NON-STATIONARY TIME SERIES AND UNIT ROOT TESTING

ECONOMETRICS C ♦ LECTURE NOTE 5

HEINO BOHN NIELSEN

FEBRUARY 6, 2012

T his 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.

OUTLINE

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

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \quad t = 1, 2, ..., T, \tag{1}$$

can be analyzed using standard tools, and most of the results for regression of independent and identical (IID) observations hold. The technical argument is that there exists 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 econometric 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 §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 §3 we review the properties of a stationary autoregressive process and discuss the implications of a unit root. We then proceed in §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

This section discusses some examples of non-stationarity typically observed 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
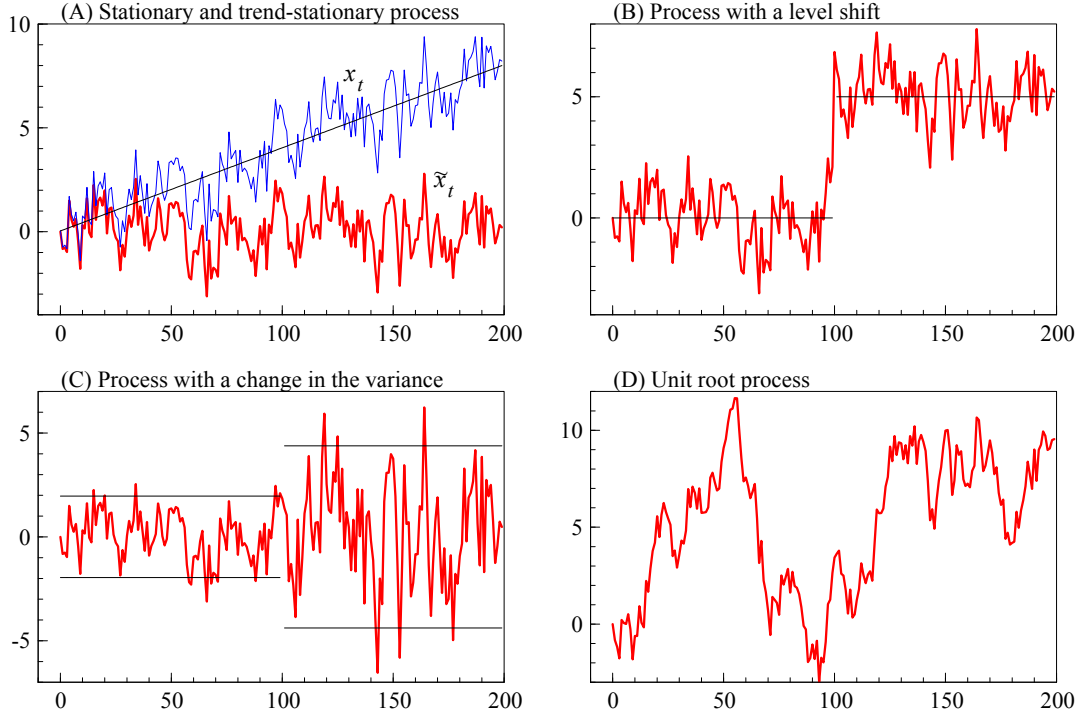
Figure 1: *Simulated examples of non-stationary time series.*

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 consider the stationary AR(1) process with a zero mean,

$$\widetilde{x}_t = \theta \widetilde{x}_{t-1} + \epsilon_t, \quad t = 1, 2, ..., T,$$

with $\mid \theta \mid < 1$ and $\widetilde{x}_0 = 0$, and a new process, x_t, defined as the stationary process plus a linear trend term and a constant,

$$x_t = \widetilde{x}_t + \mu_0 + \mu_1 t. \tag{2}$$

Since $\widetilde{x}_t$ is a stationary process, x_t is stationary around the trending mean, $E[x_t] = \mu_0 + \mu_1 t$, i.e. it is trend-stationary. A single realization of $T = 200$ observations of the processes $\widetilde{x}_t$ and x_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 using the regression

