KAN-CEAPV2505U
Econometric Analysis of Firm Data

July 30, 2025

Copenhagen Business School
Department of Economics
Ralf A. Wilke
Mette Franck
2025/2026, Autumn semester

Problem Set 3

Proxy and Instrumental Variables

1. Use the data in WAGE2.dta for this exercise.

   (a) Use the variable KWW (the "knowledge of the world of work" test score) as a proxy for ability in place of IQ in the lecture's example. What is the estimated return to education in this case?

   (b) Now, use IQ and KWW together as proxy variables. What happens to the estimated return to education?

(c) In part (b) are IQ and KWW individually significant? Are they jointly significant?

2. Use the data in WAGE2.dta for this exercise.

(a) We use the variable *sibs* (number of siblings) as an instrument for *educ* in the following simple wage regression:

$$log(wage) = \beta_0 + \beta_1 educ + u.$$

The IV estimate of the return to education is 0.122. To convince yourself that using *sibs* as an IV for *educ* is not the same as just plugging *sibs* in for *educ* and running an OLS regression, run the regression of $log(wage)$ on *sibs* and explain your findings.

(b) The variable *brthord* is birth order (*brthord* is one for a first-born child, two for a second born child, and so on). Explain why *educ* and *brthord* might be negatively correlated. Regress *educ* on *brthord* to determine whether there is a statistically significant negative correlation.

(c) Use *brthord* as an IV for *educ* in the model of part (a). Report and interpret the results.

(d) Now, suppose that we include number of siblings as an explanatory variable in the wage equation; this controls for family background, to some extent:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 sibs + u.$$

Suppose that we want to use *brthord* as an IV for *educ*, assuming that *sibs* is exogenous. The reduced form for *educ* is

$$educ = \pi_0 + \pi_1 sibs + \pi_2 brthord + \nu.$$

State and test the identification assumption.

(e) Estimate the model from part (d) using *brthord* as an IV for *educ* (and *sibs* as its own IV). Comment on the standard errors for $\hat{\beta}_{educ}$ and $\hat{\beta}_{sibs}$.

(f) Using the fitted values from part (d), $\widehat{educ}$, compute the correlation between $\widehat{educ}$ and *sibs*. Use this result to explain your findings from part (e).

3. Use the data in 401KSUBS.dta for this exercise. The equation of interest is a model with a binary dependent variable:

$$pira = \beta_0 + \beta_1 p401k + \beta_2 inc + \beta_3 inc^2 + \beta_4 age + \beta_5 age^2 + u$$

This model is called linear probability model which is a conditional mean model (with heteroscedastic error structure). The goal is to test whether there is a tradeoff between participating in a 401(k) plan (this is a defined-contribution pension plan in the US) and having having an individual retirement account (IRA, this is another type of retirement arrangement in the US). Therefore, we want to estimate β_1.

(a) Estimate the model by OLS (obtain heteroskedasticity robust standard errors). β_1 is the increase in the probability of having an IRA if the individual participates in a 401(k) plan (everything else equal).

(b) For the purpose of estimating the ceteris paribus tradeoff between participation in two different types of retirement savings plan, what might be a problem with OLS?

(c) The variable $e401k$ is a binary variable equal to one if a worker is eligible to participate in a 401(k) plan. Explain what is required for $e401k$ to be a valid IV for $p401k$. Do these assumptions seem reasonable?

(d) Estimate the reduced form for $p401k$ and verify that $e401k$ has significant partial correlation with $p401k$ (use heteroskedasticity robust standard errors).

(e) Now, estimate the structural equation by IV and compare the estimate of β_1 with the OLS estimate. Again, you should obtain heteroskedasticity-robust standard errors.

(f) Test the null-hypothesis that $p401k$ is in fact exogenous, using a heteroskedasticity robust test.

These problems are taken from the Wooldridge (2020) textbook.