Introduction to Econometrics (Midterm Exam, 2 hours, 90 pts.)

1. (20 pts) X is a random variable with prob(X = 0) = 1 − k, prob(X = 1) = k. Let m be an estimator of k and

$$m = X_1^{(X_2+4)}$$

where X_1 and X_2 are two independent samples of X.

- (5 pts.) Find prob($X_1 + X_2 = 1$) in term of k
- (7 pts.) Find the probability distribution of m in term of k.
- (8 pts.) Is m biased? What is the variance of m?

2. (30 pts.) (3 unrelated short questions)

a) (10 pts.) Explain the relationship between E(X) and $\overline{X}$. Discuss their similarity and difference. Explain the situation that E(X) and $\overline{X}$ are equal.

b) (10 pts.) Somsak use a least-square estimator to estimate a linear relationship between (X,Y) with 3 observations. The value of Y in his observation are 1, 2 and 3. The error terms are $e_1 = 1$, $e_2 = 0$. (The value of e_3 is missing.) Find the value of R^2 of his estimation.

c) (10 pts.) Find first order condition for the value of b that minimizes

$$\sum_i (Y_i - bX_i - 1)^2$$

Find b in terms of Y_i and X_i. Also find the value of b under the following samples: $(X_1, Y_1) = (0,0)$, $(X_2, Y_2) = (1,2)$

3. (20 pts.) Consider the following assumption of linear regression model

A0: X is non-random

A1: $Y_i = \beta_1 + \beta_2 X_i + u_i$

A2: Some Xi's are different.

A3: E(u_i) = 0

A4: Var(u_i) = σ_u^2

A5: Cov(u_i, u_j) = 0 for i ≠ j

A6: u_i is normally distributed.

A7: $\beta_1 = 0$

a) (5 pts.) Under assumption A0-A7, how does the graph of Y_i and X_i look like? Draw a graph to answer your question.

b) (8 pts.) Suppose we have 2 observations of (X, Y): (1,Y_1) and (2, Y_2) (The value of X_1 is 1. The value of X_2 is 2.). We then want to estimate β_2 using the two following estimators

$$m_1 = \frac{Y_1}{2X_1} + \frac{Y_2}{2X_2} \quad \text{and} \quad m_2 = \frac{1}{3}\frac{Y_1}{X_1} + \frac{2}{3}\frac{Y_2}{X_2}$$

Which estimator (m_1 or m_2) is better? Explain in details.

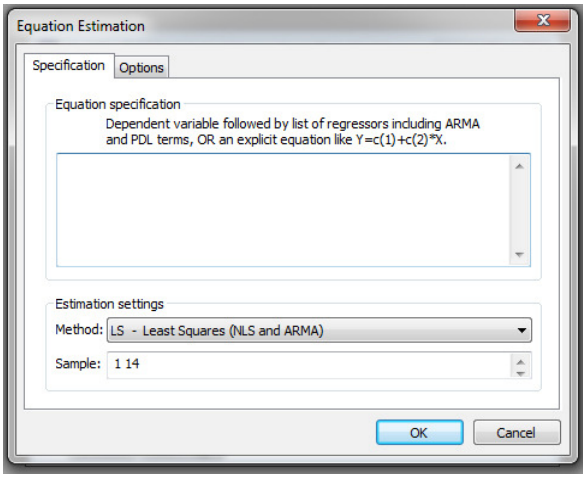
c) (7 pts.) Consider $m_3 = \lambda_1 \frac{Y_1}{X_1} + \lambda_2 \frac{Y_2}{X_2}$

Find the value of λ_1 and λ_2 that make m_3 unbiased and has minimum variance.

4. (20 pts.) You hypothesize that you could estimate the gross national product (GNP) in each year by using the money stock. Since there are several definitions of the

money stock, you selected two definitions: $M1$ and $M2$. You collected data of the variables -- $GNP, M1$ and $M2$ -- and did estimations in EViews.

4.1 (5 pts.) You imported data into the EViews and labeled your variable as gnp, m1 and m2 for the $GNP, M1$ and $M2$. You want to estimate the model $GNP = a + b * M1$ which a and b are coefficients of the model. As presented in Fig. 1, you open the window to type the model. What will you type?



[Fig. 1]

4.2 (10 pts.) You do the OLS estimation. As presented in Fig. 2, you get the estimation result.

- You wanted to save the result. Which function on the menu of the window will you click?
- Interpret the relationship between GNP and $M1$. (Both GNP and $M1$ are recorded in the unit of $ billion)
- From the result $R^2 \approx 0.991$, what does it mean?
- In box 1 where p-value of C(2) is 0.00, what is the null hypothesis?

Equation: EQ01 Workfile: EXAM::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: GNP
Method: Least Squares
Date: 07/12/12 Time: 12:32
Sample: 1 14
Included observations: 14
GNP=C(1)+C(2)*M1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-787.4724	77.96645	-10.10014	0.0000
C(2)	8.086305	0.219728	36.80141	0.0000 (1)

R-squared	0.991217	Mean dependent var	1981.971
Adjusted R-squared	0.990486	S.D. dependent var	782.0447
S.E. of regression	76.28221	Akaike info criterion	11.63832
Sum squared resid	69827.71	Schwarz criterion	11.72961
Log likelihood	-79.46824	Hannan-Quinn criter.	11.62987
F-statistic	1354.344	Durbin-Watson stat	1.096984
Prob(F-statistic)	0.000000		

[Fig. 2]

4.3. (5 pts.) You did another estimation of the model $GNP = c + d * M2$. As presented in Fig. 3, you got the result of the estimation. Compare between the models $GNP = a + b * M1$ and $GNP = c + d * M2$, which model is the best to explain the GNP? Why?

