Policy Evaluation in a Nutshell

Joachim Merz*

November 2, 2016

Content

*Univ.-Prof. Dr. Joachim Merz, Leuphana University Lüneburg, Department of Economics, Research Institute on Professions (Forschungsinstitut Freie Berufe, FFB), Chair ‚Statistics and Professions', CREPS (Center for Research in Entrepreneurship, Professions and Small Business Economics), IZA (Institute for the Study of Labour, Bonn), Scharnhorststr. 1, 21332 Lüneburg, Tel.: +49 4131 / 677- 2051, Fax: +49 4131 / 677- 2059, E-Mail: merz@uni.leuphana.de, (www.leuphana.de/ffb).

Policy Evaluation in a Nutshell

Joachim Merz

References: Morgan and Winship (2010), Angrist and Pischke (2009), Angrist and Krüger (2009), Imbens and Wooldridge (2008), Blundell and Dias (2009). Nichols (2007), Caliendo (2006), Caliendo and Hujer (2006), Rossi, Lipsey and Freeman (2004), Heckman, Lalonde and Smith (1999); more non-technical: Bauer, Fertig and Schmidt 2009, Hujer 2011, Schlotter et al. 2010

1 Policy and Program Evaluation – General Concern, Causality, Counterfactual State and Non-Experimental Studies

Policy evaluation - The range of causal questions

“The most challenging empirical questions in economics involve "what if" statements about counterfactual outcomes. Classic examples of "what if" questions in labour market research concern the effects of career decisions like college attendance, union membership, and military service.

Interest in these questions is motivated by immediate policy concerns, theoretical considerations, and problems facing individual decision makers. For example, policy makers would like to know [whether a labour market program will reduce unemployment] whether military cutbacks will reduce the earnings of minority men who have traditionally seen military service as a major career opportunity. Additionally, many new high school graduates would like to know what the consequences of serving in the military are likely to be for them. Finally, the theory of on the job training generates predictions about the relationship between time spent serving in the military and civilian earnings.

Counterfactual states and potential outcomes

Regardless of the motivation for studying the effects of career decisions, *the causal relationships at the heart of these questions involve comparisons of counterfactual states of the world.* Someone - the government, an individual decision maker, or an academic economist - would like to know what outcomes would have been observed if a variable were manipulated or changed in some way.

Lewis's (1986) study of the effects of union wage effects gives a concise description of this type of inference problem (p. 2): "At any given date and set of working conditions, there is for each worker a pair of wage figures, one for unionized status and the other for non-union status". Differences in these two *potential outcomes* define the causal effects of interest in Lewis's work, which uses regression to estimate the average gap between them (See also Rubin (1974, 1977) and Holland (1986) for formal discussions of counterfactual outcomes in causal research). …

Randomized experiments and non-experimental, observational studies

Even if ambiguities in the definition of counterfactual states can be resolved, it is still *difficult to learn about differences in counterfactual outcomes because the outcome of one scenario is all*

that is ever observed for any one unit of observation (e.g., a person, state, or firm). Given this basic difficulty, how do researchers learn about counterfactual states of the world in practice? In many fields, and especially in medical research, the prevailing view is that the best evidence about counterfactuals is generated by randomized trials because randomization ensures that outcomes in the control group really do capture the counterfactual for a treatment group. …

Unfortunately, economists rarely have the opportunity to randomize variables like educational attainment, immigration, or minimum wages. Empirical researchers must therefore rely on observational studies that typically fail to generate the same force of evidence as a randomized experiment.

But the object of an observational study, like an experimental study, can still be to make comparisons that provide evidence about causal effects. *Observational studies attempt to accomplish this by controlling for observable differences between comparison groups using regression or matching techniques, using pre-post comparisons on the same units of observation to reduce bias from unobserved differences, and by using instrumental variables as a source of quasi-experimental variation."*
Angrist and Krüger (1999, 1282-1284)

Some program/policy evaluation examples to illustrate the wide range of applications:

- Do German Welfare-to-Work Programmes reduce welfare dependency and employment (Huber, Lechner, Wunsch and Walter 2010)?
- Homogene und heterogene Teilnahmeeffekte des Hamburger Kombilohnmodells: Ein Verfahrensvergleich von Propensity-Score Matching und linearer Regression (Pfeiffer 2009)
- Human Capital Investment in Children: A Comparative Analysis of the Role of Parent-Child Shared Time in Selected Countries (Österbacka, Merz and Zick 2012)
- Earnings and employment effects of off-the job training in East Germany after unification (Lechner 1999)
- How does education effects income (Harmon and Walker 1995)?
- Do gifts have an influence on recruiting donations (Falk 2007)?
- Does a voluntarily military service have impacts on later income (Angrist 1998)?
- …

2 Alternative Approaches to Policy Evaluation

Under a general perspective the evaluation of any policy encompass three main steps (Caliendo and Hujer 2006): First, an impact analysis on the individual (Microeconomic Evaluation); second, it should be examined if the impacts are large enough and yield net social gains for the society (Macroeconomic Evaluation); third, it should be examined if the best outcome that could have been achieved for the money spent (Cost-Benefit Analysis). A broad overview of evaluation in general with many policy evaluation examples is provided by Rossi, Lipsey and Freeman 2004.

We focus on the microeconomic evaluation with its main question if the outcome for an individual is affected by a participation on a programme (e.g. a labour market program) or not.

The outcome difference of the participant's actual situation compared to the outcome if he had not participated is in the focus of interest.

The most popular policy evaluation methods in empirical microeconomics are discussed by Blundell and Costa Dias (2009). They distinguish: social experiments, natural experiments, matching, instrumental variables, discontinuity design, and control functions (regression). The first will be closest to "theory free" methods relying on randomized assignment (treatment) rules. The control function approach will be closest to the structural econometric approach with direct modelling the assignment rule to control for selection in observational (non-experimental) data.

Social experiments: Constructs a control (comparison) group as a randomized subset of the eligible population. Most convincing method; problem of "randomized in" may don't want to participate; no placebo treatment possible.

Natural experiments: control group found by a "natural" experiment like a policy program with randomized selected regional grouping, or a certain law starting by a certain date ("Kündigungsschutzgesetz") with new valid entitlements.

DID methods compare the difference in average behaviour before and after the reform for the eligible group with the before and after contrast for a comparison group.

Discontinuity design methods can also be classified as a natural experiment but one where the probability of enrollment into treatment changes discontinuously with some continuous variable. Eligibility for a scholarship only if a certain percentage of test scores are given; or job protection law is taking effect after a certain number of employees in a firm.

Matching: Line-up comparison individuals according to sufficient observable factors to remove systematic differences in the evaluation outcome between treated and nontreated. Find a statistical twin ideally according to all characteristics; Different measures of similarity: nearest-neighbour method, radius-matching etc. If there are many observable factors for identification then a sample has to be sufficient large to meet a statistical sibling in all its dimensions (dimension problem). The treatment effect then might be evaluated as some aggregates of the individual twin differences.

Instrumental variables (IV): standard econometric approach to endogeneity. The instrument is a variable which serves as a substitute for the treatment in the outcome equation. The IV is a determinant of the assignment rule but is not correlated to the outcome and other observable and not observable variables in the outcome equation. The IV estimator identifies the treatment effect removed of all the biases that emanate from a nonrandomized control.

Control function method: It uses the full specification of the assignment rule together with an excluded "instrument" when included in the outcome regression equation controlling for endogenous selection. The control function method is directly related to the Heckman selectivity estimator.

The different approaches might be organized also by selection on observables or selection on unobservables, like

Selection on observables
In the selection on observables case it is assumed that all variable which have an influence on the outcome variable as well as on the participation of a measure are observable (unconfoundness or conditional independence assumption). With the assumption that the potential outcomes as well as the treatment variable are only dependent on observables then both are independent when knowing the observables.

Matching
Regression respecting causality, control function method

Selection on unobservables

Cross Section Comparison
Before and after
Differences-in Differences
Instrument variables
Regression discontinuity design
Panel analyses

Single evaluation methods will be discussed in the subsequent chapters.

3 Program Evaluation - The Framework

The framework of program evaluation will be discussed under the basic topics

- Program Evaluation, Treatment and Potential Outcomes,
- Causal Effect of a Treatment,
- The Fundamental Evaluation Problem,
- Treatment Effects and Their Measuring in Principle
- Observed Difference of Treated and Nontreated Individuals, Average Treatment Effect on the Treated (ATT) and Selection Bias
- Why Random Assignment of an Experiment Eliminates the Selection Bias

In section 4 then an overview of different appropriate identifying methods with their respective treatment effects calculations is given.

3.1 Program Evaluation, Treatment and Potential Outcomes

The central program evaluation question is: How large is the impact of a program to a certain output, how large is the causal effect of a treatment?

Health example: Do hospitals make people healthier (Angrist and Pischke 2009)?

Potential outcomes
The cross-road situation, if one or the other way somebody had gone, the ‘what if’ situation, irrespective of whether he or she actually went, is connected with potential outcomes of the alternative paths.

The answer to the question of the impact of a program, of the effect of a treatment on an outcome y_i is addressed by: What might have happened to someone who was treated by something if that person had not been treated (counterfactual situation), and vice versa (potential outcome concept, Rubin 1974).

Thus and as Rubin 2004, p. 344 pointed out "The critical inferential idea … is to view the problem of causal inference as one of drawing an inference about the missing value".

Health example: The answer to the question of an effect on the health status y_i is addressed by: What might have happened to someone who went to the hospital if that person had not gone, and vice versa.

For an individual i, the potential outcome (Rubin 1974) (*potential health status*) is dependent on the occurrence of an event (treatment, $D_i = 1$, *hospital treatment*) or no occurrence (no occurrence, $D_i = 0$, *no hospital treatment*).

Observable and unobservable treatment situations are summarized in Table 1 with actual treatment if $D_i = 1$ (and no treatment $D_i = 0$ and with outcome after treatment by y_{1i} respectively before treatment by y_{0i}.

Table 1: Observable and Unobservable Treatment Situations

	Participation $D_i = 1$	Non-Participation $D_i = 0$
Outcome after treatment y_{1i}	observable	unobservable (counterfactual)
Outcome before treatment y_{0i}	unobservable (counterfactual)	observable

3.2 Causal Effect of a Treatment

What is the causal effect of a treatment, the difference between after versus before a treatment, the difference between y_{1i} and y_{0i}?

Health example: What is the difference between y_{0i}, the health status if he had not gone to the hospital, and y_{1i} if he goes.

The observed outcome (*health status*) is a composed outcome of the before and after situation:

$$\text{Observed outcome } y_i = \begin{cases} y_{1i} & \text{if } D_i = 1 \\ y_{0i} & \text{if } D_i = 0 \end{cases}$$

$$y_i = y_{0i} + (y_{1i} - y_{0i})D_i \quad \text{or} \quad y_i = D_i y_{1i} + (1 - D_i) y_{0i}$$

The **causal effect of a treatment** (*hospitalized*) is the individual difference

$$y_{1i} - y_{0i};$$

it is the effect if we could go back in time and change *that* person's treatment.

Concentrating on a single individual requires that there is no effect by the participation of another individual (stable unit treatment value assumption, SUTVA); any cross and general equilibrium effects then are excluded.

