EAE6029 - Econometria I
Lista 1

Prof: Pedro Forquesato
pforquesato@usp.br

Faculdade de Economia, Administração e Contabilidade
Universidade de São Paulo

2024/1

Exercise list 1

Analytical problems

1. If $\mathbb{E}\left[Y|X\right] = a + bX$, find $\mathbb{E}\left[XY\right]$ as a function of moments of X.
2. Show that in the linear regression model (see slides) for any function $h(x)$ of the covariates, $\mathbb{E}\left[h(X)e\right] = 0$ as long as it is finite.
3. True of false. If $Y = X\beta + e$, $X \in \mathbb{R}$ and $\mathbb{E}\left[Xe\right] = 0$, then $\mathbb{E}\left[X^2 e\right] = 0$
4. Consider X and Y such that their joint density is $f(x,y) = 3/2(x^2 + y^2)$ on $[0,1]^2$. Compute the coefficients of the best linear predictor $Y = \alpha + \beta X + e$. Compute the conditional expectation function $m(x) = \mathbb{E}\left[Y|X = x\right]$. Are they different?
5. Consider the long and short projections $Y = X_1\gamma_1 + e$ and $Y = X_1\beta_1 + X_1^2\beta_2 + u$. When is $\gamma_1 = \beta_1$? What if we consider $Y = X_1\theta_1 + X_1^3\theta_2 + v$, is there a situation when $\theta_1 = \gamma_1$?
6. Consider the estimation of the (sample-wide) linear regression $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$. Now change the regressors to $\mathbf{Z} = \mathbf{XC}$, where $\mathbf{C}$ is a $k \times k$ non-singular matrix. How does this affect (i) the OLS estimates, and (ii) residuals of this regression?
7. Consider $\widetilde{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Find the OLS coefficient of a regression of $\widetilde{\mathbf{Y}}$ on $\mathbf{X}$.
8. Show that $\mathbf{M}$ is idempotent.
9. Show that if $\mathbf{X} = [\mathbf{X_1 X_2}]$, and $\mathbf{X_1}'\mathbf{X_2} = 0$, then $\mathbf{P} = \mathbf{P_1} + \mathbf{P_2}$
10. A friend suggests that the assumption that observations (Y_i, X_i) are i.i.d. implies that the errors e of the regression $Y = X'\beta + e$ are homoskedastic. Do you agree? Explain why.
11. Consider the model $Y = X'\beta + e$ and the (very important nowadays) *ridge regression* estimator:

$$\widehat{\beta}_{\text{RIDGE}} = \left(\sum_{i=1}^{n} X_i X_i' + \lambda I_k\right)^{-1} \left(\sum_{i=1}^{n} X_i Y_i\right)$$

Is $\widehat{\beta}_{\text{RIDGE}}$ unbiased estimator of β? Is it consistent?

12. Take a regression model with i.i.d. observations $(Y_i, X_i)_i$ and $X \in \mathbb{R}$, such that $Y = X\beta + e$, $\mathbb{E}[e|X] = 0$ and define $\Omega = \mathbb{E}\left[X^2 e^2\right]$. If $\widehat{\beta}$ is the OLS estimator and $\hat{e}_i$ the OLS residuals, find the assymptotic distribution $\sqrt{n}(\widehat{\Omega} - \Omega)$ of the estimators:

$$\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 e_i^2$$

$$\widetilde{\Omega} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \hat{e}_i^2$$

Computational/interpretative problems

For this list, we will use the `cps09mar.dta` file provided by the textbook author. You should **provide the code** and the results *together*.

```
library(haven)
library(kableExtra)
library(knitr)

# setwd() <- this might help

#  you can download the file directly in R, or just do it manually
# url <- "https://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics%20Data.zip"
# download.file(url, "./econ_data.zip")
# unzip("./econ_data.zip")

cps09mar <- read_dta("./cps09mar.dta")
knitr::kable(head(cps09mar, 10))
```

age	female	hisp	education	earnings	hours	week	union	uncov	region	race	marit
52	0	0	12	146000	45	52	0	0	1	1	1
38	0	0	18	50000	45	52	0	0	1	1	1
38	0	0	14	32000	40	51	0	0	1	1	1
41	1	0	13	47000	40	52	0	0	1	1	1
42	0	0	13	161525	50	52	1	0	1	1	1
66	1	0	13	33000	40	52	0	0	1	1	5
51	0	0	16	37000	44	52	0	0	1	1	1
49	1	0	16	37000	44	52	0	0	1	1	1

age	female	hisp	education	earnings	hours	week	union	uncov	region	race	marit
33	0	0	16	80000	40	52	0	0	1	1	1
52	1	0	14	32000	40	52	0	0	1	1	1

We are interested in running (linear) **Mincerian regressions** of the type $\ln(\text{wage}) = X'\beta + e$, where X is a vector of worker characteristics, the most important of which (for this kind of regression) is education.

1. Run a linear regression of log earnings on age, age squared, sex and education. What is the expected log earnings of a 20 years old woman as a function of her education? What is the average partial effect of another year of age on log earnings?

2. In the job market, an important predictor of wages is *experience*. Unfortunately, that is a variable almost universally missing from data sets, which at most have tenure at current work. So many applied economists **proxy** for experience as age - 15 (minimum working age at the data). Add this variable to the regression. What happens? Calculate $(\mathbf{X}'\mathbf{X})$ and explain.

For what folllows, remove age variables and leave only experience and experience squared. (Also: now that you are already half-way there, also calculate $\mathbf{X}'\mathbf{Y}$ and calculate by hand both $\widehat{\beta}$ and $\mathrm{Var}(\widehat{\beta})$)

3. Compute homoskedastic and heteroskedastic-robust standard errors of the estimators. Do they differ? Which one do you prefer? Now estimate cluster-robust standard errors at the level of the region. Discuss whether this is a reasonable approach in this case.

4. Now use the Frisch-Waugh-Lovell theorem to estimate the effect of education on wages, while partialling out the controls in (1). Is the estimated coefficient the same? What about its standard deviation?

5. Consider a 18 years-old prospective college student deciding whether to study or work. He wants to know the *ratio* θ between returns of education and returns to experience, given his age. Write θ as a function of the parameters β and estimate it.

6. Write out the formula for the asymptotic variance of $\hat{\theta}$ as a function of the variance-covariance matrix and find the standard deviation of the estimator.

7. Construct a 90% confidence interval for $\hat{\theta}$.

8. Test (jointly) whether β_{educ} equals $\beta_{\text{exp}} + 6\beta_{\text{exp}^2}$ and β_{educ} for men equals β_{educ} for women using a Wald statistic. Interpret.