3

(1) we could therefore de-trend the variables by running the two OLS regressions:

$$y_t = \varphi_0 + \varphi_1 t + \text{residual} \quad \text{and} \quad x_t = \mu_0 + \mu_1 t + \text{residual}.$$

We then define the de-trended variables as the residuals, $\widetilde{y}_t = y_t - \widehat{\varphi}_0 - \widehat{\varphi}_1 t$ and $\widetilde{x}_t = x_t - \widehat{\mu}_0 - \widehat{\mu}_1 t$, and consider the (stationary) linear regression

$$\widetilde{y}_t = \beta_0 + \beta_1 \widetilde{x}_t + \epsilon_t.$$

Alternatively, we could consider a regression augmented with a linear trend term, i.e.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 t + \epsilon_t,$$

and we note that the two approaches give identical results, cf. the celebrated *Frisch-Waugh-Lovell* theorem.

A leading example of 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,

$$\mathsf{LPROD}_t = \theta \cdot \mathsf{LPROD}_{t-1} + \delta + \gamma t + \epsilon_t,$$

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. ♦

	Coefficient	Std. Error	t−value
LPROD_{t-1}	0.561273	0.07056	7.95
Constant	0.090890	0.01382	6.58
t	0.002412	0.00039	6.15
$\widehat{\sigma}$	0.016907	log-likelihood	366.09
R^2	0.994	T	137
	Statistic	[p−val]	Distribution
No autocorrelation of order 1-2	3.80	[0.15]	$\chi^2(2)$
Normality	2.52	[0.28]	$\chi^2(2)$
No heteroskedasticity	6.53	[0.16]	$\chi^2(4)$
Correct functional form (RESET)	2.66	[0.11]	F(1,133)

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

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, where the estimates of the two inverse characteristic roots are given by $\hat{\phi}_1 = 0.896$ and $\hat{\phi}_2 = -0.233$. In §4 we will formally test whether the deviations are stationary, i.e. whether LCONS_t is a trend-stationary process. ◆

	Coefficient	Std.Error	t−value
LCONS_{t-1}	0.662660	0.08548	7.75
LCONS_{t-2}	0.208583	0.08578	2.43
Constant	0.763562	0.29720	2.57
t	0.000436	0.00017	2.58
$\widehat{\sigma}$	0.017502	log-likelihood	359.23
R^2	0.983	T	136
	Statistic	[p−val]	Distribution
No autocorrelation of order 1-2	3.57	[0.17]	$\chi^2(2)$
Normality	52.44	[0.00]	$\chi^2(2)$
No heteroskedasticity	11.80	[0.07]	$\chi^2(6)$
Correct functional form (RESET)	2.19	[0.14]	F(1,131)

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

2.2   LEVEL SHIFTS AND STRUCTURAL BREAKS

Another type of non-stationarity in a time series is if there is a change is the unconditional mean at a given point in time. As an example the mean of a time series could be μ_1 for the first half of the sample and μ_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 regression 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 (1) to take account of a level shift in y_t or x_t:

