UNIUNEA EUROPEANĂ

GUVERNUL ROMÂNIEI

Instrumente Structurale
2007 - 2013

# Counterfactual Impact Assessment

## Day 3


## Klara Major



PROJECT COFINANCED BY EFRD THROUGH TAOP 2007-2013

# The counterfactual setting and the self-selection problem

- The counterfactual impact of a policy action
- natural- and quasi-experiments
- selection bias
- estimation strategies: matching models, econometric models

- quantifying whether a given intervention produces the desired effects on some pre-established dimension of interest
- Examples:
  - Do R&D subsidies increase the level of R&D expenditure by subsidized firms?
  - Do targeted ERDF funds increase per capita income of the assisted areas?
  - Do urban renewal programmes contribute to the economic development of urban neighbourhoods?
  - Does support to SMEs increase their employment levels? Does investment in new public infrastructure increase housing values?
- finding a credible approximation to what would have occurred in the absence of the intervention, and to compare it with what actually happened

# Counterfactual impact: unobservable



- The counterfactual situation is purely hypothetical, thus can never be directly observed
- Impact can be inferred as if data makes it possible to approximate the counterfactual
- What can we observe then?
  - For the beneficiary of the program („treated”): before and after
  - For the comparison between treated and non-treated
- Critical: interpretation of these differencies
  - Can we infer that these differencies indicate the presence of a casual relationship?

# Before-after: what’s wrong with it?



- The observed before-after differencis of the treated units are the result of
  - the program
  - and all other changes.
- How to separate these?
- Under normal circumstances units are supposed to develop, eg. invest, search job (some of them even find one), and so on. We might assume that without the program there would be some progress (positive or negative).
- Doing before-after comparison the only way to separate these effects is to take an assumption on the second
  - Let’s assume that without the training of the unemployed their job finding rate would remain the same. Based on this assumption we can infer the policy impact.

# Comparing the treated and the non-treated



- The most common strategy to estimate the counterfactual impact is to compare the perfomance of the treated with that of the non-treated
- The observed difference is a result of two different effect
  - the program
  - Selection bias
- There is selection bias if treated units were different even if there were no program
  - E.g.: support given to firms to invest in new equipment.
  - First selection: applicant, non-applicant
  - Second selection: supported, not supported
- If it is reasonable to assume that supported and not supported had differed even if the first had not received the support

# Elimination of selection bias



- One possible solution: to choose supported and not supported randomly
- This eliminates selection bias
- Politically unfeasible
- Practical situation the handling of the selection bias uses one of the following strategy:
  - Natural experiment
  - Quasi experiment
  - Matching models
  - Econometric models

# Quasi experiments

- Planned experiments
- The chance of being treated is the same as the chance of being non-treated
- The treated group and the non-treated group (the control group) both are random
- In this case there is no selection bias
- The treatment effect is simply the difference between the treated and non-treated average
- Homogeneous effect: the impact of the policy is the same for all unit and can be quantified
- Heterogeneous effect: the impact of the policy is different for different unit

- What do we measure by this difference?
- Average Treatment Effect (ATE)
- If the effect is supposed to be homogeneous, then it is the expected effect of the program for each individual/firm taking part in it
- If the effect is supposed to be heterogenous, than it is not possible to identify the effect on the individual.
  - Suppose unemployed are trained how to find a job. The effect of the program is said to be homogenious, if the best of the treated group were the best in the non-treated group, that is the order of the individuals do not change due to the training.
  - More general case: some of them will perform much better due to the training than others. In this case the effect is heterogenous
- Even in case of heterogenous effect, ATE is possible to quantify

# Pilot programs

- Possibility of quasi-experiments in social sciences are very limited
- Experiments with humans raise moral and political problems
- Pilot programs:
  - Before the full introduction of a program it is „tested" in some region or county.
  - This pilot program can be considered as a quasi-experiment if the region is chosen randomly
  - Based on the observations from the pilot program plus some other regions' outside the program ATE is possible to quantify
  - However regions might differ substantially from each other

- Observational study
- The assigment to treated or non-treated can be considered as random
- Natural experiments are most useful when there has been a clearly defined and large change in the treatment (or exposure) to a clearly defined subpopulation, so that changes in responses may be plausibly attributed to the change in treatments (or exposure)

