Econometrics II: Econometric Modelling

Jürgen Meinecke

Research School of Economics, Australian National University

12 October, 2018

Assignment 2

Deadline: 17 October at 12:00pm (noon, mid-day)

Reminder: my deadlines are *very* sharp!
(If you submit at 12:01pm you will receive a mark of zero!)

No extensions given under any circumstances!

Note: I do not offer any help on solving the assignment!

Final Exam

Scheduled for 13 November

After week 12 I will offer extra office hours

Will make specific announcement in last lecture

Roadmap

When the independent variable Y_i is binary, we estimate a model for the probability of success conditional on X_i: $\Pr(Y_i = 1|X_i)$

We have looked at three alternative ways of modeling this:

$$\Pr(Y_i = 1|X_i) = \begin{cases} \beta_0 + \beta_1 X_i & \text{LPM} \\ \Phi(\beta_0 + \beta_1 X_i) & \text{Probit} \\ F(\beta_0 + \beta_1 X_i) & \text{Logit} \end{cases}$$

Recall that

- $\Phi(\cdot)$ is the cdf of the standard normal distribution
- $F(\cdot)$ is the cdf of the logistic distribution
- $\Phi(\cdot)$ and $F(\cdot)$ look similar; they both produce an "S"-shaped look

Roadmap

How should we define marginal effects in these models?

Definition

In models with binary outcome variables, the **marginal effect** of the explanatory variable X_i at x is given by

$$\Pr(Y_i = 1 | X_i = x + 1) - \Pr(Y_i = 1 | X_i = x).$$

(If X_i is a continuous random variable, then the marginal effect would be defined as the partial derivative with respect to X_i)

Important: the marginal effect is the effect of changing X on the *probability* that Y is equal to one

Here x is a particular value of X_i

Does the marginal effect vary in x? Let's take a closer look…

Let's calculate the marginal effect at x for each of the three models:

$$\Pr(Y_i = 1|X_i = x + 1) - \Pr(Y_i = 1|X_i = x)$$
$$= \begin{cases} \beta_0 + \beta_1(x+1) - (\beta_0 + \beta_1 x) = \beta_1 & \text{LPM} \\ \Phi(\beta_0 + \beta_1(x+1)) - \Phi(\beta_0 + \beta_1 x) & \text{Probit} \\ F(\beta_0 + \beta_1(x+1)) - F(\beta_0 + \beta_1 x) & \text{Logit} \end{cases}$$

For the LPM, the marginal effect collapses to β_1

For probit and logit this does not happen because $\Phi(\cdot)$ and $F(\cdot)$ are not linear functions (for example, $\Phi(2+1) - \Phi(2) \neq \Phi(1)$)

Going back to our question from two slides earlier "Does the marginal effect vary in x?", the marginal effect does

- not vary in x for the LPM
- vary in x for the probit and logit models

The nonlinearity of probit and logit models makes the estimation of marginal effects a lot more complicated

The effect of changing X by one unit on the probability that $Y = 1$ is different for every value of X

Therefore, for every value of X the marginal effect is different!

So far, I have just presented the case of one explanatory variable

But in general, there could be several explanatory variables $X_1, \ldots, X_k$

In the probit and logit models, the marginal effects will then be functions of $X_1 = x_1$ up to $X_k = x_k$

There will be countless possibilities

To make this tractable in practice, you would only focus on a small subset of possible values that $X_1, \ldots, X_k$ can take on

LPM example:

$$\begin{aligned}\Pr(\text{Employed}_i = 1|\text{Male}_i, \text{Educ}_i, \text{Age}_i) \\ = \beta_0 + \beta_1\text{Male}_i + \beta_2\text{Educ}_i + \beta_3\text{Age}_i\end{aligned}$$

The marginal effect of increasing age from 20 to 30 is equal to

$$\begin{aligned} & \Pr(\text{Employed}_i = 1|\text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 30) \\ & - \Pr(\text{Employed}_i = 1|\text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 20) \\ = \; & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot 30 \\ & -(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot 20) \\ = \; & 10\beta_3\end{aligned}$$

- That marginal effect is the same for both men and women
- It is also the same for a high-school graduate and a college graduate
- It is the same irrespective of the values of x_1 and x_2

The marginal effect of increasing age from 30 to 40 is equal to

$$\begin{aligned} & \Pr(\text{Employed}_i = 1 | \text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 40) \\ & - \Pr(\text{Employed}_i = 1 | \text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 30) \\ = \quad & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot 40 \\ & -(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot 30) \\ = \quad & 10\beta_3 \end{aligned}$$