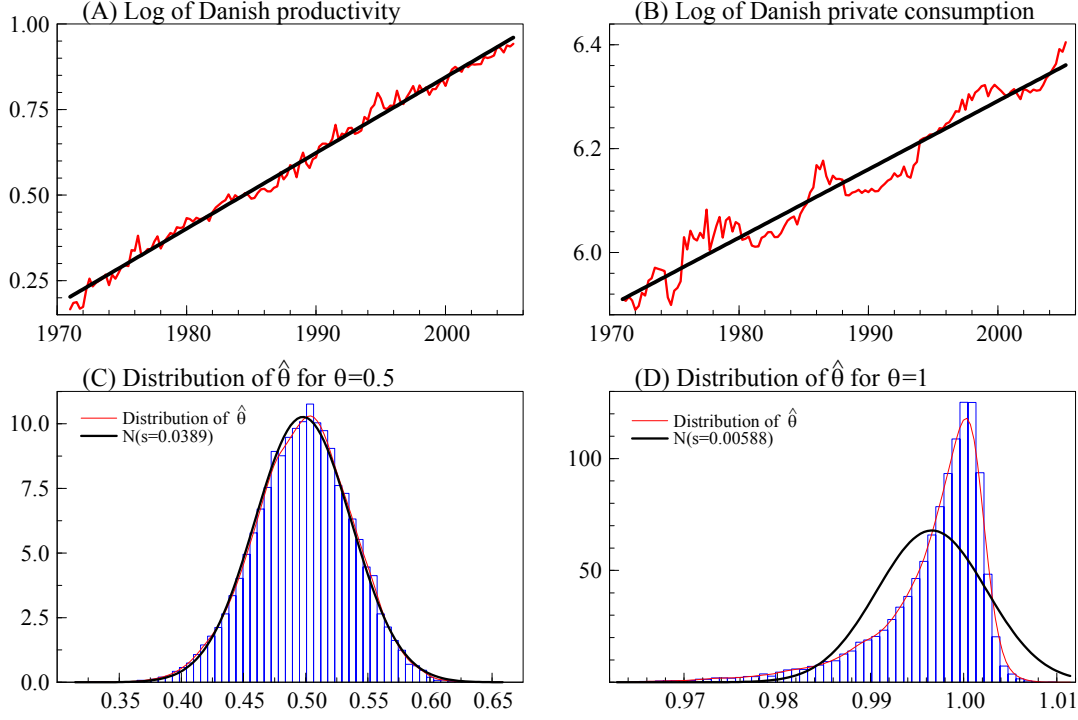$$y_t = \beta_0 + \beta_1 x_t + \beta_2 D_t + \epsilon_t. \tag{3}$$

Figure 2: *(A)-(B): Examples of non-stationary time series. (C)-(D): Distribution of the OLS estimator, $\widehat{\theta}$, in an AR(1) model when the true parameter is $\theta = 0.5$ and $\theta = 1$. Simulated with $T = 500$ observations and 20000 replications.*

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-stationarity 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.

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, $\widehat{\theta}_m$. 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 $\widehat{\theta}$, 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 $\mathsf{MCSD}(\widehat{\theta}) \approx 0.04$. Figure (D) depicts the distribution of $\widehat{\theta}$ 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 $\widehat{\theta}$ 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 $\mathsf{MCSD}(\widehat{\theta}) \approx 0.006$. This reflects that the estimator is consistent, i.e. $p \lim(\widehat{\theta}) = \theta$, 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

$$y_t = \theta y_{t-1} + \epsilon_t, \tag{4}$$

for $t = 1, 2, ..., T$, where ϵ_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 (4) is stationary if $|\theta| < 1$.

The solution to (4) in terms of the initial value and the error terms can be found from recursive substitution, i.e.

$$
\begin{aligned}
y_t &= \theta y_{t-1} + \epsilon_t \\
&= \theta(\theta y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \epsilon_t + \theta \epsilon_{t-1} + \theta^2 y_{t-2} \\
&= \epsilon_t + \theta \epsilon_{t-1} + \theta^2(\theta y_{t-3} + \epsilon_{t-2}) \\
&= \epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + \theta^3 y_{t-3} \\
&\vdots \\
&= \epsilon_t + \theta \epsilon_{t-1} + \theta^2 \epsilon_{t-2} + ... + \theta^{t-1} \epsilon_1 + \theta^t y_0.
\end{aligned}
\tag{5}
$$

An important characteristic of this process is that a shock to ϵ_t has only transitory effects because θ^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 (5) we can find the properties of y_t directly. The mean is

$$
E[y_t \mid y_0] = \theta^t y_0 \rightarrow 0.
$$

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

$$
V[y_t \mid y_0] = \sigma^2 + \theta^2 \sigma^2 + \theta^4 \sigma^2 + ... + \theta^{2(t-1)} \sigma^2 \rightarrow \frac{\sigma^2}{1 - \theta^2},
$$

and remember that the autocorrelation function is given by

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

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 (4) is unity, $\theta = 1$, i.e.

$$
y_t = y_{t-1} + \epsilon_t.
\tag{6}
$$

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 (5) with $\theta = 1$, i.e.

