Lecture 8: Experiments, LPM, Logit and Probit

Prof. dr. Wolter Hassink
Utrecht University School of Economics
w.h.j.hassink@uu.nl

Contents:

- Experimental design (not in Wooldridge)
- Regression discontinuity (not in Wooldridge)
- Difference-in-differences estimator
- Binary choice models
- Linear Probability Model
- Logit and Probit – motivation
- The logit model
- The Probit model

Material:

Section 13.2
Chapter 7: 7.5
Chapter 17: 17.1

Experimental design

Motivation - Example 1: Performance pay

Claim (Lazear): Work incentives may affect labour productivity.

Mechanism: Performance pay by firms: leads to

a) higher productivity (may lead to more motivated workers) and
b) selection effects (may attract more productive workers)

Kim, H., Kim, S., & Kim, T. (2017). The selection and causal effects of work incentives on labor productivity: Evidence from a two-stage randomized controlled trial in Malawi.

Incentives are essential to promote labor productivity. We implemented a two-stage field experiment to measure effects of career and wage incentives on productivity through self-selection and causal effect channels. First, workers were hired with either career or wage incentives. After employment, a random half of workers with career incentives received wage incentives and a random half of workers with wage incentives received career incentives. **We find that career incentives attract higher-performing workers than wage incentives but do not increase productivity for existing workers. Instead, wage incentives increase productivity for existing workers**. Observable characteristics are limited in explaining the selection effect.

Motivation - Example 2: Subsidies to attract multinational corporations

Chor, D. (2009). Subsidies for FDI: Implications from a model with heterogeneous firms. *Journal of International Economics*, *78*(1), 113-125.

This paper analyzes the welfare effects of subsidies to attract multinational corporations when firms are heterogeneous in their productivity levels. I show that the use of a small subsidy raises welfare in the FDI host country, with the consumption gains from attracting more multinationals exceeding the direct cost of funding the subsidy program through a tax on labor income. **This welfare gain stems from a selection effect, whereby the subsidy induces only the most productive exporters to switch to servicing the host's market via FDI.** I further show that for the same total subsidy bill, a subsidy to variable costs delivers a larger welfare gain than a subsidy to the fixed cost of conducting FDI, since a variable cost subsidy also raises the inefficiently low output levels stemming from each firm's markup pricing power.

Experimental design - motivation

Empirical question on government programme evaluation: What is the **causal effect** of a subsidy of innovation on the profit of a firm?

- **Causal effect:** profit if the firm receives a subsidy minus the profit if the firm receives no subsidy.

- **Fundamental problem 1** of causal inference: we do not have a **good comparison group** of firms that did not receive a subsidy. We often have no information on firms that applied but were declined to receive a subsidy.

- **Fundamental problem 2** of causal inference boils down to a **"what if" question**. We do not observe the **potential outcome** for the subsidized firm as if the firm received no subsidy, which did not take place. Not receiving a subsidy is a **counterfactual outcome** for the firm.

Experimental design – a formal framework (I)

- We start with an assignment mechanism. Firms may YES/NO receive a subsidy by the government. Let's assume that the subsidy is provided to the firms after random assignment.

- The variable D is a "yes-or-no treatment" dummy variable
  $D_i = 1$ if firm i is **assigned** to receive a subsidy (**treatment**)
  $D_i = 0$ if firm i is assigned not to receive a subsidy (no treatment/**control**).

- Let's assume that the outcome variable Y is a binary variable. It can take two values:
  $Y = 1$: firm makes a profit
  $Y = 0$: firm makes no profit

- Without any subscript, Y refers to the **observed outcome** "profit/no profit".

- Let's add a subscript to Y. $Y(D_i)$ is the **potential outcome** according to the treatment.

- The subscript 1 refers to $D_i = 1$:
  Y_{1i}: potential outcome if firm i is assigned to receive a treatment "subsidy". Hence, the potential outcome may be either profit or no profit.
  The subscript 0 refers to $D_i = 0$:
  Y_{0i}: potential outcome if firm i is assigned to receive no treatment "no subsidy". Hence, the potential outcome may be either profit or no profit.

- **Issue:** we cannot observe both potential outcomes for the same individual firm. The unobserved potential outcome: **counterfactual**.

- The observed outcome for each firm can take **two potential outcomes**:

$$\underset{\text{observed outcome}}{Y_i} = \underset{\text{subsidy}}{D_i Y_{1i}} + \underbrace{(1-D_i)Y_{0i}}_{\text{no subsidy}} \qquad (1)$$

- The **causal effect** for individual/firm *i* is the difference of the two potential outcomes. The potential outcome of the firm if the firm receives a treatment minus the potential outcome if the firm receives no treatment.

$$\delta = Y_{1i} - Y_{0i}$$

The difference cannot be observed, because the counterfactual cannot be observed by definition. Furthermore, the causal effect is different (or heterogeneous) across firms.

There are the following possibilities:

1) If $Y_{1i} = Y_{0i} = 1$ $\delta = 0$ Firm will make profit, both for subsidy and without subsidy: no effect

2) If $Y_{1i} = Y_{0i} = 0$ $\delta = 0$ Firm will make no profit, both for subsidy and without subsidy: no effect

3) If $Y_{1i} = 1; Y_{0i} = 0$ then $\delta = 1$ Firm will make profit due to subsidy: positive effect of subsidy

4) If $Y_{1i} = 0; Y_{0i} = 1$ then $\delta = -1$ Firm will make loss due to subsidy: negative effect of subsidy

- The causal effect can differ across individual firms. **The average causal effect**:

$$E(\delta) = E(Y_{1i} - Y_{0i})$$

- However: it is impossible to observe for the same individual; *i* the values $D_i = 1$ and $D_i = 0$ as well as the values Y_{1i} and Y_{0i} and, therefore, it is **impossible to observe the causal effect** of *D* on *Y* for individual *i* (Holland, 1986). In other words: we have **no counterfactual evidence**.

		Potentional outcome	
		Y_{1i}	Y_{0i}
Subsidy to firm	**Treatment:** $D_i = 1$ YES	observable	counterfactual cannot be observed
	Control $D_i = 0$ NO	counterfactual cannot be observed	observable

Role of selection effects

- **Average causal effect:** are the profits for the firms that received a subsidy **on average** higher?

$$E\left(\delta\right)=E\left(Y_{1i}-Y_{0i}\right)$$

- Or is there any **selection effect**? Firms do not receive a subsidy, because of selectivity:
  - they are better organized, and they don't need it anyway (negative bias)
  - they are less motivated to chase for subsidies (positive bias)

- Solution to get rid of selection effect: **randomization** of firms that receive a subsidy.
- If there is random assignment into treatment and control (thus, no selection effects), the Average Treatment Effect is equal to the average causal effect.

$$\underbrace{E\left(Profit_{1i}\mid D_i=1\right)-E\left(Profit_{0i}\mid D_i=0\right)}_{\text{Observed difference in profit, the Average Treatment Effect}}=E\left(Profit_{1i}\right)-E\left(Profit_{0i}\right)$$

$$\underbrace{E\left(Y_{1i}\mid D_i=1\right)-E\left(Y_{0i}\mid D_i=0\right)}_{\text{Average Treatment Effect}}=E\left(Y_{1i}\right)-E\left(Y_{0i}\right)$$

Our conclusion is that we require the following two assumptions:

Assumption 1: firms in the treatment group must be on average identical to firms in the control group with respect to their potential outcomes. The individual assignment to treatment and control is random. It does not depend on the potential outcomes.

Assumption 2: Stable Unit Treatment Value Assumption (SUTVA, Rubin). The potential outcome for any firm does not depend on the number of firms (or participants) in the intervention. There are no hidden variations of treatments across firms. No interference: treatment (subsidy) of one firm depends on treatment (subsidy) of another firm.

First-best method: social experiment

Rule of assignment: the researcher randomly assigns the participants of the treatment group and the control group, under the assumptions 1 and 2 of above.

If D_i is randomly assigned, the selection bias will disappear.

