Regression with Time Series Problem Set

1. Data on the copper industry for the years 1951 – 1980 are given in the Stata file “copper data.dta.” Use these data, and the definitions of the variables within the file, to answer the following questions.

a) Estimate using our basic OLS the following regression model:



b) Diagnose the presence of autocorrelation (AR1) in your model in (a) by:

i) calculating the correlation coefficient between the current and lagged residuals.

ii) constructing the correlogram of the residuals.

iii) conducting the Lagrange Multiplier (Breusch-Godfrey) test

c) Estimate the model with HAC (Newey-West) standard errors.

d) Conduct **both** a Cochrane-Orcutt and a Prais-Winsten version of this model that takes account of AR(1) errors – again, regardless of your findings in (b).

1. We are going to reexamine the model you estimated in (1). Your goal in this question is to use the same model shown in (1a) but to determine whether any serial correlation that exists can be eliminated through the use of some variation of an ARDL model.

a) Estimate the original model, an ARDL(0,1) model, an ARDL(1,0) model, and an ARDL(1,1) model. In *all* cases where you are to use a distributed lag you are to use a single lag of the 12-month average price of aluminum and include the other variables as they appear in the original equation.

b) Compare the four models based on the AIC and the SC/BIC criteria

c) Compare the four models based on the goal of eliminating evidence of serial correlation. Use the Lagrange Multiplier test (at the .05 level of significance and with Stata) to see if there is evidence of the continuing presence of autocorrelation or not in each specification.

d) Why in (a) did you need to drop the first observation for the original model based on what was done in this problem?

3. Now we will turn to forecasting with a pure autoregressive model (**no** independent variables – only lags of the dependent variable).

a) Forcast **lnA** for three periods beyond the last period in the data.

b) Then construct the 95% interval estimates for each forecast in (a).

4. Finally, you are to do use one more forecasting approach, exponential smoothing.

a) Forecast lnA for one year beyond the last time point in the data based on minimizing the sum of squares of within-sample forecast errors.

b) Now do exactly the same forecast of lnA for one year beyond the last time point in the data, but base it on a set value of α = 0.60.

5. We continue to use the “copper” data but now focus on just two variables, ln*C* and ln*I*.

a) Estimate the model



and use the Lagrange Multiplier test (with one lag) to assess the presence of serial correlation.

b) Create two graphs, each a side-by-side graph. The first graph will have ln*C* and the differenced values of ln*C* next to it. The second graph will have ln*I* and the differenced values of ln*I* next to it.

c) Carefully explain how your results in (b) affect your faith in your results in (a).

d) Test for unit roots, using Stata or manually, for ln*C* and then for ln*I* using the Augmented Dickey-Fuller (ADF) test including one lagged difference.

e) Now, regardless of your findings in (d), use the ADF test with no lags to test whether the differenced ln*C* and differenced ln*I* show evidence of unit roots.

f) Finally, test for cointegration using the original model in (a) of this question. You are to test for this at the **0.10** level of significance.