7 Integrated Processes

7.1 Definition, Properties and Interpretation

Up to now the discussion concentrated on stationary processes and in particular ARMA processes. According to the Wold decomposition theorem (see Theorem 3.1) every purely non-deterministic processes possesses the following representation:

$$X_t = \mu + \Psi(\mathrm{L})Z_t,$$

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Typically, we model X_t as an ARMA process so that $\Psi(\mathrm{L}) = \frac{\Theta(\mathrm{L})}{\Phi(\mathrm{L})}$. This representation implies:

- $\mathbb{E}X_t = \mu$,
- $\lim_{h\to\infty} \mathbb{P}_t X_{t+h} = \mu$.

The above property is often referred to as mean reverting because the process moves around a constant mean. Deviations from this mean are only temporary or transitory. Thus, the best long-run forecast is just the mean of the process.

This property is often violated by economic time series which typically show a tendency to growth. Classic examples are time series for GDP (see Fig. 1.3) or some stock market index (see Fig. 1.5). This trending property is not compatible with stationarity as the mean is no longer constant. In order to cope with this characteristic of economic time series, two very different alternatives have been proposed. The first one consists in letting the mean μ be a function of time $\mu(t)$. The most popular specification for $\mu(t)$ is a linear function, i.e. $\mu(t) = \alpha + \delta t$. In this case we get:

$$X_t = \underbrace{\alpha + \delta t}_{\text{linear trend}} + \Psi(\mathrm{L})Z_t$$

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_7

The process $\{X_t\}$ is then referred to as a *trend-stationary process*. In practice one also encounters quadratic polynomials of t or piecewise linear functions. For example, $\mu(t) = \alpha_1 + \delta_1 t$ for $t \leq t_0$ and $\mu(t) = \alpha_2 + \delta_2 t$ for $t > t_0$. In the following, we restrict ourself to linear trend functions.

The second alternative assumes that the time series becomes stationary after differentiation. The number of times one has to differentiate the process to achieve stationarity is called the order of integration. If d times differentiation is necessary, the process is called integrated of order d and is denoted by $X_t \sim \mathrm{I}(d)$. If the resulting time series, $\Delta^d X_t = (1 - \mathrm{L})^d X_t$, is an ARMA(p,q) process, the original process is called an ARIMA(p,d,q) process. Usually it is sufficient to differentiate the time series only once, i.e. $d = 1$. For expositional purposes we will stick to this case.

The formal definition of an I(1) process is given as follows.

Definition 7.1. *The stochastic process $\{X_t\}$ is called* integrated of order one *or* difference-stationary, *denoted as $X_t \sim \mathrm{I}(1)$, if and only if $\Delta X_t = X_t - X_{t-1}$ can be represented as*

$$\Delta X_t = (1 - \mathrm{L})X_t = \delta + \Psi(\mathrm{L})Z_t, \qquad \Psi(1) \neq 0,$$

with $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$.

The qualification $\Psi(1) \neq 0$ is necessary to avoid trivial and uninteresting cases. Suppose for the moment that $\Psi(1) = 0$, then it would be possible to write $\Psi(\mathrm{L})$ as $(1 - \mathrm{L})\widetilde{\Psi}(\mathrm{L})$ for some lag polynomial $\widetilde{\Psi}(\mathrm{L})$. This would, however, imply that the factor $1 - \mathrm{L}$ could be canceled in the above definition so that $\{X_t\}$ is already stationary and that the differentiation would be unnecessary. The assumption $\Psi(1) \neq 0$ thus excludes the case where a trend-stationary process could be regarded as an integrated process. For each trend-stationary process $X_t = \alpha + \delta t + \widetilde{\Psi}(\mathrm{L})Z_t$ we have $\Delta X_t = \delta + \Psi(\mathrm{L})Z_t$ with $\Psi(\mathrm{L}) = (1-\mathrm{L})\widetilde{\Psi}(\mathrm{L})$. This would violate the condition $\Psi(1) \neq 0$. Thus a trend-stationary process cannot be a difference-stationary process.

The condition $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ implies $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ and is therefore stronger than necessary for the Wold representation to hold. In particular, it implies the Beveridge-Nelson decomposition of integrated processes into a linear trend, a random walk, and a stationary component (see Sect. 7.1.4). The condition is automatically fulfilled for all ARMA processes because $\{\psi_j\}$ decays exponentially to zero.

Integrated processes with $d > 0$ are also called unit-root processes. This designation results from the fact that ARIMA processes with $d > 0$ can be viewed as ARMA processes, whereby the AR polynomial has a d-fold root of one.[1] An important prototype of an integrated process is the random walk with drift δ:

$$X_t = \delta + X_{t-1} + Z_t, \qquad Z_t \sim \mathrm{WN}(0, \sigma^2).$$

[1] Strictly speaking this does not conform to the definitions used in this book because our definition of ARMA processes assumes stationarity.

Trend-stationary and difference-stationary processes have quite different characteristics. In particular, they imply different behavior with respect to the long-run forecast, the variance of the forecast error, and the impulse response function. In the next section, we will explore these properties in detail.

7.1.1 Long-Run Forecast

The optimal forecast in the least-squares sense given the infinite past of a trend-stationary process is given by

$$\widetilde{\mathbb{P}}_t X_{t+h} = \alpha + \delta(t+h) + \psi_h Z_t + \psi_{h+1} Z_{t-1} + \ldots$$

Thus we have

$$\lim_{h\to\infty} \mathbb{E}\left(\widetilde{\mathbb{P}}_t X_{t+h} - \alpha - \delta(t+h)\right)^2 = \sigma^2 \lim_{h\to\infty} \sum_{j=0}^{\infty} \psi_{h+j}^2 = 0$$

because $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Thus the long-run forecast is given by the linear trend. Even if X_t deviates temporarily from the trend line, it is assumed to return to it. A trend-stationary process therefore behaves in the long-run like $\mu(t) = \alpha + \delta t$.

The forecast of the differentiated series is

$$\widetilde{\mathbb{P}}_t \Delta X_{t+h} = \delta + \psi_h Z_t + \psi_{h+1} Z_{t-1} + \psi_{h+2} Z_{t-2} + \ldots$$

The level of X_{t+h} is by definition

$$X_{t+h} = (X_{t+h} - X_{t+h-1}) + (X_{t+h-1} - X_{t+h-2}) + \ldots + (X_{t+1} - X_t) + X_t$$

so that

$$\begin{aligned}
\widetilde{\mathbb{P}}_t X_{t+h} &= \widetilde{\mathbb{P}}_t \Delta X_{t+h} + \widetilde{\mathbb{P}}_t \Delta X_{t+h-1} + \ldots + \widetilde{\mathbb{P}}_t \Delta X_{t+1} + X_t \\
&= \delta + \psi_h Z_t + \psi_{h+1} Z_{t-1} + \psi_{h+2} Z_{t-2} + \ldots \\
&\quad + \delta + \psi_{h-1} Z_t + \psi_h Z_{t-1} + \psi_{h+1} Z_{t-2} + \ldots \\
&\quad + \delta + \psi_{h-2} Z_t + \psi_{h-1} Z_{t-1} + \psi_h Z_{t-2} + \ldots \\
&\quad + \ldots + X_t \\
&= X_t + \delta h \\
&\quad + \left(\psi_h + \psi_{h-1} + \ldots + \psi_1\right) Z_t \\
&\quad + \left(\psi_{h+1} + \psi_h + \ldots + \psi_2\right) Z_{t-1} \\
&\quad \ldots
\end{aligned}$$

This shows that also for the integrated process the long-run forecast depends on a linear trend with slope δ. However, the intercept is no longer a fixed number, but given by X_t which is stochastic. With each new realization of X_t the intercept changes so that the trend line moves in parallel up and down. This issue can be well illustrated by the following two examples.

Example 1. Let $\{X_t\}$ be a random walk with drift δ. Then best forecast of X_{t+h}, $\mathbb{P}_t X_{t+h}$, is

$$\mathbb{P}_t X_{t+h} = \delta h + X_t.$$

The forecast thus increases at rate δ starting from the initial value of X_t. δ is therefore the slope of a linear trend. The intercept of this trend is stochastic and equal to X_t. Thus the trend line moves in parallel up or down depending on the realization of X_t.

