ECONOMETRICS I: PROBLEM SET 4

ASYMPTOTIC PROPERTIES AND HYPOTHESIS TESTING

1. Consider Model 1 where $K = 1$:

$$y = \beta x + u$$

where we assume $\lim_{T\to\infty} T^{-1}x'x = c \neq 0$ and (u_i) are i.i.d. Obtain the probability limit of $\hat{\beta}_R = \frac{y'y}{x'x}$. (Note that this estimator is obtained by minimizing the sum of squares in the direction of the x-axis).

- Is $\hat{\beta}_R$ consistent?
- What is required for $\hat{\beta}_R$ to be consistent?

2. Suppose that, from a sample of 63 observations, the least squares estimates and the corresponding estimated covariance matrix are given by

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} \qquad \widehat{\text{cov}}(b_1, b_2, b_3) = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

Using a 5% significance level, and an alternative hypothesis that the equality does not hold, test each of the following null hypotheses:

   (a) $\beta_2 = 0$

   (b) $\beta_1 + 2\beta_2 = 5$

   (c) $\beta_1 - \beta_2 + \beta_3 = 4$

3. The following regression equation is estimated as a production function for Q:

$$\begin{aligned} \ln Q = 1.37 + &\underset{(0.257)}{0.632} \ln K + \underset{(0.219)}{0.452} \ln L \\ R^2 = 0.98 \qquad &cov(b_k, b_l) = 0.055 \end{aligned}$$

where the standard errors are given in parentheses. Test the following hypotheses:

   (a) The capital and labor elasticities of output are identical.

(b) There are constant returns to scale.

Note: The problem does not state the number of sample observations. Does this omission affect your conclusions? What can you say about Type I and Type II errors?

4. Perform three simulations in R to illustrate the efficiency, consistency, and asymptotic normality of OLS estimates. Consider the following DGP:

$$y_i = 1 + 2x_{i1} + \epsilon_i$$

where $\epsilon \sim N(0, 1)$.

Show how the following affect

- $E[X_1 e] \neq 0$; and
- the presence of high multicollinearity, both on the consistency and efficiency of the asymptotic estimation.
- Now suppose that $E(X_1 e) = 0$; $E(X_2 e) \neq 0$; and $corr(X_1, X_2) \to 1$. Show the process $\hat{\beta}_1 \xrightarrow{p} \beta_1$. What is your conclusion?

5. (Optional) Suppose that $\{x_N\}$ is a sequence of random vectors of dimension K such that $x_N \xrightarrow{p} \beta$ and $\sqrt{N}(x_N - \beta) \xrightarrow{d} z$. Assume that $a(\cdot) : \mathbb{R}^K \to \mathbb{R}^r$ is once differentiable, with $A(\beta) \equiv \frac{\partial a(\beta)}{\partial \beta'}$ denoting the matrix of first derivatives of dimension $r \times K$, evaluated at β.

   a) Show that $\sqrt{N}(a(x_N) - a(\beta)) \xrightarrow{d} A(\beta)z$.

   b) Further assume that $\sqrt{N}(x_N - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$. What is z in this particular case?

6. (Optional) Let $\{y_i : i = 1, 2, ...\}$ be an independent, identically distributed sequence with $\mathbb{E}(y_i^2) < \infty$. Let $\mu = \mathbb{E}(y_i)$ and $\sigma^2 = \text{Var}(y_i)$.

   (a) Let $\overline{y}_N$ denote the sample average based on a sample size of N. Find $\text{Var}[\sqrt{N}(\overline{y}_N - \mu)]$.

   (b) What is the asymptotic variance of $\sqrt{N}(\overline{y}_N - \mu)$?

   (c) What is the asymptotic variance of $\overline{y}_N$? Compare this with $\text{Var}(\overline{y}_N)$.

   (d) What is the asymptotic standard deviation of $\overline{y}_N$?

   (e) How would you obtain the asymptotic standard error of $\overline{y}_N$?

APPLIED TOPIC

1. The following equation explains weekly hours of television viewing by a child in terms of the child's age, mother's education, father's education, and number of siblings:

$$tvhours^* = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 mothereduc + \beta_5 siblings + u$$

We are worried that $tvhours^*$ is measured with error in our survey. Let tvhours denote the reported hours of television viewing per week.

- Do you think the CEV assumptions are likely to hold? Explain.
- What do the classical errors-in-variables (CEV) assumptions require in this application?
- Show formally the bias that the measurement error in $tvhours*$ generates.

POTENTIAL OUTCOMES

1) Imagine you own a snickers store. An employee suggests giving "little gifts" (i.e. key rings) to clients to make them return to the store. You have 200 stores; an economist thus creates an experiment to evaluate the effect of the "little gifts" on revenues.

```
library(pacman)
p_load(tidyverse, glue)

# experimental data at the store level dataset
set.seed(22)
n_stores <- 200
true_gift_effect <- 100
noise <- 50
data_downstream <- tibble(store_id = 1:n_stores) %>%
  mutate(  #mutate allows to create new_columns in data frame.
    # treatment (random in this case)
    gives_gift=rbinom(n_stores, 1, prob =  0.5),
    # return rate increased by 20% if given gifts
    return_rate=rnorm(n_stores, mean = 0.5, sd=0.1) + gives_gift*0.1,
    # outcome (influenced by return rate)
    # gifs impact revenue through return rate
    revenue= 50 + true_gift_effect*10*return_rate + rnorm(n_stores, mean=0, sd=noise)
  )

# plot to visualize the relationship
data_downstream %>%
  mutate(treatment=ifelse(gives_gift==1, "gift", "no gift")) %>%
  ggplot(aes(return_rate, revenue, color=treatment)) +
  geom_point() +
  labs(title=glue(" ")) + geom_rug()
```

a. Discuss the data depicted in the graph plot and draw a Direct Acyclical Graph (check: https://mixtape.scunning.com/03-directed_acyclical_graphs) on the relationship between gifts, return rates, and revenues.

b. Run a regression of gifts on return rates. Does the gift make customers return to the store?

c. What is the regression a well-trained economist would run after the experiment to know the effect of the gifts?

d. Describe formally, this is with the use of the potential outcomes notation, why controlling for "return rates" in the experimental regression creates a bias.

e. Show how controlling for return rates biases our coefficient in a regression setting (i.e., run a regression controlling for return rates). Discuss your result and show formally, with the use of potential outcomes notation, why the regressor *gifts* is downward biased.