Second Midterm Exam
Economics 401
Thurs., Feb. 28, 2008

Show All Work. Only partial credit will be given for correct answers if we can not figure out how they were derived.

Note as well that the points are not evenly distributed across problems as some are more involved than others.

Points:

Problem 1:	30
Problem 2:	40
Problem 3:	15
Problem 4:	15
Total:	100

Problem 1: You have data on the amount of per capita gasoline consumption (in automobiles) by county across the united states.

Consider the following two different regressions:

$$G_i = \alpha_0 + \alpha_1 P_i + \varepsilon_i$$
$$G_i = \beta_0 + \beta_1 P_i + \beta_2 I_i + u_i$$

where

- G_i is per capita gasoline consumption
- P_i is the average price of gasoline in the county
- I_i is the average annual income in the county

a) What sign do you think your estimates of β_1 and β_2 would take if you ran this regression on real data?

I think that richer counties almost certainly buy more gas so I think $\beta_2 > 0$. β_1 is harder to sign. It depends on what varies more across counties, demand or supply. In general I think prices probably vary because the cost of providing gas varies a lot and simple economics tells us that when the price is high, people should consume less gas which implies that $\beta_1 < 0$.

b) First think of the model in a *causal* way. From this perspective why would you prefer the second regression to the first? How do you think α_1 and β_1 would compare?

I prefer the second to the first because income is an omitted variable in the first regression. The omitted variable bias is $\beta_2 \delta_1$ where δ_1 is the coefficient from a regression of Income on gas price. I think income and price are probably positively related because rich people are willing to pay more for gas (lower elasticity). Since $\beta_2 > 0$ this means that the omitted variable bias is positive, so when we include Income the coefficient on price should fall. This means that I think we will find $\alpha_1 > \beta_1$.

c) Now think of the model in a *descriptive* way. From this perspective explain how the interpretation of α_1 and β_1 differ.

From a descriptive perspective α_1 describes the relationship between gasoline prices and consumption. Thus if you show up into two towns and the price of gas is \$0.10 more in the second town, you would expect $0.1\alpha_1$ more gas consumption in that place. β_1 does the same thing except it conditions on income. Thus if you show up into two towns and the price of gas is \$0.10 more in the second town and the average annual income is identical, you would expect $0.1\beta_1$ more gas consumption in that place.

d) From a *forecasting* perspective, suppose you ran the second regression. If the price of gasoline in Madison were 3.00 and the average income in Madison was \$30,000, what would your forecast of gasoline consumption be?

My forecast would be

$$\begin{aligned} & \beta_0 + \beta_1 P_i + \beta_2 I_i \\ = \quad & \beta_0 + \beta_1 3 + \beta_2 30000. \end{aligned}$$

Problem 2: I used the data CRIME2 from Wooldridge for the following regression

$$\text{lpcinc}_i = \beta_0 + \beta_1 \text{loffic}_i + \beta_2 \text{lpop}_{2i} + \beta_3 + \text{unem}_{3i} + u_i$$

where lpcinic is the log of the crime rate, loffic is the log of the number of police officers, lpop is the log of the population, and unem is the unemployment rate. You get the following results using 92 observations:

regressor	coefficient	standard error
intercept	8.373	0.584
loffic	-0.110	0.069
lpop	0.144	0.077
unem	-0.052	0.006
R^2	0.5238	

You also run regressions of Y_i on subsets of the data and get the following results (all regressions include an intercept):

Included Regressors	R^2
loffic,lpop	0.0652
loffic,unem	0. 5052
lpop,unem	0.5101
loffic	0.0111
lpop	0.0003
unem	0.5041

a) Form a 95% confidence interval for β_2

Looking at the T-table, the degrees of freedom is approximately 90 which gives a critical value of 1.99. The the confidence interval should look like this:

$$\begin{aligned}&(\widehat{\beta}_2 - 1.99se(\widehat{\beta}_2), \widehat{\beta}_2 + 1.99se(\widehat{\beta}_2))\\=&(0.144 - 1.99 \times 0.077, 0.144 + 1.99 \times 0.077)\\=&(-0.009, 0.297)\end{aligned}$$

b) Form a 90% confidence interval for β_0

Looking at the t-table, the degrees of freedom is approximately 90 which gives a critical value of 1.66

$$\begin{aligned}&(\widehat{\beta}_0 - 1.66se(\widehat{\beta}_0), \widehat{\beta}_0 + 1.66se(\widehat{\beta}_0))\\=&(8.373 - 1.66 \times 0.584, 8.373 + 1.66 \times 0.584)\\=&(7.404, 9.342)\end{aligned}$$

