

LECTURE 5 Q&A
PUBL0050 ADVANCED QUANTITATIVE METHODS

Indraneel Sircar & Julia de Romémont

Term 2 2021-22
UCL Department of Political Science

Slide 34 of Lecture 5, would you be able to elaborate further on the variable which is adjoint to the Beta coefficient in the equation please?

Estimator (Regression 2)

With panel data we can use regression with first differences:

$$\Delta Y_i = \alpha + \delta \cdot D_i + \mathbf{X}'_i \beta + u_i,$$

where $\Delta Y_i = Y_i(1) - Y_i(0)$.

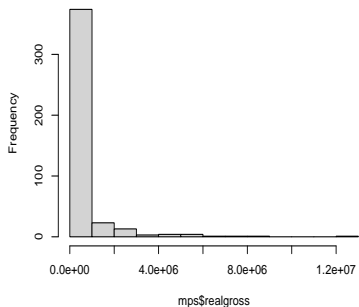
- ▶ With two periods gives identical result as other regressions
- ▶ Where $\mathbf{X}'_i \beta$ is (the inverse of) a matrix of control variables \mathbf{X}'_i multiplied by a vector of β coefficients

When and why should we take the natural logarithm?

- ▶ Linear regression assumes that variables (if sampled infinitely) are normally distributed
- ▶ Some variables however are very skewed, with a long tail; for example: income or wealth
- ▶ Taking the logarithm 'normalises' these

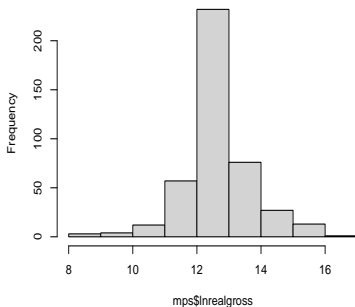
```
hist(mps$realgross)
```

Histogram of mps\$realgross



```
hist(mps$lnrealgross)
```

Histogram of mps\$lnrealgross



Could we combine Matching with DiD method? Could it possible to show some code?

- ▶ Yes we can! In fact, there have been some innovations in this respect recently
 - ▶ See the papers on [this page](#) and the associated package [PanelMatch](#)
 - ▶ The details there are a bit complicated and this applies more to contexts when treatment does not happen all at the same time

- Where treatment happens to all treated observations at once, we can use the tools we have already learned about:

```
# subset to pre-treatment period
sun92 <- sun[sun$year==1992,]
```

```
# match on covariates AND pre-treatment outcome
library(MatchIt)
match.out <- matchit(reads_sun ~ voted_lab + wkclass +white +
                    parent_labour, data= sun92,
                    replace=F,
                    ratio = 1:1)

summary(match.out)$nn
```

##	Control	Treated
## All (ESS)	1382	211
## All	1382	211
## Matched (ESS)	211	211
## Matched	211	211
## Unmatched	1171	0
## Discarded	0	0

```
# extract matches
```

```
matches <- match.data(match.out)
```

```
# keep only matches
```

```
sun_matched <- sun[sun$id %in% matches$id,]
```

```
# regression
```

```
dd_mod_match <- lm(voted_lab ~ reads_sun*as.factor(year),  
                  data = sun_matched)
```

```
dd_mod <- lm(voted_lab ~ reads_sun*as.factor(year), data = sun)
```

	Voted Labour	
	full sample	matched sample
	(1)	(2)
Intercept	0.32*** (0.01)	0.39*** (0.03)
Reads the Sun	0.07* (0.04)	-0.005 (0.05)
1997	0.11*** (0.02)	0.14*** (0.05)
Reads the Sun x 1997	0.09* (0.05)	0.05 (0.07)
Observations	3,186	844
R ²	0.02	0.03

Could you give an example of repeated cross-section data?

- ▶ Panel data is a (representative) sample of a population, that is observed a several time points
- ▶ Repeated cross-sectional data also are observed at several time points, but at each time period it is a different (representative) sample of the target population
 - ▶ Eg. Opinion polls etc. . .

*I did not see treatment2 being used again in the stargazer code.
(muni_1516\$treatment2)*

- ▶ That is a typo in the pdf version of the seminar - it's been corrected

Can you clarify the refugee per capita?

- ▶ `trarrprop` is a continuous variable that measures *intensity* of treatment rather than whether a unit is treated or not

I don't understand these 'interaction_model <- lm(gdvote ~ ever_treated + post_treatment + treatment,' three variables to calculate the interaction model?

- ▶ The variable `treatment` already is the interaction of `ever_treated` and `post_treatment`
- ▶ We need the main effects and the interaction for the model to make sense

Why is the interaction not: `lm(data ~ variable a + variable b + variable a*b)` for `interaction_model <- lm(voted ~ treatment*female, data = gerber)?`

SEMINAR 2

- ▶ `treatment*female` automatically includes both the interaction and the main effects

```
interaction_model <- lm(voted ~ treatment*female, data = gerber)
interaction_model2 <- lm(voted ~ treatment + female + treatment:female,
                        data = gerber)
```

```
# are the coefficients the same?
```

```
coef(interaction_model) == coef(interaction_model2)
```

```
##           (Intercept)           treatmentCivic Duty
##                TRUE                TRUE
##   treatmentHawthorne   treatmentSelf
##                TRUE                TRUE
##   treatmentNeighbors           female
##                TRUE                TRUE
## treatmentCivic Duty:female treatmentHawthorne:female
##                TRUE                TRUE
##   treatmentSelf:female   treatmentNeighbors:female
##                TRUE                TRUE
```

```
# yes!
```

The stargazer function did not work for me to make the table.
stargazer(post_treatment_period_regression, ...