Midterm Examination

Total points: 140
Instructor: In Choi

1. (30 points) Consider the linear regression model

$$y_t = \alpha + \beta t + u_t,\ u_t \sim iid\ N(0, \sigma^2), (t = 1, ..., T).$$

An estimator of β is

$$b = \frac{\sum_{t=1}^T y_t}{\sum_{t=1}^T t}.$$

Note that $\sum_{t=1}^T t = \frac{T(T+1)}{2}$.

(a) Show that b is a consistent estimator of β.
Ans. Write

$$b = \frac{T\alpha}{\sum_{t=1}^T t} + \beta + \frac{\sum_{t=1}^T u_t}{\sum_{t=1}^T t}.$$

Since $\left(\sum_{t=1}^T u_t\right)/T \xrightarrow{p} 0$ by the WLLN and $\left(\sum_{t=1}^T t\right)/T^2 \to \frac{1}{2}$, we have $\frac{\sum_{t=1}^T u_t}{\sum_{t=1}^T t} = \frac{\left(\sum_{t=1}^T u_t\right)/T^2}{\left(\sum_{t=1}^T t\right)/T^2} \xrightarrow{p} 0$ as $T \to \infty$. Moreover, $\frac{T\alpha}{\sum_{t=1}^T t} \to 0$ as $T \to \infty$. Thus, $b \xrightarrow{p} \beta$.

(b) Show that the estimator b is biased.
Ans. We have

$$Eb = \frac{T\alpha}{\sum_{t=1}^T t} + \beta \neq \beta.$$

This shows that b is biased.

(c) What is the probability limit of $T(b - \beta)$?
Ans. Write

$$T(b-\beta) = \frac{\alpha}{\left(\sum_{t=1}^T t\right)/T^2} + \frac{\left(\sum_{t=1}^T u_t\right)/T}{\left(\sum_{t=1}^T t\right)/T^2} \xrightarrow{p} 2\alpha + \frac{0}{1/2} = 2\alpha.$$

2. (10 points) Let $\{X_n,\ n = 1, 2, ...\}$ be a sequence of random variables taking values e^n and 0 with probability n^{-2} and $1 - n^{-2}$, respectively. Show that the probability limit of the sequence is zero.
Ans. Because for $\varepsilon > 0$

$$P[|X_n| > \varepsilon] = P[X_n = e^n] = n^{-2} \to 0 \text{ as } n \to \infty,$$

$X_n \xrightarrow{p} 0$.

3. (40 points) Comment on the validity of the following statements.

(a) BIC tends to select a smaller model (or a model with a less number of regressors) than AIC.
Ans. The penalty terms for AIC and BIC are $\frac{2(k+1)}{n}$ and $\frac{(k+1)\ln n}{n}$, respectively, where k is the number of regressors (excluding the constant term) and n is the sample size. Since $\ln n$ tends to be larger than 2, BIC has a larger penalty term for model size than AIC. This means that BIC tends to select smaller models than AIC.

(b) The regression model one is using is

$$y_t = \alpha + \beta x_t + u_t.$$

He/she decided to add another regressor z_t to the model. The regression results with the added regressor will show that the standard error for x_t decreases.
Ans. The SE of the OLS estimator of β is

$$Var(\hat{\beta} \mid all\ x) = \sqrt{\frac{\sum_{t=1}^{n} \hat{u}_t^2/n}{\sum_{t=1}^{n}(x_t - \bar{x})^2}}$$

When z is added as a regressor,

$$Var(\tilde{\beta} \mid all\ x)\sqrt{\frac{\sum_{t=1}^{n} \tilde{u}_t^2/n}{\sum_{t=1}^{n}(x_t - \bar{x})^2(1 - R_{xz}^2)}}.$$

Since $\sum_{t=1}^{n} \tilde{u}_t^2 \leq \sum_{t=1}^{n} \hat{u}_t^2$ and $1 - R_{xz}^2 \leq 1$, it is uncertain which standard error is smaller or larger.

(c) The small sample approach for statistical inference should be preferred to the large sample approach because it can provide precise inference results at each sample size.

Ans. If the normality assumption required for the small sample approach holds true in reality, the small sample approach is preferred to the large sample approach. But if it doesn't, there is no reason to prefer the small sample approach.

(d) In a regression model having both x and x^2 as regressors, the dependent variable tends to increase as x increases to a certain level and then starts to decrease if the coefficients of x and x^2 take positive and negative values, respectively.
Ans. Since $y = c_1x^2 + c_2x + c_3 = c_1(x + \frac{c_2}{2c_1})^2 + c_3 - \frac{c_2^2}{4c_1^2}$, y increases as x increases up to $x = -\frac{c_2}{2c_1}$ and then it decreases. Thus, the statement is true.

4. (60 points) Using the California school data, the following regression was run.

(a) One considers eliminating the regressor EL_PCT. What will happen to the sum of squared residuals if he/she does so? What do you expect will happen to the value of AIC?
Ans. The sum of squared residuals increases because R^2 decreases. AIC may or may not increase.

(b) Test the significance of the coefficient of "log(AVGINC)" at the 1% level.
Ans. The p-value is close to 0. Thus the coefficient is significant at the 1% level.

Dependent Variable: LOG(TESTSCR)
Method: Least Squares
Date: 10/14/15 Time: 00:26
Sample: 1 420
Included observations: 420

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.411799	0.025210	254.3396	0.0000
LOG(STR)	-0.010707	0.008031	-1.333212	0.1832
LOG(AVGINC)	0.042914	0.002130	20.14796	0.0000
EL_PCT	-0.000669	4.57E-05	-14.62454	0.0000

R-squared	0.712811	Mean dependent var	6.482924
Adjusted R-squared	0.710740	S.D. dependent var	0.029116
S.E. of regression	0.015659	Akaike info criterion	-5.466041
Sum squared resid	0.102007	Schwarz criterion	-5.427562
Log likelihood	1151.869	F-statistic	344.1751
Durbin-Watson stat	1.307265	Prob(F-statistic)	0.000000

(c) If AVGINC (average incomes) increases by 100 percent, what do you expect will happen to TESTSCR (test score)?
Ans. 4.29%

(d) Test the joint significance of all the regressors at the 1% level in the first regression. Use the small sample approach.
Ans. The p-value of the F-statistic is less than 0.01. Thus, the regressors are jointly significant at the 1% level.

(e) What does the "S.E. of regression" mean in the EViews output?
Ans. It is the square root of the sum of squared residuals divided by the sample size or the sample size minus the number of regressors including the constant term.