$$y_t = y_0 + \Delta y_1 + \Delta y_2 + ... + \Delta y_t = y_0 + \epsilon_1 + \epsilon_2 + ... + \epsilon_t = y_0 + \sum_{i=1}^{t} \epsilon_i. \qquad (7)$$

Note the striking differences between (5) and (7). 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

$$E[y_t \mid y_0] = y_0,$$

and the initial plays the role of a constant term. Secondly, the shocks to the process, ϵ_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 variance now increases with t,

$$V[y_t \mid y_0] = V\left[\sum_{i=1}^{t} \epsilon_i \mid y_0\right] = t\sigma^2,$$

and the process is clearly non-stationary. The first-differenced 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

$$
\begin{aligned}
Cov(y_t, y_{t-s} \mid y_0) &= E[(y_t - y_0)(y_{t-s} - y_0) \mid y_0] \\
&= E[(\epsilon_1 + \epsilon_2 + ... + \epsilon_t)(\epsilon_1 + \epsilon_2 + ... + \epsilon_{t-s}) \mid y_0] \\
&= (t - s)\sigma^2,
\end{aligned}
$$

and the autocorrelation is given by

$$Corr(y_t, y_{t-s} \mid y_0) = \frac{Cov(y_t, y_{t-s} \mid y_0)}{\sqrt{V[y_t \mid y_0] \cdot V[y_{t-s} \mid y_0]}} = \frac{(t-s)\sigma^2}{\sqrt{t\sigma^2 \cdot (t-s)\sigma^2}} = \frac{t-s}{\sqrt{t(t-s)}} = \sqrt{\frac{t-s}{t}},$$

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 (4) 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 (7), 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.
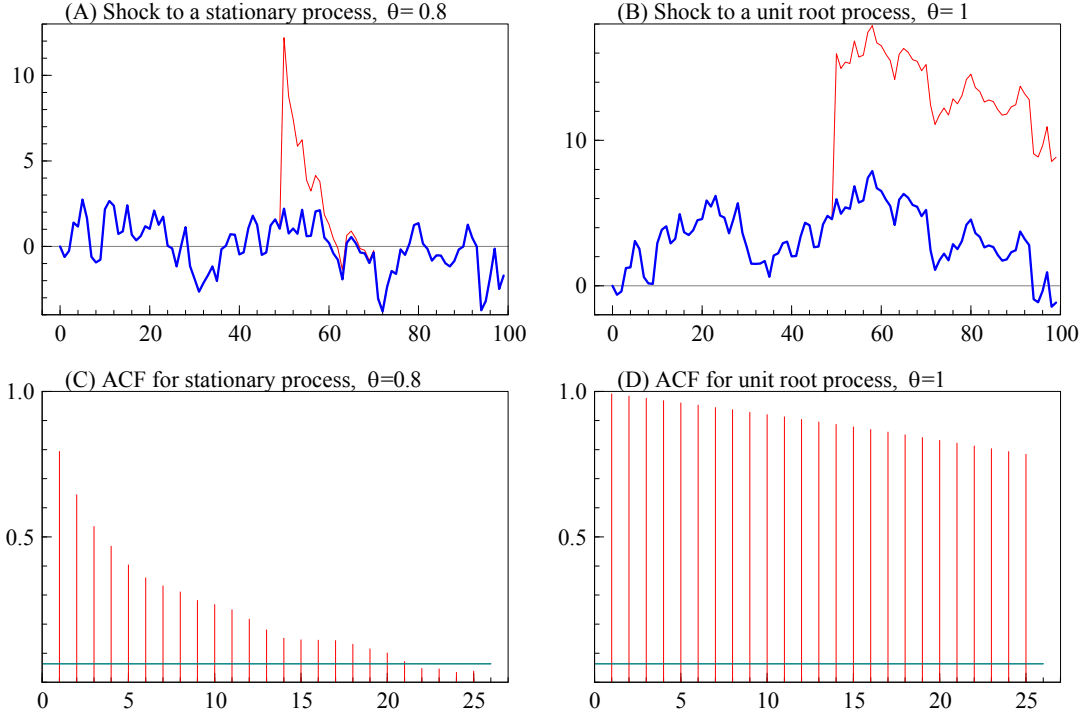
Figure 3: *Differences between stationary and non-stationary time series. (A) and (B) show one realization of a stationary and non-stationary time series respectively and illustrates the temporary and permanent impact of the shocks. (C) and (D) show the estimated autocorrelation function.*

