Testing for Unit Roots.

In the statistical literature it has long been known that unit root processes behave differently from stable processes.
For example in the scalar AR(1) model, consider the distribution of the OLS estimator of the parameter a in the simple first order process,

$$(1)\ y_t \ = \ a\ y_{t-1} \ + \ e_t \ .$$

If e_t are independently identically normally distributed (niid) variables and a_N denotes the least squares estimator of b based on $y_0, y_1, .., y_N$, then Mann and Wald (1943) showed that $N^{1/2}(a_N - a)$ has a limiting normal distribution if $a < 1$. White (1958) showed that $|a|\ N(a^2 - 1)\ (a_N - a)$ has a limiting Cauchy distribution if $a > 1$, whereas $N(a_N - 1)$ has a limiting distribution that can be written in terms of a ratio of two functionals of a Wiener process, when $a = 1$. In the later years a lot of theoretical work has been done on the distribution of least squares estimators in the presence of unit roots. Some notable early contributions are Fuller (1976), Dickey and Fuller (1979,1981), and Evans and Savin (1981,1984). The authoritative paper by P.C.B. Phillips (1987) sums up most of the theory.

It was the paper by Nelson and Plosser (1982) that sparked the huge surge in interest for unit root models among economists. They examined time series for some of the most important U.S. aggregate economic variables and concluded that almost all them were better described as being integrated of order one rather than stable. (They further went on to suggest that this favored real-business-cycle type of classical models in favor of monetarist and Keynesian models. In my opinion it is plain crazy to try and derive such sweeping conclusions from atheoretical time series modeling, and it is even crazier to do so on aggregate data; but maybe the huge interest that this paper generated is partly due to this provocative statement).

The reason why unit roots are so important is that the limiting distributions of estimates and test statistics are very different from the stationary case. Most importantly, one can not (in general) obtain limiting χ^2 (or t- or F-) distributions. Rather one obtains limiting distributions that can be expressed as functionals of Brownian motions. Also notice that in the case where you have a stable model that is "close" to an integrated process, then the distributions of estimators will look more like the distributions from unit root models than it will look like the asymptotic (normal type) distribution in small samples. This phenomenon is treated in the literature under the heading of near-integrated (or near unit-root, or nearly non-stationary) models. We may not have time to go into this; but the reading list contains a quite detailed bibliography (since I happened to have written a paper in that area - I had the references typed up already). I personally doubt whether many series in economics are best thought of as genuinely non-stationary, and I don't think that one really can decide that on the basis of the statistical evidence (there has been written lots of papers on that question since the influential paper of Nelson and Plosser (1982)). My point of view is that it does not really matter. The models will have very different predictions for the very long run whether they truly have unit roots or not; but to cite Lord Keynes: "In the long run we are all dead"; or to say it less dramatically - I do not think that simple time series models are useful for forecasting 20 years ahead under any circumstances. What matters is that the small sample distributions look like the asymptotic unit root distributions, so if you do not use those you will make wrong statistical inferences.

0.1 Brownian Motions and Stochastic Integrals.

The easiest way to think of the Brownian motions is in the following way (which corresponds exactly to the way that you will simulate Brownian motions on the computer): Let

$$B_N(t) \ = \frac{1}{\sqrt{N}} \ (\ e_1 \ + \ e_2 \ + \ ... \ + e_{[Nt]}) \ ; \ t \in [0, T] \ ,$$

where $e_1, ..., e_{[NT]}$ are iid $N(0, 1)$. The notation $[Nt]$ means the integer part of Nt, i.e. the largest integer less than of equal to Nt. Note that $B_N(t)$ is a stochastic function from the closed interval $[0, T]$ to the real numbers. If N is large $B_N(.)$ is a good approximation to the Brownian motion $B(t); t \in [0, T]$ which is defined as

$$B(t) \ = \ \lim_{N \to \infty} \ B_N(t) \ .$$

For a fixed value of t it is obvious that $B_N(t)$ converges to a normally distributed random variable with mean zero and variance t. To show that $B_N(t)$ converges as a function to a continuous function $B(t)$ takes a large mathematical apparatus. You can find that in the classical text of Billingsley (1968); but be warned that this is a book written for mathematicians (but given that it is very well written).

