Unit Root Tests

A shock is usually used to describe an unexpected change in a variable or in the value of the error terms at a particular time period.

When we have a stationary system, effect of a shock will die out gradually. But, when we have a non-stationary system, effect of a shock is permanent.

We have two types of non-stationarity. In an AR(1) model we have:
- Unit root: $|\phi_1| = 1$: homogeneous non-stationarity
- Explosive root: $|\phi_1| > 1$: explosive non-stationarity

In the last case, a shock to the system become more influential as time goes on. It can never be seen in real life. We will not consider them.
**Autoregressive Unit Root**

- Consider the AR($p$) process:
  \[ \phi(L)y_t = \mu + \epsilon_t \]
  where \[ \phi(L) = 1 - \phi_1 L - L^2 \phi_2 - \ldots - \phi_p L^p \]

  As we discussed before, if one of the \( \phi_j \)'s equals 1, \( \Phi(1)=0 \), or
  \[ \phi_1 + \phi_2 + \ldots + \phi_p = 1 \]

- We say \( y_t \) has a *unit root*. In this case, \( y_t \) is non-stationary.

**Example:** AR(1): \( y_t = \mu + \phi_1 y_{t-1} + \epsilon_t \) \( \Rightarrow \) Unit root: \( \phi_1 = 1 \).

\[ \begin{align*}
  H_0 \text{ (} y_t \text{ non-stationarity): } & \quad \phi_1 = 1 \quad (\text{or, } \phi_1 - 1 = 0) \\
  H_1 \text{ (} y_t \text{ stationarity): } & \quad \phi_1 < 1 \quad (\text{or, } \phi_1 - 1 < 0) 
\end{align*} \]

- A *t-test* seems natural to test \( H_0 \). But, the ergodic theorem and MDS CLT do not apply: the \( t \)-statistic does not have the usual distributions.

---

**Autoregressive Unit Root**

- Now, let’s reparameterize the AR(1) process. Subtract \( y_{t-1} \) from \( y_t \):

  \[ \Delta y_t = y_t - y_{t-1} = \mu + (\phi_1 - 1) y_{t-1} + \epsilon_t \]
  \[ = \mu + \alpha_0 y_{t-1} + \epsilon_t \]

- Unit root test: \( H_0: \alpha_0 = \phi_1 - 1 = 0 \)
  \( H_1: \alpha_0 < 0. \)

- Natural test for \( H_0: t-test \). We call this test the *Dickey-Fuller (DF)* test. But, what is its distribution?

- Back to the general, AR($p$) process: \( \phi(L)y_t = \mu + \epsilon_t \)

  We rewrite the process using the *Dickey-Fuller reparameterization*:

  \[ \Delta y_t = \mu + \alpha_0 y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \ldots + \alpha_{p-1} \Delta y_{t-(p-1)} + \epsilon_t \]

- Both AR($p$) formulations are equivalent.
Autoregressive Unit Root – Testing

• AR(p) lag Φ(L): \[ \phi(L) = 1 - \phi_1 L^1 - \phi_2 L^2 - ... - \phi_p L^p \]

• DF reparameterization:
\[ (1 - L) - \alpha_0 - \alpha_1 (L - L^2) + \alpha_2 (L^2 - L^3) - ... - \alpha_{p-1} (L^{p-1} - L^p) \]

• Both parameterizations should be equal. Then, \( \Phi(1) = -\alpha_0 \)
\[ \Rightarrow \text{unit root hypothesis can be stated as } H_0: \alpha_0 = 0. \]

Note: The model is stationary if \( \alpha_0 < 0 \) \( \Rightarrow \text{natural } H_1: \alpha_0 < 0. \)

• Under \( H_0: \alpha_0 = 0 \), the model is AR(p-1) stationary in \( \Delta y_t \). Then, if \( y_t \) has a (single) unit root, then \( \Delta y_t \) is a stationary AR process.

• We have a linear regression framework. A \( t \)-test for \( H_0 \) is the \textit{Augmented Dickey-Fuller} (ADF) test.

Autoregressive Unit Root – Testing: DF

• The \textit{Dickey-Fuller} (DF) test is a special case of the ADF: No lags are included in the regression. It is easier to derive. We gain intuition from its derivation.

• From our previous example, we have:
\[ \Delta y_t = \mu + (\phi - 1)y_{t-1} + \epsilon_t = \mu + \alpha_0 y_{t-1} + \epsilon_t \]

• If \( \alpha_0 = 0 \), system has a unit root: \( H_0: \alpha_0 = 0 \)
\[ H_1: \alpha_0 < 0 \quad (|\alpha_0| < 0) \]

• We can test \( H_0 \) with a \( t \)-test:
\[ t_{b=1} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})} \]

• There is another associated test with \( H_{0b} \), the \( q \)-test: \( (T-1)(\hat{\phi} - 1) \).
Review: Stochastic Calculus

- Kolmogorov Continuity Theorem
  - If for all \( T > 0 \), there exist \( a, b, \delta > 0 \) such that:
    \[
    E\left( |X(t_1, \omega) - X(t_2, \omega)|^a \right) \leq \delta |t_1 - t_2|^{(1 + b)}
    \]
  - Then \( X(t, \omega) \) can be considered as a continuous stochastic process.

