Exercises Week 3

Econometrics

1. Let X, Z be matrix of regressors and consider the projection matrices P_X that projects onto the space generated by X, and $P_{X,Z}$ that projects onto the space generated by both X and Z.

   Prove that the matrix $Q_{XZ} = P_{X,Z} - P_X$ is an orthogonal projection matrix.

2. In the lectures, we show that the variance of the j element of the OLS estimator is given by

   $$Var(\hat{\beta}_j) = \sigma^2 (X_j^T M_{-j} X_j)^{-1},$$

   where X_j is the column of regressors associated to the j variable, and M_{-j} is the maker of residuals (the projection off) of the space generated by all columns of the matrix X besides the j one.

   Show that the variance of $\hat{\beta}_j$ can also be written as

   $$Var(\hat{\beta}_j) = \frac{\sigma^2}{(n-1)Var(X_j)} \left(\frac{1}{1 - R^2_{X_j|X_{-j}}} \right),$$

   where $R^2_{X_j|X_{-j}}$ is the R^2 from the regression of X_j on all other regressors.

   Hint: Start by writing $R^2_{X_j|X_{-j}}$ and recall that $M_\iota Z = Z - \bar{Z}$, where $\bar{Z}$ is the sample mean of Z.

3. Generate a sample of size 50 from the model

   $$y_t = \beta_1 + \beta_2 x_t + u_t,$$

   with $\beta_1 = 1$ and $\beta_2 = 0.8$. For simplicity, assume that x_t are $NID(2, 2)$ and that the u_t are $NID(0, 1)$. Now generate noisy versions of the regressors and regressand $y_t^* = y_t + u_{y,t}$, $x^* = x_t + u_{x,t}$, with $u_{y,t} \sim N(0, \sigma_y^2)$, $u_{x,t} \sim N(0, \sigma_x^2)$ for some values σ_y^2 and σ_x^2 of your choosing. Estimate by OLS the equations

   $$y_t^* = \beta_1 + \beta_2^y x_t + u_t,$$

   and

   $$y_t = \beta_1 + \beta_2^x x_t^* + u_t,$$

   and compare the value of the estimator for β_2 and its standard deviation between equations.

   Repeat at least 100 times and find the averages and variance of $\hat{\beta}_2^x$ and $\hat{\beta}_2^y$. Use these averages to estimate the bias of the OLS estimators of β_2.

   What happens to the bias and variance if you increase σ_y^2 and/or σ_x^2?

4. **Ex. 4.18 in ETM**: The file *classical.csv* contains 50 observations on three variables: Y, X_2, and X_3. These are artificial data generated from the classical linear regression model

   $$Y = \beta_1 \iota + \beta_2 X_2 + \beta_3 X_3 + U, \quad U \sim N(0, \sigma^2 I).$$

   Compute a t statistic for the null hypothesis that $\beta_3 = 0$. On the basis of this test statistic, perform an exact test. Then perform parametric and semiparametric bootstrap tests using 99, 999, and 9999 simulations. How do the two types of bootstrap P values correspond with the exact P value? How does this correspondence change as B (the number of simulations) increases?

5. **Econometric Model for Crime**: Cornwell and Trumbull (1994) estimated an economic model of crime using data on 90 counties in North Carolina observed over the years 1981–87. This data set is available on the *Ecdat* package under *Crime*. For this exercise, we consider the data as cross-sectional and reconsider the panel data structure later in the course.

The data contains information on (among others):

crmrte - crimes committed per person

prbarr - probability of arrest

prbconv - probability of conviction

prbpris - probability of prison sentence

avgsen - average sentence, days

polpc - police per capita

density - people per square mile

We are interested in estimating the model

$$crmrte = \beta_0 + \beta_1 prbarr + \beta_2 prbconv + \beta_3 prbpris + \beta_4 avgsen + \beta_5 polpc + \beta_6 density.$$

What should the expected signs for the estimators be? Explain your reasoning.

Estimate the model by OLS and compare the value and significance of the parameters with your expectation. Do they agree with your intuition?

According to the estimator, if we want to decrease the crime rate in our district, should we increase or decrease the number of police officers in the district? Why do you think we obtain this result?