Example 2. Let $\{X_t\}$ be an ARIMA(0,1,1) process given by $\Delta X_t = \delta + Z_t + \theta Z_{t-1}$ with $|\theta| < 1$. The best forecast of X_{t+h} is then given by

$$\mathbb{P}_t X_{t+h} = \delta h + X_t + \theta Z_t.$$

As before the intercept changes in a stochastic way, but in contrary to the previous example it is now given by $X_t + \theta Z_t$. If we consider the forecast given the infinite past, the invertibility of the process implies that Z_t can be expressed as a weighted sum of current and past realizations of ΔX_t (see Sects. 2.3 and 3.1).

7.1.2 Variance of Forecast Error

In the case of a trend-stationary process the forecast error is

$$X_{t+h} - \widetilde{\mathbb{P}}_t X_{t+h} = Z_{t+h} + \psi_1 Z_{t+h-1} + \ldots + \psi_{h-1} Z_{t+1}.$$

As the mean of the forecast error is zero, the variance is

$$\mathbb{E}\left(X_{t+h} - \widetilde{\mathbb{P}}_t X_{t+h}\right)^2 = \left(1 + \psi_1^2 + \psi_2^2 + \ldots + \psi_{h-1}^2\right)\sigma^2.$$

For h going to infinity this expression converges to $\sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty$. This is nothing but the unconditional variance of X_t. Thus the variance of the forecast error increases with the length of the forecasting horizon, but remains bounded.

For the integrated process the forecast error can be written as

$$X_{t+h} - \widetilde{\mathbb{P}}_t X_{t+h} = Z_{t+h} + (1 + \psi_1) Z_{t+h-1} + \ldots + (1 + \psi_1 + \psi_2 + \ldots + \psi_{h-1}) Z_{t+1}.$$

The forecast error variance therefore is

$$\mathbb{E}\left(X_{t+h}-\widetilde{\mathbb{P}}_t X_{t+h}\right)^2=\left[1+(1+\psi_1)^2+\ldots+(1+\psi_1+\ldots+\psi_{h-1})^2\right]\sigma^2.$$

This expression increases with the length of the forecast horizon h, but is no longer bounded. It increases linearly in h to infinity.[2] The precision of the forecast therefore not only decreases with the forecasting horizon h as in the case of the trend-stationary model, but converges to zero. In the example above of the ARIMA(0,1,1) process the forecasting error variance is

$$\mathbb{E}\left(X_{t+h}-\mathbb{P}_t X_{t+h}\right)^2=\left[1+(h-1)\left(1+\theta\right)^2\right]\sigma^2.$$

This expression clearly increases linearly with h.

7.1.3 Impulse Response Function

The impulse response function (dynamic multiplier) is an important analytical tool as it gives the response of the variable X_t to the underlying shocks. In the case of the trend-stationary process the impulse response function is

$$\frac{\partial \widetilde{\mathbb{P}}_t X_{t+h}}{\partial Z_t}=\psi_h \longrightarrow 0 \quad \text{for } h\to\infty.$$

The effect of a shock thus declines with time and dies out. Shocks have therefore only transitory or temporary effects. In the case of an ARMA process the effect even declines exponentially (see the considerations in Sect. 2.3).[3]

In the case of integrated processes the impulse response function for ΔX_t implies:

$$\frac{\partial \widetilde{\mathbb{P}}_t X_{t+h}}{\partial Z_t}=1+\psi_1+\psi_2+\ldots+\psi_h.$$

For h going to infinity, this expression converges $\sum_{j=0}^{\infty}\psi_j=\Psi(1)\neq 0$. This implies that a shock experienced in period t will have a long-run or permanent effect. This long-run effect is called *persistence*. If $\{\Delta X_t\}$ is an ARMA process then the persistence is given by the expression

[2]*Proof*: By assumption $\{\psi_j\}$ is absolutely summable so that $\Psi(1)$ converges. Moreover, as $\Psi(1)\neq 0$, there exists $\varepsilon>0$ and an integer m such that $\left|\sum_{j=0}^{h}\psi_j\right|>\varepsilon$ for all $h>m$. The squares are therefore bounded from below by $\varepsilon^2>0$ so that their infinite sum diverges to infinity.

[3]The use of the partial derivative is just for convenience. It does not mean that X_{t+h} is differentiated in the literal sense.

$$\Psi(1) = \frac{\Theta(1)}{\Phi(1)}.$$

Thus, for an ARIMA(0,1,1) the persistence is $\Psi(1) = \Theta(1) = 1 + \theta$. In the next section we will discuss some examples.

7.1.4 The Beveridge-Nelson Decomposition

The Beveridge-Nelson decomposition represents an important tool for the understanding of integrated processes.[4] It shows how an integrated time series of order one can be represented as the sum of a linear trend, a random walk, and a stationary series. It may therefore be used to extract the cyclical component (business cycle component) of a time series and can thus be viewed as an alternative to the HP-filter (see Sect. 6.5.2) or to more elaborated so-called structural time series models (see Sects. 17.1 and 17.4.2).

Assuming that $\{X_t\}$ is an integrated process of order one, there exists, according to Definition 7.1, a causal representation for $\{\Delta X_t\}$:

$$\Delta X_t = \delta + \Psi(L) Z_t \qquad \text{with } Z_t \sim \text{WN}\left(0, \sigma^2\right)$$

with the property $\Psi(1) \neq 0$ and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$. Before proceeding to the main theorem, we notice the following simple, but extremely useful polynomial decomposition of $\Psi(L)$:

$$\begin{aligned}
\Psi(L) - \Psi(1) &= 1 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \psi_4 L^4 + \ldots \\
&\quad - 1 - \psi_1 - \psi_2 - \psi_3 - \psi_4 - \ldots \\
&= \psi_1(L-1) + \psi_2(L^2-1) + \psi_3(L^3-1) + \psi_4(L^4-1) + \ldots \\
&= (L-1)\left[\psi_1 + \psi_2(L+1) + \psi_3(L^2+L+1) + \ldots\right] \\
&= (L-1)\big[(\psi_1 + \psi_2 + \psi_3 + \ldots) + (\psi_2 + \psi_3 + \psi_4 + \ldots)L \\
&\quad + (\psi_3 + \psi_4 + \psi_5 + \ldots)L^2 + \ldots\big].
\end{aligned}$$

We state this results in the following Lemma:

Lemma 7.1. *Let* $\Psi(L) = \sum_{j=0} \psi_j L^j$, *then*

$$\Psi(L) = \Psi(1) + (L-1)\widetilde{\Psi}(L)$$

where $\widetilde{\Psi}(L) = \sum_{j=0}^{\infty} \widetilde{\psi}_j L^j$ *with* $\tilde{\psi}_j = \sum_{i=j+1}^{\infty} \psi_i$.

[4] Neusser (2000) shows how a Beveridge-Nelson decomposition can also be derived for higher order integrated processes.

As $\{X_t\}$ is integrated and because $\Psi(1) \neq 0$ we can express X_t as follows:

$$\begin{aligned} X_t &= X_0 + \sum_{j=1}^{t} \Delta X_j \\ &= X_0 + \sum_{j=1}^{t} \left\{\delta + \left[\Psi(1) + (\mathrm{L}-1)\widetilde{\Psi}(\mathrm{L})\right] Z_j\right\} \\ &= X_0 + \delta t + \Psi(1)\sum_{j=1}^{t} Z_j + \sum_{j=1}^{t} (\mathrm{L}-1)\widetilde{\Psi}(\mathrm{L}) Z_j \\ &= \underbrace{X_0 + \delta t}_{\text{linear trend}} + \quad \Psi(1) \underbrace{\sum_{j=1}^{t} Z_j}_{\text{random walk}} + \underbrace{\widetilde{\Psi}(\mathrm{L})Z_0 - \widetilde{\Psi}(\mathrm{L})Z_t}_{\text{stationary component}}. \end{aligned}$$

This leads to the following theorem.