3.3 The Fundamental Evaluation Problem

The same person i, however, cannot be in both states, be treated *and* nontreated. Thus the individual causal effect can never be observed. What is observable is only that outcome which is composed by the treatment effect (D_i=1) and the situation without treatment (D_i=0).

As Caliendo and Hujer (2006) state: "The fundamental evaluation problem arises because we can never observe both states (participation and non-participation) for the same individual at the same time, i.e. one of the states is counterfactual. Therefore finding an adequate control group and solving the problem of selection bias is necessary to make a comparison possible".

Because observed as well as the potential outcome for both situations can never be seen for any **one** person i, the causal effect of a treatment has to be analysed by the *averages* of those who were and were not treated. So, the average program effect of a group will be of central importance.

3.4 Treatment Effects and Their Measuring in Principle

In an experimental setting as a social or natural experiment, typically the only two quantities to be estimated are the sample Average Treatment Effect (ATE) or the population ATE. Both are measured as a difference in averages of the treated and nontreated.

In a nonexperimental setting several other ATEs are commonly of interest: the ATE on the treated (ATT), the ATE on the nontreated or control group (ATNT), the ATE for the whole population and a variety of local ATEs (LATE) - local to some range of values or some subpopulation. One might construct at least 2^N different ATE estimates in a sample of N observations, restricting attention to two possible weights for each observation (Nichols 2007, 511).

The marginal treatment effect (MTE) in addition measures the effect of an additional small amount of a continuous treatment.

3.5 Observed Difference of Treated and Nontreated Individuals, Average Treatment Effect on the Treated (ATT) and Selection Bias

Two treatment effects are prominent in empirical studies: The population average treatment effect (ATE) and the average treatment effect of the treated (ATT).

The **population average treatment effect (ATE)** is

$$ATE = E(y_{1i}) - E(y_{0i})$$

which is the effect if individuals in the population were randomly assigned for treatment.

Of special interest however is the group which participates on a program. With focus on this group and neglecting further influences on all other (neglecting general equilibrium effects) the 'treatment on the treated' is of special interest and discussed in the following.

The observed difference of the average outcomes of the treated and the nontreated is formally composed by

$$E(y_i \mid D_i = 1) - E(y_i \mid D_i = 0) = E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 1) + E(y_{0i} \mid D_i = 1) - E(y_{0i} \mid D_i = 0)$$

where the second and third term is added not changing the equation. Now the average treatment effect on the treated (ATT) plus the selection bias is separated.

Average treatment effect on the treated (ATT)
The first right hand side term

$$ATT = E(y_{1i} - y_{0i} \mid D_i = 1) = E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 1)$$

is the important **average causal effect of a treatment on those who are treated (ATT)**, describing the difference of the treated $E(y_{1i} \mid D_i = 1)$ (*hospitalized*) and what would have happened to the same persons if not treated $E(y_{0i} \mid D_i = 1)$. ATT thus quantifies how large would be the expected effect of the treated against what would be expected if this group would not be treated.

The observed average outcome difference of the treated and untreated ($E(y_{1i} - y_{0i} \mid D_i = 1)$) is biased when selectivity into the treatment is given, when there is a selection bias (second right hand side term).

Selection bias
The second right hand side term, the selection bias,

$$E(y_{0i} \mid D_i = 1) - E(y_{0i} \mid D_i = 0),$$

is the difference of the average outcome of the nontreated (control group) when we would have treated the nontreated compared to the average outcome of the nontreated control group. It is the difference in average y_{0i} between those who and those who were not treated (*hospitalized*). The selection bias arises because participants and non-participants are selected groups with different outcomes even when the programme is absent.

The question is, when the self-selection bias of the decision to be treated *(to go to the hospital)* will vanish, i.e., that the treated group then would have the same structure (apart from the treatment) as the control group who is not treated (*did not go to the hospital*). Then a comparison of both groups with the same structure would deliver the desired causal effect.

Selectivity bias in other words: if certain characteristics of the persons will influence both, the participation on a program (to be treated) as well as the output (*health status*), then the output of both groups will be different regardless any treatment. Then the selectivity bias will forbid the simple comparison of the average output indicator of the participants and the non-participants.

Health example: Self-selection possibility in the hospital situation: "Because the sick are more likely than the healthy to seek treatment, those who were hospitalized have worse values of, making selection bias negative in this example" Angrist and Pischke 2009, 15.

3.6 Why Random Assignment of an Experiment Eliminates the Selection Bias

The random assignment of the treatment D_i makes D_i independent of potential outcomes. What is the effect on the selection bias? The right hand side of the observed differences between treated and nontreated (see above)

$$\begin{aligned} E(y_i \mid D_i = 1) - E(y_i \mid D_i = 0) = {} & E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 1) + \\ & E(y_{0i} \mid D_i = 1) - E(y_{0i} \mid D_i = 0) \end{aligned}$$

can be reduced back to

$$E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 0).$$

Because of the independence of y_{0i} and D_i we can swap $E(y_{0i} \mid D_i = 1)$ for $E(y_{0i} \mid D_i = 0)$ and the average treatment effect *with experimental data* will simplify further to

$$\begin{aligned} E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 0) &= E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 1) \\ &= E(y_{1i} - y_{0i} \mid D_i = 1) \\ &= E(y_{1i} - y_{0i}) \end{aligned}$$

The effect of randomization thus has eliminated the selection bias and the average treatment effect can be measured by the average observable outcomes of the participants of a program (treated) minus that of the non-participants (nontreated).

The empirical problem to identify the treatment effect
Unfortunately, most economic data are non-experimental with no random assignment of the treatment. Therefore the selection bias might prohibit to take just the average outcome $E(y_{0i} \mid D_i = 0)$ of the non-participants, the simplest control or counterfactual group, as an estimate for $E(y_{0i} \mid D_i = 1)$ of the control group of non available 'otherwise comparable' persons.

The challenge with non-experimental, observational data thus is to avoid the selection bias. There are several program evaluation approaches to do this like a before-after comparison, a difference-in-difference estimator, a fixed effect approach or a matching approach by propensity scores.

4 Evaluating Treatment Effects by Different Methods

As shown, the selection bias is eliminated, when an adequate control group is found whose characteristics – despite the treatment – are similar to the treated. Then the characteristics which influence the treatment and the output indicator should not be different on average of the treated as well as of the nontreated (no selectivity bias). And, as a solution to the empirical problem to measure the causal effect, the observed outcomes then can be used to determine the average treatment effect.

We discuss in the following some evaluation methods for constructing the adequate control group using different identifying assumptions for the hypothetical population based on the observed population. As mentioned, they might be divided by methods with *selection on observables*, like matching (exact or by balancing scores like the propensity scores) and regression, and with *selection on unobservables*, too (like before and after, difference-in-differences, instrumental variables, regression discontinuity models and panel analyses).

4.1 General Matching Approaches

One possibility to find an adequate control group is a **matching procedure** where each participant in a program (treated) is assigned/matched to a control group person (nontreated) who has the same characteristics X. If such pairs of treated and untreated persons with similar characteristics different only by treatment or not, then the difference of the averages of the respective outcomes will measure the average treatment effect (ATT) without selection bias.

Matching requires the identifying assumption that the outcome is independent of being a participant or non-participant (D) condition on some characteristics, covariates X, which is called **unconfoundedness** (with Π or $\perp$ for independence)

$$(D_i \perp (y_{0i}, y_{1i}) \mid X).$$

Unconfoundedness implies that there are no unobserved characteristics which are correlated with the treatment as well as with the potential outcome. In particular, the potential outcome after

being treated y_{1i} is assumed to be independent of the participation decision. This assumption is also known as **selection on observables** or as **conditional independence assumption** (CIA).

A second assumption is also required: **Overlap / Common support**

$$0 < \Pr(D_i = 1 | X_i = x) < 1 \quad for\ all\ x,$$

which means that for any value of x there are observations with and without a treatment.

Then **the average treatment effect of the treated ATT^{match}** will be

$$ATT^{match} = E(y_{1i} | X, D_i = 1) - E(y_{0i} | X, D_i = 0)$$

where the first term can be estimated from the treatment group and the second term form the matched comparison group.

An **exact matching** is given if indeed all covariates X are identical for the treated as well as for the untreated. However, conditioning on all relevant covariates is restricted in case of a high dimensional vector X. If they are not exactly identical, then, based on some balancing scores, there are several **other matching procedures** which minimize a respective distance, e.g. nearest-neighbour matching, radius matching or kernel matching. The choice which one has to be chosen is not trivial because there is a trade-off between bias and variance (see the overview by Smith and Todd 2005).

4.2 Matching by Propensity Scores

The propensity score method adjusts for selection bias caused by observed variables by balancing treatment and comparison groups on a set of covariates, which are reduced to a single variable. This method may be preferable to the general matching method with its dimensionality problem, and regression because it does not rely on a linear functional form to adjust for potential confounding variables.

The **propensity score method (PSM) (**Rosenbaum and Rubin 1985, Guo and Fraser 2010 for a recent book about) uses matching to find the appropriate pairs by reducing the set of characteristics X to a single variable, a one dimensional probability $P(D_i=1|X)$, where the probability could be estimated by a Probit (or Logit) regression approach. Several mentioned matching methods, like nearest neighbour, Caliper, Radius or Kernel (Caliendo and Kopeinig 2008, Smith and Todd 2005), then bring together the most similar treated and untreated persons, which are represented by their propensity scores.

4.2.1 Propensity Score Approach – CIA, Common Support and SUTVA

To meet the causal effects of a treatment by the propensity score method several assumptions must be hold: the Conditional Independence Assumption (CIA), the Common Support (overlapping) Condition and the Stable Unit-Treatment Value Assumption (SUTVA).

The **Conditional Independence Assumption (CIA)** states that the potential outcome is statistical independent of the treatment assignment. If the CIA holds, i.e. the selection process into the program and control group is completely explained by the characteristics X (selection on observables), then the outcome indicator of the nontreated will be independent of the participation status $(y_{0i} \perp \mathrm{D} \mid X)$. Given X then

$$E(y_{0i} \mid D_i = 1, X) = E(y_{0i} \mid D_i = 0, X).$$

Then the selection bias vanishes and the average difference of the treated group can be compared to the nontreated control group, the average causal effect of the treatment then is found.

Rosenbaum and Rubin (1983a) now have shown that under CIA (unconfoundedness given by $(D_i \perp (y_{0i}, y_{1i}) \mid X)$) the independence of potential outcomes and treatment indicators also holds if matching is done by propensity scores (binary decision). The same propensity scores of the treated and nontreated vipes out the selectivity bias problem and thus equates to the random approach.

"Hence, within subpopulations with the same value for the propensity score, covariates are independent of the treatment indicator and thus cannot lead to biases (the same way in a regression framework omitted variables that are uncorrelated with included covariates do not introduce bias). Since under unconfoundedness all biases can be removed by adjusting for differences in covariates, this means that within subpopulations homogenous in the propensity score there are no biases in comparisons between treated and control units." Imbens and Wooldridge (2008, pp 31).

