

Empirical Economics with R

5 Instrumental Variable Estimation, Potential Outcomes, and the Impact of Intensive Job Counseling

Uni Ulm

Prof. Dr. Sebastian Kranz

WiSe 20/21

Experiment on Job Search Counseling in France

2 / 31

- Traditionally in France the public employment agency provides job search counseling for unemployed job seekers.
- After a law change in 2005 also private firms could get reimbursed by the state for providing job counseling.
- In 2007-2008 a large scale randomized experiment was conducted to compare 3 different job counseling programs:
 - An intensive counseling program by private firms.
 - A new intensive counseling program by the public unemployment agency.
 - The standard job counseling by the public unemployment agency.
- In the private and intensive public program one case worker did assist at most 40 job seekers, while in the standard job counseling a case worker assists on average 120 job seekers.

Analysis of the job-counseling experiment

3 / 31

- The experiment is analyzed in detail in the article "Private and Public Provision of Counseling to Job Seekers: Evidence from a Large Controlled Experiment." by Behaghel et al. (2014)
- We will replicate some key steps of the analysis in this chapter and in the corresponding RTutor problem set.
- First we only explain and analyse the difference between the intensive public program and the standard public job counseling, we look at the private programme briefly later.

Selection of program participants

4 / 31

- 3385 job seekers were randomly selected and given the *option* to participate in the public intensive counseling program.
- Each job seeker with the treatment option could then either accept the intensive counseling program or reject it and take the standard counseling instead. (Job seekers were not forced to accept the intensive counseling.)
 - Only 31.7% of job seekers did accept the intensive public counseling program.
- A control group of 4565 job seekers only got the standard counseling without any option to pick the intensive counseling.

Main dependent and explanatory variables.

5 / 31

- We are interested in how the treatment "intensive job counseling" affects job seekers' probability to find a job.
- More precisely, the main dependent variable `job_6m` studied by Behaghel et al. (2014) describes whether within 6 months a job seeker was no longer registered as unemployed because he found a job.
 - Whether or not indeed a job was found after 6 months could be determined only for roughly 90% of subjects. Our analysis is restricted to that subsample.
- Let treat_option_i be a dummy variable that is 1 if job seeker i was given the option for the intensive counseling treatment and 0 if not.
- Let treated_i be a dummy variable that is 1 if job seeker i indeed got the intensive counseling treatment, i.e. he got the option and accepted, and 0 if not.

Two OLS Regressions

6 / 31

- The first regression on the right regresses `job_6m` on `treat_option`. The coefficient 0.032 estimates the so called **intention to treat effect** (ITT effect). Interpretation: Giving job seekers the option for intensive counseling increases the average probability to find a job within 6 months by 3.2 percentage points.
- However, recall that only 31.7% of job seekers with the treatment option accepted the treatment. Hence the intention to treat effect probably underestimates the effect of the intensive counseling treatment on those job seekers who actually got it.
- The second regression regresses `job_6m` on the dummy `treated` that describes whether intensive counseling actually was received. Discuss why our estimate 0.079 probably is biased, i.e. why the 2nd regression does not consistently estimate the true treatment effect.

	<i>Dependent variable:</i>	
	<code>job_6m</code>	
	(1)	(2)
<code>treat_option</code>	0.032*** (0.010)	
<code>treated</code>		0.079*** (0.014)
Constant	0.201*** (0.006)	0.204*** (0.005)
Observations	7,224	7,224
R ²	0.002	0.004
Note:	* p<0.1; ** p<0.05; *** p<0.01	

The endogeneity problem

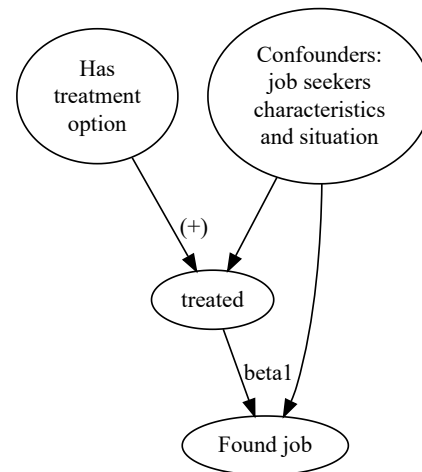
7 / 31

- We have estimated the regression

$$\text{job_6m}_i = \beta_0 + \beta_1 \text{treated}_i + u_i$$

where β_1 shall measure the causal effect of the treatment on the probability to find a job.

