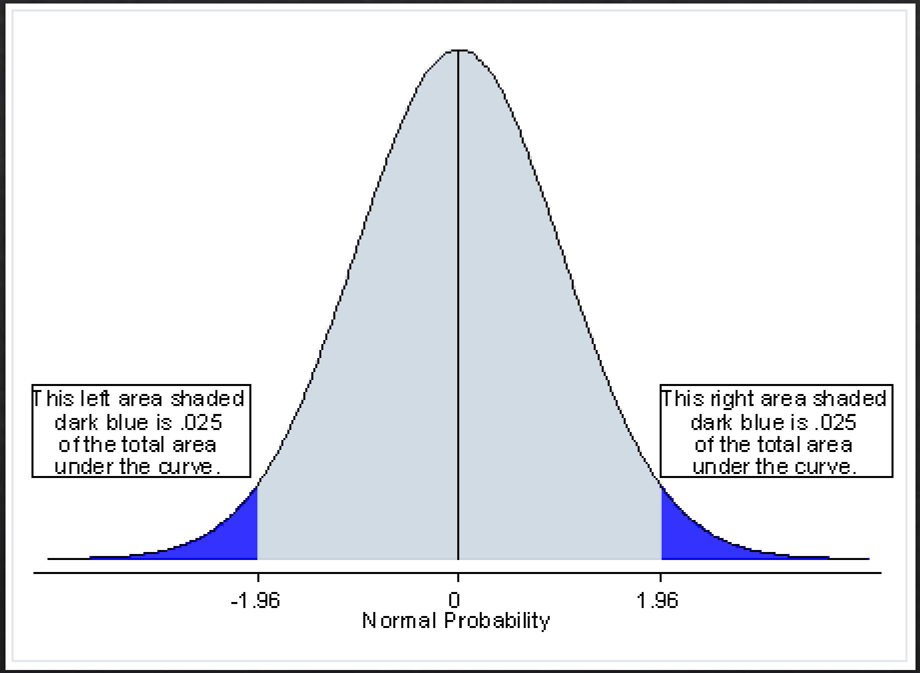
Econometric Methods: lecture 25/10

Prof. K. Drakos

More on Hypothesis Testing

- Consider the following regression model:
- $y = a + \beta * X + \gamma * Z + \delta * W + u$
- If you want to test the significance of individual parameters, eg. whether the parameter γ is significant, you may employ the t-test.
  - H_0: γ = 0


This left area shaded dark blue is .025 of the total area under the curve.
This right area shaded dark blue is .025 of the total area under the curve.
-1.96
0
1.96
Normal Probability

- If you want to test the overall significance of the model, i.e whether all slope parameters are significant, then you employ the F-overall test.
  - H_0: $\beta=\gamma=\delta=0$
- If you want to test whether a subset (not all of them) of the parameters is significant you need to employ a version of the F test, known as restricted-unrestricted.
- Let us spend some time on the background of this. Suppose that you want to test whether β and δ are jointly zero.
- H_0: $\beta=\delta=0$

- If you impose these restrictions on the original model, you obtain the following:
  - $y = a + \gamma * Z + u$
- The terminology we use is the following: the original model is called **Unrestricted**, signifying that there are no restrictions placed on it, while the model above, by the same token is called **Restricted** (in the sense that the zero restrictions have been imposed)
- In addition, the two models are also linked in the following manner: the restricted model is **nested** in the unrestricted model, since we may move from the unrestricted to the restricted by placing the zero restrictions
- If we think about the intuition of the test that we will employ, essentially, we compare a larger (the unrestricted) model with a smaller model and effectively we want to select which of the two should be chosen.