Equation: EQ02 Workfile: EXAM::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: GNP
Method: Least Squares
Date: 07/12/12 Time: 12:43
Sample: 1 14
Included observations: 14
GNP=C(3)+C(4)*M2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(3)	-44.06262	61.01338	-0.722180	0.4840
C(4)	1.587463	0.044833	35.40867	0.0000

R-squared	0.990520	Mean dependent var	1981.971
Adjusted R-squared	0.989730	S.D. dependent var	782.0447
S.E. of regression	79.25474	Akaike info criterion	11.71478
Sum squared resid	75375.76	Schwarz criterion	11.80607
Log likelihood	-80.00343	Hannan-Quinn criter.	11.70632
F-statistic	1253.774	Durbin-Watson stat	0.987363
Prob(F-statistic)	0.000000		

[Fig. 3]

Solution

1. prob($X_1 + X_2 = 1$) = prob($X_1 = 1, X_2 = 0$) + prob($X_1 = 0, X_2 = 1$)
   = k(1-k) + (1-k)k = 2k(1-k)

   $m = X_1^{(X_2+4)}$. Because X_1 = 0, 1. $X_1^{(X_2+4)} = X_1$.
   Then m = 1. Probability distribution of m is prob. distribution of X_1.
   prob(m = 0) = 1 − *k*, prob(m = 1) = *k*.

   E(m) = E(X_1) = E(X) = k, m is unbiased. Var(m) = V(X_1) = k(1-k)

2. a) $\overline{X}$ is an estimator of E(X). $E(\overline{X})$ = E(X). The two are equal in case the number of sample is very large.

   b) From $e_1 + e_2 + e_3 = 0$, $e_3 = -1$. $R^2 = 1 - \frac{\sum e_i^2}{\sum (Y_i - \overline{Y})^2} = 0$

   c) Let $S = \sum_i (Y_i - bX_i - 1)^2$

   $$\frac{dS}{db} = 2\sum_i (Y_i - bX_i - 1)(-X_i) = 0$$

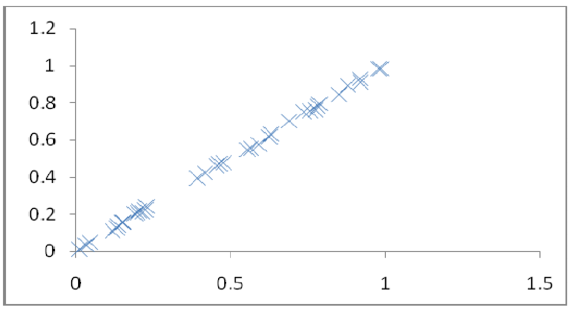
   $$\sum_i (X_iY_i - bX_i^2 - X_i) = 0$$

   $$b = \frac{\sum (X_iY_i - X_i)}{\sum X_i^2}$$

   $(X_1, Y_1) = (0,0)$, $(X_2, Y_2) = (1,2)$

   $$b = \frac{\sum (X_iY_i - X_i)}{\sum X_i^2} = \frac{(0.0 - 0) + (1.2 - 1)}{0^2 + 1^2} = 1$$

3)
a)



b) $m_1 = \frac{Y_1}{2X_1} + \frac{Y_2}{2X_2} = \frac{\beta_2 X_1 + u_1}{2X_1} + \frac{\beta_2 X_2 + u_2}{2X_2} = \beta_2 + \frac{u_1}{2X_1} + \frac{u_2}{2X_2} = \beta_2 + \frac{u_1}{2} + \frac{u_2}{4}$

$E(m_1) = \beta_2$. $V(m) = \sigma_u^2(\frac{1}{4} + \frac{1}{16}) = \frac{5}{16}\sigma_u^2$

Similarly, $m_2 = \beta_2 + \frac{u_1}{3} + \frac{u_2}{3}$. $E(m_2) = \beta_2$. $V(m) = \frac{2}{9}\sigma_u^2$

Both estimators are unbiased. m_2 is better because its variance is lower.

c) $m_3 = \lambda_1 \frac{Y_1}{X_1} + \lambda_2 \frac{Y_2}{X_2} = \lambda_1 Y_1 + \lambda_2 \frac{Y_2}{2}$

For unbiased $E(m_3) = \beta_2$ **=>** $\lambda_1 + \lambda_2 = 1$

$m_3 = \lambda_1 Y_1 + (1 - \lambda_1)\frac{Y_2}{2}$

$V(m_3) = \lambda_1^2 \sigma_u^2 + (1 - \lambda_1)^2 \frac{\sigma_u^2}{4}$

$\frac{dV(m_3)}{d\lambda_1} = 0 \Rightarrow \lambda_1 = \frac{1}{5}$ and $\lambda_2 = \frac{4}{5}$

4

4.1 gnp=c(1)+c(2)*m1 หรือใช้เป็น c(?) อะไรก็ได้ที่ไม่ใช่ c(0)

4.2 You do the OLS estimation. As presented in Fig. 2, you get the estimation result.

- You wanted to save the result. Which function on the menu of the window will you click?

Name

- Interpret the relationship between GNP and $M1$. (Both GNP and $M1$ are recorded in the unit of $ billion)

Holding other variables as constants, increasing of M1 for 1 $ billion will increase GNP for about 8.1 $ billion.

- From the result $R^2 \approx 0.991$, what does it mean?

99.1% of GNP can be explained by the model.

- In box 1 where p-value of C(2) is 0.00, what is the null hypothesis?

H0: C(2)=0

4.3. $GNP = a + b * M1$ is the best, since the model's $R^2 = 99.1$ is higher than the other model's.