Assignment 1

1. Consider the data file `civwar.dta` available on the course web site. We will consider the dependent variable *war* which is 1 if there is a civil war in the country and year of the observation and 0 otherwise, and we will model this dependent variable using the linear probability model. Consider the following independent variables:

lpop – log of population

polity2 – a measure of the level of democracy in the country, with -10 denoting the authoritarian end of the scale and 10 denoting the democratic end of the scale

gdpen – GDP per capita

western – dummy variable for Western Europe

eeurop – dummy variable for Eastern Europe

lamerica – dummy variable for Latin America

ssafrica – dummy variable for Southern Africa

asia – dummy variable for Asia

nafrme – dummy variable for Northern Africa and Middle East

colbrit – dummy variable for former British colony

colfra – dummy variable for former French colony

mtnest – percent of mountainous terrain

muslim – percent of Muslim population

(i) Run a linear probability model with civil war as the dependent variable, controlling for log of population, democracy score, GDP per capita, region, former colonial status, mountainous terrain, and percent muslin. Interpret each of the estimated coefficients.

(ii) Are classical or robust standard errors more appropriate here?

(iii) Predict the probability of civil war in a country with 5 million people, a Democracy score of 5, GDP per capita of 5,000, in Asia, which is not a former British or French colony, which is 30% mountainous and 40% Muslim.

(iv) Are there any drawbacks to the linear probability model when applied to this data set?

(v) Run a logit model with civil war as the dependent variable, controlling for log of population, democracy score, GDP per capita, region, former colonial status, mountainous terrain, and percent muslin. Interpret each of the estimated coefficients.

(vi) Report the odds ratios and standard errors for each of the independent variables and interpret the results.

(vii) Report substantive effect sizes and standard errors for each of the independent variables and interpret the results.

(viii) Report the McFadden and McKelvey/Zavoina pseudo-R-squares and interpret the results.

(ix) Suppose you believe that extremely authoritarian and extremely Democratic countries are less likely to experience civil war than countries at neither end of the continuum. Provide an appropriate regression and provide a formal hypothesis test relevant to this claim.

2. Use the delta method to derive a formula for the standard error of an odds ratio, e^{β_k}.

(Hint: define the function $c(\beta) = e^{\beta_k}$, take the derivative of $c(\beta)$, and use the formula,

$\sqrt{N}(c(\hat{\beta}) - c(\beta_0)) \xrightarrow{dist,} N(0, c_\beta(\beta_0) V c_\beta(\beta_0)')$ where V is the variance-covariance matrix of $\hat{\beta}$ and c_β is the derivative of c.

3. Consider the latent variable formulation for the logit model, $y_n = \begin{cases} 1, & y_n^* \geq 0 \\ 0, & y_n^* < 0 \end{cases}$ where $y_n^* = \beta' x_n + \varepsilon_n$ and $\varepsilon_n \sim Logistic(0,1)$. Demonstrate that $\Pr(y_n = 1 | x_n; \beta) = \frac{e^{\beta' x_n}}{1 + e^{\beta' x_n}}$ using the fact that the CDF for the Logistic(0,1) distribution is given by $F_X(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x}$.