Economics 311 Fall 2017
Problem Set #6 Due: Wednesday, November 15

This problem set again asks you to use Stata to perform some regressions and to interpret the results. Assemble your answers *along with all relevant Stata output* into a Word document and send it to me as an email attachment in .docx or .pdf format.

This assignment uses dataset bangla.dta, containing 34 years of data on sugar-cane prices (relative to the alternative crop jute) and the area planted with sugar cane in a region of Bangladesh. Apart from year numbers, these are the only two variables in the data set and they are called p and a, respectively. Since this is a time-series sample, you will need to use tsset year to tell Stata that year is your time variable.

You are to find and estimate an appropriate lag specification for the relationship between the log of area planted (dependent) and the log of price (regressor), providing detailed justification for your choices. Assume that the appropriate model is log-log (you will need to generate logged versions of both variables), so the only specification decisions are the lag structure of the equation and the appropriate time-series specification for the error term. You should explore (and report on) simple distributed-lag models, a Koyck ("dynamic") lag model, and perhaps models with lags of both the dependent and independent variables.

As in Daily Problem #18, you should choose your specification based on significance of coefficients, goodness of fit, information criteria, and absence of serial correlation in the residual. After reg commands, you can use the estat bgodfrey command to test for serial correlation in the residuals and estat ic to obtain the AIC and BIC. If you need to, you can use Prais-Winsten rather than OLS, but recall the problem with using this model when there is a lagged dependent variable in the equation. (You can use the ssesearch option in the prais command to avoid this problem.)

You should consider the following in your exploration:

- When comparing the performance of models using information criteria or other standards, always make sure that the regressions being compared use identical samples. By default Stata will use all of the available observations for every regression, so lengthening lags will decrease the number of observations used and make the numbers incomparable unless you manually set a fixed sample. Use the "if" or "in" clause in your estimation command to specify the sample. The easiest way to do this is to use the in clause. The command reg ... in 5/34 will use only the 5th through the 34th observations. Or you can use the year variable that you created and say reg ... if year>4 & year<35.
- When estimating models, test whether the error term is autocorrelated using a Breusch-Godfrey test. If appropriate, correct for autocorrelation. You can use the

estat bgodfrey command after reg, but not after prais. To test for serial correlation in the residuals after a Prais-Winsten regression you will have to run the regression of the current residual on the regressors and the lagged residual and use the formula in class. This requires several steps as outlined below:

- You need to perform the Breusch-Godfrey test on the *transformed* (quasi-differenced) residuals: $\hat{u}_t = e_t - \hat{\rho} e_{t-1}$, where $e_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t - \ldots$. If you use the prais command in Stata and then use predict ... , resid to get residuals, the residual variable created will be the *untransformed* residuals *e*, not the transformed residuals $\hat{u}$. So to perform the Breusch-Godfrey test, you'll need to do something like predict e , resid to get the untransformed residuals, then gen uhat = e - ##### * l.e to compute the transformed residuals, where ##### is the estimated $\hat{\rho}$ from your Prais-Winsten output.
- You will also need to do this for each of the regressors on the right-hand side of the regression to prepare for the Breusch-Godfrey regression. For example, if x is a regressor, you would gen transx = x - ##### * l.x to get a transformed (quasi-differenced) version of x. (You'd do this for each of the regressors in the prais command.)
- Once you have quasi-differenced the residuals and all of the regressors, you can run reg uhat transx l.uhat and compute the Breusch-Godfrey statistic as NR^2 where N is the number of observations in the residual (uhat) regression. Note that if there is more than one x there will be more than one transx and that if the lagged dependent variable l.y appears on the right-hand side then the transformation will be something like gen transly = l.y - ##### * l2.y.
- This procedure is complicated, but important if you want to be assured that there is no remaining serial correlation after using Prais-Winsten.

- If there are alternative specifications that suggest themselves, go ahead and explore them.

Your report should be a short paper describing your results, not a set of answers to specific questions.

- Provide a narrative of the specifications you tried (including the results) and why you chose your preferred model.
- For your preferred model, examine the dynamic effects of price on area planted in detail, computing the effects of *both temporary and permanent* 10% increases in price on area planted over time (first-year effect, second-year effect, third-year effect, …, long-run effect).
- Summarize your results in a conclusion with an assessment of how confident you are of their validity (and why).