2: Discrete Choice

GECO 6281 Advanced Econometrics 1 (Lab)

Patrick Mokre

Fall 2020

Limited Dependent Variables

- **What exactly does an OLS estimation coefficient capture?**
- **Under which technical conditions is that estimation BLUE?**
- **Under which conditions does that kind of modeling make sense intuitively/in a modeling context?**
- **Bonus: What is the difference between consistency and unbiasedness**?

Limited Dependent Variables 2

$\hat{\beta}_{OLS}$ is an approximation to $\frac{\partial y}{\partial X}$. Inituitively, this makes the most sense with a **continuous** dependent variable and covariates.

$\hat{\beta}_{OLS}$ is **consistent** and **efficient** und the Gauss-Markov conditions.

- $E\epsilon_i = 0 \quad \forall i \in N$
- $\epsilon_1, ..., \epsilon_n$ and $x_1, ..., x_n$ are *independent*
- $Var(\epsilon_i) = \sigma^2 \quad \forall i \in N$
- $cov(\epsilon_i, \epsilon_j) = 0 \quad \forall i, j \in N, \forall j \neq i$

Limited Dependent Variables 3

Often microeconomic data is presented in discrete or discrete mixed continuous form.

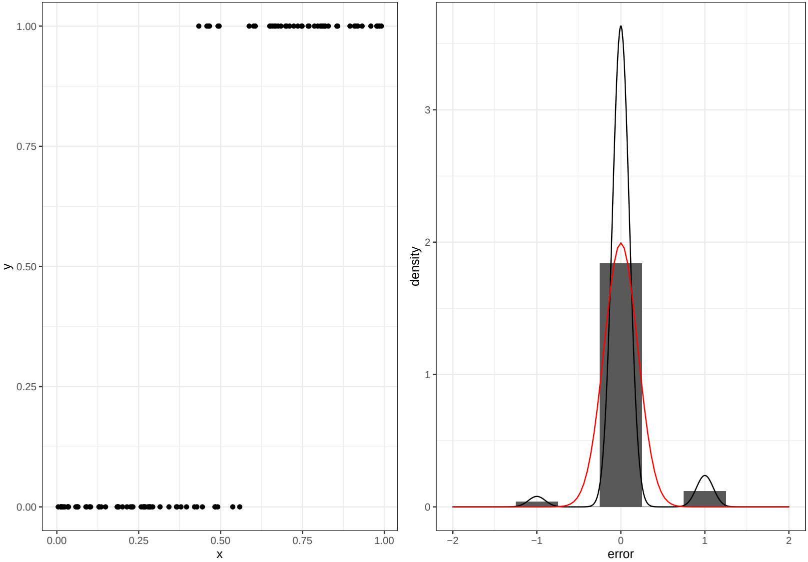
Problem 1: If one estimates binary data using OLS, $x'\beta$ must be read as a probability, which by definition can only be between 0 and 1. This is only possible if either x or β are artificially restricted.

Problem 2: Usually the error term is **not normally distributed** and suffers from **heteroskedasticity**:

$$\begin{aligned}
P(y_i = 1 \mid x_i) &= x_i'\beta \\
P(\epsilon = -x_i'\beta \mid x_i) &= P(y_i = 0 \mid x_i) = 1 - x_i'\beta \\
P(\epsilon = 1 - x_i'\beta \mid x_i) &= x_i'\beta \\
\Rightarrow V(\epsilon \mid x_i) &= x_i'\beta(1 - x_i'\beta) \neq V(\epsilon)
\end{aligned}$$

Clearly, a bipolar distribution is not Gaussian Normal, and the variance depends on the value of the covariates.

Non-Gaussian Error Distribution



Binary Choice Modeling

$$P(y_i = 1 \mid x_i) = G(x_i, \beta)$$

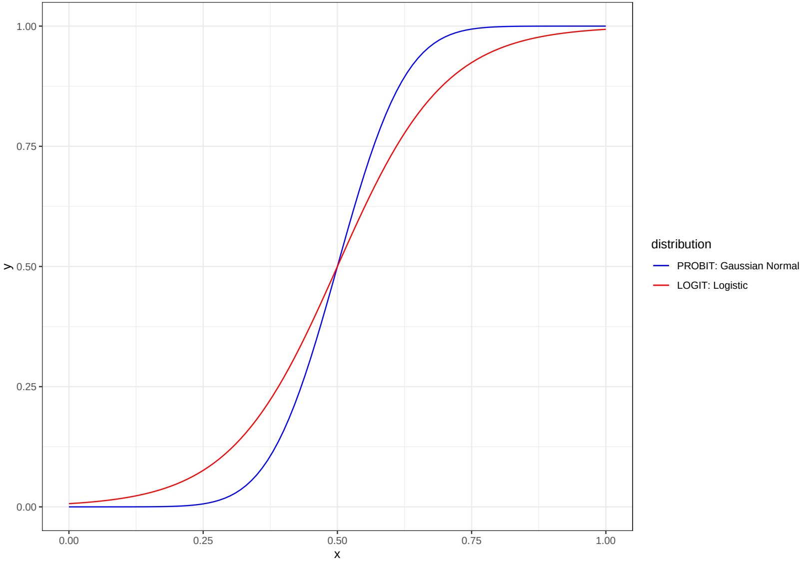
If you choose for the function $G(x_i, \beta)$ the Gaussian Normal distribution $\Phi(x_i'\beta)$, this is called a **Probit** model:

$$\frac{\partial \Phi(x_i'\beta)}{\partial x_{ik}} = \phi(x_i'\beta)\beta_k$$

The logistical distribution $\frac{exp(x_i'\beta)}{1+exp(x_i'\beta)}$ gives a **Logit model**.

$$\frac{\partial L(x_i'\beta)}{\partial x_{ik}} = \frac{exp(x_i'\beta)}{(1 + exp(x_i'\beta))^2}\beta_k$$

Normal and Logit Link Function



Latent Variable Models

One can also model a bivariate outcome as the result of a censoring process. For this, one makes behavioural assumptions on why a variable never materializes.

Let y_i^* be an underlying (latent) variable. As an example, think of a reservation wage: If a person is offered less than \$ 1500, they may not even enter the labor market.

$$\begin{aligned} y_i^* = x_i'\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \\ y_i = 1 \quad if \quad y_i^* > 0 \\ y_i = 0 \quad if \quad y_i^* < 0 \end{aligned}$$

Latent Variable Models

The model can be estimated using a simple likelihood formulation.

$$L(\beta) = \prod_{i}^{N} P(y_i = 1 \mid x_i; \beta)^{y_i} P(y_i = 0 \mid x_i; \beta)^{1-y_i}$$

Since the natural logarithm is a monotonous function, the value β that maximizes the likelihood also maximizes the log-likelihood $LL(\beta)$. Since Log-Likelihoods can be summed up rather than multiplied the procedure becomes **computationally more efficient** and does less often run into problems with **floating digits**.

$$LL(\beta) = \sum_{i}^{N} y_i log(P(y_i = 1 \mid x_i; \beta)) + (1 - y_i) log(P(y_i = 0 \mid x_i; \beta))$$

Both Logit and Probit models can be estimated using Maximum (Log-) Likelihood routines: One calculates the (log-) likelihood function for a number of parameter combinations and picks the highest.

Goodness of Fit

Goodness of Fit in probabilistic models mostly measure either precision in **calculated probabilities compared to observed frequencies** or **prediction of observed data**.

Often GOF statistics implicitly compare the model with one that includes only a constant by comparing the calculated likelihoods, L_1 and L_0 respectively.

Amemiya Pseudo-R^2:

$$1 - \frac{1}{1 + 2(logL_1 - logL_0)/N}$$

McFadden statistic:

$$1 - \frac{logL_1}{logL_0}$$

Restricted Dependent Variables: TOBIT

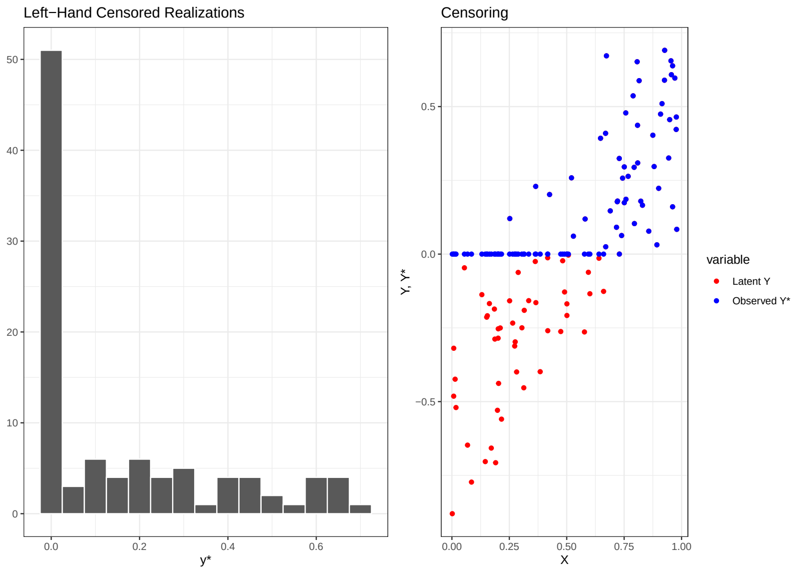
When dependent variables are continuous, but constrained, more problems arise. Examples are when a variable is zero for a large part of the population and positive for the rest (eg. expenditures, income from a certain type of activity or asset, work hours).

