Statistics
Tutorial 10

Philipp Scherer & Jens Wiederspohn

08.07.2020

Disclaimer!

- **The content of the slides partly relies on material of Philipp Prinz, a former Statistics tutor. Like us, he's just a student. Therefore we provide no guarantee for the content of the slides or other data/information of the tutorial.**
- **Please note that the slides will not cover the entire lecture content. To pass the exam, it is still absolutely necessary to deal with the Wooldridge in detail!**

Limited dependent variables

= Dependent variable whose range of values is restricted

- Common cases of restrictions:
  - Price, income,... → positive continuous values
    → Link to a selection of our finest distributions
  - Count data (e.g. votes) → non-negative integers
    → Poisson, negative binomial
  - Probabilities/shares → interval between 0 and 1
    → Logit, probit, (beta)

Logit/Probit

- Useful if we have a **binary dependent variable**
- Binary response models that give $Pr(y = 1|x)$
  - Alternative to LPM, solves problem with unit interval
  - Response probability is no longer linear in β
    - $\rightarrow$ Logit/Probit are non-linear functions
    - $\rightarrow$ We can no longer use OLS to get estimates (GM.1)
- Logit/Probit limit response probability to unit interval $[0, 1]$
  - Put x into a function $G(\beta_0 + x\beta) = G(z)$ to get $Pr(y = 1|x)$
  - For any x_j, $G(z) = Pr(y = 1|x)$ is bound between 0 and 1
- Difference between Logit/Probit: assumed **error distribution**
  1) Logit: Errors follow logistic distribution
  2) Probit: Errors follow standard normal distribution

  $\rightarrow$ Apart from that, they yield the same results
  (Economists tend to use probit, we often use logit)

Logit

- Error is assumed to have a logistic distribution
  $\rightarrow$ Use cumulative logistic function $\Lambda(z)$ for $G(z)$
- $\Lambda(z) = \frac{exp(z)}{1+exp(z)}$ is always in unit interval $[0, 1]$
  - $\Lambda(z) \rightarrow 1$ for $z \rightarrow \infty$,
  - $\Lambda(z) \rightarrow 0$ for $z \rightarrow -\infty$
- Increasing function: greater $z \rightarrow$ larger probability
  $\rightarrow$ Important for interpretation of coefficients!
- Since z stands for our usual regression equation:
  - $\Lambda(z) = \frac{exp(z)}{1+exp(z)} = \frac{exp(\beta_0+\beta_1 x_1+\beta_2 x_2)}{1+exp(\beta_0+\beta_1 x_1+\beta_2 x_2)} = Pr(y=1|x)$

Logit (graphically)

- What is the probability that a borrower will meet the debt obligations, conditional on the average bank balance?
- Logit is strictly between 0 and 1 and has non-linear β

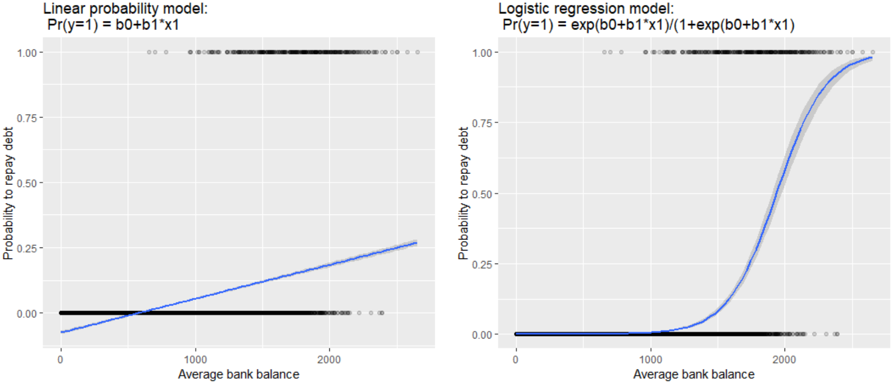


Figure 1: LPM vs. Logit model

Probit

- Error is assumed to have standard normal distribution
  $\rightarrow$ Use standard normal CDF $\Phi(z) = \int_{\infty}^{z} \phi(v)dv$ for $G(z)$
- Recall: $\Phi(z) = Pr(Z \leq z)$
  $\rightarrow$ Probability that random variable Z is smaller/equal than z
  $\rightarrow$ Allows inputs from $-\infty$ to $+\infty$ and assigns value in $[0, 1]$
    - If $z = +\infty$, Z can only be smaller but never larger
    - If $z = -\infty$, Z can never be smaller than z
- Like logit, probit model is increasing and has unit interval
  - The larger z, the larger $\Phi(z) = Pr(y = 1|X)$
  - $\Phi(z) \rightarrow 1$ for $z \rightarrow \infty$,
  - $\Phi(z) \rightarrow 0$ for $z \rightarrow -\infty$
- Again, $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, so $\Phi(z) = \Phi(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$
  - Plug in estimates from probit regression and look up Φ in table