It is the same as changing age from 20 to 30

Bottom line: marginal effect constant (because of linearity)

Probit example (logit behaves the same):

$$\Pr(\text{Employed}_i = 1|\text{Male}_i, \text{Educ}_i, \text{Age}_i)$$
$$= \Phi(\beta_0 + \beta_1 \text{Male}_i + \beta_2 \text{Educ}_i + \beta_3 \text{Age}_i)$$

The marginal effect of increasing age from 20 to 30 is equal to

$$\begin{aligned} & \Pr(\text{Employed}_i = 1|\text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 30) \\ & - \Pr(\text{Employed}_i = 1|\text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 20) \\ = \; & \Phi(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot 30) \\ & -\Phi(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot 20) \\ \neq \; & 10\beta_3 \qquad \text{b/c of non-linearity} \end{aligned}$$

- That marginal effect is not the same for men and women, high-school graduates and college graduates
- It does depend on specific values we plug in for x_1 and x_2

The marginal effect of increasing age from 30 to 40 is equal to

$$
\begin{aligned}
& \Pr(\text{Employed}_i = 1 | \text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 40) \\
& - \Pr(\text{Employed}_i = 1 | \text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 30) \\
= \quad & \Phi(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot 40) \\
& - \Phi(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot 30) \\
\neq \quad & \Pr(\text{Employed}_i = 1 | \text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 30) \\
& - \Pr(\text{Employed}_i = 1 | \text{Male}_i = x_1, \text{Educ}_i = x_2, \text{Age}_i = 20)
\end{aligned}
$$

The two marginal effects are not the same

Bottom line: marginal effect not constant (bc of non-linearity)

Roadmap

Generic estimation problem:

- we have some data $Y_1, \ldots, Y_n$
  (let's forget about $X_1, \ldots, X_n$ for a few slides)
- these data are thought of as being generated by some unknown random process
- we only know one thing about that random process: the distribution $D(Y_i|\theta)$ which generated $Y_1, \ldots, Y_n$
  (more precisely: we *assume* that we know the distribution)
- the distribution $D(\cdot|\theta)$ is only known up to parameter θ
  (note: θ can be a container for several parameters)
- we do not know θ, our goal is to estimate it

One way to estimate θ is to use OLS, another entirely new way is to use maximum likelihood estimation (MLE)

Definition

The **likelihood function** is the joint probability (or density) of the data $Y_1, \ldots, Y_n$ treated as a function of the parameters θ. The likelihood function is denoted $L(\theta | Y_1, \ldots, Y_n)$.

We read $L(\theta | Y_1, \ldots Y_n)$ as a function in the unknown parameter θ given the data $Y_1, \ldots, Y_n$

To obtain an estimator for the unknown parameter θ we only need to maximize the likelihood function…

Definition

The **maximum likelihood estimator** is the value of the parameter that maximizes the likelihood function:

$$\hat{\theta}_{MLE} := \text{argmax}_{\theta} L(\theta|Y_1, \ldots, Y_n)$$

The maximum likelihood estimator is the value of θ that maximizes the likelihood that the particular sample data available to us occurs

Let's say Y_i has the following distribution (for $i = 1, \dots, n$):

$$Y_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases}$$

(Say: student i's chance of scoring an HD in EMET3004)

(as you surely remember from STAT1008, this is called a Bernoulli distribution with probability of success p)

The data $Y_1, \dots, Y_n$ are i.i.d.

Mapping this example into the terminology from two slides earlier: the distribution $D(\cdot|\theta)$ can be dissected like:

- D is a placeholder for the Bernoulli distribution
- θ is a placeholder for p

Let's derive the likelhood function

We start with the likelihood for Y_1:

$$\Pr(Y_1 = 1) = p$$
$$\text{and } \Pr(Y_1 = 0) = 1 - p$$

therefore

$$\Pr(Y_1 = y_1) = p^{y_1}(1-p)^{1-y_1}, \qquad \text{for } y_1 \in \{0,1\}$$

The last result is just a cute way of compressing the likelihood into one single line

So far, this is merely the *marginal* likelihood for person i

We need to derive the *joint* likelihood for all persons…

Let's add person 2 now

We are savvy and exploit independence between persons 1 and 2; joint likelihood becomes

$$\begin{aligned}\Pr(Y_1 = y_1, Y_2 = y_2) &= \Pr(Y_1 = y_1) \times \Pr(Y_2 = y_2) \\ &= \left[p^{y_1}(1-p)^{1-y_1}\right] \times \left[p^{y_2}(1-p)^{1-y_2}\right] \\ &= p^{y_1+y_2}(1-p)^{2-(y_1+y_2)}\end{aligned}$$

