Unit Root Tests

Mei-Yuan Chen
Department of Finance
National Chung Hsing University

February 25, 2013
Unit Root Tests

An ARMA($p,q$) (autoregressive of order $p$ and moving average of order $q$) model of $y_t$ is typically written as

$$\Phi(B)y_t = \Psi(B)\epsilon_t,$$

where

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p,$$
$$\Psi(B) = 1 - \psi_1 B - \psi_2 B^2 - \cdots - \psi_q B^q,$$

are polynomials in $B$, the back-shift operator, and $\{\epsilon_t\}$ is white-noise.
An ARMA model is said to be “stationary” if all the roots of $\Phi(z) = 0$ are outside the unit circle, $|z| = 1$. In this case, we can obtain a stationary solution $y_t = \Phi(B)^{-1}\Psi(B)\epsilon_t$. 
Invertibility

An ARMA model is said to be “invertible” if all the roots of \( \Psi(z) = 0 \) are outside the unit circle. In this case, we can write \( \epsilon_t = \Psi(B)^{-1}\Phi(B)y_t \). For example, if \( y_t = \alpha y_{t-1} + \epsilon_t \), then \( \{y_t\} \) is stationary provided that the root of \( \Phi(z) = 1 - \alpha z = 0 \), \( z = 1/\alpha \), is outside the unit circle, i.e., \( |\alpha| < 1 \).
An ARIMA($p,d,q$) model of $y_t$ is $\Phi(B)(l - B)^d y_t = \Psi(B)\epsilon_t$, where $\Phi(B)$ and $\Psi(B)$ satisfy the stationarity and invertibility conditions. As the polynomial $A(z) \equiv \Phi(z)(1 - z)^d = 0$ has $d$ roots on the unit circle, $\{y_t\}$ is said to be an $I(d)$ process (integrated process of order $d$). That is, an $I(d)$ process must be differenced $d$ times to achieve stationarity; in particular, an $I(0)$ process is stationary.
Of particular interest to us are $I(1)$ processes of the following form: $(1 - B)y_t = u_t$, where $u_t$ has a stationary and invertible ARMA representation. To see the properties of an $I(1)$ process, consider a special case where $\{y_t\}$ is a “random walk”, i.e., the innovations $u_t$ are i.i.d. with mean zero and finite variance $\sigma_u^2$. That is,

$$y_t = y_{t-1} + u_t, \quad u_t \sim i.i.d.(0, \sigma_u^2).$$
Properties of a Random Walk

1. the effects of past $u_t$ on $y_t$ are permanent;
2. $y_t$ has unbounded variance $t\sigma_u^2$;
3. $y_t$ has a smooth sample path (in terms of level crossing) which typically wanders away and rarely crosses its mean level. By contrast, an $I(0)$ process has short memory, bounded variance, and a ragged sample path which crosses its mean level very often.
Clearly, an $I(1)$ process is non-stationary, but a non-stationary series need not be $I(1)$. If $(1 - B)y_t$ has a non-zero mean $\mu_0$ such that $(1 - B)y_t = \mu_0 + u_t$, then $\{y_t\}$ is said to be an $I(1)$ process with the drift $\mu_0$ (since $y_t = \mu_0 t + \sum_{i=0}^{t} u_{t-i}$. A major problem in time-series regressions of an $I(1)$ variable is that many known results are no longer valid.
Consider the DGP (data generating process):

\[ y_t = \alpha_0 y_{t-1} + u_t, \]

where \( u_t \) is an \( I(0) \) process. We want to test the null hypothesis of a unit root (\( \alpha_0 = 1 \)) against the alternative hypothesis that \( y_t \) is \( I(0) \) (\( |\alpha_0| < 1 \)).
Dickey-Fuller Tests

Suppose that we estimate one of the following three models.

\[
y_t = \hat{\alpha}_1 T y_{t-1} + \hat{u}_t, \quad (1)
\]

\[
y_t = \hat{\mu}_2 T + \hat{\alpha}_2 T y_{t-1} + \hat{u}_t, \quad (2)
\]

\[
y_t = \hat{\mu}_3 T + \hat{\beta}_3 T \left( t - \frac{T}{2} \right) + \hat{\alpha}_3 T y_{t-1} + \hat{u}_t, \quad (3)
\]

where \( \hat{u}_t \) is a generic notation for OLS residual.
If $u_t$ is i.i.d. $N(0, \sigma^2_u)$, Dickey & Fuller (1979) show that under the null hypothesis $T(\hat{\alpha}_{iT} - 1)$ and the $t$-statistics

$$
\tau_{\hat{\alpha}_{iT}} = (\hat{\alpha}_{iT} - 1)/[\text{s.e.}(\hat{\alpha}_{iT})], \quad i = 1, 2, 3,
$$

$$
\tau_{\hat{\mu}_{iT}} = \hat{\mu}_{iT}/[\text{s.e.}(\hat{\mu}_{iT})], \quad i = 2, 3,
$$

$$
\tau_{\hat{\beta}_{3T}} = \hat{\beta}_{3T}/[\text{s.e.}(\hat{\beta}_{3T})],
$$

