

Econometrics

Take Home Exam

1 Rules of the Take Home Exam

- The deadline to hand back the take home exam is January the 6th 2020, 23:59. The typewritten report should be handed in PDF by e-mail to Matteo Picchio (m.picchio@univpm.it) and Giulio Palomba (g.palomba@univpm.it), as well as the codes to generate your results.
- You must work alone. Any kind of cooperation between students is **NOT** permitted! We will consider it as a form of cheating during exams. If we realize that N students have cheated in answering one question, only one student randomly drawn from those N students will get points for the corresponding answer. The other cheating students will get 0 for that answer.
- To prevent free riding, you will be asked to explain in detail what you have done during an oral exam. The oral exam will take place on January the 10th 2020 at 9:30 with Matteo Picchio for the microeconometrics part, and on January 16th at 14:00 with Giulio Palomba for the time series part.
- You are suggested to carry out the empirical part in Gretl but you are not required to use Gretl. Whatever econometric software you use, we should be able to replicate what you have done. Hence, you should submit also the file containing the codes which generated your results. The codes generating your results should briefly and clearly describe what the program statements do (if not self-explaining).
- All the steps should be motivated and the comments to the results should be as clear as possible. When you are asked to “Discuss” or “Comment”, you should motivate as better as you can your results, using formal arguments and referring to the economic theory if you find it needed and appropriate. The better argued your discussion, the higher the grade assigned to the corresponding part of the exam.
- Students that do not hand the take home exam before the deadline will get 0.

2 Microeconometrics – Nonlinear models

Use the dataset *employment.dta*. It contains information on 5,638 women living in the US in 1991 and who were 18-59 years old at the moment of the interview. Suppose that **you are interested in estimating the causal effect of years of education on the probability of women being at work** (or, equivalently, of female employment).

The dependent variable is *emplo*, a dummy variable equal to 1 if the woman is at work and 0 otherwise. Then we have individual characteristics like *age*, dummies for race (*black*, *hispanic*, years of schooling (*educ*), dummies for the presence of kids in the household younger than 6 (*kidlt6*) and older than 6 (*kidge6*), the non-wife income in \$1,000, and some husband's characteristics (identified by the prefix *hus-*).

2.1 Data description

- Provide descriptive statistics of the data (mean, standard deviations, minimum, and maximum) and comment on them.
- Report the mean of the dependent variable by different levels of education: strictly less than 12, between 12 and 15, and 16 or more.
- Illustrate in a graph with *emplo* on the *y*-axis the linear relation between the employment indicator and years of education.
- Given the descriptive evidence in b. and c., is there evidence of any kind of relation between employment probability and years of education in the row data? If yes, can we interpret it as the causal effect of years of education on the probability of employment? Comment and discuss.

2.2 Linear probability model

Consider the following linear probability model (LPM) for employment, for $i = 1, \dots, N$:

$$\begin{aligned} \text{emplo}_i &= \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{age}_i + \beta_3 \text{age}_i^2 + \beta_4 \text{black}_i + \beta_5 \text{hispanic}_i \\ &+ \beta_6 \text{kidge6}_i + \beta_7 \text{kidlt6}_i + \beta_8 \text{nwifeinc}_i + u_i, \end{aligned} \quad (1)$$

where u_i is the error term.

- Estimate the model by OLS, report the estimation results in a table and comment on them.
- Estimate the average partial effect (APE) of years of education going from 12 (higher secondary school diploma) to 16 (tertiary diploma) and its standard errors. Comment on the findings.

- c. What is on average the impact of one more year of age on the probability of employment? Is it significantly different from zero? Motivate your answer.
- d. Can we interpret the estimated coefficient of education as the causal effect of years of education on the probability of employment? Comment and discuss using economic arguments.

2.3 Probit model

Imagine that you would like to move to a nonlinear model for the probability of employment.

- a. What would you gain by doing so?
- b. Estimate the nonlinear counterpart of the model in Equation (1) using the probit specification and estimate it by maximum likelihood. Report the estimation results in a table and comment on them.
- c. Estimate the APE of years of education going from 12 to 16 and its standard error. Compare it with what you got from the LPM.
- d. What is on average the impact of one more year of age on the probability of employment? Is it significantly different from zero? Compare what you get with the corresponding results from the LPM.
- e. Is the use of a nonlinear model like the probit model of help in disentangling the causal effect of education from spurious components? Motivate your answer.

2.4 Tackling endogeneity: LPM

- a. Are there reasons to suspect the years of education is an endogenous variable? What could they be? Motivate your answer.
- b. Use the control function approach to test for the endogeneity of years of education under heteroskedasticity. Use husband's years of education (*huseduc*) as excluded instrument. Explain also what requirements *huseduc* should meet to be a *valid* instrument.
- c. Estimate the linear probability model in Equation (1) using the 2SLS estimator. Report the estimated coefficients from the 2SLS estimator and the control function approach of the previous point in the same table. Do you get the same estimation results? Do you get the same standard errors? If not, why?
- d. Has the excluded instrument a strong power in explaining the endogenous regressor? Motivate your answer on the basis of a formal test.

- e.* Given the results of the endogeneity test, what estimator would you choose between OLS and 2SLS? Motivate your answer.

2.5 Tackling endogeneity: probit model

- a.* Use the two-step Rivers and Vuong (1988) approach to test for the endogeneity of years of education. Use husband's years of education (*huseduc*) as excluded instrument. Discuss what you obtain.
- b.* Provide again, but now for the model that control for endogeneity, the APE of years of education going from 12 to 16 and its standard error. Contrast it to the one that can be found by applying 2SLS to the corresponding linear probability model. Discuss.

3 Time series econometrics

The spreadsheet *USAYC.xls* contains quarterly data of US Gross Domestic Product and Federal Consumptions in USA. The available sample ranges from 1947:1 to 2017:1 ($T = 282$ observations).

Provide a full and detailed investigation about the dynamic relationships between the two time series, focussing on the model specification, the diagnostics and the economic interpretation of the results.