Econ 203B: Single Equation Models

Solutions for 2003 Midterm

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

1. (35 points) You are hired as a political analyst at a TV station. You are asked to present a study about how citizens voted at the last presidential elections. You have access to the percentage of votes per party, for the two parties, at a state level. You also know for each state whether the presidential candidate delivered a speech there within the last month before the elections or not.

a. Describe a (linear) model of states' voting for a party based on the above information and your knowledge of the states' location.

Solution Define the following variables:

$$\begin{aligned}
D_i &= \text{\% of voters in state } i \text{ voting for the Democrat} \\
C_i &= \begin{cases} 1 & \text{if the cand. delivered a speech in state } i \text{ prior to election} \\ 0 & \text{else} \end{cases} \\
N_i &= \begin{cases} 1 & \text{if state } i \text{ is in the North} \\ 0 & \text{else} \end{cases} \\
E_i &= \begin{cases} 1 & \text{if state } i \text{ is in the East} \\ 0 & \text{else} \end{cases} \\
S_i &= \begin{cases} 1 & \text{if state } i \text{ is in the South} \\ 0 & \text{else} \end{cases} \\
W_i &= \begin{cases} 1 & \text{if state } i \text{ is in the West} \\ 0 & \text{else} \end{cases}
\end{aligned}$$

An appropriate model here is:

$$D_i = \beta_1 C_i + \beta_2 N_i + \beta_3 E_i + \beta_4 S_i + \beta_5 W_i + \varepsilon_i \tag{1}$$

b. Describe how you would allow for potentially differential effects of whether the party's candidate campaigned in a state depending on the state's location (east, west, north, south).

Solution Here, I would include interaction terms between the various regions and whether or not the party's candidate campaigned in that state's. That is, I would use the following model:

$$D_i = \beta_1 N_i + \beta_2 E_i + \beta_3 S_i + \beta_4 W_i + \beta_5 N_i \cdot C_i + \beta_6 E_i \cdot C_i + \beta_7 S_i \cdot C_i + \beta_8 W_i \cdot C_i + u_i \tag{2}$$

c. How would you test for the presence of such location effects? Be as precise as possible about the type of test and critical values you would use and why.

Solution If we consider the null hypothesis that there are no such location effects, that is, $H_0 : \beta_5 = \beta_6 = \beta_7 = \beta_8$, or in matrix notation, $H_0 : \Gamma\beta = \gamma_0$ where

$$\Gamma = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \ \gamma_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

we can use the following F statistic:

$$F_0 = \frac{\left(\hat{\varepsilon}_R' \hat{\varepsilon}_R - \hat{\varepsilon}_{UR}' \hat{\varepsilon}_{UR}\right)/p}{\hat{\varepsilon}_{UR}' \hat{\varepsilon}_{UR}/\left(n-k\right)}$$

where $\hat{\varepsilon}_R'\hat{\varepsilon}_R$ is the sum of squared residuals from the estimation of (2) and $\hat{\varepsilon}_{UR}'\hat{\varepsilon}_{UR}$ is the sum of squared residuals from the estimation of (3). Here, $n = 50$, since we only have one observation from each of the 50 states. $k = 8$ since in the unrestricted model, we are estimating 8 parameters. Finally, $p = rank\,(\Gamma) = 3$, since we are testing three linearly independent hypotheses.

If I wanted to test this hypothesis at the 5% significance level, I would compare my test statistic, F_0, to $c^*_{0.05,F(3,42)}$. If $F_0 \leq c^*_{0.05,F(3,42)}$, then I would conclude that the effect of a candidate campaigning in a particular state a month prior to the election did not depend on the region in which the state was located. If $F_0 > c^*_{0.05,F(3,42)}$, then I would conclude that the region of a state mattered.

d. Suppose that in addition you also have access to the results of the previous presidential elections at a state level and you also know for each state whether the parties' leaders campaigned in that state or not during the last month before each election. Describe how you could use this additional information to improve your analysis of parts $(a) - (c)$.

Solution Since the past election presumably had at least one candidate who did not participate in the current election, I would have a difficult time incorporating the information about whether or not that past presidential-hopeful gave influential speeches, but nevertheless, define the following variables:

$$\begin{aligned}
DPast_i &= \text{\% of voters in state } i \text{ voting for the Democrat in the past election} \\
CPastD_i &= \begin{cases} 1 & \text{if in the past election, the Democratic candidate gave a speech in state } i \text{ a month prior to the election} \\ 0 & \text{else} \end{cases} \\
CPastR_i &= \begin{cases} 1 & \text{if in the past election, the Republican candidate gave a speech in state } i \text{ a month prior to the election} \\ 0 & \text{else} \end{cases}
\end{aligned}$$

and consider the following model:

$$\begin{aligned}
D_i &= \beta_1 C_i + \beta_2 DPast_i + \beta_3 CPastD_i \cdot DPast_i + \beta_4 CPastR_i \cdot DPast_i \\
&\quad + \beta_5 N_i + \beta_6 E_i + \beta_7 S_i + \beta_8 W_i + \upsilon_i
\end{aligned}$$

This specification is similar in flavor to (1), but now that we have included $DPast_i$ and $CPast_i \cdot DPast_i$ in the regression, it seems prudent to discuss the meaning of β_2 and β_3.

β_2 gives the effect of past Democrat voting percentages on current Democrat voting percentages. I would guess that $\beta_2 > 0$, since there is probably a relationship between how a state voted in the past and how it voted in recent elections, presumably due to some underlying attitude of the state.

