17.874
Spring, 2004
Problem Set 6

1. The dataset house.dta (also house.xls) is distributed to you electronically. It has the following variables.

dvs	Democratic Vote Share
dp	Democratic Party Strength
rp	Republican Party Strength
dp88	Democratic Presidential Vote 1988
dp96	Democratic Presidential Vote 1996
income1	Median Income
renter	Percent Renters
medrent	Rental Price
inc	Incumbent (-1 Republican, +1 Democrat)
gap1i	The difference between the Democrat and the Republican on an ideological scale.
midpti	The midpoint between the Democrat and the Republican on an ideological scale.

a. Regress Democratic vote share on the other variables. (Note there are redundancies and inconsistencies. Keep only cases for which the same district appears twice. Cases where district = -1 are Senate races. Drop these.) Report your results. Write a two-page description of your findings, noting statistical significance of results and substantive interpretation.
b. Regress Democratic vote share on everything but the gap1i and midpti. How does the exclusion of these variables affect the estimated incumbency advantage?
c. Drop income, renter, and median rent. How does the exclusion of these variables affect the other estimates?
d. Find the best model of the data. Justify your choice using test statistics (F-tests). Also look at plots. Do the residuals look okay, or do you see problems?
e. One hypothesis in the Congressional elections literature is that the closer the election, the more converged the candidates will be. Create an interaction of gap1i with a measure of closeness (use presidential vote). Include this interaction variable as well as the other variables. Is there evidence of an interaction?
f. A lot of races do not have measures of ideology for the candidates -- i.e., midpti and gap1i are missing. Does it look like this is a problem? Are the missing cases representative or do they differ from the included cases?

2. Greene, Chapter 6, exercises 1 and 2.