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

$$y_t = \theta^t y_0 + \sum_{i=0}^{t-1} \theta^i \epsilon_{t-i} + (1 + \theta + \theta^2 + ... + \theta^{t-1})\delta, \tag{8}$$

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

$$y_t = y_0 + \sum_{i=1}^{t}(\delta + \epsilon_i) = y_0 + \delta t + \sum_{i=1}^{t} \epsilon_i, \tag{9}$$

where the constant term is accumulated to a deterministic linear trend, δt, while the initial value, y_0, plays the role of a constant term. The process in (9) 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,

$$y_t = y_0 + \sum_{i=1}^{t} \epsilon_i. \tag{10}$$

10

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.

4 TESTING FOR A UNIT ROOT

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 hypothesis, H_0 and H_A say, 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

$$y_t = \theta y_{t-1} + \epsilon_t, \tag{11}$$

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, i.e. a one-sided hypothesis. 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},

$$\Delta y_t = \pi y_{t-1} + \epsilon_t, \tag{12}$$

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

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

11

Distribution	Quantile			
	1%	2.5%	5%	10%
$N(0,1)$	-2.33	-1.96	-1.64	-1.28
DF	-2.56	-2.23	-1.94	-1.62
DF_c	-3.43	-3.12	-2.86	-2.57
DF_l	-3.96	-3.66	-3.41	-3.13

Table 3: *Asymptotic critical values for the Dickey-Fuller unit root test. This is the one-sided test for $\pi = 0$. Reproduced from Davidson and MacKinnon (1993).*

The *Dickey-Fuller* (DF) test statistic is simply the $t-$ratio of H_0 in (11) or (12), i.e.

$$\widehat{\tau} = \frac{\widehat{\theta} - 1}{\mathsf{se}(\widehat{\theta})} = \frac{\widehat{\pi}}{\mathsf{se}(\widehat{\pi})},$$

where $\mathsf{se}(\cdot)$ denotes the estimated standard error of the coefficient. The asymptotic distribution of the test under the null hypothesis of a unit root is not standard normal, cf. also the non-standard distribution of $\widehat{\theta}$ for $\theta = 1$ illustrated in Figure 2 (D). It follows a so-called Dickey-Fuller distribution, 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 DF-distribution is -1.94, which is smaller than the corresponding -1.64 from $N(0,1)$.

It is worth noting that the DF distribution is derived under the assumption that ϵ_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

$$
\begin{aligned}
y_t - y_{t-1} &= (\theta_1 - 1)y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \epsilon_t \\
y_t - y_{t-1} &= (\theta_1 - 1)y_{t-1} + (\theta_2 + \theta_3)y_{t-2} + \theta_3(y_{t-3} - y_{t-2}) + \epsilon_t \\
y_t - y_{t-1} &= (\theta_1 + \theta_2 + \theta_3 - 1)y_{t-1} + (\theta_2 + \theta_3)(y_{t-2} - y_{t-1}) + \theta_3(y_{t-3} - y_{t-2}) + \epsilon_t \\
\Delta y_t &= \pi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t, \qquad (13)
\end{aligned}
$$

where $\pi = \theta_1 + \theta_2 + \theta_3 - 1 = -\theta(1)$, $c_1 = -(\theta_2 + \theta_3)$, and $c_2 = -\theta_3$. In equation (13) the hypothesis $\theta(1) = 0$ corresponds to