For the purpose of the present course this is all you need to know about the Brownian motion. One can show that any stationary continuous stochastic process $B(t)$ for which

$$B(t_4) - B(t_3) \ \text{ and } \ B(t_2) - B(t_1)$$

are independent for all $t_4 \geq t_3 \geq t_2 \geq t_1$ and with

$$E\{B(t_2) - B(t_1)\} \ = \ 0 \ , \ \text{ and } \ Var\{B(t_2) - B(t_1)\} \ = \ t_2 - t_1 \ ;$$

has to be a Brownian Motion. The Brownian motion is an example of a *process with identical independent increments*. You can convince yourself from the definition I gave of Brownian motion, that this formula for the variance is true. Notice that if you start the process at time 0 then

$$Var(B(t)) \ = \ t \ .$$

So it is obvious that the unconditional variance of $B(t)$ tends to infinity as t tends to infinity. This corresponds to the behavior of the discrete time random walk, which again is just an AR(1) with

an autoregressive coefficient of 1. So it is not surprising that Brownian motions show up in the (properly normalized) asymptotic distribution of estimators of AR models. Brownian motions are quite complicated if you look into some of the finer details of their sample paths. One can show that Brownian motions with probability 1 are only differentiable on a set of measure zero. You can also show that a Brownian motion that you start at zero at time zero will cross the x-axis infinitely many times in any finite interval that includes 0. Properties like those are very important in continuous time finance, so if you want to specialize in that field you should go deeper into the theory of Brownian motions. I have supplied a list of references in the reading list, that will be useful for that purpose; but for unit root theory this is not absolutely necessary.
You will also need to be able to integrate with respect to Brownian motions. We want to give meaning to the symbol $\int_0^1 f(s)dB(s)$, where f is a function that will often be stochastic itself. In many asymptotic formulae $f(s)$ is actually the same as $B(s)$. We will define the so-called Ito-integral (named after the Japanese mathematician Kiyosi Ito):

$$\int_0^1 f(s)dB(s) = \lim_{K\to\infty} \sum_{k=0}^{K} f(\frac{k-1}{K})\Delta B(\frac{k}{K}) \ ,$$

where $\Delta B(\frac{k}{K}) = B(\frac{k}{K} - B(\frac{k-1}{K})$, and where the limit is *in probability.* You can *not* obtain convergence sample path by sample path almost surely (with probability one). If we temporarily call the stochastic integral on the left hand side for I and the approximating sum on the right hand side for I_K then the convergence in probability means that there exists a stochastic variable I such that

$$\lim_{K\to\infty} P\{|I - I_K| > \epsilon\} = 0$$

for any given $\epsilon > 0$. This is however a probability statement and it does not preclude that I_K for a given sample path can be found arbitrarily far from I for arbitrarily large K. For our purpose that does not really matter, since convergence in probability is all we need. If you want to go through the mathematics (which is detailed in the probability books in the reading list, with the book by Øksendal as the most accessible), then the hard part is to show the existence of the limit I to which I_K converges, and if you don't want to go through the mathematics you should just take it for a fact that the limit exists in a well defined sense.

Notice that the sum in the approximating sum is over values of the function multiplied by a "forward looking" increment in the integrator B. This is essential and you will not get the right answer if you do not do it like that. This is in contrast to the standard Riemann-Stieltjes integral where this does not matter. The definition of the stochastic integral is given in the way that you can actually simulate the distribution, and this is the way it is done in many Monte Carlo studies in the literature.

Ito's Lemma

The main tool of modern stochastic calculus is Ito's lemma. It is usually formulated as

$$df(X,t) \ = \ \frac{\partial f}{\partial t}dt \ + \ \frac{\partial f}{\partial X}dX \ + \ \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(dX)^2 \ \text{ where } \ dt^2 \ = \ 0 \ , \text{and} \ \ dB^2 \ = \ dt \ .$$

In this course Ito's lemma is not so central; but you may meet the following equality

$$(*) \quad \int_0^1 B(s)dB(s) \;=\; \frac{1}{2}(B(1)^2 - 1)\;,$$

which is a simple consequence of Ito's lemma. Ito's lemma is essential in continuous time finance and in more advanced examinations of unit root theory.
I leave the proof of (*) for the homework.

Example:
Find $d\log(\log B)$. We do this is 2 stages. First we set $X = \log(B)$. Then (since $(dB)^2 = dt$)

$$dX \;=\; \frac{1}{B}dB \;-\; \frac{1}{2}\frac{1}{B^2}dt\;.$$

Then

$$\begin{aligned} d\log(\log B) &= \frac{1}{B}dX \;-\; \frac{1}{2}\frac{1}{B^2}(dX)^2 \\ &= \frac{1}{B}(\frac{1}{B}dB \;-\; \frac{1}{2B^2}dt) \;-\; \frac{1}{2B^2}\frac{1}{B^2}dt \\ &= \frac{1}{B^2}dB \;-\; (\frac{1}{2B^3} \;+\; \frac{1}{2B^4})dt \end{aligned}$$

(I cannot think of an application of this particular result but it illustrates the method clearly.)