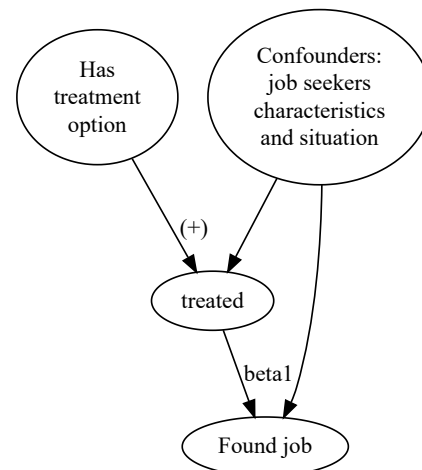
- The problem is that we most likely have unobserved confounders: Whether a job seeker accepts the intensive counseling probably depends on his characteristics and situation which also can affect the probability to find a job without intensive counseling. This means `treated` is endogenous and the OLS estimator $\hat{\beta}_1$ is biased in our short regression above.
- The sign of the bias is not obvious. One could think of different stories. E.g.
 - Job seekers who know that they can easily find themselves a good new job may more likely reject the intensive counseling because they don't find it necessary.
 - Alternatively, counseling may be systematically rejected by demotivated job seekers that have a low probability to find a job.



But we observe a source of exogenous variation

8 / 31

- While we have some background information about job seekers, like age, gender or education, it is probably impossible to control for all relevant confounders.
- However, with the variable `treat_option` we have data on a clear source of exogenous variation for our explanatory variable `treated`.
- In such a case, where one has data for a source of exogenous variation, one can consistently estimate the causal effect of an explanatory variable of interest using a method called **instrumental variable** estimation (short **IV** estimation).



Instrumental Variable Estimation in a Simple Linear Regression 9 / 31

- We will first explain IV estimation in a general setting. Afterward, we return to our job counseling application.
- Consider the simple linear regression:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

where due to confounders x is endogenous (correlated with u) so that the OLS estimator is inconsistent.

- An instrumental variable (short: instrument) z for the endogenous variable x in the simple regression is a variable that satisfies the following two conditions:
 - **Relevance:** z is correlated with the endogenous variable x : $\text{cor}(z_i, x_i) \neq 0$
 - **Exogeneity:** z is not correlated with the error term u : $\text{cor}(z_i, u_i) = 0$

IV-Estimation via “Two-Stage Least Squares” (2SLS) 10 / 31

- One can then perform the IV-estimation by a method called Two-Stage-Least-Squares, in which one runs two OLS estimations.
- 1st Stage: Regress via OLS the endogenous explanatory variable on the instrument:

$$x = \gamma_0 + \gamma_1 z + \eta$$

- Then compute the *predicted values* \hat{x} of this regression

$$\hat{x} = \hat{\gamma}_0 + \hat{\gamma}_1 z$$

- 2nd Stage: Estimate the original regression but substitute the endogenous variable by the predicted values from stage 1.

$$y = \beta_0 + \beta_1 \hat{x} + \varepsilon$$

- If the instrument z satisfies the relevance and exogeneity conditions, the OLS estimator $\hat{\beta}$ of this second stage is a consistent estimator of β .

Intuition for IV Estimation via 2SLS 11 / 31

- The first regression (asymptotically) removes from x all influences that are correlated with the error term u of the original regression and caused the endogeneity problem.
- The constructed variable \hat{x} only contains the exogenous variation induced by the instrument, which yields a consistent estimator $\hat{\beta}_1$ of the causal effect of x on y in the 2nd stage regression.
- Of course, this is just a very rough intuition. That the IV estimator indeed consistently estimates β_1 can be proven mathematically. Yet, we skip the proof, since we don't want to dig too deeply into the mathematics behind econometrics in this course.
- We will, however, illustrate that the method works with a small simulated data set in R.

