POL 603 Midterm

1. Suppose that x and y are scalars, $x \sim N(0, 3^2)$, $y \sim N(0, 2^2)$, and x and y are independent. In all cases below, justify your answer.

(a) What is the mean of $2x + 3y$?

(b) What is the variance of $2x + 3y$?

(c) What distribution does $2x + 3y$ have?

(d) What distribution does $(x/3)^2 + (y/2)^2$ have?

(e) For which of (a)-(d) would your answer change if x and y were not independent.

2. Present a heuristic proof for the consistency of the OLS estimator (i.e. show that

$\hat{\beta}_{OLS} \xrightarrow{prob.} \beta_0$). Hint: Start with $\hat{\beta}_{OLS} = \beta_0 + \left[\frac{1}{N} \sum_{n=1}^{N} x_n x_n ' \right]^{-1} \left[\frac{1}{N} \sum_{n=1}^{N} x_n \varepsilon_n \right]$, apply the law of large numbers twice, and apply the continuous mapping theorem to combine these results. You may assume that $x_n x_n '$ and $x_n \varepsilon_n$ are iid and that they have sufficient moments in order to allow for application of the WLLN (i.e. don't worry about the technical details, just present the basic argument).

3. Comment on each of the following assumptions we made when deriving the properties of the linear model with stochastic X's. Which properties are very likely to hold and which are restrictive and are made for tractability.

(B1) $y_n = \beta_0 ' x_n + \varepsilon_n$

(B2) $E[\varepsilon_n | x_n] = 0$

(B3) (x_n, ε_n) are identically and independently distributed

(B4) $E\|x_n x_n'\|_1 < \infty$ and $E\|x_n \varepsilon_n\|_1 < \infty$

(B5) $E[x_n x_n']$ is positive definite

(B6) $Var(x_n \varepsilon_n)$ is finite

(B7) $E[\varepsilon_n{}^2 \mid x_n] = \sigma_{\varepsilon 0}^2$

(B8) ε_n are normally distributed

4. Consider the data file `veto.dta` available on the course web site (https://sites.google.com/a/stonybrook.edu/mperess/teaching/classdata/veto.dta). This file contains data on the number of presidential vetoes exercised by the president in each congress. The independent variables include:

`pres_party_h`: the percentage of House members from the President's party

`pres_party_s`: the percentage of Senate members from the President's party

`repub`: 1 for a Republican president and 0 for a Democratic president

`cham_diff`: a measure of the ideological differences between the chambers, calculated as $|H_n - S_n|$ where H_n is the median DW-Nominate Common Space score in the House and S_n is the median DW-Nominate Common Space score in the Senate. If $|H_n - S_n| = 0$, then the position of the median in the House and Senate is the same. If $|H_n - S_n| = .5$, then the House median is .5 units to the right or .5 units to the left of the Senate median. The scale for the DW-Nominate Common Space scores is between -1 and 1.

`party_diff_avg`: the difference between the median DW-Nominate Common Space scores for the Democrats and Republicans, averaged across the chambers. That is, suppose that the median House Democrat is -.3, the median House Republican is .4, the median Senate Democrat is -.7, and the median Senate Republican is .5. The party difference would be .4-(-.3)=.7 in the House and .5-(-.7)=1.2 in the Senate, so `party_diff_avg`=(.7+1.2)/2=1.05.

`start`: 1 for the first two years of the presidential term and 0 for the last 2 years.

(i) Run a regression with `veto` as the dependent variable and `pres_party_h`, `pres_party_s`, `repub`, `cham_diff`, `party_diff_avg`, and `start` as independent variables. Report the results (coefficient estimates and standard errors).

(ii) Interpret the constant term.

(iii) Interpret the effect of each of the independent variables on the dependent variable (and comment on statistical significance).

(iv) Interpret the R-squared of the regression.

(v) Interpret the F-statistic reported in the regression table (make sure to state the null and alternative hypotheses formally).

(vi) Test the joint hypothesis that the population coefficients on `pres_party_h` and `pres_party_s` are zero (make sure to state the null and alternative hypotheses formally).

(vii) Is the assumption of homoskedasticity reasonable here? Provide the results of a formal hypothesis test.

(viii) Is the assumption of normal error terms reasonable here? Provide the results of a formal hypothesis test.

(ix) Considering what you found in (vii) and (viii), should robust standard errors be used here? Why or why not?

5. Under what conditions is OLS efficient in small samples? Under what conditions is OLS efficient in large samples?