Econ 203B: Single Equation Models

Solutions for 1999 Midterm

Michael Powell
Department of Economics, UCLA
February 11th, 2006

1. (30 points) In a study of turnover in the labor market, the following results were obtained for the period $1950 - I$ to $1979 - IV$ (here, the Roman numbers denote quarters):

$$\begin{aligned}\widehat{\ln Y_t} &= \hat{\beta}_0 + \hat{\beta}_1 \ln X_{1t} + \hat{\beta}_2 \ln X_{2t} + \hat{\beta}_3 \ln X_{3t} + \hat{\beta}_4 X_{4t} \\ &= \underset{(1.04)}{4.47} - \underset{(0.06)}{0.34} \ln X_{1t} + \underset{(0.40)}{1.22} \ln X_{2t} + \underset{(0.73)}{0.80} \ln X_{3t} - \underset{(0.002)}{0.0054} X_{4t}\end{aligned}$$

Numbers in parenthesis are estimated standard errors. The variables are:

Y = quit rate in manufacturing (number of people leaving jobs voluntarily per 100 employees)

X_1 = adult male unemployment rate.

X_2 = percentage of employees younger than 25.

X_3 = percentage of women employees.

X_4 = time trend (1950-I = 1)

a. (10 points) What is the meaning (interpretation) of the estimated slope parameters? Give an intuition for their sign.

Solution $\hat{\beta}_1 = -0.34$ suggests that the elasticity of the quit rate with respect to the adult male unemployment rate is -0.34. That is, a one percentage point increase in the adult male unemployment rate is associated with a 0.34 percentage point decrease in the quit rate in manufacturing. This suggests that, when there is more unemployment, people are less likely to quit their job, presumably because they believe that finding a replacement job will be more difficult.

$\hat{\beta}_2 = 1.22$ suggests that the elasticity of the quit rate with respect to the percentage of employees younger than 25 is 1.22. That is, a one percentage point increase in the percentage of employees younger than 25 is associated with a 1.22 percentage point increase in the quit rate in manufacturing. This is probably a result of the fact that younger people have higher turnover in general.

$\hat{\beta}_3 = 0.80$ suggests that the elasticity of the quit rate with respect to the percentage of employees who are women is 0.80. That is, a one percentage point increase in the percentage of employees who are women is associated with a 0.80 percentage point increase in the quit rate in manufacturing. Thankfully, this coefficient is not statistically significant so that I am not required to say anything politically incorrect.

The last coefficient is a bit more tricky, since it deals with semi-logs:

$$-0.54 = 100\hat{\beta}_4 = 100\frac{\partial \widehat{\ln Y_t}}{\partial X_{4t}} = \frac{100 \cdot \frac{\partial Y_t}{Y_t}}{\partial X_{4t}} \approx \frac{\%\Delta Y_t}{\Delta X_{4t}}$$

That is, a one unit increase in time is associated with a 0.54 percentage point decrease in the quit rate in manufacturing. (i.e. each quarter, the quit rate is expected to decrease by 0.54 percentage points)

b. (10 points) Which of the coefficients are statistically significant at the 5% significance level?

Solution Since $n = 120$, I will use the critical value for the standard normal: $c^*_{0.05,N(0,1)} = 1.96$. Performing five $t - tests$ for statistical significance of the coefficients, we have:

$$\left|t_0^{\hat{\beta}_0}\right| = \left|\frac{\hat{\beta}_0}{\widehat{se\left(\hat{\beta}_0\right)}}\right| = \left|\frac{4.47}{1.04}\right| = 4.2981 > 1.96 = c^*_{0.05,N(0,1)}$$

$$\left|t_0^{\hat{\beta}_1}\right| = \left|\frac{\hat{\beta}_1}{\widehat{se\left(\hat{\beta}_1\right)}}\right| = \left|-\frac{0.34}{0.06}\right| = 5.6667 > 1.96 = c^*_{0.05,N(0,1)}$$

$$\left|t_0^{\hat{\beta}_2}\right| = \left|\frac{\hat{\beta}_2}{\widehat{se\left(\hat{\beta}_2\right)}}\right| = \left|\frac{1.22}{0.40}\right| = 3.05 > 1.96 = c^*_{0.05,N(0,1)}$$

$$\left|t_0^{\hat{\beta}_3}\right| = \left|\frac{\hat{\beta}_3}{\widehat{se\left(\hat{\beta}_3\right)}}\right| = \left|\frac{0.80}{0.73}\right| = 1.0959 \leq 1.96 = c^*_{0.05,N(0,1)}$$

$$\left|t_0^{\hat{\beta}_4}\right| = \left|\frac{\hat{\beta}_4}{\widehat{se\left(\hat{\beta}_4\right)}}\right| = \left|-\frac{0.0054}{0.002}\right| = 2.7 > 1.96 = c^*_{0.05,N(0,1)}$$

