Tutorial 5

Hong Xiang Yue

28/03/2019

Admin and Part A

Midsemester test

Figure 1: Hopefully this isn't you before the test

Part B

Question 1: CAPM

$$E[R_i] - rf = \beta_i(E[R_m] - rf)$$

- Excess return of an asset is a function of the market's excess return
- While risk = return, under CAPM, you are only rewarded for taking on *systematic* risk and not *idiosyncratic* risk which can be diversified away
- How much you are rewarded by is given by β the sensitivity of the asset's excess returns to the market's excess returns
- β is calculated by running a regression of the asset's returns against the market's returns
- In practice the marekt portfolio will be some sort of index like the All Ords or the SP500

Question 1a

Suppose we have the excess returns of Qantas which we denote y_t and the excess returns of the All Ords x_t. We are interested in running a regression of the form:

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t = 1, .., n$$

We assume that $E[\mathbf{u}|\mathbf{X}] = 0$ and we use $\tilde{\beta}_1 = \frac{y_n - y_1}{x_n - x_1}$ as an estimate of β_1.

Show that

$$E[\tilde{\beta}_1] = \beta_1 + E[\frac{u_n - u_1}{x_n - x_1}]$$

.

Hint: In place of y_n and y_1 substitute their expressions given by the regression model and rearrange the fraction.

Question 1a: Continued

Using $E[\mathbf{u}|\mathbf{X}] = 0$ now show that $E[\tilde{\beta}_1] = \beta_1$ i.e. $E[\tilde{\beta}_1]$ is an unbiased estimate of β_1.

Hint: Use the fact that since $E[\mathbf{u}|\mathbf{X}] = 0$, then $E[\mathbf{u}] = 0$. i.e. If the *conditional* expectation is a constant then the *unconditional* expectation is also a constant.

Unbiasedness

An estimator is unbiased if the expected value of the estimator is equal to the true value of the parameter which it is trying to estimate.

- e.g. The sample mean $\bar{X}$ is an unbiased estimator of the population mean μ_X because $E[\bar{X}] = \mu_X$

In our case, we've shown that $E[\tilde{\beta}_1] = \beta_1$ so $\tilde{\beta}_1$ is an unbiased estimator of β_1

- The OLS estimator is also an unbiased estimator of β_1 provided assumptions E.1, E.2 and E.3 hold

A bit more on assumption E.3: Zero conditional mean

The assumption that $E[\mathbf{u}|\mathbf{X}] = 0$ is required for unbiasedness. It can seem more abstract than the other assumptions, but essentially it means that:

- The theoretical errors of our model should on average be zero, AND are not a function of the explanatory variables
- There is no correlation between the errors of our model and explanatory variables

A bit more on assumption E.3: Zero conditional mean

E.3 can be violated when:

- We omit a variable which is associated with **y** but also correlated with one or more explanatory variables in **X**
- e.g. A model of the form $wages = \beta_0 + \beta_1 education + u$ with *ability* omitted because we can't measure it
- We misspecify the functional form of the dependent variable
- e.g. We model y as a linear function of x, when in fact it is also a function of $x^2, x^3, ...$, etc.

Question 1b

Now, also assume that $Var(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I}_n$.

Can $\tilde{\beta}_1$ have a smaller variance than the OLS estimator? i.e. Is $\tilde{\beta}_1$ more precise than $\hat{\beta}_{1,OLS}$?

Hint: Use the Gauss-Markov assumptions from lecture 4 to prove or disprove this.

https://flux.qa/LDPPHD

Question 2

- Download the file hprice.wf1 and estimate the model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u$$

- What is the estimated equation?
- What is the estimated price increase for a house with one more bedroom if we hold *sqrft* constant?
- What is the estimated price increase for a house with an additional bedroom that is 140 square feet?
- What percentage of the variation in house price is explained by sqrft and bdrms? https://flux.qa/LDPPHD
- What is the estimated price of the first house in our sample? (Use the regression coefficients)
- What is the residual of the prediction for the first observation? Does this suggest that the buyer over or underpaid?

Question 3

- Download the bodyfat.wf1 workfile and create a scatterplot matrix between the three variables
- Based on the scatterplots, if we ran two regression with dependent variable body_fat and explanatory variable either wgk OR abdomen, what we would be the sign of the slope coefficients?
- Which of the regression would be a better fit?
- If we ran a regression of body_fat with TWO explanatory variables abdomen AND wkg, what would be the sign of the coefficient on wkg?
- https://flux.qa/LDPPHD

Question 4: verification of question 3

- Now let's actually run these regression and see what we get:

1. Estimate $body\_fat = \beta_0 + \beta_1 wkg$
2. Estimate $body\_fat = \delta_0 + \delta_1 abdomen$
3. Estimate $body\_fat = \gamma_0 + \gamma_1 abdomen + \gamma_2 wkg$

- Why is the sign of the coefficient on wkg negative in the third regression?
- If weight was measured in pounds rather than in kilograms, how would the coefficients change in regression 3? https://flux.qa/LDPPHD