Lecture 25 – Unit Root Tests III
The Ng-Perron Unit Root Test
(Improved Finite Sample Performance)

The ADF and PP unit root tests are known (from MC simulations) to suffer potentially severe finite sample power and size problems.

1. Power – The ADF and PP tests are known to have low power against the alternative hypothesis that the series is stationary (or TS) with a large autoregressive root. (See, e.g., DeJong, et al, *J. of Econometrics*, 1992.)

2. Size – The ADF and PP tests are known to have severe size distortion (in the direction of over-rejecting the null) when the series has a large negative moving average root. (See, e.g., Schwert. *JBES*, 1989: MA = -0.8, size = 100%!)

A variety of alternative procedures have been proposed that try to resolve these problems, particularly, the power problem, but the ADF and PP tests continue to be the most widely used unit root tests. That may be changing.
Digression –

Recall that there are two ways of thinking about the roots of the polynomial associated with the MA or AR component of a time series. Suppose,

\[ y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid} \]

Sometimes we define the roots of the AR as the solutions to

\[ 1 - a_1 z - \ldots - a_p z^p = 0 \]

(and, for the AR process, the stationarity condition is these roots are greater than one in modulus)

Sometimes we define the roots of the AR as the solutions to

\[ z^p - a_1 z^{p-1} - \ldots - a_p = 0 \]

(and, for the AR process, the stationarity condition is these roots are less than one in modulus)

(How are the roots in these two approaches related to one another?)
When we talk about $y_t$ having a large AR root we are thinking in terms of the second approach and referring to the largest root being close to one.

Similarly, we can calculate the roots of an MA polynomial, $b(L)$, in either of these two ways. When we talk about $y_t$ having a large MA root we are thinking in terms of the second approach and referring to the largest root being close to one.

Suppose, for example,

$$y_t = \varepsilon_t - 0.9\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{iid} (0, \sigma^2)$$

The MA root defined by the first approach is $1/0.9 = 1.1$. The MA root defined by the second approach is $0.9$, a large positive root.

This process can be approximated by an AR($p$) for sufficiently large $p$ with AR coefficients: $0.9, 0.9^2, \ldots$ Note that the larger the MA root, the larger the value of $p$ that will be needed for a good approximation.
Suppose,

\[ y_t = \varepsilon_t + 0.9\varepsilon_{t-1} \]

The MA root (defined by the second approach) is -0.9, a large negative root.

This process can be approximated by an AR(p) for sufficiently large p with AR coefficients: -0.9, (-0.9)^2,... Note, again, that the larger the MA root, the larger the value of p that will be needed for a good approximation.

Among the macroeconomic time series whose differences are thought to be characterized by a large negative MA root:

- inflation rate, unemployment rate

Insert Table 7 from Ng and Perron

Insert Table 1 from Phengpis and Crowder

End Digression
Ng and Perron *(Econometrica, 2001)*, building on some of their own work (Perron and Ng, *Rev. of Econ. Studies*, 1996) and work by Elliott, Rothenberg, and Stock (*Ecta*, 1996), new tests to deal with both of these problems. Their tests, in contrast to many of the other “new” unit root tests that have been developed over the years, seems to have caught on as a preferred alternative to the traditional ADF and PP tests.

(Note that for a test to become widely used, it must work well and be relatively easy to apply.)
The family of NP tests (which includes among others, modified DF and PP test statistics) share the following features –

First, the time series is de-meaned or detrended by applying a GLS estimator. This step turns out to improve the power of the tests when there is a large AR root and reduces size distortions when there is a large negative MA root in the differenced series.

