This note discusses some central issues in the analysis of non-stationary time series. We begin by showing examples of different types of non-stationarity, involving trends, level shifts, variance changes, and unit roots. The three first deviations from stationarity may complicate the econometric analysis, but the tools developed for stationary variables may be adapted to the new situation. The presence of unit roots, however, changes the asymptotic behavior of estimators and test statistics, and a different set of tools for unit root processes has to be applied. We continue to illustrate the properties of a unit root time series, and discuss the issue of unit root testing. In practical applications, testing for unit roots is particularly important, because the conclusion determines what kind of tool-kit that is appropriate for a given problem: For stationary time series we can use the standard tools from linear regression; for unit root time series we have to think on how to combine unit root time series. The latter is called cointegration and is discussed later in the course.

1 STATIONARY AND NON-STATIONARY TIME SERIES

Recall that a time series, $y_t$, is weakly stationary if the mean and variance are the same for all $t = 1, 2, ..., T$, and if the autocovariance, $\gamma_s = Cov(y_t, y_{t-s})$, depends on $s$ but not on $t$. Also recall that if $y_t$ and $x_t$ are stationary and weakly dependent time series, then the linear regression model

$$y_t = x_t' \beta + \epsilon_t, \quad t = 1, 2, ..., T,$$

can be analyzed using standard tools, and most of the results for regression of independent and identical (IID) observations still hold. The technical argument is that there exist a
standard law of large numbers (LLN) and a central limit theorem (CLT) for stationary
time series so that the estimators are consistent and asymptotically normally distributed.

The same thing does not hold for non-stationary time series in general; and econo-
metric analysis of non-stationary time series should always be performed with care. In
macroeconomics this is particularly important, because most observed time series do not
seem to be well characterized as stationary processes.

This note gives an intuitive account for non-stationarity in economic time series. In
Section 2 we argue that time series can be non-stationary in many different ways, and we
present some typical deviations from the stationarity assumption: namely deterministic
trends, level shifts, and changing variances. For each case we briefly discuss how the non-
stationary time series could be treated in practical applications, and it turns out that the
required modifications to the usual regression tools are minor. We proceed by introducing
the concept of unit roots. The presence of a unit root more fundamentally changes the
properties of the time series, and the usual tools no longer apply. In Section 3 we review
the properties of a stationary autoregressive process and discuss the implications of a
unit root. To keep the notation simple we consider a first order autoregression, AR(1),
but parallel results prevail for higher order processes. We then proceed in Section 4 to
discuss the issue of unit root testing, i.e. how a unit root process can be distinguished
from a stationary process. This issue is particularly important in applications, because it
determines the kind of tools that we should apply to the data: For stationary time series
we can apply the usual tools from regression and the interpretation is straightforward.
For unit root processes the tools should be modified and the econometric model should
be interpreted in terms of cointegration; we return to this issue later in the course.

2 Non-Stationarity in Economic Time Series

Many of the macro-economic time series we encounter in practice seem to behave in a non-
stationary manner. There are many types of non-stationarity, however, and this section
briefly discusses some typical examples of non-stationarity in economic data.

2.1 Deterministic Trends and Trend-Stationarity

Macro-economic variables are often trending, i.e. they have a tendency to systematically
increase or decrease over time. As examples you could think of GDP, consumption, prices
etc. The trending behavior means that the unconditional expectation changes over time,
which is not in accordance with the assumption of stationarity.

In some cases the trend is very systematic so that the deviations from the trend is a
stationary variable. In this case we can analyze the deviation from trend, the so-called de-
trended variable, instead of the original one, and since that is stationary process, the usual
results apply. A time series that fluctuates in a stationary manner around a deterministic
linear trend is called a trend-stationary process. As an example you could think of the
stationary AR(1) process with a zero mean,

\[ x_t = \theta x_{t-1} + \epsilon_t, \quad |\theta| < 1, \]

and a new process, \( y_t \), defined as the stationary process plus a linear trend term and a constant,

\[ y_t = x_t + \mu_0 + \mu_1 t. \tag{1} \]

Since \( x_t \) is a stationary process, \( y_t \) is stationary around the trending mean, \( E[y_t] = \mu_0 + \mu_1 t \), i.e. it is trend-stationary. A single realization of \( T = 200 \) observations of the processes \( x_t \) and \( y_t \) (with \( \theta = 0.5 \)) is illustrated in Figure 1 (A).

The main point of a trend-stationary process is that the stochastic part is still stationary, and the non-stationarity is deterministic. In an empirical analysis we could therefore de-trend the variable by running the OLS regression corresponding to (1) and analyze the de-trended variable, \( \hat{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t \). Alternatively we could consider a regression augmented with a linear trend term, i.e.

\[ y_t = \theta y_{t-1} + \delta + \gamma t + \epsilon_t. \]

The two approaches would give similar results, cf. the celebrated Frisch-Waugh-Lovell theorem.
A main explanation for linear trends in economic variables is that productivity increases over time, implying growth in GDP, consumption etc. To illustrate the idea of trend-stationarity, we consider two examples below.