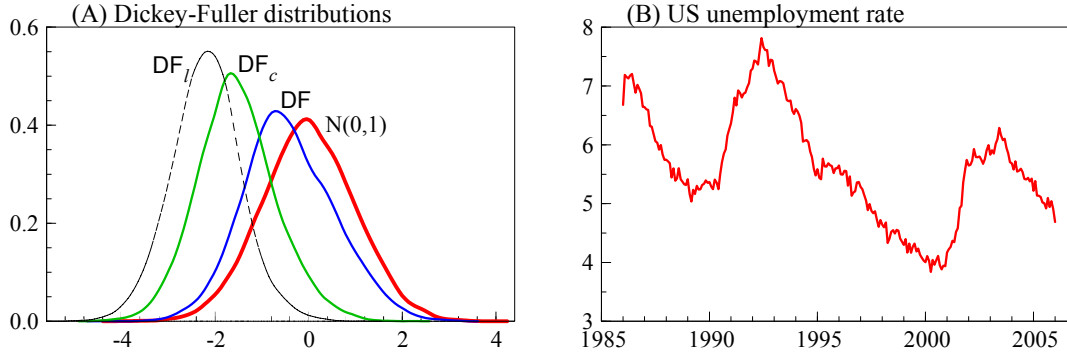$$H_0 : \pi = 0 \quad \text{against} \quad H_A : -2 < \pi < 0.$$

Figure 4: *(A) Dickey-Fuller distributions. (B) US unemployment rate.*

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).

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 (13), 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, $p_{\max}$, 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

$$\Delta y_t = \delta + \pi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t. \tag{14}$$

Distribution	Quantile			
	1%	2.5%	5%	10%
$\chi^2(2)$	9.21	7.38	5.99	4.61
DF_c^2	12.73	10.73	9.13	7.50
DF_l^2	16.39	14.13	12.39	10.56

Table 4: *Asymptotic critical values for the Likelihood Ratio test for $\pi = 0$ and $\delta = 0$.*

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

$$\widehat{\tau}_c = \frac{\widehat{\pi}}{\mathsf{se}(\widehat{\pi})}.$$

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 (8) and (9), 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 (14) with the model

$$\Delta y_t = c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t. \tag{15}$$

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 (15) and (14), respectively. Due to the presence of a unit root under the null hypothesis this also has a non-standard distribution; referred to as DF_c^2, under the null. Critical values for the LR test is given in Table 4. Note that the test of the joint hypothesis is two-sided and rejects for large values of LR.

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, $1985 : 6 - 2005 : 6$. 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,

$$\Delta \text{UNR}_t = \delta + \pi \cdot \text{UNR}_{t-1} + \sum_{i=1}^{p-1} c_i \cdot \Delta \text{UNR}_{t-i} + \epsilon_t.$$

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 5.

The augmented Dickey-Fuller test is just the $t-$test $\widehat{\tau}_{\pi=0} = -1.94$. 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.

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

$$\Delta \text{UNR}_t = \underset{(-1.60)}{-0.105} \Delta \text{UNR}_{t-1} + \underset{(2.60)}{0.168} \Delta \text{UNR}_{t-2} + \underset{(3.49)}{0.225} \Delta \text{UNR}_{t-3} + \underset{(1.79)}{0.114} \Delta \text{UNR}_{t-4} + \epsilon_t,$$

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

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

which is much smaller than the critical value of 9.13.

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 loose 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,

$$\Delta y_t = \delta + \gamma t + \pi y_{t-1} + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t. \tag{16}$$

The hypothesis of a unit root is still $H_0 : \pi = 0$, and the DF $t-$test is again just the $t-$ratio

$$\widehat{\tau}_l = \frac{\widehat{\pi}}{\text{se}(\widehat{\pi})}.$$

The presence of a trend shifts the asymptotic distribution, DF_l, further to the left as illustrated in Figure 4 (A). By looking at the distribution we note that even if the true value of π is zero, $\pi = 0$, the estimate, $\widehat{\pi}$, is always negative. This reflect the large bias

	Coefficient	Std.Error	t−value
UNR_{t-1}	−0.016803	0.00868	−1.94
ΔUNR_{t-1}	−0.106844	0.06537	−1.63
ΔUNR_{t-2}	0.165137	0.06422	2.57
ΔUNR_{t-3}	0.228119	0.06451	3.54
ΔUNR_{t-4}	0.123870	0.06372	1.94
Constant	0.087189	0.04945	1.76
$\widehat{\sigma}$	0.123707	log-likelihood	161.37
R^2	0.102	T	236
	Statistic	[p−val]	Distribution
No autocorrelation of order 1-2	5.40	[0.07]	$\chi^2(2)$
Normality	0.79	[0.67]	$\chi^2(2)$
No heteroskedasticity	10.77	[0.38]	$\chi^2(4)$
Correct functional form (RESET)	1.36	[0.24]	F(1,229)

Table 5: *Modelling* ΔUNR_t *by OLS for* $t = 1985:11 - 2005:6$.

