EXERCISE 7

This exercise considers models of the starting salaries for graduates from law schools. The data set used is a cross-section of law schools given in LAWSCH85.txt with the variable descriptions in LAWSCH85.des. These files are available on the course website as is sample Matlab code for OLS estimation and calculation of multicollinearity specific statistics. The data are taken from Green and have been edited to eliminate observations with missing data. There are still missing data for the AGE variables.

In making hiring and starting salary decisions law-firms seek to predict the quality of the candidate. This quality can depend on the natural abilities of the candidate and the quality of the education received.

1. We first consider the quality of the education received. An argument can be made that the cost paid by law-school graduates is a good market-based measure of the value-added by the educational experience at that school. Accordingly, we start with the regression of the log of the starting salary $s_i$ on the log of the cost of the degree $c_i$:

$$s_i = \alpha + \beta c_i + u_i$$  \hspace{1cm} (2)

where $u_i$ is an unknown disturbance term. Run an OLS regression and report your results.

(a) Why do you think the regression was specified in terms of logs? What is the interpretation of $\hat{\beta}$? Test the null hypothesis that $\hat{\beta}$ is zero?

An alternative model of quality might be

$$s_i = \alpha + \beta r_i + u_i$$  \hspace{1cm} (3)

where $r_i$ is the ranking of the law school. Run an OLS regression based on this specification and report your results.

(b) What is the interpretation of $\hat{\beta}$ for this regression? Test the null hypothesis that $\hat{\beta}$ is zero?

Consider the combined model

$$s_i = \alpha + \beta c_i + \gamma r_i + u_i.$$  \hspace{1cm} (4)

Run a regression on this model and report your results.

(c) How might we choose between these three models?

(d) Suppose coefficient estimates on one of the explanatory variables in (3) is insignificant. Discuss how collinearity could cause this to happen. Is there in fact, any evidence that collinearity is causing a problem?
2. A strong argument can be made that the cardinal measure $r_i$ is not the appropriate way to introduce rankings into the model. There is no reason to believe that the effect of a school being ranked 41 instead of 40 is the same as being ranked 11 instead of 10. Accordingly we consider the specification

$$s_i = \alpha + \beta c_i + \gamma_1 r_1 i + \gamma_2 r_2 i + \gamma_3 r_3 i + \gamma_4 r_4 i + u_i$$  \(5\)

where $(r_1 i, r_3 i, r_3 i, r_4 i)$ are the dummy variables for ranking 1-10, 11-25, 25-40, and 41-60, respectively. Run the indicated regression and report your results.

(a) What is the interpretation for the coefficient estimate $\hat{\gamma}_1$? How can we estimate the combined impact of being in the top-sixty vs. not?

(b) How might you choose between (3) and (4)?

(c) Should $c_i$ be included in this regression?

(d) Test the joint null hypothesis that $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$.

3. Consider now additional covariates that might measure the intrinsic quality of the candidate rather than the value added of the school. Accordingly, we consider

$$s_i = \alpha + \beta c_i + \gamma_1 r_1 i + \gamma_2 r_2 i + \gamma_3 r_3 i + \gamma_4 r_4 i + \delta_1 L_i + \delta_2 G_i + u_i$$

where $L_i$ is the median LSAT score for the school and $G_i$ is the median college GPA. Run the indicated regression and report your results.

(a) Test the null hypotheses that $\gamma_1 = 0$ and $\gamma_2 = 0$ individually and jointly. Which should be the more powerful test?

(b) Does multicollinearity seem to be a problem for any of these estimates?

We now introduce some additional covariates that might be indicative of the quality of the education received at the school:

$$s_i = \alpha + \beta c_i + \gamma_1 r_1 i + \gamma_2 r_2 i + \gamma_3 r_3 i + \gamma_4 r_4 i + \delta_1 L_i + \delta_2 G_i + \delta_3 F_i + \delta_4 A_i + \delta_5 Z_i + \delta_6 L_i + u_i$$

where $F_i$ is number of faculty at the school, $A_i$ is age of the school, in years, $Z_i$ is size of the entering class, and $L_i$ is the log of the number of volumes in the library. Run the indicated regression and report your results.

(d) Test all the $\delta$ coefficients being zero separately and jointly. Does collinearity seem to be a problem for this regression?