R Simulation for IV estimation

12 / 31

```

n = 10000
u = rnorm(n,0,1) # error term
# z does not depend on u (exogeneity condition)
z = rnorm(n,0,1)
# x depends on u (endogeneous)
# and on z (z is relevant)
x = z+u+rnorm(n,0,1)
beta0 = 0; beta1 = 1
y = beta0 + beta1*x + u

coef(lm(y~x))

```

```

(Intercept)          x
-0.01906361  1.33450412

```

```

# OLS estimator beta1.hat has positive bias
# because cor(x,u) > 0

```

```

# IV estimation via 2SLS
# 1st Stage
reg1 = lm(x~z)
x.hat = fitted(reg1)
# 2nd Stage
reg2 = lm(y~x.hat)
coef(reg2) # consistent!

```

```

(Intercept)          x.hat
-0.02185464  1.01542039

```

```

# Direct IV estimation with function ivreg
# in AER package yields same estimates:
library(AER)
coef(ivreg(y~x|z))

```

```

(Intercept)          x
-0.02185464  1.01542039

```

Manual 2SLS approach vs ivreg function: different standard errors

13 / 31

- As in our simulation above, manually computing the 2SLS estimator or using the `ivreg` function for IV regression yield the same estimated coefficients.
- However, the standard errors are different: you should use the `ivreg` function for correct standard errors (and correct confidence intervals, and p-values)
- If you want to compute robust standard errors (or cluster-robust standard errors) for IV regression, you can use the function `felm` in the `lfe` package or the function `iv_robust` in the `estimatr` package.

IV regression with multiple instruments and control variables

14 / 31

- One can also perform IV regression with multiple instruments and control variables. Consider e.g. the regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

where x_1 is an endogenous variable of interest and x_2 and x_3 are two exogenous control variables.

- Assume we also have two valid instruments z_1 and z_2 .
- The Two-Stage-Least Squares approach would then be as follows:
 - 1st Stage: Regress the endogenous explanatory variable on all exogenous explanatory variables and all instruments:

$$x_1 = \gamma_0 + \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 x_2 + \gamma_4 x_3 + \eta$$

- 2nd Stage: As before estimate the original regression but just replace the endogenous variable x_1 by its fitted value \hat{x}_1 from the 1st stage regression.

$$y = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

- The `ivreg` command (assuming data is in 'dat') in this case would be:

```
ivreg(y ~ x1+x2+x3 | z1+z2+x2+x3, data=dat)
```

Before the `|` we write the original regression. Behind the `|` we add all instruments and all exogenous explanatory variables.

Optional: Some further facts about IV regression with control variables and multiple instruments

15 / 31

- As for the case of a simple linear regression also in the case with control variables and multiple instruments, each instrument z_k must satisfy the exogeneity condition: $cor(u_i, z_{k,i}) = 0$.
- The relevance condition is slightly modified. In the first stage regression the coefficient γ_k for instrument z_k should be unequal zero. It suffices if the relevance condition is satisfied for at least one instrument.
 - In practise the instruments should affect the endogenous variable x_1 sufficiently strong in the first stage regression to avoid a so called *weak instrument problem* that can lead to substantial biases in small samples. There exist formal *weak instrument tests* for this condition, which we don't discuss further here.
- Sometimes the exogenous control variables x_2 and x_3 are also called instruments (since they satisfy the exogeneity and relevance conditions). One then refers to z_1 and z_2 as *excluded instruments*.
 - If there is one endogenous variable one needs at least one excluded instrument.
- One can also perform IV regression with more than one endogenous variable. One then needs at least as many excluded instruments as endogenous variables.

IV Regression in our Job Counseling Application

16 / 31

- We now want to estimate via IV the regression

$$\text{job_6m}_i = \beta_0 + \beta_1 \text{treated}_i + u_i$$

with `treat_option` as instrument for `treated`.