Example:
The average treatment effect:

$$\left(\overline{Y} \mid D=1\right)-\left(\overline{Y} \mid D=0\right)$$

or

$$Y_i = \delta_0 + \delta_1 D_i + u_i$$

Inclusion of additional explanatory variables *X* is irrelevant if there is random assignment

$$Y_i = \delta_0 + \delta_1 D_i + \beta X_i + u_i$$

How to measure a causal effect?
Alternatives to social experiment: a ranking of the statistical methods

- We often cannot apply social experiments.
    - It may be **unethical** to perform an experiment: REBO faculty has an ethical review committee (chaired by Jacob Jordaan)
    - Performing the experiment is too costly or too time consuming
- What if no social experiment is possible?

The following four approaches are based on a quasi-experimental design. The estimators are can be considered as second-best approaches. They all deliver estimates of causal effects.

1. Regression discontinuity (explained below)
2. Differences-in-differences (explained below and Wooldridge, Section 13.2)
3. Instrumental variables/ 2SLS (see weeks 6 and 7 of this course)
4. Propensity score matching (no part of material of this course)

   There are the following alternative methods (again ranking is in terms of relevance)

5. Panel data (weeks 4 and 5 of this course)
6. Regression analysis based on cross-section (week 1 of this course)
7. Survey data/ qualitative research methods (not in this course)

Regression discontinuity

Regression discontinuity – Example

Gehrsitz, M. (2017). Speeding, punishment, and recidivism: Evidence from a regression discontinuity design. *The Journal of Law and Economics*, *60*(3), 497-528.

This paper estimates the effects of temporary driver's license suspensions on driving behavior. A little known rule in the German traffic penalty catalogue maintains that drivers who commit a series of speeding transgressions within 365 days should have their license suspended for one month. **My regression discontinuity design exploits the quasi-random assignment of license suspensions caused by the 365-days cut-off and shows that 1-month license suspensions lower the probability of recidivating within a year by 20 percent.** This is largely a specific deterrence effect driven by the punishment itself and not by incapacitation, information asymmetries, or the threat of stiffer future penalties.

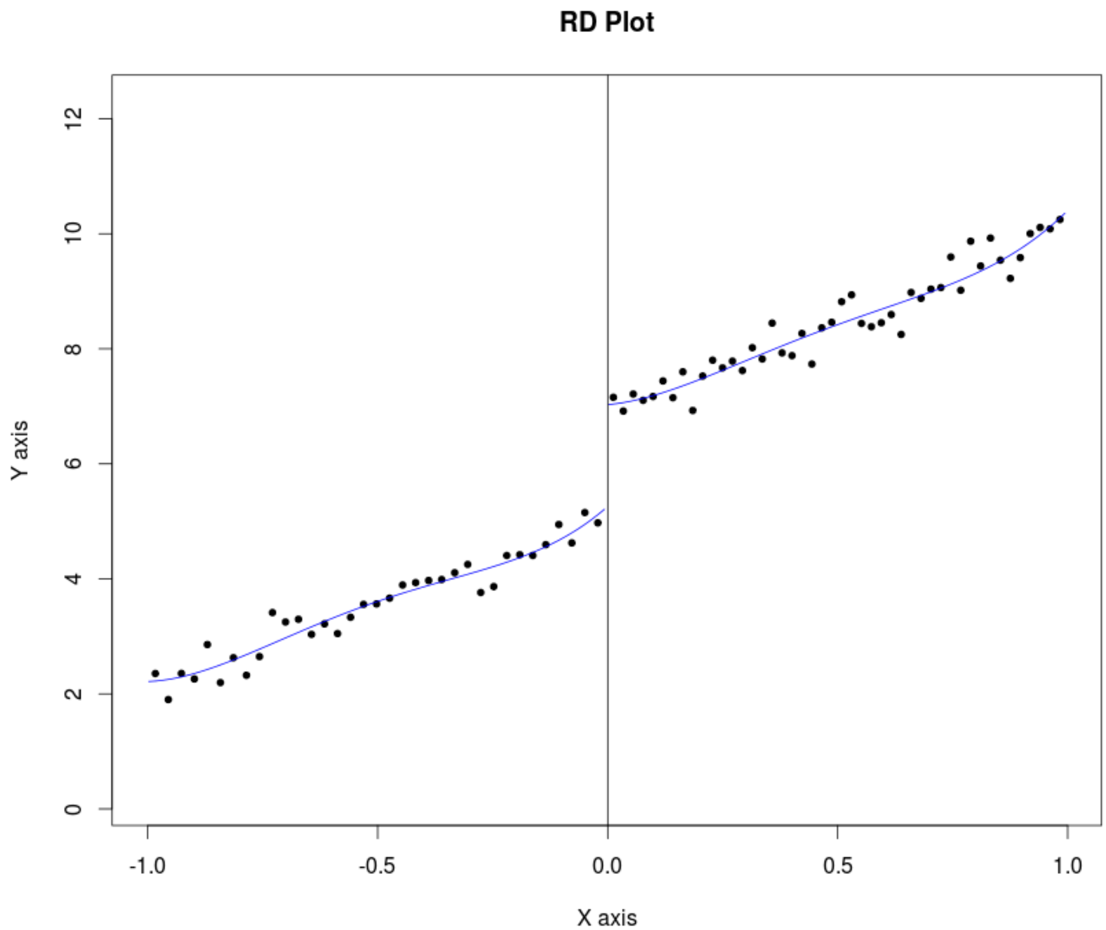
Statistical Method: Regression Discontinuity

- **Example:** Let's assume that there is a government subsidy promoting innovation for firms. The innovation subsidy is limited to firms that have at least 50 employees.

- Question: is there any difference in the outcome variable between firms just above and below the threshold (discontinuity)?

- Regression discontinuity estimator (RD-estimator) is based on the following idea: the difference in profits of e.g. firms with 51 employees and 49 employees can fully be attributed to the discontinuity. There are no other explanatory factors that may explain a difference after period 2, because both groups of firms are very similar.

- The RD-estimator compares the averages in dependent variable y in a small region either side of the discontinuity.



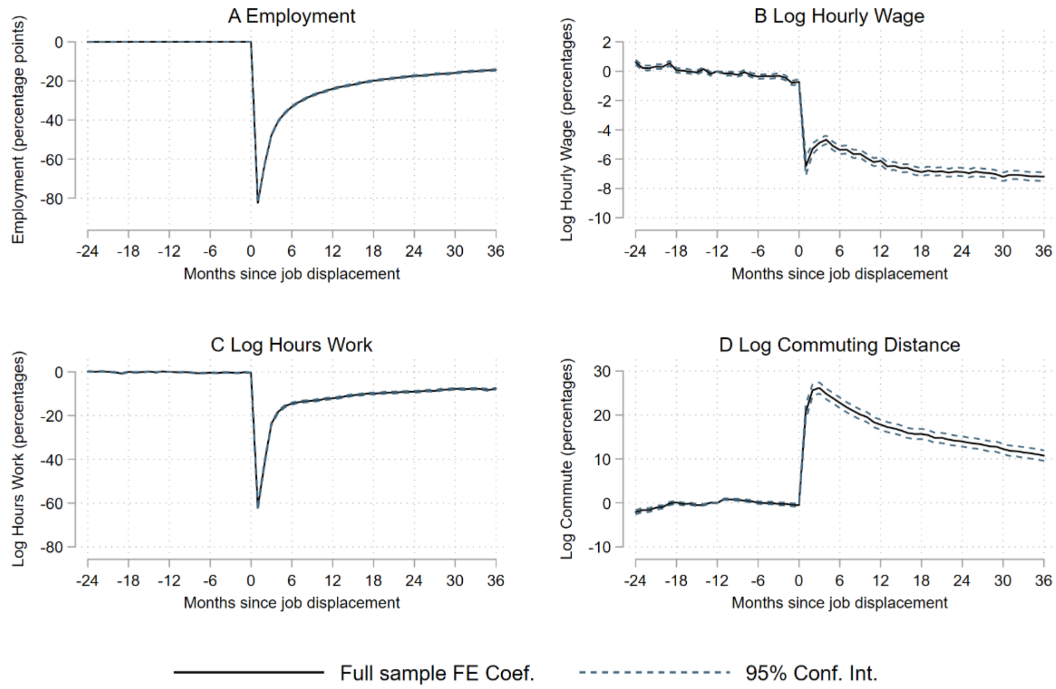
- For regression discontinuity, the regression equation is:
  $Y_i = \beta_0 + \beta_1 X_i + \rho D_i + u_i$
- 0-1 dummy D_i is one for firms above the discontinuity.
- Estimate this equation for a subsample of firms around the discontinuity. $\hat{\rho}$ is the RD-estimator of the subsidy. $\hat{\beta}_1$ should be statistically insignificant.
- Problem: At the threshold (the discontinuity) there may be not enough observations.
- Check that there are no substantial changes in the *x*-variables around the discontinuity.
- The RD-estimator is a **local** estimator. It captures only the effect around the discontinuity.