**Example 1 (productivity):** Let $LPROD_t$ denote the log of Danish hourly productivity, 1971 : 1 – 2005 : 2, compiled as the log of real output per hour worked. The time series is depicted in Figure 2 (A). The trend looks very stable and the deviation from the trend looks like a stationary process. To estimate an AR(1) model for the time series we augment the model with a linear trend,

$$LPROD_t = \theta \cdot LPROD_{t-1} + \delta + \gamma t + \epsilon_t,$$

and obtain the results reported in Table 1. The autoregressive parameter is $\theta = 0.56$, and according to the misspecification tests, the model seems to be a reasonable description of the data.

**Example 2 (consumption):** Next, let $LCONS_t$ be the log of private aggregate consumption in Denmark, 1971 : 1 – 2005 : 2, see Figure 2 (B). Again there is a positive trend, but it is less stable, and consumption shows large and persistent deviations from the trend. From a visual inspection it is not clear that the deviations are stationary. If we estimate a second order autoregressive model allowing for a linear trend, we obtain the results in Table 2. In Section 4 we will formally test whether the deviations are stationary, i.e. whether $LCONS_t$ is a trend-stationary process.

---

**Table 1: Modelling $LPROD_t$ by OLS for $t = 1971 : 2 − 2005 : 2$.**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>$t$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LPROD_{t-1}$</td>
<td>0.561273</td>
<td>0.07056</td>
</tr>
<tr>
<td>Constant</td>
<td>0.090890</td>
<td>0.01382</td>
</tr>
<tr>
<td>$t$</td>
<td>0.002412</td>
<td>0.00039</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.016907</td>
<td>log-likelihood</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.994</td>
<td>$T$</td>
</tr>
<tr>
<td><strong>Statistic</strong></td>
<td><strong>[p-val]</strong></td>
<td><strong>Distribution</strong></td>
</tr>
<tr>
<td>No autocorrelation of order 1-2</td>
<td>3.80</td>
<td>[0.15]</td>
</tr>
<tr>
<td>Normality</td>
<td>2.52</td>
<td>[0.28]</td>
</tr>
<tr>
<td>No heteroskedasticity</td>
<td>6.53</td>
<td>[0.16]</td>
</tr>
<tr>
<td>Correct functional form (RESET)</td>
<td>2.66</td>
<td>[0.11]</td>
</tr>
</tbody>
</table>

---
Table 2: Modelling LCONS_t by OLS for t = 1971 : 3 – 2005 : 2.

Figure 2: (A)-(B): Examples of non-stationary time series. (C)-(D): Distribution of the OLS estimator, \( \hat{\theta} \), in an AR(1) model when the true parameter is \( \theta = 0.5 \) and \( \theta = 1 \).

Simulated with \( T = 500 \) observations and 20000 replications.
2.2 Level Shifts and Structural Breaks

Another type of non-stationarity in a time series is if there is a change in the unconditional mean at a given point in time. As an example the mean of a time series could be \( \mu_1 \) for the first half of the sample and \( \mu_2 \) for the second half. Such a case is illustrated in Figure 1 (B). The level shift may be associated with a change in the economic structures, e.g. institutional changes, changes in the definition or compilation of the variables, of a switch from one regime to another. As an typical example you could think of the German reunification, where we might expect the time series to behave differently for unified Germany as compared to Western and Eastern Germany.

From a modelling point of view we often consider the change in the mean as deterministic and include a dummy variable in the model. Defining a dummy variable, \( D_t = 1(t \geq T_0) \), to be zero before observation \( T_0 \) and unity after, we can augment the regression model to take account of the level shift,

\[
y_t = \theta y_{t-1} + \delta + \varphi D_t + \epsilon_t. \tag{2}
\]

If we think that the structural break is more fundamental, e.g. changing all parameters in the model, we may want to model the two regimes separately. That requires sufficient observations in both sub-samples.

2.3 Changing Variances

A third type of non-stationary is related to changes in the variance. Figure 1 (C) illustrates an example:

\[
y_t = 0.5 \cdot y_{t-1} + \epsilon_t,
\]

where \( \epsilon_t \sim N(0, 1) \) for \( t = 1, 2, ..., 100 \), and \( \epsilon_t \sim N(0, 5) \) for \( t = 101, 102, ..., 200 \). Again the interpretation is that the time series covers different regimes, where one regime appears to be more volatile than the other.

If the sub-samples are long enough, a natural solution is again to model the regimes separately. An alternative solution is to try to model the changes in the variance, and later in the course we return to a particular class of models for changing variance, the so-called autoregressive conditional heteroskedasticity (ARCH) models.

2.4 Unit Roots

The final type of non-stationarity presented here is generated by unit roots in autoregressive models. Figure 1 (D) illustrates a so-called random walk,

\[
y_t = y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, 1),
\]

that has a unit root in the characteristic polynomial. Note from the graph that the random walk has no attractor, and wanders arbitrarily far up and down. Unit root processes seem
to be a good description of the behavior of actual time series in many cases, and it is the main focus in the rest of this note.

