Session 4: Model specification

MGT 581 | Introduction to econometrics

Michaël Aklin

PASU Lab | EPFL

Last time...

- Descriptive statistics: first, second moment
- Association: covariance, correlation
- Linear model (theory)
- OLS with > 1 independent variables

Plan for today

- Model specification: nonlinearities, transformations (log), heterogeneous treatment effects, interaction effects
- Readings: Stock and Watson (2011) (ch8), Verbeek (2018) (ch3)

Model specification

- **Linear** model (linear in parameters)
- Two potential issues

1. If all vars are dummy (w/ full interaction): linearity is met by construction. Draw!

- But not necessarily true with continuous treatment

2. True theoretical model is not linear

- Eg:

$$\text{Revenues}_i = L_i^{\beta} K_i^{\lambda}$$

- Not linear in parameters... but sometimes can be transformed into LM!

$$\log(\text{Revenues})_i = \beta \log(\text{L})_i + \lambda \log(\text{K})_i$$

- But again, fine if we have dichotomous vars and full interactions

Model specification

1. Linearity of treatment
2. Logs
3. Quadratic functions
4. Interaction effects

Linearity of treatment

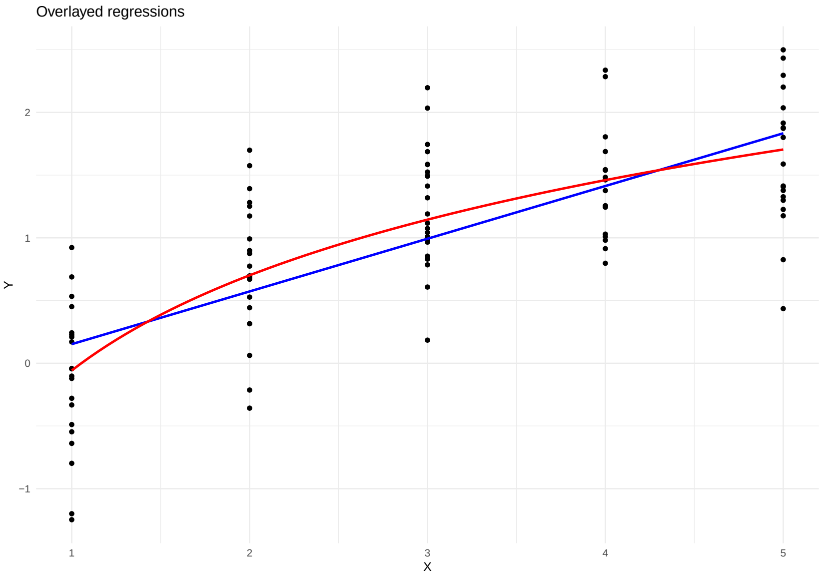
- Suppose the true model is:

$$Y = \beta log(X) + \varepsilon$$

- But by error we estimate:

$$Y = \gamma X + \varepsilon$$

- What will happen?


Overlayed regressions
Y
X
2
1
0
−1
1
2
3
4
5

Logarithms

- Natural logarithms are very common in economic/policy applications
- Reason 1: fits well with microeconomic theory (elasticities)
- Reason 2: often derived from multiplicative models
- Eg Cobb-Douglas function

$$\text{Revenues}_i = L_i^{\beta} K_i^{\lambda}$$

- Can be transformed:

$$\log(\text{Revenues})_i = \beta \log(\text{L})_i + \lambda \log(\text{K})_i$$

Four cases

There exist 4 models:

- Linear-linear (conventional LM)
- Linear-log
- Log-linear
- Log-log

Linear-log

$$Y_i = \alpha + \beta \text{log}(X_i) + \varepsilon_i$$

- $log(X) + 1 = log(X) + log(e) = log(eX)$
- This means: adding 1 is the same as multiplying X by 2.72 (e).
- It's the same as increasing X by 172%.
- For other proportions p: multiply β by log([100+p]/100).
- Why? We want $log([1 + p/100] \cdot X)$
- For instance: $log([1 + 1/100] \cdot X) = log(X \cdot 101/100)$

- Eg: 10% requires multiplying β by $log(1.1) = log(110/100)$.
- Eg: 15% requires multiplying β by $log(1.15) = log(115/100)$.
- Trick: for small p: $log([100 + p]/100) \approx p/100$.
- Thus: divide p by 100.
- Eg treatment+1% $\rightarrow \beta * 1/100$.

Log-linear

$$\log(Y_i) = \alpha + \beta X_i + \varepsilon_i$$

- An increase of X by 1 equiv to multiply Y by $exp(\beta)$. For c units: $exp(c\beta)$.
- For small β: $exp(\beta) = 1 + \beta$.
- This means that Y is multiplied by $1 + \beta$.
- Thus: if $\beta = 0.05$, then Y is multiplied by $1 + 0.05$, i.e., increases by 5%.
- In other words: Y increases by 100β percent.