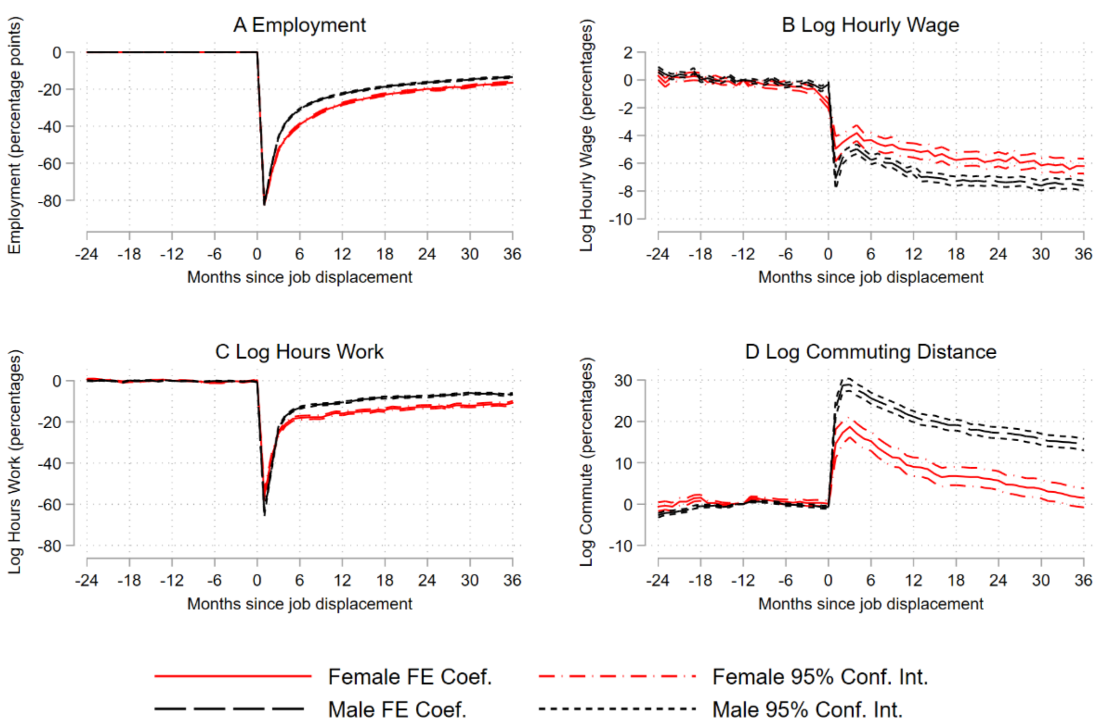
Difference-in-differences estimator

Differences in differences – Example

Meekes, J., & Hassink, W. H. (2024). Gender differences in job flexibility: Commutes and working hours after job loss. *Journal of Urban Economics*, *129*, 103425.

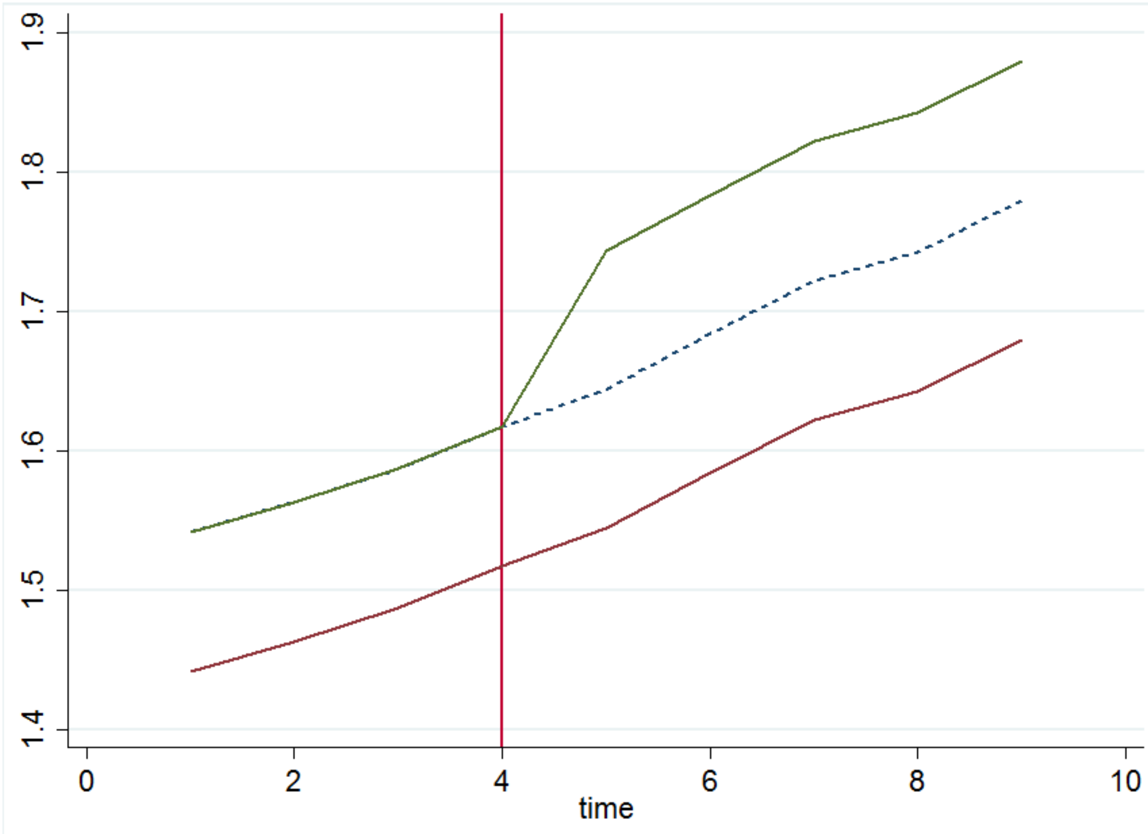
This paper studies whether women and men cope with job loss differently. Using 2006-2017 Dutch administrative monthly microdata and **a quasi-experimental empirical design involving job displacement because of firm bankruptcy, we find that displaced women are more likely than displaced men to find a flexible job with limited working hours and short commutes**. Relative to displaced men, displaced women tend to acquire a job with an 8 percentage points larger loss in working hours and an 8 percentage points smaller increase in commuting. However, displaced women experience longer unemployment durations and comparable hourly wage losses. Job loss thus widens gender gaps in employment, working hours and commuting distance. Further, results point out that displaced expectant mothers experience relatively high losses in employment and working hours, amplifying child penalty effects. The findings show that firm bankruptcy for expectant mothers widens gender gaps in employment and working hours.


A Employment
Employment (percentage points)
Months since job displacement
B Log Hourly Wage
Log Hourly Wage (percentages)
Months since job displacement
C Log Hours Work
Log Hours Work (percentages)
Months since job displacement
D Log Commuting Distance
Log Commute (percentages)
Months since job displacement
Full sample FE Coef.
95% Conf. Int.


A Employment
Employment (percentage points)
Months since job displacement
B Log Hourly Wage
Log Hourly Wage (percentages)
Months since job displacement
C Log Hours Work
Log Hours Work (percentages)
Months since job displacement
D Log Commuting Distance
Log Commute (percentages)
Months since job displacement
Female FE Coef.
Female 95% Conf. Int.
Male FE Coef.
Male 95% Conf. Int.

