Mick Jagger
AREC 815
Problem Set 0
Fall 2016

Problem Set 0 is due by the end of the day on September 6.

1. **Getting Started.** To get started with this problem set, create a `PS0` subfolder in the `AREC815` shared folder. Copy all the materials from the `PS0` subfolder into your new folder. Please make sure that you don't move the contents of the `PS0` folder into your folder or over-write the contents of the `PS0` folder.

   Once you've done this, open the `arec815ps0template.tex` file in WinEdt. Hit the PDFTexify button (with the lightning bolt on it) to compile the document. This should work (since you haven't changed the template yet). If it doesn't, please let both me and Jeff Cunningham know immediately. Now that you know that your LaTeX file actually compiles, look at it in WinEdt. I've included a bunch of comments to help you find your way around if you've never used LaTeX before. Replace the name of the world's most famous economics student with your name, and make sure the document still compiles. You can also go ahead and delete the lines of code indicating when the problem set is due. In general, if you recompile after every step, you shouldn't be able to get yourself into too much LaTeX trouble.

2. **Summary statistics.** Now open Stata, and open the `ps0code` do file in the do file editor. Change the file path so that Stata sets the working directory to your personal folder.

   One of the first things the do file does is open the data set `arec815ps0data`. This is a data set on 2,451 rural households from 26 villages in Kenya. We'll be looking at this data in this problem set.

   As you can see, the do file calls up a Stata program that generates summary statistics and outputs them directly to LaTeX. Right now, the program only reports summary statistics for the female and age variables. As you should see when you look at the Stata code, the current version of the program reports the mean, SD, median, and N for each variable, and allows you to specify what text should describe each.

   The four lines of LaTeX code below the

   `\begin{tabular}`

   command (in the LaTeXdocument) are commented out. Uncomment them to include an appropriate header for your table. Notice that the first line of the table specifies how many columns the table has, and the centering within each column. Each subsequent row must contain that many columns, separated by the "&" symbol, and a pair of backslashes at the end of the row. You can use the multicolumn command to merge columns within a row, as in the row indicating the title for the table.

Female	0.62	0.49	1	2451
Age	36.69	14.24	34	2451

Modify the table to report more of the summary statistics from the data set. You can report those that interest you, but be sure to include the variables yearschool, employed, farmhousehold, and fertilizer variables, which we'll be using later.

Once you've got your expanded summary statistics table compiling, it is time to modify it. Revise the Stata code to include two additional columns reporting the min and max for each variable in your table. You'll also need to modify the LaTeXcode to accommodate a table that is seven, rather than five, columns wide.

Discuss a few things that stand out about the summary statistics.

3. **Regressions in Stata.** The Stata do file `ps0code.do` also calls up the `ps0-regressions` program, which executes OLS regressions of the probability of employment on the female, age, and yearschool variables. The latter regression uses heteroskedasticity-robust standard errors. The code uses the Stata commands `est2vec` and `est2texOWO`, which allow you to save regression results formatted as a LaTeX table. The `est2texOWO` command is a modified version of the `est2tex` command, which does the same things, but with some bugs and weaknesses. The Stata code only outputs the main contents of the table; you have to include the header and footer rows yourself (but this means you can have several different panels in any given table).

Modify the table by interacting female with age and education and including these new variables. Create one regression without robust standard errors, and one with them. To output the coefficients of the new variables to LaTeX, you'll need to add them to the "vars" list in the first `est2vec` command. Notice that labeling the variables in Stata causes the table to use the labels rather than the variable names (unless your variable name includes an underscore). To recompile the LaTeX code using the new Stata results, you'll need to modify the first line of the table to include four centered regression output columns (instead of two).

Add two additional regression specifications to the table which estimate the same model as in the 3rd and 4th columns, but in a probit framework. Add proper header and footer rows to your table that explain what is going on in each column, what the dependent variable is, etc. You can also use the code below to add notes at the end of your table:

```
\multicolumn{7}{p{10.0cm}}{\footnotesize{Robust standard errors in parentheses
in even-numbered columns.   $^{***}$ indicates significance at the 99 percent
level; $^{**}$ indicates significance at the 95 percent level; and $^{*}$
indicates significance at the 90 percent level.}}\\
```

You'll need to modify the width (in centimeters) to fit your table. Discuss the results presented in your table.

Female	-0.0794***	-0.0794***
	(0.0119)	(0.0141)
Age	-0.0012***	-0.0012***
	(0.0004)	(0.0004)
Years of formal schooling	0.0011	0.0011
	(0.0017)	(0.0021)
Constant	0.1676***	0.1676***
	(0.0254)	(0.0314)
Observations	2451	2451
R^2	0.0228	0.0228

4. **Regressions in MATLAB.**

The last line in the Stata do file exports the variables used in the regressions in comma-delimited format, which can be easily read into MATLAB. You'll (eventually) need to modify this to also export your new interaction terms in order to complete the next task, which is to replicate the 3rd and 4th columns of your regression table using MATLAB.

(a) The program `ps0ols.m` replicates the results reported in the first column of the Stata regression table in MATLAB, making use of the relevant formulas:

$$\hat{\beta}_{ols} = \left(\mathbf{X}'\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{y}$$

and

$$V\left[\hat{\beta}_{ols} \right] = \left(\frac{\sum_i \hat{u}_i^2}{N - K} \right) \left(\mathbf{X}'\mathbf{X} \right)^{-1}.$$

Open MATLAB and run `ps0ols.m` to confirm that the coefficients and standard errors match those reported in the main regression table. Read through the program to make sure that you understand what is happening, particularly the calculation of the standard errors. Then, modify the Stata and matlab code so that you replicate the 3rd regression specification included in the main table (the one including the interaction terms). Once you've confirmed that your MATLAB code is generating the correct coefficients and standard errors, paste the results into the space below.

(b) Of course, we would almost never report non-robust standard errors, so we need to make sure our MATLAB code can replicate Stata's output when we use the `robust` option. To generate robust standard errors in MATLAB, we use the formula for sandwich estimator of the variance of $\hat{\beta}_{ols}$:

$$V\left[\hat{\beta}_{ols} \right] = \left(\mathbf{X}'\mathbf{X} \right)^{-1} \mathbf{X}'\hat{\mathbf{\Omega}}\mathbf{X} \left(\mathbf{X}'\mathbf{X} \right)^{-1}$$

where $\hat{\mathbf{\Omega}} = \text{Diag}\left[\hat{u}_i^2 \right]$ normalized by $N/(N-K)$. The program `ps0robustols.m` reproduces the robust standard errors in MATLAB.

Go through the calculation of the robust standard errors in MATLAB and make sure you understand the code. Then modify the program as necessary to report the results of the 4th specification reported in the main table, which includes the interaction terms. Confirm that your Stata and MATLAB results line up, and then paste the MATLAB output below.

5. The MATLAB program `ps0probit.m` estimates a probit model of employment as a function of age and gender (using the same data set). Run the program and confirm that the coefficients and standard errors match the coefficients you'd get using Stata. Then modify the MATLAB code to include the yearschool variable. Report your results below (after confirming that they line up with the results that you would get using Stata.