- The larger model performs better in terms of explanatory power, while in contrast performs worse in terms of the depletion of the model's degrees of freedom
- In other words, we are confronted with a typical cost-benefit problem, with the question being whether the cost is higher than the benefit or vice versa.
- To put the same statement in the context of the modelling procedure, the question that arises is whether the two variables (X,W) whose parameters are involved in the hypothesis test should stay or be removed from the model.
- In other words, if their contribution in the explanatory power (the benefit from their inclusion) is higher (lower) from the cost the bring to the model (the depletion of degrees of freedom), then these variables should be retained in (excluded from) the model.
- This is exactly the basis on which we design an appropriate F-test, which takes the following form:

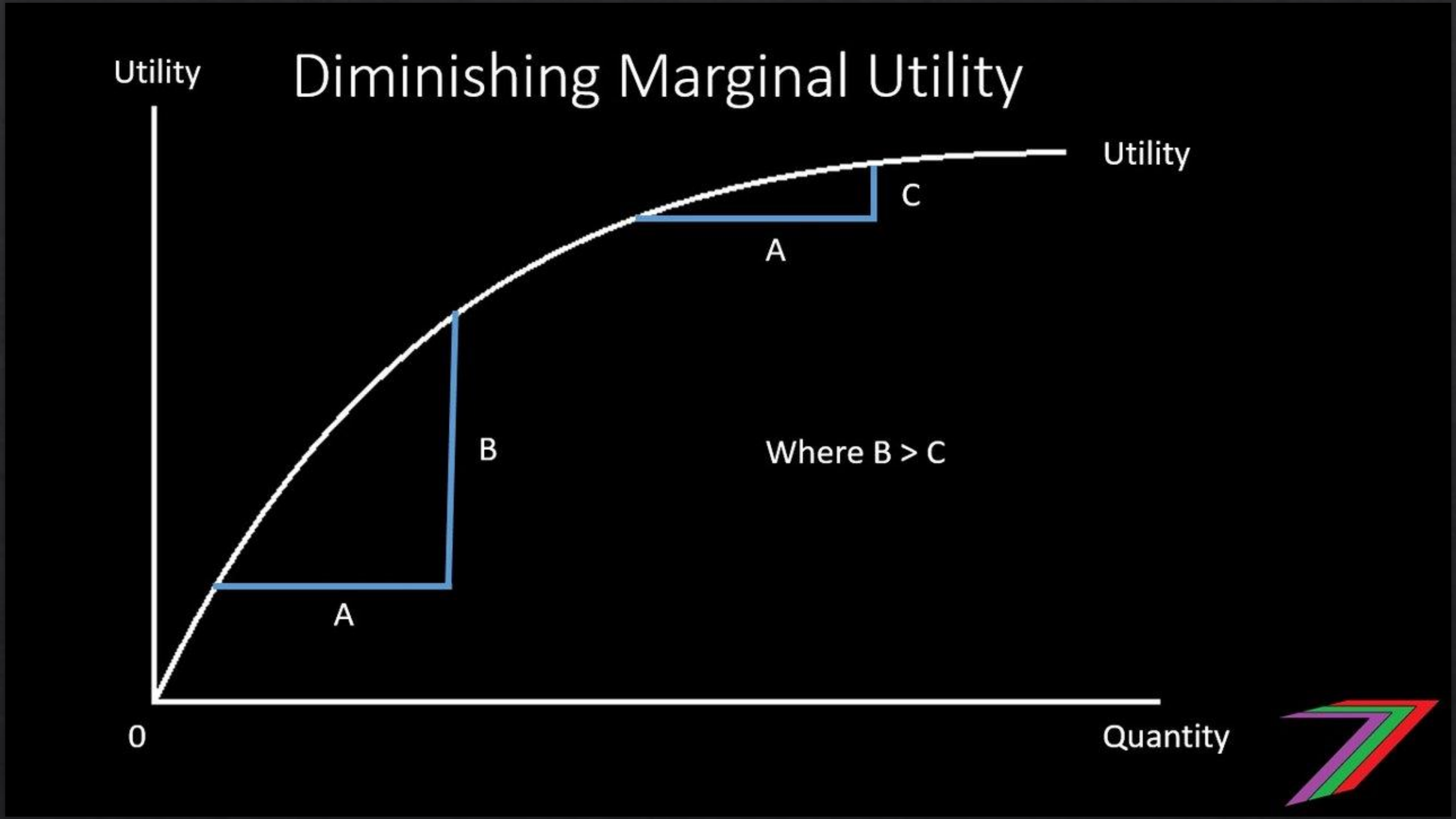
$$\frac{(R_u^2 - R_{ur}^2)/q}{(1 - R_u^2)/_{n-k-1}}$$

