Trends and Unit Roots[1]

As we have discussed, the estimation techniques of this class require that each time series be stationary. Unfortunately, many time series do not possess this property. These notes thus demonstrate how we can test for and fix non-stationary time series. We consider two cases, the first is that of a deterministic trend:

Deterministic Trends:

Consider the following example:

$$y_t = \sum_{j=0}^{m} \delta_j t^j + x_t \tag{1}$$

where x_t is assumed to be a stationary time series. y_t thus consists of the sum of a stationary process and a separate summation that consists of time dependent terms. A special case of (1) is the case of a linear trend where $m = 1$:

$$y_t = \delta t + x_t \tag{2}$$

The inclusion of the summation, however, allows us to consider much more complex trends. The solution for dealing with a deterministic trend is simply to de-trend. This consists of subtracting off the trend. For example, suppose that we are dealing with the linear trend from (2):

$$\tilde{y}_t = y_t - \delta t = x_t \tag{3}$$

By assumption, x_t is stationary and $\tilde{y}_t$ therefore is as well. More generally, we must eliminate the entire trend:

$$\tilde{y}_t = y_t - \sum_{j=0}^{m} \delta_j t^j = x_t \tag{4}$$

In practice, it is difficult to know the true nature of the trend. One common approach is the *Hodrick-Prescott* filter. Suppose that we are sure that a non -stationary time series y_t has a trend, but we are unsure of its exact nature. The H-P filter splits the time series into two parts, the trend (x_t) and the detrended series ($y_t - x_t = \tilde{y}_t$):

[1] These are undergraduate lecture notes. They do not represent academic work. Expect typos, sloppy formatting, and occasional (possibly stupefying) errors.

$$min \left[\frac{1}{T} \sum_{t=1}^{T} (y_t - x_t)^2 + \frac{\lambda}{T} \sum_{t=2}^{T-1} [x_{t+1} - x_t) - (x_t - x_{t-1})]^2 \right] \tag{5}$$

This minimization problem is usually complex. Consider a pair of extreme cases:

1. $\lambda = 0$. Here, the minimization problem is solved by setting $y_t = x_t$. In other words, the trend is the series. We are thus assuming that fluctuations in the series are unrelated to random shocks.

2. $\lambda \to \infty$. Here the minimization problem is solved by setting δx_t equal to a constant. In words, the trend is linear.

The parameter λ thus captures the degree of non-linearity in the trend. In practice, it is usually calibrated to about $\lambda = 100$ for annual data.

If we successfully de-trend the time series, then the transformed variable will be stationary. We can then use the econometric techniques of this class, provided there are no other sources of misspecification. De-trending, for example, will not solve omitted variable bias.

Unit Roots

The other major source of non-stationarity is when an ARMA(p,q) process violates its stability conditions. In the case of an AR(1) this takes the form:

$$y_t = \alpha y_{t-1} + u_t \tag{6}$$

where, crucially, $|\alpha| > 1$. The solution to dealing with unit roots is to difference. Subtract y_{t-1} from both sides of (6):

$$\Delta y_t = (1 - \alpha) y_{t-1} + u_t \tag{7}$$

It can be shown that (7) is a stationary process if and only if $|(1 - \alpha)| < 1$. If so, then the first-difference of y_t is stationary and we can work with this time series instead of the original time series. The time series y_t is then said to be integrated of order 1 or I(1). A stationary time series is thus I(0).

Most common macroeconomic time series, such as prices and output, are non-stationary and must be differenced at least once. One cost of differencing is that the differenced series has one fewer observation than the original.

Suppose that $\alpha > 2$. In this case, (7) is also non-stationary. The solution is then to difference for a second time:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = (1-\alpha)y_{t-1} - \Delta y_{t-1} + u_t \tag{8}$$

If (8) is stationary, then the series is said to be I(2). In general, if we must difference d times to yield a stationary process, then the time series is said to be I(d). An ARMA(p,q) process that must be differenced d times to be stationary is also known as an ARIMA(p,d,q) process.