Given the CIA, then the potential outcome is independent of the participation assignment given by $P(D_i=1|X)$ and

$$E(y_{0i} \mid D_i = 1, P(D_i = 1 \mid X)) = E(y_{0i} \mid D_i = 0, P(D_i = 1 \mid X))$$

and the average causal effect then is the difference of the average outcome of the treated compared to the average observed outcome of the control group, the nontreated

$$\begin{aligned} &E(y_{1i} \mid D_i = 1, P(D_i = 1 \mid X)) - E(y_{0i} \mid D_i = 0, P(D_i = 1 \mid X)) = \\ &E(y_{1i} \mid D_i = 1, P(D_i = 1 \mid X)) - E(y_{0i} \mid D_i = 1, P(D_i = 1 \mid X)) \\ &\qquad = E(y_{1i} - y_{0i} \mid D_i = 1) \\ &\qquad = E(y_{1i} - y_{0i}) \end{aligned}$$

Within an empirical application the problem is to find such a vector X (selection on observables): An expansive list of covariates in the propensity score matching method guards against broader forms of omitted variable bias, the selection bias. Although it is less likely to violate the conditional independence assumption, a greater number of covariates may narrow the range of common support, i.e. may diminish the possibility to find a control group person.

Besides the weaker Conditional Mean Independence Assumption (CMIA) by the condition on $P(D_i=1|X)$ two additional assumptions must hold to apply the propensity score approach: the Common Support Condition and the Stable Unit-Treatment Value Assumption (SUTVA).

The **Common Support (Overlap) Condition** asks – for persons with identical characteristics – to have positive probabilities for a participant as well as for a member of the control group.[1]

The **Stable Unit-Treatment Value Assumption (SUTVA)** requires that there is no spillover effect, i.e. the outcome and the own participation of a person depends on the own participation only and not on the treatment of the other persons.

4.2.2 Propensity Score Approach – Procedure

Once these assumptions are fulfilled then the **average treatment effect of the treated** can be estimated by the propensity score approach and is defined by the difference of the average outcome of the treated and the assigned nontreated which have the same (or near) participation probability:

Average treatment effect of the treated ATT^{PSM}

$$ATT^{PSM} = E(y_{1i} \mid D_i = 1, P(D_i = 1 \mid X)) - E(y_{0i} \mid D_i = 0, P(D_i = 1 \mid X)),$$

where the propensity score is estimated by a logit or probit procedure. It is discussed if the propensity score is specified just for estimation a best fit or if it is driven by a respective substantive theory.

Thus, after matching a treated individual with the "statistical sibling" (control, with the same or near propensity score as the treated individual) either the mean of the individual outcome differences of the treated and the control group or the difference of both means provide the **average treatment effect of the treated ATT^{PSM}** .

Summarizing the propensity score method:
"The propensity score represents the predicted probability of participating in the treatment, based on the observed and measured characteristics used in the prediction equation. Each member of the sample receives a propensity score, which will range between zero and one, representing either a low or high likelihood of receiving the treatment. The probability of receiving treatment is estimated as a function of the observed characteristics one wishes to balance between the treatment and comparison group. Once the propensity score is calculated, a treatment group member can be paired to a comparison group member who has a similar propensity score.

A key feature of the propensity score method is that matching is done only on observed characteristics; unlike randomization, propensity scores cannot match on unobserved characteristics. Propensity scores address the bias from unobserved characteristics by assuming that all factors related to selecting into a treatment […] are observed and measured.

[1] ATT overlapping: $0 < P(D_i = 1 \mid X) < 1$. See Caliendo (2006, 4, footnote 6 and Wooldridge (2002)

If these factors are taken into account, then conditioned on these factors, an outcome is said to be "ignorable" of treatment status (Rosenbaum and Rubin 1983, 42). That is, an outcome is not influenced by those same factors that also influenced someone to take up a treatment. This assumption, which is variously known as "selection on observables" (Heckman and Vytlacil 2001, 107), the "unconfoundedness assumption" (Imbens 2004, 5), or the "conditional independence assumption" (Black and Smith 2004, 102), is critical to the validity of propensity scores. If this assumption is violated and unobserved characteristics influence both the use of [the treatment and the outcome], the treatment and comparison groups may differ in unobserved ways, and between-group differences may reflect those characteristics rather than the treatment (Bryson, Dorsett, and Purdon 2002).

… The treatment effect can be estimated only for the common support [to meet the counterfactual person], the area of overlap between the propensity scores (Imbens 2004; Smith and Todd 2005)." (Gibson-Davis and Foster 2006, 8)

4.2.3 Propensity Score Applications

- The outcome of coaching and training for self-employment. A statistical evaluation of outside assistance support programs for unemployed business founders in Germany: Oberschachtsiek, D. and P. Scioch (2015),
- Human Capital Investments in Children: A Comparative Analysis of the Role of Parent-Child Shared Time in Selected Countries: Österbacka, Merz and Zick (2012)
- Do German Welfare-to-Work Programmes Reduce Welfare Dependency and Increase Employment? Huber, Lechner, Walter and Wunsch 2010
- Teilnahmeeffekte des Hamburger Kombilohnmodells: Pfeifer 2009
- Evaluation of German Labour Market Program Effects: Caliendo, Hujer and Thomsen (2004)
- Food stamp effects on food insecurity: Gibson-Davis and Foster 2006
- …

4.3 Classical Linear Regression Approach and Treatment Effect

We discuss regression and treatment first under the substantive problem behind, then define the average treatment effect under regression, discuss a two step selectivity correction treatment effects estimation and the panel data fixed effects regression to circumvent the selection bias and provide regression applications.

4.3.1 Regression and Treatment – The Substantive Problem Behind

A treatment D might be specified as a dummy in the Classical Linear Regression context with some explanatory variables x_i and respecting an error term ε_i via

$$y_i = X_i\beta + D_i\alpha + \varepsilon_i$$

where the estimate of α is the difference between the mean (expectation) of y_i for $D_i = 1$ (the treatment group) and the mean of y_i for $D_i = 0$ (the control group, non-treated). If and only if D is not correlated with the error term, then α would be the average treatment (ATE). However,

and this will be the topic of the next paragraphs, it is to be expected that D_i is correlated with the error term; an OLS estimate then would result in a bias treatment effect estimation.

The selection on observables implicates that the Classical Linear Regression (CLR) assumption $E(\varepsilon_i \mid X_i, D_i) = 0$, i.e. that the error term is uncorrelated with the X_is and the treatment D_i has to be valid for unbiased estimates. In addition, the effect of the treatment D_i and the variables X_i on y_i has to be linear, additive separable and constant over all observations. If all of this holds, then a counterfactual for an observation might be built by a linear combination of all observations even if there is no observation of the respective opposite treatment status (no common support).

However, if a treatment effect is specified simply as a dummy in the outcome equation biased results under the Classical Linear Regression assumptions have to be expected because of possible correlation with the error term and the outcome itself.

A key assumption is that observable characteristics are the only reason why the treatment and the error term are correlated (selection on observables, unconfoundedness, conditional independence assumption (CIA)).

Take as an example the study: Human Capital Investments in Children: A Comparative Analysis of the Role of Parent-Child Shared Time in Selected Countries by Österbacka, Merz and Zick (2012). With their hypothesis: Parents spend more time for a specific leisure/cultural activity when a child is present, the treatment in the linear regression approach would be a child dummy

time spent in a certain leisure/cultural activity = f(X, child dummy, ε).

The CLR assumption would be that X and the child dummy are not correlated with the error term ε. However, if there would be a correlation, than we face an *endogeneity problem* (see Excursus Endogeneity) and the model would be misspecified (omitted variable problem, selectivity problem, unobserved heterogeneity) with biased coefficients.

Is the presence of a child effect causal when measured by the coefficient of the child dummy (treatment)?

Within such a 'normal' regression approach it is disputable, probably not: Due to self-selection this effect might be spurious: Higher educated men/women may select themselves (or are selected) to have no children (influence on the treatment, child dummy). In addition higher educated men/women might spend less time for a leisure activity because of time intensive working hours (influence on the outcome, time spent in an activity). If education is not included in X (or in the case of neglected ability which is (often) not measured at all) there is potential for an *omitted variable bias* (unobserved heterogeneity) with the consequence of misleading results and downward (or upward) bias of the estimated coefficients.

The omitted variable bias is a bias arising from unobserved and uncontrolled differences in the outcome potential between the treatment and non-treatment group. The omitted bias is the

selection bias which is - like in the other causality approaches - of particular concern in the causality analysis with the regression and treatment effects approach.

4.3.2 Regression and Treatment Effects

There is a solution to handle the selection problem: The mentioned abbreviated **control function estimator**, based on Heckman 1976, 1979) but extended, takes the selection rule explicit into consideration and handles the endogeneity of D (D is correlated with the error term ε_i) problem in the linear regression framework (Heckman, Lalonde and Smith 1999, Blundell and Costas Dias 2009). A key assumption is that observable characteristics are the only reason why the treatment and the error term are correlated (selection on observables, unconfoundedness, conditional independence assumption (CIA)).

Let us first detect the selection problem within the regression approach combined with the potential outcome approach and its treatment effect.

Regression and the selection problem

The observed outcome as in chapter 3.2 is

$$y_i = y_{0i} + (y_{1i} - y_{0i})D_i \quad \textit{respectively} \quad y_i = y_{0i} + D_i(y_{1i} - y_{0i})$$

with $(y_{1i} - y_{0i})$ as the causal effect of the treatment. Because any regression is formulated as

$$y_i = E[y_i] + (y_i - E[y_i])$$
$$y_i = E[y_i] + \varepsilon_i = \mu(X_i) + \varepsilon_i = X_i\beta + \varepsilon_i$$

with the error term ε_i as the difference between the expected (estimated) value and its observed value, we can write the single regressions on the treated y_{1i} and the non-treated y_{0i} as

$$y_{1i} = \mu_1(X_i) + \varepsilon_{1i} = X_i\beta_1 + \varepsilon_{1i}$$
$$y_{0i} = \mu_0(X_i) + \varepsilon_{0i} = X_i\beta_0 + \varepsilon_{0i}.$$

with $\mu_j(X_i)$ as the expectation of y_i and the parameters β_j to be estimated $(j = 0,1)$. If we assume constant treatment effects $(y_{1i} - y_{0i}) = \alpha$, effects which would be the same for everyone, and impute the regression for y_{0i} into the observed outcome equation we get

$$\begin{aligned} y_i &= y_{0i} + D_i(y_{1i} - y_{0i}) \\ &= \mu_0(X_i) + D_i\alpha + \varepsilon_{0i}. \end{aligned}$$

The respective conditional equations for $D_i = 1$ and $D_i = 0$ are

$$E(y_i \mid D_i = 1) = \mu_0(X_i) + \alpha + E(\varepsilon_{0i} \mid D_i = 1)$$
$$E(y_i \mid D_i = 0) = \mu_0(X_i) \quad + E(\varepsilon_{0i} \mid D_i = 0).$$

The difference of both is the **general treatment effect (ATE)**, which as in chapter 3.2 is the average outcome of the treated minus the average outcome of the non-treated

$$\begin{aligned} & E(y_i \mid D_i = 1) - E(y_i \mid D_i = 0) \\ & = \mu_0(X_i) + \alpha + E(\varepsilon_{0i} \mid D_i = 1) - \mu_0(X_i) - E(\varepsilon_{0i} \mid D_i = 0) \\ & = \quad \alpha \quad + \quad E(\varepsilon_i \mid D_i = 1) - E(\varepsilon_i \mid D_i = 0) \\ & \quad \textit{treatment effect} \qquad \textit{selection bias.} \end{aligned}$$

The selection bias thus amounts to correlation between the regression error ε_i term and the regressor D_i. Equating the selection bias formulation as in chapter 3.2 to this result

$$E(\varepsilon_i \mid D_i = 1) - E(\varepsilon_i \mid D_i = 0) = E(y_{0i} \mid D_i = 1) - E(y_{0i} \mid D_i = 0)$$

it will be obvious that "this correlation reflects the difference in (no treatment) potential outcomes between those who get treated and those who don't" (Angrist and Pischke 2009, pp.22, 23) (*the hospitalized have poorer health outcomes in the no-treatment state*).