Theorem 7.1 (Beveridge-Nelson Decomposition)**.** *Every integrated process* $\{X_t\}$ *has a decomposition of the following form:*

$$X_t = \underbrace{X_0 + \delta t}_{\text{linear trend}} + \quad \Psi(1) \underbrace{\sum_{j=1}^{t} Z_j}_{\text{random walk}} + \underbrace{\widetilde{\Psi}(\mathrm{L})Z_0 - \widetilde{\Psi}(\mathrm{L})Z_t}_{\text{stationary component}}.$$

The above representation is referred to as the Beveridge-Nelson decomposition.

Proof. The only substantial issue is to show that $\widetilde{\Psi}(\mathrm{L})Z_0 - \widetilde{\Psi}(\mathrm{L})Z_t$ defines a stationary process. According to Theorem 6.4 it is sufficient to show that the coefficients of $\widetilde{\Psi}(\mathrm{L})$ are absolutely summable. We have that:

$$\sum_{j=0}^{\infty} |\tilde{\psi}_j| = \sum_{j=0}^{\infty} \left| \sum_{i=j+1}^{\infty} \psi_i \right| \leq \sum_{j=0}^{\infty} \sum_{i=j+1}^{\infty} |\psi_i| = \sum_{j=1}^{\infty} j|\psi_j| < \infty,$$

where the first inequality is a consequence of the triangular inequality and the second inequality follows from the Definition 7.1 of an integrated process. □

Shocks of a random walk component have a permanent effect. This effect is measured by the persistence $\Psi(1)$, the coefficient of the random walk component. In macroeconomics aggregate supply shocks are ascribed to have a long-run effect as they affect productivity. In contrast monetary or demand shocks are viewed to have temporary effects only. Thus the persistence $\Psi(1)$ can be interpreted as a measure for the importance of supply shocks (see Campbell and Mankiw (1987), Cochrane (1988) or Christiano and Eichenbaum (1990)). For a critical view from an econometric standpoint see Hauser et al. (1999). A more sophisticated multivariate approach to identify supply and demand shocks and to disentangle their relative importance is provided in Sect. 15.5.

In business cycle analysis it is often useful to decompose $\{X_t\}$ into a sum of a trend component μ_t and a cyclical component ε_t:

$$X_t = \mu_t + \varepsilon_t.$$

In the case of a difference-stationary series, the cyclical component can be identified with the stationary component in the Beveridge-Nelson decomposition and the trend component with the random walk plus the linear trend. Suppose that $\{\Delta X_t\}$ follows an ARMA process $\Phi(\mathrm{L})\Delta X_t = c + \Theta(\mathrm{L})Z_t$ then $\Delta\mu_t = \delta + \Psi(1)Z_t$ can be identified as the trend component. This means that the trend component can be recursively determined from the observations by applying the formula

$$\mu_t = \frac{\Phi(\mathrm{L})}{\Theta(\mathrm{L})}\Psi(1)X_t.$$

The cyclical component is then simply the residual: $\varepsilon_t = X_t - \mu_t$.

In the above decomposition both the permanent (trend) component as well as the stationary (cyclical) component are driven by the same shock Z_t. A more sophisticated model would, however, allow that the two components are driven by different shocks. This idea is exploited in the so-called structural time series analysis where the different components (trend, cycle, season, and irregular) are modeled as being driven by separated shocks. As only the series $\{X_t\}$ is observed, not its components, this approach leads to serious identification problems. See the discussion in Harvey (1989), Hannan and Deistler (1988), or Mills (2003). In Sects. 17.1 and 17.4.2 we will provide an overall framework to deal with these issues.

Examples

Let $\{\Delta X_t\}$ be a MA(q) process with $\Delta X_t = \delta + Z_t + \ldots + \theta_q Z_{t-q}$ then the persistence is given simply by the sum of the MA-coefficients: $\Psi(1) = 1 + \theta_1 + \ldots + \theta_q$. Depending on the value of these coefficients. The persistence can be smaller or greater than one.

If $\{\Delta X_t\}$ is an AR(1) process with $\Delta X_t = \delta + \phi\Delta X_{t-1} + Z_t$ and assuming $|\phi| < 1$ then we get: $\Delta X_t = \frac{\delta}{1-\phi} + \sum_{j=0}^{\infty}\phi^j Z_{t-j}$. The persistence is then given as $\Psi(1) = \sum_{j=0}^{\infty}\phi^j = \frac{1}{1-\phi}$. For positive values of ϕ, the persistence is greater than one. Thus, a shock of one is amplified to have an effect larger than one in the long-run.

If $\{\Delta X_t\}$ is assumed to be an ARMA(1,1) process with $\Delta X_t = \delta + \phi\Delta X_{t-1} + Z_t + \theta Z_{t-1}$ and $|\phi| < 1$ then $\Delta X_t = \frac{\delta}{1-\phi} + Z_t + (\phi + \theta)\sum_{j=0}^{\infty}\phi^j Z_{t-j-1}$. The persistence is therefore given by $\Psi(1) = 1 + (\phi + \theta)\sum_{j=0}^{\infty}\phi^j = \frac{1+\theta}{1-\phi}$.

The computation of the persistence for the model estimated for Swiss GDP in Sect. 5.6 is more complicated because a fourth order difference $1 - \mathrm{L}^4$ has been used instead of a first order one. As $1-\mathrm{L}^4 = (1-\mathrm{L})(1+\mathrm{L}+\mathrm{L}^2+\mathrm{L}^3)$, it is possible to extend the above computations also to this case. For this purpose we compute the persistence for $(1 + \mathrm{L} + \mathrm{L}^2 + \mathrm{L}^3) \ln \mathrm{BIP}_t$ in the usual way. The long-run effect on $\ln \mathrm{BIP}_t$ is therefore given by $\Psi(1)/4$ because $(1 + \mathrm{L} + \mathrm{L}^2 + \mathrm{L}^3) \ln \mathrm{BIP}_t$ is nothing but four times the moving-average of the last four values. For the AR(2) model we get a persistence of 1.42 whereas for the ARMA(1,3) model the persistence is 1.34. Both values are definitely above one so that the permanent effect of a one-percent shock to Swiss GDP is amplified to be larger than one in the long-run. Campbell and Mankiw (1987) and Cochrane (1988) report similar values for the US.

7.2 Properties of the OLS Estimator in the Case of Integrated Variables

The estimation and testing of coefficients of models involving integrated variables is not without complications and traps because the usual asymptotic theory may become invalid. The reason being that the asymptotic distributions are in general no longer normal so that the usual critical values for the test statistics are no longer valid. A general treatment of these issues is beyond this text, but can be found in Banerjee et al. (1993) and Stock (1994). We may, however, illustrate the kind of problems encountered by looking at the Gaussian AR(1) case[5]:

$$X_t = \phi X_{t-1} + Z_t, \qquad t = 1, , 2, \ldots,$$

where $Z_t \sim \mathrm{IID}\,\mathrm{N}(0, \sigma^2)$ and $X_0 = 0$. For observations on $X_1, X_2, \ldots, X_T$ the OLS-estimator of ϕ is given by the usual expression:

$$\hat{\phi}_T = \frac{\sum_{t=1}^T X_{t-1}X_t}{\sum_{t=1}^T X_{t-1}^2} = \phi + \frac{\sum_{t=1}^T X_{t-1}Z_t}{\sum_{t=1}^T X_{t-1}^2}.$$

For $|\phi| < 1$, the OLS estimator of ϕ, $\hat{\phi}_T$, converges in distribution to a normal random variable (see Chap. 5 and in particular Sect. 5.2):

$$\sqrt{T}\left(\hat{\phi}_T - \phi\right) \xrightarrow{d} \mathrm{N}\left(0, 1 - \phi^2\right).$$

The estimated density of the OLS estimator of ϕ for different values of ϕ is represented in Fig. 7.1. This figure was constructed using a Monte-Carlo simulation of the above model for a sample size of $T = 100$ using $10,000$ replications for

[5] We will treat more general cases in Sect. 7.5 and Chap. 16.