The most important complication from the introduction of unit root processes is that standard versions of LLN and CLT do not apply. Consequently, we have to develop new statistical tools for the analysis of this case.

As an illustration consider a small Monte Carlo simulation with data generating process (DGP) given by

$$y_t = \theta y_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, 1),$$

for $t = 1, 2, ..., T$, and $y_0 = 0$. For each simulated time series, $m = 1, 2, ..., M$, we estimate an AR(1) model and collect the OLS estimates, $\hat{\theta}_m$. In the simulation we take $T = 500$ observations to illustrate the behavior in large samples, and $M = 20000$ replications.

Figure 2 (C) depicts the distribution of $\hat{\theta}$, when the true parameter is $\theta = 0.5$. This is the standard stationary case, and the distribution is close to normal with a standard deviation of $MCSD(\hat{\theta}) \approx 0.04$. Figure (D) depicts the distribution of $\hat{\theta}$ when the true parameter is $\theta = 1$, i.e. in the presence of a unit root. Note that the distribution is highly skewed compared to the normal distribution, with a long left tail. This reflect that the asymptotic distribution of $\hat{\theta}$ is non-normal in the unit root case. Also note that the distributions is much more condensed (compare the scales of graph (C) and (D)) with $MCSD(\hat{\theta}) \approx 0.006$. This reflects that the estimator is consistent, i.e. $p\lim(\hat{\theta}) = \theta$, and the convergence to the true value is much faster for $\theta = 1$ than for $\theta = 0.5$. This phenomenon of fast convergence is referred to as super-consistency.

3 Stationary and Unit Root Autoregressions

In this section we discuss the properties of an autoregressive model under the stationarity condition and under the assumption of a unit root. To make the derivations as simple as possible we focus on the first order autoregression, but parallel results could have been derived for more general models. We first analyze the case with no deterministic terms, and then discuss the interpretation of deterministic terms in the model.

3.1 Stationary Autoregression

Consider the first order autoregressive, AR(1), model given by

$$y_t = \theta y_{t-1} + \epsilon_t,$$

for $t = 1, 2, ..., T$, where $\epsilon_t$ is an IID($0, \sigma^2$) error term, and the initial value, $y_0$, is given. The characteristic polynomial is given by $\theta(z) = 1 - \theta z$, and the characteristic root is $z_1 = \theta^{-1}$ with inverse root $\phi_1 = z_1^{-1} = \theta$. Recall that the stationarity condition is that the inverse root is located inside the unit circle, and it follows that the process in (3) is stationary if $|\theta| < 1.$
The solution to (3) in terms of the initial value and the error terms can be found from recursive substitution, i.e.

\[ y_t = \theta y_{t-1} + \epsilon_t \]
\[ = \theta(\theta y_{t-2} + \epsilon_{t-1}) + \epsilon_t \]
\[ = \epsilon_t + \theta \epsilon_{t-1} + \theta^2 y_{t-2} \]
\[ = \epsilon_t + \theta \epsilon_{t-1} + \theta^2(\theta y_{t-3} + \epsilon_{t-2}) \]
\[ = \epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + \theta^3 y_{t-3} \]
\[ \vdots \]
\[ = \epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + \ldots + \theta^{t-1} \epsilon_1 + \theta^t \epsilon_0. \]

(4)

An important characteristic of this process is that a shock to \( \epsilon_t \) has only transitory effects because \( \theta^s \) goes to zero for \( s \) increasing. We say that the process has an attractor, and if we could set all future shock to zero, the process \( y_t \) would converge towards the attractor.

In the present case the attractor is the unconditional mean, \( E[y_t] = 0 \).

Figure 3 (A) illustrates this idea by showing one realization of a stationary AR(1) process. We note that the process fluctuates around a constant mean. An extraordinary large shock (at time \( t = 50 \)) increases the process temporarily, but the series will return and fluctuate around the attractor after some periods.

From the solution in (4) we can find the properties of \( y_t \) directly. The mean is

\[ E[y_t] = \theta^t y_0 \to 0. \]

We note that the initial value affects the expectation in small samples, but the effect vanishes for increasing \( t \), so that the expectation is zero in the limit. Likewise, the variance is found to be

\[ V[y_t] = \sigma^2 + \theta^2 \sigma^2 + \theta^4 \sigma^2 + \ldots + \theta^{t-1} \sigma^2 \to \frac{\sigma^2}{1 - \theta^2}, \]

and remember that the autocorrelation function is given by

\[ \rho_s = \text{Corr}(y_t, y_{t-s}) = \theta^s, \]

which goes to zero for increasing \( s \), cf. Figure 3 (C).