Probit (graphically)

- Shape of standard normal CDF resembles logistic CDF
- Probit is strictly between 0 and 1 and has non-linear β

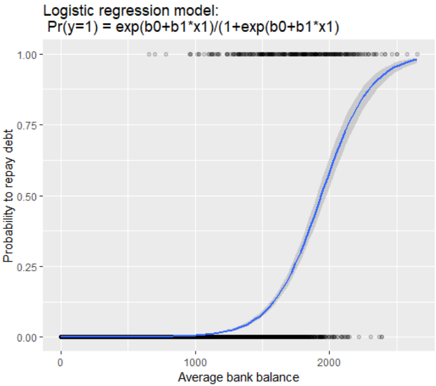
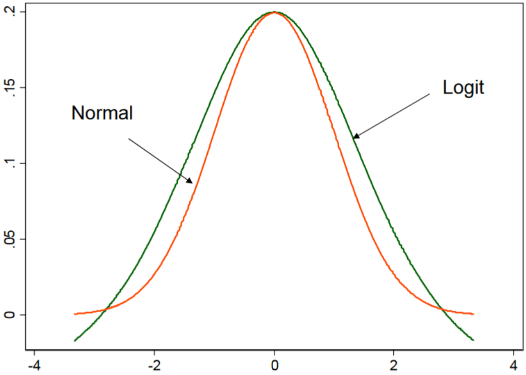


Figure 2: Logit vs. Probit model

Logistic and normal density functions

- Similar shape of standard normal and logistic density function
  - Logit has thinner tails than the normal distribution



Interpretation of logit/probit coefficients

- Estimation of logit/probit coefficients is not straightforward
  - Logit: one unit change in x changes **log-odds** of Y by β
  - Probit: one unit change in x changes **z-score** of Y by β

    $\rightarrow$ We cannot easily interpret magnitude, only sign

- What is the effect of a marginal change in x on $Pr(y = 1|x)$?
  - $\frac{\partial Pr(y=1|x)}{\partial x_j} = \frac{\partial G(\beta_0 + x\beta)}{\partial x_j} = G'(\beta_0 + x\beta) * \beta_j = g(\beta_0 + x\beta) * \beta_j$

    $\rightarrow$ Effect of x_j depends on the other x in $g(\beta_0 + x\beta)$

    $\rightarrow$ $G(z)$ is strictly increasing $\rightarrow$ $g(z)$ (= slope) always positive

    $\rightarrow$ Sign of $g(z) \cdot \beta_j$ depends on β_j

- We can determine effect of a change in x_k if we compute difference of $P(y = 1|x)$ for two given values $c_k + 1$ and c_k

  $\rightarrow$ Caution! Effect is not linear, it is different for each level of x

Logistic and probit regression models

Table 1: Regression of repaying one's debt on average bank balance

Dependent variable:	*Repay debt*	
	logistic	*probit*
Constant	-11.006^{***}	-5.579^{***}
	(0.489)	(0.232)
Bank balance	0.006^{***}	0.003^{***}
	(0.0003)	(0.0001)
Observations	6,047	6,047
Log Likelihood	-454.344	-455.461
Note:	*p<0.1; **p<0.05; ***p<0.01	

→ *What is probability for s.o. with 2000 average bank balance?*

Logistic and probit regression models

Bank balance ()	$\Lambda(z)$	$\Phi(z)$
1000	0.005	0.003
1500	0.076	0.091
2000	0.582	0.531
2500	0.960	0.932
3000	0.998	0.998

Table 2: Probabilites of $y = 1$ for logit and probit models

$\rightarrow$ In logit models, increasing the average bank balance from 1000 to 1500 Euro increases the probability of repaying one's debt from 0.5% to 7.6%.

$\rightarrow$ The effect of a 500 increase depends on x. Going from 2000 to 2500 increases $Pr(y = 1)$ from 58.2% to 96%.

Estimation of logit/probit models $\rightarrow$ MLE

- We cannot estimate the models with OLS
  - $\rightarrow$ Linearity assumption violated
  - $\rightarrow$ Use maximum likelihood estimation (MLE) instead

- Like OLS, MLE is a method for estimating population parameters from sample data
- The ML estimators are those that maximize the likelihood ($\approx$ probability) that our model produces the data that we observe
  - $\rightarrow$ Instead of the probability of the estimates given the data, we look at the **likelihood of the data given the estimates**
  - $\rightarrow$ How likely is it to observe the values y_{ij} if we have the coefficients β_1, β_2,...? Is there any other combination of the β that yield a higher likelihood to observe the data at hand?