have non-normal limiting distributions, where s.e. stands for the OLS standard errors.
The empirical distributions of $T(\hat{\alpha}_{iT} - 1)$ and $\tau_{\hat{\alpha}_{iT}}$, $i = 1, 2, 3$ are tabulated in Fuller (1976, p. 371 & p. 373); the distributions of the $t$-ratios $\tau_{\hat{\mu}_{iT}}$, $i = 2, 3$ and $\tau_{\hat{\beta}_{3T}}$ are tabulated in Dickey & Fuller (1981). From the table for $T(\hat{\alpha}_{1T} - 1)$ it can be seen that this distribution is skewed to the left and $\hat{\alpha}_{1T}$ is downward biased.
**Table** Finite-sample and asymptotic critical values of the Dickey-Fuller test.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>25</td>
<td>−3.75</td>
</tr>
<tr>
<td>50</td>
<td>−3.58</td>
</tr>
<tr>
<td>100</td>
<td>−3.51</td>
</tr>
<tr>
<td>250</td>
<td>−3.46</td>
</tr>
<tr>
<td>∞</td>
<td>−3.43</td>
</tr>
</tbody>
</table>
The distributions of $T(\hat{\alpha}_iT - \alpha_0)$ when $\alpha_0 = -1$ are just the mirror images of those for $\alpha_0 = 1$, in the sense that

$$P\{\hat{\alpha}_iT - \alpha_0 > c|\alpha_0 = 1\} = P\{\hat{\alpha}_iT - \alpha_0 < -c|\alpha_0 = -1\}.$$ 

**Homework!** Simulate and plot the empirical distributions of $\hat{\alpha}_iT$ under the DGPs (data generating process) with $\alpha_0 = 1$ and $\alpha_0 = -1$ for $T = 100$ and replications 2000.
Augmented Dickey-Fuller Tests

More generally, suppose that $u_t$ has an ARMA($p,q$) representation: $\Phi(B)u_t = \Psi(B)\epsilon_t$, where $\epsilon_t$ are i.i.d. $N(0, \sigma^2)$.

First consider the case that $p = q = 1$. Then

$u_t - \phi_1 u_{t-1} = \epsilon_t - \psi_1 \epsilon_{t-1}$, and

$\epsilon_t$

$$= u_t - \phi_1 u_{t-1} + \psi_1 \epsilon_{t-1}$$

$$= (u_t - \phi_1 u_{t-1}) + \psi_1 (u_{t-1} - \phi_1 u_{t-2} + \psi_1 \epsilon_{t-2})$$

$$= (u_t - \phi_1 u_{t-1}) + \psi_1 (u_{t-1} - \phi_1 u_{t-2}) + \psi_1^2 (u_{t-2} - \phi_1 u_{t-3} + \psi_1 \epsilon_{t-3})$$

$$= \vdots$$

$$= \sum_{j=0}^{\infty} \psi_1^j (u_{t-j} - \phi_1 u_{t-1-j}).$$
We have, since $\epsilon_{t-1} = \sum_{j=1}^{\infty} \psi_1^{j-1}(u_{t-j} - \phi_1 u_{t-1-j})$,

$y_t - y_{t-1}$

$= (\alpha_0 y_{t-1} + u_t) - y_{t-1}$

$= (\alpha_0 - 1)y_{t-1} + \phi_1 u_{t-1} - \psi_1 \epsilon_{t-1} + \epsilon_t$

$= (\alpha_0 - 1)y_{t-1} + \phi_1 u_{t-1} - \psi_1 \sum_{j=1}^{\infty} \psi_1^{j-1}(u_{t-j} - \phi_1 u_{t-1-j}) + \epsilon_t$

$= (\alpha_0 - 1)y_{t-1} + (\phi_1 - \psi_1)[u_{t-1} + \psi_1 u_{t-2} + \psi_1^2 u_{t-3} + \cdots] + \epsilon_t.$
This suggests a regression of $\Delta y_t$ on $y_{t-1}$ and $\Delta y_{t-1}$, $\Delta y_{t-2}$, \ldots, which can be approximated using an autoregression of finite order $k$:

$$
\Delta y_t \approx (\alpha_0 - 1)y_{t-1} + 
\left(\phi_1 - \pi_1\right)\left[\Delta y_{t-1} + \psi_1 \Delta y_{t-2} + \psi_1^2 \Delta y_{t-3} + \cdots \right.
+ \psi_1^{k-1} \Delta y_{t-k}\right] + \epsilon_t
= : \theta_0 y_{t-1} + \theta_1 \Delta y_{t-1} + \theta_2 \Delta y_{t-2} + \cdots + \theta_k \Delta y_{t-k} + \epsilon_t.
$$
To obtain consistent estimates, the order of autoregression, $k$, must be a function of $T$; in particular, Said & Dickey (1984) show that $k$ must be $o(T^{1/3})$ and that there exist $c > 0$ and $r > 0$ such that $ck > T^{1/r}$. 


On the other hand, the $t$-statistic for $\hat{\theta}_{0T}$ has the same limiting distribution as that of $\tau_{\hat{\alpha}_{1T}}$; hence the table in Fuller (1976) can be used. If a constant term is included in the model, the resulting $t$-statistic has the same limiting distribution as that of $\tau_{\hat{\alpha}_{2T}}$. The same idea carries over if $u_t$ follows an ARMA($p,q$) model. Tests of this type are known as the Augmented Dickey-Fuller (ADF) tests.
Phillips-Perron Tests