- Let us first discuss whether `treat_option` is a valid instrument:

- Exogeneity Condition:

$$\text{cor}(\text{treat_option}_i, u_i) = 0$$

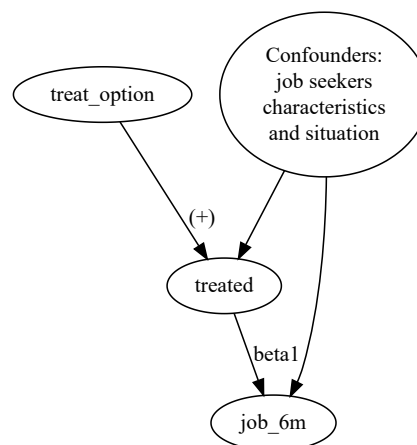
This should be satisfied if the experiment was indeed well randomized. Then whether a job seeker got the option for intensive counselling was completely random and independent of the job seekers characteristics.

- Relevance Condition:

$$\text{cor}(\text{treat_option}_i, \text{treated}_i) \neq 0$$

Should also be satisfied since only job seekers that got the option for intensive counselling can get the treatment.

- Note that the relevance condition can be checked with our data set, since we can compute that correlation. The exogeneity condition cannot be checked statistically, because we don't observe the error terms u_i . The exogeneity condition must always be assessed and discussed using your knowledge about how the data was generated.



IV Regression in our Job Counseling Application

17 / 31

- On the right you see the results of the IV regression and the previous two OLS regressions. The IV estimator shows the strongest effect of the intensive job counseling treatment.
 - Interpretation of the IV estimate: The intensive job counseling treatment increases the probability to find a job within 6 months by 10.2 percentage points.
 - We see from the constant that in the control group 20.1 percent of job seekers found a job. This means we estimate that the treatment increases the probability to find a job by roughly 50 percent. That seems quite a large amount.
- Note: If we multiply the IV estimator 0.102 by the share of job seekers that accept the treatment option 0.317, we get the intention to treat (ITT) estimate of 0.032.

	<i>Dependent variable:</i>		
	job_6m		
	OLS-ITT (1)	OLS (2)	IV (3)
treat_option	0.032*** (0.010)		
treated		0.079*** (0.014)	0.102*** (0.031)
Constant	0.201*** (0.006)	0.204*** (0.005)	0.201*** (0.006)
Observations	7,224	7,224	7,224
R ²	0.002	0.004	0.004
Note:	*p<0.1; **p<0.05; ***p<0.01		

Which average treatment effect does our IV regression estimate? 18 / 31

- In our regression specification

$$\text{job_6m}_i = \beta_0 + \beta_1 \text{treated}_i + u_i$$

we assumed that for every job seeker i the intensive counseling would have the same effect β_1 on the probability to find a job.

- As already noted in Chapter 3, one usually implicitly assumes that the coefficient of such a simple linear specification measures some average over heterogeneous treatment effects.
- In our application it seems quite likely that those job seekers who accept the intensive counseling benefit more from the treatment than those job seekers who reject the treatment would have benefited. (Probably when offered some treatment, those people are more willing to accept it who expect a larger benefit from the treatment.)
- But which average treatment effect does our IV estimator estimate? Is it the average treatment effect for all job seekers, or the average for those subgroup that is willing to accept the treatment?
- One can show that (under weak assumptions) our IV estimator estimates the average treatment effect for those job seekers that are willing to accept the treatment.