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 γt 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 (16) with the model under the null:

$$\Delta y_t = \delta + c_1 \Delta y_{t-1} + c_2 \Delta y_{t-2} + \epsilon_t. \qquad (17)$$

In the model under the null it still holds that δ is accumulated to a linear trend, which exactly match the deterministic specification under the alternative (16). 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_l^2, with critical values reported in Table 4.

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

$$\widehat{\tau}_l = \frac{\widehat{\theta} - 1}{\mathsf{se}(\widehat{\theta})} = \frac{0.561273 - 1}{0.07056} = -6.22,$$

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

$$\Delta\mathsf{LPROD}_t = \underset{(6.58)}{0.091} + \underset{(6.15)}{0.0024t} - \underset{(-6.22)}{0.439}\,\mathsf{LPROD}_{t-1} + \epsilon_t,$$

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

$$\Delta\mathsf{LPROD}_t = \underset{(3.48)}{0.0057} + \epsilon_t,$$

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

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

which strongly rejects compared to the critical value og 12.39.   ♦

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

$$\Delta\mathsf{LCONS}_t = \underset{(2.57)}{0.764} + \underset{(2.58)}{0.0004t} - \underset{(-2.56)}{0.129}\,\mathsf{LCONS}_{t-1} - \underset{(-2.43)}{0.209}\,\Delta\mathsf{LCONS}_{t-1} + \epsilon_t,$$

which produces the same likelihood as in Table 2, $\log L_A = 359.23$. The Dickey-Fuller $t-$test is given by $\widehat{\tau}_l = -2.56$, which is not significant in the DF_l distribution (but close to the critical value). 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,

$$\Delta\mathsf{LCONS}_t = \underset{(2.97)}{0.0046} - \underset{(-3.29)}{0.274}\,\Delta\mathsf{LCONS}_{t-1} + \epsilon_t,$$

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

$$LR(\pi = \gamma = 0) = -2 \cdot (\log L_0 - \log L_A) = -2 \cdot (355.869 - 359.23) = 6.722,$$

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 view 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 the problem, consider two time series, generated as

$$\Delta y_t = -0.2 \cdot y_{t-1} + 0.05 \cdot t + \epsilon_t \tag{18}$$

$$\Delta x_t = 0.25 + \epsilon_t. \tag{19}$$

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 (18) 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 (19) 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 10% level, so we expect the size to converge to 10% and the power to converge to 100% as T diverges. 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 10%. 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, (18) and (19), apart.

