QS101: Introduction to Quantitative Methods in Social Science

Week 20: The Classical Model

Dr. Florian Reiche
Teaching Fellow in Quantitative Methods
Course Director BA Politics and Sociology
Deputy Director of Student Experience and Progression, PAIS

March 12, 2015

Recap

The Classical Model

Recap

Queries

- Which test do we use to test for the significance of coefficients in a regression?

Queries

- Which test do we use to test for the significance of coefficients in a regression?
- How do we interpret the constant?

Queries

- Which test do we use to test for the significance of coefficients in a regression?
- How do we interpret the constant?
- What does *ceteris paribus* mean?

Queries

- Which test do we use to test for the significance of coefficients in a regression?
- How do we interpret the constant?
- What does *ceteris paribus* mean?
- What is our regression equation here?

```
. regress b_fimngrs_dv b_netuse b_jbhrs
```

Source	SS	df	MS
Model	1.6391e+10	2	8.1955e+09
Residual	8.8781e+10	40872	2172163.73
Total	1.0517e+11	40874	2573068.81

Number of obs	= 40875
F(2, 40872)	= 3772.95
Prob > F	= 0.0000
R-squared	= 0.1558
Adj R-squared	= 0.1558
Root MSE	= 1473.8

b_fimngrs_dv	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
b_netuse	54.86892	3.428123	16.01	0.000	48.14972	61.58811
b_jbhrs	26.82213	.355873	75.37	0.000	26.12462	27.51965
_cons	1008.081	18.51395	54.45	0.000	971.793	1044.368

The Classical Model

The Fine Print

- OLS *is* the best estimator available for regression models . . .

The Fine Print

- OLS *is* the best estimator available for regression models ...
- ... provided certain assumptions hold

The Fine Print

- OLS *is* the best estimator available for regression models ...
- ... provided certain assumptions hold
- These assumptions are referred to as the "Classical Model"

The Fine Print

- OLS *is* the best estimator available for regression models ...
- ... provided certain assumptions hold
- These assumptions are referred to as the "Classical Model"
- Your job is therefore to decide whether these assumptions hold, and whether it might be necessary to select a different method

The Classical Assumptions

1. The regression model is linear, is correctly specified, and has an additive error term.

The Classical Assumptions

1. The regression model is linear, is correctly specified, and has an additive error term.
2. The error term has a zero population mean

The Classical Assumptions

1. The regression model is linear, is correctly specified, and has an additive error term.
2. The error term has a zero population mean
3. All explanatory variables are uncorrelated with the error term

The Classical Assumptions

1. The regression model is linear, is correctly specified, and has an additive error term.
2. The error term has a zero population mean
3. All explanatory variables are uncorrelated with the error term
4. Observations of the error term are uncorrelated with each other (no serial correlation)

The Classical Assumptions (contd.)

5. The error term has a constant variance (no heteroskedasticity)

The Classical Assumptions (contd.)

5. The error term has a constant variance (no heteroskedasticity)
6. No explanatory variable is a perfect linear function of any other explanatory variable(s) (no perfect multicollinearity)

The Classical Assumptions (contd.)

5. The error term has a constant variance (no heteroskedasticity)
6. No explanatory variable is a perfect linear function of any other explanatory variable(s) (no perfect multicollinearity)
7. The error term is normally distributed (optional)

1. Correct Specification

- No omitted variabes

1. Correct Specification

- No omitted variabes
- No incorrect functional form

1. Correct Specification

- No omitted variabes
- No incorrect functional form
- A stochastic error term is added

1. Correct Specification (contd.)

The regression model is assumed to be linear

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \tag{1}$$

1. Correct Specification (contd.)

The regression model is assumed to be linear

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \quad (2)$$

This would also be true, if one of our variables had a quadratic specification:

$$x_i^* = (x_i)^2 \quad (3)$$

Then

$$Y_i = \beta_0 + \beta_1 X_{1i}^* + \beta_2 X_{2i} + \epsilon_i \quad (4)$$

is still linear.

2. Zero Population Mean of the Error Term

- Error term is added to account for variation in the dependent variable that cannot be explained by the model

2. Zero Population Mean of the Error Term

- Error term is added to account for variation in the dependent variable that cannot be explained by the model
- When the entire population of possible values for the stochastic error term is considered, then the average value of that population is zero

2. Zero Population Mean of the Error Term

- Error term is added to account for variation in the dependent variable that cannot be explained by the model
- When the entire population of possible values for the stochastic error term is considered, then the average value of that population is zero
- At least this is true for large samples.