- Brownian motion is a continuous stochastic process.

- Brownian motion (Wiener process): \( X(t, \omega) \) is almost surely continuous, has independent normal distributed (\( N(0, t-s) \)) increments and \( X(t=0, \omega) = 0 \) (“a continuous random walk”).

Review: Stochastic Calculus – Wiener process

- Let the variable \( z(t) \) be almost surely continuous, with \( z(t=0) = 0 \).
- Define \( N(\mu, \nu) \) as a normal distribution with mean \( \mu \) and variance \( \nu \).
- The change in a small interval of time \( \Delta t \) is \( \Delta z \)

- Definition: The variable \( z(t) \) follows a Wiener process if
  - \( z(0) = 0 \)
  - \( \Delta z = \varepsilon \sqrt{\Delta t} \), where \( \varepsilon \sim N(0,1) \)
  - It has continuous paths.
  - The values of \( \Delta z \) for any 2 different (non-overlapping) periods of time are independent.

Notation: \( W(t), W(t, \omega), B(t) \).

Example: \( W_T(r) = \frac{1}{\sqrt{T}} (\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \ldots + \varepsilon_{[rT]}) \); \( r \in [0,1] \)
Review: Stochastic Process: Wiener process

- What is the distribution of the change in z over the next 2 time units?
  The change over the next 2 units equals the sum of:
  - The change over the next 1 unit (distributed as N(0,1)) plus
  - The change over the following time unit --also distributed as N(0,1).
  - The changes are independent.
  - The sum of 2 normal distributions is also normally distributed.
  Thus, the change over 2 time units is distributed as N(0,2).

- Properties of Wiener processes:
  - Mean of $\Delta z$ is 0
  - Variance of $\Delta z$ is $\Delta t$
  - Standard deviation of $\Delta z$ is $\sqrt{\Delta t}$
  - Let $N=T/\Delta t$, then $z(T) - z(0) = \sum_{i=1}^{N} \varepsilon_i \sqrt{\Delta t}$

Review: Stochastic Calculus – Wiener process

Example: $W_T(r) = \frac{1}{\sqrt{T}}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \ldots + \varepsilon_{[rT]}) = \frac{1}{\sqrt{T}} S_{[rT]}$; $r \in [0,1]$

- If $T$ is large, $W_{T}(r)$ is a good approximation to $W(r)$; $r \in [0,1]$, defined:
  $W(r) = \lim_{T \to \infty} W_T(r)$  $\Rightarrow$  $E[W(r)] = 0$
  $\Rightarrow$  $\text{Var}[W(r)] = t$

- Check Billingsley (1986) for the details behind the proof that $W_{T}(r)$ converges as a function to a continuous function $W(r)$.

- In a nutshell, we need
  - $\varepsilon$, satisfying some assumptions (stationarity, $E[|\varepsilon_i|^q] < \infty$ for $q>2$, etc.)
  - a FCLT (*Functional CLT*).
  - a Continuous Mapping Theorem. (Similar to Slutsky’s theorem).
**Review: Stochastic Calculus – Wiener process**

- **Functional CLT** (Donsker’s FCLT)
  If $\varepsilon_t$ satisfies some assumptions, then
  
  $$W_T(r) \xrightarrow{D} W(r),$$
  where $W(r)$ is a standard Brownian motion for $r \in [0, 1]$.

  Note: That is, sample statistics, like $W_T(r)$, do not converge to constants, but to functions of Brownian motions.

- A CLT is a limit for one term of a sequence of partial sums $\{S_k\}$, Donsker’s FCLT is a limit for the entire sequence $\{S_k\}$ instead of one term.

**Example:** $y_t = y_{t-1} + \varepsilon_t$ (Case 1). Get distribution of $(X'X/T^2)^{-1}$ for $y_t$.

\[
T^{-2} \sum_{t=1}^{T} (y_{t+1})^2 = T^{-2} \sum_{t=1}^{T} \sum_{i=1}^{t-1} (S_i + y_0)^2 = T^{-2} \sum_{i=1}^{T} (S_i + y_0)^2
\]

\[
= T^{-2} \sum_{i=1}^{T} [(S_i)^2 + 2y_0S_i + y_0^2]
\]

\[
= \sigma^2 \sum_{i=1}^{T} \left( \frac{S_i}{\sigma \sqrt{T}} \right)^2 T^{-1} + 2y_0 \sigma T^{-1/2} \sum_{i=1}^{T} \left( \frac{S_i}{\sigma \sqrt{T}} \right) T^{-1} + T^{-1} y_0^2
\]

\[
= \sigma^2 \sum_{i=1}^{T} \int_{i-1}^{T} \left( \frac{1}{\sigma \sqrt{T}} \bar{S}_{(T)} \right)^2 dr + 2y_0 \sigma T^{-1/2} \sum_{i=1}^{T} \int_{i-1}^{T} \left( \frac{1}{\sigma \sqrt{T}} \bar{S}_{(T)} \right) dr + T^{-1} y_0^2
\]

\[
= \sigma^2 \int_0^1 \frac{1}{2} X_T(r)^2 dr + 2y_0 \sigma T^{-1/2} \int_0^1 \frac{1}{2} X_T(r) dr + T^{-1} y_0^2
\]

\[
\xrightarrow{T \to \infty} \sigma^2 \int_0^1 W(r)^2 dr,
\]
Review: Stochastic Calculus – Ito’s Theorem

• The integral w.r.t a Brownian motion, given by Ito’s theorem (integral):
  \[ \int f(t, \omega) \, dB = \sum f(t_k, \omega) \Delta B_k \text{ where } t_k \in [t_k, t_{k+1}) \text{ as } t_{k+1} - t_k \to 0. \]

As we increase the partitions of \([0, T]\), the sum \(\to^p\) to the integral.

