Applied Econometrics Using Stata, WS 2014/15

July 15, 2015

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. Please remember that cheating at the exam is an academic offence and will be punished to the fullest extent. We both wish you good luck!

1.   OLS Estimation (14 Points)

a)   Under which conditions is OLS the best linear unbiased estimator? Briefly explain what we mean by these assumptions.

b)   In the simple linear regression model $y = \beta_0 + \beta_1 x + u$, suppose that $E(u) \neq 0$. Letting $\alpha_0 = E(u)$, show that the model can always be rewritten with the same slope, but a new intercept and error, where the new error has a zero expected value.

2.   Instrumental Variables (17 Points)

(a)   What do we understand by endogeneity? List three different sources of endogeneity and briefly explain what they imply. (7 Points)

(b)   Proxy and instrumental variables can be fixes for endogeneity problems. Consider a wage equation $wage = \beta_0 + \beta_1 educ + u$ where wage is solely determined by education. We believe that ability has a major effect on wage and should have been included in such an equation. Our dataset furthermore includes the following variables: motheduc (education of the mother of the wage earner), books (number of books the wage earner owns), IQ (how well the wage earner scores on an IQ test). Precisely explain whether we should use books as a proxy or instrumental variable in our

(c)   In the case of an omitted variable, can a single variable be considered a good proxy and an instrument at the same time? (3 Points)

3.   Duration Models (8 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? (4 Points)

(b)   Imagine a proportional hazard model where the dependent variable is time until an individual runs into personal bankruptcy in Germany. The dataset contains information from January 1, 2007 till December 31, 2010. Provide one example for fixed and random censoring in such a dataset. Why might random censoring cause a problem when estimating a model based on this data? (4 Points)