2. Zero Population Mean of the Error Term (contd.)

- In small samples, the mean of ϵ might not be zero

2. Zero Population Mean of the Error Term (contd.)

- In small samples, the mean of ϵ might not be zero
- OLS forces ϵ to be zero, however, by way of the constant

Example

Consider the typical regression equation:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{5}$$

Now, suppose ϵ had a mean of 3, then

$$Y_i = (\beta_0 + 3) + \beta_1 X_i + (\epsilon_i - 3) \tag{6}$$

We can write this in turn in the form of a zero statistic mean:

$$Y_i = \beta_0^* + \beta_1 X_i + \epsilon_i^* \tag{7}$$

3. No Correlation between Explanatory Variables and Error Term

- IF an explanatory variable was correlated with the error term, then OLS would attribute some of the variation in Y attributed for by the error term

3. No Correlation between Explanatory Variables and Error Term

- IF an explanatory variable was correlated with the error term, then OLS would attribute some of the variation in Y attributed for by the error term
- This induces a bias in the estimates of our coefficient

3. No Correlation between Explanatory Variables and Error Term

- IF an explanatory variable was correlated with the error term, then OLS would attribute some of the variation in Y attributed for by the error term
- This induces a bias in the estimates of our coefficient
- Most often violated by leaving an important independent variable out

4. Observations of the Error Term are uncorrelated with each other

- Observations of the error term are drawn independently from one another

4. Observations of the Error Term are uncorrelated with each other

- Observations of the error term are drawn independently from one another
- IF systematic correlation exists, then the standard errors of the coefficients are inaccurate

4. Observations of the Error Term are uncorrelated with each other

- Observations of the error term are drawn independently from one another
- IF systematic correlation exists, then the standard errors of the coefficients are inaccurate
- Most important in time-series contexts (next year's module)

5. No Heteroskedasticity

- The variance of the distribution from which the the observations of the error term are drawn is constant

5. No Heteroskedasticity

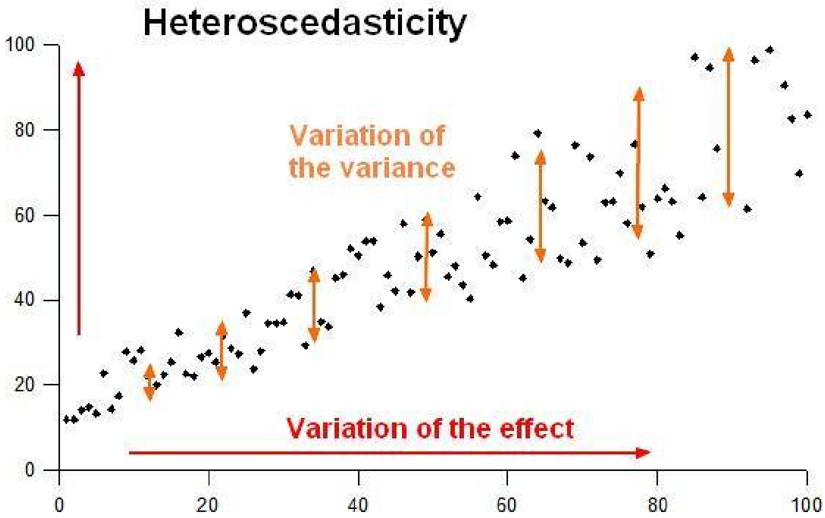
- The variance of the distribution from which the the observations of the error term are drawn is constant
- Contrary would be that the variance of the distribution of the error terms different for each observation

5. No Heteroskedasticity

- The variance of the distribution from which the the observations of the error term are drawn is constant
- Contrary would be that the variance of the distribution of the error terms different for each observation
- The actual values of the error term are not observable, but heteroskedasticity leads to inaccurate results

5. No Heteroskedasticity

- The variance of the distribution from which the the observations of the error term are drawn is constant
- Contrary would be that the variance of the distribution of the error terms different for each observation
- The actual values of the error term are not observable, but heteroskedasticity leads to inaccurate results
- Likely to occur in cross-sectional data


Heteroscedasticity
Variation of the variance
Variation of the effect
100
80
60
40
20
0
0
20
40
60
80
100

6. No Perfect Multicollinearity

- Perfect collinearity between independent variables implies that they are really the same variable
- For example: One variable is the exact multiple of another

6. No Perfect Multicollinearity

- Perfect collinearity between independent variables implies that they are really the same variable
- For example: One variable is the exact multiple of another
- Then, OLS is incapable of distinguishing these from one another

6. No Perfect Multicollinearity

- Perfect collinearity between independent variables implies that they are really the same variable
- For example: One variable is the exact multiple of another
- Then, OLS is incapable of distinguishing these from one another
- If more than two variables are affected, we speak of multicollinearity

6. No Perfect Multicollinearity (contd.)

- Way out: drop one of the variables affected

6. No Perfect Multicollinearity (contd.)

- Way out: drop one of the variables affected
- You can test for multicollinearity in STATA

6. No Perfect Multicollinearity (contd.)

- Way out: drop one of the variables affected
- You can test for multicollinearity in STATA
- Make sure to consult Section 10.7.3. in the Acock book