Regression and individual treatment effects

So far we combined the potential outcome approach with regression and showed that the selection bias results in the correlation between the treatment and the error term in a regression approach.

Let us now formulate the general regression approach without initially assuming constant treatment effects following Heckman, Lalonde and Smith (1999).

The imputation of the y_{0i} and y_{1i} regressions in the observed outcome equation yields

$$\begin{aligned} y_i &= y_{0i} \qquad\qquad + D_i(y_{1i} - y_{0i}) \\ &= \mu_0(X_i) + \varepsilon_{0i} + D_i[\mu_1(X_i) + \varepsilon_{1i} - \mu_0(X_i) - \varepsilon_{0i}] \\ &= \mu_0(X_i) \qquad + D_i[\mu_1(X_i) - \mu_0(X_i) + \varepsilon_{1i} - \varepsilon_{0i}] + \varepsilon_{0i} \\ &= X_i\beta_0 \qquad\quad + D_i[X_i(\beta_1 - \beta_0) + \varepsilon_{1i} - \varepsilon_{0i}] \qquad + \varepsilon_{0i}. \end{aligned}$$

The **individual treatment effects** is the expression to $D_i = 1$

$$T(X_i) = y_{1i} - y_{0i} = X_i(\beta_1 - \beta_0) + \varepsilon_{1i} - \varepsilon_{0i}.$$

Thus the individual treatment effects are a mixture of explanatory variables and the error terms, which is not a standard regression problem.

Different assumptions about the single components of the coefficient to D_i will result in different regression models.

Regression and the average treatment effect with homogeneous treatments

For randomly picking a person with X and moving that person from a non-treated to a treated situation the average treatment effect (Heckman, Lalonde and Smith (1999, p.21) is

$$E(y_{1i} - y_{0i} \mid X_i)$$

with the observed outcome regression (as above but rearranged)

$$\begin{aligned} y_i &= \mu_0(X_i) + D_i[\mu_1(X_i) - \mu_0(X_i) + \varepsilon_{1i} - \varepsilon_{0i}] + \varepsilon_{0i} \\ &= \mu_0(X_i) + D_i[\mu_1(X_i) - \mu_0(X_i)] + [\varepsilon_{0i} + D_i(\varepsilon_{1i} - \varepsilon_{0i})] \\ &= \mu_0(X_i) + D_i[E(y_1 - y_0)] + [\varepsilon_{0i} + D_i(\varepsilon_{1i} - \varepsilon_{0i})] \\ &= X_i\beta_0 + D_i[X_i(\beta_1 - \beta_0)] + [\varepsilon_{0i} + D_i(\varepsilon_{1i} - \varepsilon_{0i})]. \end{aligned}$$

If the parameters β_1 and β_0 are different only in the constant with $X_i(\beta_1 - \beta_0) = \alpha$ then we get the regression

$$y_i = X_i\beta_0 + D_i\alpha + [\varepsilon_{0i} + D_i(\varepsilon_{1i} - \varepsilon_{0i})]$$

which is a nonstandard regression because the error term switches off and on with D_i and has no zero mean. The selection problem, which arises from the correlation between the explanatory variable D_i and the remaining error term, again will be obvious. In such a regression there would be a homogeneous ***constant treatment effect*** which would be the same for all observations.

If the unobservables are common across the two states we have $(\varepsilon_{1i} - \varepsilon_{0i}) = 0$ and the treatment effect is

$$T(X_i) = y_{1i} - y_{0i} = \mu_1(X_i) - \mu_0(X_i) = X_i(\beta_1 - \beta_0),$$

which now is only a function of observables.

If as above is assumed that the difference between potential outcomes is a constant (α) then we get a familiar looking dummy variable regression model with ***homogenous treatment effects*** α

$$y_i = X_i\beta + D_i\alpha + \varepsilon_i \quad \text{where } E(\varepsilon_i) = 0,$$

which is on the focus of conventional econometric evaluation literature with focus on a homogeneous treatment effect α or more rarely on a heterogeneous treatment effect $X_i(\beta_1 - \beta_0)$ (Heckman, Lalonde and Smith 1999, p. 23).

However, the assumption $\varepsilon_{1i} = \varepsilon_{0i}$ and then $E(\varepsilon_i) = 0$ no longer holds when sample selection appears requiring more than traditional econometrics.

Regression and the average treatment effect on the treated

So far we analyzed the general treatment effect. Now, the average treatment effect on the treated is discussed, which again measures the gain of program participants compared to that what they would have experienced in the non-treated state

$$E(y_{1i} - y_{0i} \mid X_i, D_i = 1) .$$

This expectation of the above coefficient to D_i is the **average treatment effect of the treated (ATT^{REG})**

$$\begin{aligned} ATT^{REG} &= E(y_{1i} - y_{0i} \mid X_i, D_i = 1) \\ &= E[X_i(\beta_1 - \beta_0) + \varepsilon_{1i} - \varepsilon_{0i} \mid X_i, D_i = 1] \\ &= X_i(\beta_1 - \beta_0) + E(\varepsilon_{1i} - \varepsilon_{0i} \mid X_i, D_i = 1) \end{aligned}$$

Putting this into the above y_i observed outcome equation

$$y_i = X_i\beta_0 \quad + D_i[X_i(\beta_1 - \beta_0)] + [\varepsilon_{0i} + D_i(\varepsilon_{1i} - \varepsilon_{0i})]$$

without changing the equation (that is why the second term of the average treatment effect on the treated $E(\varepsilon_{1i} - \varepsilon_{0i} \mid X_i, D_i = 1)$ will be subtracted) yields

$$\begin{aligned} = X_i\beta_0 \quad & + D_i[X_i(\beta_1 - \beta_0) + E(\varepsilon_{1i} - \varepsilon_{0i} \mid X_i, D_i = 1)] \\ & + \varepsilon_{0i} + D_i(\varepsilon_{1i} - \varepsilon_{0i}) \; - E(\varepsilon_{1i} - \varepsilon_{0i} \mid X_i, D_i = 1) \\ = X_i\beta_0 \quad & + D_i[X_i(\beta_1 - \beta_0) + E(\varepsilon_{1i} - \varepsilon_{0i} \mid X_i, D_i = 1)] \\ & + \{\varepsilon_{0i} + D_i[(\varepsilon_{1i} - \varepsilon_{0i}) \; - E(\varepsilon_{1i} - \varepsilon_{0i} \mid X_i, D_i = 1)]\}. \end{aligned}$$

Again, this is a nonstandard regression problem which combines structural parameters $X_i(\beta_1 - \beta_0)$ with the means of unobservables where parts of the error term switches on or off with D_i. $ATT^{REG} = E(y_{1i} - y_{0i} \mid X_i, D_i = 1)$ "measures the average gain in the outcome for persons who choose to participate in a program compared to what they would have experienced in the base state. It computes the average gain in terms of both observables and unobservables. It is the latter that makes the parameter look non-standard." Heckman, Lalonde and Smith (1999, p. 1887).

The coefficient to D_i now varies and is no further constant for the treated. We have ***heterogeneous treatment effects*** which are different for different treated groups.

Heckman, Lalonde and Smith 1999, pp. 24) further discuss different regressions when different assumptions for the error and the structural parameters are made.

4.3.3 Two Step Selectivity Bias Correction Treatment Effects Estimation

The selection bias problem (non-random sampling) again occurs if $E(\varepsilon \mid X, D) \neq 0$, if there is correlation of D with the error term ε. In this case the required identifying assumption unconfoundedness ($\varepsilon_{1i}, \varepsilon_{0i} \perp X_i, D_i = 1$, independence between the error terms and X and D_i) is injured.

Based on the two steps but extended Heckman (1976, 1979) procedure to account for selectivity an unbiased OLS estimation in the outcome equation is possible. In the first step the treatment (participation in a program) as a latent variable approach is estimated with all observations by a probit approach. This step delivers selection information which by an additional selectivity correction variable (Hazard rate, Mills' ratio) is respected when the outcome is estimated by OLS of the participants only.

The *"classical" sample selection Heckman (1976, 1979) model* respects the interdependence of the selection equation (and thus the interdependence of the treatment and the outcome) by a joint normal distribution of the error terms. In this model, the treatment cannot be estimated directly in the outcome equation simply because there are only treated observations (all $D_i = 1$ for this subsample) in the estimation of the outcome equation in the second step.

The *treatment selectivity bias correction model*, however, estimates the effect of an endogenous binary treatment on a continuous, fully observed outcome variable.

"Classical" sample selection model (Heckman 1976, 1979)

Step 1: Treatment (participation) from an unobserved latent variable D^*

$$D_i^* = Z_i\gamma + \nu_i, \qquad (Z_i \text{ independent of the error term } \nu_i)$$
$$D_i = 1 \text{ if } D_i^* > 0,\ D_i = 0 \quad \text{otherwise.}$$

with explanatory variables Z_i (at least one variable has to be different to X_i) estimated by a probit regression. (participants and non-participants).

Step 2: Outcome from a selectivity correction OLS estimation of participants only

Rewriting the outcome equation $y_i = X_i\beta + \varepsilon_i$ yields

$$y_i = X_i\beta + E(\varepsilon_i \mid X_i, D_i, Z_i) + [\varepsilon_i - E(\varepsilon_i \mid X_i, D_i, Z_i)]$$
$$y_i = X_i\beta + E(\varepsilon_i \mid X_i, D_i, Z_i) + \varepsilon_i^*$$

Assuming (ε, ν) is distributed joint normal the selectivity corrected outcome equation is

$$y_i = X_i\beta + \lambda_i(\sigma_{\varepsilon v} / \sigma_v) + \varepsilon_i^*,$$

where ε_i^* is independent and normally distributed and $\lambda_i = \phi(-Z_i\gamma)/[1-\Phi(-Z_i\gamma)]$ is the additional selectivity correction variable. The estimated OLS coefficient of $(\sigma_{\varepsilon v} / \sigma_v)$ based on the participants only then informs about the selectivity significance. The appropriate asymptotic covariance matrix of the estimated coefficients ensures unbiased estimators.

Treatment selectivity bias correction model

Following Maddala (1983, pp. 120-122) and Greene (2008, p. 180) the treatment effect, however, can be estimated from the selectivity corrected outcome equation based on an endogenous binary treatment D_i in a two steps procedure with *all* observations in the second step:

Step 1: Treatment (participation) from an unobserved latent variable D^* (again)

$$D_i^* = Z_i\gamma + \nu_i, \qquad (Z_i \textit{ independent of the error term } \nu_i)$$
$$D_i = 1 \textit{ if } D_i^* > 0,\ D_i = 0 \quad \textit{otherwise.}$$

with explanatory variables Z_i (at least one variable has to be different to X_i).