Second-best method: natural experiment – Differences-in-differences method

- An institutional policy change is characterized as the ‘treatment’.
- The dependent variable, y, may change as a result of the treatment
- There are two groups of individuals. They either receive the treatment, or they do not.
- Treatment is a binary variable ($dT = 1$ if treated; $dT = 0$ otherwise)
- There are two periods of observation for both groups: t=1 and t= 2. In period 2, the second group is treated.
- Assumption: common trends assumption: both groups have same development across time, without being treated.
- We would like to know what would have happened to the treatment group if they were not treated – the counterfactual. This outcome cannot be observed, but we can infer the effect by the difference between the treatment and control groups after period 2.
- Next, we make a **common trends assumption**. It is assumed that the trend does not differ between both groups of individuals (treatment and non-treatment) in absence of treatment.
- In addition, we assume that the assignment (treatment versus non-treatment) is independent of the error term.


1.9
1.8
1.7
1.6
1.5
1.4
0
2
4
6
8
10
time

- The difference-in-differences estimator:

$$\hat{\delta}_1^{did} = \underbrace{\left(\overline{y}_2^{tr} - \overline{y}_2^{ntr}\right)}_{\substack{\text{difference in}\\ \text{period 2}}} - \underbrace{\left(\overline{y}_1^{tr} - \overline{y}_1^{ntr}\right)}_{\substack{\text{difference in}\\ \text{period 1}}} \tag{13.6}$$

For which $(\overline{y}_2^{tr} - \overline{y}_2^{ntr})$ is the average difference of the dependent variable between both groups (treated versus non treated) in the second period
$(\overline{y}_1^{tr} - \overline{y}_1^{ntr})$ is the average difference between both groups (treated versus non treated) for the first period.

- Another way to measure the difference-in-differences estimator:

$$y = \beta_0 + \delta_0 d2 + \beta_1 dT + \delta_1 d2 \cdot dT + other \;\; factors \tag{13.10}$$

d2 is a 0-1 dummy for the treatment group
dT is a 0-1 dummy for the second period (period of treatment)
$\hat{\delta}_1$ is the dif-in-dif estimator (obtained through Ordinary Least Squares)

- **Example:** two groups of firms. One group of firms receives a government subsidy to promote innovation. Both groups are observed before and after the subsidy (before and after intervention). Question: what is the effect of innovation subsidies on specific outcome variables (e.g. profit)?

Binary choice models

Binary choice models

Limited dependent variable models describe phenomena of a discrete or mixed discrete continuous nature.

Examples:

a) Binary choice models - e.g. the choice to work or not

b) Ordered response models

Do you expect that your income next year will:
1. Decrease a lot
2. Decrease a bit
3. Stay more or less the same
4. Increase a bit
5. Increase a lot

c) Multinomial response models

Which of the following is your favoured use of social media?
Facebook
LinkedIn
Twitter
Tumblr
Instagram
Pinterest
blog, vlog or Youtube channels
Google+
Snapchat

Unlike in ordered response models, there is no logical ordering of the dependent variable in multinomial models.

- Here: Binary choice models

Background: Bernoulli-distribution

Bernoulli distributions are used to understand binary choice models,

We start with a (Bernoulli) binomial distribution for n=1

$$f(Y;\theta) = \theta^{Y}(1-\theta)^{1-Y}$$

With

$$\Pr(Y=1) = \theta^{1}(1-\theta)^{0} = \theta$$

$$\Pr(Y=0) = \theta^{0}(1-\theta)^{1} = (1-\theta)$$

$$EY = 0\cdot\Pr(Y=0) + 1\cdot\Pr(Y=1) = 1\cdot\theta = \theta$$

$$EY^{2} = 0^{2}\cdot\Pr(Y=0) + 1^{2}\cdot\Pr(Y=1) = 1\cdot\theta = \theta$$

$$VarY = E(Y-EY)^{2} = EY^{2} - (EY)^{2} = \theta - \theta^{2} = \theta(1-\theta)$$

Linear Probability Model

Linear probability model (LPM)

Aim: to introduce the LPM-model, to consider the expected value of y and to interpret the parameters of the LPM.

- We begin by considering cross sectional data, without a panel structure.
- The dependent variable is a dummy variable

$y_i = 1$ if firm *i* uses social media

$y_i = 0$ if firm *i* does not use social media

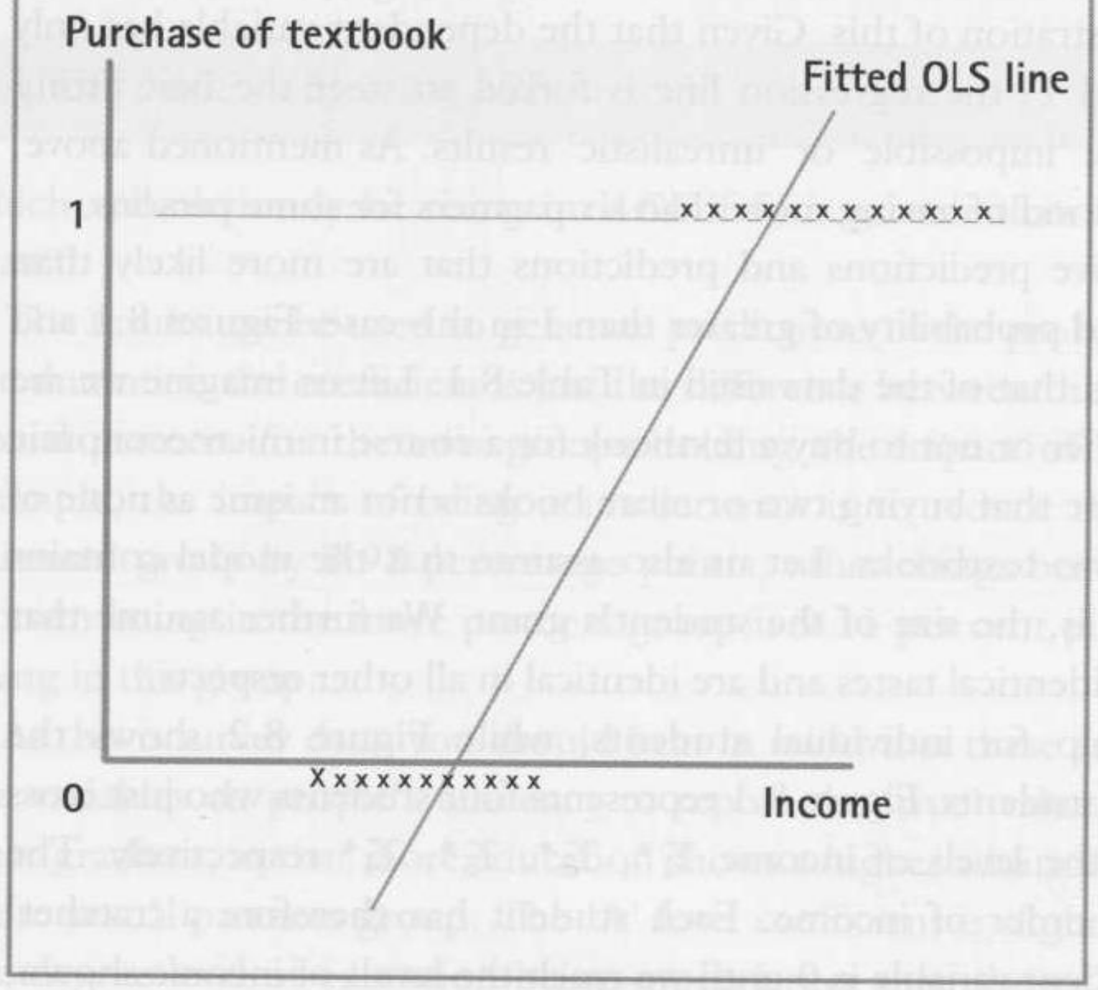
- y is a linear function of the explanatory variables using the linear regression model (the linear probability model).

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i \qquad (1)$$

where the usual assumptions regarding u hold.

- The dependent variable takes the value of either 0 or 1. The explanatory variables may take any value.