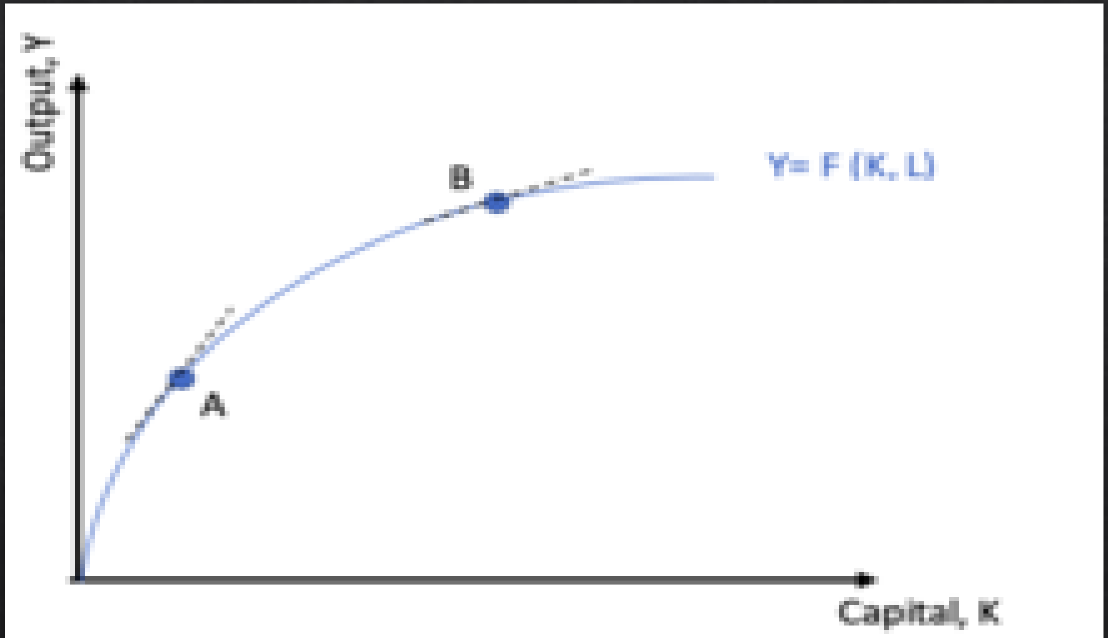
- If you focus on the numerator, you will immediately see that it takes into account the extra explanatory power of the unrestricted model in comparison to the restricted model (the extra benefit), while it penalises by dividing with q
- where q is the number of parameters set equal to zero, or in other words, the hypothesized reduction of the unrestricted model or similarly, the difference in the model degrees of freedom between the unrestricted and the restricted models
- how does the test work?
- If the test leads to a rejection, we reject that both parameters are statistically insignificant, and hence we cannot remove the two variables from the model on statistical grounds (this implies that for these two variables their contribution to the explanatory power is higher than the cost the bring)
- Modelling decision: between the unrestricted and the restricted model, the unrestricted model is qualified
- If the test does not indicate a rejection, then the two parameters are zero, and therefore the two variables must be removed from the model
- Modelling decision: between the unrestricted and the restricted model, the restricted model qualifies