The probit estimates defines the selectivity correction, the so called hazard λ_i, which is computed by

$$\lambda_i = \begin{Bmatrix} \phi(Z_i\hat{\gamma}) / \Phi(Z_i\hat{\gamma}) & D_i = 1 \\ -\phi(Z_i\hat{\gamma}) / (1 - \Phi(Z_i\hat{\gamma})) & D_i = 0 \end{Bmatrix}$$

where ϕ is the standard normal density (pdf) and Φ the standard normal cumulative distribution function (cdf).

Step 2: Outcome from a selectivity correction OLS estimation with treatment dummy for all (participants *and* (!) non-participants)

Rewriting the outcome equation of the treatment model $y_i = X_i\beta + D_i\alpha + \varepsilon_i$ yields

$$y_i = X_i\beta + D_i\alpha + E(\varepsilon_i \mid X_i, D_i, Z_i) + \varepsilon_i^*$$
$$y_i = X_i\beta + D_i\alpha + \lambda_i(\rho\sigma) + \varepsilon_i$$

where ε and ν are bivariate normal with mean zero and covariance matrix

$$\text{cov}(\varepsilon, \nu) = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}$$

with ρ as the correlation coefficient between the error terms.

The two steps parameter estimates β and α thus are obtained by the augmented equation with the hazard rate λ_i. The estimated OLS coefficient of $(\rho\sigma)$ then informs about the selectivity significance. Note, the second step is now based on all observations, participants and non-participants.

If the correlation coefficient ρ is zero then the problem reduces to one estimable by OLS which delivers the ATT^{REG} as simply α. If there is correlation, then the Classical Linear Regression OLS overestimates the treatment effect α.

An appropriate asymptotic covariance matrix of the estimated coefficients then ensures the unbiased estimators (see Greene 2008, p. 180). All the selection models are provided by the program packages LIMDEP (Econometric Software Inc. 1998, p. 716) and Stata (with *treatreg* methods and formula).

A comparable method, a 2SLS instrumental variable approach, is discussed in section 4.7.

4.3.4 Fixed Effects Regression and Treatment Effects

If panel data (with repeated observations of the same individuals) are available, the regression selectivity problem can be solved by a fixed effects regression estimation. In a fixed effects regression approach it is assumed that the unobserved effect, θ_i, is a parameter to be estimated for each observation i out of all time periods t. Thus, θ_i is the intercept for observation i (person, firm etc.) that is to be estimated along with the β for the individual characteristics x_{it} and α again as the average treatment effect of

$$y_{it} = \theta_i + X_{it}\beta + D_{it}\alpha + \varepsilon_{it} , \qquad t = 1, 2, ..., T .$$

A fixed effects estimation is possible by "brute force": generating dummies for each time invariant but different individual. Each dummy variable coefficient then would account for unobserved heterogeneity (dummy variable regression). Another approach, first differencing, or de-meaning, uses a transformation to remove the unobserved effect θ_i prior to estimation by averaging the above equation over time:

$$\bar{y}_{it} = \bar{\theta}_i + \bar{X}_{it}\beta + \bar{D}_{it}\alpha + \bar{\varepsilon}_i$$

If we subtract the mean equation from the equation with all time periods (within transformation)

$$y_{it} - \bar{y}_{it} = \theta_i - \bar{\theta}_i + (X_{it} - \bar{X}_{it})'\beta + (D_{it} - \bar{D}_{it})\alpha + \varepsilon_{it} - \bar{\varepsilon}_i \qquad (\theta_i = \bar{\theta}_i)$$

we finally achieve

$$\ddot{y}_{it} = \ddot{X}_{it}\beta + \ddot{D}_{it}\alpha + \ddot{\varepsilon}_{it} , \qquad t = 1, 2, ..., T .$$

The estimated effect α can be interpreted as the effect of a **within-observation change in treatment**. We get the time de-meaned data on y, x and ε which allows pooled OLS estimation of the regression coefficients with unbiased coefficients.

Though the unobserved effects have disappeared in the pooled de-meaned estimation, unobserved heterogeneity nevertheless is respected by the variation over time of the de-meaned data. Note, any time constant explanatory variable (like gender) also are removed along with θ_i.

The fixed effects approach controls for all observation specific factors – whether observable or unobservable – that are constant over time. All time-invariant observation-level factors as a source of omitted variable bias (selection bias) are ruled out even though we may not ever be able to observe or measure them. Each unit serves as its own control group. The identifying assumption is that the counterfactual trend in treatment and control group is the same (no time-varying omitted variables).

One of the desirable features of the fixed effects approach is that it allows for the observation-specific effect to be correlated with the Xs. Thus it explicitly accounts for one form of endogeneity, that resulting from time-invariant omitted variables.

An alternative approach is to use a random-effects model. However, the random effects model assumes that the unit-specific effect is uncorrelated with the Xs; it requires assuming no omitted bias variables. Thus from a causal inference perspective the random effects model is not particularly useful (Berry 2011).

To sum up: When individual heterogeneity is accounted for by the fixed effects regression approach, then unobserved individual heterogeneity does not cause an omitted variable bias problem, which means, it does not cause a selectivity problem anymore. Further information about fixed and random effects panel estimates is discussed e.g. in Wooldridge (2009, chapter 14).

4.3.5 Regression and Treatment Effects Applications

Example 1: OLS estimation
Homogene und heterogene Teilnahmeeffekte des Hamburger Kombilohnmodells:

Pfeifer, Chr. (2009), Homogene und heterogene Teilnahmeeffekte des Hamburger Kombilohnmodells: Ein Verfahrensvergleich von Propensity Score Matching und linearer Regression, in: Wirtschafts- und Sozialstatistisches Archiv, 3, 41-65.

The author compares in this study a propensity score approach to a regression approach to detect treatment effects of the Hamburg “Kombilohn”- model.

He is estimating homogeneous and heterogeneous treatment effects with the regression approach by OLS estimation under the specific assumption that all possible differences between the characteristics of the treated and non-treated are captured by the respected observables.

The substantive results: “Die so ermittelten Teilnahmeeffekte [propensity score matching and linear regression] variieren nur geringfügig zwischen den beiden Evaluierungsmethoden. Jedoch ist die Ermittlung und Interpretation von heterogenen Teilnahmeeffekten mit linearen Regressionen einfacher als mit Matching Methoden. Insgesamt zeigen sich stark positive Teilnahmeeffekte, die bei Problemgruppen am Arbeitsmarkt höher ausfallen. Daher scheint das

Hamburger Kombilohnmodell neben allgemein positiven Wirkungen auch zielgruppenorientiert zu sein“ (Pfeifer 2009, S. 41).

Example 2: Two step regression treatment:
Timing and Fragmentation of Daily Working Hours Arrangements and Income Inequality – An Earnings Treatment Effects Approach with German Time Use Diary Data

Merz, J., Böhm, P. and D. Burgert (2009), Timing and Fragmentation of Daily Working Hours Arrangements and Income Inequality – An Earnings Treatment Effects Approach with German Time Use Diary Data, in: electronic International Journal of Time Use Research, 6/2, 200-239.

The authors analyzed different daily working hours arrangements according to fragmentation and the timing of work as the treatment.

With a bivariate probit selection equation the influence, the treatment effect, being in such a daily working hours situation, on the earnings is estimated via a bivariate probit model for the selection incorporated then in the outcome equation, the individual earnings. The method is the above discussed treatment selectivity bias correction model.

Note: The bivariate probit estimation here allows to specify the treatment and its effect within the outcome equation.

The substantive result: there are significant income effects according to the treatment, which is the different daily working hours arrangement.

4.4 Cross Section Comparison

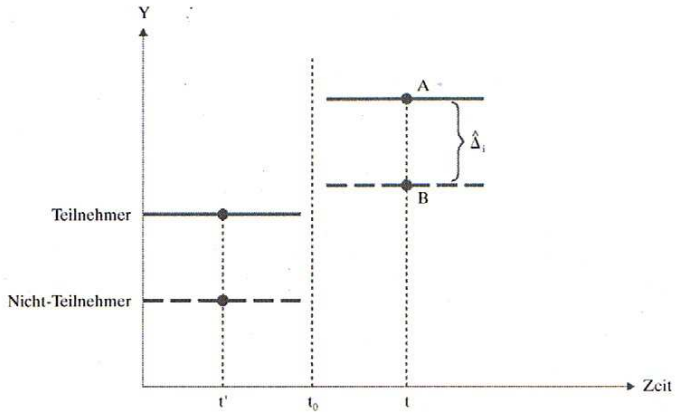
Whereas the propensity score and the regression model approach are rather complex, the following methods rely more directly on the observed data. Within the cross section comparison the non-participants simply are used as the comparison group to the participants. The expected outcome value of the non-participants is taken as the contrafactual situation for the outcome measure of the participants as non-participants. The identifying assumption behind assumes that the participant’s average outcome would have been developed similar as of the non-participants if the participants had not been treated.

Figure 1 illustrates the cross section evaluation with $(A-B)$ as the treatment effect.

The general average treatment effect on the treated (ATT)

$$ATT = E(y_{1i} - y_{0i} \mid D_i = 1) = E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 1)$$

Figure 1: Cross Section Evaluation



Source: Bauer, Fertig and Schmidt 2009, p. 158

then results in the **average treatment effect of the treated under cross section comparison ATT^{csc}** to

$$ATT^{csc} = E(y_{1i} \mid D_i = 1) - E(y_{0i} \mid D_i = 0) \ .$$

The identifying assumption

$$E(y_{0i} \mid X, D_i = 1) = E(y_{0i} \mid X, D_i = 0)$$

would be problematic if the participants based on unobservable characteristics (motivation, cognitive abilities) would select themselves systematically into one of the two groups (selectivity bias).

4.5 Before and After

In the before-after approach the participants are compared by themselves where the outcome before treatment (at time t') is compared to the outcome after treatment (at time t).

Figure 2 illustrates the before and after evaluation with $(A-B)$ as the treatment effect.

The **average treatment effect of the treated under before and after comparison ATT^{ba}** then is

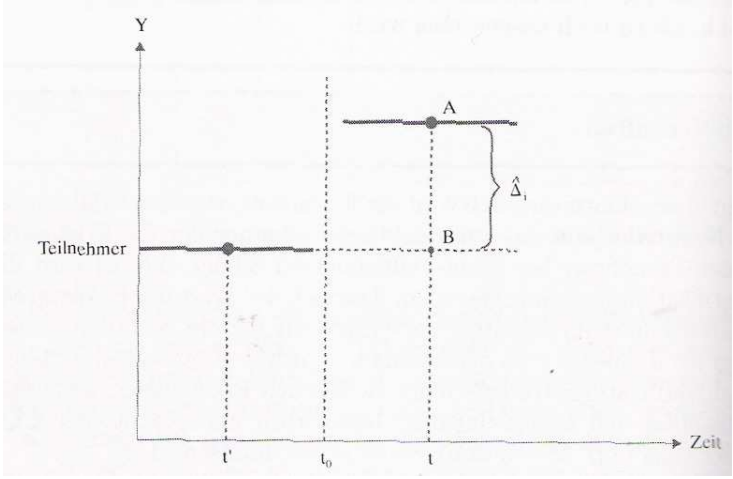
$$ATT^{ba} = E(y_{1i,t} \mid X, D_i = 1) - E(y_{0i,t'} \mid X, D_i = 1) \ .$$

The identifying assumption

$$E(y_{0i,t} \mid X, D_i = 1) = E(y_{0i,t'} \mid X, D_i = 1)$$

requires no change of the participants behaviour/situation besides the effect of the treatment.

