TI: Econometrics I 2021/2022
Assignment II

Instructions

- You are supposed to make the assignments individually or in a group of two. Of course, you may discuss the questions and your general ideas with your fellow students, but the actual answers should be given individually or in a group of two. Fraud, in the sense of copying answers, will be reported to the examination board and may have serious consequences.
- Before you start answering the questions, first read the complete exercise.
- In answering the questions, always state explicitly what you did, why you did it, and what your conclusions are. Be clear and concise in your statements.
- Submitting your answers in LaTeX is highly appreciated, though Word (converted to pdf) is also accepted.
- For all questions where R (or other software) is used also provide the code that you used to generate your results.
- Only include relevant R output in your file. Either include your code in your answers, or send it in a separate file.
- For all tests, use a significance level of 5%, unless indicated otherwise.
- Due date: 2021 November 30 at 23:59 at Canvas.

Question 1

Consider the following model

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma_i^2) \\ \sigma_i^2 &= \sigma^2 e^{\gamma z_i} \end{aligned}$$

for $i = 1, ..., 5000$. Suppose that the true parameters are $\beta_0 = 3, \beta_1 = 5, \beta_2 = 8$. Suppose that $x_{i,1}$ follows a normal distribution with mean 1 and variance 1, and that $x_{i,2}$ follows a normal distribution with mean 2 and variance 1. Let z_i follow a Gamma distribution with shape parameter 1.2 and scale parameter 1.1, and we also assume that $\sigma^2 = 1$. Set the random seed to 2021.

(a) [1 point] Assume that $\gamma = 1$. Generate 5000 observations and then use OLS to estimate the parameters in the model and calculate the related OLS standard errors, t-values and p-values. Then use OLS to estimate the model with White standard errors[1] and calculate the related White standard errors, t-values and p-values. Compare the results.

(b) [1 point] What are the procedures to perform the Breusch-Pagan test for heteroskedasticity? Perform a Breusch-Pagan test for heteroskedasticity. Provide the value of the test statistic and explain if the null hypothesis is rejected.

(c) [2 points] Assume that $\gamma = 0$. Estimate β_0, β_1 and β_2 separately using

1. OLS
2. WLS with known $\gamma = 0$
3. FWLS with estimated γ (i.e. γ is unknown)

Explain the weights you use for WLS and FWLS. Provide the coefficients and standard errors of the three estimators for three methods and compare the results. Are the estimators close to the their true values?

(d) [2 points] Now assume that $\gamma = 1$. Repeat sub-question(c). Provide the coefficients and standard errors of the three estimators for three methods and compare the results. Are the estimators close to the true ones?

(e) [1 point] Now assume that $\gamma = -1$. Repeat sub-question(c). Provide the coefficients and standard errors of the three estimators for three methods and compare the results. Are the estimators close to the true ones?

Question 2

The model below consists of two equations describing the relation among macroeconomic consumption C, disposable income D, and non-consumptive expenditures Z.

$$\begin{aligned} \text{consumption equation}: \quad & C_i = \alpha + \beta D_i + \epsilon_i \\ \text{income equation}: \quad & D_i = C_i + Z_i \end{aligned}$$

We assume that $E[Z_i \epsilon_i] = 0$ for $i = 1, ..., n$. We also assume that $\text{Var}(\epsilon) = \sigma^2$.

[1] If you use R, compute the White standard errors by setting the standard error type to "HC0" in the function *lm_robust* in the *estimatr* package; otherwise please state clearly how you compute the White standard errors.

(a) [1 point] Show that the OLS estimator of the parameter β is not consistent.

(b) [2 points] Derive plim(b) where b is the OLS estimator of β. Determine the sign of the magnitude of the inconsistency when $0 < \beta < 1$, that is, the sign of plim(b) $-\beta$ when $0 < \beta < 1$.

(c) [1 point] Find an instrumental variable (IV) for the endogenous variable D_i and argue why it could be an IV.

