Economics 466

Introduction to Econometrics

Spring 2009, Midterm Exam I

Total points - 100, Time - 1 hr. 15 min.

Note: Answer all questions clearly. Show your work for full credit.

1. (**50 points**) You are given the following data to estimate regression coefficients in several models.

$$n = 25, \bar{X} = 5, \bar{Y} = 10, \sum_i X_i^2 = 650, \sum_i Y_i^2 = 2600, \sum_i X_i Y_i = 1275. \tag{1}$$

   (a) (**5 points each**) Estimate the **intercept and slope coefficients** for the following models.
   (i) $Y_i = \beta_0 + \beta_1 X_i + u_i$
   (ii) $Y_i = \beta_0 + \beta_1 (X_i - \bar{X}) + u_i$
   (iii) $Y_i - \bar{Y} = \beta_0 + \beta_1 X_i + u_i$
   (iv) $Y_i - \bar{Y} = \beta_0 + \beta_1 (X_i - \bar{X}) + u_i$

   (b) (**12 points**) Compute the standard error of the slope coefficient and the *t-value* for testing the null hypothesis $H_0 : \beta_1 = 0$ for the model in (i). Do you think that the standard errors and the *t-values* of the slope coefficient for the models in (ii)-(iv) will be the same? Explain.

   (c) (**8 points**) Find R^2 in model (i) and show that $R^2 = r^2$ where r^2 is the correlation coefficient between the dependant and independent variables. Do you think that R^2 for the models in (ii)-(iv) will be the same? Why?

   (d) (**9 points**) Your friend Don is estimating the model (v) $Y_i = \beta_1 X_i + u_i$. Compute Don's slope coefficient and its standard error. If the true model is given in (i), which model would you prefer, (i) or (v)? Explain.

2. Using data on **53 observations**, the following model was estimated

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + u, \tag{2}$$

where Y = number of car sales (in thousands), X_1 = car price (in thousand dollars), X_2 = average family income (in thousand dollars). The regression results for this model are reported in the following table.

Coefficient	estimate	std. error	t-value
$\hat{\beta}_0$	-51.25	10.12	
$\hat{\beta}_1$	-1.45	0.29	
$\hat{\beta}_2$	1.13	0.64	
R^2	.81	–	—

(a) Compute the t-values in the last column of the table. **(6 points)**

(b) Interpret the coefficients of $\ln X_1$, $\ln X_2$ and R^2 in words. From the above table is it possible to find the marginal effects of an increase in price ($\partial\hat{Y}/\partial X_1$) and family income ($\partial\hat{Y}/\partial X_2$)? If not, what additional information do you need? **(12 points)**

(c) Explain how you would test the following hypothesis: (i) $\beta_1 = -1$, (ii) $\beta_2 = 1$, (iii) joint test on (i) and (ii). Choose your level of significance and perform the tests, when possible. **(12 points)**

(d) How would you test the hypothesis that $2\beta_1 + 3\beta_2 = 0$? Explain in details. **(6 points)**

(e) Your friend estimated the simple regression $\ln Y$ on $\ln X_1$ and found the slope coefficient to be -1.25. Using the result (discussed in class) relating simple to multiple regression coefficients ($\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\,\hat{\delta}$), find the value of $\hat{\delta}$ (which is the regression coefficient of $\ln X_2$ on $\ln X_1$). Based on this result what can you say about your friend's regression? **(8 points)**

(f) Use the results in the above table to test the hypothesis that there is no regression. **(8 points)**

Answer Key Spring 09 Exam I

Ques 1

a) i) $\hat{\beta}_1 = \dfrac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2}$

$= \dfrac{1275 - 25(5)(10)}{650 - 25(5)^2}$

$= \dfrac{25}{25} = 1$

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

$= 10 - (1)(5) = 5$

iii) $\hat{\beta}_1 = \dfrac{\Sigma X_i(Y_i - \bar{Y}) - n\bar{X}\,\overline{(Y_i - \bar{Y})}}{\Sigma X_i^2 - n\bar{X}^2}$

$= \dfrac{\Sigma X_i(Y_i - \bar{Y})}{\Sigma X_i^2 - n\bar{X}^2} = \dfrac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}^2}$

$= \dfrac{25}{25} = 1$

$\hat{\beta}_0 = \overline{(Y_i - \bar{Y})} - \hat{\beta}_1 \bar{X}$

$= -\hat{\beta}_1 \bar{X} = -5$

ii) $\hat{\beta}_1 = \dfrac{\Sigma(X_i - \bar{X})Y - n\overline{(X_i - \bar{X})}\,\bar{Y}}{\Sigma(X_i - \bar{X})^2 - n\left(\overline{X_i - \bar{X}}\right)^2}$

note that $\overline{X_i - \bar{X}} = \Sigma(X_i - \bar{X})/n = 0$

$\hat{\beta}_1 = \dfrac{\Sigma(X_i - \bar{X})Y}{\Sigma(X_i - \bar{X})^2} = \dfrac{\Sigma XY - \bar{X}\Sigma Y}{\Sigma X^2 - n\bar{X}^2}$

