ECO 375–Homework 2
University of Toronto
Due: 17 November, 2019
Late assignments will not be accepted
For full credit, please show your work

1 Theoretical Problems

1. True or false: First indicate whether the following statements are true or false and then justify your answer.

   (a) In the simple linear regression model if the R^2 is equal to one, then the linear relationship between the variables is exact and residuals are all zero.

   (b) In the simple linear regression model, if $\text{Var}(Y) = \text{Var}(X)$ then the estimated slope in a regression model of Y on X is approximately equal to the estimated slope in a regression model of X on Y.

   (c) The fact that R^2 is equal to zero indicates that variables are unrelated.

   (d) A crucial assumption of the linear model is that the sum of the residuals is zero.

   (e) The fact that residuals in the linear model estimated by least-squares have zero mean is a consequence of assuming that the expected value of the error term is zero.

   (f) The assumption that the error term is normally distributed is necessary to demonstrate that the least-squares estimator is unbiased.

2. Take $Y = \log(W)$. Assume the log-linear model $Y = \beta_0 + \beta_1 X + U$, with $E(U) = 0$. Prove the following:

   (a) Show that if $E(U|X) = 0$, then $Cov(X, U) = 0$.

   (b) Assume $Cov(X, U) = 0$. Show that $\beta_1 = Cov(X, Y)/Var(X)$.

   (c) Suppose $\widehat{\beta}_1$ is the OLS estimator of β_1. Show that $\widehat{\beta}_1 \xrightarrow{p} \beta_1 + \frac{Cov(X,U)}{Var(X)}$.

   (d) Assume $Cov(X, U) = 0$. What is the estimated *approximate* percentage change in W for a change in X, say from $X = x_0$ to $X = x_1$? And what is the estimated *exact* percentage change in W?

   (e) Assume $Cov(X, U) = 0$. Show that $e^{x\widehat{\beta}} - 1$ is a biased estimator for $e^{x\beta} - 1$. Show that $e^{x\widehat{\beta}} - 1$ is a consistent estimator for $e^{x\beta} - 1$.

2 Computer Based Problems

1. **Determinants of Income.** Use the dataset "ANES2016.dta" for this question. The data are drawn from the American National Election Survey of 2016 (available at https://electionstudies.org/data-center/2016-time-series-study/).

   The dataset includes log of income ($loginc$), gender indicator ($female$), indicators for black and hispanic ($black$, $hispanic$), age, five education dummy variables, numbered $educ0$ through $educ4$ (from "high school dropout" to "graduate or professional school"), among others. (Note the data labels on the variables.)

   (a) Take $educ0$, "high school dropout," to be the base level of education and estimate the following model using OLS:

   $$loginc_i = \beta_0 + \beta_1 female_i + \beta_2 black_i + \beta_3 age_i + \beta_4 age_i^2 + \beta_5 educ1_i + \beta_6 educ2_i + \beta_7 educ3_i + \beta_8 educ4_i + \varepsilon_i$$

   Assume all assumptions of the classical linear regression model hold. How should the coefficient on $educ1$ be interpreted? What about $educ4$?

   (b) Run the regression again, but now take $educ1$, not $educ0$, to be the base case. First, write down this regression equation, estimate the model parameters, and interpret the estimated coefficient on $educ4$. Is it possible to obtain the same result using the regression estimated in item (a)? If it is not possible, explain why. If it is possible, explain how.

   (c) Test whether age has significant impacts on income. Based on the estimated results, what is the (approximated) effect of an increase in age from 34 to 35 on income? In which age do we expect to see the maximum income level (holding all other covariates constant)?

2. **Economic Convergence.** The idea that poor countries grow faster than richer countries is a result central to many neoclassical growth models. This idea is often referred to in the literature as (absolute) β-convergence. Empirically, papers such as the influencial study by Robert J. Barro (1991, "Economic Growth in a Cross Section of Countries," published at the *Quarterly Journal of Economics*) demonstrate how β-convergence can be tested on a cross-section of economic data. For an early survey of the literature, see Sala-i-Martin (1994. "Cross-sectional Regressions and the Empirics of Economic Growth," published at the *European Economic Review*).