# Matching models



- In most cases none of the experiments settings are not relevant
- In these cases it is necessarily to choose the control group from the untreated population to estimate ATE
- Matching models performs this task directly
- Goal: match treated and non-treated such as
  - For each non-treated find a treated that is very similar to the treated in most respect, especially
  - The non-treated had the chance to become treated
  - Participation and the observed results are uncorrelated
- In these cases the counterfactual result of the treated can be approximated by the observed result of the nontreated

- Conditions
  - The more features of treated and non-treated can be observed, the better the matching
  - For each treated there is one non-treated
    - It is difficult to ensure even in large samples
    - Who forms the control group if participation in the program needs to be in accordance with some criteria and all of those who fits these criteria participate in the program?
- Propensity score models
  - Econometrically estimate the chance of being treated using observable criteria

## Econometric models



- The treated and non-treated group froms the sample for the estimation
- To control for as much criteria as is possible
- Different estimation design has been set to identfy ATE in different problems
- Two are discussed:
  - Cross-section regressions
    - Difference-in-difference regression
  - Selection models

# Matching models

- Self-selection problem
- Selection of the control group
- Propensity score models

# Purposes of the tool

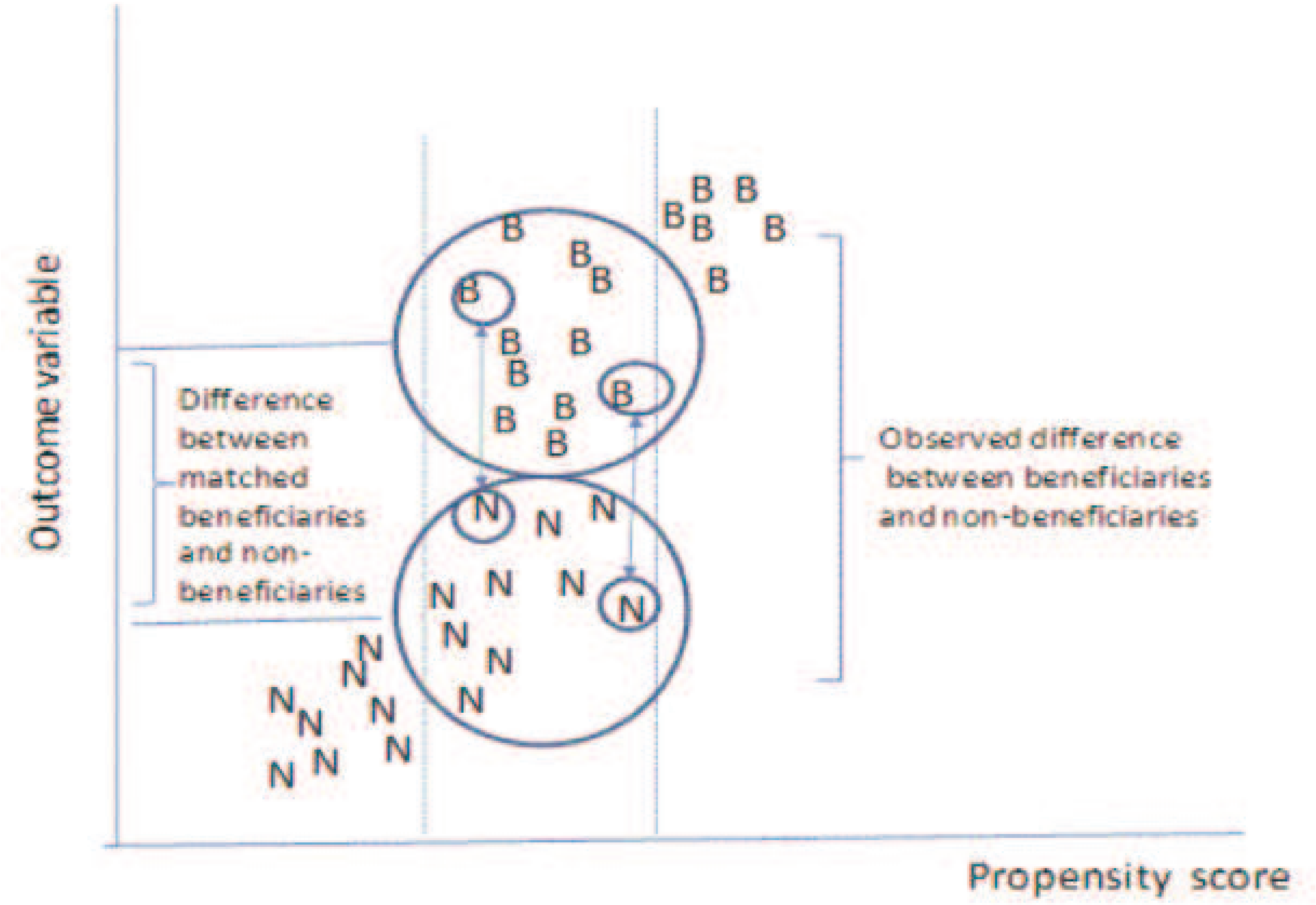
- select a group of non-beneficiaries in order to make them resemble the beneficiaries in everything, but the fact of receiving the intervention
- the effect of the intervention is estimated as the difference between the average outcomes of the two groups
- Intuitively: by constructing a control group and using difference in means, it mimics random assignment
  - However in case of random assignment the two groups similar in all respect: observable and non-observable, while in a matching framwork only in observable
- Assumption: the two groups are balances with respect to all characteristics relevant for the outcome