Mistakes with Non-Stationary Time Series

Here, we consider the consequences of getting it wrong.

#1: Doing Nothing. As discussed earlier, if we fail to deal with non-stationarity then our results are biased and probably not useful. There are some signs that suggest that we may have erroneously included such a variable in our specification. These include highly autocorrelated error terms and suspiciously high R^2s.

#2: De-trending when we should difference. Suppose that the data generating process is described by (6), a single unit root. But we mistakenly assume that the variable includes a linear trend (as in (2)) and we thus subtract off δt from (6):

$$\tilde{y}_t = y_{t-1} - \delta t + u_t \tag{9}$$

Notice that this does little to remedy the explosive dynamics of the time series. The transformed variable is thus still non-stationary and the consequences of this mistake are thus similar to #1.

#3. Differencing when you should de-trend. Now suppose that (2) is the true data generating process but we mistakenly first difference the time series instead:

$$\Delta y_t = \delta t + x_t - \delta(t-1) - x_{t-1} = \delta + x_t - x_{t-1} \tag{10}$$

If x_t is white noise, then we have transformed the time series into a stationary MA(1) process. We have thus fixed the bigger problem of non-stationarity, but added autocorrelation into the time series. This is problematic for the same reasons discussed in ECO 255. But because it does not bias the coefficients, just the standard errors, it it usually less serious than #2. For this reason, there is a tendency to difference rather than de-trend when in doubt.

#4 Overdifferencing: Now suppose that our time series is stationary but we mistakenly first difference it:

$$y_t = x_t \tag{11}$$

$$\Delta y_t = x_t - xt - 1 \tag{12}$$

The result is similar to #3 (unsurprising given that we are differencing in both cases when we should not). The process remains stationary it now exhibits autocorrrelation. But because autocorrelation does not cause bias, there is a tendency to difference when in doubt. But it is important not to automatically difference your variables because overdifferencing is a source of misspecification.

There are other ways to deal with non-stationarity. For example, when running a panel vector autoregression (PVAR), ordinary differencing causes bias. We thus must use a more complicated technique. But for now, we will limit our attention to differencing and de-trending.

Dickey-Fuller

There are a great many statistical tests for unit roots and trends. We will focus on the test is that typically appropriate for most economic time series, the (sometimes Augmented) Dickey Fuller Test.

Suppose that we are willing to impose that the time series follows an AR(1) process. This is known as imposing structure on the regression specification and may be reasonable in the face of convincing theoretical support. Write the process as:

$$y_t = \alpha + \beta t + \rho y_{t-1} + u_t \tag{13}$$

To run a Dickey Fuller test, we simply obtain the OLS regression coefficients for (13). It can also be shown, however, that the standard errors and autoreggressive coefficient are biased downward. Intuitively, suppose that $\rho = 1$ so that the time series is non stationary. We know that non stationarity leads to bias. So we can't use ordinary standard errors to reject a hypothesis that implies bias.

The null hypothesis is $H_0 : \rho = 1$, implying a unit root. The alternative is $H_a : |\rho| < 1$, implying stationarity. Dickey and Fuller (1979, 1981) derive unbiased standard errors for this test. So running the test is just like running an ordinary t-test but with different standard errors.

The Dickey Fuller critical values depend on the exact specification. Recall that for a one sided test that $\rho < 1$ at the 95% confidence level, the critical value is 1.65. Consider three cases:

1. If, as in (13), we include an intercept and trend. In this case, the critical value is 3.46. If we use unadjusted standard errors we will thus be too likely to reject unit roots that do exist and we will thus be too likely to include a non stationary variable in our specification.

2. Suppose, as is often the case in macroeconomics, we are willing to impose no trend so that $\beta = 0$. In this case, the critical value is 2.89.

3. Finally, suppose that are willing to impose both that $\beta = 0$ and $\alpha = 0$. In this case, the critical value is 1.94.