Test the following Null Hypotheses at the 5% level. (Decide yourself whether to use a one or two sided test when necessary and explain why.):

c)

$$H_0 : \beta_1 = 0$$

Since we expect more police to deter crime, the interesting alternative is $\beta_1 < 0$. Thus I use a one sided test so the critical value is -1.66. The t-statistic is

$$\frac{\widehat{\beta_1}}{se\left(\widehat{\beta_1}\right)} = \frac{-0.110}{0.069} = -0.727 > -1.66$$

so we fail to reject the null hypothesis

d)

$$H_0 : \beta_2 = 0.2$$

I don't have a strong opinion about whether β_2 should be larger or smaller than 0.2, so I will use a two sided test meaning the critical value is 1.99. The t-statistic is

$$\frac{\widehat{\beta_2} - 0.2}{se\left(\widehat{\beta_2}\right)} = \frac{0.144 - 0.2}{0.077} = 1.879 < 1.99$$

e)

$$\begin{aligned} H_0 : \beta_1 &= 0 \\ \beta_2 &= 0 \\ \beta_3 &= 0 \end{aligned}$$

This is testing the significance of the whole regression. The R^2 form of the f-statistic is:

$$F = \frac{\left(R^2\right)/3}{(1 - R^2)/(92 - 3 - 1)} = \frac{(0.5238)/3}{(1 - 0.5238)/88} = 32.27$$

To get the critical value we look at the F with 3 degrees of freedom in the numerator and approximately 90 in the denominator which is 2.71. We clearly reject the null.

f)

$$\begin{aligned} H_0 : \beta_2 &= 0 \\ \beta_3 &= 0 \end{aligned}$$

Now we want to compare the R^2 of the unrestricted regression to the $R^2$of the restricted regression. The restricted regression in this case is the one in which we only include loffic, so the R^2 in this regression is 0.011. Thus the F-statistic is

$$F = \frac{\left(R_u^2 - R_r^2\right)/2}{(1 - R_u^2)/(92 - 3 - 1)} = \frac{(0.5238 - 0.011)/2}{(1 - 0.5238)/88} = 47.38$$

Looking at the back of the book the critical value is about 3.10 so we clearly reject the null.

g) Calculate $\overline{R}^2$ for this regression (that is the full regression).
The formula is

$$\overline{R}^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - K - 1} = 1 - \frac{(1 - .5283)(92 - 1)}{92 - 3 - 1} = 0.507.$$

Problem 3:

Consider the regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + u_i$$

Explain what procedure you would follow to test the joint null hypothesis

$$\begin{aligned} H_0 : \beta_1 &= \beta_4 \\ B_3 &= 0 \end{aligned}$$

The unrestricted model is the one above and let SSR_ube the regression sum of squares from the unrestricted model.

We can impose the restriction by writing

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + u_i \\ &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_1 X_{i4} + u_i \\ &= \beta_0 + \beta_1 \left(X_{i1} + X_{i4} \right) + \beta_2 X_{i2} + u_i \end{aligned}$$

This is the unrestricted model. I can obtain SSR_r by running a regression of Y_i on $(X_{i1} + X_{i4})$ and X_{i2}. I then can test the restriction using the F-statistic

$$\frac{(SSR_r - SSR_u)/2}{SSR_u/(N-5)}$$

Problem 4:

Suppose you run a regression with the sample regression function:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \hat{u}_i.$$

What is

$$\sum_{i=1}^{N}(2 + 3X_{i1} - X_{i2})\hat{u}_i?$$

(You will get no credit for the right answer to this question unless you show your work.)

$$\begin{aligned} &\sum_{i=1}^{N}(2 + 3X_{i1} - X_{i2})\hat{u}_i \\ &= \sum_{i=1}^{N} 2\hat{u}_i + \sum_{i=1}^{N} 3X_{i1}\hat{u}_i - \sum_{i=1}^{N} X_{i2}\hat{u} \\ &= 2\sum_{i=1}^{N} \hat{u}_i + 3\sum_{i=1}^{N} X_{i1}\hat{u}_i - \sum_{i=1}^{N} X_{i2}\hat{u} \\ = 0 \end{aligned}$$

where the whole thing is zero because from the normal equations we know that

$$0 = \sum_{i=1}^{N} \hat{u}_i = \sum_{i=1}^{N} X_{i1}\hat{u}_i = \sum_{i=1}^{N} X_{i2}\hat{u}_i.$$