• But, this is a probability statement: We can find a sample path where the sum can be arbitrarily far from the integral for arbitrarily large partitions (small intervals of integration).

• You may recall that for a Riemann integral, the choice of \(t_k^*\) (at the start or at the end of the partition) is not important. But, for Ito’s integral, it is important (at the start of the partition).

• Ito’s Theorem result:
  \[ \int B(t, \omega) \, dB(t, \omega) = B^2(t, \omega)/2 - t/2. \]

Autoregressive Unit Root – Testing: Intuition

• We continue with \(y_t = y_{t-1} + \varepsilon_t\) (Case 1). Using OLS, we estimate \(\phi\):
  \[
  \hat{\phi} = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \frac{\sum_{t=1}^T (y_{t-1} + \Delta y_{t-1}) y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = 1 + \frac{\sum_{t=1}^T y_{t-1} \Delta y_{t-1}}{\sum_{t=1}^T y_{t-1}^2}
  \]

• This implies:
  \[
  T(\hat{\phi} - 1) = \frac{\sum_{t=1}^T y_{t-1} \Delta y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} = \frac{\sum_{t=1}^T (y_{t-1} / \sqrt{T})(\varepsilon_t / \sqrt{T})}{\sum_{t=1}^T (y_{t-1} / \sqrt{T})^2}
  \]

• From the way we defined \(W_{T}(\cdot)\), we can see that \(y_{t}/\sqrt{T}\) converges to a Brownian motion. Under \(H_0, y_{t}\) is a sum of white noise errors.
Autoregressive Unit Root – Testing: Intuition

- Intuition for distribution under H_0:
  - Think of \( y_t \) as a sum of white noise errors.
  - Think of \( \varepsilon_t \) as \( dW(t) \).

Then, using Billingley (1986), we guess that \( T(\hat{\phi} - 1) \) converges to

\[
T(\hat{\phi} - 1) \xrightarrow{d} \int_0^1 W(t) dW(t) - \int_0^T W(t)^2 dt
\]

- We think of \( \varepsilon_t \) as \( dW(t) \). Then, \( \sum_{k=0}^{t-1} \varepsilon_k \), which corresponds to \( \int_0^{t/T} dW(s) = W(s/T) \) (for \( W(0) = 0 \)). Using Ito’s integral, we have

\[
T(\hat{\phi} - 1) \xrightarrow{d} \frac{1}{2} W(1)^2 - \int_0^T W(t)^2 dt
\]

Note: \( W(1) \) is a \( N(0,1) \). Then, \( W(1)^2 \) is just a \( \chi^2(1) \) RV.

- Contrary to the stable model the denominator of the expression for the OLS estimator – i.e., \( (1/T)\sum \chi^2 \) -- does not converge to a constant \( a.s. \), but to a RV strongly correlated with the numerator.

- Then, the asymptotic distribution is not normal. It turns out that the limiting distribution of the OLS estimator is highly skewed, with a long tail to the left.
Autoregressive Unit Root – Testing: Intuition

• DF distribution relative to a Normal. It is skewed, with a long tail to the left.

Autoregressive Unit Root – Testing: DF

• Back to the AR(1) model. The $t$-test statistic for $H_0: \phi_0=0$ is given by

$$t_{\phi=1} = \frac{\hat{\phi} - 1}{SE(\hat{\phi})} = \frac{\hat{\phi} - 1}{\sqrt{\frac{s^2}{\sum_{i=2}^{T} y_{t-1}^2}}^2}$$

• The test is a one-sided left tail test. If $\{y_t\}$ is stationary (i.e., $|\phi| < 1$) then it can be shown

$$\sqrt{T}(\hat{\phi} - \phi) \overset{d}{\to} N(0, 1 - \phi^2)$$

• This means that under $H_0$, the asymptotic distribution of $t_{\phi=1}$ is $N(0,1)$. That is, under $H_0$:

$$\hat{\phi} \overset{d}{\to} N(0,1)$$

which we know is not correct, since $y_t$ is not stationary and ergodic.
• Under $H_0$, $y_t$ is not stationary and ergodic. The usual sample moments do not converge to fixed constants. Using the results discussed above, Phillips (1987) showed that the sample moments of $y_t$ converge to random functions of Brownian motions. Under $H_0$:

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

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

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

---

**Autoregressive Unit Root – Testing: DF**

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

• The limiting distribution of $t_{\phi=1}$ is the DF distribution, which does not have a closed form representation. Then, quantiles of the distribution must be numerically approximated or simulated.

