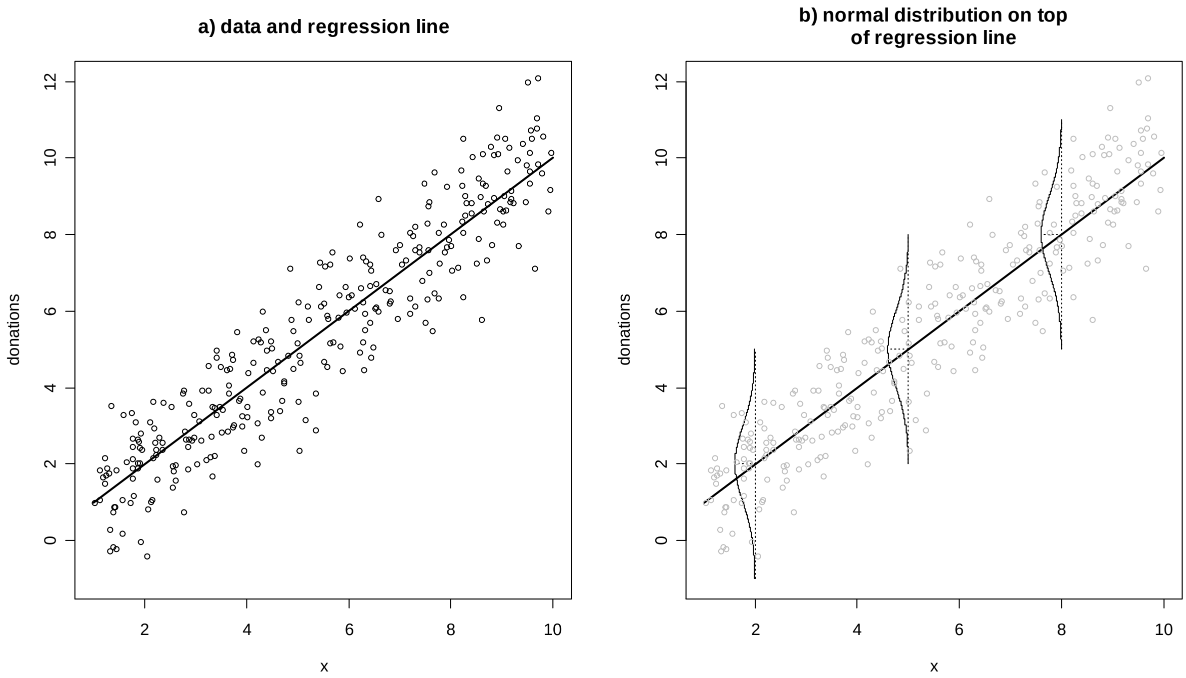

Figure 2 The distribution of data points and normality assumption in normal regression