Rubin Causal Model

- In particular in settings with binary treatments and binary instruments, the popular modern statistical approach to determine which average of treatment effects an IV estimator measures builds upon the Rubin causal model.
 - The Rubin causal model differs from our previous theoretical approach that started with writing down a regression model with a certain functional form and an error term that shall satisfy certain assumptions.
- Let us denote by Y_i the outcome for individual i . In our example, we have $Y_i = \text{job_6m}_i$.
- The Rubin causal model assumes that every individual has two *potential outcomes*:
 - $Y_i(0)$ is the potential outcome if i would not be treated
 - $Y_i(1)$ is the potential outcome if i would be treated
- No particular distribution or functional form for the potential outcomes is assumed.
- It is assumed that the potential outcomes of each individual i are not influenced by how the treatment is assigned to other individuals. This is called the *stable unit treatment value assumption (SUTVA)*.
 - It sharpens the analytic mind to think about potential reasons why the SUTVA assumption may be violated in our application. For example, assume that under intensive counseling some counselors directly contact firms with open positions and strongly promote the job seekers they counsel. This may make it harder for non-treated job seekers to fill those positions and they thus possibly would have better potential outcomes if nobody received intensive counseling. One needs to investigate what counselors actually do to rule out such concerns.
 - Since Behaghel et al. don't mention such concerns, we assume the SUTVA assumption is satisfied.
- The treatment effect for individual i is $Y_i(1) - Y_i(0)$.
- Each individual can only be treated or not. Let treated_i be a dummy variable that is 1 if and only if i has been actually treated.
- For an individual i we can only observe one of the two potential outcomes:

$$Y_i = \begin{cases} Y_i(0) & \text{if } \text{treated}_i = 0 \\ Y_i(1) & \text{if } \text{treated}_i = 1 \end{cases}$$

Since we only observe one potential outcome, we can never observe directly the treatment effect for an individual i . This is sometimes called the *fundamental problem of causal inference*.

ATE and ATT

- The **average treatment effect (ATE)** is defined as

$$ATE = E(Y_i(1) - Y_i(0))$$

where the expectation is taken over all possible individuals i in the population.

- The **average treatment effect on the treated (ATT)** is defined as

$$ATT = E(Y_i(1) - Y_i(0) | \text{treated}_i = 1)$$

In general, whether or not a subject is treated can be correlated with her potential outcomes. E.g. subjects may be more likely to select themselves into the treatment if they have a large treatment effect. The ATT is the expected treatment effect conditional that we know that i is treated.

- While we can never observe the individual treatment effect $Y_i(1) - Y_i(0)$, the ATE or ATT can sometimes be consistently estimated.
- For example, if treated_i would be assigned in a perfectly randomized experiment, it would be independently distributed of the potential outcomes of individuals. Then the ATE and ATT would be the same and can be consistently estimated by the difference of the mean outcomes in the treatment and control groups (or by running a simple OLS regression with a treatment dummy).

IV Estimator and Local Average Treatment Effect

20 / 31

- Assume we have a dummy variable treat_option_i as instrument for treated_i . The instrument shall be statistically independent from the potential outcomes of the individual. That is the case in our application where the option to be treated was assigned in a perfectly randomized fashion.
- In our application we can distinguish two different types of individuals:
 - **Compliers** take the treatment if and only if suggested by the instrument, i.e. for compliers we have $\text{treated}_i = \text{treat_option}_i$
 - **Never-takers** never take the treatment, i.e. for never-takers we have always $\text{treated}_i = 0$
- In principle there could also be two other types:
 - **Always-takers** always manage to take the treatment (even if the instrument $\text{treat_option}_i = 0$), i.e. for always-takers we have always $\text{treated}_i = 1$
 - **Defiers** only take the treatment if not suggested by the instrument, i.e. for defiers we have $\text{treated}_i = 1 - \text{treat_option}_i$
- We don't have always-takers and defiers in our data set, but in other settings with different instruments they can exist.
- Imbens and Angrist (1994) have shown that if there are no defiers then the IV estimator consistently estimates the average treatment effect of the compliers. The literature uses different names for this effect: CATE (complier average treatment effect), CACE (complier average causal effect) or LATE (local average treatment effect):

$$CATE = CACE = LATE = E(Y_i(1) - Y_i(0) | i \text{ is complier})$$

- Given that there are no always-takers in our data set, only compliers are treated. This means the IV estimator estimates in our application also the ATT (average treatment effect on the treated).

No perfect random sampling

21 / 31

- So far we have assumed that job seekers were perfectly randomly assigned to the treatment and control groups. However, Behaghel et al. write in their article (p. 148):

Once the caseworker had assessed the job seeker's eligibility, he ran an Extranet program to randomly assign her to treatment 1 (public program), treatment 2 (private program), or the control group. The probabilities of assignment to each group varied locally so as to maximize the statistical power of the evaluation while complying with the quantitative objectives of each program (each local area had targets in terms of recipients of the two programs).