Figure 8.2 Application of linear probability model (OLS) to student textbook purchase decision



- LPM: estimate the regression parameters β by OLS
- In a linear regression equation the familiar property holds:

$$E(y_i \mid x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} \qquad (2)$$

- We make use of the fact that *y* has a Bernoulli distribution (see previous slide). The expected value of a Bernoulli variable is the following:

$$E(y_i \mid x_i) = \sum_i y_i \cdot \Pr[Y = y_i \mid x_i] =$$
$$1 \cdot \Pr[y_i = 1 \mid x_i] + 0 \cdot \Pr[y_i = 0 \mid x_i] = \Pr[y_i = 1 \mid x_i] \qquad (3)$$

- Equation (2) is equal to equation (3):

  $$\Pr(y_i = 1 | x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}$$

- In this particular linear model $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}$ can be interpreted as the probability that a randomly selected family owns a car.

  $$\Delta \Pr(y_i = 1 | x_i) = \beta \Delta x_i$$

- A one unit change in the dependent variable is 100%.
- A change in the dependent variable is in percentage points!
- Calculate the marginal effect to measure the impact of a change an explanatory variable on the dependent variable

  What is the marginal effect? Take the partial derivative

  $$\frac{\partial[\Pr(y_i = 1 | x_{1i})]}{\partial x_1} = \frac{\partial[\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}]}{\partial x_1} = \beta_1$$

- Thus the coefficient on an explanatory variable gives the marginal effect

- The error terms in the LPM-model may be interpreted in the following way:

  Model:

  $$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i \qquad y = 0,1$$

  Error terms

  $$u_i = 1 - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \ldots - \beta_k x_{ki}$$

  with probability $P = \Pr[y = 1]$

  $$u_i = -(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki})$$

  with probability $1 - P = 1 - \Pr[y = 1]$

Variance of *y*

Aim: to demonstrate why robust standard errors must be computed when using the LPM.

- In the same vein one can show that

  $$Var(y_i \mid x_i) = \Pr[y_i = 1 \mid x_i] \cdot (1 - \Pr[y_i = 1 \mid x_i])$$

  This is because *y* has a Bernoulli distribution

- So the variance of y_i depends on x_i. It implies there is heteroskedasticity.

- Conclusions

  1) We can estimate the Linear Probability model
  $$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i$$

  2) We should correct for heteroskedasticity
  Calculate the robust standard errors

LPM model summary

Advantages:

- The model does not assume a specific distribution of the error term *u*.
- It is simple to compute
- The model can be easily adapted to panel data.

Disadvantages:

- The predicted value of the dependent variable (predicted value) may be outside the [0,1] range:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_k x_{ki}$$

- If *Ey* is not close to zero or one, it is likely that for many observations:

$$0 \leq \hat{y}_i \leq 1$$

- It is important to check the incidence of the predicted values (<0 and >1) after LPM.

Example: Application of LPM-model (data from prof. Jeroen de Jong)

Survey by Dutch Chamber of Commerce on the use of Social Media by Entrepreneurs

Question of research: Which social media are used by you of your firm for purposes of business?

Facebook
LinkedIn
Twitter
Tumblr
Instagram
Pinterest
blog, vlog or Youtube channels
Google+
Snapchat

Other questions:
What is the strategy of your firm to be competitive in the market?

- Low costs
- Low pricelevel
- Innovation

Are the customers other businesses, consumers or a mix of both?

```
. tabulate soc_media

  soc_media |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        360       19.16       19.16
          1 |        521       27.73       46.89
          2 |        440       23.42       70.30
          3 |        318       16.92       87.23
          4 |        139        7.40       94.62
          5 |         63        3.35       97.98
          6 |         23        1.22       99.20
          7 |         13        0.69       99.89
          8 |          2        0.11      100.00
------------+-----------------------------------
      Total |      1,879      100.00
```

Interpretation of dum_socmedia: 1 if firm is applying any form of social media. 0 is no use of social media.

```
. gen dum_socmedia = 0
. replace dum_socmedia = 1 if soc_media > 0


. sum dum_socmedia consumer costs price_level innovation

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
dum_socmedia |      1879    .8084087    .3936579          0          1
   consumers |      1879    2.465141    1.289646          1          5
       costs |      1879     .266099    .4420343          0          1
 price_level |      1879     .158595    .3653952          0          1
  innovation |      1879    .1601916    .3668813          0          1
```

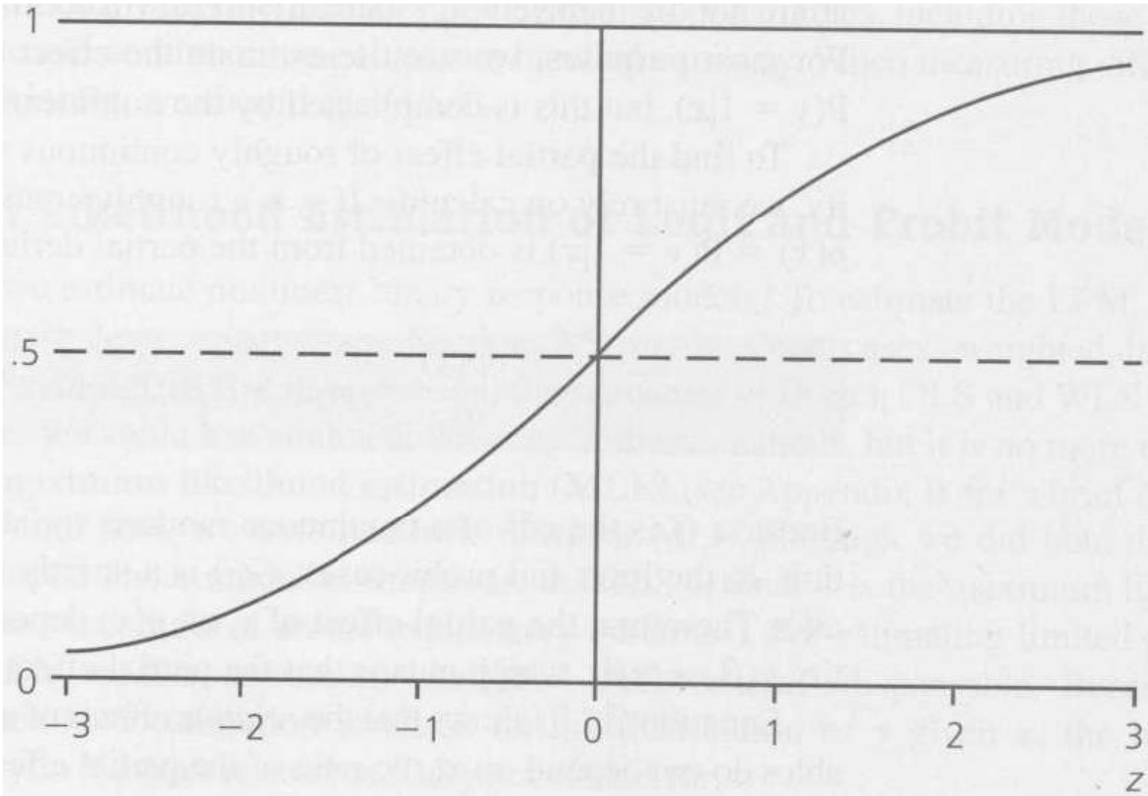
We estimate a Linear Probability Model with robust standard errors. Firms that apply low costs strategy have 6.8 percentage points lower use of social media. Firms that apply a strategy of innovation have 11.4 percentage points higher use of social media.

```
. reg dum_socmedia consumer costs price_level innovation, robust

Linear regression                                      Number of obs =    1879
                                                       F(  4,  1874) =   11.72
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0190
                                                       Root MSE      =  .39031

------------------------------------------------------------------------------
             |               Robust
dum_socmedia |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   consumers |   .0034023   .0071925     0.47   0.636    -.0107038    .0175084
       costs |  -.0683147    .021827    -3.13   0.002    -.1111224   -.0255069
 price_level |   .0156037   .0245839     0.63   0.526     -.032611    .0638185
  innovation |   .1143292    .019716     5.80   0.000     .0756616    .1529968
       _cons |   .7974108   .0212699    37.49   0.000     .7556957     .839126
------------------------------------------------------------------------------


. predict yhat
(option xb assumed; fitted values)
. gen out_of_range = 0
. replace out_of_range = 1 if yhat < 0
(0 real changes made)
. replace out_of_range = 1 if yhat > 1
(0 real changes made)
. sum out_of_range
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
out_of_range |      1879           0           0          0          0
```

The output above shows that all of the predictions are in the 0-1 range.