More general results are available when $u_t$ are weakly dependent. Let

$$\sigma^2 := \lim_{T \to \infty} \frac{1}{T} E \left[ \left( \sum_{t=1}^{T} u_t \right)^2 \right], \quad \sigma_u^2 := \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E(u_t^2).$$
Define $Y_T$ in $D[0, 1]$ as

$$Y_T(r) = \frac{1}{\sigma \sqrt{T}} \sum_{t=1}^{[Tr]} u_t,$$

where $[Tr]$ denotes the integer part of $Tr$. We assume suitable conditions under which the FCLT holds, i.e., $Y_T \Rightarrow W$, where $W$ is a standard Wiener process. More specific conditions can be found in, e.g., Phillips (1987) and Wooldridge & White (1988).
The results below are fundamental for Phillips (1987) and many other papers written by Phillips and his co-authors. Without loss of generality we set $y_0 = 0$. Let

$$W^* = W - \int_0^1 W(r) \, dr.$$ 

If $y_t = \sum_{i=1}^{t} u_i$,

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

(2) $T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 \, dr$;
\( T^{-1} \sum_{t=1}^{T} y_{t-1} u_t \Rightarrow \)
\[ \sigma^2 \int_0^1 W(r) \, dW(r) + \frac{1}{2}(\sigma^2 - \sigma_u^2) = \frac{1}{2}(\sigma^2 W(1)^2 - \sigma_u^2); \]

\( T^{-3/2} \sum_{t=1}^{T} t u_t \Rightarrow \)
\[ \sigma \int_0^1 r \, dW(r) = \sigma \left( W(1) - \int_0^1 W(r) \, dr \right); \]

\( T^{-5/2} \sum_{t=1}^{T} t y_{t-1} \Rightarrow \sigma \int_0^1 r W(r) \, dr; \]

\( T^{-2} \sum_{t=1}^{T} (y_{t-1} - \bar{y}_{-1})^2 \Rightarrow \sigma^2 \int_0^1 W^*(r)^2 \, dr; \]

\( T^{-1} \sum_{t=1}^{T} (y_{t-1} - \bar{y}_{-1}) u_t \Rightarrow \)
\[ \sigma^2 \int_0^1 W^*(r) \, dW(r) + \frac{1}{2}(\sigma^2 - \sigma_u^2). \]
Under the null hypothesis that \( y_t = y_{t-1} = u_t \), where \( \{u_t\} \) satisfies the FCLT, then

(a) For model (1),

\[
T(\hat{\alpha}_1 T - 1) \Rightarrow \frac{1}{2} \left( W(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right) \frac{1}{\int_0^1 W(r)^2 \, dr} ;
\]

\[
\tau \hat{\alpha}_1 T \Rightarrow \frac{\sigma}{2\sigma_u} \left( W(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right) \left( \int_0^1 W(r)^2 \, dr \right)^{1/2}.
\]
(b) Let $\lambda = \frac{1}{2}(\sigma^2 - \sigma_u^2)$ and $\lambda^* = \lambda/\sigma^2$. For model (2),

$$T(\hat{\alpha}_{2T} - 1) \Rightarrow \frac{\int_0^1 W^*(r) \, dW(r) + \lambda^*}{\int_0^1 W^*(r)^2 \, dr};$$

and
\[ \tau_{\hat{\alpha}_{2T}} \Rightarrow \frac{\sigma}{\sigma_u} \left( \int_0^1 W^*(r) \, dW(r) + \lambda^* \right) \left( \int_0^1 W^*(r)^2 \, dr \right)^{1/2}; \]

\[ \tau_{\hat{\mu}_{2T}} \Rightarrow \frac{\sigma}{\sigma_u} \left[ \left( \int_0^1 W(r)^2 \, dr \right) W(1) - \left( \int_0^1 W(r) \, dr \right) \left( \int_0^1 W(r) \, dW(r) + \lambda^* \right) \right] \left( \int_0^1 W^*(r)^2 \, dr \right)^{1/2} \left( \int_0^1 W(r)^2 \, dr \right)^{1/2} \]

\[ \varphi_1 \Rightarrow \frac{\sigma^2}{2\sigma_u^2} \left[ \left( \int_0^1 W^*(r) \, dW(r) + \lambda^* \right)^2 \left( \int_0^1 W^*(r)^2 \, dr \right) + W(1)^2 \right]. \]
Remarks

(1) For model (3), the limits of $T(\hat{\alpha}_{3T} - 1)$, $\tau_{\hat{\mu}_{3T}}$, $\tau_{\hat{\beta}_{3T}}$, and $\tau_{\hat{\alpha}_{3T}}$ can be found in Phillips & Perron (1988), and the limits of $\phi_2$ and $\phi_3$ can be found in Perron (1990).

(2) The OLS estimators $\hat{\alpha}_{iT}$, $i = 1, 2, 3$, are all "supper consistent" in the sense that $\hat{\alpha}_{iT} - 1$ is $O_p(T^{-1})$, whereas in the standard case such that $|\alpha_0| < 1$, $\alpha_{iT} - \alpha_0$ is $O_p(T^{-1/2})$. Moreover, the OLS estimators remain consistent even when $y_{t-1}$ and serially correlated disturbances are both present because the sample variation $\sum_{t=1}^{T} y_{t-1}^2$ grows much faster than the regressor-error correlation $\sum_{t=1}^{T} y_{t-1} u_t$. 
Note that under the null hypothesis, the estimators $\hat{\sigma}_T^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2$ and $T^{-1} \sum_{t=1}^{T} (y_t - y_{t-1})^2$ are consistent for $\sigma_u^2$ by suitable law of large numbers. Also observe that $\sigma^2$ is the limit of

$$
\frac{1}{T} E \left[ \left( \sum_{t=1}^{T} u_t \right)^2 \right] = \frac{1}{T} \left( \sum_{t=1}^{T} E(u_t^2) + 2 \sum_{\tau=1}^{T-1} \sum_{t=\tau+1}^{T} E(u_t u_{t-\tau}) \right),
$$

which can be consistently estimated using the following (non-parametric) estimator