- To get the MLE, maximize the log-likelihood function
  - $\rightarrow$ Analytical solutions ($\rightarrow$ RDII) are usually too difficult, so we use numerical methods in statistical software

Exercise 1

Which statistical model should you use to estimate the follwing effects according to Chapter 17 of Wooldridge's book?

Exercise 1a and b

The effect of German citizens' political interest on the amount of campaign contribution, where you have a random sample from all German citizens

- Tobit model
- Campaign contributions have nonnegative outcomes that may pile up at zero but may also take on a broad range of positive values. In such cases, the use of Tobit models is appropriate.

The effect of German citizens' political interest on the amount of campaign contribution, however you have only the survey of citizens who contributed at least 1000 Euro

- Truncated model
- Citizens who contributed less than 1000 Euro have been exluded

→ "A truncated regression model arises when a part of the population is excluded entirely" (Wooldridge p. 621).

Exercise 1c and d

The effect of German citizens' political interest on the amount of campaign contribution, where you have a random sample from all German citizens, however you realized that those with higher income tend not to answer the question about campaign contribution.

- Heckit model

→ "those with higher income tend not to answer the question" indicates a sample selection bias. Heckman's method helps to test and correct such problems.

The effect of German citizens' political interest on whether they turned out to vote at the last federal election

- Logit model or probit model

→ "Whether they turned out.." indicates a binary dependent variable. In such cases, Logit/Probit models seem to be very useful

Exercise 1e and f

The effect of German citizens' political interest on how many times they were exposed to candidate posters at the last federal election

- Poisson model

→ "on how many times..." indicates a count variable (which takes on only takes only nonnegative, integer values). In such cases, a Poisson model is appropriate

The effect of the presence of peace keeping operations on the time until the next armed conflict breaks out

- Duration model, or censored regression model
- "A duration is a variable that measures the time before a certain event occurs." (Wooldridge p. 611)

Exercise 2

You are interested whether and to which extent citizens' discontent with the government affect their vote intention for AfD. To this end, you conducted a survey. In your data, vote intention is coded binary with one if one intends to vote for AfD and zero otherwise. Attitude toward the government is measured with a scale from -5 to +5 where -5 is the most dissatisfied attitude and +5 is the most satisfied one. You use a logit model, which includes only intercept and attitude toward the government as independent variable:

$$\Pr(y = 1|x) = G(\beta_0 + \beta_1 x) \tag{1}$$

Exercise 2a and b

Is it appropriate to use OLS or WLS to estimate the parameters of the above model? Answer with yes or no.

- No
- Because of the nonlinear nature of $E(y|x)$, OLS and WLS are not applicable. We could use nonlinear versions of these methods, but it is no more difficult to use maximum likelihood estimation (MLE).
- Up until now, we have had little need for MLE, although we did note that, under the classical linear model assumptions, the OLS estimator is the maximum likelihood estimator (conditional on the explanatory variables)

Justify your answer above.

- The model is not linear.

Exercise 2c

You interviewed one single person and this person does not intend to vote for AfD and his attitude toward the government was measured as +3 by using the scale above. Given $\hat{\beta}_0 = 1$ and $\hat{\beta}_1 = 0.5$, calculate $\Pr(y = 1|x)$.

- $Pr(y = 1|x) = \frac{exp(z)}{1+exp(z)}$
- $Pr(y = 1|x) = \frac{exp(\beta_0+\beta_1 x_1)}{1+exp(\beta_0+\beta_1 x_1)}$
- $Pr(y = 1|x) = \frac{exp(1+0.5*3)}{1+exp(1+0.5*3)}$
- $Pr(y = 1|x) = 0.9241$

Exercise 2d

Calculate the log-likelihood given this respondent and the parameter values above.

- Given $y_1 = 0$, because the person does not intend to vote for AFD
- $\ell_1(\beta) = y_1 log[G(\beta_0 + \beta_1 x_1)] + (1 - y_1) log[1 - G(\beta_0 + \beta_1 x_1)]$
- $\ell_1(\beta) = 0 * log[0.9241] + (1 - 0) log[1 - 0.9241]$
- $\ell_1(\beta) = log[1 - 0.9241]$
- $\ell_1(\beta) = -2.57889$
- With, as always, *ln*, the natural logarithm, for *log*

Exercise 2e

Suppose you have another respondent who intend to vote for AfD and his attitude toward the government was measured as -2. Calculate the log-likelihood given this respondent and the parameter values above.