As long as we are willing to impose an AR(1) process then making the process stationary is straightforward. We run a Dickey Fuller test and then difference or de-trend as needed. If differencing, we repeat until the order of integration, d is determined and we must difference that many times.

If we are interested in testing for unit roots for a more general AR(p) process, then we can instead run an Augmented Dickey Fuller test. The specification for this test is:

$$\Delta y_t = \alpha + \beta t + (\rho - 1)y_{t-1} + \theta_1 \Delta y_{t-1} + \theta_2 \Delta y_{t-2} + ... + \theta_k \Delta y_{t-k} + u_t \qquad (14)$$

The first order of business is to choose k. This is done using an information criteria. Recall that an information criteria is a measure of fit (such as R^2) less a penalty for including additional variables. In general, this penalty is *ad-hoc*. Along literature examines which information criteria do best in different circumstances.

Conducting the test is otherwise similar to the Unaugmented Dickey Fuller test. We obtain the OLS regression coefficients. We then test for $H_0 : \rho = 1$ using Dickey Fuller critical values that depend on whether the specification includes a trend or constant. We then difference until we are able to reject a unit root.

Both (13) and (14) may exhibit autocorrelated error terms. This may be tested for and corrected in the standard manner.

Generating Dickey Fuller Critical Values

By design, this course focuses mostly on the implementation of time series techniques while spending relatively little time on the theoretical underpinnings. Example 5.1 in the textbook, however, provides a very easy example of how theoretical time series often works, in this case how it generates Dickey Fuller critical values. In applied work, of course, we cannot observe the true data generating process. We can, however, use theory to assume a specific process. Suppose that we assume that some time series, y_t has a unit root:

$$y_t = y_{t-1} + u_t \tag{15}$$

The authors then conduct 100,000 simulations of (15) for some distribution of u_t and $T = 200$. The authors then conduct a Dickey Fuller test for each of the 100,000 simulations. They find that the mean estimated autoregressive coefficient is 0.973. This is biased downward because it is less than the true value of one. A simple t-test is thus too likely to reject a unit root.[2] Dickey and Fuller correct for this bias by calculating higher critical values that make rejecting the hypothesis of a unit root harder.

Granger Causality

Econometricians are, of course, typically interested in establishing causality. This, however, is notoriously hard to do. Granger Causality represents a systematic attempt to establish causation. The textbook spends considerable time on this issue. We will not. I consider it nearly useless in economics.

Suppose that we have two time series, x_t, and y_t and that we have appropriately minimized omitted variable bias through the proper choice of controls. The general idea of Granger Causality is to regress x_t on y_{t-1} (or other lags) and vice-versa. If the regression coefficient on x_{t-1} is significant, then it is said to Granger Cause y_t and if the regression coefficient on y_{t-1} is significant, then it is said to Granger Cause x_t. Granger causation thus relies on simple temporal orderings. Granger causation depends on which variable moves first and it assumes that a variable in period t cannot cause a variable in time $t-1$.

This may be reasonable in other disciplines, especially some natural sciences. But expectations are crucial in economics and provide plausible mechanism where t can cause $t-1$. Consider some examples:

[2] If the confidence level is set at 10%, then then test should reject the null incorrectly no more than 10% of the time.

1. In early December, Christmas cards increase. Each December 25, Christmas occurs.[3] Christmas cards thus Granger cause Christmas.

2. Each summer salmon return to their birthplaces to spawn. Shortly thereafter, fall occurs. Spawning thus causes fall.

3. New housing permits often rise before the Fed announces lower interest rates. A stronger housing market thus Granger causes lower interest rates.

I chose the first two examples because they are ridiculous. In these cases, the latter event is entirely expected by households and fish respectively. Through this expectation, t clearly causes $t-1$. I chose #3 because this type of causality is not entirely implausible. But most changes to the Federal Funds rate are expected (at least in terms of the direction). So it is much more likely that builders are getting ready to increase building in expectation of lower rates. Because expectations are so important in economics, I do not consider Granger causality as useful in establishing real causation.

[3]Except for that year when the Grinch stole it.