TIME SERIES REGRESSION NOTES

Time series regression analysis is tricky!! Be careful and don't generate spurious results.

1. Transform all variables (both dependent variable and explanatory variables) to stationary form. Then your equation will be balanced.
2. Search for the correct dynamics of your regression model. My preference is to build the autoregressive part first and then try to add the dynamics of several interesting explanatory variables later.
3. You have the model dynamically correct when the residuals of your regression model are "white noise," i.e. uncorrelated. (Wooldridge refers to such models as being "dynamically complete" and "absent of serial correlation.")

See the programs spurious.sas, spurious2.sas, and spurious3.sas to understand how important it is to run time series regressions in their stationary forms. See Exercise 7 for illustrations of this point.

Different Types of Time Series Regression Models

Here it is assumed that the time series z_t and w_t are stationary time series.

For example z_t is (weakly) **stationary** if

(i) $E(z_t) = \mu_z$ for all t (constant mean)

(ii) $Var(z_t) = \sigma_z^2$ for all t (constant variance)

(iii) $Cov(z_t, z_{t+j}) = \gamma_j$ for all t and arbitrarily chosen j = 1, 2, …

Static Time Series Regression Model:

$$z_t = \beta_0 + \alpha_0 w_t + u_t$$

Distributed Lag Time Series Regression Model:

$$z_t = \beta_0 + \alpha_0 w_t + \alpha_1 w_{t-1} + \cdots + \alpha_p w_{t-p} + u_t$$

Autoregressive, Distributed Lag Time Series Regression Model:
(commonly called the ARX model)

$$z_t = \beta_0 + \delta_1 z_{t-1} + \cdots + \delta_r z_{t-r} + \alpha_0 w_t + \cdots + \alpha_p w_{t-p} + u_t$$

Obtaining a Dynamically Complete Time Series Regression Model

For a good example of time series regression modeling you should follow the "Fertility Equation" example in the book as Wooldridge works through the effect of various variables on the general fertility rate in the U.S. For an example of a spurious regression of the fertility equation see Example 10.4, page 338 in the textbook. Here we are regressing a nonstationary dependent variable (gfr) on a nonstationary explanatory variable (pe) plus two dummy variables (ww2 and pill). This regression overstates the significance of all of the explanatory variables. Wooldridge then starts the reader in the right direction in Example 11.6 on page 378. Here he suggests regressing Δ gfr on Δ pe both variables now being in stationary form. In equation (11.27) Wooldridge puts in two lags of Δ pe in the equation in addition to the contemporaneous value of Δ pe. As it turns out the effect of Δ pe only comes into play with the second lag of Δ pe. We then add the lagged value of Δ gfr as an explanatory variable (which is significant) and we add the dummy variables ww2 and pill. We now have a **dynamically complete model**. We know we have such a model when we look at the autocorrelation function of the residuals, it indicates that the residuals are white noise (i.e. uncorrelated since the autocorrelations at the various lags are statistically insignificant.) There is nothing to be gained by adding any more lags of the dependent variable or explanatory variables. As it turns out this properly estimated time series regression no longer produces a very significant effect of either the world war 2 dummy or the pill dummy. Thus you can see that time series regression analysis is tricky! See my EVIEWS program Fertil3.wf1 posted on the website for this course to follow the logic of this example.

Augmented Dickey-Fuller Unit Root Tests

But how do we know when to difference the data to make it stationary? You use the Augmented Dickey-Fuller t-statistic. For more discussion, see Wooldridge, pp. 607 – 613.

Here are the **various cases** of the test equation:

a. When the time series is **flat** (i.e. doesn't have a trend) and **potentially slow-turning around zero**, use the following test equation:

$$\Delta z_t = \theta z_{t-1} + \alpha_1 \Delta z_{t-1} + \alpha_2 \Delta z_{t-2} + \cdots + \alpha_p \Delta z_{t-p} + a_t$$

where the number of augmenting lags (p) is determined by minimizing the Schwartz Bayesian information criterion or minimizing the Akaike information criterion or lags are dropped until the last lag is statistically significant. EVIEWS allows all of these options for you to choose from. Notice that this test equation does **not** have an intercept term or a time trend. What you want to use for your test is the t-statistic associated with the Ordinary least squares estimate of θ. This is called the **Dickey-Fuller t-statistic**. Unfortunately, the Dickey-Fuller t-statistic does not follow a

standard t-distribution as the sampling distribution of this test statistic is skewed to the left with a long, left-hand-tail. EVIEWS will give you the correct critical values for the test, however. Notice that the test is **left-tailed**. The null hypothesis of the Augmented Dickey-Fuller t-test is

$$H_0 : \theta = 0 \qquad \text{(i.e. the data needs to be differenced to make it stationary)}$$

versus the alternative hypothesis of

$$H_1 : \theta < 0 \qquad \text{(i.e. the data is stationary and doesn't need to be differenced)}$$

b.   When the time series is **flat and potentially slow-turning around a non-zero value**, use the following test equation:

$$\Delta z_t = \alpha_0 + \theta z_{t-1} + \alpha_1 \Delta z_{t-1} + \alpha_2 \Delta z_{t-2} + \cdots + \alpha_p \Delta z_{t-p} + a_t .$$

Notice that this equation has an intercept term in it but no time trend. Again, the number of augmenting lags (p) is determined by minimizing the Schwartz Bayesian information criterion or minimizing the Akaike information criterion or lags are dropped until the last lag is statistically significant. EVIEWS allows all of these options for you to choose from. You then use the t-statistic on the θ coefficient to test whether you need to difference the data to make it stationary or not. Notice the test is **left-tailed**. The null hypothesis of the Augmented Dickey-Fuller t-test is

$$H_0 : \theta = 0 \qquad \text{(i.e. the data needs to be differenced to make it stationary)}$$

versus the alternative hypothesis of

$$H_1 : \theta < 0 \qquad \text{(i.e. the data is stationary and doesn't need to be differenced)}$$

c.   When the time series **has a trend in it (either up or down) and is potentially slow-turning around a trend line you would draw through the data**, use the following test equation:

$$\Delta z_t = \alpha_0 + \theta z_{t-1} + \gamma t + \alpha_1 \Delta z_{t-1} + \alpha_2 \Delta z_{t-2} + \cdots + \alpha_p \Delta z_{t-p} + a_t .$$

Notice that this equation has an intercept term **and** a time trend. Again, the number of augmenting lags (p) is determined by minimizing the Schwartz Bayesian information criterion or minimizing the Akaike information criterion or lags are dropped until the last lag is statistically

significant. EVIEWS allows all of these options for you to choose from. You then use the t-statistic on the θ coefficient to test whether you need to difference the data to make it stationary or you need to put a time trend in your regression model to correct for the variables deterministic trend. Notice the test is left-tailed.

The null hypothesis of the Augmented Dickey-Fuller t-test is

$$H_0 : \theta = 0 \qquad \text{(i.e. \textbf{the data needs to be differenced to make it stationary})}$$

versus the alternative hypothesis of

$$H_1 : \theta < 0 \qquad \text{(i.e. \textbf{the data is trend stationary and needs to be analyzed by means of using a time trend in the regression model instead of differencing the data})}$$

Just a note of caution: Sometimes if you have data that is **exponentially trending** then you might need to take the log of the data first before differencing it. In this case in your Dickey-Fuller unit root tests you will need to take the differences of the log of the series rather than just the differences of the series.