Logit and Probit – motivation

Ideal fit of probability of dependent variable



Binary response model

Define a latent dependent variable y^* and assume it relates in a linear way to the independent variables x

$$y^* = \beta_0 + x\beta + e$$

The error term is denoted by *e*. We do NOT observe y^* but only whether or not this variable takes on a value larger than zero or not. In this latter case the observed dependent variable (*y*) takes on the value 1. We formalize this as follows

$$y = \begin{cases} 0 & \text{if} \quad y^* \leq 0 \\ 1 & \text{if} \quad y^* > 0 \end{cases}$$

Rewrite as

$$y = \begin{cases} 0 & \text{if} \quad \beta_0 + x\beta + e \leq 0 \\ 1 & \text{if} \quad \beta_0 + x\beta + e > 0 \end{cases}$$

Next:

$$y = \begin{cases} 0 & \text{if} \quad e \leq -\left(\beta_0 + x\beta\right) \\ 1 & \text{if} \quad e > -\left(\beta_0 + x\beta\right) \end{cases}$$

We observe *y* and *x* and <u>if</u> we know the probability density function of the error term ***e*** we can fully specify:

$$\Pr(y = 1 \mid x; \beta) = \Pr\left(e > -\left(\beta_0 + x\beta\right)\right)$$

In words: if we know *x*, and have an estimate for β, we can calculate (predict) the probability that *y* is equal to 1.

The logit model

The Logit model

Aim: to introduce the Logit model

- For the probability of Y=1 we need a non-linear function $G(t)$, which has the following properties:
  - Continuous (i.e. smooth) non-decreasing function
  - $t = -\infty$ gives a value of 0 for $G(t)$
  - $t = \infty$ gives a value of 1 for $G(t)$

$$G(z) = \Lambda(z) = \frac{\exp(z)}{1 + \exp(z)}$$

$$z = \beta_0 + \beta_1 X_1$$

Apply a Logit model

Aim: to explain the mathematical basis for the Logit model

- The Logit model is based on a logistic function:
- $G(z) = \Lambda(z) = \dfrac{\exp(z)}{1+\exp(z)}$
- Note that in this function $\Lambda(z)$ takes a value between 0 and 1.
- $z = -\infty$: $\Lambda(z)$ gets a value of 0
- $z = \infty$: $\Lambda(z)$ gets a value of 1

- $\Lambda'(z) = \dfrac{\exp(z)(1+\exp(z)) - \exp(z) * \exp(z)}{(1+\exp(z))^2} = \dfrac{\exp(z)}{(1+\exp(z))^2}$

The derivative has a nice mathematical feature

- $\Lambda'(z) = \Lambda(z)(1-\Lambda(z))$

Reason:

$$\Lambda(z)(1-\Lambda(z)) = \frac{\exp(z)}{1+\exp(z)} * [1 - \frac{\exp(z)}{1+\exp(z)}] = \frac{\exp(z)}{1+\exp(z)} * [\frac{1}{1+\exp(z)}] = \Lambda'(z)$$

- $\Pr(Y=1 \mid X_1) = \dfrac{\exp(\beta_0 + \beta_1 X_1)}{1+\exp(\beta_0 + \beta_1 X_1)} = \Lambda(\beta_0 + \beta_1 X_1)$
- $\beta_0 + \beta_1 X_1 \to -\infty$ gives a probability of 0
- $\beta_0 + \beta_1 X_1 \to \infty$ gives a probability of 1

- Both values are asymptotic values (i.e. never exactly equal to 0 or 1)

Notes:
- There is no error term *u* in logistic function
- *Y* follows a Bernoulli distribution.

How is the marginal effect calculated for a Logit model?

Aim: to introduce the marginal effect for the logit model (for continuous explanatory variable X_1*)*

- The individual effect of an explanatory variable on the dependent is the marginal effect

- Procedure (see LPM example): take the first partial derivative of the probability with respect to X_1

- $\Pr(Y=1|X_1) = \dfrac{\exp(\beta_0+\beta_1 X_1)}{1+\exp(\beta_0+\beta_1 X_1)} = \Lambda(\beta_0+\beta_1 X_1)$

- $\dfrac{\partial[\Pr(Y=1|X_1)]}{\partial X_1} = \Lambda'(z) * \beta_1 = \Lambda(z)(1-\Lambda(z)) * \beta_1$

where $z = \beta_0 + \beta_1 X_1$

- Note that we have used the chain rule:
$$\frac{\partial z}{\partial X_1} = \beta_1$$
- This formula can easily be applied again to calculate the derivative: $\Lambda'(z) = \Lambda(z) * (1-\Lambda(z))$

How is the marginal effect calculated when we have a discrete explanatory variable?

Aim: to discuss the marginal effect for the Logit model (for discrete explanatory variable X_k)

- The first derivative cannot be calculated using the procedure above.

- Assume
  - that there are *k* explanatory variables
  - the discrete explanatory variable is X_k
  - that X_k changes from *c* towards *c*+1 (thus from 0 to 1)

- For all further *k*-1 explanatory variables we take the average: $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_{k-1}$
- For X_k we take the values *c* and *c*+1 (0 and 1)

- The marginal effect of the discrete variable:

$$\Lambda(\hat{\beta}_0 + \hat{\beta}_1 \overline{X}_1 + \hat{\beta}_2 \overline{X}_2 + \ldots + \hat{\beta}_{k-1} \overline{X}_{k-1} + \hat{\beta}_k (c+1)) -$$

$$\Lambda(\hat{\beta}_0 + \hat{\beta}_1 \overline{X}_1 + \hat{\beta}_2 \overline{X}_2 + \ldots + \hat{\beta}_{k-1} \overline{X}_{k-1} + \hat{\beta}_k c)$$

Stata: Application to Logit

```
. logit dum_socmedia consumer costs price_level innovation
Iteration 0:   log likelihood = -917.93306
Iteration 1:   log likelihood = -899.10058
Iteration 2:   log likelihood = -898.41825
Iteration 3:   log likelihood = -898.41702
Iteration 4:   log likelihood = -898.41702
Logistic regression                               Number of obs   =       1879
                                                  LR chi2(4)      =      39.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -898.41702                       Pseudo R2       =     0.0213
------------------------------------------------------------------------------
dum_socmedia |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   consumers |   .0225883   .0460796     0.49   0.624    -.0677259    .1129026
       costs |  -.4135992   .1269221    -3.26   0.001    -.6623619   -.1648365
 price_level |   .1016056   .1641313     0.62   0.536    -.2200858    .4232971
  innovation |   .9568558   .2130065     4.49   0.000     .5393706    1.374341
       _cons |   1.374571   .1391009     9.88   0.000     1.101938    1.647204
------------------------------------------------------------------------------

. mfx
Marginal effects after logit
      y  = Pr(dum_socmedia) (predict)
         =   .8160168
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
consum~s |   .0033913      .00692    0.49   0.624  -.010166  .016948   2.46514
   costs*|  -.0659355      .02131   -3.09   0.002  -.107704 -.024167   .266099
price_~l*|   .0149217      .02356    0.63   0.527  -.031264  .061107   .158595
innova~n*|    .116427      .01989    5.85   0.000   .077452  .155401   .160192
------------------------------------------------------------------------------
   (*) dy/dx is for discrete change of dummy variable from 0 to 1
```

The Probit model

Estimating the Probit Model

We take the following formulation:

$$y^* = \beta_0 + x\beta + e$$

$$y = \begin{cases} 0 & \text{if} \quad y^* \le 0 \\ 1 & \text{if} \quad y^* > 0 \end{cases}$$

e has a standard normal density: $\Pr(e < z) = \Phi(z)$ (the cumulative distribution function)

This implies (see some earlier slides):

$$\Pr(y = 1 \mid x; \beta_0, \beta) = \Phi\left(\beta_0 + x\beta\right)$$
$$\Pr(y = 0 \mid x; \beta_0, \beta) = 1 - \Phi\left(\beta_0 + x\beta\right)$$