   To investigate this issue, let $y_{i,t}$ represent the GDP per capita of country i at year t, and consider the following regression model:

   $$\log\left(\frac{y_{i,t+k}}{y_{i,t}}\right) = \alpha + \beta\log(y_{i,t}) + u_{i,t}. \tag{1}$$

   The dependent variable measures the (approximate) growth rate of GDP per capita of country i between year t and $t+k$. The model assumes that the growth rate depends on the initial level of income per capita $y_{i,t}$, and on other (unobserved) factors $u_{i.t}$. If $\beta < 0$, richer countries are expected to have smaller growth rates than poorer countries, leading to the β-convergence.

   Please use the Penn World Tables dataset, "PWT_data.dta" for this question (the original data is available at https://www.rug.nl/ggdc/productivity/pwt/). For the remainder of this question, let $t = 1975$ and $t + k = 1995$. A description of variables is provided below:

Variable	Description
GDP1975	Real GDP of country i in 1975
GDP1995	Real GDP of country i in 1995
POP1975	Population of country i in 1975
POP1975	Population of country i in 1995
HCI1975	Human capital index of country i in 1975
GCF1975	Gross capital formation shares of country i in 1975

(a) Assume the Gauss-Markov assumptions are valid. Estimate equation (1) using ordinary least squares. Interpret the results. Do you find evidence in favor or against the β-convergence?

(b) Now we will add the human capital index for country i at time t, $HC_{i,t}$ into the model:

$$\log\left(\frac{y_{i,t+k}}{y_{i,t}}\right) = \alpha + \beta_1 \log(y_{i,t}) + \beta_2 HC_{i,t} + u_{i,t} \tag{2}$$

If $\beta_1 < 0$, then the group of countries are said to be *conditionally* β-convergent. Estimate equation (2) using OLS. Based on the estimated results, do you find evidence in favor of conditional economic convergence? Interpret the results and compare them with the results you found in (a).

(c) Now add one more variable to the regression - share of gross capital formation in country i:

$$\log\left(\frac{y_{i,t+k}}{y_{i,t}}\right) = \alpha + \beta_1 \log(y_{i,t}) + \beta_2 GCF_{i,t} + \beta_3 HC_{i,t} + u_{i,t} \tag{3}$$

Interpret the results. Do your conclusions from (b) change? Are both types of capitals jointly important to explain future growth?

3. **Monte Carlo Simulation**. Simulate the following model in STATA:

$$Y = \beta_0 + \beta_1 X + U$$

where

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} -10 \\ 5 \end{pmatrix}$$

$$X \sim U(0,1),$$

that is, X is uniformly distributed between 0 and 1; and

$$U \sim N(0,5).$$

For each simulation, generate a data set $\{y_i, x_i : i = 1, ..., n\}$ with $n = 100$ observations. Then, for each sample, estimate β using OLS, make the tests described below, and save the p-values. Run $m = 1000$ simulations.

(a) In each simulated data, perform the following hypothesis test: H_0: $\beta_1 = 5$ vs H_1: $\beta_1 \neq 5$, and save the p-value. In what fraction of the simulations can you reject the null hypotheses? Most likely, you will find that the fraction of rejections is not too far from 5%. Why is that true for this test?

(b) Now, in each simulated data, perform the following hypothesis tests:

i. H_0: $\beta_1 = 4.5$ vs H_1: $\beta_1 \neq 4.5$, and

ii. H_0: $\beta_1 = 0$ vs H_1: $\beta_1 \neq 0$,

and save the corresponding p-values. In what fraction of the simulations can you reject each null hypotheses? Are those fractions close to 5%? Which one is greater? Why are these results expected for these tests?

Provide your do file and log file as part of your submission.