$= \dfrac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} = \dfrac{25}{25} = 1$

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \overline{(X_i - \bar{X})} = \bar{Y} = 10$

iv)

$\hat{\beta}_1 = \dfrac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y}) - n\bar{x}\bar{y}}{\Sigma(X_i - \bar{X})^2 - n\bar{x}}$

$\bar{x} = \dfrac{\Sigma(X_i - \bar{X})}{n}$; $\bar{y} = \dfrac{\Sigma(Y_i - \bar{Y})}{n}$

so $\bar{x} = 0$ and $\bar{y} = 0$

$\hat{\beta}_1 = \dfrac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$

$= \dfrac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2}$

$= \dfrac{25}{25} = 1$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0$

b) $se(\hat{\beta}_1) = \sqrt{\widehat{Var}(\hat{\beta}_1)}$

$$= \sqrt{\frac{\hat{\sigma}^2}{\Sigma(X_i - \bar{X})^2}}$$

$$\hat{\sigma}^2 = \frac{\Sigma \hat{u}_i^2}{n-2}$$

$$= \frac{SST - SSE}{n-2}$$

$$= \frac{\Sigma(Y_i - \bar{Y})^2 - \hat{\beta}_1^2 \Sigma(X_i - \bar{X})^2}{n-2}$$

$$= \frac{\Sigma Y^2 - n\bar{Y}^2 - \hat{\beta}_1^2 \Sigma X_i^2 + \hat{\beta}_1^2 n \bar{X}^2}{n-2}$$

$$= \frac{2600 - 25(10)^2 - 650 + 25(5)^2}{23}$$

$$= \frac{75}{23} = 3.26$$

$$se(\hat{\beta}_1) = \sqrt{\frac{3.26}{25}} = 0.361$$

t-value

$$t = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

$$= \frac{1}{0.361} = 2.77$$

2

The std errors will be the same. Why? Because the SST & SSE do not change for any of these models

The t-values will also be the same since the std errors are the same

c) $R^2 = \frac{SSE}{SST} = \frac{\hat{\beta}_1^2 \Sigma(X_i - \bar{X})^2}{\Sigma(Y_i - \bar{Y})^2}$

$$= \frac{\hat{\beta}_1^2 \Sigma X_i^2 - \hat{\beta}_1^2 n \bar{X}^2}{\Sigma Y^2 - n \bar{Y}^2}$$

$$= \frac{650 - 625}{2600 - 25(10)^2} = \frac{25}{100} = 0.25$$

$$r^2 = \frac{\{cov(x,y)\}^2}{Var X \; Var Y} = \frac{[\Sigma(X-\bar{X})(Y-\bar{Y})]^2}{\Sigma(X-\bar{X})^2 \; \Sigma(Y-\bar{Y})^2} = \frac{(\Sigma XY - n\bar{X}\bar{Y})^2}{(\Sigma X^2 - n\bar{X}^2)(\Sigma Y^2 - n\bar{Y}^2)}$$

$$= \frac{(1275 - 1250)^2}{(650 - 625)(2600 - 2500)} = \frac{25^{\cancel{2}}}{\cancel{25} \cdot 100} = 0.25 = R^2$$

The R^2 will be the same. The SSE and SST do not change for each model

d) $\tilde{\beta}_1 = \frac{\Sigma XY}{\Sigma X^2} = \frac{1275}{650} = 1.96$

$$Se(\tilde{\beta}_1) = \sqrt{\widehat{Var \, \tilde{\beta}_1}}$$

$$= \sqrt{\frac{\hat{\sigma}^2}{\Sigma X^2}}$$

$$= \sqrt{\frac{3.26}{650}} = 0.0708$$

d) cont'd It depends.
Since $\hat{\beta}_1$ is unbiased & $\tilde{\beta}_1$ is biased we cannot say that we will choose the one with minimum variance. Comparing variances will not make sense.
If unbiasedness is more important then prefer (i)
If minimum variance is more important then prefer (v)

Ques 2

a)

Coefficient	estimate	std. error	t-value
$\hat{\beta}_0$	-51.25	10.12	-5.06
$\hat{\beta}_1$	-1.45	0.29	-5
$\hat{\beta}_2$	1.13	0.64	1.766

using $t = \dfrac{\hat{\beta}_j}{se(\hat{\beta}_j)}$

b) $\hat{\beta}_1$: Holding average family income fixed, a 1 percent increase in car price will lead to a 1.45 percent decrease in the estimated number of car sales.

$\hat{\beta}_2$: Holding car price fixed, a 1% increase in average family income will lead to a 1.13 % increase in the estimated number of car sales.

