```
***** Preliminaries
capture log close _all  // Closes any log if open //

cd "C:\Users\AN.4271\Dropbox\HHS 651\Assignments\Assignment 2" // Set the wd.

log using "assignment2log", text replace   /* Starts a text-type log file called
                                            "assignment2log" */
set more off

*******************************************************************************
***********              HHS 651: Assignment 2          ********************
**************    Stata Solutions - Andrew Proctor      ********************
*******************************************************************************


***** Import Data
wbopendata, indicator(SG.GEN.PARL.ZS; TM.TAX.MRCH.WM.AR.ZS; ///
 NY.GDP.TOTL.RT.ZS; SE.SEC.CMPT.LO.ZS; SH.H2O.SAFE.ZS; SH.ANM.ALLW.ZS; ///
 IC.LGL.DURS; SE.SEC.ENRR; NY.GDP.PCAP.KD ) clear

/*  Only if you import using the CSV file:
import delimited using "./WDI_csv/WDIData.csv",  varnames(1) clear

 **** Keep only the relevant variables for analysis"
 keep if (inlist(indicatorcode, "SG.GEN.PARL.ZS", "TM.TAX.MRCH.WM.AR.ZS", ///
  "NY.GDP.TOTL.RT.ZS", "SE.SEC.CMPT.LO.ZS") ///
  | inlist(indicatorcode,  "SH.H2O.SAFE.ZS", "SH.ANM.ALLW.ZS", ///
  "IT.CEL.SETS",  ///
  "IC.LGL.DURS", "SE.SEC.ENRR", "NY.GDP.PCAP.KD"))
*/


*******************************************************************************
*******************        Data Preparation             *******************
*******************************************************************************


***** Drop Supranational areas
 *** Create list of countrycodes to drop
 local dropregions ARB CSS CEB EAR EAS EAP TEA EMU ECS ECA TEC EUU FCS HPC HIC ///
    IBD IBT IDB IDX IDA LTE LCN LAC TLA LDC LMY LIC LMC MEA MNA TMN MIC NAC  ///
    OED OSS PSS PST PRE SST SAS TSA SSF SSA TSS UMC WLD INX

    *** Drop the countrycodes in 'dropregions'
    foreach region in `dropregions' {
    drop if countrycode=="`region'"
    }

 **** Change indicator codes into readable names for variables
 replace indicatorcode = "pcgdp" if (indicatorcode=="NY.GDP.PCAP.KD")
 replace indicatorcode = "contractenftime" if (indicatorcode=="IC.LGL.DURS")
```

```
 52     replace indicatorcode = "womenlegislseats" if (indicatorcode=="SG.GEN.PARL.ZS")
 53     replace indicatorcode = "tariff" if (indicatorcode=="TM.TAX.MRCH.WM.AR.ZS")
 54     replace indicatorcode = "natresource" if (indicatorcode=="NY.GDP.TOTL.RT.ZS")
 55     replace indicatorcode = "secondeduc" if (indicatorcode=="SE.SEC.CMPT.LO.ZS")
 56     replace indicatorcode = "waterq" if (indicatorcode=="SH.H2O.SAFE.ZS")
 57     replace indicatorcode = "anemia" if (indicatorcode=="SH.ANM.ALLW.ZS")
 58     replace indicatorcode = "secondenroll" if (indicatorcode=="SE.SEC.ENRR")
 59
 60
 61        *** Explore the relationship between the Stata assign variable name and the year
 62        drop countryname
 63
 64
 65    ***** Reshaping the Data
 66                /*  Only if you import using the CSV file:
 67                    *** Drop empty variables
 68                        drop v62'
 69                */
 70
 71        *** Reshape year as a single column
 72        reshape long yr, i(countrycode indicatorcode ) j(year)
 73            /*  Only if you import using the CSV file:
 74                    replace year = year + 1955 if !missing(year)
 75            */
 76        *** Drop early years
 77        drop if year <  1970
 78
 79        *** Change v prefix to value
 80        rename yr value
 81         /*  If you import using the CSV file, this should instead be:
 82                        rename v value
 83            */
 84
 85        drop indicatorname // Drop superfluous variable
 86
 87        *** Reshape Indicator column so that indicators are each different variables
 88        reshape wide value, i(countrycode year) j(indicatorcode) string
 89        rename value* * // Remove variables
 90
 91    ***** Get log of GDP for analysis
 92        gen lnpcgdp =log(pcgdp)
 93
 94    ***** Set the Panel
 95        encode countrycode, gen(countryval) /* Encode country code so it can be used for
 96                                        xtset.                                      */
 97        xtset countryval year  // Tell Stata the panel structure.