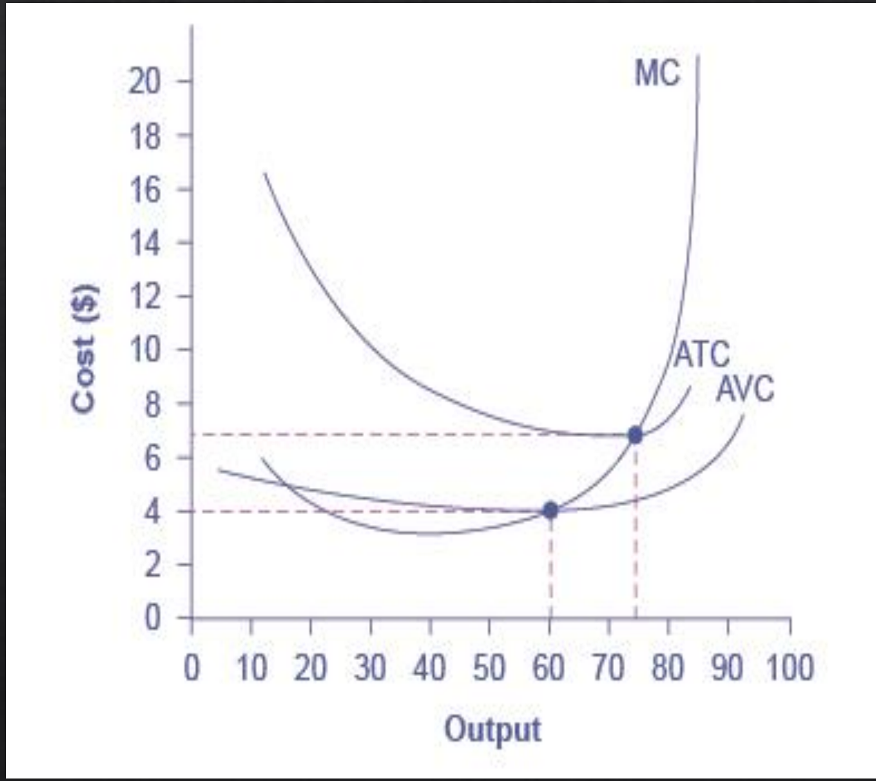
- While the application of the restricted vs. unrestricted test can be used in any situation, there are cases that it is especially useful.
- Typical application is the case where the potential explanatory variables belong to blocks, and the researches wishes to investigate whether a given block can be excluded.
- For instance, suppose that you want to model the return of a firm.
- One may think, 3 blocks of variables that may affect the firm's return:
- 1. firm-specific (financial ratios), 2. sectoral variables, 3. macroeconomic variables (GDP growth, inflation, unemployment)
- Then, using this procedure one may test whether a given block (say the macro variables) indeed affects the firm's return

Typical cases of non-linearity

- It is not rare for the relationship under consideration, to be confronted with a polynomial function
- In fact, there is a wide variety of economic/financial theories and/or phenomena, that suggest such polynomial functions


Utility
Diminishing Marginal Utility
Utility
C
A
B
Where B > C
A
0
Quantity


Output, Y
B
Y= F (K, L)
A
Capital, K



MC
ATC
AVC
Cost ($)
20
18
16
14
12
10
8
6
4
2
0
0 10 20 30 40 50 60 70 80 90 100
Output

A direct application of the restricted vs unrestricted F test on the choice of functional form

- Suppose that we suspect (from the scatter diagram) the presence of polynomial relationship
- Then we can design a procedure, based on the restricted-unrestricted test, in order to decide the best polynomial form
- Consider the following set of models:
- $y = a + \beta * X + \gamma * X^2 + \delta * X^3 + u$
- This is a cubic model

- If we impose the restriction that δ = 0, then we obtain the quadratic model
- $y = a + \beta * X + \gamma * X^2 + u$
- If we impose the restrictions that δ = γ = 0, then we obtain the linear model
- $y = a + \beta * X + u$
- If we want to compare the cubic vs. the quadratic model, we can employ the t-test, but we may also use the restricted (quadratic) vs. the unrestricted (cubic) model
- If we want to compare the cubic vs the linear model, then the only choice is the F restricted (linear) vs. unrestricted (cubic)

another case of non-linearity that one very often encounters, requires the use of a dummy variable

- suppose we consider a regression model
    - $y = a + \beta * X + u$

- And, say y is firms' profitability (say ROA), while X is firms' investment rate.
- Now, assume that our sample includes a mix of listed and non-listed firms.
- How can we investigate whether there is any difference between listed and non-listed firms?

