# 1 Unit Root Tests

Consider the trend-cycle decomposition of a time series $y_t$

$$y_t = TD_t + TS_t + C_t = TD_t + Z_t$$

The basic issue in unit root testing is to determine if $TS_t = 0$. Two classes of tests, called unit root tests, have been developed to answer this question:

- $H_0 : TS_t \neq 0$ ($y_t \sim I(1)$) vs. $TS_t = 0$ ($y_t \sim I(0)$)

- $H_0 : TS_t = 0$ ($y_t \sim I(0)$) vs. $TS_t \neq 0$ ($y_t \sim I(1)$)

## 1.1 Autoregressive unit root tests

These tests are based on the following set-up. Let

$$y_t = \phi y_{t-1} + u_t,\ u_t \sim I(0)$$

The null and alternative hypothesis are

$$\begin{aligned} H_0 &: \phi = 1 \text{ } (\phi(z) = 0 \text{ has a unit root}) \\ H_1 &: |\phi| < 1 \text{ } (\phi(z) = 0 \text{ has root outside unit circle}) \end{aligned}$$

The most popular of these tests are the Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test. The ADF and PP tests differ mainly in how they treat serial correlation in the test regressions.

1. ADF tests use a parametric autoregressive structure to capture serial correlation

$$\begin{aligned} \phi^*(L)u_t &= \varepsilon_t \\ \phi^*(L) &= 1 - \phi_1^* L - \cdots - \phi_k^* L^k \end{aligned}$$

2. PP tests use non-parametric corrections based on estimates of the long-run variance of $\Delta y_t$.

## 1.2 Stationarity Tests

These tests can be interpreted in two equivalent ways. The first is based on the Wold representation

$$\Delta y_t = \psi^*(L)\varepsilon_t,\ \varepsilon_t \sim \text{iid}(0, \sigma^2)$$

The null and alternative hypotheses are

$$\begin{aligned} H_0 &: \psi^*(1) = 0 \ (\psi^*(z) = 0 \text{ has a unit root}) \\ H_1 &: \psi^*(1) > 0 \ (\psi^*(z) = 0 \text{ has roots outside unit circle}) \end{aligned}$$

The second is based on the UC-ARIMA model

$$\begin{aligned} y_t &= \mu_t + C_t \\ \mu_t &= \mu_{t-1} + \varepsilon_t,\ \varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2) \\ \phi(L)C_t &= \theta(L)\eta_t,\ \eta_t \sim \text{iid}(0, \sigma_\eta^2) \\ cov(\varepsilon_t, \eta_t) &= 0 \end{aligned}$$

Here, the null and alternative hypotheses are

$$\begin{aligned} H_0 &: \sigma_\varepsilon^2 = 0 \ (\mu_t = \mu_0) \\ H_1 &: \sigma_\varepsilon^2 > 0 \ (\mu_t = \mu_0 + \sum_{j=1}^{t} \varepsilon_j) \end{aligned}$$

Result: Testing for a unit moving average root in $\psi^*(L)$ is equivalent to testing $\sigma_\varepsilon^2 = 0$.

Intuition: Recall the random walk plus noise model. The reduced form is an MA(1) model with moving average root given by

$$\begin{aligned} \theta &= \frac{-(q+2)+\sqrt{q^2+4q}}{2} \\ q &= \frac{\sigma_\varepsilon^2}{\sigma_\eta^2} \end{aligned}$$

If $\sigma_e^2 = 0$ then $q = 0$, $\theta = -1$ and the reduced form MA(1) model has a unit moving average root.

The most popular stationarity tests are the Kitawoski-Phillips-Schmidt-Shin (KPSS) test and the Leyborne-McCabe test. As with the ADF and PP tests the KPSS and Leyborne-McCabe tests differ main in how they treat serial correlation in the test regressions.

## 1.3 Statistical Issues with Unit Root Tests

Conceptually the unit root tests are straightforward. In practice, however, there are a number of difficulties:

- Unit root tests generally have nonstandard and non-normal asymptotic distributions.

- These distributions are functions of standard Brownian motions, and do not have convenient closed form expressions. Consequently, critical values must be calculated using simulation methods.

- The distributions are affected by the inclusion of deterministic terms, e.g. constant, time trend, dummy variables, and so different sets of critical values must be used for test regressions with different deterministic terms.