STATA: Marginal effects of probit model and that of Logit model render similar estimation results as LPM model

```
. dprobit dum_socmedia consumer costs price_level innovation
Iteration 0:   log likelihood = -917.93306
Iteration 1:   log likelihood = -898.78919
Iteration 2:   log likelihood = -898.56232
Iteration 3:   log likelihood = -898.56219

Probit regression, reporting marginal effects          Number of obs =   1879
                                                       LR chi2(4)    =  38.74
                                                       Prob > chi2   = 0.0000
Log likelihood = -898.56219                            Pseudo R2     = 0.0211
------------------------------------------------------------------------------
dum_so~a |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
consum~s |   .0030612   .0069709     0.44   0.661   2.46514  -.010602  .016724
   costs*|   -.065891   .0215829    -3.18   0.001   .266099  -.108193 -.023589
price_~l*|   .0159274   .0239574     0.65   0.514   .158595  -.031028  .062883
innova~n*|   .1158772   .0199828     4.71   0.000   .160192   .076712  .155043
---------+--------------------------------------------------------------------
  obs. P |   .8084087
 pred. P |   .8143263  (at x-bar)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| correspond to the test of the underlying coefficient being 0


. logit dum_socmedia consumer costs price_level innovation
Iteration 0:   log likelihood = -917.93306
Iteration 1:   log likelihood = -899.10058
Iteration 2:   log likelihood = -898.41825
Iteration 3:   log likelihood = -898.41702
Iteration 4:   log likelihood = -898.41702
Logistic regression                               Number of obs   =       1879
                                                  LR chi2(4)      =      39.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -898.41702                       Pseudo R2       =     0.0213
------------------------------------------------------------------------------
dum_socmedia |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   consumers |   .0225883   .0460796     0.49   0.624    -.0677259    .1129026
       costs |  -.4135992   .1269221    -3.26   0.001    -.6623619   -.1648365
 price_level |   .1016056   .1641313     0.62   0.536    -.2200858    .4232971
  innovation |   .9568558   .2130065     4.49   0.000     .5393706    1.374341
       _cons |   1.374571   .1391009     9.88   0.000     1.101938    1.647204
------------------------------------------------------------------------------

. mfx
Marginal effects after logit
      y  = Pr(dum_socmedia) (predict)
         =   .8160168
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
consum~s |   .0033913      .00692    0.49   0.624  -.010166  .016948   2.46514
   costs*|  -.0659355      .02131   -3.09   0.002  -.107704 -.024167   .266099
price_~l*|   .0149217      .02356    0.63   0.527  -.031264  .061107   .158595
innova~n*|    .116427      .01989    5.85   0.000   .077452  .155401   .160192
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

Probit versus Logit

Remark 1: there is no fundamental difference between a Logit and Probit model. The only difference is the distributional assumption. The empirical results do not fundamentally differ between these two approaches. (one is not "better" than the other)

Remark 2: the estimates of β will differ between the probit and logit models, since they are essentially not the same parameters (only up to a constant fraction due to differences in the variance of **e**), but the marginal effects of x and the choice probability are very similar. We return to this issue later on.

Remark 3: an important advantage of using a Probit or Logit model is that the predicted probabilities are always between 0 and 1.

Winding up

- Experiments and quasi experimental methods (regression discontinuity and differences-in-differences are extremely useful to get information on causal effects.

- We have considered the LPM, in which the dependent variable is a 0-1 variable

- Overall, this course has considered the following major topics:
  Time series
  Panel Data
  Instrumental variables

- Difference techniques; very useful for applied work in many fields of economics.

- It is not finished here: block 2: another course on research methods.

APPENDIX (it is NO part of the course material): Sample selection and Heckman selection – an introduction

Motivation for this appendix. The Heckman selection models are used a lot in empirical analyses of entrepreneurship. Example:

Van Balen, T., Tarakci, M., & Sood, A. (2019). Do disruptive visions pay off? The impact of disruptive entrepreneurial visions on venture funding. *Journal of Management Studies*, *56*(2), 303-342.

- We apply the selection model to the labour supply decision:
    - Two decision processes:
        1. Selection equation: Decision to work or not to work
        2. Wage equation
- We usually observe the wage only for the people who have decided to work.
- In other words: we do not observe the wage for people who do not work.
- In other words: there is a selection of the sample on the dependent variable
- OLS gives biased parameter estimates. Thus, we cannot measure a causal effect.

Sample selection - model

- We have two regression equations, which describe the selection process (yes/no work) and the other decision (wage).

- The selection equation is modeled by a Probit model (independent variables $x_1,...,x_l$)

$$h^* = \gamma_0 + \gamma_1 x_1 + ... + \gamma_l x_l + u_2$$

$$h = 1 \text{ if } h^* > 0$$

$$h = 0 \text{ if } h^* \leq 0$$

- The second equation is the wage equation: wages are only observed for people who actually work (independent variables $z_1,...,z_k$; there are fewer variables in the wage equation than in the selection equation)

$$w^* = \beta_0 + \beta_1 z_1 + ... + \beta_k z_k + u_1$$

- w^* is not observed for people who are not working (this leads to selection)

$$w = w^* \text{ if } h = 1 \text{ if } h^* > 0$$

w is not observed if $h = 0$ if $h^* \leq 0$

- Wage equation is the usual regression equation

- The conditional wage, given that a person is working:

$$E(w \mid h=1) = \beta_0 + \beta_1 z_1 + ... + \beta_k z_k + \rho \frac{\phi(\gamma_0 + \gamma_1 x_1 + ... + \gamma_l x_l)}{\Phi(\gamma_0 + \gamma_1 x_1 + ... + \gamma_l x_l)}$$

$$= \beta_0 + \beta_1 z_1 + ... + \beta_k z_k + \rho \underbrace{\lambda(\gamma_0 + \gamma_1 x_1 + ... + \gamma_l x_l)}_{\text{the inverse Mill's ratio}}$$

- $\lambda(\gamma_0 + \gamma_1 x_1 + ... + \gamma_l x_l)$: Heckman's lambda or the Inverse Mill's ratio
- There is no selection bias if the parameter $\rho = 0$. The equation becomes

$$E(w \mid h=1) = \beta_0 + \beta_1 z_1 + ... + \beta_k z_k$$

- There is sample selection bias if the parameter $\rho \neq 0$. Thus we need to construct the Inverse Mill's ratio, to estimate the parameter ρ together with the regression parameters $\beta_0, \beta_1, ..., \beta_k$

- For this method, we have to apply the following three assumptions:

  - Normality of the error terms u_1 and u_2
  - All variables included in $z_1, ..., z_k$ are also in the selection equation $x_1, ..., x_l$
  - The selection equation contains some variables $x_1, ..., x_l$ that are not included in $z_1, ..., z_k$

Two-step estimation method (Heckman 1979)

The two-step procedure is based on the following regression:

$$w = \beta_0 + \beta_1 z_1 + ... + \beta_k z_k + \rho\lambda + u$$

where $\lambda = \dfrac{\phi(\gamma_0 + \gamma_1 x_1 + ... + \gamma_l x_l)}{\Phi(\gamma_0 + \gamma_1 x_1 + ... + \gamma_l x_l)}$ (1)

Step 1: Estimate the parameters γ of the selection equation by means of probit (result: $\hat{\gamma}$).

Step 2: Construct the inverse Mill’s ratio for each individual *i*:

$$\hat{\lambda}_i = \frac{\phi(\hat{\gamma}_0 + \hat{\gamma}_1 x_{i1} + ... + \hat{\gamma}_l x_{il})}{\Phi(\hat{\gamma}_0 + \hat{\gamma}_1 x_{i1} + ... + \hat{\gamma}_l x_{il})}$$

Step 3: Estimate the following model on the “selected sample” (*h*=1)

$$w = \beta_0 + \beta_1 z_1 + ... + \beta_k z_k + \rho\hat{\lambda} + u \qquad (2)$$

- Solution to this problem: find “exclusion restrictions’ (i.e. find variables that are included in $x_1,...,x_l$ and not in $z_1,...,z_k$ (remember that all variables in $z_1,...,z_k$ are also in $x_1,...,x_l$ and that $l > k$) Such variables are hard to find. Sensitivity analysis is important.

Example from literature of entrepreneurship

Lien, W. C., Chen, J., & Sohl, J. (2024). Do I have a big ego? Angel investors' narcissism and investment behaviors. *Journal of Business Venturing*, *37*(5), 106247.

In this study, we draw on the threatened egotism theory to examine the effect of angel narcissism on their investment behaviors and the boundary condition of past investment performance. We propose that angel narcissism is positively related to deal size and portfolio industry diversification but negatively related to the number of co-investors. Moreover, past investment performance moderates these relationships such that the effects of angel narcissism on their investment behaviors are stronger when past investment performance is lower. Data from a longitudinal analysis of 133 angels from 2010 to 2019 largely supported our hypotheses. Our study, the first to examine angel narcissism, highlights the psychological foundation of angel investing.

One indicator of narcissism that we used is based on the narratives (interviews, speeches, and direct quotes). However, we obtained narratives for only a portion of our sampled angels, which might result in a bias toward the angels with a relatively great public presence. **We adopted a Heckman two-stage procedure (Heckman, 1979) to account for potential sample selection bias (Certo et al., 2016). Our first-stage probit model predicted the presence of angels in our final sample. We included each angel's gender, education, investing experience, entrepreneurial experience, and total current investments in a probit regression. Certo et al. (2016: 2644) suggested that researchers should include "at least one variable in the first stage that does not appear in the second stage." Following this advice, we included angels' elite education6 as an additional predictor in the first-stage regression.**