# The curse of dimensionality

- the list of possible variables can be too large to allow a match to be achieved on each one separately, particularly if they are continuous variables
- Rosenbaum-Rubin (1983): using propensity score to decrease dimensionality problem
- each beneficiary is matched to the non-beneficiary who is most similar in terms of probability of being a beneficiary
- the average effect is estimated for each outcome by simply computing the difference in means between the two groups

# A graphical representation of matching on the propensity score





# Conditions

- availability of characteristics observed before the intervention takes place
  - all variables affecting the selection process should be included in the list of matching variables
- existence of a substantial overlap between the characteristics of beneficiaries and non-beneficiaries
- (recently tecniques for non-binary treatment has also been developed)
- In practice, matching usually combined with difference-in-difference methods
  - Estimation in two steps:
  - 1. matcing
  - 2. diff-in-diff

1. Estimating the propensity score
2. Matching the units using propensity score
3. Assessing the quality of the match
4. Estimating ATE and its standard error

# 1. Estimating the propensity score



- Using econometric estimation method for a binary dependent variable
- Linear probability models
- Nonlinear probability models
  - Probit
  - Logit

# 2. Matching the units using propensity score



- Nearest neighbour matching
- Calliper and radius matching
  - Calliper: imposing a tolerance level on the maximum propensity score distance
  - Radius: use not only the nearest neighbour within each calliper but all of the units within the calliper
  - The advantages of these methods comes when there are numerous control but only a few treated
- Stratification Matching
  - Ranking both the treated and the control and form quintiles
  - Impact for each stratum is calculated by the difference of the mean between the 2 groups in each stratum
  - Overall impact is a weighted average of these differences, weights are propotional to the number of treated in each stratum

# Matching the units using propensity score



- Kernel matching
  - All observations in the control group are used to calculate the „matching pair”
  - Weights depend on the distance between each individual from the control group and the participant observation for which the counterfactual is estimated
    - higher weight to observations close in terms of propensity score to a treated individual and lower weight on more distant observations
  - lower variance is reached because more information is used

# The choice between matching techniques



- Trade-off: between efficiency and bias
- None of the procedure is better in all respect to others
- The best choice hardy depends upon the situation at hand

# 3. Assessing the quality of the match



- Visualization: plot the data and search for the common support
  - Or using estimate of the density function/histogram
- More formal rules:
  - Comparing the minima and maxima in both group
  - Common support is the common interval, observations outside the common support must be disregarded
  - If too much drops out the estimated effect might not be considered as representative
  - Check if distributions of main variables in the treated and control groups became balanced as a result of matching. If not, more correction is needed (eg. Interaction terms in the matching function)
  - See next table: test the hypothesis that the means are the same in the two subsample for the most interesting variables