Therefore, we can reject the four null hypotheses: $H_0 : \beta_0 = 0$, $H_0 : \beta_1 = 0$, $H_0 : \beta_2 = 0$, $H_0 : \beta_4 = 0$ (i.e. $\beta_0, \beta_1, \beta_2$, and β_4 are statistically significant), but we cannot reject the null hypothesis $H_0 : \beta_3 = 0$. (i.e. β_3 is not statistically significant.

c. (10 points) How would you test at the 5% significance level that the effect of the proportion of employees younger than 25 on the quit rate is the same as the effect of the proportion of women employees? Formulate the test, making sure you state the critical region of rejection of the null hypothesis. (Assume that the Classical Normal Regression Model holds.) Note that you cannot actually test the hypothesis with the data given. What additional information would you need to carry the test?

Solution Here, we are asked to test the hypothesis $H_0 : \beta_2 = \beta_3$. (Equivalently, $H_0 : \beta_2 - \beta_3 = 0$) For this, I would form the following $t - statistic$:

$$t_0^{\hat{\beta}_2 - \hat{\beta}_3} = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\widehat{se\left(\hat{\beta}_2 - \hat{\beta}_3\right)}}$$

If we had the estimated variance-covariance matrix, it would be possible to form this test statistic. (Actually, all we need is the estimated covariance between $\hat{\beta}_2$ and $\hat{\beta}_3$. Comparing $\left|t_0^{\hat{\beta}_2 - \hat{\beta}_3}\right|$ to $c^*_{0.05,N(0,1)} = 1.96$, I would reject the null hypothesis if $\left|t_0^{\hat{\beta}_2 - \hat{\beta}_3}\right| > 1.96$. Otherwise, I would fail to reject. This gives us the following critical region:

$$C_{0.05}^{H_0:\beta_2 - \beta_3 = 0} = \left\{X : \left|\frac{\hat{\beta}_2 - \hat{\beta}_3}{\widehat{se\left(\hat{\beta}_2 - \hat{\beta}_3\right)}}\right| > 1.96\right\}$$

2. (30 points) Consider the simple regression model:

$$y_i = \beta x_i + \varepsilon_i$$

a. What is the linear (in the $y_i's$) estimator of β with the minimum mean squared error?

Solution Let $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$. For this question, I will assume that the standard assumptions of the OLS model hold. Further, I will assume nonstochastic regressors, so as to simplify the notation. (The solution is the same either way). Here, we want to find

$$\hat{c} = \arg\min_{c} \{MSE(c'y)\}$$

I will use the result that $MSE(c'y) = Var(c'y) + (E[c'y - \beta])^2$:

$$\begin{aligned} MSE(c'y) &= Var(c'y) + (E[c'y - \beta])^2 \\ &= c'Var(y)c + (c'E[y] - \beta)^2 \\ &= c'\left(\sigma^2 I_n\right)c + (c'x\beta - \beta)^2 \\ &= \sigma^2 c'c + \beta^2 (c'x - 1)^2 \\ &= \sigma^2 c'c + \beta^2 (x'c)^2 - 2\beta^2 c'x + \beta^2 \end{aligned}$$

Where I used the fact that $c'x$ is a scalar and hence $c'x = (c'x)' = x'c$.