Tobit models are well-suited for such latent variable problems. It applies conditional probabilities ot the problem, usually introducing a Gaussian Normal density function.

$$P(y_i = 0) = P(y_i^* \leq 0) = P(\epsilon_i \leq -x_i'\beta) = 1 - \Phi(\frac{x_i'\beta}{\sigma})$$

$$E(y_i \mid y_i > 0) = x_i'\beta + E(\epsilon_i \mid \epsilon_i > -x_i'\beta) = x_i'\beta + \sigma\frac{\phi(x_i'\beta/\sigma)}{\Phi(x_i'\beta/\sigma)}$$

Restricted Dependent Variables: TOBIT 2



Restricted Dependent Variables: TOBIT 3

The parameters obtained in a Maximum-Likelihood procedure can be interpreted in two ways. Note that the ML procedure has to simultaneously estimate β and σ.

Marginal impact on the probability to observe a zero value in the dependent variable:

$$\frac{\partial P(y_i = 0)}{\partial x_{ik}} = -\phi(\frac{x_i'\beta}{\sigma})\frac{\beta_k}{\sigma}$$

Marginal impact on the expected value of the dependent variable, conditional on a positive realization:

$$\begin{aligned} E(y_i) &= x_i'\beta\Phi(x_i'\beta/\sigma) + \sigma\phi(x_i'\beta/\sigma) \\ \frac{\partial E(y_i)}{\partial x_{ik}} &= \beta_k\Phi(x_i'\beta/\sigma) \\ \frac{\partial E(y_i^*)}{\partial x_{ik}} &= \beta_k \end{aligned}$$

Specification/Estimation Issues

Violations of the distributional assumptions on ϵ_i (e.g. non-normality and heteroskedasticity) will lead to inconsistent parameter estimations.

Pagan and Vella (1989) propose a moment-based test for **normality**, as for normally distributed errors it should hold that $E(\epsilon^3/\sigma^3 \mid x_i) = 0$ and $E(\epsilon^4/\sigma^4 - 3 \mid x_i) = 0$ (absence of skewness and kurtosis).

Selection Bias: Tobit 2-Model/Heckman 2-Step Selection Model

One can argue that underlying the restriction of a continuous variable y (say: wages) lies a binary outcome h (say: to seek employment or not).

$$y_i^* = x_{1i}'\beta_1 + \epsilon_1$$
$$h_i^* = x_{2i}'\beta_2 + \epsilon_2$$
$$y_i = y*_i, h_i = 1 \quad if \quad h_i^* > 0$$
$$y_i = 0, h_i = 0 \quad if \quad h_i^* \leq 0$$

Selection Bias: Tobit 2-Model/Heckman 2-Step Selection Model 2

Under the assumption that $\epsilon_2 \sim N(0,1) \Rightarrow \sigma_2^2 = 1$:

$$E(w_i \mid h_i = 1) = x'_{1i}\beta_1 + \sigma_{12}\frac{\phi(x'_{2i}\beta_2)}{\Phi(x'_{2i}\beta_2)}$$

$$\sigma_{12} = \rho_{12}\sigma_1$$

$$\rho_{12} = Corr(\epsilon_1, \epsilon_2)$$

Selection Bias: Tobit 2-Model/Heckman 2-Step Selection Model 3

The model can be denoted as a maximum likelihood estimation.

$$
\begin{aligned}
logL_3(\beta, \sigma_1^2, \sigma_{12}) &= \sum_{i \in I_0} logP(h_1 = 0) + \sum_{i \in I_1} [logf(y_i \mid h_1 = 1) + logP(h_i = 1)] \\
&= \sum_{i \in I_0} logP(h_1 = 0) + \sum_{i \in I_1} [logf(y_i) + logP(h_i = 1 \mid y_i)]
\end{aligned}
$$

Selection Bias: Tobit 2-Model/Heckman 2-Step Selection Model 4

Heckman provides a two step estimation technique which is often applied in research.

$$y_i = x'_{1i}\beta_1 + \sigma_{12}\lambda_i + \eta_i$$
$$\lambda_i = \frac{\phi(x'_{2i}\beta_2)}{\Phi(x'_{2i}\beta_2)}$$

The only unknown in λ_i is β_2, which can be estimated in a Tobit routine to be then plugged into a linear regression for the upper equation.

Key Questions

- Which Gauss-Markov assumptions will **always** be violated with binary outcome data?
- What is the relationship between a **link function** and **marginal effects**?
- What is the difference between a **Logit and Probit** link function?
- What is censored data, and why is it a porblem?
- What is the Tobit estimator and what is the importance of the inverse Mill's ratio?
- What is the intuition behind the **Heckman selection model**?