3.2 Autoregression with a Unit Root

Now consider the case where the autoregressive parameter in (3) is unity, \( \theta = 1 \), i.e.

\[ y_t = y_{t-1} + \epsilon_t. \]

(5)

Note that unity is now a root in the characteristic polynomial, \( \theta(z) = 1 - z \), so that \( \theta(1) = 0 \), hence the name unit root process. The solution is given by (4) with \( \theta = 1 \), i.e.

\[ y_t = y_0 + \Delta y_1 + \Delta y_2 + \ldots + \Delta y_t = y_0 + \epsilon_1 + \epsilon_2 + \ldots + \epsilon_t = y_0 + \sum_{i=1}^{t} \epsilon_i. \]

(6)
Note the striking differences between (4) and (6). First, the effect of the initial value, \( y_0 \), stays in the unit root process and does not disappear for increasing \( t \). This means that the unconditional expectation of \( y_t \) is simply

\[
E[y_t] = y_0,
\]

and in a sense, the initial value plays the role of a constant term. Secondly, the shocks to the process, \( \epsilon_t \), are accumulated to a random walk component, \( \sum \epsilon_i \). This is called a stochastic trend, and it implies that shocks to the process have permanent effects. This is illustrated in Figure 3 (B), where the large shock at time \( t = 50 \) increases the level of the series permanently. More generally we note that the process is moved around by the shocks with no attractor.

Thirdly, the unconditional variance now increases with \( t \),

\[
V[y_t] = V\left[\sum_{i=1}^{t} \epsilon_i\right] = t\sigma^2,
\]

and the process is clearly non-stationary. The first difference process, \( \Delta y_t = \epsilon_t \), is stationary, however, and the process \( y_t \) is often referred to as integrated of first order, \( I(1) \), meaning that it is a stationary process that has been integrated once. More generally, a time series is integrated of order \( d \), \( I(d) \), if it contains \( d \) unit roots.

We also note that the covariance between \( y_t \) and \( y_{t-s} \) is given by

\[
\text{Cov}(y_t, y_{t-s}) = E[(y_t - y_0)(y_{t-s} - y_0)] = E[(\epsilon_1 + \epsilon_2 + \ldots + \epsilon_t)(\epsilon_1 + \epsilon_2 + \ldots + \epsilon_{t-s})] = (t-s)\sigma^2,
\]

and the autocorrelation is given by

\[
\text{Corr}(y_t, y_{t-s}) = \frac{\text{Cov}(y_t, y_{t-s})}{\sqrt{V[y_t]} \cdot \sqrt{V[y_{t-s}]}} = \frac{(t-s)\sigma^2}{\sqrt{t\sigma^2} \cdot \sqrt{(t-s)\sigma^2}} = \frac{t-s}{\sqrt{t(t-s)}},
\]

which dies out very slowly with \( s \). The autocorrelation function is illustrated for the unit root case in graph (D).

### 3.3 Deterministic Terms

The statistical model in (3) is only valid if the time series under analysis has a zero mean. This is rarely the case, and in practice it is always necessary to include a constant term, and sometimes it is also necessary to allow for a deterministic linear trend. Note from (6), however, that a unit root implies accumulation of the terms in the model, and the interpretation of the deterministic terms changes in the presence of a unit root.

Consider as an example, a model with a constant term,

\[
y_t = \delta + \theta y_{t-1} + \epsilon_t.
\]

If \( |\theta| < 1 \), the solution can be derived as

\[
y_t = \theta^t y_0 + \sum_{i=0}^{t} \theta^i \epsilon_{t-i} + (1 + \theta + \theta^2 + \ldots)\delta,
\]

(7)
Figure 3: Differences between stationary and non-stationary time series. (A) and (B) show one realization of a stationary and non-stationary time series respectively and illustrates the temporary and permanent impact of the shocks. (C) and (D) show the autocorrelation function.

where the mean converges to $(1 + \theta + \theta^2 + ... )\delta \to \delta/(1 - \theta)$. In the case of unit root, $\theta = 1$, we find the solution

$$y_t = y_0 + \sum_{i=1}^{t}(\delta + \epsilon_i) = y_0 + \delta t + \sum_{i=1}^{t} \epsilon_i,$$

(8)

where the constant term is accumulated to a deterministic linear trend, $\delta t$, while the initial value, $y_0$, plays the role of a constant term. The process in (8) is referred to as a random walk with drift. Note that if the constant term is zero, $\delta = 0$, then the solution is the random walk,

$$y_t = y_0 + \sum_{i=1}^{t} \epsilon_i,$$

(9)

and the joint hypothesis, $\theta = 1$ and $\delta = 0$, plays an important role in unit root testing.

It holds in general that the deterministic terms in the model will accumulate, and a linear trend term in an autoregressive equation with a unit root corresponds to a quadratic trend in $y_t$. 

10
4 Unit Root Testing

To test for a unit root in a time series $y_t$, the idea is to estimate a statistical model, and then to test whether $z = 1$ is a root in the autoregressive polynomial, i.e. whether $\theta(1) = 0$. The only thing which makes unit root testing different from hypothesis testing in stationary models is that the asymptotic distributions of the test statistics are not $N(0, 1)$ or $\chi^2(1)$ in general. We say that the test statistics follow non-standard distributions.