There is a large literature that tries to construct unit root tests with a higher power than the Dickey Fuller type test presented above, see Haldrup and Jansson (2006) for a survey. The most famous test is developed by Elliott, Rothenberg, and Stock (1996)
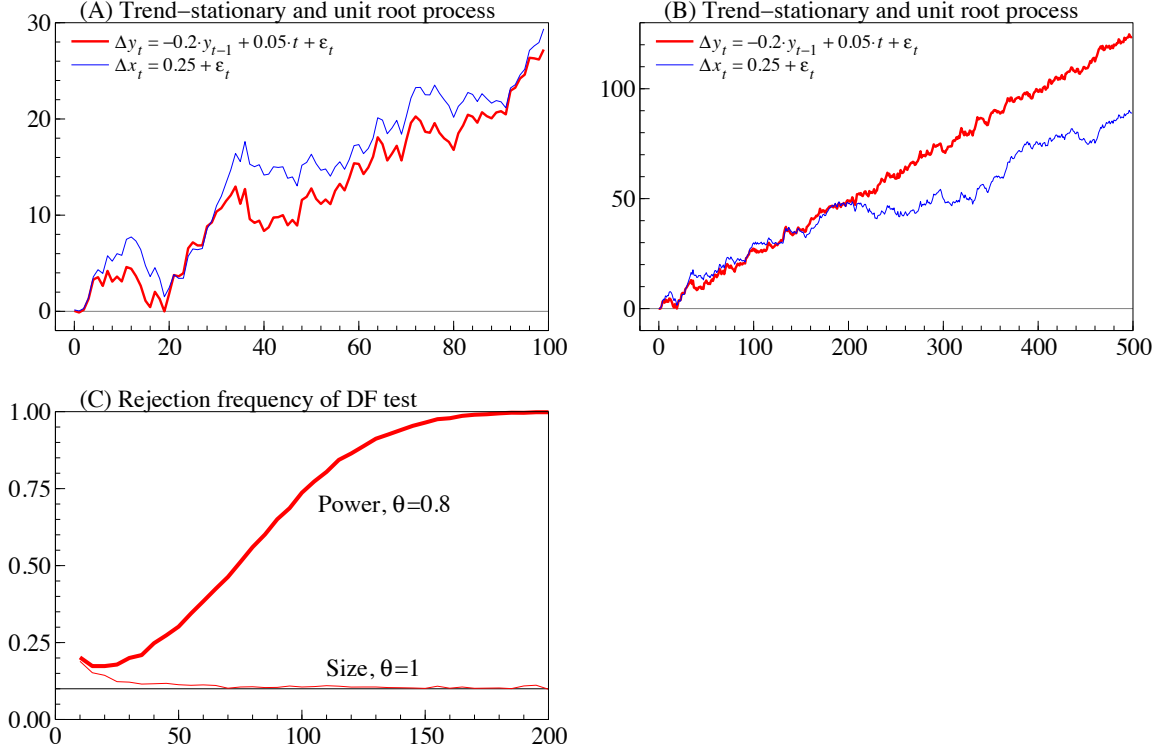
Figure 5: *Low power of unit root tests.*

(ERS), and this test is often used in applications. The idea is that the role of the initial value is fundamentally different for stationary and non-stationary variables and (correct) assumptions on the distribution of the initial value may improve the power to distinguish stationary and unit root processes. In practice the ERS test is based on an initial detrending of the variable, $\tilde{x}_t = x_t - \tilde{\gamma}_0 - \tilde{\gamma}_1 t$, where $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ are *not* OLS estimates, but more complicated estimates that take the initial value into account. The null hypothesis of a unit root is then testet on the detrended variable, $\tilde{x}_t$, using a Dickey Fuller type test. The implicit assumption on the initial value is not easy to test, however, and if it is not in accordance with the data, then the test may have lower power than the simple DF test conditional on the initial values, see Nielsen (2008).

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 (3). 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.

6   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

Banerjee, A., J. Dolado, J. W. Gailbraith, and D. Hendry (1993): *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford University Press, Oxford.

Elliott, G., T. Rothenberg, and J. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Enders, W. (2004): *Applied Econometric Time Series.* John Wiley & Sons, 2nd edn.

Haldrup, N., and M. Jansson (2006): "Improving Size and Power in Unit Root Testing," in *Palgrave Handbook of Econometrics, Vol. 1*, ed. by T. C. Mills, and K. Patterson, chapter 6, pp. 252–277. Palgrave, New York.

Hamilton, J. D. (1994): *Time Series Analysis.* Princeton University Press, Princeton.

Maddala, G. S., and I.-M. Kim (1998): *Unit Roots, Cointegration, and Structural Change.* Cambridge University Press, Cambridge.

Nielsen, B. (2008): "Power of Tests for Unit Roots in the Presence of a Linear Trend," *Oxford Bulletin of Economics and Statistics*, 70, 619–644.

Patterson, K. (2000): *An Introduction to Applied Econometrics. A Time Series Approach.* Palgrave MacMillan, New York.

VERBEEK, M. (2008): *A Guide to Modern Econometrics.* John Wiley & Sons, 3rd edn.