$$
S^2_{T,n(T)} = \frac{1}{T} \left( \sum_{t=1}^{T} \hat{u}_t^2 + 2 \sum_{\tau=1}^{n(T)} \sum_{t=\tau+1}^{T} \hat{u}_t \hat{u}_{t-\tau} \right),
$$

where $w(\cdot)$ is some kernel.
Note that $n(T)$ characterizes kernel’s bandwidth which should be growing with $T$ but at a slower rate; especially, $n(T)$ can be $o(T^{1/2})$. If we use the so-called Bartlett kernel:

$$w_{\tau n} = \begin{cases} 1 - \tau/(n(T) + 1), & \text{if } 0 \leq \tau/n(T) \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

we obtain the Newey & West (1987) estimator; if we use the so-called Parzen kernel:

$$w_{\tau n} = \begin{cases} 1 - 6[\tau/n(T)]^2 + 6[\tau/n(T)]^3, & \text{if } 0 \leq \tau/n(T) \leq 1/2, \\ 2[1 - \tau/n(T)]^3, & \text{if } 1/2 \leq \tau/n(T) \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

we obtain the Gallant (1987) estimator.
Limiting Distributions of D-F Tests

If $u_t$ are i.i.d. with mean zero and variance $\sigma^2$, then $\sigma_u^2 = \sigma^2$ and $\lambda = \lambda^* = 0$. Under the null hypothesis that $y_t = y_{t-1} + u_t$, where $\{u_t\}$ satisfies the FCLT, then:

(a) For model (1),

$$T(\hat{\alpha}_{1T} - 1) \Rightarrow \frac{1}{2} \left( \left. W'(1)^2 \right| - 1 \right) \left( \int_0^1 W(r)^2 \, dr \right);$$

$$\tau_{\hat{\alpha}_{1T}} \Rightarrow \frac{1}{2} \left( \left. W'(1)^2 \right| - 1 \right) \left( \int_0^1 W(r)^2 \, dr \right)^{1/2}.$$
(b) for model (2),

\[
T(\hat{\alpha}_{2T} - 1) \Rightarrow \frac{\int_0^1 W^*(r) \, dW(r)}{\int_0^1 W^*(r)^2 \, dr};
\]

\[
\tau_{\hat{\alpha}_{2T}} \Rightarrow \frac{\left( \int_0^1 W^*(r) \, dW(r) \right)}{\left( \int_0^1 W^*(r)^2 \, dr \right)^{1/2}};
\]

\[
\tau_{\hat{\mu}_{2T}} \Rightarrow \frac{\left( \int_0^1 W(r)^2 \, dr \right) W(1) - \left( \int_0^1 W(r) \, dr \right) \left( \int_0^1 W(r) \, dW(r) \right)}{\left( \int_0^1 W^*(r)^2 \, dr \right)^{1/2} \left( \int_0^1 W(r)^2 \, dr \right)^{1/2}};
\]

\[
\varphi_1 \Rightarrow \frac{1}{2} \left[ \frac{\left( \int_0^1 W^*(r) \, dW(r) \right)^2}{\int_0^1 W^*(r)^2 \, dr} + W(1)^2 \right].
\]
As the limits of the DF tests depend on nuisance parameters when $u_t$ are weakly dependent, critical values cannot be tabulated. Fortunately, the nuisance parameters can be eliminated even when $u_t$ are not i.i.d.
Consider

\[
Z(\hat{\alpha}_1 T) = T(\hat{\alpha}_1 T - 1) - \frac{\frac{1}{2}(s^2 T n - \hat{\sigma}^2_T)}{T^{-2} \sum_{t=1}^{T} y_{t-1}^2}
\]

\[
\Rightarrow \quad \frac{\frac{1}{2}(W(1)^2 - \sigma_u^2 / \sigma^2)}{\int_0^1 W(r)^2 \, dr} - \frac{\frac{1}{2}(\sigma^2 - \sigma_u^2)}{\sigma^2 \int_0^1 W(r)^2 \, dr}
\]

\[
= \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 \, dr},
\]

which is the same as the limit of \( T(\hat{\alpha}_1 T - 1) \). Therefore, the table in Fuller (1976) can be directly used for this test. A modified

\[
Z'_{\alpha} = Z(\hat{\alpha}_1 T) / \sqrt{2}
\]

is related to the empirical tables in Evans & Savin (1981).
Similarly,

\[
Z(\hat{\tau}_{1T}) = \frac{\hat{\sigma}_T}{s_{TN}} \tau_{\hat{\alpha}_{1T}} - \frac{1}{2} \left( \frac{s_{TN}^2 - \hat{\sigma}_T^2}{s_{TN} \left( T - 2 \sum_{t=1}^{T} y_{t-1}^2 \right)^{1/2}} \right)
\]

\[
\Rightarrow \frac{1}{2} \left( W(1)^2 - \frac{\sigma_u^2}{\sigma^2} \right) \left( \int_0^1 W(r)^2 \, dr \right)^{1/2} - \frac{1}{2} \left( \frac{\sigma^2 - \sigma_u^2}{\sigma^2} \right) \left( \int_0^1 W(r)^2 \, dr \right)^{1/2}
\]

\[
= \frac{1}{2} \left( W(1)^2 - 1 \right) \left( \int_0^1 W(r)^2 \, dr \right)^{1/2}.
\]
ERS (1996) propose a simple modification of the ADF tests in which the data are detrended so that explanatory variables are ”taken out” of the data prior to running the test regression. ERS define a quasi-difference of that depends on the value representing the specific point alternative against which we wish to test the null:

\[
d(y_t | \alpha) = \begin{cases} 
  y_t & \text{if } t = 1, \\
  y_t - ay_{t-1} & \text{if } t < 1, 
\end{cases} \quad d(x_t | \alpha) = \begin{cases} 
  x_t & \text{if } t = 1, \\
  x_t - ay_{t-1} & \text{if } t < 1, 
\end{cases}
\]
Next, consider an OLS regression of the quasi-differenced data \( d(y_t|a) \) on the quasi-differenced \( d(x_t|a) \):

\[
d(y_t|a) = d(x_t|a)'\delta(a) + u_t
\]

where \( x_t \) contains either a constant, or a constant and trend, and let \( \hat{\delta}(a) \) be the OLS estimates from this regression.
Now, all that we need know is a value for $a$. ERS recommend the use of $\bar{a}$, where:

$$\bar{a} = \begin{cases} 
1 - \frac{7}{T} & \text{if } x_t = 1, \\
1 - \frac{13.5}{T} & \text{if } x_t = (1, t)'. 
\end{cases}$$

Then, define the **GLS detrended** data, $y_t^d$ using the estimates associated with the $\bar{a}$:

$$y_t^d = y_t - x_t' \hat{\delta}(\bar{a}).$$
Then the DFGLS test involves estimating the standard ADF test equation after substituting the GLS detrended $y^d_t$ for the original $y_t$:

$$\triangle y^d_t = \alpha y^d_{t-1} + \sum_{j=1}^{p} \beta_j \triangle y^d_{t-j} + v_t.$$ 

Note that since the $y^d_t$ are detrended, we do not include the $x_t$ in the DFGLS test equation. As with the ADF test, we consider the $t$-ratio for $\alpha$ from this test equation. While the DFGLS -ratio follows a Dickey-Fuller distribution in the constant only case, the asymptotic distribution differs when you include both a constant and trend. ERS (1996, Table 1, p. 825) simulate the critical values of the test statistic in this latter setting for $T = \{50, 100, 200, \ldots, \infty\}$. 

M.-Y. Chen (I(0) vs I(1))
Homework! Simulate the empirical sizes and powers of ADF and DFGLS tests under the DGP: \( y_t = 0.1 + \alpha_0 y_{t-1} + u_t \) with \( \alpha_0 = 1 \) and \( \alpha_0 = 0.8 \). The sample size and replications are considered as \( T = 100 \) and 2000.
Unit Root Tests in gretl

The unit root tests of ADF and ADF-GLS are provided in gretl. These two tests are accessed via 「Variable (V)」 in toolbar of gretl.
Unit Root Tests in R

The following packages in R provide unit root tests:

1. urca; see twi_unit.R
2. tseries
3. fUnitRoots

In urca, the unit root tests can be implemented by using

\[ \text{urTest}(x, \text{method} = \text{c("unitroot", "adf", "urers", "urkpss", "urpp", "ursp", "urza")), title = \text{NULL}, \text{description} = \text{NULL}, ...) \]
Unit Root Tests in urca

1. **ur.df**: Augmented-Dickey-Fuller Unit Root Test
   
   \[
   \text{ur.df}(y, \text{type} = \text{c("none", "drift", "trend")}, \text{lags} = 1, \text{selectlags} = \text{c("Fixed", "AIC", "BIC")})
   \]

2. **ur.ers**: Elliott, Rothenberg & Stock Unit Root Test
   
   \[
   \text{ur.ers}(y, \text{type} = \text{c("DF-GLS", "P-test")}, \text{model} = \text{c("constant", "trend")}, \text{lag.max} = 4)
   \]

3. **ur.kpss**: Kwiatkowski et al. Unit Root Test
   
   \[
   \text{ur.kpss}(y, \text{type} = \text{c("mu", "tau")}, \text{lags} = \text{c("short", "long", "nil")}, \text{use.lag} = \text{NULL})
   \]

4. **ur.pp**: Phillips & Perron Unit Root Test
   
   \[
   \text{ur.pp}(x, \text{type} = \text{c("Z-alpha", "Z-tau")}, \text{model} = \text{c("constant", "trend")}, \text{lags} = \text{c("short", "long")}, \text{use.lag} = \text{NULL})
   \]
Unit Root Tests in urca: continue

5. `ur.sp`: Schmidt & Phillips Unit Root Test
   `ur.sp(y, type = c("tau", "rho"), pol.deg = c(1, 2, 3, 4), signif = c(0.01, 0.05, 0.1))`

6. `ur.za`: Zivot & Andrews Unit Root Test
   `ur.za(y, model = c("intercept", "trend", "both"), lag=NULL)`
Unit Root Tests in \texttt{tseries}

1. \texttt{adf.test}: Augmented Dickey-Fuller Test
   \begin{verbatim}
   adf.test(x, alternative = c("stationary", "explosive"), k =
   trunc((length(x) - 1)(1/3)))
   \end{verbatim}

2. \texttt{kpss.test}: KPSS Test for Stationarity
   \begin{verbatim}
   kpss.test(x, null = c("Level", "Trend"), lshort = TRUE)
   \end{verbatim}

3. \texttt{pp.test}: Phillips-Perron Unit Root Test
   \begin{verbatim}
   pp.test(x, alternative = c("stationary", "explosive"), type =
   c("Z(\alpha)", "Z(t_{\alpha})"), lshort = TRUE)
   \end{verbatim}
Unit Root Tests in \texttt{fUnitRoots}

1. \texttt{adfTest}: Augmented Dickey-Fuller test for unit roots,
   \[
   \texttt{adfTest}(x, \text{lags} = 1, \text{type} = \text{c("nc", "c", "ct")}, \text{title} = \text{NULL}, \text{description} = \text{NULL})
   \]