β_3 and β_4 are a little more complicated, since they are interaction terms. Taking derivatives may help provide some intuition:

$$\frac{\partial D_i}{\partial DPast_i} = \beta_2 + \beta_3 CPastD_i + \beta_4 CPastR_i$$

Here, we see that the lingering effect of previous elections on current elections depends both on whether or not Democratic candidates and Republican candidates gave speeches a month prior to the previous election. If I had to venture a guess, I would say that $\beta_3 < 0$ and $\beta_4 > 0$. What would this mean? Presumably, the fact that the Democratic candidate in the past was able to garner the votes he/she did was due in some part to his/her campaigning abilities and maybe some general positive personality traits. Such effects would be over and above the state's typical voting habits. As a result (assuming that neither candidate is an incumbent), we would expect that $\beta_3 < 0$, since this particular election is not likely to be influenced by the amiability of the past candidate. A similarly convoluted interpretation can be attributed to the β_4 coefficient.

This problem is very open ended. As a result, there are many ways in which the new information could have entered into the model.

2. (50 points) Suppose that

$$y_i = x_i^* \beta + \varepsilon_i$$
$$x_i = x_i^* \cdot \nu_i$$
$$z_i = x_i^* \cdot \eta_i$$

where $x_i^*, \varepsilon_i, \nu_i$ and η_i are independent random variables and $(x_i^*, \varepsilon_i, \nu_i, \eta_i)$ is independent and identically distributed over i.

You observe (y_i, x_i, z_i) for $i = 1, 2, \ldots, n$, and you are interested in estimating β.

Answer the following questions under the assumption that all relevant moments exist.

a. (20 points) Under what conditions is the OLS estimator

$$\hat{\beta}_{OLS} = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}$$

consistent for β?

For the remaining questions assume that $E[\nu_i] = E[\eta_i] = 1$. In addition, assume that the conditions of part (b) hold so that $\tilde{\beta}$ is consistent.

b. (20 points) Let α be a fixed real number between 0 and 1. Under what conditions is

$$\tilde{\beta} = \frac{\sum_{i=1}^n (\alpha y_i x_i + (1-\alpha) y_i z_i)}{\sum_{i=1}^n x_i z_i}$$

consistent?

c. (5 points) Derive the limiting distribution of $\sqrt{n}\left(\tilde{\beta} - \beta\right)$, where $\tilde{\beta}$ is defined in (b).

d. (5 points) What is the best choice of α?

3. (15 points) A regression model with $k = 16$ independent variables is fit using a panel of seven years of data. The sums of squares for the seven separate regressions and the pooled regression are shown below. The model with the pooled data allows a separate constant for each year. Test the hypothesis that the same coefficients apply every year.

	1984	1985	1986	1987	1988	1989	1990	All Years
Observations	65	55	87	95	103	87	78	570
ESS $\left(\hat{\varepsilon}'\hat{\varepsilon}\right)$	104	88	206	144	199	308	211	1425

Make sure that you explain the test you perform, derive the distribution of the test statistic under the assumptions of the CNR model and justify the critical value you use.

Solution For this question, I would do an F test on the null hypothesis that the coefficients for all the years are simultaneously equal to zero. The test statistic I would use would be:

$$F_0 = \frac{\hat{\varepsilon}_R'\hat{\varepsilon}_R - \hat{\varepsilon}_{UR}'\hat{\varepsilon}_{UR}}{\hat{\varepsilon}_{UR}'\hat{\varepsilon}_{UR}} \frac{n-k}{p}$$

To see the derivation of this test statistic, refer either to the solutions for the 2000 midterm or to my notes on "The F-Statistic."

Here, the restricted model is the model which pools all the years together. Therefore, we have:

$$\hat{\varepsilon}_R'\hat{\varepsilon}_R = 1425$$

And

$$\hat{\varepsilon}_{UR}'\hat{\varepsilon}_{UR} = 104 + 88 + 206 + 144 + 199 + 308 + 211 = 1260$$

To get a good feel for the number of restrictions we are imposing in this model, it is useful to write out both the restricted and the unrestricted models:

$$Y = \begin{bmatrix} X_{1984} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_{1990} \end{bmatrix} \begin{bmatrix} \beta_{1984} \\ \vdots \\ \beta_{1990} \end{bmatrix} + \varepsilon \tag{1}$$

And (pooling all the data together)

$$Y = X\beta + u \tag{2}$$

Here, (1) is the unrestricted model and (2) is the restricted model. The restriction we are imposing is $\beta_{1984} = \cdots = \beta_{1990}$ and since each β is a 16×1 vector (there are 16 independent variables), it follows that we are imposing $6 \times 16 = 96$ restrictions. How many parameters are we estimating in the unrestricted model? We are estimating a set of 16 parameters for each of the years in our regression. That is, $k = 7 \times 16 = 112$. Finally, we know that $n = 570$. Therefore, we have:

$$F_0 = \frac{1425 - 1260}{1260}\frac{458}{96} = 0.6248$$

Using the critical value $c^*_{0.05,F(96,458)} = 1.28$, we have that $F_0 \leq c^*_{0.05,F(96,458)}$ and we conclude that the same coefficients do apply each year. Alternatively, we know that as $n \to \infty$, $pF_0 \xrightarrow{d} \chi^2(p)$ and we could have used the test statistic

$$pF_0 = 96 \cdot 0.6248 = 59.981$$

And compared it to the critical value $c^*_{0.05,\chi^2(96)} = 119.89$, of course reaching the same conclusion.