PLSC 503: Solutions to Problem Set 7

Thad Dunning

Department of Political Science

Yale University

Spring 2010

1 Theoretical/Conceptual Exercises

Question 1: Positive Definite Matrices. Read section 3.4 in Freedman (2009). Work Exercise Set D in Chapter 3, Problems 1-8. (Turn in your work for exercises 5-6).

Solution: See solutions in Freedman (2009).

Question 2: Work exercise sets A and B in Chapter 8. (You do not need to turn this in).

Solution: See solutions in Freedman (2009); we worked exercises A.1 and A.2 in class.

Question 3: In equation (6) on p. 167, there is no intercept α, only intercepts a_j for each region. Why not? What would happen if you put an intercept α into the equation? Could you reformulate the model so that it has both an intercept α and regional intercepts a_j for some j?

Solution: With an intercept α in equation (6) on p. 167 of Freedman (2009), the design matrix would not have full rank: the column of 1's would be a linear combination of the columns of 1's and 0's that correspond to each α_j. If you want to include α, then leave out one of the α_j's. In this case, α is equivalent to the omitted α_j.

Question 3: On p. 175, Freedman (2009) explains that White's method for estimating the SEs in OLS (what Beck and Katz (1995) call "panel-corrected standard errors" in the time-series cross-section context) "may have the same sort of problems as plug-in SEs, because estimated covariance matrices can be quite unstable."

Can you give an explain why the estimated covariance matrix would be unstable in the settings discussed by Beck and Katz. How this would affect panel-correct standard errors? (Hint: look at p. 638 of Beck and Katz. Where does the covariance matrix of the errors appear in the formula for the covariance matrix of $\hat{\beta}$, and how is it estimated?).

Solution: Note that the variance-covariance matrix of the OLS estimator, when $\text{cov}(\epsilon|X) = G$, is

$$\text{cov}(\hat{\beta}_{OLS}|X) = (X'X)^{-1}X'GX(X'X)^{-1}. \tag{1}$$

(See Freedman 2009: p. 63, equation 8). Then, the White-like standard errors (known as "panel-corrected standard errors" in the time-series cross-section context) are given by (1), with G replaced by $\hat{G}$. That is,

$$\widehat{\text{cov}}(\hat{\beta}_{OLS}|X) = (X'X)^{-1}X'\hat{G}X(X'X)^{-1}. \tag{2}$$

Here, the typical element of $\hat{G}$ is obtained from manipulation of the OLS residuals. For example, if we assume that G has contemporaneously correlated elements (and that this contemporaneous correlation is constant over time), we can estimate the typical element $\text{cov}(\epsilon_{i,t}, \epsilon_{j,t})$ consistently as

$$\widehat{\text{cov}}(\epsilon_{i,t}, \epsilon_{j,t}) = \frac{1}{T}\sum_{i=1}^{T} e_{i,t}e_{j,t}, \tag{3}$$

where $e_{i,t}$ is the residual from the OLS fit for unit i at time t, and T is the number of time periods.

Now, notice that if there are many units, then there are many contemporaneous covariances (correlations) of the errors to estimate. Namely, there are $N(N-1)/2$ contemporaneous covariances, where N is the number of units. This can be a large number, and if T is small, the estimates will be very imprecise, and possibly biased downwards in some cases. (Remember, $\hat{G}$ is a non-linear estimator for G, due to the multiplication in equation 3).

This is why the estimated variance-covariance matrix used to find the panel-corrected standard errors can be unstable, just as it is for fGLS.

2 Computer exercises

Question a: Bootstrap. For this lab, you will bootstrap a regression model on time-series cross-section data, producing a table like 8.1 in Freedman (2009).

The first trick is to find data. Your group should try to find time-series cross-section data from a published paper that uses fGLS, ideally one in which the lagged dependent variable appears as a right-hand side variable (though this is not required). One place to start is with some of the papers mentioned in Beck and Katz (1995); you could see if the data are archived at the ISPS. If you are having trouble finding data, please see me or Mario.

Once you have identified a published paper and found the associated dataset, you should email the paper to me. If more than one regression equation appears in the paper, tell me which equation you will bootstrap following the procedure outlined in sections 8.1 and 8.2 of Freedman (2009).

You should write code that will allow you to produce a table like 8.1 for the application you have chosen. Turn in both your code and your table. For the table, you can concentrate on Columns (A)-(D); do not worry about columns (E) or (F).

Solution: Code will be posted on the class website (on the classes v2 server).