# Assessing the quality of the match



**Table 1: Assessing the Quality of the Match in the Swedish study on Regional Development Grant**

| Variable | Treated | Control | %difference | t-statistic | p-value |
|---|---|---|---|---|---|
| Employees1999 | 29.957 | 26.179 | 10,7 | 0,610 | 0,543 |
| ROA1999 | 8.207 | 7.457 | 4,5 | 0,330 | 0,745 |
| Solidity1999 | 32.783 | 33.796 | -4,4 | -0,270 | 0,784 |
| New Company | 0.072 | 0.065 | 3,3 | 0,180 | 0,856 |
| Share Higher Education | 18.735 | 18.705 | 0,4 | 0,030 | 0,977 |
| Share State Employed | 43.333 | 42.635 | 10,4 | 0,610 | 0,544 |
| Share Foreign Born | 5.802 | 6.128 | -7,3 | -0,430 | 0,665 |
| Unemployment Rate | 3.314 | 3.351 | -4,6 | -0,250 | 0,799 |
| Migration | -0.016 | -0.016 | -2,0 | -0,120 | 0,908 |
| Income | 60,937 | 60,817 | 3,1 | 0,170 | 0,863 |
| Manufacturing | 0.623 | 0.666 | -10,1 | -0,520 | 0,606 |

Source: Gadd, Hansson and Månsson (2009) "Evaluating the impact of firm subsidy using a multilevel propensity score approach"

# 4. Estimating ATE and its standard error



- Finally and easy task: compute the sample averages in both samples and compare them
- However the variance of the estimated difference might be difficult to estimate due to the two-stage estimation process

- The impact of RGD subsidy to firms seems to have positive effect on employment but no effect on the return to assets

**Table 2: The main impact estimates in the Swedish study on Regional Development Grant**

| Outcome variable | Treated | Control | Estimated impact | Standard error | t-statistics | Treated w/ common support | Treated |
|---|---|---|---|---|---|---|---|
| Employment growth between 2000 and 2003 | 3.197 | -1.500 | 4.697 | 1.975 | 2.38 | 66 | 83 |
| Return on total assets between 2000 and 2003 | -3.788 | -0.840 | -2.947 | 2.859 | -1.03 | 64 | 83 |

Source: Gadd, Hansson and Månsson (2009) "Evaluating the impact of firm subsidy using a multilevel propensity score approach".

# Strength and limitations

- Disadvantages relative to experimental techniques
  1. Assumption of conditional independence
  2. Matching can only estimate the treatment effect where there is overlap between beneficiary and non-beneficiary of the program
- Advantages over experimental techniques
  1. No ethical questions concerning the random support
- Advantages over regression tecniques:
  1. It highlights the problem of common support
  2. No assumption on the functional form of the relationship

# Binary choice dependent variable models



- Linear probability models
- Logit models

# Estimation of the propensity score



- Decisions have to be made in the beginning of the estimation:
  - Estimation method
    - Linear probability model
    - Logit/Probit
  - Variables to include into the regression

# Which variables to choose?

- Goal of estimating propensity score is to eliminate selection bias
- All variables are important that have an impact on
  - Decision to participate AND
  - Result of participation (outcome)
- Omitting important variable leads to a biased estimation
- This requires preliminary research on the participation decision and institutional settings
- Only variables that unaffected by participation should be included
  - Constant variables (do not change during the program)
  - Measured before participation
- Data of beneficiaries and non-beneficiaries should come from the same source

# Estimation methods

- Linear models
    - Linear probability model
- Nonlinear models
    - Logit/Probit models

# Linear probability model

- We want to use regression to explain a qualitative event (participate, not participate)
- Formally we may write the equations of the estimation similarly former OLS estimates:

$$y=\beta_0+\beta_1x_1+\ldots+\beta_kx_k+u$$

- Where now y is a binary variable: it takes either 1 or 0

# Estimated coefficients

- How to interpret the estimated coefficients in this case?
- For the whole sample we might say that the values of y in average changes with the values of x in average, therefore the estimated beta coefficients shows the *probability of succes*.
- Basically what we assume is that the probability of the success is linear function of  the observed explanatory variables.
- Betas called *response probability*
- Based on the estimated coefficients we might ”forecast” y values. These are the predicted probabilities (of success).