Now consider the process (1) again. Notice that if $a = 1$ then

$$y_t \;=\; y_{t-1} \;+\; e_t \;=\; \sum_{k=1}^{t} e_k\;.$$

We will consider the least squares estimator

$$\hat{a} \;=\; \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} \;=\; \frac{\sum_{t=1}^T (y_{t-1} + \Delta y_t) y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} \;=\; 1 \;+\; \frac{\sum_{t=1}^T y_{t-1}\Delta y_t}{\sum_{t=1}^T y_{t-1}^2}$$

Now notice that this implies that

$$T(\hat{a} \;-\; 1) \;=\; T\frac{\sum_{t=1}^T y_{t-1}\Delta y_t}{\sum_{t=1}^T y_{t-1}^2} \;=\; \frac{\sum_{t=1}^T (y_{t-1}/\sqrt{T})(e_t/\sqrt{T})}{\sum_{t=1}^T (y_{t-1}/\sqrt{T})^2 \frac{1}{T}}$$

One can now see from the way that we defined the Brownian motion that $y_{t-1}/\sqrt{T}$ converges to a Brownian motion, and from the way that we defined the stochastic integral, one can see (at least one would guess) that

$$T(\hat{a} \;-\; 1) \;\Rightarrow\; \frac{\int_0^1 B(s)dB(s)}{\int_0^1 B(s)^2 ds}\;.$$

Intuitively you should always think of e_t as $dB(t)$ and y_t which under the null of a unit root is equal to $\sum_{k=0}^t e_t$ corresponds then to $\int_0^{t/T} dB(s) = B(s/T)$ (for $B(0) = 0$). From our application of Ito's lemma one can see that another expression would be

$$T(\hat{a} \;-\; 1) \;\Rightarrow\; \frac{1}{2}\frac{B(1)^2 - 1}{\int_0^1 B(s)^2 ds}\;.$$

Notice that $B(1)$ is just a standard normal distribution, so that $B(1)^2$ is just a standard $\chi^2(1)$ distributed random variable. Contrary to the stable model the denominator of the expression for the least squares estimator does not converge to a constant almost surely; but rather to a stochastic variable that is strongly correlated with the numerator. For these reasons the asymptotic distribution does not look like a normal, and it turns our that the limiting distribution of the least squares estimator is highly skewed, with a long tail to the left - look at the graph of the distribution in e.g. Evans and Savin (1981).

OK, if you couldn't quite follow the "derivation" of the limiting distribution, don't despair. One needs quite a bit more machinery and notation to make sure that all the limit operations are legal; but all you need is the basic intuition, so try and get that. Most of the part of the unit root literature that is concerned with asymptotic theory contains the limiting distribution given here. We will refer to the distribution of

$$\frac{\int_0^1 B(s)dB(s)}{\int_0^1 B(s)^2 ds}$$

as *the unit root distribution*. It is not possible to find any simple expression for the density of this distribution but one can find the characteristic function, which can be inverted in order to tabulate the distribution function - (see Evans and Savin (1981)). You can also evaluate the distribution by Monte Carlo simulation, which is performed by choosing a large value of T, and then drawing the innovation terms from a pseudo random number generator (this is very easy in GAUSS) and the generating the series y_t from the defining equation (1). For large T the distribution of the LS estimator is close to the limiting distribution, which can be graphed by repeating this exercise like 10- or 20,000 times and plotting the result.

TS and DS models

If you look at a plot of a typical macro economic time series, like real GNP, it is obvious that it displays are very pronounced *trend*. What is a trend? Well, for many years that question was not considered for many seconds - a trend was simply assumed to be linear function of time, and econometricians would routinely "detrend" their series by using the residuals from a regression on time (and a constant) rather than the original series. This practice was challenged by the Box-Jenkins methodology, which became somewhat popular in economics in the seventies, although it originated from engineering. The Box-Jenkins methodology had as one of its major steps the "detrending" of variables by the taking of differences. In the 80ies a major battle between thes two approaches raged, with the difference-detrenders seemingly having the upper hand in the late 80ies, although challenged from many sides - the Bayesians being the most aggressive.