## 1.4 Distribution Theory for Unit Root Tests

Consider the simple AR(1) model

$$y_t = \phi y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

The hypotheses of interest are

$$\begin{aligned} H_0 &: \phi = 1 \text{ (unit root in } \phi(z) = 0) \Rightarrow y_t \sim I(1) \\ H_1 &: |\phi| < 1 \Rightarrow y_t \sim I(0) \end{aligned}$$

The test statistic is

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{\text{SE}(\hat{\phi})}$$
$$\hat{\phi} = \text{least squares estimate}$$

If $\{y_t\}$ is stationary (i.e., $|\phi| < 1$) then

$$\begin{aligned} &\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} N(0, (1 - \phi^2)) \\ &\hat{\phi} \overset{A}{\sim} N\left(\phi, \frac{1}{T}(1 - \phi^2)\right) \\ &t_{\phi=\phi_0} \overset{A}{\sim} N(0, 1) \end{aligned}$$

However, under the null hypothesis of nonstationarity the above result gives

$$\hat{\phi} \overset{A}{\sim} N(1, 0)$$

which clearly does not make any sense.

Problem: under the unit root null, $\{y_t\}$ is not stationary and ergodic, and the usual sample moments do not converge to fixed constants. Instead, Phillips (1987) showed that the sample moments of $\{y_t\}$ converge to random functions of Brownian motion:

$$T^{-3/2} \sum_{t=1}^{T} y_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r)dr$$

$$T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr$$

$$T^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \xrightarrow{d} \sigma^2 \int_0^1 W(r)dW(r)$$

where $W(r)$ denotes a standard Brownian motion (Wiener process) defined on the unit interval.

A Wiener process $W(\cdot)$ is a continuous-time stochastic process, associating each date $r \in [0, 1]$ a scalar random variable $W(r)$ that satisfies:

1. $W(0) = 0$

2. For any dates $0 \leq t_1 \leq \cdots \leq t_k \leq 1$ the changes $W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_k) - W(t_{k-1})$ are independent normal with

$$W(s) - W(t) \sim N(0, (s - t))$$

3. $W(s)$ is continuous in $s$.

Intuition: A Wiener process is the scaled continuous-time limit of a random walk

Using the above results Phillips showed that under the unit root null $H_0 : \phi = 1$

$$T(\hat{\phi} - 1) \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}$$

$$t_{\phi=1} \xrightarrow{d} \frac{\int_0^1 W(r)dW(r)}{\left(\int_0^1 W(r)^2 dr\right)^{1/2}}$$

For example,

$$\begin{aligned}
\hat{\phi} - 1 &= \left(\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} \sum_{t=1}^{T} y_{t-1}\varepsilon_t \\
&\Rightarrow T(\hat{\phi} - 1) = \left(T^{-2}\sum_{t=1}^{T} y_{t-1}^2\right)^{-1} T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \\
&\xrightarrow{d} \left(\int_0^1 W(r)^2 dr\right)^{-1} \int_0^1 W(r)dW(r)
\end{aligned}$$

Phillips' derivations yield some surprising results:

- $\hat{\phi}$ is *super-consistent*; that is, $\hat{\phi} \xrightarrow{p} \phi$ at rate $T$ instead of the usual rate $T^{1/2}$.

- $\hat{\phi}$ is not asymptotically normally distributed, and $t_{\phi=1}$ is not asymptotically standard normal.

- The limiting distribution of $t_{\phi=1}$ is called the *Dickey-Fuller* (DF) distribution and does not have a closed form representation. Consequently, quantiles of the distribution must be computed by numerical approximation or by simulation.

- Since the *normalized bias* $T(\hat{\phi} - 1)$ has a well defined limiting distribution that does not depend on nuisance parameters it can also be used as a test statistic for the null hypothesis $H_0 : \phi = 1$.

**Quantiles of standard normal and DF distribution**

```
> qnorm(c(0.01,0.05,0.10))
[1] -2.326 -1.645 -1.282

> qunitroot(c(0.01,0.05,0.10), trend="nc")
[1] -2.565 -1.941 -1.617
```

The usual one-sided 5% critical value for standard normal is $-1.645$

The one-sided 5% critical value for the DF distribution is $-1.941$

Note: $-1.645$ is the 9.45% quantile of the DF distribution

## 1.5 Trend Cases

When testing for unit roots, it is crucial to specify the null and alternative hypotheses appropriately to characterize the trend properties of the data at hand.

