

Advanced Quantitative Methods

Homework 3

Johan A. Elkink
jos.elkink@ucd.ie

Due February 28, 2012, 9am

Please submit the answers in PDF format, with a separate R file for the R code.

Relevant variables in data set `p4factor.csv`:

<code>dem</code>	democracy score based on factor analysis of Polity IV indicators
<code>bautlag</code>	lag of binary autocracy indicator
<code>laggppc</code>	lag of GDP per capita
<code>energy2</code>	energy consumption per capita
<code>pnbdem</code>	proportion of neighboring democracies
<code>cwar</code>	civil war (1 is civil war on territory)
<code>ipyyears</code>	years without conflict on territory
<code>egr</code>	economic growth rate
<code>propdem</code>	global proportion of democracies
<code>nbtd</code>	neighboring transition to democracy
<code>tatdem</code>	time at democracy
<code>tataut</code>	time at autocracy

The dependent variable is based on the Polity IV data set (Jaggers and Gurr, 1995; Marshall and Jaggers, 2002); the independent variables are from the replication data set of Gleditsch and Ward (2006).

1. We will look at regression model

$$dem_i = \beta_0 + \beta_1 energy2_i + \beta_2 laggdppc_i + \beta_3 pnbdem_i + \beta_4 ipyears_i + \beta_5 egr_i + \varepsilon_i$$

in 1998 by executing the following code:

```
p4 <- read.csv(file.choose())  
p4 <- p4[p4$year.x == 1998, ]
```

```
summary(m <- lm(dem ~ energy2 + laggdppc + pnbdem + ipyears + egr, data = p4))
```

- (5%) In a publishable format, present estimated coefficients, standard errors, t -values, and p -values.
- (5%) Do an F -test, testing whether the coefficients for `energy2` and `laggdppc` might both be zero.
- (5%) Do a Chow test, comparing the difference in coefficients between those countries that experienced civil war in that year and those that did not (variable `cwar`).
- Perform (4%) Goldfeld-Quandt, (4%) Breusch-Pagan and (4%) White tests for heteroscedasticity.
- (10%) Write a short (approximately one page) substantive interpretation of the results, which has to be consistent with the above results, but which should be phrased in layman terms, i.e. interpretable for anyone who does not understand regression analysis.

2. Using the dataset `cpds2007.csv` (see codebook for variable descriptions):¹

(a) (5%) Estimate and present in publishable table format the results for:

$$tot_taxrev_i = \beta_0 + \beta_1 gov_left2_i + \beta_2 gov_right2_i + \beta_3 judrev_i + \beta_4 vturn_i + \varepsilon_i$$

(b) (5%) Investigate whether there are any indications of nonlinearities that are not modeled properly.

(c) (5%) Investigate whether there are any problematic cases (outliers, high leverage, high influence).

(d) (5%) Investigate whether there is any multicollinearity problem.

(e) (5%) Perform a Goldfeld-Quandt test of heteroscedasticity.

(f) (5%) Using the `hccm()` function, calculate and present standard errors of the HC0 and HC3 variations. Perform the t -tests using these revised residuals.

(g) (8%) In layman terms, interpret the results of this analysis (approximately one page).

3. Write a small Monte Carlo simulation (10%) using matrix notation rather than `lm()` with:

$$n = 30$$

$$m = 500$$

$$\begin{aligned} \mathbf{y} &= \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \beta_3 \mathbf{x}_3 + \varepsilon \\ &= 0.1 + 0.2\mathbf{x}_1 + 0.1\mathbf{x}_2 + 0\mathbf{x}_3 + \varepsilon \end{aligned}$$

$$\mathbf{x}_1 \sim N(3, 0.5)$$

$$\mathbf{x}_2 \sim N(1 + 2\mathbf{x}_1, 0.25)$$

$$\mathbf{x}_3 \sim N(0.1\mathbf{x}_2, 0.25)$$

$$\varepsilon \sim N(0, 2),$$

where n is the sample size and m the number of iterations.

(a) (5%) Report the proportion of times that the simulation commits a Type I error, rejecting $H_0 : \beta_3 = 0$, using a t -test with $\alpha = .05$.

(b) (5%) Report the proportion of times that the simulation commits a Type II error, not rejecting $H_0 : \beta_2 = 0$, using a t -test with $\alpha = .05$.

(c) (5%) Report the proportion of times that the simulation commits a Type II error, not rejecting $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$, using an F -test with $\alpha = .05$.

Gleditsch, Kristian S. and Michael D. Ward. 2006. "Diffusion and the international context of democratization." *International Organization* 60(4): 911–933.

Jagers, K. and T.R. Gurr. 1995. "Transitions to democracy: tracking democracy's third wave with the Polity III data." *Journal of Peace Research* 32: 469–482.

Marshall, M.G. and K. Jagers. 2002. "Polity IV project: political regime characteristics and transitions, 1800–2002."

URL: <http://www.bsos.umd.edu/cidcm/polity/>

¹The codebook can be found at:

http://www.ipw.unibe.ch/content/team/klaus_armingeon/comparative_political_data_sets/index_eng.html.

Note that this dataset has been slightly manipulated to remove missing information on tax revenues.