Figure 2: Before and After Evaluation

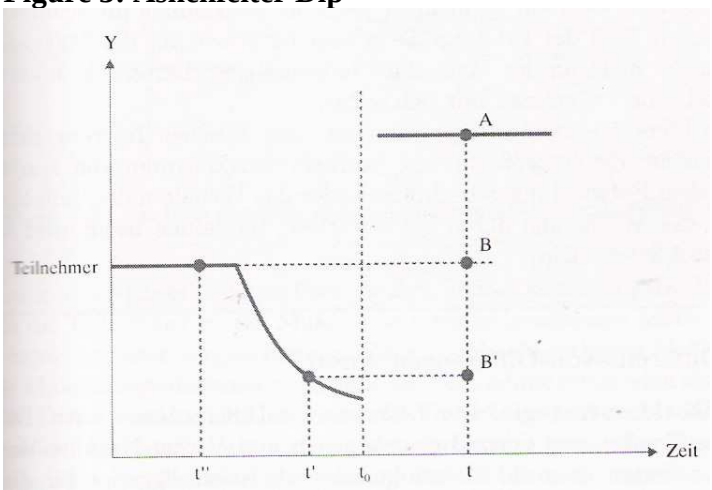


Source: Bauer, Fertig and Schmidt 2009, p. 160

The before-after treatment effect, which needs two time period observations, before and after a treatment, might be inconsistent if in the meantime there are business cycle changes or other changes in the behaviour of the participants themselves, like if the participants are already influenced by the expected treatment ("Ashenfelter's dip"; see Figure 3).
Ashenfelter (1978, p. 55) in his study on effects of training programs on earnings observed, that "all of the trainee groups suffered unpredicted earnings declines in the year prior to training. … This suggests that simple before and after comparisons of trainees earnings may be seriously misleading evidence on the effect of training on earnings …". Neglecting this decline in earnings therefore leads to an overestimation of the job training effect.

In the program effects literature such a decline of an outcome measure just prior to a program start– whatever reason is behind - since then is known as Ashenfelter's dip and subsequent research finds this regularity (Ashenfelter and Card 1985).

Figure 3: Ashenfelter Dip



Source: Bauer, Fertig and Schmidt 2009, p. 161

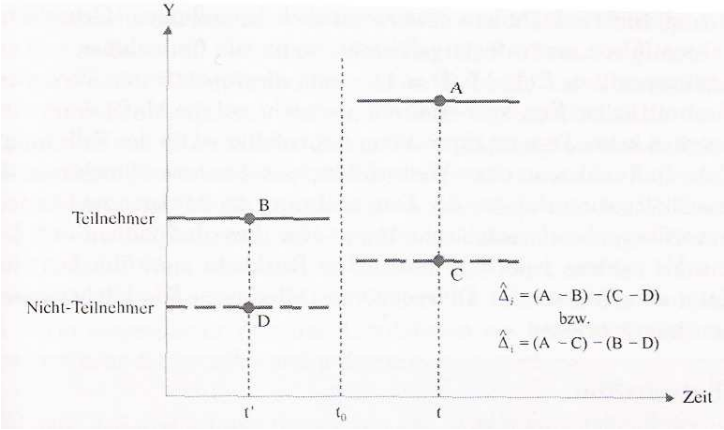
4.6 Differences-in-Differences

The differences-in differences (DiD) estimator is an extension of the before-after approach and a combination of the cross-section and before-after comparison: The non-participants (as in the cross-section approach) are the control group of the participants, and the change of the outcome variable between t' and t between both groups are compared (as in the before-after comparison). With required longitudinal data the changes in the outcome for participants are contrasted with the corresponding changes for the non-treated individuals.

Thus, the DiD estimator is defined as the difference in average outcome in the treated group before and after treatment *minus* the difference in average outcome in the control group before and after treatment; it is a difference of differences.

Figures 4a,b illustrates the treatment effect as $(A-B)-(C-D)$ respectively as $(A-C)-(B-D)$.

Figure 4a: Diff-in-Diff Evaluation



Source: Bauer, Fertig and Schmidt 2009, p. 163

The outcome difference between participants and non-participants before the treatment ($B-D$, in t') is compared to the outcome difference between both groups after the treatment ($A-C$, in t). The difference of both differences then is the DiD average treatment effect:

$$ATT^{DiD} = (A-C)-(B-D) = (A-B)-(C-D).$$

ATT^{DID} takes into account pre-existing differences between treatment and control group and a general time trend.

The **average treatment effect by differences-in-differences ATT^{DiD}** then is

$$ATT^{DiD} = [E(y_{1i,t} \mid X, D_i = 1) - E(y_{0i,t'} \mid X, D_i = 1)]$$
$$-[E(y_{0i,t} \mid X, D_i = 0) - E(y_{0i,t'} \mid X, D_i = 0)] \qquad (A-B)-(C-D)$$

$$= E[(y_{1i,t} - y_{0i,t'} \mid D_i = 1) - (y_{0i,t} - y_{0i,t'} \mid D_i = 0)]$$
$$= E(y_{1i,t} - y_{0i,t'} \mid D_i = 1) - E(y_{0i,t} - y_{0i,t'} \mid D_i = 0).$$

The identifying assumption

$$E(y_{1i,t} \mid X, D_i = 1) - E(y_{0i,t} \mid X, D_i = 1)$$
$$= E(y_{1i,t} - y_{0i,t'} \mid X, D_i = 1) - E(y_{0i,t} - y_{0i,t'} \mid X, D_i = 0)$$

respectively

$$E[(y_{0i,t} - y_{0i,t'} \mid D_i = 1) - (y_{0i,t} - y_{0i,t'} \mid D_i = 0)] = 0$$

postulates that the outcome of the participants has changed in the same way as the outcome of the non-participants if the participants had not been treated (over time permanent same difference between treated and non-treated). Differencing the differences between participants and non-participants then eliminates the time-invariant linear selection bias. This would be best achieved if a randomized situation, a natural experiment, constitutes the treatment process.

The key assumption for any DiD strategy is that the outcome in treatment and control group would follow the same trend in the absence of the treatment. This identifying assumption will be distorted if business cycle or other developments would *differently* influence the outcome of participants and non-participants in the course of time. Also the Ashenfelter Dip problem could yield wrong results.

Diff-in-Diff and Regression

Another approach to quantify the Diff-in-Diff approach is regression. A regression approach provides the $\mathbf{ATT^{DiD}}$ with an estimated regression coefficient in the following outcome specification

$$y_{iT} = \beta_0 + \beta_1 T_i + \beta_2 D_i + \beta_3 (T_i * D_i) + X_i \gamma + \varepsilon_{it}$$

where $T_i = 0$ characterizes the before and $T_i = 1$ the after situation, $D_i = 1$ if in a program (difference between the treatment and control group prior to a policy change) and $D_i = 0$ if not and $(T_i * D_i)$ as the interaction term (T_i multiplied by D_i is the dummy for those observed in the treatment group in the second period) whose coefficient β_3 is the estimated Diff-in-Diff average treatment effect (Villa 2012); again X_i are individual characteristics and γ a vector of estimated parameters using the data pooled over both years. Why β_3 is the estimated Diff-in-Diff average treatment effect is illustrated in the following.

Meaning of estimated coefficients:

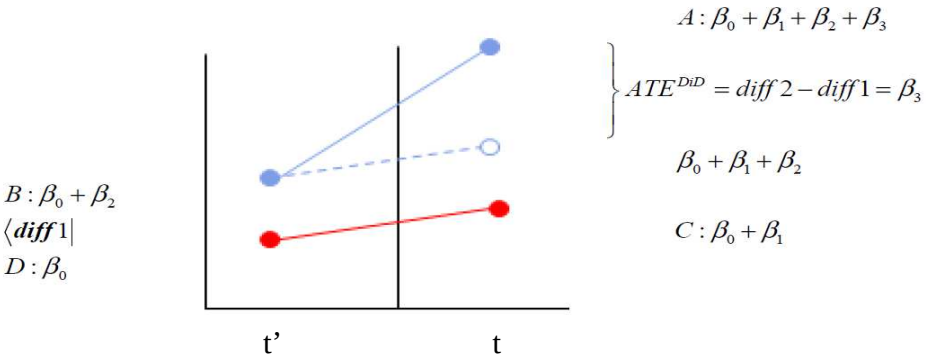
β_0: outcome of control group at T=0

$\beta_0 + \beta_1$:	outcome of control group at T=1
β_2:	difference between treatment and control group at T=0
$\beta_0 + \beta_2$:	outcome of treatment group at T=0
$\beta_0 + \beta_1 + \beta_2 + \beta_3$:	outcome of treatment group at T=1
β_3:	**ATT^{DiD} average treatment effect**

β_3 can be interpreted as a causal effect if the observations (persons or firms etc) cannot select (1) into or out of the treatment group and also (2) not into or out of the treatment period, and (3) if both, treatment and control group would have faced the same trend in the absence of the treatment.

Figure 4b and Table 2 illustrate the DiD differencing procedure and regression.

Figure 4b: Diff-in-Diff Evaluation and Regression



Source: Gong 2009 and own illustration

$$(\boldsymbol{B}-\boldsymbol{D}) \quad \boldsymbol{diff}\,1: (\beta_0 + \beta_2) - (\beta_0) = \beta_2$$

$$(\boldsymbol{A}-\boldsymbol{C}) \quad \boldsymbol{diff}\,2: (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_1) = \beta_2 + \beta_3$$

$$\boldsymbol{ATT}^{DID} = \boldsymbol{DiD} = (\boldsymbol{A}-\boldsymbol{C}) - (\boldsymbol{B}-\boldsymbol{D}) = \boldsymbol{diff}\,2 - \boldsymbol{diff}\,1 = (\beta_2 + \beta_3) - (\beta_2) = \beta_3$$

Table 2: Diff-in-Diff Evaluation Before and After

ATT^{DiD}	before	after	after-before
treated	$B: \beta_0 + \beta_2$	$A: \beta_0 + \beta_1 + \beta_2 + \beta_3$	$(\boldsymbol{A}-\boldsymbol{B}) = \beta_1 + \beta_3$
untreated	$D: \beta_0$	$C: \beta_0 + \beta_1$	$(\boldsymbol{C}-\boldsymbol{D}) = \beta_1$
differences	$(\boldsymbol{B}-\boldsymbol{D})$ $\boldsymbol{diff}\,1 = \beta_2$	$(\boldsymbol{A}-\boldsymbol{C})$ $\boldsymbol{diff}\,2 = \beta_2 + \beta_3$	$(\boldsymbol{A}-\boldsymbol{C}) - (\boldsymbol{B}-\boldsymbol{D}) = \boldsymbol{diff}\,1 - \boldsymbol{diff}\,2 = \beta_3$

DiD: ATT or ATE?

The question arises if the DiD estimates in the standard model corresponds to the average treatment effect of the treated ATT or to the more general population average treatment effect ATE.

