ECNS561, Fall 2011

Take-home Graded Practice Opportunity 5

Due date: November 29, 2011, 5:00 p.m.

Practice with testing restrictions

1. You are working with the function: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$. Variables $\{X_1, X_2, X_3\}$ are continuous and X_4 is a categorical variable. After estimating the model, you retrieve the following results:

β_0	β_1	β_2	β_3	β_4
-2.47	1.82	0.57	0.19	-3.32

R^2	$\hat{\varepsilon}'\hat{\varepsilon}$	n
0.91	8.97	9

$$(\boldsymbol{X'X})^{-1} = \begin{bmatrix} 8.82 & -0.65 & -0.55 & -0.38 & -1.21 \\ -0.65 & 0.06 & 0.05 & 0.02 & 0.06 \\ -0.55 & 0.05 & 0.05 & 0.01 & 0.02 \\ -0.38 & 0.02 & 0.01 & 0.05 & 0.12 \\ -1.21 & 0.06 & 0.02 & 0.12 & 0.75 \end{bmatrix}$$

   (a) Set up and test the joint hypothesis that $X_1 = 0$ and $X_2 = 0$.

   (b) Set up and test the joint hypothesis that $X_2 = X_3$ and $X4 = -3$.

Putting it all together

You are a consultant working for a major airport. You are tasked with determining manners in which to decrease the average waiting time for travelers in airports. To do so, you decide to perform an empirical analysis of waiting times at three other major airports. The U.S. Customs and Border Protection agency posts these times for a number of airports, and these data can be accessed at:

http://apps.cbp.gov/awt/index.asp

(use the "Date range average" option and retrieve data across the entire available time period).

You decide to randomly select the Hartsfield-Jackson Atlanta International Airport (ATL), the Chicago O'Hare Airport (ORD), and the George Bush Intercontinental/Houston Airport (IAH). After looking at the data, you decide that instead of dealing with "Time of Day" ranges, you will simply round up to the nearest hour. For example, the range "Midnight to 1 A.M." will be "Hour = 1;" the range "1 A.M. to 2 A.M." will be "Hour = 2." Furthermore, you believe that there are distinct differences in the airports and decide to control for these differences using categorical variables. Your initial model is therefore:

$$Avg.\ wait\ time = \beta_0 + \beta_1 Hour + \beta_2 Passengers + \beta_3 Booths + \sum_i \beta_{ji} Airport + \varepsilon$$

where $Hour$ represents the hour of day, $Passengers$ is the average number of arriving passengers, $Booths$ is the number of open service booths, and $Airport$ is a categorical variable representing the airport.

Your goals are the following:

1. Collect, clean, and organize the data.
2. Import the data into your favorite statistical analysis software and explore it using visual and statistical summaries. Provide a brief interpretation of the output and potential economic and econometric implications.
3. Estimate the model proposed above. Provide economic interpretation of the results.
4. Show that there is multicollinearity in the above model. What are the economic reasons for the multicollinearity? How would you attempt to deal with the multicollinearity?
5. One potential manner in which to reduce waiting times is to increase the number of service booths during peak hours. How would you test whether this would be an effective strategy? Re-estimate the model (after dealing with multicollinearity) that can be used to answer the question. Interpret the results.

6. Lastly, you wish to know whether separate models must be used to estimate average waiting times for off-peak and peak hours. That is, you believe that there are structural differences in waiting times occurring between 22:00 (10 p.m.) and 3:00 (3 a.m.), and waiting times occurring between 4:00 (4 a.m.) and 21:00 (9 p.m.). Test this hypothesis and provide an economic interpretation of the findings.

7. What final recommendations can you make to the airport?