The second feature of the NP tests is a modified lag selection (or truncation selection) criteria. It turns out that the standard lag selection procedures used in specifying the ADF regression (or for calculating the long run variance for the PP statistic) tend to underfit, i.e., choose too small a lag length, when there is a large negative MA root. This creates additional size distortion in unit root tests. The NP modified lag selection criteria accounts for this tendency.
Example – NP’s DF-GLS test using a modified AIC (MAIC) to select the lag length for the ADF regression.

\[ H_0: \, Y_t \sim I(1), \text{ no drift} \]
\[ H_A: \, Y_t \sim I(0), \text{ no restriction on the mean} \]

\( Y_t = \text{inflation rate, unemployment rate, interest rate, \ldots} \)

Step 1 – De-mean the data by applying GLS using a “local-to-unity” transformation:

\[
Y_1^* = Y_1, \\
Y_t^* = Y_t - (1-7/T)Y_{t-1} \text{ for } t = 2,\ldots,T \\
1_1^* = 1 \\
1_t^* = 1 - (1-7/T) \text{ for } t = 2,\ldots,T
\]

Regress \( Y_t^* \) on \( 1_t^* \), bhat; \( y_t = Y_t - b\text{hat} \).
Step 2 – Find the optimal lag length, p, for the ADF regression of $y_t$ on $y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$. (Note that there will not be an intercept in the ADF regression. Why not?)

- Select $p_{\text{max}}$
- Fit the ADF regression for $p = 1, \ldots, p_{\text{max}}$ (holding the sample size fixed: $\hat{\rho}_p, \hat{\epsilon}_{t,p}$)
- Choose $p$ to minimize the MAIC:

$$\text{MAIC}(p) = \log \hat{\sigma}_p^2 + \frac{2(r_p + p)}{T - p_{\text{max}}}$$

where

$$\hat{\sigma}_p^2 = \frac{1}{T - p_{\text{max}}} \sum_{p_{\text{max}}+1}^{T} \hat{\epsilon}_{t,p}^2$$

$$r_p = (\hat{\rho}_p - 1)^2 * \sum_{p_{\text{MAX}}+1}^{T} y_{t-1}^2 / \hat{\sigma}_p^2$$
• Fit the ADF regression (no intercept or trend) to $y_t$ using the optimal $p$:

$$\rho\text{-hat, } \tau_{\mu}^{\text{ADF-GLS}}$$

where $\tau_{\mu}^{\text{ADF-GLS}}$ is the t-statistic associated with $H_0: \rho = 1$ in the ADF regression (or $H_0: \rho = 0$ in the regression of $\Delta y_t$ on $y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$).

• The asymptotic null distribution of $\tau_{\mu}^{\text{ADF-GLS}}$ is the DF distribution for $\tau_{\mu}$. 
Note – If the original series $Y_t$ is assumed to have a linear deterministic trend component under $H_0$ and $H_A$, the procedure above is modified as follows:

1. Revise Step 1 – De-mean the data by applying GLS using a “local-to-unity” transformation:

$$Y_1^* = Y_1,$$

$$Y_t^* = Y_t - (1-13.5/T)Y_{t-1} \text{ for } t = 2,\ldots,T$$

$$1_1^* = 1$$

$$1_t^* = 1 - (1-13.5/T) \text{ for } t = 2,\ldots,T$$

$$t_1^* = 1$$

$$t_t^* = t-(1-13.5/T)*(t-1) \text{ for } t = 2,\ldots,T$$

Regress $Y_t^*$ on $1_t^*, t_t^*$: $bhat; y_t = Y_t - bhat$. 
2. Revise Step 2

- Fit the ADF regression (no intercept or trend) to $y_t$ using the optimal $p$:

  $$\rho\text{-hat, } \tau_t^{ADF-GLS}$$

  where $\tau_t^{ADF-GLS}$ is the t-statistic associated with $H_0: \rho = 1$ in the ADF regression (or $H_0: \rho = 0$ in the regression of $\Delta y_t$ on $y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$).

- The asymptotic null distribution of $\tau_t^{ADF-GLS}$ is not the DF distribution for $\tau_t$. Percentiles of the asymptotic null distribution of $\tau_t^{ADF-GLS}$ are provided in NP, Table 1 (See the $p = 1$ case.)