**Econ 637, Spring 2016**

**John Vandivier**

**Problem Set 6**

**Due: May 3, 2016**

**Endogeneity Issues**

Age and agesq are endogenous, however they are intended to measure different effects. Age and gen20\* are endogenous because a young person may not have been previously eligible to vote. Education and age are endogenous because a young person may not have finished higher education and because in the past there was lower participation in higher education. The gender life expectancy gap indicates endogeneity between male and age, and gender also correlates with education.

**Model Regressed**

The initial long model was of the form:

reg vote anylive age agesq treatment lafollow gen20\* i.state i.education i.group income lnincome incomesq male, robust

The final long model was of the form:

reg vote age agesq lafollow gen20\* i.state i.education male, robust

The difference in the long models is due to the omission of statistically insignificant variables. Various short and intermediate models were also checked for variables robustness.

agesq was a generated variables which is the square of age. It appears age had a significant marginal effect. Income was also checked individually as income, income squared, and the log of income, but these variables were insignificant.

Certain variables were omitted from the model or dropped from the set. gen2014 was dropped as the same data is included in vote and by dropping this variable I was able to leverage the gen20\* syntax for the right hand of the equation. Precinct was omitted because it resulted in an r(103) error. Anylive, livefollow, and answerfollow were omitted because these variables were captured in lafollow, which was used. The gender variable was omitted because the male variable was used.

**Results**

One clear result is the significance of the lafollow variable, which indicates whether an individual had either a live or answering machine pickup for at least one of the follow-up calls. The result is also clear that the treatment was insignificant, as well as the number of calls as indicated by the group variable.

I think we can conclude that holistically robocalls did have a significant effect, when we include both the treatment and the follow-up calls.

**Tabled Regressions**

State, gen20\*, education, and group were omitted due to the number of categorical variables. State, education, and gen20\* are significant while group is not.