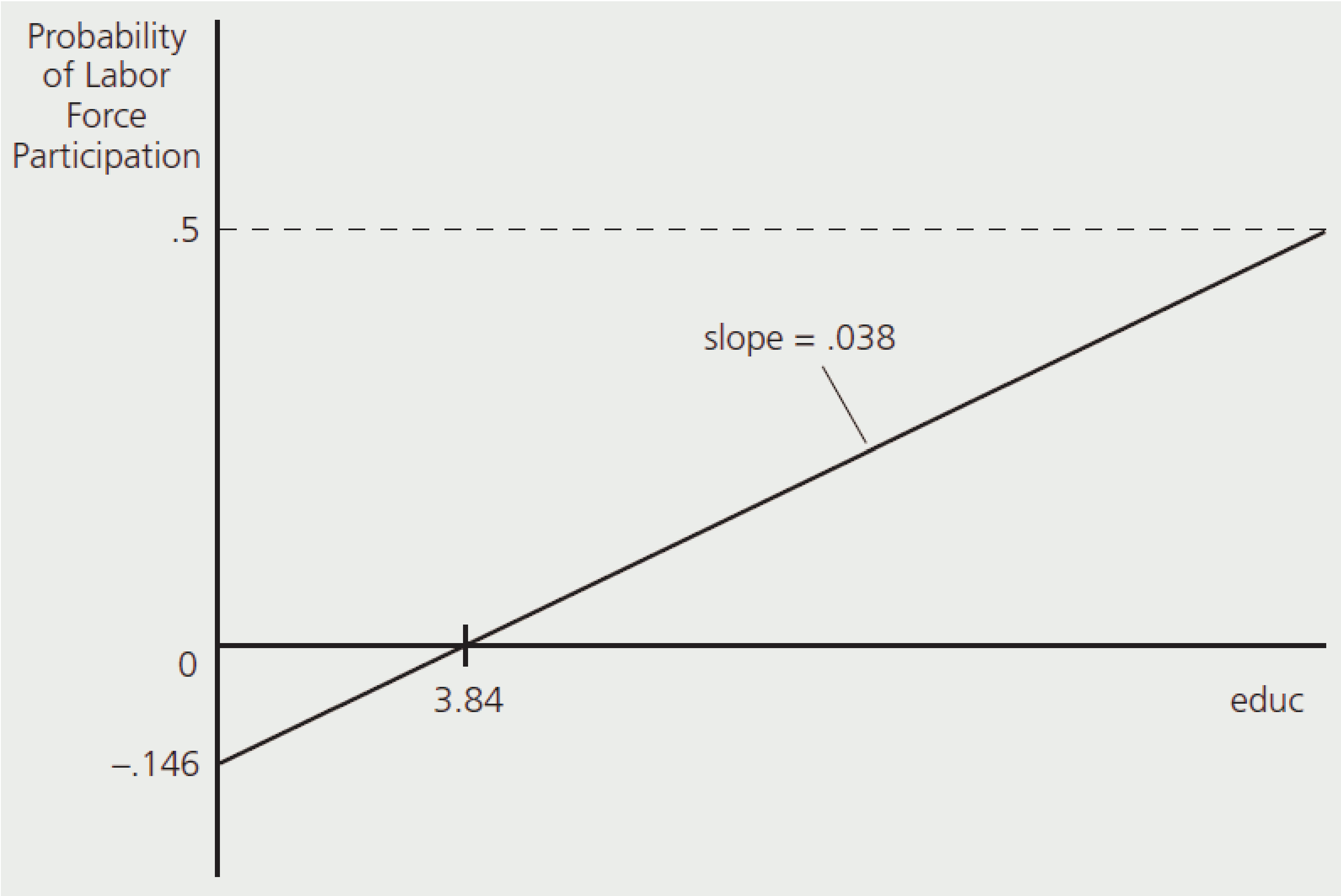
# Example

- Let inlf be a binary variable indicating labor force participation by married woman during 1975.
- We assume that labor force participation depends on other sources of income including husbands' income (nwifeinc), years of education (educ), past years of labor market participation (exper), age, number of children less than six years old (kidslt6) and number of kids between 6 and 18 (kidsge6).
- Based on the data the following equation estimated:

$$\begin{aligned}\widehat{inlf} = \underset{(.154)}{.586} - \underset{(.0014)}{.0034}\, nwifeinc + \underset{(.007)}{.038}\, educ + \underset{(.006)}{.039}\, exper \\ - \underset{(.00018)}{.00060}\, exper^2 - \underset{(.002)}{.016}\, age - \underset{(.034)}{.262}\, kidslt6 + \underset{(.0132)}{.0130}\, kidsge6 \\ n = 753,\ R^2 = .264.\end{aligned}$$

## Estimated relationship between the probability of being in the labor force and years of education, with other explanatory variables fixed





- Advantages:
  1. easy to estimate
  2. easy to interpret
- Disadvantages:
  1. It is easy to have an estimated probability outside (0,1) interval
  2. Probabilities can not be in linear relationship for the explanatory variables for all their possible values.
     - For example it seems reasonable to assume that the first child decreases the probability of labor market participation more, than the second child
- Still this model is useful and often applied

## Exercise: A linear probability model of arrests



- Open CRIME1.GDT!
- Let arr86 be a binary variable equal to unity if a man was arrested during 1986, and zero otherwise.
- The population is a group of young men in California born in 1960 or 1961 who have at least one arrest prior to 1986.
- A linear probability model for describing *arr86* is

$$arr86=\beta_0+\beta_1 pcnv+\beta_2 avgsen+\beta_3 tottime+\beta_4 ptime86+\beta_5 qemp86+u,$$

- *pcnv* is the proportion of prior arrests that led to a conviction,
- *avgsen* is the average sentence served from prior convictions (in months),
- *tottime* is months spent in prison since age 18 prior to 1986,
- *ptime86* is months spent in prison in 1986, and
- *qemp86* is the number of quarters (0 to 4) that the man was legally employed in 1986.

# Exercise: A linear probability model of arrest



- How to interpret the intercept?
- How to interpret the slope for each variables? Do their sign reasonable?

# Nonlinear probability models

- Nonlinear probability models solve the drawbacks of the linear model
- They apply special functional forms in estimation that guarantee for the estimated value to lie between 0 and 1
- General form of the estimator:

$$y = G(\beta_0+\beta_1 x_1+\ldots+\beta_k x_k)$$

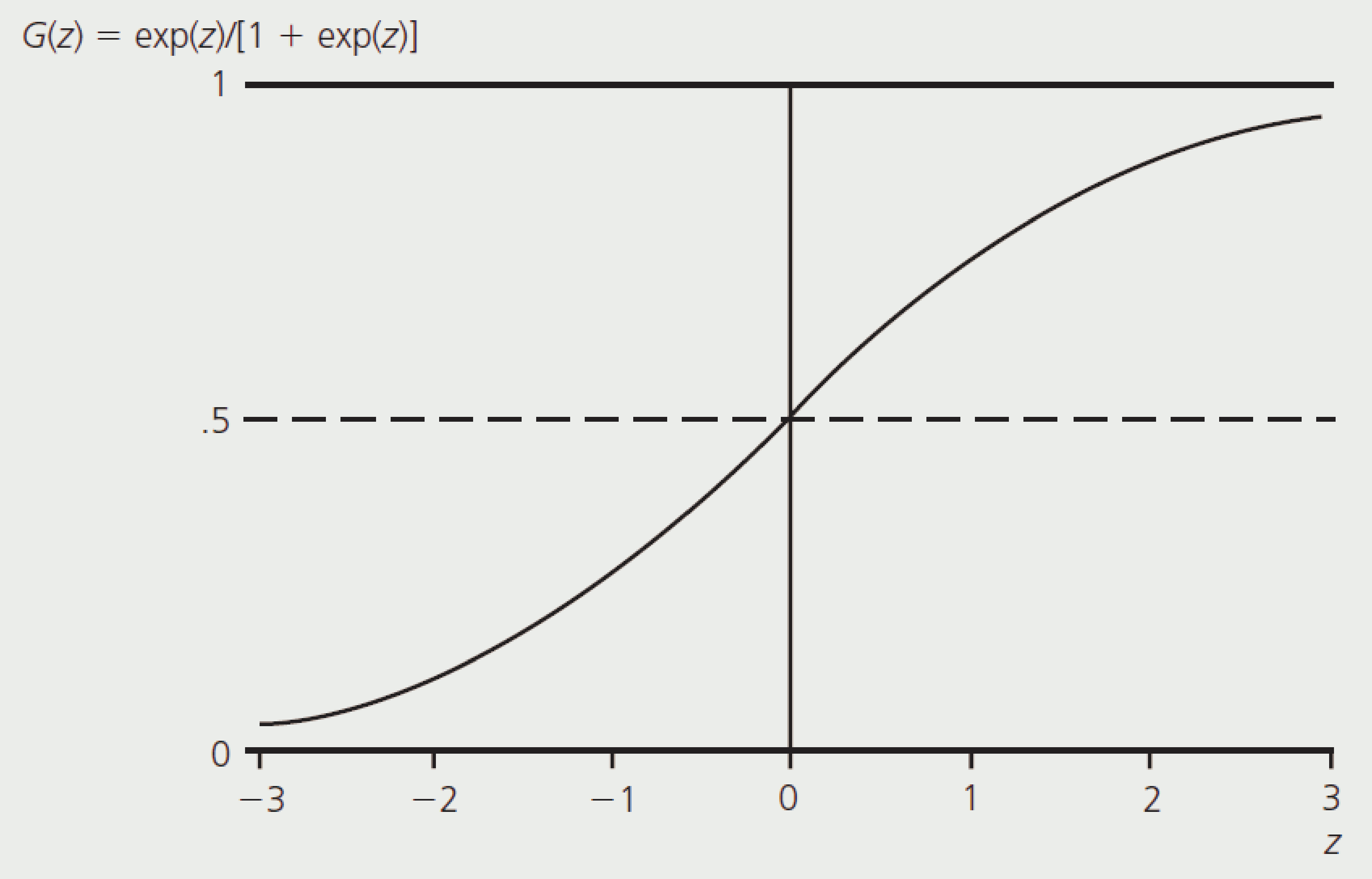
- Where G() is a special non-linear function with values between 0 and 1

- In case of logit model the G is the logistic function:

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

- Probit model is called when G() is the distribution function of the standard normal distribution

# Graph of the logistic function



# Interpretations of the logit and probit models



- These models are common to interpret as models with a latent variable
- That is let y* a latent variable with

$$y^* = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

- whereas the observable y variable is 1 if y*>0 and 0 if y*<0.

- These models are easy to estimate using econometrics softwares. Nonetheless it is very difficult to interpret (especially the estimated coefficients)
- Interpretation of the estimated coefficients:

  *They show the direction of the relationship but not its strength!*

# Goodness of fit

- Hard to check
- Usual approach: percent correctly predicted

- Married women’s labor force participation
- Open MROZ.GDT to estimate the labor force participation model.
- Estimate the probability of married women’s labor force participation rate using LPM, logit and probit models.
- Explanatory variables should be:
    - Nwifeinc
    - Educ
    - Exper
    - Exper^2
    - Age
    - Kidslt6
    - Kidsge6
    - constant

# Application of binary choice models in a matching framework



- Exercise on the effects of job training on earnings

- The data in JTRAIN2.GDT are from a job training experiment in the 1970s. The response variable is real earnings in 1978, measured in thousands of dollars. Real earnings are zero for men who did not work during the year. Training began up to two years prior to 1978.
- The factors influcencing income are real earnings in 1974 and 1975, age (in quadratic form), a binary high school degree indicator (nodegree), marital status, and binary variables for black and Hispanic.

- Estimate the probability of training using different probability models (linear model and probit).
- Forecast these models to have the propensity score.
- Determine the control group based on the propensity score using at least one method.
- Estimate the average treatment effect by calculating the difference between the average of these groups.

AAM CONSULTING

# Thank you for your attention!

Bucharest 30.10.2011