Assignment 4

1. Consider the data file `wages_full_time.dta` available on the course web site. The data file contains a sample of full time workers along with their hourly wage, age, school years (i.e. level of education), and gender.

(i)   Run a regression with log-wage as the dependent variable and male as the independent variable. Interpret the results. Do there appear to be differences in the male and female wages? Why might this not be evidence for wage discrimination against women?

(ii)   Control for age in the analysis using a quadratic specification for age. Interpret the results (including providing the age at which log-wage is expected to be the highest or lowest).

(iii)   What happens to the coefficient on male when age is included in the specification? Why does this change occur?

(iv)   Control for age in the analysis using a series of dummy variables (use the age groups 18-34, 35-44, 45-64, and 65+). What can be said about the effect of age on wages. How do these results differ from what was found in part (ii).

(v)   Test the models from (ii) and (iv) using the Ramsey reset test. Interpret the results of the model tests.

(vi)   Use the Adjusted R-Squared, AIC, and BIC to compare the models estimated in (ii) and (iv).

(vii)   Would a test between these models be a nested or non-nested test? And if non-nested, would it be overlapping or non-overlapping?

(viii)   Compare the two models using a test based on an encompassing model.

(ix)   Compare the two models using the J-test.

(x)   Compare the two models using the Cox-Pesaran test.

(xi)   Compare the two models using the Vuong test (to solve this problem, you will have to

modify my r code for the civil war example. Since this one is hard, I will grade it as extra credit).

(xii)   Now control for years of education using a quadratic specification (and using a quadratic

specification for age). What can be said of the effect of education on wages.

(xiii)   Control for education, this time using dummy variables for education (and using dummy

variables for age as well).

(xiv)   Considering the regression you ran in (xiii), is their still evidence for a gender disparity in

wages.

(xv)   Are females rewarded differently than males for increasing levels of education? Conduct

an analysis using interaction terms.

2.   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 regression 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)   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.

(vi)   Suppose you believe that the percentage of mountainous terrain affects the probability of civil war in poor countries but not rich countries. Provide an appropriate regression and provide a formal hypothesis test relevant to this claim.