Escalating further, we get joint likelihood for all persons:

$$\begin{aligned}&\Pr(Y_1 = y_1, \ldots, Y_n = y_n) \\ &\qquad = \Pr(Y_1 = y_1) \times \ldots \times \Pr(Y_n = y_n) \\ &\qquad = \left[p^{y_1}(1-p)^{1-y_1}\right] \times \ldots \times \left[p^{y_n}(1-p)^{1-y_n}\right] \\ &\qquad = p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i}\end{aligned}$$

We have just derived the likelihood function:

$$L(p|Y_1, \ldots, Y_n) = p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i}$$

To get the maximum likelihood estimator for p we now need to maximize $L(p|Y_1, \ldots, Y_n)$

To do that, it is often easier to work with the logarithm of L (this is called the log-likelihood, denoted $\ln L$)

Because the logarithm is a monotone transformation, the maximizer of $\ln L$ coincides with the maximizer of L

We can therefore re-define the maximum likelihood estimator to be the argmax of the log-likelihood

In our example, because

$$L(p|Y_1,\ldots,Y_n) = p^{\sum_{i=1}^{n} y_i}(1-p)^{n-\sum_{i=1}^{n} y_i}$$

$$\Rightarrow \ln L(p|Y_1,\ldots,Y_n) = \left[\sum_{i=1}^{n} y_i\right] \ln p + \left[n - \sum_{i=1}^{n} y_i\right] \ln(1-p)$$

Obtaining the first derivative of $\ln L$

$$\frac{\partial \ln L(p|Y_1,\ldots,Y_n)}{\partial p} = \left[\sum_{i=1}^{n} y_i\right] \frac{1}{p} + \left[n - \sum_{i=1}^{n} y_i\right] \frac{-1}{1-p}$$

Setting equal zero results in the MLE

$$0 = \left[\sum_{i=1}^{n} y_i\right] \frac{1}{\hat{p}_{MLE}} + \left[n - \sum_{i=1}^{n} y_i\right] \frac{-1}{1-\hat{p}_{MLE}}$$

Cleaning up, you will find that…

$$\hat{p}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} y_i =: \bar{Y}$$

In other words: when Y_i is i.i.d. Bernoulli, the MLE is equal to the sample average

This summarizes the MLE primer

Probit and logit models are also estimated using the MLE approach

This will be a little bit more complicated than the Bernoulli example presented here

But it is still not all too difficult

Roadmap

I will illustrate the main idea using the probit model

Now we also have one explanatory variable X_i
(generalizing to k explanatory variables would not be difficult)

$$\Pr(Y_i = 1|X_i) = \Phi(\beta_0 + \beta_1 X_i)$$

How can we estimate β_0 and β_1?

What is the sampling distribution of the estimators?

How can we construct standard errors?

Answer: use maximum likelihood estimation

In analogy to the Bernoulli example earlier, we first derive the likelihood for person 1

Given the explanatory variable X_1 that likelihood is equal to

$$\Pr(Y_1 = 1|X_1) = \Phi(\beta_0 + \beta_1 X_1)$$

$$\Pr(Y_1 = 0|X_1) = 1 - \Phi(\beta_0 + \beta_1 X_1)$$

combining in one line results in

$$\Pr(Y_1 = y_1|X_1) = \Phi(\beta_0 + \beta_1 X_1)^{y_1} \left[1 - \Phi(\beta_0 + \beta_1 X_1)\right]^{1-y_1}$$

We read this as a function in the two parameters β_0 and β_1

The probit likelihood function is the joint density of $Y_1, \ldots, Y_n$ given $X_1, \ldots, X_n$, treated as a function of β_0, β_1

Let's add person 2 now

Again we are exploiting independence between persons

$$
\begin{aligned}
\Pr(Y_1 = y_1, Y_2 = y_2 | X_1, X_2) \\
&= \Pr(Y_1 = y_1 | X_1) \times \Pr(Y_2 = y_2 | X_2) \\
&= \Phi(\beta_0 + \beta_1 X_1)^{y_1} [1 - \Phi(\beta_0 + \beta_1 X_1)]^{1-y_1} \times \\
&\Phi(\beta_0 + \beta_1 X_2)^{y_2} [1 - \Phi(\beta_0 + \beta_1 X_2)]^{1-y_2}
\end{aligned}
$$