Some textbooks and computer programs present unit root tests as a misspecification test that should be routinely applied to time series. This ‘automatic’ approach has the danger that the user forgets the properties of the models involved and the interpretation of the unit root test itself. It is therefore recommended to take a more standard approach, and to consider the unit root hypothesis as any other hypothesis in econometrics.

The first step is to set up a statistical model for the data. To do that we have to determine which deterministic components we want to include and we should ensure that the statistical model is an adequate representation of the structure in the time series, e.g. by testing the specification of the model. Based on the statical model we can test for a unit root by comparing two hypotheses, $H_0$ and $H_A$, bearing in mind the properties of the model under the null ($H_0$) and under the alternative ($H_A$).

4.1 Dickey-Fuller Test in an AR(1)

First consider an AR(1) model given by

$$y_t = \theta y_{t-1} + \epsilon_t,$$

for $t = 1, 2, ..., T$. A unit root implies that $\theta = 1$. The null hypothesis of a unit root is tested against a stationary alternative by comparing the hypotheses

$$H_0 : \theta = 1 \quad \text{against} \quad H_A : -1 < \theta < 1.$$

Note that the alternative is explicitly a stationary model. A test could also be devised against an explosive alternative, but that is rare in empirical applications, and will not be discussed. An alternative but equivalent formulation is obtained by subtracting $y_{t-1}$,

$$\Delta y_t = \pi y_{t-1} + \epsilon_t,$$

where $\pi = \theta - 1 = -\theta(1)$ is the characteristic polynomial evaluated in $z = 1$. The hypothesis $\theta(1) = 0$ translates into

$$H_0 : \pi = 0 \quad \text{against} \quad H_A : -2 < \pi < 0.$$

The Dickey-Fuller (DF) test statistic is simply the $t-$ ratio of $H_0$ in (10) or (11), i.e.

$$\hat{\tau} = \frac{\hat{\theta} - 1}{\text{se}(\hat{\theta})} = \frac{\hat{\pi}}{\text{se}(\hat{\pi})}.$$
so-called Dickey-Fuller distribution, $\text{DF}$, the standard normal distribution in Figure 4 (A). The case of We note again that a unit root in $\theta$ hypothesis $\pi$ where also the non-standard distribution of $\theta$ for $\theta = 1$ illustrated in Figure 2 (D). It follows a so-called Dickey-Fuller distribution, $\text{DF}$, which is tabulated in Table 3 and compared to the standard normal distribution in Figure 4 (A). The 5% asymptotic critical value in the $\text{DF}$-distribution is $-1.94$, which is smaller than the corresponding $-1.64$ from $N(0, 1)$.

It is worth noting that the $\text{DF}$ distribution is derived under the assumption that $\epsilon_t$ is IID. If that is not the case, e.g. if there is autocorrelation, the statistical model could be augmented with more lags, which is the next topic.

### 4.2 Dickey-Fuller Test in an AR(p)

The DF test is easily extended to an autoregressive model of order $p$. Here we consider the case of $p = 3$ lags:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \epsilon_t.$$  

We note again that a unit root in $\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \theta_3 z^3$ corresponds to $\theta(1) = 0$. To avoid testing a restriction on $1 - \theta_1 - \theta_2 - \theta_3$, which involves all $p = 3$ parameters, the model is rewritten as

$$\begin{align*}
y_t - y_{t-1} &= (\theta_1 - 1)y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \epsilon_t \\
y_t - y_{t-1} &= (\theta_1 - 1)y_{t-1} + (\theta_2 + \theta_3)y_{t-2} + \theta_3(y_{t-3} - y_{t-2}) + \epsilon_t \\
y_t - y_{t-1} &= (\theta_1 + \theta_2 + \theta_3 - 1)y_{t-1} + (\theta_2 + \theta_3)(y_{t-2} - y_{t-1}) + \theta_3(y_{t-3} - y_{t-2}) + \epsilon_t \\
\Delta y_t &= \pi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t, \quad \text{(12)}
\end{align*}$$

where $\pi = \theta_1 + \theta_2 + \theta_3 - 1 = -\theta(1)$, $c_1 = -(\theta_2 + \theta_3)$, and $c_2 = -\theta_3$. In equation (12) the hypothesis $\theta(1) = 0$ corresponds to

$$H_0 : \pi = 0 \quad \text{against} \quad H_A : -2 < \pi < 0.$$  

The test statistic is again the $t-$ratio for $H_0$ and it is denoted the augmented Dickey-Fuller (ADF) test. The asymptotic distribution is the same as for the DF test in an AR(1).