- Given $y_2 = 1$, because the person does intend to vote for AFD
- $\ell_2(\beta) = y_2 log[G(\beta_0 + \beta_1 x_1)] + (1 - y_2) log[1 - G(\beta_0 + \beta_1 x_1)]$
- $\ell_2(\beta) = y_2 log[\frac{exp(\beta_0+\beta_1 x_1)}{1+exp(\beta_0+\beta_1 x_1)}] + (1 - y_2) log[1 - (\frac{exp(\beta_0+\beta_1 x_1)}{1+exp(\beta_0+\beta_1 x_1)})]$
- $\ell_2(\beta) = y_2 log[\frac{exp(1+0.5*(-2))}{1+exp(1+0.5*(-2)}] + (1 - y_2) log[1 - (\frac{exp(1+0.5*(-2))}{1+exp(1+0.5*(-2))})]$
- $\ell_2(\beta) = y_2 log[0.5] + (1 - y_2) log[1 - 0.5]$
- $\ell_2(\beta) = 1 * log[0.5] + (1 - 1) log[1 - 0.5]$
- $\ell_2(\beta) = log[0.5]$
- $\ell_2(\beta) = -0.69315$

Exercise 2f

Calculate the log-likelihood given the two respondents above and the parameter values above.

- The log-likelihood for a sample size of n is obtained by summing $\ell_i(\beta) = y_i log[G(\beta_0 + \beta_1 x_1)] + (1 - y_i) log[1 - G(\beta_0 + \beta_1 x_1)]$ across all observations: $\mathcal{L}(\beta) = \sum \ell_i(\beta)$.
- $\mathcal{L}(\beta) = \sum \ell_i(\beta)$
- $\mathcal{L}(\beta) = \ell_1(\beta) + \ell_2(\beta)$
- $\mathcal{L}(\beta) = (-2.57889) + (-0.6931)$
- $\mathcal{L}(\beta) = -3.27204$

Exercise 2g

Calculate the log-likelihood given the two respondents above and $\hat{\beta}_0 = 1$ and $\hat{\beta}_1 = -0.5$.

- $\mathcal{L}(\beta) = \sum \ell_i(\beta)$
- $\mathcal{L}(\beta) = \ell_1(\beta) + \ell_2(\beta)$
- $\mathcal{L}(\beta) = [y_1 log[\frac{exp(\beta_0+\beta_1 x_1)}{1+exp(\beta_0+\beta_1 x_1)}] + (1 - y_1) log[1 - (\frac{exp(\beta_0+\beta_1 x_1)}{1+exp(\beta_0+\beta_1 x_1)})] + [y_2 log[\frac{exp(\beta_0+\beta_1 x_1)}{1+exp(\beta_0+\beta_1 x_1)}] + (1 - y_2) log[1 - (\frac{exp(\beta_0+\beta_1 x_1)}{1+exp(\beta_0+\beta_1 x_1)})]$
- $\mathcal{L}(\beta) = [0 * log[\frac{exp(1-0.5*3)}{1+exp(1-0.5*3)}] + (1 - 0) log[1 - (\frac{exp(1-0.5*3)}{1+exp(1-0.5*3)})] + [1 * log[\frac{exp(1-0.5*(-2))}{1+exp(1-0.5*(-2))}] + (1 - 1) log[1 - (\frac{exp(1-0.5*(-2))}{1+exp(1-0.5*(-2))})]$
- $\mathcal{L}(\beta) = 1 * log[1 - 0.377541] + 1 * log[0.880797]$
- -0.6010056

Exercise 2h

Given $\beta_0 = 1$, which of $\hat{\beta}_1 = 0.5$ and $\hat{\beta}_1 = -0.5$ is the better estimate for β?

- -0.5 is the better estimate for β, because the MLE of β, denoted by $\hat{\beta}$, maximizes the log-likelihood and the log-likelihood is bigger with $\hat{\beta}_1 = -0.5$.

Exercise 2i and j

What is the name of the function G in the model above?

- The logistic function
- In the logit model, G is the logistic function: $G(z) = \frac{exp(z)}{1+exp(z)} = \Lambda(z)$

If you replace the function G with the standard normal cumulative distribution function, what model can you obtain?

- Probit model
- In the probit model, G is the standard normal cumulative distribution function (cdf), which is expressed as an integral:
  $G(z) = \Phi(z) = \int_{\infty}^{z} \phi(v)dv$

Exercise 2k

What test corresponds to the F-test which compares a restricted and an unrestricted linear regression model?

- Likelihood ratio (LR) test
- The LR test is based on the same concept as the F test in a linear model.
- The F test measures the increase in the sum of squared residuals when variables are dropped from the model. The LR test is based on the difference in the log-likelihood functions for the unrestricted and restricted models.
- The likelihood ratio statistic is twice the difference in the log-likelihoods: $LR = 2(\mathcal{L}_{ur} + \mathcal{L}_r)$

Exercise 2l

Which distribution does the test statistic for the above test have?

- Chi-square distribution
- The multiplication by two in $LR = 2(\mathcal{L}_{ur} + \mathcal{L}_r)$ is needed so that LR has an approximate chi-square distribution under H_0.