- Being listed in the stock market, is a qualitative property (not a number), which divides the firms in our sample in two mutually exclusive (and exhaustive) sets:
- A firm is listed or not, it cannot satisfy both conditions at the same time
- Based on this, we define an indicator variable, which we call dummy, that flags whether given firm is listed or not
- $D = \begin{cases} 1 \; if \;\; the \; firm \; is \; listed \\ \quad 0 \; otherwise \end{cases}$
- Note that the two states (listed, non-listed) can be adequately represented by only one dummy
- This is the M-1 rule (any M-dimensional mutually exclusive and exhaustive situation, can be represented by M-1 dummies)

- Now consider the following model:
  - $y = a + \beta * X + \gamma * D + \delta * D * X + u$
- Note how the original model and the one above are related, restricted-unrestricted, nested
- If $\gamma = \delta = 0$, then we obtain the original model (i.e. being listed or not plays no role on profitability
- Let's play a little bit with the augmented (unrestricted) model
- Suppose that X = 0
  $y = \alpha + \gamma * D + u$
- but D takes 0 or 1 values, hence
- $y = \alpha + u \quad \text{if } D = 0$, and
- $y = \alpha + \gamma + u \quad \text{if } D = 1$

- what is the meaning of this? Well, it boils down to the parameter γ
- If γ is zero, then the two expressions are identical
- If γ is different from zero, there are two different expressions (one for listed and another for non-listed firms)
- Recall from a past lecture that the constant, when estimated, gives us the sample mean of the dependent variable
- In other words, if γ is zero, we conclude that the sample average y (profitability in our example) is statistically the same for listed and non-listed firms
- However, if γ is different from zero we have evidence for a differential profitability between listed and non-listed firms
- In fact,
- if $\gamma > 0$, then we conclude that profitability is higher on average for listed firms
- if $\gamma < 0$, then we conclude that profitability is lower on average for listed firms
- also note that diagrammatically, a significant γ parameter, would imply that the regression line has different intercept, depending on whether we are dealing with a listed or a non-listed firm

- now, let's turn our attention on the parameter δ
- if we take the derivative of y with respect to X, we obtain the following
- $\partial Y / \partial X = \beta + \delta * D$
- but D takes 0 or 1 values, hence
- $\partial Y / \partial X = \beta$ if $D = 0$, and
- $\partial Y / \partial X = \beta + \delta$ if $D = 1$
- what is the meaning of this?
- Well, it boils down to the parameter δ
- If δ is zero, then the two expressions are identical
- If δ is different from zero, there are two different expressions (one for listed and another for non-listed firms)
- Recall from a past lecture that we are now talking about the slope of the regression model

- In other words, if δ is zero, we conclude that the slope (how profitability reacts to investment rate) is statistically the same for listed and non-listed firms
- However, if δ is different from zero we have evidence for a differential response of profitability to investment between listed and non-listed firms
- In fact,
- if $\delta > 0$, then we conclude that profitability exhibits a higher sensitivity to investment for listed firms
- if $\delta < 0$, then we conclude that profitability exhibits a lower sensitivity for listed firms
- also note that diagrammatically, a significant δ parameter, would imply that the regression line has different slopes, depending on whether we are dealing with a listed or a non-listed firm

γ different from zero (δ zero)

γ zero, δ different from zero

γ, δ different from zero