<table>
<thead>
<tr>
<th>Distribution</th>
<th>1%</th>
<th>2.5%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N(0,1)$</td>
<td>-2.33</td>
<td>-1.96</td>
<td>-1.64</td>
<td>-1.28</td>
</tr>
<tr>
<td>$\text{DF}$</td>
<td>-2.56</td>
<td>-2.23</td>
<td>-1.94</td>
<td>-1.62</td>
</tr>
<tr>
<td>$\text{DF}_c$</td>
<td>-3.43</td>
<td>-3.12</td>
<td>-2.86</td>
<td>-2.57</td>
</tr>
<tr>
<td>$\text{DF}_l$</td>
<td>-3.96</td>
<td>-3.66</td>
<td>-3.41</td>
<td>-3.13</td>
</tr>
</tbody>
</table>

*Table 3: Asymptotic critical values for the Dickey-Fuller unit root test. Reproduced from Davidson and MacKinnon (1993).*
We note that it is only the test for $\pi = 0$ that follows the DF distribution, while tests related to $c_1$ and $c_2$ have standard asymptotics. The reason for this difference is that the hypothesis, $c_1 = 0$, does not introduce any unit roots.

For practical purposes, a unit root test is therefore just performed as a test for $\pi = 0$ in the regression (12), where we include sufficient lags to ensure that the errors are IID. To determine the number of lags, $p$, we can use the standard procedures. One approach is to use general-to-specific testing. One starts with a maximum lag length, $p_{\text{max}}$, and insignificant lags are then deleted. Most people prefer to remove the longest lags first and to avoid holes in the lag structure, but that is not necessary. Another possibility is to use information criteria to select the best model. In any case it is important to ensure that the model is well-specified before the unit root test is applied.

Some authors suggest to calculate the DF test for all values of $p$, and to look at the whole range. This is presented as a robustness check, but the interpretation is not as simple as it sounds. If the regression model includes fewer lags than the true model, then there is autocorrelation by construction and the DF distribution is no longer valid. And if the regression model includes too many lags, the parameters are imprecisely estimated, which also deteriorates the test. In practice it is therefore more relevant to carefully model the process, and to perform the DF test in the preferred model.

### 4.3 Dickey-Fuller Test with a Constant Term

In practice we always include deterministic variables in the model, and the unit root test has to be adapted to this situation. The DF regression with a constant term is given by

$$
\Delta y_t = \delta + \pi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t.
$$

The hypothesis of a unit root is unchanged $H_0 : \pi = 0$, and as a test statistic we can use the $t-$ratio

$$
\hat{\tau}_c = \frac{\hat{\pi}}{\text{se}(\pi)}.
$$
There are two important things to note. First, the presence of the constant term in the regression changes the asymptotic distribution. The asymptotic distribution that allows for a constant, $DF_c$, is illustrated in Figure 4 (A), and the critical values are reported in Table 3. We note that the constant term shifts the distribution to the left and the 5% critical values is $-2.86$.

Secondly, under the null hypothesis, $\pi = 0$, the constant term is accumulated to a linear trend, and the DF $t$-test actually compares the model in (7) and (8), i.e. a stationary model with a non-zero level with a random walk with drift. This is not a natural comparison, and there is an implicit assumption that also $\delta = 0$ under the null so that the drift disappears; but this restriction is not imposed in estimation or testing.

A more satisfactory solution is to impose the joint hypothesis $H^*_0: \pi = \delta = 0$, i.e. to compare (13) with the model

$$\Delta y_t = c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t. \quad (14)$$

The joint hypothesis can be tested by a LR test,

$$LR(\pi = \delta = 0) = -2 \cdot (\log L_0 - \log L_A),$$

where $\log L_0$ and $\log L_A$ denote the log-likelihood values from the models in (14) and (13), respectively. Due to the presence of a unit root under the null hypothesis this also has a non-standard distribution; and the LR test statistic follows the square of the DF distribution, $DF^2_c$, under the null. Instead of separately tabulating critical values for the $DF^2_c$ distribution, we can use the signed LR test,

$$\hat{\omega}_c = \text{sign}(\hat{\pi}) \cdot \sqrt{LR(\pi = \delta = 0)}.$$ 

This statistic follows the $DF_c$ distribution reported in Table 3.

**Example 3 (unemployment):** To illustrate the use of the DF test consider in Figure 4 (B) the US unemployment rate, calculated as the number of unemployed in percentage of the labour force, 1986 : 1 – 2006 : 1. We denote the variable $UNR_t$. From an economic point of view many economists would object to a linear trend in the model, and we consider a regression with a constant term,

$$\Delta UNR_t = \delta + \pi \cdot UNR_{t-1} + \sum_{i=1}^{p-1} c_i \cdot \Delta UNR_{t-i} + \epsilon_t.$$

To satisfactorily model the time series we use an AR(5), which means that we allow for four lagged first differences in the Dickey-Fuller regression. The results are reported in Table 4.

The augmented Dickey-Fuller test is just the $t$-test $\hat{\tau}_{\pi = 0} = -1.94$. The asymptotic distribution is $DF_c$, with a 5% critical values of $-2.86$. Based on the DF $t$-test we therefore cannot reject the null hypothesis of a unit root, and we conclude that the time series for US unemployment is likely to be generated as a unit-root non-stationary process.