Regard again the DiD regression

$$y_{iT} = \beta_0 + \beta_1 T_i + \beta_2 D_i + \beta_3 (T_i * D_i) + X_i \gamma + \varepsilon_{it}$$

The expected outcome without treatment is

$$E(y_{0i,T}) = \beta_0 + \beta_1 T_i + \beta_2 D_i$$

with D_i measuring just the difference between the control group and the treatment group before treatment.
The expected outcome in the presence of the intervention is

$$E(y_{1i,T}) = E(y_{0i,T}) + \beta_3 (T_i * D_i)$$

neglecting the socio-economic X_i influences for the ease of exposure.

The average treatment of the treated ATT is the expected difference in $y_{1i,T} - y_{0i,T}$ for those who are treated in time 1, i.e. with $D_i = 1$ and $T_i = 1$. Putting $D_i = 1$ and $T_i = 1$ in the above expectation equations then we get

$$\begin{aligned} ATT^{DiD} &= E(y_{1i,T} - y_{0i,T} \mid D_i = 1, T = 1) \\ &= E(y_{1i,T}) \mid D_i = 1, T = 1) - E(y_{0i,T}) \mid D_i = 1, T = 1) \\ &= E(y_{0i,T}) + \beta_3 (T)_i * D_i)) - E(y_{0i,T}) \\ &= \beta_3 (T_i * D_i) \qquad\qquad T_i = 1; D_i = 1 \\ &= \beta_3 . \end{aligned}$$

Thus, the **standard DiD** delivers the **average treatment effect of the treated ATT^{DiD}**.

Diff-in-Diff and Panel Regression
A simple panel regression model with two time periods and a treatment indicator, D_{it}, which is unity if unit i participates in a program at time t, is

$$y_{it} = \delta_i + \beta_1 T_{it} + \beta_3 (T_{it} * D_{it}) + X_{it}^{'} \gamma + \varepsilon_{it},$$

where δ_i is an unobserved effect, $T_{it} = 1$ if $t = 2$ and zero otherwise, and ε_{it} an idiosyncratic error or time-varying error[2]. The treatment effect is the coefficient β_3.

An estimation procedure is to first difference to remove the time invariant individual effects δ_i

$$y_{i2} - y_{i1} = \beta_1 (T_{i2} - T_{i1}) + \beta_3 (T_{i2} * D_{i2} - T_{i1} * D_{i1}) + (X_{i2} - X_{i1}) \gamma + \varepsilon_{i2} - \varepsilon_{i1}$$

or

[2] An idiosyncatic error or time-vaying error represents unobserved factors that change over time and affect y_{it} (Wooldridge 2009, 456).

$$\Delta y_i = \beta_1 + \beta_3 \Delta D_i + \Delta X_i \gamma + \Delta \varepsilon_i .$$

If the change in the treatment is not correlated with the change of the error term, i.e. $E(\Delta D_i \Delta \varepsilon_i) = 0$, then OLS applied is consistent (Imbens and Wooldridge 2007, 8). When $D_{i1} = 0$ for all i then no units are exposed to the program in the initial period. With

$$\Delta y_{i,treated} = \beta_1 + \beta_3 D_i + \Delta X_i \gamma + \Delta \varepsilon_i$$
$$\Delta y_{i,control} = \beta_1 \qquad + \Delta X_i \gamma + \Delta \varepsilon_i ,$$

then the OLS estimator of the **average treatment effect** is

$$\beta_3 = \Delta \bar{y}_{i,treated} - \Delta \bar{y}_{i,control}$$

which is a **difference-in-difference** estimate except that the means of the same units over time are differented.

Thus, in case of **panel data** β_2 from the pooled regression approach is substituted by an individual specific effect δ_i resulting to the above panel outcome DiD regression

$$y_{it} = \delta_i + \beta_1 T_{it} + \beta_3 (T_{it} * D_{it}) + X_{it} \gamma + \varepsilon_{it}$$

again with β_3 as the average treatment effect ATE^{DiD}. β_3 can be interpreted as a causal effect if the observations cannot select into or out neither of the treatment group nor the treatment period and that both treatment and control group would have experienced the same trends in the absence of treatment. Note, as with fixed effects panel estimation δ_i is cancelled out after de-meaning of the variables.

General DiD requirements: the exogeneous treatment at the best should be a natural experiment to ensure randomness. Two cross sections (before and after) are needed. Sometimes with the help of individual age variables the before and after situation can be constructed out of one cross section only (Valente 2013). Differencing over time and groups eliminates a common trend and unobserved heterogeneity. As in the before and after approach, if some business cycle developments (trends) have different influences on the treated and untreated, or some developments are already expected at t', then the results are no more consistent.

Examples DiD

David Card (1990), **The impact of the Mariel Boatlift on the Miami labor market**, in: Industrial and Labor Relations Review, 43/2, 242-257.

The question there is how the labour markets do react by immigration shocks. The case: the immigration from Cuba to Miami (USA) of about 125.000 cubans; comparison of four cities with a similar economic development and a similar population structure. The result: the treatment, the immigration, had almost no effect. There is almost no influence on wages and labour supply of non-cuban lower skilled workers.

Card, D. and A.B. Krueger (1994), **Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania**, in: The American Economic Review, Vol 84, No 4,772-793.

"On April 1, 1992, New Jersey's minimum wage rose from $4.25 to $5.05 per hour. To evaluate the impact of the law we surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. Comparisons of employment growth at stores in New Jersey and Pennsylvania (where the minimum wage was constant) provide simple estimates of the effect of the higher minimum wage. We also compare employment changes at stores in New Jersey that were initially paying high wages (above $5) to the changes at lower-wage stores. We find no indication that the rise in the minimum wage reduced employment. (JEL 530, 523)"

Braakmann, N. and J. Wagner (2013), **Labor market adjustments after a long import shock: Evidence from the German clothing industry and the WTO Agreement on Textiles and Clothing,** in: Hirschel, D., Paic, P. and M. Zwick (Eds.), Daten in der wirtschaftswissenschaftlichen Forschung, Festschrift zum 65. Geburtstag von Prof. Dr. Joachim Merz, SpringerGabler, Wiesbaden.

In this paper the consequences of a large import shock that hit the German clothing industry in 2002 and 2005. Using social security data on both the individual and the firm level and workers and firms from an unaffected, but similar industry as the control group, they showed that, despite a high level of firm mortality, individual employment prospects were harmed to a much lesser extent. Their method is a difference-in-difference panel data regression approach.

4.7 Instrumental Variables

The instrumental variable (IV) approach in regression analyses deals with the endogeneity problem of explanatory variables and is related to the omitted variable problem and its consequences for the selection bias.

As discussed in our regression section 6 unbiased results is the consequence, when the Classical Linear Regression (CLR) assumption of independence of the explanatory variable and the error term is not valid. As we have discussed in the case of a treatment by a dummy variable in the regression approach, however, the treatment dummy is correlated with the error term.

The IV approach tries to solve the problem by the following: Instead of using the treatment dummy directly a substitute is incorporated in the outcome equation which provides consistent (but not unbiased) results (Wooldridge 2009, 506 pp).

Consider again the linear outcome equation $y_i = X_i\beta + D_i\alpha + \varepsilon_i$. If we have an observable variable z which satisfies two assumptions:

First: z is uncorrelated with the outcome error term ε $(\text{cov}(z,\varepsilon) = 0)$ and
Second: z is correlated with the treatment D $(\text{cov}(z, D) \neq 0)$,

then z can serve as an instrument instead of D as the exogenous variable in the outcome equation

$$y_i = X\beta + z_i\alpha + \varepsilon_i$$

with desired properties for a consistent estimation.

The problem is to find such an instrument which affects participation only (which is closely correlated, negatively or positively, with the treatment) but does not affect the outcome and its error term.

That z has to be uncorrelated with any other determinant of the outcome (equivalent to $\text{cov}(z,\varepsilon) = 0$) "is called an *exclusion restriction*, since z can be said to be excluded from the causal model of interest " (Angrist and Pischke 2009, p.116).

If z is estimated by some other variables, then a two stage least square (2SLS) estimation procedure with the first equation explaining z (dichotomous or continuous in the case of heterogeneous treatments) and using its estimates in the second stage, the outcome estimation. A 2SLS IV estimation e.g. with the program package LIMDEP is described in the LIMDEP User's Manual Version 7 (Econometric Software Inc. 1998, pp. 716-717).

Example IV

Estimating the Return to Education for Married Women (Wooldridge 2009, 512 example):

Women's log(wage) there is estimated simply by education. Since ability might be the omitted variable, therefore father's education is used as an instrument. A regression of father's education on the education of the women delivers significant correlation between the instrument and the treatment, woman's education. The result: the regression of women's log(wage) on the instrument delivers a return to education which is half as large as the original regression with women's education.

4.8 Regression Discontinuity

The regression discontinuity model (RDM, or regression discontinuity design RDD) approach can be seen as a type of instrumental variable identification. Discontinuities in the selection process are used to identify the causal effects. The treatment depends on some observed variables and a certain threshold determines the participation.

For a deterministic assignment to a treatment an indicator variable x_i switches from 0 to 1 if the selection variable z_i exceeds a threshold z_0

$$x_i = \begin{cases} 0 & if \quad z_i < z_0 \\ 1 & \quad\;\; z_i \geq z_0 \end{cases}.$$

The observed outcome y_i then can be described by (Burgert 2006, 128)

$$y_i = y_{0i} + x_i(z_i)\beta_i,$$

with y_i as the outcome variable, y_{0i} the non-treatment case and β_i as the treatment effect of the program. The main advantage of using RDD is "that one does not need to assume a model to hold neither in its specification of variables included nor a parametric form" (Burgert 2006, 128).

Hahn et al. (2001) show how to identify the treatment effects for common effects ($\beta_i = \beta$) and for variable treatment effects (β_i). In both cases the **local average treatment effect (LATE)** of a program at point z_0 is

$$\beta = y^+ - y^- \quad \text{respectively} \quad E(\beta_i \mid z_i = z_0) = y^+ - y^-$$

with $y^+ = \lim_{z \to z_0^+} E(y \mid z_i = z)$ and $\lim_{z \to z_0^-} E(y \mid z_i = z)$.

When the treatments are heterogeneous over firms a local continuity assumption of $E(\beta_i \mid z_i = z)$ at $z = z_0$ is required, and conditional of z being close to z_0, x_i and the individual treatment effect have to be independent.

For estimation of the LATE

$$\hat{\beta} = \hat{y}^+ - \hat{y}^-$$

a local linear regression is used where the limit y^+ is estimated by the coefficient a of

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum (y_i - a - b(z_i - z_0))^2 K\left((z_i - z_0)/h\right) I(z_i > z_0)$$

with K as a Kernel function, h ist bandwidth and I(.) as an index function (Hahn et al. 2001).

Note: Because of the regression nature there is no common support for participants and non-participants which makes any matching impossible.

For more details see Hahn, Todd and van de Klaauw 2001 and Heckman, Lalonde and Smith 1999.

Example RDM

Burgert, D. (2006), **The Impact of German Job Protection Legislation on Job Creation in Small Establishments - An Application of the Regression Discontinuity Design**, in: Applied Economics Quarterly 52/2, 123-140

There being eligible under the job protection legislation the discontinous selection variable number of employee is above a certain firm size from which the Kündigungsschutzgesetz is applied as a quasi random experiment..