- If the observed data does not exhibit an increasing or decreasing trend, then the appropriate null and alternative hypotheses should reflect this.

- The trend properties of the data *under the alternative hypothesis* will determine the form of the test regression used.

- The type of deterministic terms in the test regression will influence the asymptotic distributions of the unit root test statistics.

### 1.5.1 Case I: Constant Only

The test regression is

$$y_t = c + \phi y_{t-1} + \varepsilon_t$$

and includes a constant to capture the nonzero mean under the alternative. The hypotheses to be tested are

$$\begin{aligned} H_0 &: \phi = 1, c = 0 \Rightarrow y_t \sim I(1) \text{ without drift} \\ H_1 &: |\phi| < 1 \Rightarrow y_t \sim I(0) \text{ with nonzero mean} \end{aligned}$$

This formulation is appropriate for non-trending economic and financial series like interest rates, exchange rates, and spreads.

The test statistics $t_{\phi=1}$ and $T(\hat{\phi}-1)$ are computed from the above regression. Under $H_0 : \phi = 1, c = 0$ the asymptotic distributions of these test statistics are influenced by the presence, but not the coefficient value, of the constant in the test regression:

$$\begin{aligned} T(\hat{\phi}-1) &\Rightarrow \frac{\int_0^1 W^{\mu}(r)dW(r)}{\int_0^1 W^{\mu}(r)^2 dr} \\ t_{\phi=1} &\Rightarrow \frac{\int_0^1 W^{\mu}(r)dW(r)}{\left(\int_0^1 W^{\mu}(r)^2 dr\right)^{1/2}} \end{aligned}$$

where

$$W^{\mu}(r) = W(r) - \int_0^1 W(r)dr$$

is a "de-meaned" Wiener process. That is,

$$\int_0^1 W^{\mu}(r) = 0$$

Note: derivation requires special trick from Sims, Stock and Watson (1989) ECTA.

## Quantiles of DF Distribution with Constant

Note: inclusion of a constant pushes the distributions of $t_{\phi=1}$ and $T(\hat{\phi}-1)$ to the left:

```
> qunitroot(c(0.01,0.05,0.10), trend="c")
[1] -3.430 -2.861 -2.567
> qunitroot(c(0.01,0.05,0.10), trend="c",
+           statistic="n")
[1] -20.62 -14.09 -11.25
```

Note: $-1.645$ is the 45.94% quantile of the DF$^{\mu}$ distribution!

### 1.5.2 Case II: Constant and Time Trend

The test regression is

$$y_t = c + \delta t + \phi y_{t-1} + \varepsilon_t$$

and includes a constant and deterministic time trend to capture the deterministic trend under the alternative. The hypotheses to be tested are

$$H_0 : \phi = 1, \delta = 0 \Rightarrow y_t \sim I(1) \text{ with drift}$$
$$H_1 : |\phi| < 1 \Rightarrow y_t \sim I(0) \text{ with deterministic time trend}$$

This formulation is appropriate for trending time series like asset prices or the levels of macroeconomic aggregates like real GDP. The test statistics $t_{\phi=1}$ and $T(\hat{\phi} - 1)$ are computed from the above regression.

Under $H_0 : \phi = 1, \delta = 0$ the asymptotic distributions of these test statistics are influenced by the presence but not the coefficient values of the constant and time trend in the test regression.

$$T(\hat{\phi} - 1) \Rightarrow \frac{\int_0^1 W^\tau(r)dW(r)}{\int_0^1 W^\tau(r)^2 dr}$$

$$t_{\phi=1} \Rightarrow \frac{\int_0^1 W^\tau(r)dW(r)}{\left(\int_0^1 W^\tau(r)^2 dr\right)^{1/2}}$$

where

$$W^\tau(r) = W^\mu(r) - 12(r - \frac{1}{2}) \int_0^1 (s - \frac{1}{2}) W(s) ds$$

is a "de-meaned" and "de-trended" Wiener process.

The inclusion of a constant and trend in the test regression further shifts the distributions of $t_{\phi=1}$ and $T(\hat{\phi}-1)$ to the left.

```
> qunitroot(c(0.01,0.05,0.10), trend="ct")
[1] -3.958 -3.410 -3.127
```

```
> qunitroot(c(0.01,0.05,0.10), trend="ct",
+            statistic="n")
[1] -29.35 -21.70 -18.24
```

Note: $-1.645$ is the 77.52% quantile of the DF$^{\tau}$ distribution!