14
The likelihood ratio test for the joint hypothesis $LR(\pi = \delta = 0)$ can be obtained by performing the regression under the null,

$$\Delta \text{UNR}_t = -0.105 \Delta \text{UNR}_{t-1} + 0.168 \Delta \text{UNR}_{t-2} + 0.225 \Delta \text{UNR}_{t-3} + 0.114 \Delta \text{UNR}_{t-4} + \epsilon_t,$$

where the numbers in parentheses are $t$–values. The log-likelihood value from this regression is $\log L_0 = 159.06$, and the likelihood ratio test is given by

$$LR(\pi = \delta = 0) = -2 \cdot (\log L_0 - \log L_A) = -2 \cdot (159.06 - 161.37) = 4.62.$$

The signed LR statistic is therefore given by

$$\hat{\omega}_c = -\sqrt{LR(\pi = \delta = 0)} = -\sqrt{4.62} = -2.15,$$

which is not significant compared to the critical values of $-2.86$.

The conclusion that unemployment has a unit root has important consequences from an economic point of view. It implies that shocks to the labour market have permanent effects. An explanation could be that if people get unemployed, then they gradually lose their ability to work, and it is very difficult to be reemployed. This hypothesis is known as *hysteresis* in the labour market.

### 4.4 Dickey-Fuller Test with a Trend Term

If the variable in the analysis is trending, the relevant alternative is in many cases trend-stationarity. The test is in this case based on the regression model with a trend,

$$\Delta y_t = \delta + \gamma t + \pi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t.$$  \hspace{1cm} (15)
The hypothesis of a unit root is still $H_0 : \pi = 0$, and the DF $t$–test is again just the $t$–ratio
\[ \hat{\tau}_t = \frac{\hat{\pi}}{se(\hat{\pi})}. \]

The presence of a trend shifts the asymptotic distribution, $DF_t$, further to the left as illustrated in Figure 4 (A). By looking at the distribution we note that even if the true value of $\pi$ is zero, $\pi = 0$, the estimate, $\hat{\pi}$, is always negative. This reflects the large bias when the autoregressive parameter is close to one. From Table 3 we see that the 5% critical values is now $-3.41$.

It holds again that if $\pi = 0$ then $\gamma t$ is accumulated to produce a quadratic trend in the model for $y_t$. To avoid this we may consider the joint hypothesis, $H^*_0 : \pi = \gamma = 0$, i.e. to compare (15) with the model under the null:
\[ \Delta y_t = \delta + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t. \]
In the model under the null it still holds that $\delta$ is accumulated to a linear trend, which exactly matches the deterministic specification under the alternative (15). The test is a comparison between a trend-stationarity model under the alternative and a random walk with drift under the null. It seems reasonable to allow the same deterministic components under the null and under the alternative, and the joint hypothesis, $H^*_0$, is in most cases preferable in empirical applications. The joint hypothesis can be tested by a LR test
\[ LR(\pi = \gamma = 0) = -2 \cdot (\log L_0 - \log L_A), \]
where $\log L_0$ and $\log L_A$ again denote the log-likelihood values from the two relevant models. The asymptotic distribution of $LR(\pi = \gamma = 0)$ is $DF_t^2$, but we can again consider the signed square root
\[ \hat{\omega}_t = \text{sign}(\hat{\pi}) \cdot \sqrt{LR(\pi = \gamma = 0)}, \]
which follows the same $DF_t$ distribution as the $t$–test.

**Example 4 (productivity):** To formally test whether the Danish productivity in Example 1 is trend-stationary, we want to test for unit root in the model in Table 1. One possibility is to use the $t$–test
\[ \hat{\tau}_t = \frac{\hat{\theta} - 1}{se(\hat{\theta})} = \frac{0.561273 - 1}{0.07056} = -6.22, \]
from Table 1. This is much smaller than the critical value, and we clearly reject the null hypothesis of unit root, thus concluding that productivity appears to be a trend-stationary process. We can note that exactly the same result would have been obtained if we considered the transformed regression
\[ \Delta \text{LPROD}_t = 0.091 + 0.0024 t - 0.439 \text{LPROD}_{t-1} + \epsilon_t, \]
\[ \text{Table 1.} \]
where we recognize the \( t \)-ratio in parenthesis.

To test the joint hypothesis, \( H_0^* : \pi = \gamma = 0 \), we may run the regression under the null

\[
\Delta \text{LPROD}_t = 0.0057 + \epsilon_t,
\]

which gives a log-likelihood value of 348.63. The likelihood ratio test is given by

\[
LR(\pi = \gamma = 0) = -2 \cdot (\log L_0 - \log L_A) = -2 \cdot (348.63 - 366.09) = 34.92,
\]

and the signed LR test is

\[
\hat{\omega}_t = -\sqrt{LR(\pi = \gamma = 0)} = -\sqrt{34.92} = -5.91,
\]

and we clearly reject the unit root hypothesis.

**Example 5 (Consumption):** We also test the hypothesis that private consumption in Example 2 is trend-stationary. Since the regression in that case in an AR(2), there is no way to derive the test statistic from the output of Table 2 alone. Instead we run the equivalent regression in first differences

\[
\Delta \text{LCONS}_t = 0.764 + 0.0004t - 0.129 \text{ LCONS}_{t-1} - 0.209 \Delta \text{LCONS}_{t-1} + \epsilon_t,
\]

which produces the same likelihood as in Table 2, \( \log L_A = 359.23 \). The Dickey-Fuller \( t \)-test is given by \( \hat{\tau}_t = -2.56 \), which is not significantly in the DF\( _t \) distribution. We conclude that private consumption seems to behave as a unit-root non-stationary process.

