DATA 5610

Time Series Notes IV

Tyler J. Brough

Department of Data Analytics & Information Systems

UtahStateUniversity

Beginning Time Series Topics IV

Unit Roots Continued

The random walk with drift

$$x_t = \mu + x_{t-1} + \varepsilon_t$$

where $\mu = E(x_t - x_{t-1}) = \mu$ and $\{\varepsilon_t\}$ is white noise. The constant term μ represents the time trend of x_t and is called the drift.

Assume the initial value of x_t is x_0, then

$$x_1 = \mu + x_0 + \varepsilon_1$$

$$x_2 = \mu + x_1 + \varepsilon_2 = 2\mu + x_0 + \varepsilon_1 + \varepsilon_2$$

$$\vdots$$

$$x_t = t\mu + x_0 + \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1$$

The last equation shows that $\{x_t\}$ consists of a time trend $t\mu$ and a pure random-walk process $\sum_{i=1}^{t} \varepsilon_i$.

$Var(\sum_{i=1}^{t} \varepsilon_i) = t\sigma_\varepsilon^2$ where σ_ε^2 is the variance of ε_t.

The conditional standard deviation of x_t is $\sqrt{t}\sigma_\varepsilon^2$, which grows at a slower rate than the conditional expectation of x_t. Therefore, if we graph x_t against the time index t, we have a time trend with slope μ

Let's look at some actual market data for IBM from 1947 to 1997.

Trend-Stationary Time Series

A closely related model that exhibits linear trend is the trend-stationary time series model:

$$x_t = \beta_0 + \beta_1 t + z_t$$

where z_t is a stationary time series (e.g. a stationary AR(p) series). Here x_t grows linearly in time with rate β_1 and hence can exhibit behavior similar to a random walk with drift.

There is one major difference between the random walk with drift and the trend-stationary series:

Random Walk with Drift

$$E(x_t) = x_0 + \mu t \quad \text{and}$$
$$Var(x_t = t\sigma_\varepsilon^2$$

which clearly is not stationary because the variance is directly time dependent. While

Trend-Stationary Series

$$E(x_t) = \beta_0 + \beta_1 t \quad \text{and}$$
$$Var(x_t = Var(z_t)$$

which is finite and time-independent.

General Unit-Root Nonstationary Models

Consider an ARMa model. If we extend the model by allowing the AR polynomial to have 1 as a characteristic root, then the model becomes the Autoregressive Integrated Moving Average (ARIMA) model.

An ARIMA model is said to be unit-root nonstationary because its AR polynomial has a unit root.

A conventional approach for handling unit-root nonstationarity is differencing.

Differencing

A time series x_t is said to be an ARIMA(p,1,q) process if the change series

$$c_t = x_t - x_{t-1} = (1 - L)x_t$$

follows a stationary and invertible ARMA(p,q) process.

Ex: in finance price series are commonly believed to be nonstationary, but the log-return series $r_t = \ln(p_t) - \ln(p_{t-1})$ is stationary. Here the price series $\{p_t\}$ is unit-root nonstationary and hence can be treated as an ARIMA process.

The idea of transforming a nonstationary series into a stationary one by considering its change series is called *differencing* in the time series literature.

Formally, $c_t = x_t - x_{t-1}$ is referred to as the first differenced series of x_t.

In some fields a time series x_t may contain multiple unit roots. For example, if both x_t and its first differenced series $c_t = x_t - x_{t-1}$ are unit-root nonstationary, but $s_t = c_t - c_{t-1} = x_t - 2x_{t-1} + x_{t-2}$ is weakly stationary, then x_t has double unit roots, and s_t is the second differenced series of x_t.

If s_t follows an ARMA(p,q) model then x_t is an ARIMA(p, 2, q) process.

Testing For Unit Roots

Q: Do economic variables such as GNP, employment, and interet rates tend to revert back to a long-run trend after a shock, or do they follow random walks?

The question is important for two reasons:

1. If these variables follow random walks, a regression of one against another can lead to spurious results.

For example, suppose two series are generated by independent random walks:

$$x_t = x_{t-1} + \epsilon_t$$

$$y_t = y_{t-1} + \nu_t$$

and $E(\epsilon_t \nu_t) = 0$ for all t, s.