## 1.6 Dickey-Fuller Unit Root Tests

- The unit root tests described above are valid if the time series $y_t$ is well characterized by an AR(1) with white noise errors.

- Many economic and financial time series have a more complicated dynamic structure than is captured by a simple AR(1) model.

- Said and Dickey (1984) augment the basic autoregressive unit root test to accommodate general ARMA($p, q$) models with unknown orders and their test is referred to as the *augmented Dickey-Fuller* (ADF) test

Basic model

$$\begin{aligned} y_t &= \boldsymbol{\beta}'\mathbf{D}_t + \phi y_{t-1} + u_t \\ \phi(L)u_t &= \theta(L)\varepsilon_t,\ \varepsilon_t \sim \mathsf{WN}(0, \sigma^2) \end{aligned}$$

The ADF test tests the null hypothesis that a time series $y_t$ is $I(1)$ against the alternative that it is $I(0)$, assuming that the dynamics in the data have an ARMA structure. The ADF test is based on estimating the test regression

$$y_t = \boldsymbol{\beta}'\mathbf{D}_t + \phi y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta y_{t-j} + \varepsilon_t$$
$$\mathbf{D}_t = \text{deterministic terms}$$
$$\Delta y_{t-j} \text{ captures serial correlation}$$

The ADF t-statistic and normalized bias statistic are

$$\begin{aligned}\mathsf{ADF}_t &= t_{\phi=1} = \frac{\hat{\phi}-1}{\mathsf{SE}(\phi)} \\ \mathsf{ADF}_n &= \frac{T(\hat{\phi}-1)}{1-\hat{\psi}_1-\cdots-\hat{\psi}_p}\end{aligned}$$

Result: $\mathsf{ADF}_t$,$\mathsf{ADF}_n$ have same asymptotic distributions as $t_{\phi=1}$ and $T(\hat{\phi}-1)$ under white noise errors provided $p$ is selected appropriately.

Intuition: Re-parameterize AR(2) model

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\ &= \phi_1 y_{t-1} + (\phi_2 y_{t-1} - \phi_2 y_{t-1}) + \phi_2 y_{t-2} + \varepsilon_t \\ &= (\phi_1 + \phi_2) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t \\ &= \phi y_{t-1} + \psi \Delta y_{t-1} + \varepsilon_t \end{aligned}$$

where

$$\begin{aligned} \phi &= (\phi_1 + \phi_2) \\ \psi &= -\phi_2 \end{aligned}$$

Remarks:

- $y_{t-1} \sim I(1) \Rightarrow \hat{\phi}$ has non-normal distribution

- $\Delta y_{t-1} \sim I(0) \Rightarrow \hat{\psi}$ has normal distribution!

- Derivation requires trick from Sims, Stock and Watson (1989) ECTA

Important results:

- In the AR(2) model with a unit root

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

  the model may be reparameterized such that $\phi_2$ is the coefficient on an $I(0)$ variable

$$y_t = (\phi_1 + \phi_2) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t$$

  The Sims, Stock and Watson trick then shows that $\hat{\phi}_2$ has an asymptotic normal distribution.

- The model cannot be reparameterized such that $\phi = \phi_1 + \phi_2$ is the coefficient on an $I(0)$ variable. It is the coefficient on an $I(1)$ variable. Therefore, $\hat{\phi}$ has an asymptotic "unit root" distribution.

Alternative formulation of the ADF test regression:

$$\begin{aligned}\Delta y_t &= \boldsymbol{\beta}'\mathbf{D}_t + \pi y_{t-1} + \sum_{j=1}^{p} \psi_j \Delta y_{t-j} + \varepsilon_t \\ \pi &= \phi - 1\end{aligned}$$

Under the null hypothesis,

$$\Delta y_t \sim I(0) \Rightarrow \pi = 0.$$

The ADF t-statistic and normalized bias statistics are

$$\begin{aligned}\text{ADF}_t &= t_{\pi=0} = \frac{\hat{\pi}}{\text{SE}(\phi)} \\ \text{ADF}_n &= \frac{T\hat{\pi}}{1 - \hat{\psi}_1 - \cdots - \hat{\psi}_p}\end{aligned}$$

and these are equivalent to the previous statistics.