• The distribution of the DF test is non-standard. It has been tabulated under different scenarios.

1) with a constant: $y_t = \mu + \Phi y_{t-1} + \varepsilon_t$.

2) with a constant and a trend: $y_t = \mu + \delta t + \Phi y_{t-1} + \varepsilon_t$.

3) no constant: $y_t = \Phi y_{t-1} + \varepsilon_t$.

• The tests with no constant are not used in practice.
Autoregressive Unit Root – Testing: DF

• Critical values of the DF test under different scenarios.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>0.01 0.025 0.05 0.10</td>
</tr>
<tr>
<td>Model without constant</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>-2.60 -2.24 -1.95 -1.61</td>
</tr>
<tr>
<td>250</td>
<td>-2.58 -2.23 -1.95 -1.62</td>
</tr>
<tr>
<td>500</td>
<td>-2.58 -2.23 -1.95 -1.62</td>
</tr>
<tr>
<td>∞</td>
<td>-2.58 -2.23 -1.95 -1.62</td>
</tr>
<tr>
<td>Model with constant</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>-3.51 -3.17 -2.89 -2.58</td>
</tr>
<tr>
<td>250</td>
<td>-3.46 -3.14 -2.88 -2.57</td>
</tr>
<tr>
<td>500</td>
<td>-3.44 -3.13 -2.87 -2.57</td>
</tr>
<tr>
<td>∞</td>
<td>-3.43 -3.12 -2.86 -2.57</td>
</tr>
<tr>
<td>Model with time trend</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>-4.04 -3.73 -3.45 -3.15</td>
</tr>
<tr>
<td>∞</td>
<td>-3.96 -3.66 -3.41 -3.12</td>
</tr>
</tbody>
</table>

Autoregressive Unit Root – DF: Case 2

• Case 2. DF with a constant term in DGP: \( y_t = \mu + \phi y_{t-1} + \epsilon_t \)

The hypotheses to be tested:

\[ H_0 : \phi = 1, \mu = 0 \Rightarrow Y_t \sim I(1) \text{ without drift} \]

\[ H_1 : |\phi| < 1 \Rightarrow Y_t \sim I(0) \text{ with zero mean} \]

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 \( (T-1)(\hat{\phi} - 1) \) are computed from the estimation of the AR(1) model with a constant.
Autoregressive Unit Root – DF: Case 2

• Under $H_0: \Phi = 1, \mu = 0$, the asymptotic distributions of these test statistics are influenced by the presence, but not the coefficient value, of the constant in the test regression:

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

where $W^\mu(r) = W(r) - \int_0^1 W(r) dr$ is a demeaned Wiener process, i.e., $\int_0^1 W^\mu(r) dr = 0$

• Inclusion of a constant pushes the tests’ distributions to the left.

Autoregressive Unit Root – DF: Case 3

• Case 3. With constant and trend term in the DGP. The test regression is $y_t = \mu + \delta t + \phi y_{t-1} + \epsilon_t$

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

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

• 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 $(T-1)(\hat{\phi}-1)$ are computed from the above regression.
Autoregressive Unit Root – DF: Case 3

• Again, under $H_0 : \Phi = 1, \delta = 0$, the asymptotic distributions of both test statistics are influenced by the presence of the constant and time trend in the test regression. Now, we have:

$$\frac{(T - 1)(\hat{\delta} - 1)}{2} \rightarrow \int_0^1 W^\mu(r)^2 dr$$

$$\int_0^1 W^\mu(r) dW(r) \rightarrow T - 1 \left( \frac{1}{2} \right) W(r)dr$$

where $W^\mu(r) = W^\mu(r) - 12 \left( r - \frac{1}{2} \right) \left( s - \frac{1}{2} \right) W(r)dr$ is a demeaned and detrended Wiener process.

Autoregressive Unit Root – DF: Case 3

• Again, the inclusion of a constant and trend in the test regression further shifts the distributions of $t_{\Phi=1}$ and $(T - 1)(\Phi - 1)$ to the left.
• Which version of the three main variations of the test should be used is not a minor issue. The decision has implications for the size and the power of the unit root test.

• For example, an incorrect exclusion of the time trend term leads to bias in the coefficient estimate for $\Phi$, leading to size distortions and reductions in power.

• Since the normalized bias $(T-1)(\hat{\Phi} - 1)$ 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 $H_0 : \Phi = 1$.

**Autoregressive Unit Root – Testing: DF**

• Back to the general, AR($p$) process. We can rewrite the equation as the *Dickey-Fuller reparameterization*:

$$\Delta y_t = \mu + \alpha_0 y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \ldots + \alpha_{p-1} \Delta y_{t-(p-1)} + \varepsilon_t$$

• The model is stationary if $\alpha_0 < 0 \Rightarrow$ natural $H_1$: $\alpha_0 < 0$.

• Under $H_0$: $\alpha_0 = 0$, the model is AR($p$-1) stationary in $\Delta y_t$. Then, if $y_t$ has a (single) unit root, then $\Delta y_t$ is a stationary AR process.

• The $t$-test for $H_0$ from OLS estimation is the Augmented Dickey-Fuller (ADF) test.

• Similar situation as the DF test, we have a non-normal distribution.