R^2: 81% of the total sample variation in # of car sales is explained by car price and avg family income.

$$\frac{\partial \ln Y}{\partial \ln X} = \frac{\frac{\partial Y}{Y}}{\frac{\partial X}{X}} = \frac{\partial Y}{\partial X} \cdot \frac{X}{Y} \Rightarrow \frac{\partial Y}{\partial X} = \frac{\partial \ln Y}{\partial \ln X} \cdot \frac{Y}{X}$$

Since β_i $i=1,2$ captures only $\dfrac{\partial \ln Y}{\partial \ln X_i}$ then you need more information. You also need $\dfrac{Y}{X_i}$ $i=1,2$

Question 2 cont'd

C) i) step 1: set up null hypothesis & alternative hypothesis

$$H_0: \beta_1 = -1$$

$$H_1: \beta_1 \neq -1$$

step 2: then calculate t-value

$$t = \frac{\hat{\beta}_1 + 1}{se(\hat{\beta}_1)} = \frac{-1.45 + 1}{} = -1.55$$

choosing 10% level of significance at 50 df, $t_{50} = 1.675905$
choosing 5% " " " " 50 df, $t_{50} = 2.008559$
" 1% " " " " 50 df, $t_{50} = 2.677793$

we fail to reject the null hypothesis since $|t\text{-stat}| < t_c$

ii) step 1: set up null hypothesis & alternative hypothesis

$$H_0: \beta_2 = 1$$

$$H_1: \beta_2 \neq 1$$

step 2: then calculate t-value

$$t = \frac{\hat{\beta}_2 - 1}{se(\hat{\beta}_2)} = \frac{1.13 - 1}{0.64} = 0.203$$

choosing 10% level of significance at 50 df, $t_{50} = 1.675905$
" 5% " " " " 50 df, $t_{50} = 2.008559$
" 1% level " " " 50 df, $t_{50} = 2.677793$

we fail to reject the null hypothesis since $|t\text{-stat}| < t_c$

iii) i) & ii) jointly

step 1: $H_0: \beta_1 = -1, \ \beta_2 = 1$

H_1: null is not true

This is a joint test so it requires the use of the F_dist.

step 2: Calculati F-stat.

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/n-k-1}$$

Important note:
R^2 cannot be used in this case, because the SST for the restricted regression changes.

ie $\beta_1 = -1$ & $\beta_2 = 1 \Rightarrow \ln Y + \ln X_1 - \ln X_2 = \beta_0 + u$

$\Rightarrow \ln Y^* = \beta_0 + u$ where $Y^* = \frac{X_1 Y}{X_2}$

Step 3: Find $F_{2,30}$ at 5%, 1% or 10%, then compare F-stat. If F-stat falls outside critical region, we reject the null.

d) set $\theta_1 = 2\beta_1 + 3\beta_2$ $\quad \hat{\theta} = 2\hat{\beta}_1 + 3\hat{\beta}_2$

$H_0: \theta_1 = 0$

$H_1: \theta_1 \neq 0$

In order to get $\hat{\theta}$ we first have to modify the equation

$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + u$

$$\beta_1 = \frac{\theta - 3\beta_2}{2}$$

$$\ln Y = \beta_0 + \frac{\theta_1 - 3\beta_2}{2} \ln X_1 + \beta_2 \ln X_2 + u$$

$$2\ln Y = 2\beta_0 + \theta_1 \ln X_1 + \beta_2 (2\ln X_2 - 3\ln X_1) + u$$

let $Y^* = Y^2$; $\beta_0^* = \beta_0$; $X_2^* = X_2^2 / X_1^3$

then $Y^* = \beta_0^* + \theta_1 \ln X_1 + \beta_2 \ln X_2^* + u$

Run this regression to get

$$\hat{Y}^* = \hat{\beta}_0^* + \hat{\theta}_1 \ln X_1 + \hat{\beta}_2 \ln X_2^*$$

$\hat{\theta}_1$ is the coefficient on $\ln X_1$. Extract this coefficient & its corresponding std. error when you run this regression in a software package such as Excel.

then we can calculate the t-stat.

$t = \dfrac{\hat{\theta}}{se(\hat{\theta})}$. We compare this t-stat to the critical value, t_{50} at our chosen significance level to determine whether we reject or fail to reject the null hypothesis.

e) $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \hat{\delta} \Rightarrow \hat{\delta} = \dfrac{\tilde{\beta}_1 - \hat{\beta}_1}{\hat{\beta}_2} = \dfrac{-1.25 + 1.45}{1.13} = 0.177$

First since $\hat{\delta} \neq 0$, we can say that there is correlation between X_1 and X_2 so my friend's regression (simple) is not identical to the multiple regression. We can also assume that the partial relationship between X_2 & Y is positive & the correlation between X_1 & X_2 is also positive. So it would seem there is a positive bias.

f) $H_0: \beta_1 = 0, \beta_2 = 0$

H_1: null is not correct

$$F = \frac{R^2_{UR}/q}{(1 - R^2_{UR})/n-k-1} = \frac{0.81/2}{0.19/50} = 106.58$$

at 5% level of significance, $F_{2,50} = 3.1826$. Hence we reject the null. Given the large value of the F-stat, we can safely say that we can reject the null at any level of significance.