

Homework 3

Fernando Rios-Avila

Part I: Instrumental Variables (40pts)

Consider the dataset `labsup` (available using `frause`). This dataset has information on the number of hours worked per week, hourly wage, and years of education for a sample black and hispanic women. The goal is to determine the effect of an additional child has on the number of hours worked per week.

To do this, estimate a model of the form:

$$hours = b_0 + b_1kids + b_2educ + b_3age + b_4age^2 + b_5hispanic + b_6black + b_7nonmomi + u \quad (1)$$

(see variable description in the datafile)

Q1-1. (10pts) Estimate the model using OLS. What is the effect of an additional child on the number of hours worked per week? Is this effect statistically significant? What about the rest of the variables?

Q1-2. (10pts) It is believed that Number of childrens may be correlated with the model error (endogenous). Explain why this may be the case. (You can consult the [original](#) paper)

Q1-3. (20pts) There are two possible candidates for IV in this case. `samesex` (if the first two children had the same sex), and `multi2nd` - if the family had twins the second during the second pregnancy. Estimate Equation 1, using each of these variables as IV separately, and both of them together. While doing this answer:

- i. Are the instruments strong individually? Are they strong together? (F-statistic of first stage)
- ii. How does the effect of kids on Hours worked change when using the different IV's. Are there any differences?
- iii. Based on all specifications, is there any evidence that the number of children is endogenous?

extra (5pts): Should be worry about overidentification here? why? Test for it.

Part II: MLE and Nonlinear models (30pts)

Consider the dataset `smoke` (available using `frause`). This dataset has information on few demographic characteristics, cigarette prices, income, as well as number of cigarettes smoked per day.

In this case, “Cigarettes smoked per day” is a kind of limited dependent variable, because not everyone will smoke, and the number of cigarettes people smoke is an integer. Under this considerations estimate the following:

Q2-1. (10pts) Estimate a Linear probability model (LPM) and probit model, analyzing the probability of smoking as function of demographics, log of income and log of prices. Analyze the results and compare the two models (magnitudes of the effects). What would you say the impact of restaurant smoking bans is on the probability of smoking?

Q2-2. (10pts) Now, say that you are also interested in analyzing the relations of the different factors on the number of cigarettes smoked per day. Estimate a Poisson model and a Tobit model (with data censored at 0). Analyze the results and compare the two models (magnitudes of the effects). What would you say the impact of restaurant smoking bans is on the number of cigarettes smoked per day? what about the impact of a 10% increase in cigarette prices?

Note: For the Tobit model, you can use the `tobit`, `ll(0)`, and will need to use `margins`, `margins, dydx(*) ystar(0,.)` to get the marginal effects that are comparable across models.

Q2-3. (10pts) You are also asked to use the Tobit results to analyze the probability of smoking. To do this, you will need to use the marginal effects from the Tobit model `margins, dydx(*) pr(0,.)`. Compare the results of this exercise with the results from the LPM and probit models. Would you reach the same conclusions? if not, what are the implications for the tobit assumptions of a tobit model?

Part III: Panel data (30pts)

Consider the dataset `driving` (available using `frause`). This dataset has information at the state level, from 1980 to 2004, on statistics regarding traffic accidents, fatalities, and laws.

See datafile for variable description.

Q3-1. (10pts) Estimate a model of the form:

$$totfatrte = b_0 + b_1\mathbb{1}(seatbealt = 1) + b_2\mathbb{1}(seatbealt = 2) + b_3minage + b_3zerotol + b_4unem + b_5perc14_24 + b_6v$$

(2)

And interpret the results. is there any effect that seems unexpected? Explain why this may be the case.

Q3-2. (5pts) One of the reasons one may be finding unexpected results is because of unobserved heterogeneity across states. Can you explain what kind of factors are there that could be related to both accidents and the laws? (think about the political process of passing laws)

Q3-3. (10pts) To control for unobserved heterogeneity, estimate three models:

- Random effect model
- Fixed effect model
- Correlated Random effect model

How do the results from the Random effect model and Fixed effect model compare to each other? what about compared to Equation 2?

Q3-4. (5pts) Using the results from the Correlated Random effect model, test for which of the models is more appropriate, between Random Effects and Fixed effects. What do you conclude?