**Autoregressive Unit Root – Testing: ADF**

...
The asymptotic distribution is:

\[ T \hat{a}_0 \xrightarrow{d} (1 - \alpha_1 - \alpha_2 - \cdots - \alpha_{k-1}) DF_{\alpha_0} \]

\[ ADF = \frac{\hat{a}_0}{s(\hat{a}_0)} \xrightarrow{} DF_t. \]

The limit distributions \( DF_\alpha \) and \( DF_t \) are non-normal. They are skewed to the left, and have negative means.

• First result: \( \alpha_0^* \) converges to its true value (of zero) at rate \( T \); rather than the conventional rate of \( \sqrt{T} \) ➞ superconsistency.

• Second result: The t-statistic for \( \alpha_0^* \) converges to a non-normal limit distribution, but does not depend on \( \alpha \).

The ADF distribution has been extensively tabulated under the usual scenarios: 1) with a constant; 2) with a constant and a trend; and 3) no constant. This last scenario is seldom used in practice.

Like in the DF case, which version of the three main versions of the test should be used is not a minor issue. A wrong decision has potential size and power implications.

One-sided \( H_1 \): the ADF test rejects \( H_0 \) when \( ADF < c \) where \( c \) is the critical value from the ADF table.

Note: The \( SE(\alpha_0^*) = s \sqrt{\sum_y^2} \), the usual (homoscedastic) SE. But, we could be more general. Homoskedasticity is not required.
Autoregressive Unit Root – Testing: ADF

- We described the test with an intercept. Another setting includes a linear time trend:

\[ \Delta y_t = \mu_1 + \mu_2 t + \alpha_0 y_{t-1} + \alpha_1 \Delta y_{t-1} + \ldots + \alpha_{p-1} \Delta y_{t-(p-1)} + \epsilon_t \]

- Natural framework when the alternative hypothesis is that the series is stationary about a linear time trend.

- If \( t \) is included, the test procedure is the same, but different critical values are needed. The ADF test has a different distribution when \( t \) is included.

Autoregressive Unit Root – DF: Example 1

- Monthly USD/GBP exchange rate, \( S_n \) (1800-2013), \( T=2534 \).
- Case 1 (no constant in DGP):

| Parameter       | DF | Estimate   | Error     | t Value | Pr > |t| |
|------------------|----|------------|-----------|---------|-------|----|
| x1               | 1  | 0.99934    | 0.00061935| 1613.52 | <.0001|

\( (T-1)(\hat{\phi} -1)=2533*(1-.99934)=-1.67178 \)

Critical values at 5% level: –8.0 for \( T=500 \)
-8.1 for \( T=\infty \)

- Cannot reject \( H_0 \) => Take 1st differences (changes in \( S \)) to model the series.

- With a constant, \( \hat{\phi} =0.99631 \). Similar conclusion (Critical values at 5% level: –14.0 for \( T=500 \) and –14.1 for \( T=\infty \)): Model changes in \( S \).
**Autoregressive Unit Root – DF: Example 2**

  - No constant in DGP (unusual case, called Case 1): $y_t = \phi y_{t-1} + \varepsilon_t$

| Parameter | DF | Estimate | Error | t Value | Pr > |t| |
|-----------|----|----------|-------|---------|-------|---|
| x1        | 1  | 1.00298  | 0.00088376 | 1134.90 | <.0001 |

$(T-1)(\hat{\phi} - 1) = 2533 \times 0.00298 = 7.5483$ (positive, not very interesting)

Critical values at 5% level: $-8.0$ for $T=500$, $-8.1$ for $T=\infty$

- Cannot reject $H_0 =>$ Take 1st differences (returns) to model the series.

- With a constant, $\hat{\phi} = 1.00269$. Same conclusion.

---

**Autoregressive Unit Root – DF-GLS**

- Elliott, Rothenberg and Stock (1992) (ERS) study *point optimal invariant tests* (POI) for unit roots. An invariant test is a test invariant to nuisance parameters.

- In the unit root case, we consider invariance to the parameters that capture the stationary movements around the unit roots - i.e., the parameters to AR($\phi$) parameters.

- Consider: $y_t = \mu + \delta t + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t$.

- ERS show that the POI test for a unit root against $q = q^*$ is:

  $$M_T = \frac{s^2_{\rho=1}}{s^2_{\rho=q^*}}$$
Autoregressive Unit Root – DF-GLS

- $M_T = \frac{s_{\rho=1}^2}{s_{\rho=\rho^*}^2}$

where $s^2$ is the variances residuals from the GLS estimation under both scenarios for $\rho$, $\rho = 1$ and $\rho = \rho^*$, respectively:

- The critical value for the test will depend on $c$ where $\rho^* = 1 - c/T$.

Note: When dynamics are introduced in the $u_t$ equation, $\Delta u_t$ lags, the critical values have to be adjusted.

- In practice $\rho^*$ is unknown. ERS suggest different values for different cases. Say, $c=-13.5$, for the case with a trend, gives a power of 50%.

### Autoregressive Unit Root – DF-GLS

- It turns out that if we instead do the GLS-adjustment and then perform the ADF-test (without allowing for a mean or trend) we get approximately the POI-test. ERS call this test the DF-GLS$_t$ test.