Escalating further, we get joint likelihood for all persons

$$
\begin{aligned}
\Pr(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n | X_1, X_2, \ldots, X_n) \\
=& \Phi(\beta_0 + \beta_1 X_1)^{y_1} [1 - \Phi(\beta_0 + \beta_1 X_1)]^{1-y_1} \times \\
& \Phi(\beta_0 + \beta_1 X_2)^{y_2} [1 - \Phi(\beta_0 + \beta_1 X_2)]^{1-y_2} \times \\
& \quad \vdots \qquad\qquad\qquad \vdots \\
& \Phi(\beta_0 + \beta_1 X_n)^{y_n} [1 - \Phi(\beta_0 + \beta_1 X_n)]^{1-y_n}
\end{aligned}
$$

We have just derived the likelihood function:

$$
\begin{aligned}
&L(\beta_0, \beta_1 | Y_1, \ldots, Y_n, X_1, \ldots, X_n) = \\
&\qquad = \Phi(\beta_0 + \beta_1 X_1)^{y_1} [1 - \Phi(\beta_0 + \beta_1 X_1)]^{1-y_1} \times \\
&\qquad\quad \Phi(\beta_0 + \beta_1 X_2)^{y_2} [1 - \Phi(\beta_0 + \beta_1 X_2)]^{1-y_2} \times \\
&\qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
&\qquad\quad \Phi(\beta_0 + \beta_1 X_n)^{y_n} [1 - \Phi(\beta_0 + \beta_1 X_n)]^{1-y_n}
\end{aligned}
$$

To get the maximum likelihood estimator for β_0, β_1 we now need to maximize $L(\beta_0, \beta_1 | Y_1, \ldots, Y_n, X_1, \ldots, X_n)$

Again use log-likelihood

$\hat{\beta}_0$ and $\hat{\beta}_1$ maximize this likelihood function

But we can't solve for the maximum explicitly!
Must be maximized using numerical methods

Luckily, Stata will solve this for us

Maximum Likelihood Estimation has some desirable theoretical properties

In large samples $\hat{\beta}_0$ and $\hat{\beta}_1$ are

- consistent
- normally distributed
- efficient

(assuming the probit model is actually the correct model)

Computation of standard errors is also done by Stata
(the actual derivation goes beyond EMET3004)

Once you accept Stata's standard errors, testing and confidence interval construction proceed as ususal

Roadmap

- Mortgages (home loans) are an essential part of buying a home
- Is there differential access to home loans by race?
- If two otherwise identical individuals, one white and one black, applied for a home loan, is there a difference in the probability of denial?

HMDA Data Set

- Data on individual characteristics, property characteristics, and loan denial/acceptance
- The mortgage application process circa 1990-1991:
  - go to a bank or mortgage company
  - fill out an application (personal and financial info)
  - meet with the loan officer
- Then the loan officer decides, by law, in a race-blind way
- Presumably, the bank wants to make profitable loans, and (if the incentives inside the bank or loan origination office are right - a big if during the mid-2000s housing bubble!) the loan officer doesn't want to originate defaults

Loan Officer's Decision

- uses key financial variables:
  - P/I ratio
  - housing expense-to-income ratio
  - loan-to-value ratio
  - personal credit history
- decision rule is nonlinear:
  - loan-to-value ratio $> 80\%$
  - loan-to-value ratio $> 95\%$ (what happens in default?)
  - credit score

We will study: $\Pr(deny = 1|black, other\ X_i's)$

Compare the three models

- linear probability model
- probit
- logit

Main problem with last week's results: omitted variable bias

The following variables enter the loan officer decision and may be correlated with race:

- wealth, type of employment
- credit history
- family status

Fortunately, the HMDA data set is very rich...

TABLE 11.1 Variables Included in Regression Models of Mortgage Decisions

Variable	Definition	Sample Average
Financial Variables		
P/I ratio	Ratio of total monthly debt payments to total monthly income	0.331
housing expense-to-income ratio	Ratio of monthly housing expenses to total monthly income	0.255
loan-to-value ratio	Ratio of size of loan to assessed value of property	0.738
consumer credit score	1 if no "slow" payments or delinquencies 2 if one or two slow payments or delinquencies 3 if more than two slow payments 4 if insufficient credit history for determination 5 if delinquent credit history with payments 60 days overdue 6 if delinquent credit history with payments 90 days overdue	2.1
mortgage credit score	1 if no late mortgage payments 2 if no mortgage payment history 3 if one or two late mortgage payments 4 if more than two late mortgage payments	1.7
public bad credit record	1 if any public record of credit problems (bankruptcy, charge-offs, collection actions) 0 otherwise	0.074
Additional Applicant Characteristics		

denied mortgage insurance	1 if applicant applied for mortgage insurance and was denied, 0 otherwise	0.020
self-employed	1 if self-employed, 0 otherwise	0.116
single	1 if applicant reported being single, 0 otherwise	0.393
high school diploma	1 if applicant graduated from high school, 0 otherwise	0.984
unemployment rate	1989 Massachusetts unemployment rate in the applicant's industry	3.8
condominium	1 if unit is a condominium, 0 otherwise	0.288
black	1 if applicant is black, 0 if white	0.142
deny	1 if mortgage application denied, 0 otherwise	0.120

TABLE 11.2 Mortgage Denial Regressions Using the Boston HMDA Data

Dependent variable: *deny* = 1 if mortgage application is denied, = 0 if accepted; 2380 observations.

