POLS/CSSS 503:
Advanced Quantitative Political Methodology

Problem Set 5 (Optional)

Professor: Chris Adolph, Political Science and CSSS

Spring Quarter 2014

Due in section, 6 June 2014

General instructions for homeworks: Homework can be handwritten or typed. For any exercises done with `R` or other statistical packages, you should attach all code you have written and all (interesting) output. Materials should be stapled together in order by problem. The most readable and elegant format for homework answers incorporates student comments, code, output, and graphics into a seamless narrative, as one would see in a textbook.

Problem 1: Assessing dynamics of mystery time series, part 1

In the file `mysteryts.csv`, you will find ten columns of time series data. Each column is an independent time series generated by your instructor to have a particular AR(p) or MA(q) structure. For each series, you task is to make your best guess of what process and order the time series was generated from, and whether the series is covariance-stationary. For each series **a.** to **j.**, you should conduct the following tasks:

- Graph the time series against time using `plot()`. There is no need to show this graph in your write up, but you may describe whether it appears mean-reverting or not.
- Produce the correlograms of autocorrelations and partial autocorrelations using `ACF()` and `PACF()`. You do not need to show the graphs, but do write a sentence

describing what sort of patterns would be consistent with the sample ACF and PACF you found.

- Make you best guess as to the underlying process: whether AR or MA, what order, your approximate guesses as the values of the ϕ's or ρ's, and whether the process is stationary.

Problem 2: Assessing dynamics of mystery time series, part 2

In this part, we will focus more closely on the problem of stationarity. In the file `mysteryts2.csv`, you will find five additional time series, **k.** to **o.** Once again, each column is an independent time series generated to have a particular structure. All of these series follow some AR(p) process. For each time series, conduct the following tasks:

- Subset the first 20 observations. Plot them against time, and plot the ACF and PACF. (There is no need to show these graphs on your write-up.) Based on these plots, guess the order of the AR(p), and the approximate values of the ϕ's.
- Subset the first 100 observations, and repeat the analysis.
- Conduct your analysis on all 1000 observations. Did any of your conclusions change? If so, what implications does this have for assessing stationarity in comparative politics time series?

Problem 3: Analyzing a Panel Data Set: Oil and Democracy, Revisited

For this problem, we return to the dataset employed by Michael Ross in "Does Oil Hinder Democracy?" *World Politics*, 2001. n that paper, Ross estimated a time series cross-section model of Polity scores regressed on oil exports and a battery of controls. In Homework 4, we worked with a single year of data; now, we will work with the full dataset, `rossoildata.csv`.

A description of the included variables follows:

Variable	Description
`regime1`	1–10 scale increasing in democracy; computed from Polity components
`oil`	Fuel exports as a proportion of GDP
`metal`	Ore and mineral exports as a proportion of GDP
`GDPpc`	per capita GDP in PPP dollars, lagged 5 years
`islam`	Muslims as a proportion of population, 1970 data
`oecd`	dummy for rich industrialized countries
`cty_name`	the name of the country observed
`id`	a three character abbreviation of the country name
`id1`	a numeric country code
`year`	the year of the observation

We will use this dataset and the baseline regression from Ross (2001) to explore our tools for specifying regression models and improving fit.

a. Load the dataset `rossoildata.csv`. We need to construct 5-year lags of the variables `oil`, `metal`, and `GDPpc`. The dataset seems to include such lags under names like `oil_5`. (*i*) Examine these preconstructed lags and explain why they shoudl not be used. (*ii*) Use the `lagpanel()` function in the `simcf` library to create your own 5-year lags of these variables (call them `oilL5`, etc., to distinguish them from Ross's "lags"), and check that you have constructed these lags correctly.

b. Check the time series properties of $\texttt{regime1}_{it}$ using all the relevant methods we have learned (time series plots, ACFs, PACFs, unit root tests, AIC tests on estimated ARIMA models), and specify an ARIMA(p,d,q) specification for $\texttt{regime1}_{it}$. For modeling purposes, you may assume that all units share the same ARIMA(p,d,q) process, but note if this assumption seems faulty based on your data exploration.

c. Choose and justify either a random effects, fixed effects, or mixed effects intercept for the model.

d. Estimate a panel regression model of `regime1` on `oilL5`, `metalL5`, `GDPpcL5`, `islam`, and `oecd`. Record the estimate coefficients – including those for any AR and/or MA terms – in a table. You do not need to show the fixed effects (if you included them), but indicate whether the model includes fixed or random intercepts. The table should make clear whether you have differenced the outcome variable or not, and should indicate the number of units, time periods, and observations included, as well as the standard error of the regression and AIC.

e. For each covariate in x_{it}, interpret the long and short run conditional expectation of y_{it} given hypothetical values of that covariate, with all other covariates held constant. You may calculate either an "expected value" like $\mathrm{E}(y_{\mathrm{hyp},t}|x_{\mathrm{hyp},t})$ for some hypothetical $x_{\mathrm{hyp},t}$ or a "first difference" $\mathrm{E}(y_{\mathrm{post},t} - y_{\mathrm{pre},t}|x_{\mathrm{pre},t}, x_{\mathrm{post},t})$ for two different hypothetical values, $x_{\mathrm{pre},t}$ and $x_{\mathrm{post},t}$. Present these results nicely in graphics, tables, or sentences, as you would for a paper. Explain whether and why you trust or distrust the long run estimates for your model.

f. (Bonus) Compare the estimated short- and long-run first difference in `regime1` given a persistent change in `oilL5` from the 50th percentile to the 95th percentile of the fully observed data, all else equal. Show this result graphically. Then re-estimate your model using Ross's problematic lagged covariates instead of your own, and recompute these short- and long-run first differences. Do the results differ?