- The critical values depend on $T$.

<table>
<thead>
<tr>
<th>T</th>
<th>1%</th>
<th>5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-3.77</td>
<td>-3.19</td>
</tr>
<tr>
<td>100</td>
<td>-3.58</td>
<td>-3.03</td>
</tr>
<tr>
<td>200</td>
<td>-3.46</td>
<td>-2.93</td>
</tr>
<tr>
<td>500</td>
<td>-3.47</td>
<td>-2.89</td>
</tr>
<tr>
<td>$\infty$</td>
<td>-3.48</td>
<td>-2.89</td>
</tr>
</tbody>
</table>

- Check ERS for critical values for other scenarios.
Important issue: lag $p$.

- Check the specification of the lag length $p$. If $p$ is too small, then the remaining serial correlation in the errors will bias the test. If $p$ is too large, then the power of the test will suffer.

- Ng and Perron (1995) suggestion:
  1. Set an upper bound $p_{\text{max}}$ for $p$.
  2. Estimate the ADF test regression with $p = p_{\text{max}}$.
     - If $|t_{a(p)}| > 1.6$ set $p = p_{\text{max}}$ and perform the ADF test.
     - Otherwise, reduce the lag length by one. Go back to (1)

- Schwert’s (1989) rule of thumb for determining $p_{\text{max}}$:
  \[ p_{\text{max}} = \left[ 12 \left( \frac{T}{100} \right)^{1/4} \right] \]

**Autoregressive Unit Root – Testing: PP Test**

- The Phillips-Perron (PP) unit root tests differ from the ADF tests mainly in how they deal with serial correlation and heteroskedasticity in the errors.

- The ADF tests use a parametric autoregression to approximate the ARMA structure of the errors in the test regression. The PP tests correct the DF tests by the bias induced by the omitted autocorrelation.

- These modified statistics, denoted $Z_t$ and $Z_{\delta}$, are given by
  \[
  Z_t = \sqrt{\frac{\hat{\sigma}^2}{\lambda^2 \delta_{a}^2}} \cdot \frac{1}{2} \left( \frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2} \right) \left( \frac{T \left( \text{SE}(\hat{a}_0) \right)}{\hat{\sigma}^2} \right)
  \]
  \[
  Z_{\delta} = T\hat{a}_0 - \frac{1}{2} \frac{T^2 \left( \text{SE}(\hat{a}_0) \right)}{\hat{\sigma}^2} \left( \hat{\lambda}^2 - \hat{\sigma}^2 \right)
  \]
**Autoregressive Unit Root – Testing: PP Test**

• The terms $\sigma^2$ and $\hat{\lambda}$ are consistent estimates of the variance parameters:

$$\sigma^2 = \lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E\left( \varepsilon_t^2 \right) \quad \hat{\lambda}^2 = \lim_{T \to \infty} \sum_{t=1}^{T} E \left( \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \right)$$

• Under $H_0: \alpha_0 = 0$, the PP $Z_t$ and $Z_{\alpha_0}$ statistics have the same asymptotic distributions as the DF t-statistic and normalized bias statistics.

• PP tests tend to be more powerful than the ADF tests. But, they can suffer severe size distortions (when autocorrelations of $\varepsilon_t$ are negative) and they are more sensitive to model misspecification (order of ARMA model).

---

**Autoregressive Unit Root – Testing: PP Test**

• Advantage of the PP tests over the ADF tests:
  - Robust to general forms of heteroskedasticity in the error term $\varepsilon_t$.
  - No need to specify a lag length for the ADF test regression.
• The ADF and PP unit root tests are very popular. They have been, however, widely criticized.

• Main criticism: Power of tests is low if the process is stationary but with a root close to the non-stationary boundary.

• For example, the tests are poor at distinguishing between $\phi = 1$ or $\phi = 0.976$, especially with small sample sizes.

• Suppose the true DGP is $y_t = 0.976 y_{t-1} + \varepsilon_t$ => $H_0: \alpha_0 = 0$ should be rejected.

• One way to get around this is to use a stationarity test (like KPSS test) as well as the unit root ADF or PP tests.

Autoregressive Unit Root – Testing: Criticisms

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

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

2. Size – Both tests are known to have severe size distortion (in the direction of over-rejecting $H_0$) when the series has a large negative MA root. (See, Schwert, JBE,$\dagger$, 1989: MA = -0.8 => size = 100%)
Autoregressive Unit Root – Testing: KPSS

• A different test is the KPSS (Kwiatkowski, Phillips, Schmidt and Shin) Test (1992). It can be used to test whether we have a deterministic trend vs. stochastic trend:

\[ H_0: Y_t \sim I(0) \quad \rightarrow \text{level (or trend) stationary} \]
\[ H_1: Y_t \sim I(1) \quad \rightarrow \text{difference stationary} \]

• Setup

\[ y_t = \mu + \delta t + r_t + u_t \]
\[ r_t = r_{t-1} + \varepsilon_t \]

where \( \varepsilon_t \sim \text{WN}(0,\sigma^2) \), uncorrelated with \( u_t \sim \text{WN} \). Then,
\( H_0 \) (trend stationary): \( \sigma^2 = 0 \)
\( H_0 \) (\( y_t \) (level) stationary): \( \sigma^2 = 0 \) & \( \delta = 0 \).