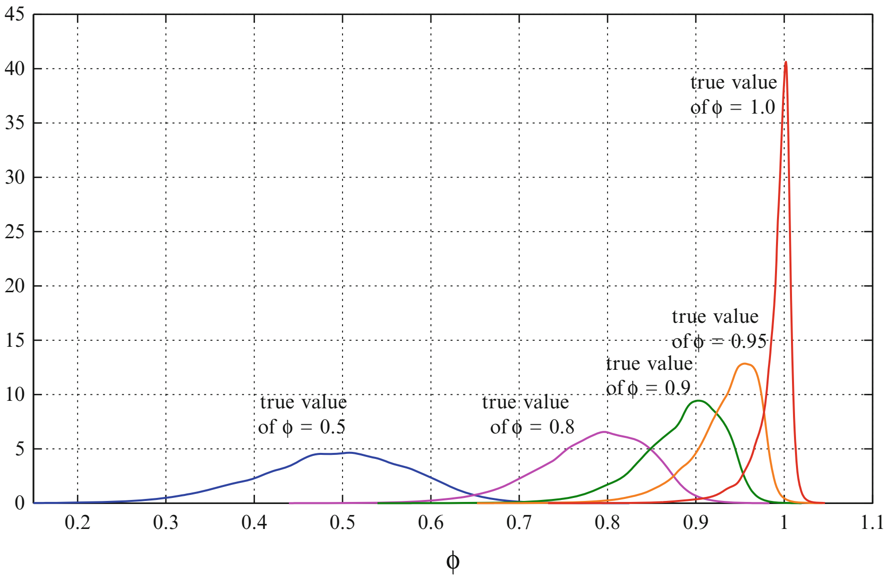

Fig. 7.1 Distribution of the OLS estimator of ϕ for $T = 100$ and 10,000 replications

each value of ϕ.[6] The figure shows that the distribution of $\hat{\phi}_T$ becomes more and more concentrated if the true value of ϕ gets closer and closer to one. Moreover the distribution gets also more and more skewed to the left. This implies that the OLS estimator is downward biased and that this bias gets relatively more and more pronounced in small samples as ϕ approaches one.

The asymptotic distribution would be degenerated for $\phi = 1$ because the variance approaches zero as ϕ goes to one. Thus the asymptotic distribution becomes useless for statistical inferences under this circumstance. In order to obtain a non-degenerate distribution the estimator must be scaled by T instead by $\sqrt{T}$. It can be shown that

$$T\left(\hat{\phi}_T - \phi\right) \xrightarrow{d} \nu.$$

This result was first established by Dickey and Fuller (1976) and Dickey and Fuller (1981). However, the asymptotic distribution ν need no longer be normal. It was first tabulated in Fuller (1976). The scaling with T instead of $\sqrt{T}$ means that the OLS-estimator converges, if the true value of ϕ equals one, at a higher rate to $\phi = 1$. This property is known as *superconsistency*.

[6]The densities were estimated using an adaptive kernel density estimator with Epanechnikov window (see Silverman (1986)).

In order to understand this result better, in particular in the light of the derivation in the Appendix of Sect. 5.2, we take a closer look at the asymptotic distribution of $T\left(\hat{\phi}_T - \phi\right)$:

$$T\left(\hat{\phi}_T - \phi\right) = \frac{\frac{1}{\sigma^2 T}\sum_{t=1}^T X_{t-1}Z_t}{\frac{1}{\sigma^2 T^2}\sum_{t=1}^T X_{t-1}^2}.$$

Under the assumption $\phi = 1$, X_t becomes a random walk so that X_t can be written as $X_t = Z_t + \ldots + Z_1$. Moreover, as a sum of normally distributed random variables X_t becomes itself normally distributed as $X_t \sim \mathrm{N}(0, \sigma^2 t)$. In addition, we get

$$\begin{aligned}
X_t^2 &= (X_{t-1} + Z_t)^2 = X_{t-1}^2 + 2X_{t-1}Z_t + Z_t^2 \\
&\Rightarrow X_{t-1}Z_t = \left(X_t^2 - X_{t-1}^2 - Z_t^2\right)/2 \\
&\Rightarrow \sum_{t=1}^T X_{t-1}Z_t = \frac{X_T^2 - X_0^2}{2} - \frac{\sum_{t=1}^T Z_t^2}{2} \\
&\Rightarrow \frac{1}{T}\sum_{t=1}^T X_{t-1}Z_t = \frac{1}{2}\left[\frac{X_T^2}{T} - \frac{\sum_{t=1}^T Z_t^2}{T}\right] \\
&\Rightarrow \frac{1}{\sigma^2 T}\sum_{t=1}^T X_{t-1}Z_t = \frac{1}{2}\left(\frac{X_T}{\sigma\sqrt{T}}\right)^2 - \frac{1}{2\sigma^2}\frac{\sum_{t=1}^T Z_t^2}{T} \xrightarrow{d} \frac{1}{2}\left(\chi_1^2 - 1\right).
\end{aligned}$$

The numerator therefore converges to a χ_1^2 distribution. The distribution of the denominator is more involved, but its expected value is given by:

$$\mathbb{E}\sum_{t=1}^T X_{t-1}^2 = \sigma^2 \sum_{t=1}^T (t-1) = \frac{\sigma^2 T(T-1)}{2},$$

because $X_{t-1} \sim \mathrm{N}\left(0, \sigma^2(t-1)\right)$. To obtain a nondegenerate random variable one must scale by T^2. Thus, intuitively, $T(\hat{\phi}_T - \phi)$ will no longer converge to a degenerate distribution.

Using similar arguments it can be shown that the t-statistic

$$t_T = \frac{\hat{\phi}_T - 1}{\hat{\sigma}_{\hat{\phi}}} = \frac{\hat{\phi}_T - 1}{\sqrt{\frac{s_T^2}{\sum_{t=1}^T X_{t-1}^2}}}$$

with $s_T^2 = \frac{1}{T-2}\sum_{t=2}^T\left(X_t - \hat{\phi}_T X_{t-1}\right)^2$ is not asymptotically normal. Its distribution was also first tabulated by Fuller (1976). Figure 7.2 compares its density with the standard normal distribution in a Monte-Carlo experiment using again a sample of

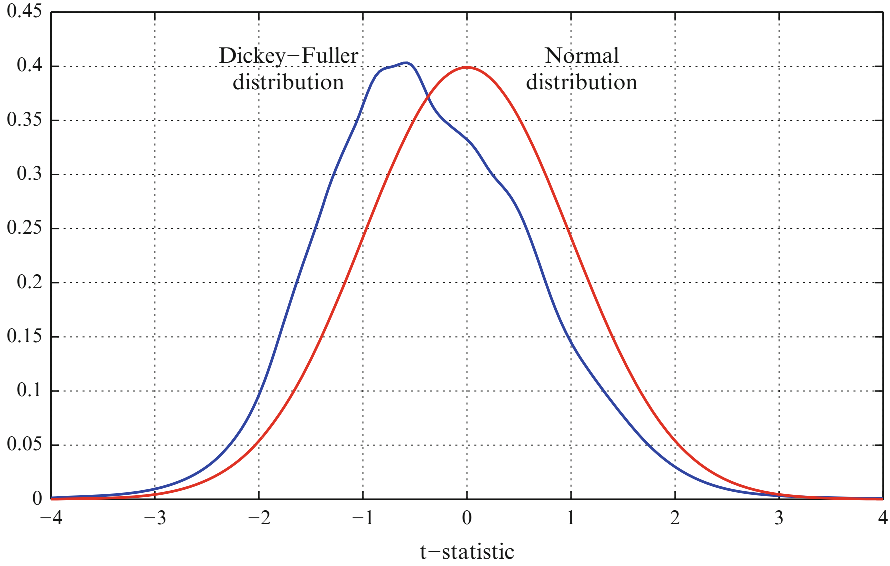

Fig. 7.2 Distribution of t-statistic for $T = 100$ and 10,000 replications and standard normal distribution

$T = 100$ and $10,000$ replications. It is obvious that the t-distribution is shifted to the left. This implies that the critical values will be absolutely higher than for the standard case. In addition, one may observe a slight skewness.