 98
 99    ***** Create Lower Secondary Completion Average
100        /* Create a variable equal to secondary school education if nonmissing,
101           otherwise 0.   */
102        gen compsecforavg = secondeduc
```

```
103       replace compsecforavg = 0 if missing(secondeduc)
104
105       gen compsectotal = 0 // Sum of completion variable
106       gen nonmissing_compsec = 0  // Count of nonmissing completion variable
107
108       *** Compute sum and count
109       forval i=3/15 {
110           replace nonmissing_compsec = nonmissing_compsec + 1 if ///
111               !missing(l`i'.secondeduc)
112           replace compsectotal = (compsectotal + l`i'.compsecforavg)
113   }
114       *** Generate average
115       gen compsecavg = compsectotal / nonmissing_compsec
116
117
118
119
120   ********************************************************************************
121   ********************              Analysis                ********************
122   ********************************************************************************
123
124   ***** Exploratory Analysis
125       *** Check for missing observations in variables
126       mdesc
127
128       *** Get Descriptive Statistics
129       summarize
130
131
132   ***** Regressional Analysis
133       *** 3a: Basic Regression
134           reg lnpcgdp compsecavg contractenftime anemia waterq tariff womenlegislseats natresource
135
136       *** 3b: Regression with country and year fixed-effects
137           reg lnpcgdp compsecavg contractenftime anemia waterq tariff womenlegislseats  ///
138               natresource i.countryval i.year
139
140           /*  Commentary:
141               The sign of the effect changes for contract enforcement time, anemia,
142               women's seats in legislature, and natural resources.  The magnitude
143               of the effects change significantly, while contract enforcement time
144               becomes highly statistical significant (from non-significance), while
145               women's legislative seats and natural resources become less
146               significant.
147           */
148
149       *** Part 3c:  Test for heteroskedasticity
150       hettest, iid
151
152           /*  Commentary:
153               The p-value for the generalized Breusch-Pagan test is 0.7587, hence
```

```
154             we fail to reject the null hypothesis of homoskedasticity.  Note that
155             this is not the same thing as accepting homoskedasticity.
156         */
157
158     *** Part 3d:  Repeat the regression with heteroskedasticity-robust standard errors
159     reg lnpcgdp compsecavg  anemia waterq tariff womenlegislseats natresource ///
160             i.countryval i.year, robust
161
162         /*  Commentary:
163             The parameter estimates do not change, but the standard errors all
164             increase except for water quality and tariffs. Women's legislative
165             seats and natural resources lose significance.
166         */
167
168     *** Part 3e:  Repeat the regression allowing for correlation in errors.
169     reg lnpcgdp compsecavg contractenftime anemia waterq tariff womenlegislseats  ///
170             natresource i.countryval i.year, cluster(countryval)
171
172         /*  Commentary:
173             Once again, allowing for autocorrelation in standard errors will not
174             affect point estimates.  The standard errors increase even more in
175             this case, with anemia statistically significant and contract
176             enforcement time significant only at the p=0.1 level.
177         */
178
179     *** Part 3f: Basic fixed effects regression.
180     xtreg lnpcgdp compsecavg contractenftime anemia waterq tariff womenlegislseats  ///
181             natresource i.year, fe
182     *** 3g: Test for autocorrelation using the Inoue and Solo (2006) Test
183     xtistest
184
185         /*  Commentary:
186             With a p-value of 0.000, we reject the null of no autocorrelation.
187         */
188
189     *** Part 3h: Fixed effects regression allowing for autocorrelated errors.
190     xtreg lnpcgdp compsecavg contractenftime anemia waterq tariff womenlegislseats  ///
191             natresource i.year, fe robust
192         /*  Commentary:
193            A 1 percentage point change in the average lower secondary school completion rate
194          is estimate to increase GDP per capita by 0.209 percentage points (significant
195          at p=0.05 level).
196
197          A increase in the average time to enforcce a contract in court
198          of 1 day is estimated to reduce GDP by 0.01974 percentage points (significant at
199          p=0.1 level).
200
201          A 1 percentage point increase in the women aged 15-49 is estimated
202          to increase GPD per capita by 0.40422 percentage points, but this is not
203          statistically significant.
204
```

```
205           A 1 percentage point increase in the % of the population
206           with acess to an improved water source is estimate to increase GDP per capita by
207           1.23581 percentage points (significant at p=0.01 level).
208
209          A 1 percentage point
210          increase in the weight mean tariff rate rates for all products is estimated to
211          decrease GDP per capita by 0.52052 percentage points (significant at p=0.01 level).
212
213           A 1 percentage point increase in the proportion of seats held by women in the
214           national parliament/legislature is estimated to decrease GDP per capita by 0.11545
215           percentage points, but this is not statistically significant.
216
217           A 1 percentage point increase in
218           natural resource rents as a share of GDP is estimated to reduce GDP per capita by
219           0.1318 percentage points but this is not statistically significant.