Under \( H_1 \): \( \sigma^2 \neq 0 \), there is a RW in \( y_t \).

Autoregressive Unit Root – Testing: KPSS

• Under some assumptions (normality, \textit{i.i.d.} for \( u_t \) & \( \varepsilon_t \)), a one-sided LM test of the null that there is no random walk (\( \varepsilon_t = 0 \), for all \( t \)) can be constructed with:

\[ KPSS = T^{-2} \sum_{i=1}^{T} \frac{S_i}{S_u} \]

where \( S_u^2 \) is the variance of \( u_t \) (“long run” variance) estimated as

\[ S_u^2(l) = \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t^2 + \frac{2}{T} \sum_{t=1}^{T} w(t, l) \sum_{t=2+1}^{T} \tilde{u}_t \tilde{u}_{t-l} \]

where \( w(l, l) \) is a kernel function, for example, the Bartlett kernel. We also need to specify the number of lags, which should grow with \( T \).

• Under \( H_0 \), \( \tilde{u}_t \) can be estimated by OLS.
Autoregressive Unit Root – Testing: KPSS

• Easy to construct. Steps:
  1. Regress \( y_t \) on a constant and time trend. Get OLS residuals, \( u^\hat{} \).
  2. Calculate the partial sum of the residuals: \( S_i = \sum_{i=1}^{T} \hat{u}_i \).
  3. Compute the KPSS test statistic

\[
KPSS = T^{-2} \sum_{i=1}^{T} \frac{S_i}{s_u^2}
\]

where \( s_u^2 \) is the estimate of the long-run variance of the residuals.
  4. Reject \( H_0 \) when KPSS is large (the series wander from its mean).

• Asymptotic distribution of the test statistic is non-standard—it can be derived using Brownian motions, appealing to FCLT and CMT.

Autoregressive Unit Root – Testing: KPSS

• KPSS converges to three different distribution, depending on whether the model is trend-stationary (\( \delta \neq 0 \)), level-stationary (\( \delta = 0 \)), or zero-mean stationary (\( \delta = 0, \mu = 0 \)).

• For example, if a constant is included (\( \delta = 0 \)) KPSS converges to

\[
KPSS \overset{d}{\to} \int_{0}^{1} [W(r) - W(1)] dr
\]

Note: \( V = W(r) - rW(1) \) is called a standard Brownian bridge. It satisfies \( V(0) = V(1) = 0 \).

• It is a very powerful unit root test, but if there is a volatility shift it cannot catch this type non-stationarity.
Autoregressive Unit Root – Structural Breaks

• A stationary time-series may look like non-stationary when there are structural breaks in the intercept or trend.

• The unit root tests lead to false non-rejection of the null when we do not consider the structural breaks. A low power problem.

• A single known breakpoint was studied by Perron (Econometrica, 1989). Perron (1997) extended it to a case of unknown breakpoint.

• Perron considers the null and alternative hypotheses

\[ H_0: y_t = a_0 + y_{t-1} + \mu_1 D_p + \varepsilon_t \quad (y_t \sim ST \text{ with a jump}) \]

\[ H_1: y_t = a_0 + a_2 t + \mu_2 D_L + \varepsilon_t \quad (y_t \sim TS \text{ with a jump}) \]

Pulse break: \( D_p = 1 \) if \( t = T_B + 1 \) and zero otherwise,
Level break: \( D_L = 0 \) for \( t = 1, \ldots, T_B \) and one otherwise.

---

Unit Root – Single Structural Break: Perron

• Power of ADF tests: Rejection frequencies of ADF–tests

<table>
<thead>
<tr>
<th>Model: ( a_0 = a_2 = 0.5 ) and ( \mu_2 = 10 )</th>
<th>1% level</th>
<th>5% level</th>
<th>10% level</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF–tests</td>
<td>0.004</td>
<td>0.344</td>
<td>0.714</td>
</tr>
<tr>
<td>Model: ( a_0 = a_2 = 0.5 ) and ( \mu_2 = 12 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADF–tests</td>
<td>0.000</td>
<td>0.028</td>
<td>0.264</td>
</tr>
</tbody>
</table>

• Observations:
  - ADF tests are biased toward non-rejection of the non-stationary \( H_0 \).
  - Rejection frequency is inversely related to the magnitude of the shift.

• Perron estimated values of the AR coefficient in the DF regression. They were biased toward unity and that this bias increased as the magnitude of the break increased.
Unit Root – Single Structural Break: Perron

- Perron’s suggestion: Running the following OLS regression:

\[ y_t = a_0 + a_1 y_{t-1} + a_2 t + \mu_1 D_L + \gamma_1 D_L + \varepsilon_t \]

\[ H_0: a_1 = 1 \quad \Rightarrow \text{use } t\text{-ratio, DF unit root test.} \]

- Perron shows that the asymptotic distribution of the \( t \)-statistic depends on the location of the structural break, \( \lambda = T_B/T \).