2. \texttt{unitrootTest}: the same based on McKinnons’s test statistics.
   \[
   \texttt{unitrootTest}(x, \text{lags} = 1, \text{type} = \text{c("nc", "c", "ct")}, \text{title} = \text{NULL}, \text{description} = \text{NULL})
   \]

3. \texttt{urdfTest}: Augmented Dickey-Fuller test for unit roots,

4. \texttt{urersTest}: Elliott-Rothenberg-Stock test for unit roots,

5. \texttt{urkpssTest}: KPSS unit root test for stationarity,

6. \texttt{urppTest}: Phillips-Perron test for unit roots,

7. \texttt{urspTest}: Schmidt-Phillips test for unit roots,

8. \texttt{urzaTest}: Zivot-Andrews test for unit roots.
Testing Stationarity Against a Unit Root

We have learned that the DF tests usually have low power against trend stationarity. To complement the aforementioned unit-root tests, it is quite natural to consider tests from an opposite direction, i.e., tests of the null hypothesis of (trend) stationarity against a unit root.
KPSS Tests

Consider a linear regression model

\[ y_t = x_t a_t + z_t' b_0 + \epsilon_t, \quad t = 1, \cdots, T, \]

where \( a_t = a_{t-1} + v_t \) with the initial value \( a_0 \), and \( v_t \) are i.i.d. \( N(0, \sigma^2_v) \) independent of \( \epsilon_t \) which are also i.i.d. \( N(0, \sigma^2_\epsilon) \).

Under the null that \( a_t = a_0, \sigma^2_v = 0 \). Under the alternative of random walk,

\[ y_t = x_t a_0 + z_t' b_0 + x_t \sum_{j=1}^{t} v_j + \epsilon_t, \quad t = 1, \cdots, T. \]
Nabeya & Tanaka (1988) derive the locally best invariant test for \( \sigma_v^2 = 0 \), which is based on the ratio of two quadratic forms:

\[
\frac{e' D_x A_T D_x e}{e'e},
\]

where \( D_x = \text{diag}(x_1, \ldots, x_T) \), \( A_T \) is such that its \((s, t)\)th element is \( \min(s, t) \), and \( e \) is the residual vector from regressing \( y_t \) on \( x_t \) and \( z_t \).
Kwiatkowski, Phillips, Schmidt, & Shin (1992) apply this test to test for the null of (trend) stationarity. Consider the DGP:

\[ y_t = a_0 + b_0 t + \epsilon_t, \]

which is a special case of the null model of Nabeya & Tanaka (1988) with \( x_t = 1 \) and \( z_t = t \); when \( b_0 = 0 \), this is just a level-stationary model. Under the alternative that \( y_t \) is an \( I(1) \) process with drift,

\[ y_t = a_0 + b_0 t + y_t^*, \]

where \( y_t^* = \sum_{j=1}^{t} v_j \).
Note that $A_T$ can be written as $C'_T C_T$, where $C_T$ is an upper triangular matrix with all elements on and above the main diagonal being 1. In this case, $D_x = I_T$ so that $e' D_x A_T D_x e = e' C'_T C_T e$, and it can be verified that $C_T e$ is a vector containing reversed partial sums of $e_t$, i.e., the $t$-th element is $R_t = \sum_{i=t}^{T} e_i$. Let $S_t$ denote partial sums of $e_t$. Then, $R_1 = S_T = 0$ and $S_t = -R_{t+1}$ for $t = 1, \ldots, T - 1$. 
Nabeya and Tanaka’s statistics can be written as:

\[ \eta_i = \frac{T^{-2} \sum_{t=1}^{T} R_{t}^2}{\hat{\sigma}_T^2} = \frac{T^{-2} \sum_{t=1}^{T} S_{t}^2}{\hat{\sigma}_T^2}, \quad i = 1, 2, \]

with \( \hat{\sigma}_T^2 = e'e/T \), where \( e \) is obtained from regressing \( y_t \) on the constant 1 for the null of level stationarity \((i = 1)\) and from regressing \( y_t \) on 1 and \( t \) for the null of trend stationarity \((i = 2)\).
To allow for weakly dependent and heterogeneous $\epsilon_t$, the statistics become

$$\eta_i = \frac{T^{-2} \sum_{t=1}^{T} S_t^2}{S_{Tn}^2}, \quad i = 1, 2,$$

where $S_{Tn}^2$ is the Newey-West or Gallant estimator of

$$\sigma^2 = \lim_T T^{-1} \text{IE}(S_T^2);$$

these two tests will be referred to as the KPSS tests.
Theorem: KPSS

In the level-stationary model,

$$\eta_1 \Rightarrow \int_0^1 W^0(r)^2 \, dr,$$

where $W^0$ is a Brownian bridge; in the trend-stationary model,

$$\eta_2 \Rightarrow \int_0^1 V(r)^2 \, dr;$$

where

$$V(r) = W(r) + (2r - 3r^2)W(1) - (6r - 6r^2) \int_0^1 W(s) \, ds.$$
These tests suffer similar problems as the PP test. When $\epsilon_t$ are highly correlated, the empirical sizes are too high if the truncation lag $n$ for $s^2_{Tn}$ is too small, say, 4. A large truncation lag, however, has an adverse effect on the power of test.
Table 2: Asymptotic critical values of the KPSS tests.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>1%</th>
<th>2.5%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_1$</td>
<td>0.739</td>
<td>0.574</td>
<td>0.463</td>
<td>0.347</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>0.216</td>
<td>0.176</td>
<td>0.146</td>
<td>0.119</td>
</tr>
</tbody>
</table>
Panel Unit Root Tests

Consider the model

$$y_{it} = \rho_{it-1} y_{it-1} + z'_{it} \gamma + u_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T,$$  \hspace{1cm} (4)

where $z_{it}$ is the deterministic component and $u_{it}$ is a stationary process. $z_{it}$ could be zero, one, the fixed effects, $\mu_i$, or fixed effects as well as a time trend, $t$. 