- This means the probabilities to end up in a treatment group or the control group can depend on the region the job seeker lives in.

Resulting endogeneity problem in our IV regression

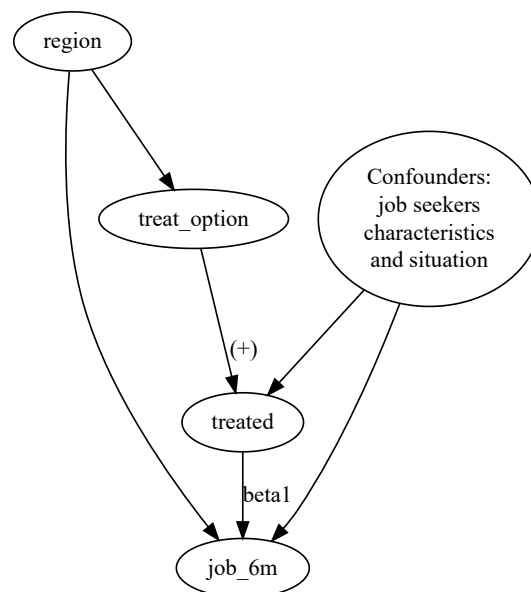
22 / 31

- Assume a job seekers region affects his treatment assignment, i.e. `treat_option` but also directly her chances to find a job (see graph on right).
- If we estimate the short regression

$$\text{job_6m} = \beta_0 + \beta_1 \text{treated} + u$$

via IV with `treat_option` as instrument, then `treat_option` would not satisfy the exogeneity condition and be an invalid instrument:

- If the `region` affects the chance to find a job, it is part of the error term u in our regression and we have $\text{cor}(\text{treat_option}_i, u_i) \neq 0$



Possible solution: Add region fixed effects to IV regression

23 / 31

- An solution to the endogeneity problem is to add region fixed effects, i.e. dummy variables for every region and estimate via IV the regression:

$$\text{job_6m} = \beta_0 + \beta_1 \text{treated} + \text{region dummies} + \varepsilon$$

- Now the region effects are "taken out" of the error term and the instrument `treat_option` is exogenous: $\text{cor}(\text{treat_option}_i, \varepsilon_i) = 0$
- As long as we are willing to make the assumption that the treatment effect itself is homogeneous across all regions (i.e. intensive counseling increases the probability to find a job by the same amount in all regions) this is a perfectly fine solution.

Alternative solution: Weighted regressions with inverse probability weighting

24 / 31

- Behaghel et al. don't add region fixed effects but use a different way to solve the endogeneity problem called *inverse probability weighting*. They write (slightly adapted):

In the regressions, we will use weights computed as the inverse of [...] the estimated assignment probabilities (which differ across regions) [...], to avoid imbalances between assignment groups. (p. 150)

In the presence of heterogeneous treatment effects, this is arguably better than using [region] fixed effects. One can indeed show that the fixed effect estimator artificially gives more weights to [regions] where assignment probabilities to the control and the treatment groups are close to each other. (footnote 16, working paper version)

- The following slides try to explain how this weighting is done, why it works, and how it can be better than adding region specific fixed effects.

Weighted Least Squares

25 / 31

- A weighted least squares (WLS) estimator solves the following optimization problem:

$$\hat{\beta}^{WLS} = \arg \min_{\hat{\beta}} \sum_{i=1}^n w_i \hat{u}_i(\hat{\beta})^2$$

where w_i is some specified weight for observation i .

- One can also perform weighted IV regression, e.g. via 2SLS with the weights w_i in both stages.
- We continue with an example that shall illustrate that if assignment probabilities to treatment and control groups differ by regions, computing an average treatment effect effectively boils down to give different weights to different observations.

