Exercise 3
QuantII
Due: Feb. 11

1. tab9s.dta is a Stata data set based on congressional elections in 1992; it has all elections where the incumbent ran for reelection. The depvar is cv92, the challenger's vote in the 1992 election. the iv's are all labelled in obvious ways, choose 2 iv's (plus a constant, of course).

a. Run a regression of cv92 on your two variables and a constant. Note the output of interest (what to note will be clear when you do the rest of the exercise.)

a' - Just to make sure you can use MATA in this simple example, rerun a in MATA and make sure you can reproduce the relevant results.

b. (This is not just an exercise in whether you can copy from the notes, I trust you already have good handwriting, so the only way you are going to learn something is to make sure you can do this yourself - thus you should not do in a group, but of course you may ask for help from others, or check notes if needed). Write down the likelihood for a regression problem. You should do this in scalar notation for a model with constant and 2 iv's, then in straight matrix notation.

c. From b, take the log likelihood (scalar and vector). What interesting properties of estimating linear models do you observe?

d. Now using the log likelihood for a regression (independent observations) use Mata to estimate the same model via ml.

This code uses numerical derivatives and Hessians. Run it and compare your results with a. What is identical? What is close but not identical? Why? (Does identical mean identical to 7 decimal places? Probably not. But it means more than to 1 significant digit.)

e. What is the gradient for the ols loglikelihood function? What is the Hessian? Take its expectation and then inverse. What does this tell you? Show that the negative of the expectation of the outer product of the gradient equals the negative of the information matrix (remember this is done at the true parm values, so $E(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) = E(\boldsymbol{\epsilon}'\boldsymbol{\epsilon}) = N\sigma^2$ and $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2 I$ (I is the identity matrix, NOT the information matrix).

f. (In general we will end with something substantive). In less than one paragraph, what does your regression tell you about what impacts challenger vote? So what is the effect of a unit change of an iv? don't forget your estimate is uncertain, so report the uncertainty.

2. Exercises on elementary ml/logit/probit. This exercises use a small subset of 1992 American National Election Study data. The codebook is in nes92.codebook.pdf.

The 1992 election had three candidates. For part of this exercise we only want a binary DV, so I have created nes92nomissclb.dta which only has people who reported voting for Clinton or Bush. The dv is v1 (Bush/Clinton) where Bush is scored 0, Clinton 1.

Missing data is a critical issue, and one we may deal with towards the end of the course. For now we are going to take the worst way (almost) out and just eliminate all observations with missing data. The nes92nomissclb.dta has no missing data.

a. Using Mata commands , fit a model with 3 iv's to predict Clinton vote. Make sure your results are same as those from a probit command in Stata.

b. Now rewrite your code to do a logit. Again compare your ml results to the Stata logit command (they should be identical).

c. What results change between probit and logit? what do not (very much)?

d. Use Stata to compute the P(Clinton vote) in probit and logit models. Compare the probs. (Do this using the spreadsheet approach, that is, use the formula to compute the predicted probabilities. Then compare the results to what you get from the Stata predict command and make sure they are the same.

e. Use Stata to draw a graph of the response of the probability of voting for Clinton against some continuous variable of interest (obviously setting the value of other variables at some place of interest; if you did not do logit with a continuous iv, just do this now using the logit command, no need to redo the ml code for this).

f We noted that any cdf would yield an admissible binary dv model. The next most common after logit/probit is cloglog (complementary log log if you exponential both sides). It is nice because is not symmetric (as we discuss in class). The p function for the cloglog is p = 1-exp(-exp(xb)), the 1- makes it complementary. Change the likelihood fn to use the clog. Check your results with Stata. What changes? what doesn't change much?

g. Take your probit code from a. Instead of a standard normal, allow the variance to vary, that is, make σ^2 a paramet'er. What happens. Why?

h. Take your probit code. Drop the constant term and make the threshold a parameter (so errors are $N(\mu, 1)$. Estimate that model. What happnes.