M.-Y. Chen I(0) vs I(1)

Levin and Lin (LL) tests assume that $\rho_i = \rho$ for all $i$ and are interesting in testing

$$H_0 : \rho = 1 \quad \text{v.s.} \quad H_a : \rho < 1.$$ 

Denote $\hat{\rho}$ as the OLS estimator of $\rho$ in (4), Levin and Lin (2002) show that

$$\sqrt{NT}(\hat{\rho} - 1) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \tilde{y}_{i,t-1} \tilde{u}_{it} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} \tilde{y}_{i,t-1}^2$$

$$t_{\hat{\rho}} = \frac{(\hat{\rho} - 1) \sqrt{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{i,t-1}^2}}{s_e},$$

$$s_e^2 = \frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{u}_{it}^2.$$
As $T \to \infty$ and then $N \to \infty$,

<table>
<thead>
<tr>
<th>$z_{it}$</th>
<th>$\hat{\rho}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\sqrt{NT}(\hat{\rho} - 1) \Rightarrow N(0, 2)$</td>
</tr>
<tr>
<td>1</td>
<td>$\sqrt{NT}(\hat{\rho} - 1) \Rightarrow N(0, 2)$</td>
</tr>
<tr>
<td>$\mu_i$</td>
<td>$\sqrt{NT}(\hat{\rho} - 1) + 3\sqrt{N} \Rightarrow N(0, 51/5)$</td>
</tr>
<tr>
<td>$(\mu_i, t)$</td>
<td>$\sqrt{NT}(\hat{\rho} - 1) + 7.5) \Rightarrow N(0, 2895/112)$</td>
</tr>
</tbody>
</table>
As $T \to \infty$ and then $N \to \infty$,

<table>
<thead>
<tr>
<th>$z_{it}$</th>
<th>$t_{\hat{\rho}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$t_{\hat{\rho}} \Rightarrow N(0, 1)$</td>
</tr>
<tr>
<td>1</td>
<td>$t_{\hat{\rho}} \Rightarrow N(0, 1)$</td>
</tr>
<tr>
<td>$\mu_i$</td>
<td>$\sqrt{1.25}t_{\hat{\rho}} + \sqrt{1.875N} \Rightarrow N(0, 1)$</td>
</tr>
<tr>
<td>$(\mu_i, t)$</td>
<td>$\sqrt{448/277}(t_{\hat{\rho}} + \sqrt{3.75N}) \Rightarrow N(0, 1)$</td>
</tr>
</tbody>
</table>

In case $u_{it}$ is stationary, the asymptotic distributions of $\hat{\rho}$ and $t_{\hat{\rho}}$ need to be modified due to the presence of serial correlation.
Harris and Tzavalis (1999)’s Tests

Harris and Tzavalis (1999) also derived unit root tests for (4) with \( z_{it} = \{0\}, \{\mu_i\}, \) or \( \{\mu_i, t\} \) when the time dimension of the panel, \( T \), is fixed. This is typical case for micro panel studies. The main results are

\[
\begin{align*}
\text{when } & z_{it} = \{0\}, \{\mu_i\}, \text{ or } \{\mu_i, t\}, \\
& \sqrt{N} (\hat{\rho} - 1) \Rightarrow N \left( 0, \frac{2}{T(T-1)} \right) \quad (i = 1, \ldots, N) \\
& \sqrt{N} (\hat{\rho} - 1 + \frac{3}{T+1}) \Rightarrow N \left( 0, \frac{3(17T^2 - 20T + 17)}{5(T-1)(T+1)^3} \right) \quad (i = 1, \ldots, N) \\
& \sqrt{N} (\hat{\rho} - 1 + \frac{15}{2(T+2)}) \Rightarrow N \left( 0, \frac{15(193T^2 - 728T + 1147)}{112(T+1)^3(T-2)} \right) \quad (i = 1, \ldots, N) \\
\end{align*}
\]
Im, Pesaran and Shin (2003)’s Tests

The tests of Levin and Lin (1992) are restrictive in the sense that it requires $\rho$ to be homogeneous across $i$. Im, Pesaran and Shin (2003) allow for heterogeneous coefficient of $y_{it-1}$ and proposed an alternative testing procedure based on the augmented DF tests when $u_{it}$ is serially correlated with different serial correlation properties across cross-sectional units, i.e., $u_{it} = \sum_{j=1}^{p_i} \psi_{ij} u_{it-j} + \epsilon_{it}$. Substituting this $u_{it}$ in (4), we get

$$y_{it} = \rho_i y_{it-1} + \sum_{j=1}^{p_i} \psi_{ij} \Delta y_{it-j} + z'_{it} \gamma + \epsilon_{it}, i = 1, \ldots, N, t = 1, \ldots, T.$$
The null hypothesis is

\[ H_0 : \rho_i = 1 \]

for all \( i \) against the alternative hypothesis

\[ H_a : \rho_i < 1 \]

for at least one \( i \). The \( t \)-statistic suggested by Im, Pesaran and Shin (2003) is defined as