During that period the following terminology took hold. The model

$$(DS)\ y_t \ = \ \mu \ + \ \ y_{t-1} \ + \ e_t$$

is called Difference Stationary (DS) since it is stationary after the application of the differencing operation, and the model and

$$(TS)\ y_t \ = \ \mu \ + \ \beta t \ + \ a\, y_{t-1} \ + \ e_t \ ; \quad a < 1 \ ,$$

is called Trend Stationary (TS) since it is stationary after good old-fashioned detrending by regressing on a time-trend. Most tests for unit roots are formulated as testing TS versus DS.

0.2 Unit Root tests

0.2.1 Dickey-Fuller tests

The most famous of the unit root tests are the ones derived by Dickey and Fuller and described in Fuller (1976).
Dickey and Fuller considered the estimation of the parameter a from the models

$$(1) \ \ y_t \ = \ \rho\, y_{t-1} \ + \ e_t \ ,$$

$$(2) \ \ y_t \ = \ \mu \ + \ \rho\, y_{t-1} \ + \ e_t \ .$$

and

$$(3) \ \ y_t \ = \ \mu \ + \ \beta t \ + \ \rho\, y_{t-1} \ + \ e_t \ .$$

The parameter ρ is just the AR-parameter that we have denoted by a so far; but here I use the notation of Fuller (1976). It is assumed that $y_0 = 0$.

The simplest Dickey-Fuller test is simply to estimate (1) by least squares and compare $T(\hat{\rho} - 1)$ to a table of the distribution derived from a Monte Carlo study (or, as shown in Evans and Savin (1981), one can find the characteristic function and invert it). This test is sometimes known as the Dickey-Fuller ρ-test. The critical values for this and the following Dickey-Fuller (DF) tests can be found in Fuller (1976), p. 373. Simplified versions of the tables from Fuller can be found many places, f. ex. in Maddala (1992). (The Monte Carlo simulations were actually done as part of David Dickey's Ph.D. thesis). In practice the model (1) is often too simple and one would like to allow for a mean term in the estimation. Dickey and Fuller suggested that one estimates (2) by first calculating the average $\bar{y}$ of the observations $y_2, ..., y_T$ and the average $\bar{y}_0$ of $y_1, ..., y_{T-1}$ and then calculates the least squares estimator $\hat{\rho}_\mu$ by regressing $y_t - \bar{y}$ on $y_{t-1} - \bar{y}_0$. Comparing $T(\hat{\rho}_\mu - 1)$ to the critical values in Fuller is known as the Dickey-Fuller ρ_μ-test. Dickey and Fuller also suggested estimating $\hat{\rho}_\tau$ from model (3) by a standard regression and they tabulated the critical values of $T(\hat{\rho}_\tau - 1)$ under the composite null hypothesis $\rho = 1$ and $\beta = 0$.
Note that this last test is a test for the DS model against the TS model, and it is known as Dickey-Fuller ρ_τ-test.

It is often not realistic that a data series should follow as simple a model as (1), (2), or (3) with iid error terms. In most economic data series there will also be substantial short term autocorrelations, so it may be more reasonable to assume the model

$$(4) \quad y_t \;=\; a_1 y_{t-1} \;+ ... +\; a_p y_{t-p} \;+\; e_t \;,$$

where e_t is iid normal, rather than model (1) (and equivalent for models (2) and (3)). An equivalent way of writing model (4) is

$$(4') \quad \Delta y_t \;=\; \theta_1 y_{t-1} \;+\; \theta_2 \Delta y_{t-1} + ... +\; \theta_p \Delta y_{t-p+1} \;+\; e_t \;.$$

It is simple to show the equivalence of (4) and (4'). For 2 lags, we can subtract y_{t-1} from both sides and get

$$\Delta y_t \;=\; (a_1 - 1) y_{t-1} \;+\; a_2\, y_{t-2} \;+\; e_t \;.$$

By adding and subtracting $a_2\, y_{t-1}$ we get

$$\Delta y_t \;=\; (a_2 + a_1 - 1)\, y_{t-1} \;-\; a_2\,(y_{t-1} - y_{t-2}) \;+\; e_t \;.$$

The null of a unit root means that unity is a root in $1 - a_1 * z - a_2 z^2$ or $1 - a_1 - a_2 = 0$ which obviously implies $a_2 + a_1 - 1 = 0$; i.e., the θ_1 term in (4') is nill. This calculation generalize easily to more lags, so a unit root in (4) will show up as $\theta_1 = 0$ in (4'). It turns out that if you instead of using the distribution of the coefficient you use the distribution of the t-statistic, then the test will *not* depend on the form of the autoregression. You can find critical values for the t-values under the null of a unit root in many places.