Taking first order conditions:

$$(c) : 2\sigma^2 \hat{c} + 2\beta^2 x'\hat{c}x - 2\beta^2 x = 0$$

Rearranging,

$$\begin{aligned} 2\sigma^2 \hat{c} + 2\beta^2 \hat{c}'xx - 2\beta^2 x &= 0 \\ \sigma^2 \hat{c} + \beta^2 \hat{c}'xx - \beta^2 x &= 0 \\ \sigma^2 \hat{c} + \beta^2 (\hat{c}'x - 1)x &= 0 \end{aligned} \tag{1}$$

Premultiplying (1) by x', we have:

$$\begin{aligned} 0 &= \sigma^2 x'\hat{c} + \beta^2 (\hat{c}'x - 1)x'x \\ &= \sigma^2 \hat{c}'x + \beta^2 \hat{c}'xx'x - \beta^2 x'x \\ &= \hat{c}'x\left(\sigma^2 + \beta^2 x'x\right) - \beta^2 x'x \end{aligned}$$

And, solving for $\hat{c}'x$

$$\hat{c}'x = \frac{\beta^2 x'x}{\sigma^2 + \beta^2 x'x}$$

Substituting this back into (1) gives us:

$$\begin{aligned} 0 &= \sigma^2 \hat{c} + \beta^2 \left(\frac{\beta^2 x'x}{\sigma^2 + \beta^2 x'x} - \frac{\sigma^2 + \beta^2 x'x}{\sigma^2 + \beta^2 x'x}\right)x \\ &= \sigma^2 \hat{c} - \frac{\beta^2 \sigma^2 x}{\sigma^2 + \beta^2 x'x} \end{aligned}$$

Solving for $\hat{c}$:

$$\hat{c}\sigma^2 = \frac{\beta^2 \sigma^2 x}{\sigma^2 + \beta^2 x'x}$$

Or

$$\hat{c} = \frac{\beta^2 x}{\sigma^2 + \beta^2 x'x}$$

The minimum mean squared error estimator of β is therefore:

$$\hat{\beta} = \hat{c}'y = \frac{\beta^2 x'y}{\sigma^2 + \beta^2 x'x}$$

b. For the estimator in part (a), show that the ratio of its mean squared error to that of the Ordinary Least Squares estimator is

$$\frac{\tau^2}{1+\tau^2} \text{ where } \tau^2 = \frac{\beta^2}{\sigma^2/x'x}$$

Here x is an $n \times 1$ vector, where n is the sample size, and σ^2 is the variance of y_i.

Solution Recall that under the typical OLS assumptions, $Var\left(\hat{\beta}_{OLS}\right) = \sigma^2 \left(x'x\right)^{-1}$, and therefore,

$$\begin{aligned} MSE\left(\hat{\beta}_{OLS}\right) &= Var\left(\hat{\beta}_{OLS}\right) + \left(E\left[\hat{\beta}_{OLS} - \beta\right]\right) \\ &= Var\left(\hat{\beta}_{OLS}\right) = \frac{\sigma^2}{x'x} = \frac{\beta^2}{\tau^2} \end{aligned}$$

Now all that remains is to solve for $MSE\left(\hat{c}'y\right)$:

$$MSE\left(\hat{\beta}\right) = Var\left(\hat{c}'y\right) + \left(E\left[\hat{c}'y - \beta\right]\right)^2$$

Breaking this up into two parts, we have:

$$\begin{aligned} Var\left(\hat{c}'y\right) &= Var\left(\frac{\beta^2 x'y}{\sigma^2 + \beta^2 x'x}\right) = \left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 Var\left(x'y\right) \\ &= \left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 x' Var\left(y\right) x = \left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 x'\left(\sigma^2 I_n\right) x \\ &= \sigma^2 x'x \left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 \end{aligned}$$

And,

$$\begin{aligned} E\left[\hat{c}'y - \beta\right] &= E\left[\frac{\beta^2 x'y}{\sigma^2 + \beta^2 x'x} - \beta\right] \\ &= E\left[\frac{\beta^2 x'\left(x\beta + \varepsilon\right)}{\sigma^2 + \beta^2 x'x} - \frac{\beta\sigma^2 + \beta^3 x'x}{\sigma^2 + \beta^2 x'x}\right] \\ &= E\left[\frac{\beta^3 x'x - \beta\sigma^2 - \beta^3 x'x}{\sigma^2 + \beta^2 x'x} + \frac{\beta^2 x'\varepsilon}{\sigma^2 + \beta^2 x'x}\right] \\ &= \frac{1}{\sigma^2 + \beta^2 x'x} E\left[-\beta\sigma^2 + \beta^2 x'\varepsilon\right] = -\frac{\beta\sigma^2}{\sigma^2 + \beta^2 x'x} \end{aligned}$$

