

# Applied Econometrics Using Stata, WS 2015/16

July 27, 2016

Jun. Prof. Dr. Lars Hornuf and Cornelia Strüwing

The exam consists of two parts. You have 45 minutes to complete the first part of the exam. This part comprises a pencil-and-paper test and you may not use anything besides a pencil. We will notify you when the first part of the exam ends and collect your answer sheets. Thereafter, we will move to the computer-based exam, for which you will have another 75 minutes. During this part of the exam, you may use your textbooks, course materials and personal notes. Please remember that cheating is an academic offence and will be punished to the fullest extent. We both wish you good luck!

## 1. OLS Estimation (18 Points)

- a) Name three desirable properties of an estimator and briefly describe what they imply? (6 Points)
- b) Given that  $R^2$  is defined as the *explained sum of squares (SSE) / total sum of squares (SST)* or  $1 - (\text{residual sum of squares (SSR)} / \text{total sum of squares (SST)})$ , why is it useful to report the adjusted  $R^2$ ? Will the adjusted  $R^2$  usually be larger or smaller than the  $R^2$ ? Why is that the case? (4 Points)
- c) Let's consider a simply wage equation  $wage = \beta_0 + \beta_1 educ + \beta_2 abil + u$  where ability remains unobserved and is not included in the estimated regression. Analytically derive the direction of the possible bias. Would you expect to overestimate or underestimate the effect of education on wages? (5 Points).
- d) What is an event study? Briefly explain how it could be used to find out whether the recent death of the Unister CEO Thomas Wagner affects the valuation of the company. Which steps would have to be conducted in an econometric analysis in order to estimate the effect Thomas Wagner's sudden death? (3 Points).

## 2. Panel data (16 Points)

- (a) Write down in abstract form a difference-in-difference regression equation. Explain which particular effect the different coefficients capture. For which type of analyses are these models usually employed? Let's assume we want to learn about the effectiveness of a new education program. Under which conditions could the difference-in-difference yield inconsistent estimates? (7 Points)
- (b) What is the difference between a balanced and an unbalanced panel? Provide two reasons why a panel dataset could be unbalanced? Why could that be problematic for your analysis? (4 Points)
- (c) State five assumptions you need to make for the fixed effects estimator to be unbiased. Briefly explain what they mean. (5 Points)

**3. Proxy and Instrumental Variables (7 Points)**

- (a) Explain the difference between a proxy variable and the instrumental variable approach? Which assumptions need to be fulfilled if you want to use these approaches to solve an endogeneity problem? Hint: You might want to use path diagrams to explain the differences.

**4. Duration Models (4 Points)**

- (a) Duration models are estimated when the dependent variable measures duration of time. Please explain what econometricians understand under right and left censoring. What is the difference between fixed and random censoring if cases are right censored?