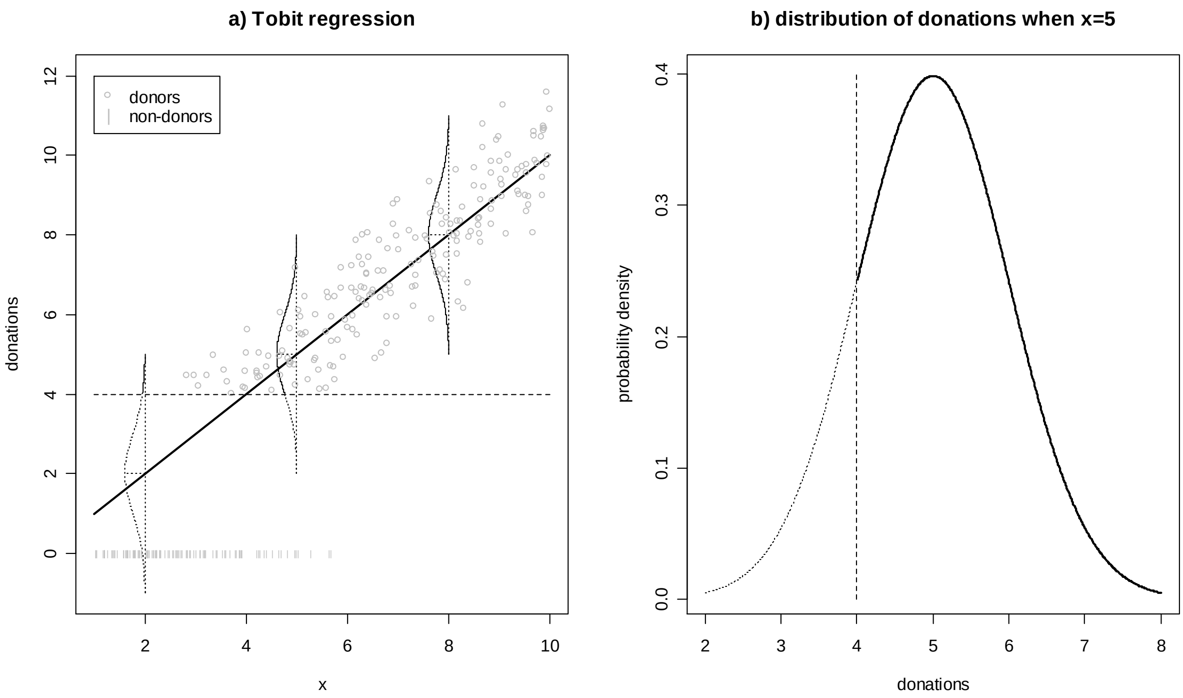

Figure 3 Distribution of data points and normality assumption in Tobit regression

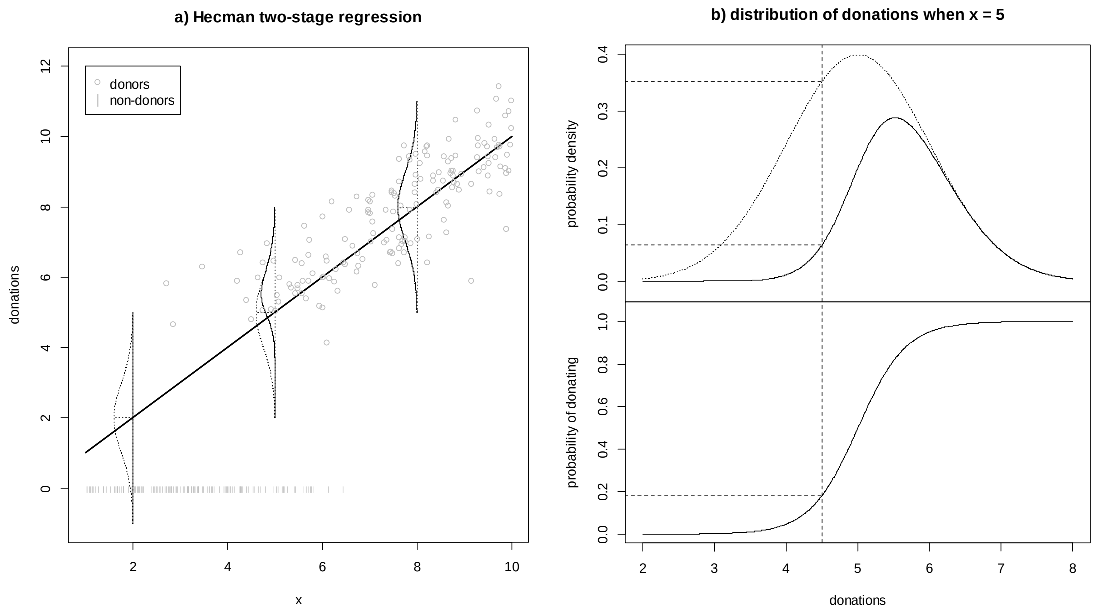

Figure 4 Distribution of data points and normality assumption in Heckman two-stage regression