Regression Model Regressor	LPM (1)	Logit (2)	Probit (3)	Probit (4)	Probit (5)	Probit (6)
black	0.084** (0.023)	0.688** (0.182)	0.389** (0.098)	0.371** (0.099)	0.363** (0.100)	0.246 (0.448)
P/I ratio	0.449** (0.114)	4.76** (1.33)	2.44** (0.61)	2.46** (0.60)	2.62** (0.61)	2.57** (0.66)
housing expense-to-income ratio	−0.048 (.110)	−0.11 (1.29)	−0.18 (0.68)	−0.30 (0.68)	−0.50 (0.70)	−0.54 (0.74)
medium loan-to-value ratio (0.80 ≤ *loan-value ratio* ≤ 0.95)	0.031* (0.013)	0.46** (0.16)	0.21** (0.08)	0.22** (0.08)	0.22** (0.08)	0.22** (0.08)
high loan-to-value ratio (*loan-value ratio* > 0.95)	0.189** (0.050)	1.49** (0.32)	0.79** (0.18)	0.79** (0.18)	0.84** (0.18)	0.79** (0.18)
consumer credit score	0.031** (0.005)	0.29** (0.04)	0.15** (0.02)	0.16** (0.02)	0.34** (0.11)	0.16** (0.02)
mortgage credit score	0.021 (0.011)	0.28* (0.14)	0.15* (0.07)	0.11 (0.08)	0.16 (0.10)	0.11 (0.08)
public bad credit record	0.197** (0.035)	1.23** (0.20)	0.70** (0.12)	0.70** (0.12)	0.72** (0.12)	0.70** (0.12)
denied mortgage insurance	0.702** (0.045)	4.55** (0.57)	2.56** (0.30)	2.59** (0.29)	2.59** (0.30)	2.59** (0.29)

self-employed	0.060** (0.021)	0.67** (0.21)	0.36** (0.11)	0.35** (0.11)	0.34** (0.11)	0.35** (0.11)
single				0.23** (0.08)	0.23** (0.08)	0.23** (0.08)
high school diploma				−0.61** (0.23)	−0.60* (0.24)	−0.62** (0.23)
unemployment rate				0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
condominium					−0.05 (0.09)	
black × P/I ratio						−0.58 (1.47)
black × housing expense-to-income ratio						1.23 (1.69)
additional credit rating indicator variables	no	no	no	no	yes	no
constant	−0.183** (0.028)	−5.71** (0.48)	−3.04** (0.23)	−2.57** (0.34)	−2.90** (0.39)	−2.54** (0.35)

(continued)

(*Table 11.2 continued*)

F-Statistics and p-Values Testing Exclusion of Groups of Variables

	(1)	(2)	(3)	(4)	(5)	(6)
applicant single; high school diploma; industry unemployment rate				5.85 (< 0.001)	5.22 (0.001)	5.79 (< 0.001)
additional credit rating indicator variables					1.22 (0.291)	
race interactions and black						4.96 (0.002)
race interactions only						0.27 (0.766)
difference in predicted probability of denial, white vs. black (percentage points)	8.4%	6.0%	7.1%	6.6%	6.3%	6.5%

These regressions were estimated using the $n = 2380$ observations in the Boston HMDA data set described in Appendix 11.1. The linear probability model was estimated by OLS, and probit and logit regressions were estimated by maximum likelihood. Standard errors are given in parentheses under the coefficients, and p-values are given in parentheses under the F-statistics. The change in predicted probability in the final row was computed for a hypothetical applicant whose values of the regressors, other than race, equal the sample mean. Individual coefficients are statistically significant at the *5% or **1% level.

Summary of HDMA Results

- Coefficients on the financial variables make sense
- Black is statistically significant in all specifications
- Race-financial variable interactions are not significant
- Including the covariates sharply reduces the effect of race on denial probability
- LPM, probit, logit: similar estimates of effect of race on the probability of denial
- Estimated effects are large in a “real world” sense