Putting this back together:

$$\begin{aligned}
MSE\left(\hat{\beta}\right) &= Var\left(\hat{c}'y\right) + \left(E\left[\hat{c}'y - \beta\right]\right)^2 \\
&= \sigma^2 x'x\left(\frac{\beta^2}{\sigma^2 + \beta^2 x'x}\right)^2 + \left[-\frac{\beta\sigma^2}{\sigma^2 + \beta^2 x'x}\right]^2 \\
&= \frac{\beta^4\sigma^2 x'x}{\left(\sigma^2 + \beta^2 x'x\right)^2} + \frac{\beta^2\sigma^4}{\left(\sigma^2 + \beta^2 x'x\right)^2} \\
&= \frac{\sigma^2\beta^2\beta^2 x'x + \sigma^4\beta^2}{\left(\sigma^2 + \beta^2 x'x\right)^2}
\end{aligned}$$

Recognizing that $\tau^2 = \frac{\beta^2}{\sigma^2/x'x} \Rightarrow \beta^2 x'x = \tau^2\sigma^2$, we have:

$$\begin{aligned}
MSE\left(\hat{\beta}\right) &= \frac{\sigma^2\beta^2\left(\tau^2\sigma^2\right) + \beta^2\sigma^4}{\left(\sigma^2 + \sigma^2\tau^2\right)^2} \\
&= \frac{\beta^2\sigma^4\left(1+\tau^2\right)}{\sigma^4\left(1+\tau^2\right)^2} = \frac{\beta^2}{1+\tau^2}
\end{aligned}$$

Therefore, we have:

$$\frac{MSE\left(\hat{c}'y\right)}{MSE\left(\hat{\beta}_{OLS}\right)} = \frac{\frac{\beta^2}{1+\tau^2}}{\frac{\beta^2}{\tau^2}} = \frac{\tau^2}{1+\tau^2}$$

3. (30 points) Gasoline sales in a regional market were modeled by the following regression, estimated with quarterly data:

$$\widehat{Q} = 70 - 0.01P + 0.2Y - 1.5S_1 + 3.6S_2 + 4.7S_3$$

where Q is gasoline sales, P is gasoline price, Y is disposable income, and S_j $(j = 1, 2, 3, 4)$ are quarterly (seasonal) dummy variables (i.e. $S_j = 1$ if observation is from the $j'th$ quarter and 0 otherwise). The expected paths of P and Y for the next year are as follows:

Quarter	1	2	3	4
P	110	116	122	114
Y	100	102	104	103

a. (10 points) Calculate the sales of gasoline to be expected in each quarter of next year. How would you calculate standard errors for your predicted values?

Solution Quarter 1

$$\begin{aligned}
\widehat{Q} &= 70 - 0.01P + 0.2Y - 1.5S_1 + 3.6S_2 + 4.7S_3 \\
&= 70 - 0.01\,(110) + 0.2\,(100) - 1.5 = 87.4
\end{aligned}$$

Quarter 2

$$\begin{aligned}
\widehat{Q} &= 70 - 0.01P + 0.2Y - 1.5S_1 + 3.6S_2 + 4.7S_3 \\
&= 70 - 0.01\,(116) + 0.2\,(102) + 3.6 = 92.84
\end{aligned}$$

Quarter 3

$$\begin{aligned}
\widehat{Q} &= 70 - 0.01P + 0.2Y - 1.5S_1 + 3.6S_2 + 4.7S_3 \\
&= 70 - 0.01\,(122) + 0.2\,(104) + 4.7 = 94.28
\end{aligned}$$

Quarter 4

$$\begin{aligned}\widehat{Q} &= 70 - 0.01P + 0.2Y - 1.5S_1 + 3.6S_2 + 4.7S_3 \\ &= 70 - 0.01\,(114) + 0.2\,(103) = 89.46\end{aligned}$$