Finally, we also want to investigate the autocovariance function of a random walk. Using similar arguments as in Sect. 1.3 we get:

$$\begin{aligned}\gamma(h) &= \mathbb{E}\left(X_T X_{T-h}\right)\\ &= \mathbb{E}\left[\left(Z_T + Z_{T-1} + \ldots + Z_1\right)\left(Z_{T-h} + Z_{T-h-1} + \ldots + Z_1\right)\right]\\ &= \mathbb{E}\left(Z_{T-h}^2 + Z_{T-h-1}^2 + \ldots + Z_1^2\right) = (T-h)\sigma^2.\end{aligned}$$

Thus the correlation coefficient between X_T and X_{T-h} is:

$$\rho(h) = \frac{\gamma(h)}{\sqrt{\mathbb{V}X_T}\sqrt{\mathbb{V}X_{T-h}}} = \frac{T-h}{\sqrt{T(T-h)}} = \sqrt{\frac{T-h}{T}}, \qquad h \leq T.$$

The autocorrelation coefficient $\rho(h)$ therefore monotonically decreases with h, holding the sample size T constant. The rate at which $\rho(h)$ falls is, however, smaller than for ARMA processes for which $\rho(h)$ declines exponentially fast to zero. Given h, the autocorrelation coefficient converges to one for $T \to \infty$. Figure 7.3 compares the theoretical and the estimated ACF of a simulated random walk with $T = 100$. Typically, the estimated coefficients lie below the theoretical ones. In addition, we

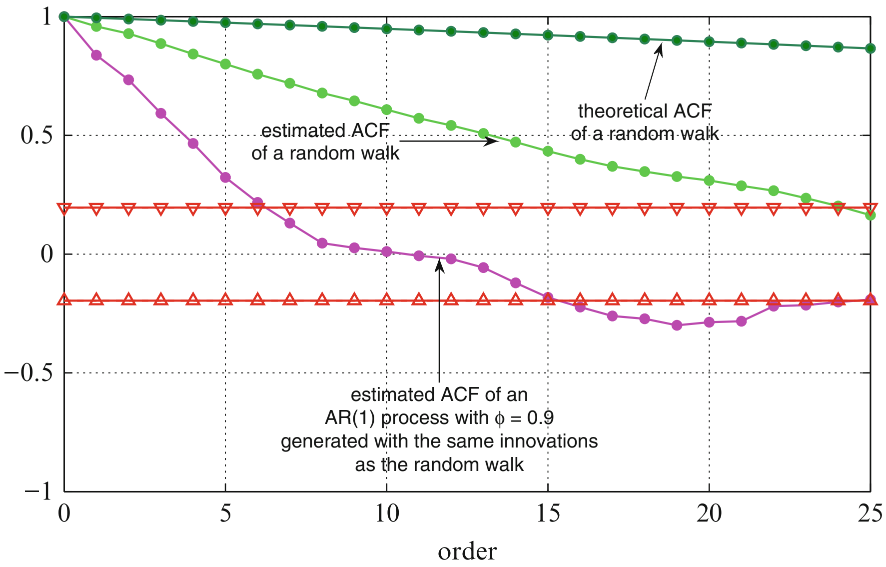

Fig. 7.3 ACF of a random walk with 100 observations

show the estimated ACF for an AR(1) process with $\phi = 0.9$ and using the same realizations of the white noise process as in the construction of the random walk. Despite the large differences between the ACF of an AR(1) process and a random walk, the ACF is only of limited use to discriminate between an (stationary) ARMA process and a random walk.

The above calculation also shows that $\rho(1) < 1$ so that the expected value of the OLS estimator is downward biased in finite samples: $\mathbb{E}\hat{\phi}_T < 1$.

7.3 Unit-Root Tests

The previous Sects. 7.1 and 7.2 have shown that, depending on the nature of the non-stationarity (trend versus difference stationarity), the stochastic process has quite different algebraic (forecast, forecast error variance, persistence) and statistical (asymptotic distribution of OLS-estimator) properties. It is therefore important to be able to discriminate among these two different types of processes. This also pertains to standard regression models for which the presence of integrated variables can lead to non-normal asymptotic distributions.

The ability to differentiate between trend- and difference-stationary processes is not only important from a statistical point of view, but can be given an economic interpretation. In macroeconomic theory, monetary and demand disturbances are alleged to have only temporary effects whereas supply disturbances, in particular technology shocks, are supposed to have permanent effects. To put it in the language of time series analysis: monetary and demand shocks have a persistence of zero whereas supply shocks have nonzero (positive) persistence. Nelson and Plosser (1982) were the first to investigate the trend properties of economic time series from this angle. In their influential study they reached the conclusion that, with the important exception of the unemployment rate, most economic time series in the US are better characterized as being difference stationary. Although this conclusion came under severe scrutiny (see Cochrane (1988) and Campbell and Perron (1991)), this issue resurfaces in many economic debates. The latest discussion relates to the nature and effect of technology shocks (see Galí (1999) or Christiano et al. (2003)).

The following exposition focuses on the Dickey-Fuller test (DF-test) and the Phillips-Perron test(PP-test). Although other test procedures and variants thereof have been developed in the meantime, these two remain the most widely applied in practice. These types of tests are also called unit-root tests.

Both the DF- as well as the PP-test rely on a regression of X_t on X_{t-1} which may include further deterministic regressors like a constant or a linear time trend. We call this regression the *Dickey-Fuller regression*:

$$X_t = \begin{array}{c}\text{deterministic}\\ \text{variables}\end{array} + \phi X_{t-1} + Z_t. \tag{7.1}$$

Alternatively and numerically equivalent, one may run the Dickey-Fuller regression in difference form:

$$\Delta X_t = \begin{array}{c}\text{deterministic}\\ \text{variables}\end{array} + \beta X_{t-1} + Z_t$$

with $\beta = \phi - 1$. For both tests, the null hypothesis is that the process is integrated of order one, difference stationary, or has a unit-root. Thus we have

$$\mathbf{H}_0 : \phi = 1 \qquad \text{or} \qquad \beta = 0.$$

The alternative hypothesis $\mathbf{H}_1$ is that the process is trend-stationary or stationary with constant mean and is given by:

$$\mathbf{H}_1 : -1 < \phi < 1 \qquad \text{or} \qquad -2 < \beta = \phi - 1 < 0.$$

Thus the unit root test is a *one-sided test*. The advantage of the second formulation of the Dickey-Fuller regression is that the corresponding t-statistic can be readily read off from standard outputs of many computer packages which makes additional computations unnecessary.

7.3.1 The Dickey-Fuller Test (DF-Test)

The Dickey-Fuller test comes in two forms. The first one, sometimes called the ρ-test, takes $T(\hat{\phi} - 1)$ as the test statistic. As shown previously, this statistic is no longer asymptotically normally distributed. However, it was first tabulated by Fuller and can be found in textbooks like Fuller (1976) or Hamilton (1994b). The second and much more common one relies on the usual t-statistic for the hypothesis $\phi = 1$:

$$t_{\hat{\phi}} = (\hat{\phi}_T - 1)/\hat{\sigma}_{\hat{\phi}}.$$

This test-statistic is also not asymptotically normally distributed. It was for the first time tabulated by Fuller (1976) and can be found, for example, in Hamilton (1994b). Later MacKinnon (1991) presented much more detailed tables where the critical values can be approximated for any sample size T by using interpolation formulas (see also Banerjee et al. (1993)).[7]

The application of the Dickey-Fuller test as well as the Phillips-Perron test is obfuscated by the fact that the asymptotic distribution of the test statistic (ρ- or t-test) depends on the specification of the deterministic components and on the true data generating process. This implies that depending on whether the Dickey-Fuller regression includes, for example, a constant and/or a time trend and on the nature of the true data generating process one has to use different tables and thus different critical values. In the following we will focus on the most common cases listed in Table 7.1.

In case 1 the Dickey-Fuller regression includes no deterministic component. Thus, a rejection of the null hypothesis implies that $\{X_t\}$ has to be a mean zero stationary process. This specification is, therefore, only warranted if one can make sure that the data have indeed mean zero. As this is rarely the case, except, for example, when the data are the residuals from a previous regression,[8] case 1 is

Table 7.1 The four most important cases for the unit-root test

Data generating process (null hypothesis)	Estimated regression $T(\hat{\phi} - 1)$	ρ-test: (Dickey-Fuller regression)	t-test
$X_t = X_{t-1} + Z_t$	$X_t = \phi X_{t-1} + Z_t$	Case 1	Case 1
$X_t = X_{t-1} + Z_t$	$X_t = \alpha + \phi X_{t-1} + Z_t$	Case 2	Case 2
$X_t = \alpha + X_{t-1} + Z_t$, $\alpha \neq 0$	$X_t = \alpha + \phi X_{t-1} + Z_t$		N(0,1)
$X_t = \alpha + X_{t-1} + Z_t$	$X_t = \alpha + \delta t$		
	$+\phi X_{t-1} + Z_t$	Case 4	Case 4

[7]These interpolation formula are now implemented in many software packages, like EVIEWS, to compute the appropriate critical values.