220
221           Comparing this to the model in 3f, we see that whereas all parameter estimates are
222           statistically significant at the p=0.05 level (and significant at p<0.01), now
223           estimates for anemaia rates, women's share of seats in legislature, and natural
224           resource rents are all not statistically significant.
225          */
226
227      *** Part 3i: Discussion of Fixed Effects Assumptions in Model
228      /*
229      Commentary:
230      In short, this model is quite problematic.  As with any fixed effects or
231      difference-in-differences model, we are able to control for only those factors
232      that are fixed within a country over time, or which are common to all countries
233      in a given time period (year).  This panel notably contains the 2008 recession,
234      which hit different countries in very different ways.  Additionally, however,
235      one might note the following two observations (in addition to the concern in 3j):
236
237      (1)  We should very much question what is the basis for the variation in our
238      explanatory variables.  To the extent that these variables are relevant in the
239      fixed effects regression, this indicates that the variables change both between
240      regions and over time - and without a strong a priori basis for thinking these
241      changes are random. In truth, these variables are likely endogenous to other
242      characteristics of the country that are also not captured by the year or country
243      fixed effects.
244
245      (2) The explanatory variables are likely to have persistent effects (i.e. an effect
246      not only in the current period, but period (t+1,t+2,...).  Each of the issues listed
247      above indicate a violation of the strict exogeneity assumption:  E[u_it|X_it, a_t]=0.
248
249      */
250
251   *** Part 3j:  Concern with Fixed Effects in this context
252
253      /*
254      Commentary:
255      Because determinants of growth are likely to change only slowly, the fixed effects are
```

```
256       likely to absorb much of the effects of the true underlying features captured by the
257       proxy variables we choose to include.  In this case, instead of identifying the effects
258       of, for example, (relatively fixed) natural resource endowments, we instead capture the
259       effects of the change in value of these fixed resource endowments.  This change in value
260       is likely to occur for several reasons (and hence may be endogenous to a number of other
261       issues). In any case, it is unclear that estimates based on these minor fluctuations in
262       relatively static determinants will identify the same effects as estimates based on the
263       underlying variables themselves.
264       */
265
266   *** Part 4a:  Inclusion of lagged dependent variable in fixed effects model
267       xtreg lnpcgdp L.lnpcgdp compsecavg contractenftime anemia waterq tariff womenlegislseats  ///
268              natresource i.year, fe robust
269
270   *** Part 4b:  Discussion of lagged dependent variable
271   /*  Commentary:  See theory solution for Stock and Watson question , part (g).           */
272
273   *** Part 4d:  Discussion of inclusion of alternative explanatory variables
274       /*    Commentary:
275
276       Each of these variables are proxies for determinants of growth.
277
278       Variables like sanitation rate, tuburculosis rate, and female mortality
279       rate address issues related to the health environment of a country, which may help
280       determine it's human capital stock.  Each of these variables address different types
281       of health problems, consequentially producing different estimated effects and implicating
282       different types of endogeneity concerns.
283
284       Domestic credit and the tax rate address the financial institutional setting of a country,
285       which one might consider related to the tariff rate, although 'openness,' 'competitiveness,'
286       and 'financial access' or all somewhat different.  Once again, each of these concerns
287       implicate different types of endogeneity/OVB, but are also each likely to be sources of OVB in
288       the original FE regression.
289
290       Tertiary enrollment captures a human capital effect similar to the lower secondary school
291       completion rate, but it is also different in that the effects of basic and higher educational
292       are likely quite different.
293
294       Finally, cellphone usage rates proxies communication and infrastructure development, which has
295       generally not been addressed up by variables in the main regressions.
296
297       The different but correlated effects of these additional variables relative to the original
298       variables suggests we might want to include them in the analysis, but the fundamental concern
299       that we fail to isolate exogenous variation in the explanatory variables is still a major issue.
300
301       Access to captial for example is likely a major determinant of growth, but it is likely
302       endogenous to income/wealth (e.g. GDP) and demand for capital, expectations about the likelihood
303       of repayment (including features likely likelihood of conflict or recession), as well as to the
304       political and historical context of the banking sector.
305
306       In similar fashion, the tax rate is likely a response to issues like income, inequality, and
```

```
307       changes in the macroeconomic environemtn, as well as a host of politica and cultural/historical
308       determinants of the size of government and scope of redistribution.  We could continue listing
309       such concerns for each of these variables.                                                  */
310
311   *** Part 4e:  Discussion of the effect of including additional explanatory variables on number of
312   ***        repeated observations and implications
313
314       /* Commentary:
315
316       Including additional variables in this case will often make the number of observations
317       (years) per group (country) quite small, which will tend to increase standard errors and
318       reduce the power of hypothesis tests.
319
320       */
321
```