The result of this treatment effects analysis: Against a discussed negative effect of the job protection law (Kündigungsschutzgesetz) on employment, no significant effects were visible by this (and other) microeconometric studies.

4.9 Panelanalyses

Paneldata provide repeated observation of the same individuals. A prominent example is the German Socio-Economic Panel (SDOEP) or the US Panel Study of Income Dynamics (PSID). Paneleconometric approaches consider the longitudinal aspect which e.g. is necessary for the differences-in-differences method. But also for almost all other approaches, like dynamic matching or the regression approach, panel data allows deeper insights into the dynamic treatment effects.

One particular problem, the unobserved heterogeneity of the individuals (observations), which might disturb CLR regression assumptions, can be handled to a certain degree by fixed effects panel models (e.g.) which estimate the single additional individual influence out of the longitudinal information (see chapter 4.3.4).

Further discussion about panel econometrics and causality is given in Lechner 2002 or Caliendo and Hujer 2006.

Excursus: Endogeneity

Endogeneity describes the presence of an endogeneous explanatory variable. An endogeneous explanatory variable is an explanatory variable in a multiple regression model that is correlated with the error term, either because of an omitted variable, measurement error, or simultaneity (Wooldridge 2009, p. 838).

Endogeneity results in a biased and inconsistent estimator. For illustration let us regard a single regress with mean centered x- and y-values (the respective means are substracted from the original data). Then for the regression

$$\mathbf{y} = \beta_1 \mathrm{x} + \varepsilon$$

the OLS estimator $\hat{\beta}_1$ of β_1 is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

Put y into the OLS estimator then

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(\beta_1 x_i + \varepsilon_i)}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n}(\beta_1 x_i^2 + x_i\varepsilon_i)}{\sum_{i=1}^{n} x_i^2} = \frac{\beta_1 \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} x_i\varepsilon_i}{\sum_{i=1}^{n} x_i^2} = \beta_1 + \frac{\sum_{i=1}^{n} x_i\varepsilon_i}{\sum_{i=1}^{n} x_i^2}.$$

If x und ε are correlated, then the expectation of the second term is not zero and the nominator will not converge to zero if the sample size increases.

Therefore, the estimator of β_1 is biased by the second term and the bias will not be outweighed by an increasing sample.

References

Angrist, J.A. and J.St. Pischke (2009), Most Harmless Econometrics – An Empiricist's Companion, Princeton University Press, Princeton, USA.

Angrist, J.D. (1998), Estimating the labor market impact of voluntary military service using social security data on military applicants, in: Econometrica 66(2), pp. 249-288.

Angrist, J.D. and A.B. Krueger (1999), Empirical Strategies in Labor Economics, in: Handbook of Labor Economics, Volume 3, Chapter 23, Edited by O. Ashenfelter and D. Card, Elsevier Science B.V. , The Netherlands, 1277-1365

Ashenfelter, O. (1978), Estimating the Effect of Training Programs on Earnings, in: Review of Economics and Statistics, 60/1, 47-57.

Ashenfelter, O. and D. Card (1985), Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs, in: The Review of Economics and Statistics, 67, 648-660.

Bauer, Th.K., Fertig, M. and Chr.M. Schmidt (2009), Empirische Wirtschaftsforschung: eine Einführung, Springer.

Berry, Christopher (2011), Fixed Effects, Empirical Implications of Theoretical Models (EITM) summer school ppt, The Harris School, The University of Chicago.

Black, D. A. and J. Smith (2004), How Robust Is the Evidence on the Effects of College Quality? Evidence from Matching, in: Journal of Econometrics, 121 (1–2), 99–124.

Blundell, R. and M. Costa Dias (2009), Alternative Approaches to Evaluation in Empirical Microeconomics, in: The Journal of Human Resources, 44/1, 565-640.

Braakmann, N. and J. Wagner (2013), Labor market adjustments after a long import shock: Evidence from the German clothing industry and the WTO Agreement on Textiles and Clothing, in: Hirschel, D., Paic, P. and M. Zwick (Eds.), Daten in der wirtschaftswissenschaftlichen Forschung, Festschrift zum 65. Geburtstag von Prof. Dr. Joachim Merz, SpringerGabler, Wiesbaden, 201-224.

Bryson, A., Dorsett, R. and S. Purdon (2002), The Use of Propensity Score Matching in the Evaluation of Active Labour Market Policies, Working Paper no. 4, Department for Work and Pensions, London.

Burgert, D. (2006), The Impact of German Job Protection Legislation on Job Creation in Small Establishments - An Application of the Regression Discontinuity Design, in: Applied Economics Quarterly 52/2, 123-140

Caliendo, M. (2006), Microeconometric Evaluation of Labour Market Policies, Springer, Berlin.

Caliendo, M. and R. Hujer (2006), The microeconometric estimation of treatment effects – An overview, in: Allgemeines Statistisches Archiv, 90, 199-215.

Caliendo, M. and S. Kopeinig (2008), Some practical guidance for the implenmentation of propensity score matching, in: Journal of Economic Surveys, 22, 31-72.

Caliendo, M., Hujer, R. and S. Thomsen (2004), Evaluation der Eingliederungseffekte von Arbeitsbeschaffungsmaßnahmen in reguläre Beschäftigung für Teilnehmer in Deutschland, Journal for Labour Market Research / Zeitschrift für ArbeitsmarktForschung, 37(3), 211-237.

Card, D. (1990), The impact of the Mariel Boatlift on the Miami labor market, in: Industrial and Labor Relations Review, 43/2, 242-257.

Card, D. and A.B. Krueger (1994), Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania, in: The American Economic Review, Vol 84, No 4,772-793.

Econometric Software Inc. (1998), LIMDEP User's Manual Version 7, Plainview, USA

Falk, A. (2007), Gift exchange in the field, in: Econometrica 75/5, 1501-1511.

Gangl, Markus (2010), Causal Inference in Sociological Research (August 2010). Annual Review of Sociology, Vol. 36, 21-47.

Gibson-Davis, C. M., and E.M. Foster (2006), A Cautionary Tale: Using Propensity Scores to Estimate the Effect of Food Stamps on Food Insecurity, in: Social Service Review, 80, 93-126.

Greene, W. H. (2008). Econometric Analysis. 6th ed. Upper Saddle River, NJ: Prentice–Hall.

Guo, S. and M.W. Fraser (2010), Propensity Score Analysis – Statistical Methods and Applications, Sage Publications, Thousand Oaks, USA.

Hahn, J., Todd, P. and W. Van der Klaauw (2001), Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design, in: Econometrica, 69 (1), 201-209.

Harmon, C. and I. Walker (1995): Estimates of the Economic Return to Schooling for the United Kingdom, in: The American Economic Review, 85 (3), 1278-1286

Heckman, J.J. (1976), The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables, and a Simple Estimator for Such Models, in: Annals of Economic and Social Measurement, 5, 475-92

Heckman, J.J. (1979), Sample Selection Bias as a Specification Erroe, in: Econometrica, 47/1, 153-61.

Heckman, J.J., and E. Vytlacil (2001), Policy-Relevant Treatment Effects, in: American Economic Review 91 (2), 107–11.

Heckman, J.J., Lalonde, R.J. and J.A. Smith, (1999): The Economics and Econometrics of Active Labor Market Programs. In: Ashenfelter, O./Card, D. (Hrsg.): Handbook of Labour Economics, 3, Elsevier, Amsterdam, 1865-2097.

Huber, M., Lechner M., Walter, T and C. Wunsch (2010): Do German Welfare-to-Work Programmes Reduce Welfare Dependency and Increase Employment?, in: German Economic Review 12 (2): 182-204

Hujer, R. (2011), Mikroökonometrie und Kausalität: Zur Bedeutung arbeitsmarktpolitischer Evaluationsstudien, in: AStA Wirtschafts- und Sozialstatistisches Archiv, 5. 5-18.

Imbens, G.W. (2004), Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review, in: The Review of Economics and Statistics, February 2004, 86(1): 4–29.

Imbens, Guido W., and Jeffrey M. Wooldridge (2007), What's New in Econometrics? NBER Lecture 10.

Imbens, Guido W., and Jeffrey M. Wooldridge (2009), Recent Developments in the Econometrics of Program Evaluation, in: Journal of Economic Literature, Vol. 47(1), 5-86.

Lechner, M. (1999), Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification, in: Journal of Business & Economic Statistics, 17, Nr. 1, S. 74–90.

Lechner, M. (2002), Eine Übersicht über gängige Modelle der Panelökonometrie und ihre kausale Interpretation, in: Allgemeines Statistisches Archiv, 86, 125-143.

Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

Merz, J., Böhm, P. and D. Burgert (2009), Timing and Fragmentation of Daily Working Hours Arrangements and Income Inequality – An Earnings Treatment Effects Approach with German Time Use Diary Data, in: electronic International Journal of Time Use Research, 6/2, 200-239.

Morgan, Stephen L. and Christopher Winship (2010), Counterfactuals and Causal Inference – Methods and Principles for Social Research, Cambridge University Press, Cambridge (UK).

Nichols, Austin (2007), Causal inference with observational data, in: The Stata Journal 7(4), pp. 507-541.

Oberschachtsiek, D. and P. Scioch (2015), The outcome of coaching and training for self-employment. A statistical evaluation of outside assistance support programs for unemployed business founders in Germany, in: Journal of Labour Market Research, 48, 1-25, DOI 10.1007/s12651-014-0161-6.

Österbacka, E., Merz, J. and E., Zick, C.D. (2012), Human Capital Investment in Children – A Comparative Analysis of the Role of Parent-Child Shared Time in Selected Countries, in: electronic International Journal of Time Use Research, Vol. 9, 120-143.

Pfeifer, Chr. (2009), Homogene und heterogene Teilnahmeeffekte des Hamburger Kombilohnmodells: Ein Verfahrensvergleich von Propensity Score Matching und linearer Regression, in: Wirtschafts- und Sozialstatistisches Archiv, 3, 41-65.

Rosenbaum, P. and D. Rubin, (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika, 70, 41-55.

Rosenbaum, P.R. and D.B. Rubin (1985), Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, in: Journal of the American Statistical Association, 39, 33-38.

Rossi, P.H., Lipsey, M.W. and H.E. Freeman (2004), Evaluation – A Systematic Approach, 7th edition, Thousand Oaks, CA: Sage Publications.

Schlotter, M., Schwerdt, G. and L. Woessmann (2010), Econometric Methods for Causal Evaluation of Education Policies and Practices: A Non-Technical Guide, IZA DP No. 4725; Bonn.

Smith, J.A. and P.E. Todd (2005), Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?, in: Journal of Econometrics 125 (1–2), 305–53.

Sobel, Michael E. (2000), Causal Inference in the Social Sciences, in: Journal of the American Statistical Association, Vol. 95, No. 450, 647-651.

Valente, Christine (2013), Education and Civil Conflict in Nepal, The World Bank Economic Review: lht014.

Villa, Juan M. (2012), Simplifying the estimation of differences in differences treatment effects with Stata, in: MPRA Paper (43943), 1-12.

Winship, Christopher and Stephen L. Morgan (1999), The Estimation of Causal Effects from Observational Data, in: Annual Review of Sociology, Vol. 25, 659-706.

Wooldridge, J.M. (2009, 2006), Introductory Econometrics, A Modern Approach, Fourth, Third Edition, Thomson, South-Western, Canada.