[8]This fact may pose a problem by itself.

very uncommon in practice. Thus, if the data do not display a trend, which can be checked by a simple time plot, the Dickey-Fuller regression should include a constant. A rejection of the null hypothesis then implies that $\{X_t\}$ is a stationary process with mean $\mu = \frac{c}{1-\phi}$. If the data display a time trend, the Dickey-Fuller regression should also include a linear time trend as in case 4. A rejection of the null hypothesis then implies that the process is trend-stationary. In the case that the Dickey-Fuller regression contains no time trend and there is no time trend under the alternative hypothesis, asymptotic normality holds. This case is only of theoretical interest as it should a priori be clear whether the data are trending or not. In the instance where one is not confident about the trending nature of the time series see the procedure outlined in Sect. 7.3.3.

In the cases 2 and 4 it is of interest to investigate the joint hypothesis $\mathbf{H}_0 : \alpha = 0$ and $\phi = 1$, and $\mathbf{H}_0 : \delta = 0$ and $\phi = 1$ respectively. Again the corresponding F-statistic is no longer F-distributed, but has been tabulated (see Hamilton (1994b, Table B7)). The trade-off between t- and F-test is discussed in Sect. 7.3.3.

Most economic time series display a significant amount of autocorrelation. To take this feature into account it is necessary to include lagged differences $\Delta X_{t-1}, \ldots, \Delta X_{t-p+1}$ as additional regressors. The so modified Dickey-Fuller regression then becomes:

$$X_t = \begin{array}{c}\text{deterministic}\\ \text{variables}\end{array} + \phi X_{t-1} + \gamma_1 \Delta X_{t-1} + \ldots + \gamma_{p-1} \Delta X_{t-p+1} + Z_t.$$

This modified test is called the augmented Dickey-Fuller test (ADF-test). This autoregressive correction does not change the asymptotic distribution of the test statistics. Thus the same tables can be used as before. For the coefficients of the autoregressive terms asymptotic normality holds. This implies that the standard testing procedures (t-test, F-test) can be applied in the usual way. This is true if instead of autoregressive correction terms moving-average terms are used instead (see Said and Dickey (1984)).

For the ADF-test the order p of the model should be chosen such that the residuals are close to being white noise. This can be checked, for example, by looking at the ACF of the residuals or by carrying out a Ljung-Box test (see Sect. 4.2). In case of doubt, it is better to choose a higher order. A consistent procedure to find the right order is to use the Akaike's criterion (AIC). Another alternative strategy advocated by Ng and Perron (1995) is an iterative testing procedure which makes use of the asymptotic normality of the autoregressive correction terms. Starting from a maximal order $p - 1 = p_{max}$, the method amounts to the test whether the coefficient corresponding to the highest order is significantly different from zero. If the null hypothesis that the coefficient is zero is not rejected, the order of the model is reduced by one and the test is repeated. This is done as long as the null hypothesis is not rejected. If the null hypothesis is finally rejected, one sticks with the model and performs the ADF-test. The successive test are standard t-tests. It is advisable to use a rather high significance level, for example a 10 % level. The simulation results by Ng and Perron (1995) show that this procedure leads to a smaller bias compared to using the AIC criterion and that the reduction in power remains negligible.

7.3.2 The Phillips-Perron Test (PP-Test)

The Phillips-Perron test represents a valid alternative to the ADF-test. It is based on the simple Dickey-Fuller regression (without autoregressive correction terms) and corrects for autocorrelation by modifying the OLS-estimate or the corresponding value of the t-statistic. The simple Dickey-Fuller regression with either constant and/or trend is:

$$X_t = \begin{array}{c}\text{deterministic}\\ \text{variables}\end{array} + \phi X_{t-1} + Z_t,$$

where $\{Z_t\}$ need no longer be a white noise process, but can be any mean zero stationary process. $\{Z_t\}$ may, for example, be an ARMA process. In principle, the approach also allows for heteroskedasticity.[9]

The first step in the Phillips-Perron unit-root test estimates the above appropriately specified Dickey-Fuller regression. The second step consists in the estimation of the unconditional variance $\gamma_Z(0)$ and the long-run variance J of the residuals $\hat{Z}_t$. This can be done using one of the methods prescribed in Sect. 4.4. These two estimates are then used in a third step to correct the ρ- and the t-test statistics. This correction would then take care of the autocorrelation present in the data. Finally, one can use the so modified test statistics to carry out the unit-root test applying the same tables for the critical values as before.

In case 1 where no deterministic components are taken into account (see case 1 in Table 7.1) the modified test statistics according to Phillips (1987) are:

$$\rho\text{-Test}: \qquad T\left(\hat{\phi} - 1\right) - \frac{1}{2}\left(\widehat{\mathrm{J}}_T - \hat{\gamma}_Z(0)\right)\left(\frac{1}{T^2}\sum_{t=1}^{T} X_{t-1}^2\right)^{-1}$$

$$\text{t-Test}: \qquad \sqrt{\frac{\hat{\gamma}_Z(0)}{\widehat{\mathrm{J}}_T}}\, t_{\hat{\phi}} - \frac{1}{2}\left(\widehat{\mathrm{J}}_T - \hat{\gamma}_Z(0)\right)\left(\frac{\widehat{\mathrm{J}}_T}{T^2}\sum_{t=1}^{T} X_{t-1}^2\right)^{-1/2}.$$

If $\{Z_t\}$ would be white noise so that $J = \gamma(0)$, respectively $\widehat{\mathrm{J}}_T \approx \hat{\gamma}_Z(0)$ one gets the ordinary Dickey-Fuller test statistic. Similar formulas can be derived for the cases 2 and 4. As already mentioned these modifications will not alter the asymptotic distributions so the same critical values as for the ADF-test can be used.

The main advantage of the Phillips-Perron test is that the non-parametric correction allows for very general $\{Z_t\}$ processes. The PP-test is particularly appropriate if $\{Z_t\}$ has some MA-components which can be only poorly approximated by low order autoregressive terms. Another advantage is that one can avoid the exact modeling of the process. It has been shown by Monte-Carlo studies that the PP-test has more power compared to the DF-test, i.e. the PP-test rejects the null hypothesis more often when it is false, but that, on the other hand, it has also a higher size distortion, i.e. that it rejects the null hypothesis too often.

[9] The exact assumptions can be read in Phillips (1987) and Phillips and Perron (1988).

7.3.3 Unit-Root Test: Testing Strategy

Independently whether the Dickey-Fuller or the Phillips-Perron test is used, the specification of the deterministic component is important and can pose a problem in practice. On the one hand, if the deterministic part is underrepresented, for example when only a constant, but no time trend is used, the test results are biased in favor of the null hypothesis, if the data do indeed have a trend. On the other hand, if too many deterministic components are used, the power of the test is reduced. It is therefore advisable to examine a plot of the series in order to check whether a long run trend is visible or not. In some circumstances economic reasoning may help in this regard.

Sometimes, however, it is difficult to make an appropriate choice a priori. We therefore propose the following testing strategy based on Elder and Kennedy (2001).

X_t **has a long-run trend**: As X_t grows in the long-run, the Dickey-Fuller regression

$$X_t = \alpha + \delta t + \phi X_{t-1} + Z_t$$

should contain a linear trend.[10] In this case either $\phi = 1$, $\delta = 0$ and $\alpha \neq 0$ (unit root case) or $\phi < 1$ with $\delta \neq 0$ (trend stationary case). We can then test the joint null hypothesis

$$\text{H}_0 : \quad \phi = 1 \text{ and } \delta = 0$$

by a corresponding F-test. Note that the F-statistic, like the t-test, is not distributed according to the F-distribution. If the test does not reject the null, we conclude that $\{X_t\}$ is a unit root process with drift or equivalently a difference-stationary (integrated) process. If the F-test rejects the null hypothesis, there are three possible situations:

(i) The possibility $\phi < 1$ and $\delta = 0$ contradicts the primary observation that $\{X_t\}$ has a trend and can therefore be eliminated.
(ii) The possibility $\phi = 1$ and $\delta \neq 0$ can also be excluded because this would imply that $\{X_t\}$ has a quadratic trend, which is unrealistic.
(iii) The possibility $\phi < 1$ and $\delta \neq 0$ represents the only valid alternative. It implies that $\{X_t\}$ is stationary around a linear trend, i.e. that $\{X_t\}$ is trend-stationary.

