

Test Exercise 4

Notes:

- See website for how to submit your answers and how feedback is organized
- This exercise uses the datafile TestExer4_Wage and requires a computer.
- The dataset TestExer4_Wage is available on the website.

Goals and skills being used:

- Obtain insight in consequences of endogeneity
- Practice with identifying causes of endogeneity
- Practice with identifying valid instruments
- Obtain hands-on experience with applying 2SLS and the Sargan test

Questions

A challenging and very relevant economic problem is the measurement of the returns to schooling. In this question we will use the following variables on 3010 US men:

- `logw`: log wage
- `educ`: number of years of schooling
- `age`: age of the individual in years
- `exper`: working experience in years
- `smsa`: dummy indicating whether the individual lived in a metropolitan area
- `south`: dummy indicating whether the individual lived in the south
- `nearc`: dummy indicating whether the individual lived near a 4-year college
- `dadeduc`: education of the individual's father (in years)
- `momeduc`: education of the individual's mother (in years)

This data is a selection of the data used by D. Card (1995)¹

(a) Use OLS to estimate the parameters of the model

$$\log w = \beta_1 + \beta_2 \text{educ} + \beta_3 \text{exper} + \beta_4 \text{exper}^2 + \beta_5 \text{smsa} + \beta_6 \text{south} + \varepsilon.$$

Give an interpretation to the estimated β_2 coefficient.

(b) OLS may be inconsistent in this case as `educ` and `exper` may be endogenous. Give a reason why this may be the case. Also indicate whether the estimate in part (a) is still useful.

¹"Using Geographic Variation in College Proximity to Estimate the Return to Schooling". In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press, 1995.

- (c) Give a motivation why age and age² can be used as instruments for exper and exper².
- (d) Run the first-stage regression for educ for the two-stage least squares estimation of the parameters in the model above when age, age², nearc, dadeduc, and momeduc are used as additional instruments. What do you conclude about the suitability of these instruments for schooling?
- (e) Estimate the parameters of the model for log wage using two-stage least squares where you correct for the endogeneity of education **and** experience. Compare your result to the estimate in part (a).
- (f) Perform the Sargan test for validity of the instruments. What is your conclusion?