An example with unequal assignment probabilities

26 / 31

- On the right we compute the sample ATE by hand for a simple two-region example.
- In this example the treatment shall be exogenously fixed. This means we don't have an instrumental variable setting where just the option to be treated is exogenously specified.
- Job seekers in region A have a lower baseline probability to find job (30% in control group) and a lower treatment effect.

	Region A		Region B	
	Control	Treated	Control	Treated
Sample Size	10	10	5	15
Share found job	30%	40%	60%	80%
Treatment effect	40% − 30% = 10%		80% − 60% = 20%	
Average Treatment Effect (ATE)	$\frac{20}{40} \cdot 10\% + \frac{20}{40} \cdot 20\% = 15\%$			

- We will see that this ATE computation can be replicated with a weighted linear regression. To see why different observations have different weights in our ATE calculation, answer the following question: By how much would the ATE change if an additional individual from region A / B in the control / treatment group would find a job? Which individuals have the largest impact on the ATE? How is an individual's impact on the ATE related to her region-specific probability to end up in her group (treatment or control)?

Solution

Let us compare the 4 cases for the additional individual that finds a job:

Region A, Control Group: Now 4 instead of 3 of the 10 control group members find a job. This leads to reduction of the ATE by 5 percentage points:

	Region A		Region B	
	Control	Treated	Control	Treated
Sample Size	10	10	5	15
Share found job	40%	40%	60%	80%
Treatment effect	40% − 40% = 0%		60% − 50% = 20%	
Average Treatment Effect (ATE)	$\frac{1}{2} \cdot 0\% + \frac{1}{2} \cdot 20\% = 10\%$			

Region A, Treatment Group: Now 5 instead of 4 of the 10 treatment group members find a job. This leads to an increase of the ATE by 5 percentage points:

	Region A		Region B	
	Control	Treated	Control	Treated
Sample Size	10	10	5	15
Share found job	30%	50%	60%	80%

Treatment effect	$50\% - 30\% = 20\%$	$80\% - 60\% = 20\%$
Average Treatment Effect (ATE)	$\frac{1}{2} \cdot 20\% + \frac{1}{2} \cdot 20\% = 20\%$	

Region B, Control Group: Now 4 instead of 3 of the 5 control group members find a job. This leads to reduction of the ATE by 10 percentage points:

	Region A		Region B	
	Control	Treated	Control	Treated
Sample Size	10	10	5	15
Share found job	30%	40%	80%	80%
Treatment effect	40% − 30% = 10%		80% − 80% = 0%	
Average Treatment Effect (ATE)	$\frac{1}{2} \cdot 10\% + \frac{1}{2} \cdot 0\% = 5\%$			

Region B, Control Group: Now 13 instead of 12 of the 15 treatment group members find a job. This leads to an increase of the ATE by just 3.33 percentage points.

	Region A		Region B	
	Control	Treated	Control	Treated
Sample Size	10	10	5	15
Share found job	30%	40%	60%	86.67%
Treatment effect	30% − 40% = 10%		86.67% − 60% = 26.67%	
Average Treatment Effect (ATE)	$\frac{1}{2} \cdot 10\% + \frac{1}{2} \cdot 26.67\% = 18.33\%$			

We see that the outcome of a control group member in region B has the largest effect on the estimated ATE in our example. Because we have so few members in this subgroup, a single member has a larger impact.

	Region A		Region B	
	Control	Treated	Control	Treated
Sample Size	10	10	5	15
Probability to be selected in control / treatment group	1/2	1/2	1/4	3/4
Inverse of the selection probability	2	2	4	4/3
Absolute impact on ATE of one individual in percentage points	5	5	10	3.33

We see that the impact of one individual on the ATE is proportional (here multiplied by 2.5) to the inverse of the probability that the individual is selected in its group (treatment or control).

This means our computation of the ATE essentially weights observations by the inverse of those selection probabilities.