Similar conclusions can be reached if, instead of the F-test, a t-test is used to test the null hypothesis $\text{H}_0 : \phi = 1$ against the alternative $\text{H}_1 : \phi < 1$. Thereby a non-rejection of H_0 is interpreted that $\delta = 0$. If, however, the null hypothesis H_0 is rejected, this implies that $\delta \neq 0$, because $\{X_t\}$ exhibits a long-run trend.

[10] In case of the ADF-test additional regressors, $\Delta X_{t-j}, j > 0$, might be necessary.

The F-test is more powerful than the t-test. The t-test, however, is a one-sided test, which has the advantage that it actually corresponds to the primary objective of the test. In Monte-Carlo simulations the t-test has proven to be marginally superior to the F-test.

X_t **has no long-run trend**: In this case $\delta = 0$ and the Dickey-Fuller regression should be run without a trend[11]:

$$X_t = \alpha + \phi X_{t-1} + Z_t.$$

Thus we have either $\phi = 1$ and $\alpha = 0$ or $\phi < 1$ and $\alpha \neq 0$. The null hypothesis in this case therefore is

$$\mathrm{H}_0 : \quad \phi = 1 \text{ and } \alpha = 0.$$

A rejection of the null hypothesis can be interpreted in three alternative ways:

(i) The case $\phi < 1$ and $\alpha = 0$ can be eliminated because it implies that $\{X_t\}$ would have a mean of zero which is unrealistic for most economic time series.
(ii) The case $\phi = 1$ and $\alpha \neq 0$ can equally be eliminated because it implies that $\{X_t\}$ has a long-run trend which contradicts our primary assumption.
(iii) The case $\phi < 1$ and $\alpha \neq 0$ is the only realistic alternative. It implies that the time series is stationary around a constant mean given by $\frac{\alpha}{1-\phi}$.

As before one can use, instead of a F-test, a t-test of the null hypothesis $\mathrm{H}_0 : \phi = 1$ against the alternative hypothesis $\mathrm{H}_1 : \phi < 1$. If the null hypothesis is not rejected, we interpret this to imply that $\alpha = 0$. If, however, the null hypothesis H_0 is rejected, we conclude that $\alpha \neq 0$. Similarly, Monte-Carlo simulations have proven that the t-test is superior to the F-test.

The trend behavior of X_t **is uncertain**: This situation poses the following problem. Should the data exhibit a trend, but the Dickey-Fuller regression contains no trend, then the test is biased in favor of the null hypothesis. If the data have no trend, but the Dickey-Fuller regression contains a trend, the power of the test is reduced. In such a situation one can adapt a two-stage strategy. Estimate the Dickey-Fuller regression with a linear trend:

$$X_t = \alpha + \delta t + \phi X_{t-1} + Z_t.$$

Use the t-test to test the null hypothesis $\mathrm{H}_0 : \phi = 1$ against the alternative hypothesis $\mathrm{H}_1 : \phi < 1$. If H_0 is not rejected, we conclude the process has a unit root with or without drift. The presence of a drift can then be investigated by a simple regression of ΔX_t against a constant followed by a simple t-test of the

[11] In case of the ADF-test additional regressors, $\Delta X_{t-j}, j > 0$, might be necessary.

null hypothesis that the constant is zero against the alternative hypothesis that the constant is nonzero. As ΔX_t is stationary, the usual critical values can be used. [12] If the t-test rejects the null hypothesis H_0, we conclude that there is no unit root. The trend behavior can then be investigated by a simple t-test of the hypothesis $H_0 : \delta = 0$. In this test the usual critical values can be used as $\{X_t\}$ is already viewed as being stationary.

7.3.4 Examples of Unit-Root Tests

As our first example, we examine the logged real GDP for Switzerland, $\ln(\text{BIP}_t)$, where we have adjusted the series for seasonality by taking a moving-average. The corresponding data are plotted in Fig. 1.3. As is evident from this plot, this variable exhibits a clear trend so that the Dickey-Fuller regression should include a constant and a linear time trend. Moreover, $\{\Delta \ln(\text{BIP}_t)\}$ is typically highly autocorrelated which makes an autoregressive correction necessary. One way to make this correction is by augmenting the Dickey-Fuller regression by lagged $\{\Delta \ln(\text{BIP}_t)\}$ as additional regressors. Thereby the number of lags is determined by AIC. The corresponding result is reported in the first column of Table 7.2. It shows that AIC chooses only one autoregressive correction term. The value of t-test statistic is -3.110 which is just above the 5-% critical value. Thus, the null hypothesis is not rejected. If the autoregressive correction is chosen according to the method proposed by Ng and Perron five autoregressive lags have to be included. With this specification, the value of the t-test statistic is clearly above the critical value, implying that the null hypothesis of the presence of a unit root cannot be rejected (see second column in Table 7.2).[13] The results of the ADF-tests is confirmed by the PP-test (column 3 in Table 7.2) with quadratic spectral kernel function and band width 20.3 chosen according to Andrews' formula (see Sect. 4.4).

The second example, examines the three-month LIBOR, $\{\text{R3M}_t\}$. The series is plotted in Fig. 1.4. The issue whether this series has a linear trend or not is not easy to decide. On the one hand, the series clearly has a negative trend over the sample period considered. On the other hand, a negative time trend does not make sense from an economic point of view because interest rates are bounded from below by zero. Because of this uncertainty, it is advisable to include in the Dickey-Fuller regression both a constant and a trend to be on the safe side. Column 5 in Table 7.2 reports the corresponding results. The value of the t-statistic of the PP-test with Bartlett kernel function and band width of 5 according to the Newey-West rule of thumb is -2.142 and thus higher than the corresponding 5-% critical of -3.435.

[12] Eventually, one must correct the corresponding standard deviation by taking the autocorrelation in the residual into account. This can be done by using the long-run variance. In the literature this correction is known as the Newey-West correction.

[13] The critical value changes slightly because the inclusion of additional autoregressive terms changes the sample size.

Table 7.2 Examples of unit root tests

	$\ln(\text{BIP}_t)$	$\ln(\text{BIP}_t)$	$\ln(\text{BIP}_t)$	R3M_t	R3M_t
Test	ADF	ADF	PP	PP	PP
Autoregressive correction	AIC	Ng and Perron	Quadratic spectral	Bartlett	Bartlett
Band width			20.3	5	5
α	0.337	0.275	0.121	0.595	−0.014
δ	0.0001	0.0001	0.0002	−0.0021	
ϕ	0.970	0.975	0.989	0.963	−0.996
γ_1	0.885	1.047			
γ_2		−0.060			
γ_3		−0.085			
γ_4		−0.254			
γ_5		0.231			
$t_{\hat{\phi}}$	−3.110	−2.243	−1.543	−2.142	−0.568
Critical value (5 %)	−3.460	−3.463	−3.460	−3.435	−2.878

Critical values from MacKinnon (1996)

Thus, we cannot reject the null hypothesis of the presence of a unit root. We therefore conclude that the process $\{\text{R3M}_t\}$ is integrated of order one, respectively difference-stationary. Based on this conclusion, the issue of the trend can now be decided by running a simple regression of ΔR3M_t against a constant. This leads to the following results:

$$\Delta\text{R3M}_t = \underset{(0.0281)}{-0.0315} + e_t.$$

where e_t denotes the least-squares residual. The mean of ΔR3M_t is therefore −0.0315. This value is, however, statistically not significantly different from zero as indicated by the estimated standard error in parenthesis. Note that this estimate of the standard error has been corrected for autocorrelation (Newey-West correction). Thus, $\{\text{R3M}_t\}$ is not subject to a linear trend. One could have therefore run the Dickey-Fuller regression without the trend term. The result of corresponding to this specification is reported in the last column of Table 7.2. It confirms the presence of a unit root.