\[ \bar{t} = \frac{1}{N} \sum_{i=1}^{N} t_{\hat{\rho}_i}, \quad (6) \]

where \( t_{\hat{\rho}_i} \) is the individual \( t \)-statistic of testing \( H_0 : \rho_i = 1 \) in (5). It is known that for a fixed \( N \),

\[ t_{\hat{\rho}_i} \Rightarrow \frac{\int_0^1 W_iZ dW_iZ}{\left[ \int_0^1 W_i^2 Z \right]^{1/2}} = t_i T \quad (7) \]

as \( T \to \infty \).
Im, Pesaran and Shin (2003) assume that $t_{iT}$ are i.i.d. and have finite means and variances. Then

$$\sqrt{N} \left( \frac{1}{N} \sum_{i=1}^{N} t_{iT} - E[t_{iT} | \rho_i = 1] \right) \sim \sqrt{\text{var}[t_{iT} | \rho_i = 1]} \Rightarrow N(0, 1)$$

as $N \rightarrow \infty$ by the Lindeberg-Levy central limit theorem.
Hence, the test statistic of Im, Pesaran and Shin (2003) has the limiting distribution as

\[ t_{IPS} = \frac{\sqrt{N}(\bar{t} - E[t_{iT}|\rho_i = 1])}{\sqrt{\text{var}[t_{iT}|\rho_i = 1]}} \Rightarrow N(0, 1) \]

as \( T \to \infty \) followed by \( N \to \infty \) sequentially. The values of \( E[t_{iT}|\rho_i = 1] \) and \( \text{var}[t_{iT}|\rho_i = 1] \) have been computed by Im, Pesaran and Shin (2003) via simulations for different values of \( T \) and \( p_i \)'s.
Combining $p$-value Tests

Let $p_i$ be the $p$-value of a unit root test for cross-section $i$, Maddala and Wu (1999) and Choi (1999) proposed a Fisher type test as:

$$P = -2 \sum_{i=1}^{N} \ln p_i$$

which combines the $p$-value from unit root tests for each cross-section $i$ to test for unit root in panel data. $P$ is distributed as $\chi^2$ with $2N$ degrees of freedom as $T_i \to \infty$ for all $N$. When $p_i$ closes to 0 (null hypothesis is rejected), $\ln p_i$ closes to $-\infty$ so that large value $P$ will be found and then the null hypothesis of existing panel unit root will be rejected. In contrast, when $p_i$ closes to 1 (null hypothesis is not rejected), $\ln p_i$ closes to 0 so that small value $P$ will be found and then the null hypothesis of existing panel unit root will not be rejected.
Choi (1999) pointed out the advantages of the Fisher test: (1) the cross-sectional dimension, $N$, can be either finite or infinite, (2) each group can have different types of nonstochastic and stochastic components, (3) the time series dimension, $T$ can be different for each $i$ (imbalance panel data), and (4) the alternative hypothesis would allow some groups to have unit roots while others may not. A main disadvantage involved is that the $p$-value have to be derived by Monte Carlo simulations.

When $N$ is large, Choi (1999) also proposed a $Z$ test,\

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (-2 \ln p_i - 2) \tag{9}$$

since $E[-2 \ln p_i] = 2$ and $\text{var}[-2 \ln p_i] = 4$. Assume $p_i$’s are i.i.d. and use the Lindeberg-Levy central limit theorem to get 

$$Z \Rightarrow N(0, 1)$$

as $T_i \to \infty$ followed by $N \to \infty$. 
Hadri (1999)’s Test: KPSS Type

Let \( \hat{e}_{it} \) be the residuals from the regression:

\[
y_{it} = z_{it}' \gamma + e_{it}
\]  

(10)

and \( \hat{\sigma}_e^2 \) be the estimate of the error variance. Also, let \( S_{it} \) be the partial sum process of the residuals,

\[
S_{it} = \sum_{j=1}^{t} \hat{e}_{ij}.
\]

Then the LM statistic is

\[
LM = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} S_{it}^2 \frac{\hat{\sigma}_e^2}{\hat{\sigma}_e^2}.
\]
It can be shown that

\[ LM \xrightarrow{p} E \left[ \int W_{iZ} \right] \]

as \( T \to \infty \) followed by \( N \to \infty \) provided \( E[\int W_{iZ}^2] < \infty \).

Also,

\[
\frac{\sqrt{N}(LM - E[\int W_{iZ}^2])}{\sqrt{\text{var}[\int W_{iZ}^2]}} \Rightarrow N(0, 1)
\]

as \( T \to \infty \) followed by \( N \to \infty \).
Panel Unit Roots Tests in R: CADFtest

The asymptotic p-values of the Hansen’s (1995) Covariate-Augmented Dickey Fuller (CADF) test for a unit root are computed using the approach outlined in Costantini et al. (2007).

1. \textbf{CADFpvalues}: p-values of the CADF test for unit roots
   \[ \text{CADFpvalues}(t0, \text{rho2} = 0.5, \text{type}=c(”trend”, ”drift”, ”none” )) \]

2. \textbf{CADFtest}: Hansen’s Covariate-Augmented Dickey Fuller (CADF) test for unit roots
1. **purtest**: Unit root tests for panel data

   ```r
   purtest(object, data = NULL, index = NULL, test =
   c("levinlin", "ips", "madwu", "hadri"), exo = c("none", "intercept", "trend"), lags = c("SIC", "AIC", "Hall"),
   pmax = 10, Hcons = TRUE, q = NULL, dfcor = FALSE,
   fixedT = TRUE, ...)
   ```

Panel Unit Roots Tests in R: **plm**