(d) [2 points] Derive b_{IV}, the IV estimator of β in terms of the variables C, D, and Z step by step.

(e) [2 points] Use the expression of b_{IV} to show that it is consistent.

Question 3

The China Shock You are going to replicate some of the results from a seminal study of the effects of increased competition [1]. The replication files for the paper are available at `https://www.openicpsr.org/openicpsr/project/112670/version/V1/view`. Download the datasets "workfile_china_long.dta", "workfile_china.dta" and "workfile_china_preperiod.dta".

For the first four questions, it is easier to use the "workfile_china_long.dta" dataset. For the final question, use "workfile_china_preperiod.dta" to reproduce Table 2, and "workfile_china.dta" to reproduce Table 3.

You can read these into R easily using the 'read_dta' command in the 'haven' package.

The paper measures the effect of import exposure on jobs across US commuting zones. They do this using 'shift-share' measures, which are very common in applied work in immigration and trade - see e.g `https://ftp.iza.org/dp11307.pdf`.

(a) [1 point] Plot the distribution of the growth rate of employment and of import exposure 1990-2007 across US commuting zones.

(b) [1 point] Regress import exposure on the growth rate of employment from 1990-2007. Plot your results. You should be able to reproduce panel B of Figure 2. Compute normal OLS standard errors and HAC standard errors clustered by the state levels (hint use the vcovHAC command from the sandwich package)and compare them.[2]

(c) [1 point] Is this a good causal estimate of the effect of import exposure on employment? Give a reason why or why not.

(d) [2 points] The authors construct an instrument for import exposure using the growth rate of Chinese imports in eight other similar countries.

Construct the instrumental variable estimate of the effect of the growth of import exposure on the growth of employment using the instrument from the data in "workfile_china.dta". Do so in two ways. First, use a package. Then use matrix multiplication. Present regression results for both. Do not include any additional controls for now.

[2]Note that the clustered standard errors should be slightly different from the ones reported in the paper because Stata automatically implements a slightly different form of the HAC standard errors to R - if you want the same specify "type=HC1".

To show that you have done the matrix multiplication is correct, report the third entry of the projection matrix of the instrument times the endogenous variable i.e of $P_z X$.

(e) [2 points] You might notice that your results are different from the results in the paper. The authors use weighted estimates, where the weights are shares of manufacturing employment.

Now reproduce the results in Table 2 and Table 3 in the paper exactly as they do (so two tables containing all the coefficient estimates). For the first, you will need to use $workfile\_china\_preperiod.dta$. For the second, use "$workfile\_china.dta$" instead.

Report the first-stage F statistics. Is the instrument a good instrument?

Question 4

Imagine we have a binary dependent variable $y_i \in \{0, 1\}$ and some fixed explanatory variables x_i.

(a) [1 point] Imagine we fit a linear probability model of $y_i = \alpha + \beta x_i + \epsilon_i$. Derive the distribution of the error terms. Will our least-squares parameter estimate $\hat{\beta}$ be unbiased? Will it still be the most efficient estimator?

(b) [1 point] Now imagine that we want to estimate this regression model for a given distribution of the errors $F()$ (e.g the logistic distribution) using maximum likelihood. Write out the distribution of y_i.

(c) [2 points] Use the distribution to write out the log-likelihood function. Then, write out the first-order condition for maximisation with respect to β.

(d) [2 points] Imagine we assume a logistic distribution of the errors. Show that our expression above simplifies to

$$\sum_i (y_i - \frac{1}{1 + e^{\beta x_i}}) x_i = 0.$$

(e) [2 points] Finally, use the value of $F(\beta x_i)$ to write the log of the odds ratio as a function of the parameters of the model. Thus, give an interpretation of the value of β.

References

[1] David Dorn David Autor and Gordon Hansen. "The China Syndrome: the Local Labour Market Effects of Import Competition in the United States". In: *American Economic Review* 103.6 (2013), pp. 2121–2168.