Log-log

$$\text{log}(Y_i) = \alpha + \beta\text{log}(X_i) + \varepsilon_i$$

- Combines the two previous interprations.
- β is % change in Y when X increases by 1%.
- In microeconomics, this is the elasticity of Y.

```
> model = lm_robust(data=data_combined, log(GDP_Per_Capita) ~ pm25)
> summary(model)

Call:
lm_robust(formula = log(GDP_Per_Capita) ~ pm25, data = data_combined)

Standard error type:  HC2

Coefficients:
            Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper  DF
(Intercept)  9.96203   0.142759  69.782 2.670e-162  9.68082 10.24324 242
pm25        -0.04259   0.004651  -9.157  2.322e-17 -0.05175 -0.03343 242

Multiple R-squared:  0.3113 ,   Adjusted R-squared:  0.3085
F-statistic: 83.85 on 1 and 242 DF,  p-value: < 2.2e-16
```

An increase of PM2.5 by one micro g/m3 leads to a reduction of (expected) GDP per capita by 4%.

```
> model = lm_robust(data=data_combined, pm25 ~ log(GDP_Per_Capita))
> summary(model)

Call:
lm_robust(formula = pm25 ~ log(GDP_Per_Capita), data = data_combined)

Standard error type:  HC2

Coefficients:
                    Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
(Intercept)           92.307     6.6385   13.90 1.063e-32   79.231  105.384 242
log(GDP_Per_Capita)   -7.309     0.7254  -10.08 3.631e-20   -8.738   -5.881 242

Multiple R-squared:  0.3113 ,   Adjusted R-squared:  0.3085
F-statistic: 101.5 on 1 and 242 DF,  p-value: < 2.2e-16
```

Increase in GDP/capita by 1% $\rightarrow$ reduction in average PM25 by 0.07 micro g/m3

```
> model = lm_robust(data=data_combined, log(GDP) ~ log(Population))
> summary(model)

Call:
lm_robust(formula = log(GDP) ~ log(Population), data = data_combined)

Standard error type:  HC2

Coefficients:
                Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper  DF
(Intercept)      11.1036    0.40484   27.43  3.622e-78  10.3064  11.9009 256
log(Population)   0.8621    0.02456   35.10 7.712e-100   0.8138   0.9105 256

Multiple R-squared:  0.7914 ,   Adjusted R-squared:  0.7906
F-statistic:  1232 on 1 and 256 DF,  p-value: < 2.2e-16
```

Increase in pop by 1% $\rightarrow$ increase in GDP by 0.8%

Quadratic functions

Consider the following model:

$$Y_i = \alpha + \beta X_i + \gamma X_i^2 + \varepsilon_i$$

- Relation between X and Y is nonlinear (quadratic).
- If $\gamma > 0$: U
- If $\gamma < 0$: inverted-U
- Marginal effect is not a constant:

$$\frac{\partial E[Y]}{\partial X} = \beta + 2\gamma X$$

- We can solve for a max or min:

$$\frac{\partial Y}{\partial X} = \beta + 2\gamma X = 0$$
$$X = -\frac{\beta}{2\gamma}$$

- General rule: very strong parametric assumption (symmetry)
- Seldom a realistic data-generating process
- Note: you could add higher powers (x^3, x^4, etc.) for more flexibility
- Good alternative: bin X and let the data show you how parabolic it really is
- Alternative: semi-parametric models (eg local regressions):

$$y = f(x) + \gamma Z + \varepsilon$$

- Equivalent to run regressions on small sections connected by splines

Interaction effects

- Two types of heterogeneity...

1. Idiosyncratic heterogeneity in treatment effects: TE are not constant
2. Systematic heterogeneity: TE varies by group.

- If unknown group: machine learning
- If known+observable (ie can be measured) group: interaction effects
- Idea: is the ATE different for some groups than for others?
- Eg: is effect of *ads* (D) on *support for Presidential Candidate A* (Y) different for men and women?

$$E[Y|D, X = i] \neq E[Y|D, X = j]$$

- Systematic heterogeneity can be captured by **interaction effects**

$$Y = \alpha + \beta D_i + \lambda X_i + \gamma D_i X_i + \varepsilon_i$$

- Suppose that X is a dummy. Then we have two different effects (one for $X = 0$ and one for $X = 1$):

$$\frac{\partial Y}{\partial D} = \beta | X = 0$$
$$\frac{\partial Y}{\partial D} = \beta + \gamma | X = 1$$

- If X is continuous, then we have:

$$\frac{\partial Y}{\partial D} = \beta + \gamma X_i$$

$\rightarrow$ TE of D varies depending on value of X_i

Example

```
> summary(lm_robust(data=growth, growth ~ yearsschool + oil + yearsschool:oil))

Call:
lm_robust(formula = growth ~ yearsschool + oil + yearsschool:oil,
    data = growth)

Standard error type:  HC2

Coefficients:
                Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
(Intercept)       0.8708    0.45352  1.9201 0.059528 -0.03608   1.7776 61
yearsschool       0.2657    0.08568  3.1006 0.002922  0.09433   0.4370 61
oil               1.3729    2.27675  0.6030 0.548750 -3.17978   5.9255 61
yearsschool:oil  -0.2730    0.34481 -0.7919 0.431505 -0.96253   0.4164 61

Multiple R-squared:  0.1191 ,   Adjusted R-squared:  0.07576
F-statistic: 3.212 on 3 and 61 DF,  p-value: 0.02905
```

$$\text{Growth} = 0.9 + 0.27\text{Schooling} + 1.4\text{Oil} - 0.27\text{School*Oil}$$

Interpretation: positive effect of schooling disappears in oil-producing countries

Concluding on interactions

- Note: in the bivariate case, with $D \in \{0, 1\}$: linearity holds by construction
- In the multiple regression case: as long as all variables are dummy, and you include all interactions: linearity holds as well!
- Just need to make sure that you have a combination of parameters for each category

Questions?

References

Stock, James H., and Mark W. Watson. 2011. *Introduction to Econometrics, 3rd Edition*. Pearson.

Verbeek, Marno. 2018. *A Guide to Modern Econometrics 5th Edition*. Wiley.