Midterm Exam

Economics 401

Thurs., Feb. 28, 2008

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

Points:

Problem 1:	25
Problem 2:	25
Problem 3:	25
Problem 4:	25
Total:	100

Problem 1: Your goal is to forecast the amount of snow that we will have in March based on the amount of snow we have had in February. Let f_i denote the amount snow in February and m_i the amount of snow in March (both measured in inches).

Suppose you have a friend that works for the weather channel. She is not able to get you all of the data in march, but for Madison in the last N years she tells you

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{M} f_i &= \overline{f} = 10 \\
\frac{1}{N}\sum_{i=1}^{M} m_i &= \overline{m} = 5 \\
\frac{1}{N}\sum_{i=1}^{N} \left(f_i - \overline{f}\right)\left(f_i - \overline{f}\right) &= 150 \\
\frac{1}{N}\sum_{i=1}^{N} \left(f_i - \overline{f}\right)\left(m_i - \overline{m}\right) &= -30
\end{aligned}$$

Further suppose that you know that the snowfall in February was 20 inches. Using this data, forecast the amount of snow in march.

Solution: We know from class that

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\frac{1}{N}\sum_{i=1}^{N}\left(f_i - \overline{f}\right)\left(m_i - \overline{m}\right)}{\frac{1}{N}\sum_{i=1}^{N}\left(f_i - \overline{f}\right)\left(f_i - \overline{f}\right)} \\
&= \frac{-30}{150} \\
&= -\frac{1}{5}
\end{aligned}$$

$$\begin{aligned}
\widehat{\beta}_0 &= \overline{m} - \widehat{\beta}_1 \overline{f} \\
&= 5 - \left(-\frac{1}{5}\right) 10 \\
&= 7
\end{aligned}$$

Then we can write the predicted value as

$$\begin{aligned}
\widehat{\beta}_0 + \widehat{\beta}_1 20 &= 7 - \frac{1}{5} 20 \\
&= 3
\end{aligned}$$

Problem 2: Suppose you have data on the wages (w_i) and number of hours per week (h_i) of workers in the population. You regress hours on wages (measured in dollars per hour) and get the following population regression function

$$\widehat{h_i} = 34.06 + 0.52 w_i.$$

a) Think of the model as a *descriptive* model. Interpret the estimated slope coefficient (0.52) in this context.

If I randomly drew the wage of a worker making \$5 per hour and randomly drew a worker making \$6 per hour I would expect the person making \$6 per hour to work 0.52 hours per week more than the other worker.

b) Now suppose your model is *causal.* Interpret the coefficient in this case.

If I increase the wage of a particular worker by a dollar, that person will work an extra 0.52 hours per week.

c) Why does this not necessarily represent a causal effect? What specifically are other variables that could be related to this relationship?

There are a lot of other variables that could be correlated with wages and are important in determining hours. Examples: gender, education, test scores, family background, religion.

d) Choose one of the variables you mentioned *c*). If you could get data on in and included it in the regression, how would the new estimate compare to 0.52? Please be explicit in your answer (i.e. explain it using the formula for omitted variable bias).

Consider education. The omitted variable bias is $\beta_2 \delta_1$, where in this case β_2 picks up the effect of education on hours and δ_1 picks up the relationship between hours and wages. I would expect more educated people likely work longer ($\beta_2 > 0$) and we know that educated people tend to earn more ($\delta_1 > 0$). Putting these together means that the omitted variable bias is positive. Therefore, if I included education in the model I would expect the coefficient to fall.

Problem 3: Suppose that you have data on x, y, w, z where

$$
\begin{aligned}
w &= 10y \\
z &= 3x
\end{aligned}
$$

Consider the following 4 sample regression functions

$$
\begin{aligned}
\widehat{y}_i &= \widehat{a}_0 + \widehat{a}_1 x_i \\
\widehat{w}_i &= \widehat{b}_0 + \widehat{b}_1 x_i \\
\widehat{y}_i &= \widehat{c}_0 + \widehat{c}_1 z_i \\
\widehat{w}_i &= \widehat{d}_0 + \widehat{d}_1 z_i
\end{aligned}
$$

a) What is the relationship between $\widehat{b}_0$ and $\widehat{a}_0$ and between $\widehat{b}_1$ and $\widehat{a}_1$?

Doing things as in class

$$
\begin{aligned}
a_0 + a_1 x &= E(y \mid X = x) \\
&= \frac{1}{10} E(w \mid X = x) \\
&= \frac{1}{10} [b_0 + b_1 x]
\end{aligned}
$$

so

$$
\begin{aligned}
\widehat{b}_0 &= 10\widehat{a}_0 \\
\widehat{b}_1 &= 10\widehat{a}_1
\end{aligned}
$$

b) What is the relationship between $\widehat{c}_0$ and $\widehat{a}_0$ and between $\widehat{c}_1$ and $\widehat{a}_1$?

$$
\begin{aligned}
a_0 + a_1 x &= E(y \mid X = x) \\
&= E(y \mid Z = 3x) \\
&= [c_0 + c_1 3x]
\end{aligned}
$$

so

$$
\begin{aligned}
\widehat{c}_0 &= \widehat{a}_0 \\
\widehat{c}_1 &= \frac{1}{3}\widehat{a}_1
\end{aligned}
$$

c) What is the relationship between $\widehat{d}_0$ and $\widehat{a}_0$ and between $\widehat{d}_1$ and $\widehat{a}_1$?

$$
\begin{aligned}
a_0 + a_1 x &= E\left(y \mid X = x\right) \\
&= \frac{1}{10} E\left(y \mid Z = 3x\right) \\
&= \frac{1}{10}\left[d_0 + d_1 3x\right]
\end{aligned}
$$

so

$$
\begin{aligned}
\widehat{d}_0 &= 10\widehat{a}_0 \\
\widehat{d}_1 &= \frac{10}{3}\widehat{a}_1
\end{aligned}
$$

Problem 4: Suppose all of the assumptions of the classical linear regression model apply, except that rather than assume $E(u_i \mid x_i) = 0$, you assume that

$$E(u_i \mid x_i) = \alpha x_i.$$

What is the expected value of the OLS slope coefficient $(\widehat{\beta}_1)$ in this case?

$$\begin{aligned}
E\left(\widehat{\beta}_1 \mid x_1, ..., x_N\right) &= E\left(\beta_1 + \frac{\sum (x_i - \overline{x}) u_i}{\sum (x_i - \overline{x})^2} \mid x_1, ..., x_N\right) \\
&= \beta_1 + \frac{\sum (x_i - \overline{x}) E(u_i \mid x_1, ..., x_N)}{\sum (x_i - \overline{x})^2} \\
&= \beta_1 + \frac{\sum (x_i - \overline{x}) \alpha x_i}{\sum (x_i - \overline{x})^2} \\
&= \beta_1 + \alpha \frac{\sum (x_i - \overline{x}) x_i}{\sum (x_i - \overline{x})^2} \\
&= \beta_1 + \alpha
\end{aligned}$$

Then, we obtain $E[\hat{\beta}_1] = E\Big[E[\hat{\beta}_1 | x_1, x_2 \cdots, x_n]\Big] = \beta_1 + \alpha$.