Advanced Quantitative Methods
Homework 1: OLS & Diagnostics

Johan A. Elkink
jos.elkink@ucd.ie

Due February 21, 2018, 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. Alternatively, you can integrate R code and responses in R Markdown and submit both in PDF format (easiest is to use HTML format and then in the web browser save as PDF).*

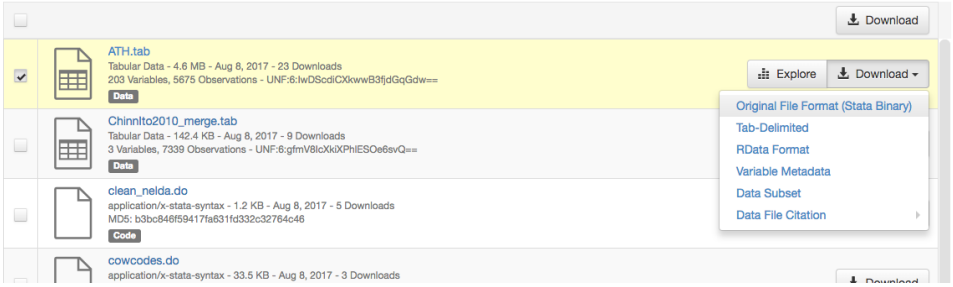
Note that if you do not use Latex, there is an "Equation" entry in Microsoft Word under "Insert" that will allow you to include matrices and other mathematical equations. You can also hand-write questions 4 and 5 and submit a PDF scanned version.

Percentages with an asterisk indicate that positive rather than negative marking will be applied.

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

Data

This homework is based on the replication data for Escribà-Folch, Meseguer and Wright (2015), which you can access at their Harvard Dataverse record—probably the first study you find when searching for the keyword "remittance".[1]



The "explore" button might also be very helpful to get a feel for the data.

[1] `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NFMGFD`

Questions

1. In line with Escribà-Folch, Meseguer and Wright (2015), we will be investigating the relationship between remittances and democracy. However, while they have a binary dependent variable capturing transitions to democracy, we will have a continuous dependent variable capturing a scale from -10 for authoritarian regime to 10 for democracies, called Polity IV. Note that our data set contains a number of countries over a number of years, so this is a panel data set. We will use what is called the "pooled model", basically ignoring the panel data structure of the data.
   The baseline model is:
   $$P_{it} = \beta_1 + \beta_2 \log(R)_{it} + \varepsilon_{it},$$
   where P stands for the Polity IV score (**pol4_polity2** in the data set) and R for remittances (**wdi_remitc**); i for the country and t for the year. Note the logarithmic transformation.

   (a) Perform the above regression. We will refer to this as Model 1.

   (b) Add control variables for GDP per capita (**mad_lgdppc**) and population size (**mad_lpop**).[2] We refer to this as Model 2.

   (c) In addition, add control variable for intra-country conflict (**prio_lconflict_intra**), inter-country conflict (**prio_lconflict_inter**) and democracy in neighbouring states (**neighbor_pol4_polity2**). This is Model 3.

   (d) (10*%) Produce a publishable regression table, with one column for each of the above three regressions.[3]

   (e) (20*%) Write a 500 word report on your findings, discussing this study of the relation between remittances and democracy. Discuss the statistical results (coefficients, t-tests, R^2, etc.), but focus on the substantive interpretation—translate it to an essay for political scientists on your findings.

2. Using Model 3, we proceed with a number of diagnostic tests related to heteroscedasticity and multicollinearity.

   (a) (5%) Use residual plots against fitted values and against **wdi_remitc** to investigate potential heteroscedasticity.[4]

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

   (c) (5%) Calculate "HC3" robust standard errors and recalculate the t- and p-values using these. Do the results of the regression change?

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

[2] Note that the "l" in the variable name indicates that these are logged values as well, so there is no need to take the logarithm explicitly—this is already done.

[3] See `http://www.joselkink.net/files/regression-tables.pdf`.

[4] 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.

3. We will continue with the same model, but using only observations from Algeria (to be selected using the **cowcode** variable, which should be equal to 615.[5]

   (a) (3%) Re-estimate Model 3 using only the Algerian cases (**cowcode** is 615).

   (b) (3%) Plot the residuals over time (the data should be sorted, so this should just be `plot(residuals(mdl))`.

   (c) (5%) Perform a Breusch-Godfrey test for autocorrelation in the residuals. What do you conclude?

   (d) (5%) Perform a Durbin-Watson test for autocorrelation on the residuals. What do you conclude?

4. Consider a matrix $\mathbf{X}$ containing the test scores of $n = 5$ students on $k = 2$ tests and $\mathbf{u}$ a vector of n ones. $\mathbf{D}$ contains the inverse of the standard deviations of $\mathbf{X}$.

$$\mathbf{X} = \begin{bmatrix} 3 & 9 \\ 4 & 10 \\ 6 & 7 \\ 10 & 7 \\ 3 & 1 \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{D}^{-1} = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{bmatrix} = \begin{bmatrix} 0.3390 & 0 \\ 0 & 0.2863 \end{bmatrix}$$

   Write down the result of

   (a) (4%) $\mathbf{u}'\mathbf{X}$

   (b) (4%) $\mathbf{M} = \frac{1}{n}\mathbf{u}\mathbf{u}'\mathbf{X}$

   (c) (4%) $\mathbf{A} = \mathbf{X} - \mathbf{M}$

   (d) (4%) $\mathbf{Z} = \mathbf{A}\mathbf{D}$ (you will see that $\mathbf{Z}$ contains the standardized columns of $\mathbf{X}$)

5. When we run an OLS regression, we estimate the coefficients using $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. We can subsequently calculate the residuals as

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{M}\mathbf{y}.$$

   (a) (6%) Show that $\mathbf{M}$ is symmetric ($\mathbf{M}' = \mathbf{M}$).

   (b) (6%) Show that $\mathbf{M}$ is idempotent ($\mathbf{M}'\mathbf{M} = \mathbf{M}$).

Grade conversion scheme

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

Escribà-Folch, Abel, Covadonga Meseguer and Joseph Wright. 2015. "Remittances and Democratization." *International Studies Quarterly* 59(4):571–586.

[5] COW stands for Correlates of War, see `http://www.correlatesofwar.org/data-sets/cow-country-codes`.