To test the joint hypothesis, \( H_0^* : \pi = \gamma = 0 \), we use the regression under the null,

\[
\Delta \text{LCONS}_t = 0.0046 - 0.274 \Delta \text{LCONS}_{t-1} + \epsilon_t,
\]

with a log-likelihood value of 355.869. For the consumption series, the likelihood ratio test for a unit root is given by

\[
LR(\pi = \gamma = 0) = -2 \cdot (\log L_0 - \log L_A) = -2 \cdot (355.869 - 359.23) = 6.722.
\]

The signed LR test is \( \hat{\omega}_t = -\sqrt{6.722} = -2.59 \), and we conclude in favour of a unit root.

### 5 Further Issues in Unit Root Testing

This section contains some concluding remarks on unit root testing.

#### 5.1 The Problem of Low Power

The decision on the presence of unit roots or not is important. From an economic point of vies it is important to know whether shocks have permanent effects or not, and from a statistical point of view it is important to choose the appropriate statistical tools. It
should be noted, however, that the decision is difficult in real life situations. We often say that the unit root test has low power to distinguish a unit root from a large (but stationary) autoregressive root.

To illustrate this, consider two time series, generated as

\[
\Delta y_t = -0.2 \cdot y_t + 0.05 \cdot t + \epsilon_t
\]
\[
\Delta x_t = 0.25 + \epsilon_t.
\]

Process \(y_t\) is trend-stationary, while \(x_t\) is a random walk with drift. Figure 5 (A) depicts \(T = 100\) observations from one realization of the processes. We note that the two series are very alike, and from a visual inspection it is impossible to distinguish the unit root from the trend-stationary process. This illustrates that in small samples, a trend-stationary process can be approximated by a random walk with drift, and vice versa. That makes unit root testing extremely difficult! Figure 5 (B) shows the same series, but now extended to \(T = 500\) observations. For the long sample the difference is clear: \(y_t\) has an attractor, \(x_t\) does not.

Using the equation (17) as a Monte Carlo DGP, we can illustrate the power of the unit root hypothesis, \(H_0 : \pi = 0\), in the model

\[
\Delta y_t = \delta + \gamma t + \pi y_{t-1} + \epsilon_t,
\]

i.e. how often \(\pi = 0\) is rejected given that the true value is \(\pi = -0.2\). Similarly we can use equation (18) as a DGP to illustrate the size of the test, i.e. how often \(\pi = 0\) is rejected if it is true in the DGP. Figure 5 (C) depict the size and power of \(H_0\) as a function of the number of observations. All tests are performed at a 5% level, so we expect the size to converge to 5% and the power to converge to 100. We note that the actual size is too large in small sample, so that we reject a true hypothesis too often. As the number of observations increases, the actual size converges to 5%. The power is increasing relatively slowly in the number of observations. To reject the false unit root hypothesis 50% of the times, we need close to 100 observations. In small samples, e.g. \(T = 50\), it is extremely difficult to tell the two processes, (17) and (18), apart.

For comparison graph (D) shows the corresponding results for the LR test of the joint hypothesis, \(H_0^* : \pi = \gamma = 0\). We note that the size properties of the combined test are clearly preferable, and the true hypothesis is rejected 5% of the times for all sample lengths. Based on this we would this test in practice, although the power is still low.

5.2 Importance of Special Events

The core of a unit root test is to assess whether shocks have transitory or permanent effects, and that conclusion is very sensitive to a few large shocks.

As discussed above, a stationary time series with a level shift is no-longer stationary. A unit root test applied to a time series like the one in Figure 1 (B) is therefore likely not to reject the unit root. From an intuitive point of view the process in graph (B) can be
approximated by a random walk, while a stationary autoregression will never be able to change its level to track the observed time series. That will bias the conclusion towards the finding of unit roots. If an observed time series has a level shift, the only solution is to model it using dummy variables, and to test for a unit root in a model like (2). This is complicated, however, because the distribution of the unit root test depends on the presence of the dummy in the regression model, and new critical values have to be used.

If the time series have many large isolated outliers, the effects on the unit root test will be the opposite. Large outliers make the time series look more stable than it actually is and that will bias the test towards stationarity. The solution is again to model the outliers with dummy variables, but the issue is complicated because the dummies accumulate under the null hypothesis.

5.3 Further Readings

The literature on unit root testing is huge, and most references are far more technical than the present note. Textbooks on time series econometrics with sections on unit root testing include Patterson (2000), Enders (2004), Banerjee, Dolado, Gailbraith, and Hendry (1993), and Hamilton (1994), where the latter is rather technical. The review of the unit root literature in Maddala and Kim (1998) is a good starting point for further reading.
References


