Advanced Quantitative Methods
Homework 2

Johan A. Elkink

jos.elkink@ucd.ie

Due March 3, 2014, 5pm

Please submit by email in PDF format. Add R code in a separate `.R` *file, or SPSS code in a separate* `.sps` *file, or Stata code in a separate* `.do` *file, or the code for any other package you use separately. Note that if you do not use Latex, there is an "Equation" entry in Microsoft Word under "Insert" that will allow you to include matrices and other mathematical equations.*

(10%) of the grade is used for an overall evaluation of the clarity and presentation of your code.

Data

The last part of this homework is based on the replication data for Ross (2004).[1] While Ross (2004) uses panel data – a set of countries observed over a number of years – we will be using a cross-section – only one year – for 1980, as well as one country over time. The data is available as `ross_1980.dta` and `ross_subsaharan.dta` ("taxation and representation") on the teaching data page.

1. For this question you will use the 1980 data set. We will be investigating the following model:

$$democracy_i = \beta_1 + \beta_2 loggdppc_i + \beta_3 logpopulation_i$$
$$+ \beta_4 islam_i + \beta_5 urban_i + \varepsilon_i \quad (1)$$

   (a) (5%) Use residual plots (against fitted values and independent variables) to investigate potential heteroscedasticity.[2]

   (b) (5%) Perform a Breusch-Pagan test for heteroscedasticity. What do you conclude?

---

[1] `http://hdl.handle.net/1902.1/14305 UNF:5:8Ymx2XXoRulfSgFb5Xq1uQ==`.

[2] It will help to use `na.action = na.exclude` as an option to the `lm()` command, so that the `residuals()` command will return a vector of the same length as the original data vectors. Alas, if you want to use the `sandwich` package to calculate the "HC3" standard errors, you need to estimate the model without the `na.action` parameter as well.

(c) (5%) Perform a White test for heteroscedasticity. What do you conclude?

(d) (5%) Calculate "HC3" robust standard errors and recalculate the t- and p-values using these. Do the results of the regression change? Include this as a separate regression (2nd column) in the same regression table.

(e) (5%) Calculate Variance Inflation Factors. What do you conclude?

2. For this question we will use the Subsaharan data set. Perform all analysis for this question on the data for the Union of the Comoros. You can create a subset as follows:

```
rossCOM <- subset(rossSSA, country == "COM")
rossCOM <- rossCOM[order(rossCOM$year), ]
```

We will use this data to further investigate model (1), but without the `islam` variable.

(a) (5%) Re-estimate the model using this data set. Perform a Breusch-Godfrey test for autocorrelation in the residuals. What do you conclude?

(b) (5%) Add a lagged dependent variable to the model and re-estimate. You can calculate a lagged variable using:[3]

```
T <- dim(data)[1]
data$Ldemocracy <- c(NA, data$democracy[-T])
```

(c) (5%) Add the lagged version of the `loggdppc` variable to the model and re-estimate (keep the lagged dependent variable), so the model is now:

$$
\begin{aligned}
democracy_t =& \beta_1 + \beta_2 loggdppc_t + \beta_3 logpopulation_t + \beta_5 urban_t \\
& + \beta_6 democracy_{t-1} + \beta_7 loggdppc_{t-1} + \varepsilon_t
\end{aligned}
\tag{2}
$$

(d) (10%) Calculate the short-term and long-term impact of an increase in the log of GDP per capita on democracy.[4] Work out the what this means if GDP per capita increases by 1%.[5]

(e) (5%) Perform an Adjusted Dickey-Fuller test on the residuals of the regression. What do you conclude?

(f) (5%) Calculate first differences for all variables in Equation (1), except `islam`, and re-estimate the model using these first differenced variables (without the lagged variables added in Equation (2). You can calculate first differences (the amount of change in the variable at a given time point) using:

```
data$Ddemocracy <- data$democracy - data$Ldemocracy
```

---

[3]This assumes the data contains only one time-series and is sorted by time, which should be the case for this data set.

[4]In applied work, you would probably not proceed because the t-tests are insignificant. This is due to the tiny sample size. One can nevertheless calculate these short- and long-term based on the estimated coefficients.

[5]Note that this means GDP is multiplied by 1.01 and note that $log(1.01a) = log(1.01) + log(a)$, so the increase in the log of GDP per capita is $log(1.01)$.

(g) (5%) Perform a Durbin-Watson test for autocorrelation in the residuals of the first difference model. What do you conclude?

3. Combining results from the two questions:

(a) (10%) Present one regression table, with one column per model, including: 1) the baseline model of Q1 as in Equation (1); 2) the Q1 model with robust standard errors; 3) the baseline model of Q2; 4) the Q2 model with a lagged dependent variable; 5) the Q2 model as specified in Equation (2).

(b) (15%) Write a one page report on your findings, discussing this study of the relation between GDP per capita and democracy, including a clear interpretation of the short- and long-term impacts.

Grade conversion scheme

Score	Grade	Score	Grade	Score	Grade	Score	Grade
97-100%	A+	85-87%	B	74-76%	C-	54-64%	E+
94-96%	A	83-84%	B-	71-73%	D+	44-53%	E
91-93%	A-	80-82%	C+	68-70%	D	33-43%	E-
88-90%	B+	77-79%	C	65-67%	D-	0-32%	F

Ross, Michael. 2004. "Does taxation lead to representation?" *British Journal of Political Science* 34:229–249.