4 Regressions for our example

We run 4 regressions with our example data set from the previous slide:

1. The simple OLS regression yields a biased estimate of the treatment effect because region B has both a higher share of treated individuals and larger baseline values of the dependent variable.
2. The 2nd regression is the weighted least squares regression using the inverse selection probabilities as weights. The coefficient for `treated` estimates consistently the average treatment effect.
3. The third regression does not weight observations but just controls for `region`. If we would have homogeneous treatment effects, that would be perfectly fine. With heterogeneous treatment effects, coefficient for `treated` is still a slightly biased estimator of the ATE. one can show that with region fixed effects those regions with more equal shares in the treatment and control groups get a larger weight. Here the coefficient leans a bit closer to the treatment effect in region A (10) than to that in region B (20).
4. The 4th column adds the interaction term between `treated` and `region`. Given that each region has the sample size, we can compute the ATE from this regression by the coefficient before `treated` (0.1) plus half the coefficient of the interaction term `treated:regionB` ($0.5 \cdot 0.1$). This also yields 0.15. Using interaction terms to account for the heterogeneity, also yields a consistent estimator of the ATE.

	SAMPLE ATE = 0.15			
	OLS (1)	Weighted (2)	OLS (3)	OLS (4)
<code>treated</code>	0.240	0.150	0.143	0.100
<code>regionB</code>			0.364	0.300
<code>treated:regionB</code>				0.100
Constant	0.400	0.450	0.279	0.300
Observations	40	40	40	40
Note:	*p<0.1; **p<0.05; ***p<0.01			

Discussion: Inverse Probability Weighting

28 / 31

- Our example illustrated that inverse probability weighting allows to consistently estimate the average treatment effect if probabilities to be assigned to treatment differ among regions or other subgroups of subjects.
- In our example, we took the assignment probabilities as exogenously given. If one does not know the assignment probabilities, one can also estimate them in an initial step by regressing the `treatment` dummy on the region dummies or other characteristics that determine subgroups. (Usually, one would use a logit or probit regression in this step to guarantee that the estimated probabilities are between 0 and 1. You won't need to know those methods for this course).
- Behaghel et al. use such inverse probability weighting in an instrumental variable context. It allows them to consistently estimate the LATE/CATE (complier average treatment effect) even though the probability to be treated differs among regions.
- Instead of using inverse probability weighting, one could also estimate a regression that contains all relevant interaction effects in order to estimate the ATE (or LATE in an IV context).
- This blog post gives some additional information on the topic:
<https://declaredesign.org/blog/biased-fixed-effects.html>

Results for Intensive Job-Counseling Program

29 / 31

- The 3rd column in the regression table shows the results of the weighted IV regression of Behaghel et al.
- In our application, the estimated treatment effects differ only very little between the weighted and unweighted IV regression. This suggests that imbalances in the treatment assignment don't play an important role in our data set.

	<i>Dependent variable:</i>		
	job_6m		
	OLS	IV	IV weighted
	(1)	(2)	(3)
treated	0.079*** (0.014)	0.102*** (0.031)	0.100*** (0.032)
Constant	0.204*** (0.005)	0.201*** (0.006)	0.230*** (0.007)
Observations	7,224	7,224	7,224
R ²	0.004	0.004	0.003
Note:	*p<0.1; **p<0.05; ***p<0.01		

- Another treatment in the job counseling experiment was intensive counseling by private firms. That treatment effect can be analyzed using the same methodology (IV regression with inverse probability weights).
- The results on the right show that the intensive public treatment had a much stronger positive treatment effect for job seekers than the intensive private counseling (more than twice as large). This result has important policy implications for the question in how far job counseling should be outsourced to private providers.

	<i>Dependent variable:</i>	
	Public	Private
	(1)	(2)
treated	0.100*** (0.032)	0.042** (0.017)
Constant	0.230*** (0.007)	0.230*** (0.007)
Observations	7,224	34,883
Note:	* p<0.1; ** p<0.05; *** p<0.01	

References

- Behaghel, Luc, Bruno Crépon, and Marc Gurgand. (2014). "Private and Public Provision of Counseling to Job Seekers: Evidence from a Large Controlled Experiment." *American Economic Journal: Applied Economics*, 6 (4): 142-74.
- Imbens, G., & Angrist, J. (1994). "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2), 467-475. doi:10.2307/2951620