More on Unit Root Tests

Given a simple AR(1) process:

$$y_t = \rho y_{t-1} + x_t'\delta + \varepsilon_t ,$$

the Dickey-Fuller test is simply the result of some algebra:

$$\Delta y_t = \alpha y_{t-1} + x_t'\delta + \varepsilon_t , \text{ where } \alpha = \rho - 1.$$

Dickey and Fuller (1979) showed that under the null hypothesis of a unit root, the t-statistic for α does not follow a conventional t-distribution, and derive asymptotic results and simulate critical values for various test and sample sizes.

McKinnon (1991, 1996) implements a much larger set of simulations than those given by Dickey and Fuller. Moreover, he estimates response surfaces for the simulation results, permitting the calculation of DF critical values and p-values for arbitrary sample sizes. It is McKinnon’s values that are most commonly used now.

The Augmented Dickey-Fuller test constructs a parametric correction for higher-order correlation by assuming that the y series follows an AR(p) process, and adds p lagged differences of y to the RHS of the test regression.

This raises the problem of choosing the number of lags p. This is done by a variety of tests. In practical terms, you'd like to add enough terms so that the errors are white noise. Tests for optimal lag lengths that are used include:

- Schwartz Information Criteria
- Akaike Information Criteria
- Hannan-Quinn Criteria
- Modified forms of these criteria

You must also decide whether or not to include a time trend or constant term. Again the issue is to take all the information out of the residuals, to leave them white noise.

Another related test is the Dickey-Fuller test with GLS detrending (DFGLS). It involves detrending y with a GLS technique and substituting the detrended y into the ADF test.

The Phillips-Perron (PP) test offer an alternative method for correcting for serial correlation in unit root testing. Basically, they use the standard DF or ADF test, but modify the t-ratio so that the serial correlation does not affect the asymptotic distribution of the test statistic.

In the PP test, you have to decide whether or not to include a constant and/or time trend. You also have to choose a method for computing an estimator of the residual spectrum at frequency zero. This is often done by a sum-of-covariances approach or an autoregressive spectral density estimation.

The Kwiatkowski, Phillips, Schmidt, and Shin (or KPSS test (1992) uses a null hypothesis that the series is trend stationary. Once again, this requires an estimator of the residual spectrum at frequency zero, and a set of exogenous regressors.

The Elliot, Rothenberg, and Stock (ERS) test and the Ng and Perron (NP) test also require frequency zero spectrum estimates and exogenous regressors.

Perron (1988) suggests running an array of tests because some tests results are biased under certain structures.

Table 3.2 Perron's (1988) testing procedure using the DF test (unknown d.g.p.).

Step and model	Null hypothesis	Test statistic	Critical values*
(1) $\Delta y_t = \mu_c + \gamma_c t + (\rho_c - 1)y_{t-1} + u_t$	$(\rho_c - 1) = 0$	τ_τ	Fuller (table 8.5.2, block 3)
(2) $\Delta y_t = \mu_c + \gamma_c t + (\rho_c - 1)y_{t-1} + u_t$	$(\rho_c - 1) = \gamma_c = 0$	Φ_3	Dickey and Fuller (table VI)
(2a) $\Delta y_t = \mu_c + \gamma_c t + (\rho_c - 1)y_{t-1} + u_t$	$(\rho_c - 1) = 0$	t	Standard normal
(3) $\Delta y_t = \mu_b + (\rho_b - 1)y_{t-1} + u_t$	$(\rho_b - 1) = 0$	τ_μ	Fuller (table 8.5.2, block 2)
(4) $\Delta y_t = \mu_b + (\rho_b - 1)y_{t-1} + u_t$	$(\rho_b - 1) = \mu_b = 0$	Φ_1	Dickey and Fuller (table IV)
(4a) $\Delta y_t = \mu_b + (\rho_b - 1)y_{t-1} + u_t$	$(\rho_b - 1) = 0$	t	Standard normal
(5) $\Delta y_t = (\rho_a - 1)y_{t-1} + u_t$	$(\rho_a - 1) = 0$	τ	Fuller (table 8.5.2, block 1)

* Fuller (1976) and Dickey and Fuller (1981).

The issue is that if the testing procedure contains more deterministic features in it than are present in the data generating process, the results become biased. Under some of these situations, the test statistic becomes better described as asymptotically normal, and needs to be tested as such.

In finite samples, it has been shown that any trend-stationary process can be approximated arbitrarily well by a unit root process in the sense that the autocovariance structures and spectra will be arbitrarily close. Likewise, any unit root process can be approximated by a trend-stationary process when the sample size is small. (Campbell and Perron, 1991).

This implies that a unit root test with high power against any stationary alternative will necessarily have a high probability of false rejection of a unit root when applied to near stationary processes. A near stationary process is one whose properties are closer to stationary white noise than to a nonstationary random walk.