Here,I used the estimator $\hat{Q}_0 = X_0\hat{\beta}$. How do we calculate the standard errors for these estimates? Under the standard assumptions of OLS:

$$\begin{aligned}Var\left(\hat{Q}_0\middle| X_0, X\right) &= Var\left(X_0\hat{\beta}\middle| X_0, X\right) \\ &= X_0 Var\left(\hat{\beta}\middle| X_0, X\right) X_0' \\ &= X_0\sigma^2\left(X'X\right)^{-1} X_0' \\ &= \sigma^2 X_0\left(X'X\right)^{-1} X_0'\end{aligned}$$

This, of course, is not feasible, so we would use the estimate

$$se\left(\hat{Q}_0\middle| X_0, X\right) = \sqrt{\widehat{Var\left(\hat{Q}_0\middle| X_0, X\right)}} = \sqrt{\hat{\sigma}^2 X_0\left(X'X\right)^{-1} X_0'}$$

Where

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k}$$

b. (10 points) Suppose another researcher proposes to use the same data to estimate an equation of the same form except that she wishes to employ the seasonal dummy variables S_2, S_3, S_4 instead of S_1, S_2, S_3. Write down the equation that will come from her calculations with the appropriate estimated coefficients.

Solution Write the estimated model as:

$$\hat{Q} = \hat{\alpha} + \hat{\beta}_1 P + \hat{\beta}_2 Y + \hat{\gamma}_1 S_1 + \hat{\gamma}_2 S_2 + \hat{\gamma}_3 S_3$$

Since we know that $S_1 + S_2 + S_3 + S_4 = 1$ (this holds $\forall i$, but we have suppressed the indices), we have that $S_1 = 1 - S_2 - S_3 - S_4$, which gives us:

$$\begin{aligned}\hat{Q} &= \hat{\alpha} + \hat{\beta}_1 P + \hat{\beta}_2 Y + \hat{\gamma}_1\left(1 - S_2 - S_3 - S_4\right) + \hat{\gamma}_2 S_2 + \hat{\gamma}_3 S_3 \\ &= \left(\hat{\alpha} + \hat{\gamma}_1\right) + \hat{\beta}_1 P + \hat{\beta}_2 Y + \left(\hat{\gamma}_2 - \hat{\gamma}_1\right) S_2 + \left(\hat{\gamma}_3 - \hat{\gamma}_1\right) S_3 + \hat{\gamma}_1 S_4\end{aligned}$$

Plugging in the estimates, we have:

$$\begin{aligned}\hat{Q} &= (70 - 1.5) - 0.01P + 0.2Y + (3.6 + 1.5)\, S_2 + (4.7 + 1.5)\, S_3 + 1.5S_4 \\ &= 68.5 - 0.01P + 0.2Y + 5.1S_2 + 6.2S_3 + 1.5S_4\end{aligned}$$

c. (10 points) Yet another investigator proposes to use all four seasonal dummies. Write down the estimated equation for this case.

Solution Clearly, since $S_1 + S_2 + S_3 + S_4 = 1$, we would have perfect collinearity if we were to include the constant in addition to the four seasonal dummies. Therefore, for this model, I will not include the intercept term (I will substitute it out):

$$\begin{aligned}\hat{Q} &= \hat{\alpha}\cdot 1 + \hat{\beta}_1 P + \hat{\beta}_2 Y + \hat{\gamma}_1 S_1 + \hat{\gamma}_2 S_2 + \hat{\gamma}_3 S_3 \\ &= \hat{\alpha}\left(S_1 + S_2 + S_3 + S_4\right) + \hat{\beta}_1 P + \hat{\beta}_2 Y + \hat{\gamma}_1 S_1 + \hat{\gamma}_2 S_2 + \hat{\gamma}_3 S_3 \\ &= \hat{\beta}_1 P + \hat{\beta}_2 Y + \left(\hat{\alpha} + \hat{\gamma}_1\right) S_1 + \left(\hat{\alpha} + \hat{\gamma}_2\right) S_2 + \left(\hat{\alpha} + \hat{\gamma}_3\right) S_3 + \hat{\alpha} S_4\end{aligned}$$

Plugging in the estimates, we have:

$$\begin{aligned}\hat{Q} &= -0.01P + 0.2Y + (70 - 1.5)\, S_1 + (70 + 3.6)\, S_2 + (70 + 4.7)\, S_3 + 70S_4 \\ &= -0.01P + 0.2Y + 68.5S_1 + 73.6S_2 + 74.7S_3 + 70S_4\end{aligned}$$