EC 339–001

Problem Set 3

Prof. Santetti

Fall 2023

INSTRUCTIONS: Carefully read all problems. You must submit a single STATA do-file with your *first name* (mine would be `marcio.do`). In case you submit your files with different names, you will lose 1 point.

You can find templates for your do-files on `theSpring`, under the "Templates" module. Please consider using it.

I should be able to fully replicate your code to answer the questions, as well as fully understand your written interpretations to the proposed problems.

Avoid using unnecessary code in your submission files. It is totally fine to do other things by yourself that may help you better understand the data and the problems. However, for grading purposes, I am only interested in the commands and interpretations that actually answer the questions. You may keep a separate file for yourself with your additional explorations.

`Assignment due 11/17, 1:00 PM`.
`Points Possible: 30`

- You have more than 2 weeks to complete this assignment. In accordance with our `course syllabus`, no late submissions will be accepted.
- Be honest. Don't cheat.
- As a Skidmore student, always recall your votes of academic integrity, and the **`Honor Code`** you have abided by:

> *"I hereby accept membership in the Skidmore College community and, with full realization of the responsibilities inherent in membership, do agree to adhere to honesty and integrity in all relationships, to be considerate of the rights of others, and to abide by the college regulations."*

Have fun!

Problem 1

Download the `oecd.dta` data set (available at the "Data files" module on `theSpring`), used by `Everaert and Pozzi (2014)` to examine the predictability of consumption growth in 15 OECD countries. For this exercise, however, we will only concentrate on US data.

(a) After importing the data set, filter only observations for the US economy (labeled "US").

(b) Estimate the following model:

$$csumptn_t = \beta_0 + \beta_1 hours_t + \beta_2 gov_t + \beta_3 r_t + \beta_4 inc_t + u_t$$

(c) Run a *RESET* test (up to the 4^{th} power) to verify whether (b)'s model is well specified. Assume $\alpha = 0.05$. What do you conclude?

(d) Next, estimate a new model. Remove the statistically *insignificant* variable(s) from (b)'s model (hold the intercept!). Also, add the first *lag* of the dependent variable on the right-hand side.

(e) Then, estimate a *RESET* test for (d)'s model. What do you conclude? Explain.

Problem 2

Using the same data set from **Problem 1**, answer the following questions:

(a) Does the model from Problem 1, part (d) suffer from multicollinearity? Estimate Variance Inflation Factors (VIFs) and interpret your results.

(b) Does the model from Problem 1, part (d) suffer from serial correlation? Run a Durbin-Watson test and report your inference from this test.

(c) Following a similar procedure as in the previous part, run a Breusch-Godfrey test for first-degree serial correlation and report your inference from this test.

Problem 3

Using time-series data on five different countries, `Atkinson and Leigh (2013)` examine changes in inequality measured as the percentage income share held by those with the top 1% of incomes. A subset of their annual data running from 1921 to 2000 can be found in the data file `inequality.dta`, available at the "Data files" module on `theSpring`.

(a) Using data *only* for the US economy, estimate a regression model for $share_t$, controlling for tax_t.

(b) From (a)'s model, extract its residuals. Plot these over time.

(c) Still using (a)'s model residuals, plot the autocorrelation function (ACF) for this series. For *how many lags* is the autocorrelation coefficient (ρ) statistically significant? Explain.

(d) Does (a)'s model suffer from serial correlation? Estimate *Durbin-Watson* and *Breusch-Godfrey* tests, assuming $\alpha = 0.05$. Interpret your results.

(e) Run a *RESET* test (up to the 4^{th} power) to verify for possible functional form misspecification. Assume $\alpha = 0.05$ and interpret your results.

(f) In case your answers to (d) were *positive* for serial correlation, use the *Cochrane-Orcutt* estimator for (a)'s model.

(g) Interpret the *slope* coefficient from (f).