Now suppose we run y_t on x_t by OLS

$$y_t = \alpha + \beta x_t + u_t$$

The assumptions underlying the CLRM are violated. In this case you tend to see "significant" β more often than the OLS formula say you should.

2. If affects our understanding of the economy and our ability to make forecasts:
    - If a variable such as GNP follows a random walk, then the effects of a temporary shock (e.g. increase in oil prices or an increase in government spending) not dissapate after several years but will instead have permanent effects.
    - If stock prices follow random walks they should not be forecastable.

Nelson & Plosser

NP found evidence that GNP and other macro variables behave like random walks. This spurred a huge literature to investigate whether or not economic and financial variables are random walks or are trend-reverting. Several of these studies show that many economic time series do appear to be random walks or at least have random walk components.

Most of these studies use unit-root tests introduced by Dicky & Fuller (1979) JASA.

Suppose we believe that a variable Y_t, which has been growing over time, can be described by the following equation:

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \epsilon_t$$

One possibility is that Y_t has been growing because Y_t has a positive time trend ($\beta > 0$) but would be stationary after detrending (i.e. $\rho < 1$) In this case Y_t could be used in a regression and all of the results and tests of the CLRM would apply.

Another possibility is that Y_t has been growing because it follows a random walk with a positive drift (i.e. $\alpha > 0$, $\beta = 0$, and $\rho = 1$). In this case we would need to work with ΔY_t (change series).

Detrending would not make the series stationary, and the inclusion of Y_t in a regression would lead to spurious results.

One might think that the equation could be estimated by OLS and that the t statistic on $\hat{\rho}$ could be used to test $H_0 : \rho = 1$. However, if the true value is indeed 1 then OLS would lead to spurious results, which could mean we could incorrectly reject the random walk hypothesis.

Dickey & Fuller derived the distribution for the estimator $\hat{\rho}$ that holds when $\rho = 1$ and generated statistics for an F-test of the random walk hypothesis, i.e. the hypothesis that $\beta = 0$ and $\rho = 1$.

The **Dickey-Fuller Test** works as follows, supposing

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \epsilon_t$$

First, using OLS run the (unrestricted) regression

$$Y_t - Y_{t-1} = \alpha\beta t + (\rho - 1) Y_{t-1}$$

and then the (restricted) regression

$$Y_t - Y_{t-1} = \alpha$$

Then calculate the F-ratio

$$F = \frac{(SSR_R - SSR_{UR})}{SSR_{UR}} \frac{N-k}{q}$$

where SSR_R is the sum of squared residuals of the restricted model and SSR_{UR} likewise for the unrestricted model. $(N-k)$ is the degrees of freedom of the unrestricted model and q is the number of restrictions placed on the restricted model.

This ratio is not distributed as a standard F distribution under the null hypothesis. Instead one must use the distributions tabulated by Dickey and Fuller.

Note: critical values from the Dickey-Fuller distribution are much larger than for the standard F-distribution.

The Augmented Dickey-Fuller Test

The original Dickey-Fuller test implicitly makes the assumption of no serial correlation in ϵ_t. Often we would like to allow for serial correlation in ϵ_t and still test for a unit root. This can be done with the augmented Dickey-Fuller test.

This test is carried out by extending the data-generating process (DGP) to include lagged changes in Y_t on the right-hand side:

$$Y_t = \alpha + \beta t + \rho Y_{t-1} + \sum_{j=1}^{p} \lambda_j \Delta y_{t-j} + \epsilon_t$$

where $\Delta Y_t = Y_t - Y_{t-1}$.

The unit-root test proceeds as before:

1. Using OLS, run the unrestricted regression

$$Y_t - Y_{t-1} = \alpha + \beta t + (\rho - 1) Y_{t-1} + \sum_{j=1}^{p} \lambda_j Y_{t-j}$$

2. And then the restricted regression

$$Y_t - Y_{t-1} = \alpha + \sum_{j=1}^{p} \lambda_j Y_{t-j}$$

3. Form the F-statistic to test if the restrictions hold ($\beta = 0$ and $\rho = 1$)

Phillips-Perron Test

Consider the following two regressions:

$$y_t = \mu + \alpha y_{t-1} + \epsilon_t \quad (*)$$

$$y_t = \mu + \beta(t - \frac{1}{2}T) + \alpha y_{t-1} + \epsilon_t \quad (**)$$