```
reg lnrealwage age age2 dmale dprimeduc dseceduc
dhigheduc dy* if dloon > 0, robust

Linear regression              Number of obs =  421995
                               F( 19,421975) =67313.78
                               Prob > F      =  0.0000
                               R-squared     =  0.7522
                               Root MSE      =  .64233

---------------------------------------------------------
|               Robust
  lnrealwage |      Coef.   Std. Err.      t    P>|t|
[95% Conf. Interval]
-------------+-------------------------------------------
         age |    .071126   .0007483    95.05
        age2 |  -.0007515   .0000101   -74.53
       dmale |   .4112667   .0024081   170.79
   dprimeduc |   .3479937   .0062099    56.04
    dseceduc |   .9442209    .006184   152.69
   dhigheduc |   1.402113    .006716   208.77
         dy2 |   .0849304   .0069224    12.27
         dy4 |   .1078749   .0066755    16.16
         dy5 |   .1208934   .0063125    19.15
         dy6 |   .1563493   .0063101    24.78
         dy7 |   .1936037   .0063184    30.64
         dy8 |   .2313503   .0064229    36.02
         dy9 |   .2909641    .006421    45.31
        dy10 |   .3085834   .0063427    48.65
        dy12 |   .0005565   .0078209     0.07
        dy13 |   .1465183   .0084218    17.40
        dy14 |   .4556952   .0074471    61.19
        dy15 |   .4891831   .0064967    75.30
       _cons |    8.67865   .0149631   580.00
---------------------------------------------------------
```

```
dprobit dloon age age2 dmale djawa dmale dprimeduc
dseceduc dhigheduc dy*

note: dmale dropped due to collinearity
note: dy1 dropped due to collinearity
note: dy12 dropped due to collinearity
Iteration 0:   log likelihood = -993218.79
Iteration 1:   log likelihood = -852483.19
Iteration 2:   log likelihood = -849662.92
Iteration 3:   log likelihood = -849655.61
Iteration 4:   log likelihood = -849655.61

Probit regression, reporting marginal effects
Number of obs =1849272
LR chi2(20)   = 2.9e+05
Prob > chi2   = 0.0000
Log likelihood = -849655.61
Pseudo R2     = 0.1445

-----------------------------------------
 dloon |      dF/dx   Std. Err.      z
---------+-------------------------------
     age |   .0200792   .0002036    98.48
    age2 |  -.0002776   2.68e-06  -103.45
   dmale*|    .182046   .0006103   291.59
   djawa*|   .1040514   .0006239   165.35
dprime~c*|   .0186775   .0011737    15.88
dseceduc*|   .2116248   .0014303   157.07
dhighe~c*|   .5625979   .0021255   235.60
     dy2*|   .0024567   .0021559     1.14
     dy4*|    .012912   .0022051     5.95
     dy5*|   .0179425   .0021436     8.55
     dy6*|   .0165399   .0021403     7.88
     dy7*|    .016286   .0021392     7.76
     dy8*|   .0228274   .0021676    10.81
     dy9*|   .0280902   .0022162    13.09
    dy10*|    .023367   .0021866    10.98
    dy11*|   .0045432   .0025509     1.79
    dy13*|   -.055039   .0019707   -25.21
    dy14*|  -.0227163   .0021503   -10.19
    dy15*|  -.0423776   .0018613   -21.32
---------+--------------------------------
  obs. P |   .2281952
 pred. P |    .195292  (at x-bar)
------------------------------------------
 (*) dF/dx is for discrete change of dummy variable from
0 to 1
```

heckman lnrealwage age age2 dmale dprimeduc dseceduc dhigheduc dy*, select(dloon = age age2 dmale djawa dprimeduc dseceduc dhigheduc dy*) robust

```
Heckman selection model                           Number of obs      =   1849272
(regression model with sample selection)          Censored obs       =   1427277
                                                  Uncensored obs     =    421995
                                                  Wald chi2(19)      =  1.11e+06
Log pseudolikelihood =  -1260633                  Prob > chi2        =    0.0000
-----------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z
-------------+-------------------------------
lnrealwage   |
         age |   .0838211   .0009344    89.71
        age2 |  -.0009294   .0000128   -72.65
       dmale |   .5363367   .0059678    89.87
   dprimeduc |   .3626675   .0061771    58.71
    dseceduc |   1.070063   .0079998   133.76
   dhigheduc |   1.662124   .0129022   128.82
         dy2 |    .054732   .0069508     7.87
         dy4 |    .085015   .0067851    12.53
         dy5 |   .1007509   .0064978    15.51
         dy6 |   .1356917   .0065032    20.87
         dy7 |   .1725471   .0065143    26.49
         dy8 |    .213692   .0065876    32.44
         dy9 |   .2799888   .0065753    42.58
        dy10 |   .2952577   .0065107    45.35
        dy11 |   .0019386   .0080073     0.24
        dy13 |   .0796449   .0090849     8.77
        dy14 |   .4181726   .0075845    55.14
        dy15 |   .4337457   .0070553    61.48
       _cons |   8.017047   .0321029   249.73
-------------+-------------------------------
dloon    age |   .0726341   .0007526    96.52
        age2 |  -.0010016   9.86e-06  -101.62
       dmale |   .6531768   .0022966   284.41
       djawa |   .3957876   .0022121   178.92
   dprimeduc |   .0676257   .0042857    15.78
    dseceduc |   .7002714   .0045548   153.74
   dhigheduc |   1.562436   .0068349   228.60
         dy2 |   .0188039    .007954     2.36
         dy3 |   .0089388   .0079339     1.13
         dy4 |    .054076   .0079242     6.82
         dy5 |    .072204   .0076397     9.45
         dy6 |   .0669711   .0076508     8.75
         dy7 |   .0662307   .0076521     8.66
         dy8 |   .0905873   .0076524    11.84
         dy9 |   .1068973   .0077084    13.87
        dy10 |   .0897727   .0076902    11.67
        dy11 |   .0173561   .0094231     1.84
        dy13 |  -.2105484   .0090186   -23.35
        dy14 |  -.0779851   .0085253    -9.15
        dy15 |  -.1564175   .0078284   -19.98
       _cons |  -2.914405   .0160144  -181.99
-------------+-------------------------------
     /athrho |    .406521   .0176167    23.08
    /lnsigma |  -.3887694   .0057713   -67.36
-------------+-------------------------------
         rho |   .3855147   .0149985
       sigma |   .6778906   .0039123
      lambda |   .2613368   .0115003
Wald test of indep. eqns. (rho = 0): chi2(1) =   532.50
-------------------------------------------------------
```