- Perron (1989) derives critical values for different cases. For example:

\[ H_0: y_t = a_0 + y_{t-1} + \mu_1 D_L + \gamma_1 D_L + \varepsilon_t \quad (y_t \sim ST \text{ with a jump & break}) \]

\[ H_1: y_t = a_0 + a_2 t + \mu_2 D_L + \gamma_2 D_L + \varepsilon_t \quad (y_t \sim TS \text{ with a jump & break}) \]

- Main problem with this test procedure: structural breaks are not known, they need to be estimated from data.

Unit Root – Single Structural Break: ZA

- Main problem with this test: Structural breaks are not known, they need to be estimated. Many papers dealing with endogenous structural breaks: Zivot and Andrews (ZA, 1992), Lumsdaine and Papell (1998), Lee and Strazicich (2003).

- ZA’s test is a sequential ADF test, using a different dummy variable for each possible break date. The break date is selected where the \( t \)-statistic from the ADF test is a minimum (most negative) – break date is chosen where the evidence is least favorable for the unit root null.

- ZA’s critical values are different from Perron’s (1989). In general, ZA provide more evidence for unit roots than under Perron’s.
Unit Root – Multiple Structural Breaks

• Lumsdaine and Papell (1998) and Lee and Strazicich (2003) allow for multiple breaks in their tests.

• Lumsdaine and Papell extend ZA, by allowing two structural breaks under the alternative hypothesis of the unit root test and additionally allow for breaks in level and trend.

• The derivation of critical values on ZA and Lumsdaine and Papell (1998) assumes no breaks under the null hypothesis. This assumption may lead to conclude incorrectly ( spuriously) reject H0 (unit root) when, in fact, the series is difference-stationary with breaks.

Unit Root – Multiple Structural Breaks

• The derivation of critical values on ZA and Lumsdaine and Papell (1998) assumes no breaks under the null hypothesis. This assumption may lead to conclude incorrectly ( spuriously) reject H0 (unit root) when, in fact, the series is difference-stationary with breaks.

• To deal with this issue, Lee and Strazicich (2003) propose a LM unit-root test, incorporating structural breaks under H0 (& H1), with DGPs (augmenting with $p$ first-difference AR lags works well):

$$H_0: y_t = a_0 + \gamma_{t,1} + \mu_1 D_{p,1} + \mu_2 D_{p,2} + \gamma_{1} D_{L,1} + \gamma_{2} D_{L,2} + \varepsilon_t$$

$$H_1: y_t = a_0 + (1-a_1) \gamma_{t,1} + a_2 \gamma_{2} + \mu_1 D_{p,1} + \mu_2 D_{p,2} + \gamma_{1} D_{L,1} + \gamma_{2} D_{L,2} + \varepsilon_t$$

• In general, using Lee and Strazicich (2003), we tend to reject more H0 (unit root).
Autoregressive Unit Root - Relevance

• We can always decompose a unit root process into the sum of a random walk and a stable process. This is known as the Beveridge-Nelson (1981) (BN) composition.

• Let $y_t \sim I(1)$, $r_t \sim RW$ and $c_t \sim I(0)$.
  
  $y_t = r_t + c_t$

Since $c_t$ is stable it has a Wold decomposition:

$$(1 - L) y_t = \psi(L) \epsilon_t$$

Then,

$$(1-L)y_t = \psi(L)\epsilon_t = \psi(1)\epsilon_t + (\psi(L) - \psi(1))\epsilon_t$$

$= \psi(1)\epsilon_t + \psi(L)*\epsilon_t$

where $\psi(1)=0$. Then,

$$y_t = \psi(1)(1-L)^{-1}\epsilon_t + \psi(L)* (1-L)^{-1}\epsilon_t = r_t + c_t$$

Autoregressive Unit Root - Relevance

• Usual finding in economics: Many time series seem to have unit roots. But, there is debate over power of unit root tests and the effect of structural breaks.

Example: Consumption, output, stock prices, interest rates, unemployment, size, compensation are usually $I(1)$.

• Sometimes a linear combination of $I(1)$ series produces an $I(0)$. For example, (log consumption– log output) is stationary. This situation is called cointegration.

• Practical problems with cointegration:
  - Asymptotics change completely.
  - Not enough data to definitively say we have cointegration.
Autoregressive Unit Root – Structural Breaks 2

- Nelson and Plosser (1982) tested using ADF 14 macroeconomic series (GNP, IP, employment, etc) for unit roots: Rejected $H_0$ for only one. Summary of results from tests allowing for structural breaks, from Glyn et al. (2007):

Table 1: Unit Root Tests with the Nelson and Plosser’s Data (1982) Set

<table>
<thead>
<tr>
<th>Empirical Studies by:</th>
<th>Model</th>
<th>Unit Root (with possible breaks)</th>
<th>Stationary (with possible breaks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson and Plosser (1982)</td>
<td>ADF test with no break</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Perron (1989)**</td>
<td>Exogenous with one break</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Zivot and Andrews (1992)*</td>
<td>Endogenous with one break</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Lumsdaine and Papell (1997)*</td>
<td>Endogenous with two breaks</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Lee and Strazicich (2003)**</td>
<td>Endogenous with two breaks</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

* Assume no breaks under the null hypothesis of unit root.
** Assume breaks under both the null and the alternative hypothesis.