Advanced Econometrics:
Home Assignment 2

2025

Instructions

- The problem set is due November 26, 2025, 23:59. A late submission automatically means 0 points.
- The solutions should be sent to martina.luskova@fsv.cuni.cz, with the following subject: `'AE_HW_groupxy_surname1_surname2_surname3'`. Please send only one solution per group and include all group members in copy (cc).
- Send me the solution as one Jupyter notebook (.ipynb), named `'AE_HW1_groupxy_surname1_surname2_surname3.ipynb'` which should contain the main analysis together with commented code. Do not forget to use the full potential of Jupyter notebooks. Show graphs and results as an output of your code. Write the reasoning and other text in markdown cells (which supports headers, Latex equations, pictures, etc.). It is also possible to send me one pdf file with the analysis and one R script (following the naming convention).
- The empirical problems do not necessarily have a unique solution in terms of numbers. You are assessed based on the execution of the analysis not on the right numbers that you should get from the output. The emphasis is put mainly on meaningful presentation and the extent of your knowledge.
- Please, use `'set.seed()'` function, so I can replicate your results.
- If you have any questions concerning the homework, do contact me by mail. Do it rather sooner than later.
- If you use any AI tool, you need to make a declaration at the beginning of the file: During the preparation of this output, the author used [`name of tool/service`] in order to [`state reason`].
- Be aware that suspiciously identical-looking solutions will receive automatically 0 points.

Problem 1: GMM (2 points)

In this task we will work with MA(2) model:

$$y_t = \mu + e_t + \theta_1 * e_{t-1} + \theta_2 * e_{t-2}$$

$$e_t \sim iid(0, \sigma^2)$$

(a) For 1002 observations simulate the process e_t: $e_t \sim N(0,1)$. Also get e_{t-1} and e_{t-2}. Check that all 3 variables (e_t, e_{t-1}, e_{t-2}) have the same number of observations (1000). Model MA(2) process with the following coefficients:

$$\mu = 0$$

$$\theta_1 = 0.75$$

$$\theta_2 = 0.5$$

(b) Prove that the modelled process is indeed MA(2) (hint: you can do this by plotting the ACF).

(c) Derive the moment conditions function for MA(2) process. Use more moment conditions than is the number of coefficients that you want to estimate $(\mu, \theta_1, \theta_2, \sigma^2)$.

(d) Estimate the model using GMM and both identity and optimal weighting matrix. Provide the output and interpret the coefficient significance and the J-test statistics.

Problem 2: Bootstrap (2 points)

Assume a random sample from exponential distribution. $X \sim exp(\lambda)$.

(a) Generate 200 observations from the exponential distribution for $\lambda = 2$.

(b) Check the sample mean and variance, and compare them to the theoretical values.

(c) Plot the histogram of generated data, kernel density approximation, theoretical density of exponential distribution and exponential Q-Q plot. Discuss.

(d) Use brute force bootstrapping to obtain the Bootstrap mean, standard errors and bias-reduced estimate of the mean. Use 100,000 bootstrap replications.

(e) Now, use the `'boot'` function with 100,000 bootstrap replications.

(f) Obtain the confidence intervals, try different types and compare the results. Which CI type is appropriate to use in this case?

Problem 3: Endogeneity (2 points)

Let us follow the idea of the first exercise from Seminar 6 but for now we create another artificial dataset containing 300 observations (note that although variance of RVs is specified below, R commands often require to specify standard deviation instead):

$$z_1 \sim N(2, 3^2); z_2 \sim N(2, 1.5^2); z_3 \sim N(0, 2^2); z_4 \sim N(1.8, 2.5^2)$$

$$\epsilon_{1,2,3} \sim N(0, 1.5^2)$$

$$x_1 = 0.3z_1 - z_2 + 0.9z_4 + 0.75\epsilon_1$$

$$x_2 = 0.75z_2 + +0.75\epsilon_2$$

$$x_3 \sim N(0, 1)$$

$$y = 1 + 2.5x_1 - x_2 + 0.45x_3 + \epsilon_3$$

We should estimate the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + e$$

(a) Discuss the nature of the endogeneity problem in the system above. You might check important correlations and you should explain the difference between x_1 and x_3. Do you expect to observe any bias within the OLS estimation? Explain why.

(b) Estimate the model by OLS and interpret.

(c) The data set includes some potential IV candidates: z_1; z_2; z_3; z_4. What assumptions need to be satisfied to have a 'good' instrument? Which of these candidates seem to be 'good' instruments and why? Test their relevancy statistically. Is there any invalid, irrelevant, or weak instrument?

(d) Based on section c), choose the best instrument and run the IV regression. Run also 2SLS regression using all 'good' instruments. Compare coefficient estimates and standard errors.

(e) Finally, test for the endogeneity using the Hausman test. Report and interpret the results.

(f) Using the extended dataset (simulate more data from the data generating process), show that OLS is not a consistent estimator of β_1 and β_3. Show that 2SLS provides consistent results.

Problem 4: Mix (2 points)

Download the dataset `hw2data`. It contains random variable which is χ^2 with d degrees of freedom.

(a) Make a preliminary guess about the degree of freedom using only the graphical method. (Hint: plot the histogram and kernel density approximation. Compare it with the theoretical χ^2 density function with different degrees of freedom).

(b) The mean of χ^2 random variable Y with d degrees of freedom is $E(Y) = d$ and the variance is $Var(Y) = 2d$. Based on these two moment conditions specify 2 sample moment conditions that can be utilized in the GMM

(c) Use the 2 sample moment conditions to estimate d using GMM using both identity and optimal weighting matrix.

(d) Let's model variable X next way:

$$X = 0.3 * V_1 + 0.2 * V_2$$

Where $V_1 \sim N(0,1)$ and $V_2 \sim \chi^2(2)$. Now, let's consider linear model:

$$Y = \beta_1 + \beta_2 X + \epsilon$$

Let's denote Φ in the following way:

$$\Phi = (\beta_1)^2 - exp(\beta_2^3)$$

Estimate the value of Φ and also the variance of Φ, $Var(\Phi)$. To estimate $Var(\Phi)$ use the next methods:

(a) Delta method (by hand)

(b) Delta method (using `'deltamethod'` function)

(c) Bootstrap (with 1000, 10000 replications)

(d) Compare the results.