7.4 Generalizations of Unit-Root Tests

7.4.1 Structural Breaks in the Trend Function

As we have seen, the unit-root test depends heavily on the correct specification of the deterministic part. Most of the time this amounts to decide whether a linear trend is present in the data or not. In the previous section we presented a rule how

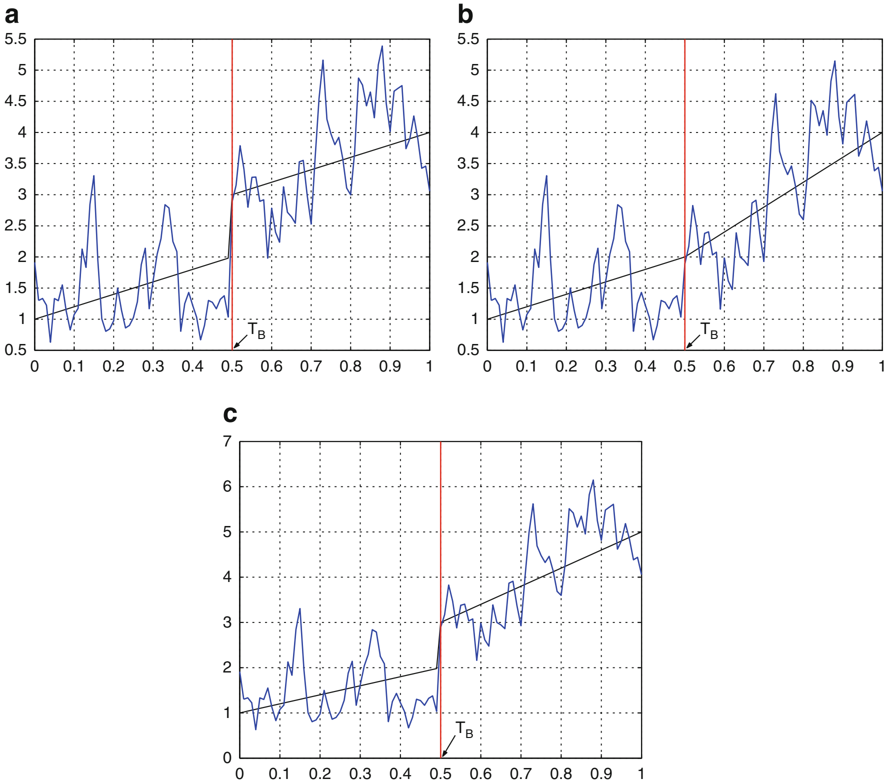

Fig. 7.4 Three types of structural breaks at T_B. (**a**) Level shift. (**b**) Change in slope. (**c**) Level shift and change in slope

to proceed in case of uncertainty about the trend. Sometimes, however, the data exhibit a structural break in their deterministic component. If this structural break is ignored, the unit-root test is biased in favor of the null hypothesis (i. e. in favor of a unit root) as demonstrated by Perron (1989). Unfortunately, the distribution of the test statistic under the null hypothesis, in our case the t-statistic, depends on the exact nature of the structural break and on its date of occurrence in the data. Following Perron (1989) we concentrate on three exemplary cases: a level shift, a change in the slope (change in the growth rate), and a combination of both possibilities. Figure 7.4 shows the three possibilities assuming that a break occurred in period T_B. Thereby an AR(1) process with $\phi = 0.8$ was superimposed on the deterministic part.

The unit-root test with the possibility of a structural break in period T_B is carried out using the Dickey-Fuller test. Thereby the date of the structural break is assumed to be known. This assumption, although restrictive, is justifiable in many applications. The first oil price shock in 1973 or the German reunification in 1989 are examples of structural breaks which can be dated exactly. Other examples would

Table 7.3 Dickey-Fuller regression allowing for structural breaks

	Model A: Level Shift
$\mathbf{H}_0$:	$X_t = \alpha + \mathbf{1}_{\{t=T_B+1\}}\delta_B + X_{t-1} + Z_t$
$\mathbf{H}_1$:	$X_t = \alpha + \delta t + \mathbf{1}_{\{t>T_B\}}(\alpha_B - \alpha) + \phi X_{t-1} + Z_t, \quad \phi < 1$
	Model B: Change in Slope (Change in Growth Rate)
$\mathbf{H}_0$:	$X_t = \alpha + \mathbf{1}_{\{t>T_B\}}(\alpha_B - \alpha) + X_{t-1} + Z_t$
$\mathbf{H}_1$:	$X_t = \alpha + \delta t + \mathbf{1}_{\{t>T_B\}}(\delta_B - \delta)(t - T_B) + \phi X_{t-1} + Z_t, \quad \phi < 1$
	Model C: Level Shift and Change in Slope
$\mathbf{H}_0$:	$X_t = \alpha + \mathbf{1}_{\{t=T_B+1\}}\delta_B + \mathbf{1}_{\{t>T_B\}}(\alpha_B - \alpha) + X_{t-1} + Z_t$
$\mathbf{H}_1$:	$X_t = \alpha + \delta t + \mathbf{1}_{\{t>T_B\}}(\alpha_B - \alpha) + \mathbf{1}_{\{t>T_B\}}(\delta_B - \delta)(t - T_B) + \phi X_{t-1} + Z_t, \quad \phi < 1$

$\mathbf{1}_{\{t=T_B+1\}}$ and $\mathbf{1}_{\{t>T_B\}}$ denotes the indicator function which takes the value one if the condition is satisfied and the value zero otherwise

include changes in the way the data are constructed. These changes are usually documented by the data collecting agencies. Table 7.3 summarizes the three variants of Dickey-Fuller regression allowing for structural breaks.[14]

Model A allows only for a level shift. Under the null hypothesis the series undergoes a one-time shift at time T_B. This level shift is maintained under the null hypothesis which posits a random walk. Under the alternative, the process is viewed as being trend-stationary whereby the trend line shifts parallel by $\alpha_B - \alpha$ at time T_B. Model B considers a change in the mean growth rate from α to α_B at time T_B. Under the alternative, the slope of time trend changes from δ to δ_B. Model C allows for both types of break to occur at the same time.

The unit-root test with possible structural break for a time series X_t, $t = 0, 1, \ldots, T$, is implemented in two stages as follows. In the first stage, we regress X_t on the corresponding deterministic component using OLS. The residuals $\widetilde{X}_0, \widetilde{X}_1, \ldots, \widetilde{X}_T$ from this regression are then used to carry out a Dickey-Fuller test:

$$\widetilde{X}_t = \phi \widetilde{X}_{t-1} + Z_t, \qquad t = 1, \ldots, T.$$

The distribution of the corresponding t-statistic under the null hypothesis depends not only on the type of the structural break, but also on the relative date of the break in the sample. Let this relative date be parameterized by $\lambda = T_B/T$. The asymptotic distribution of the t-statistic has been tabulated by Perron (1989). This table can be used to determine the critical values for the test. These critical values are smaller than those from the normal Dickey-Fuller table. Using a 5 % significance level, the critical values range between -3.80 and -3.68 for model A, between -3.96 and -3.65 for model B, and between -4.24 and -3.75 for model C, depending on the value of λ. These values also show that the dependence on λ is only weak.

In the practical application of the test one has to control for the autocorrelation in the data. This can be done by using the Augmented Dickey-Fuller (ADF) test. This

[14] See Eq. (7.1) and Table 7.1 for comparison.

amounts to the introduction of $\Delta\widetilde{X}_{t-j}$, $t = 1, 2, \ldots, p-1$, as additional regressors in the above Dickey-Fuller regression. Thereby the order p can be determined by Akaike's information criterion (AIC) or by the iterative testing procedure of Ng and Perron (1995). Alternatively, one may use, instead of the ADF test, the Phillips-Perron test. In this case one computes the usual t-statistic for the null hypothesis $\phi = 1$ and corrects it using the formulas in Phillips and Perron (1988) as explained in Sect. 7.3.2. Which of the two methods is used, is irrelevant for the determination of the critical values which can be extracted from Perron (1989).

Although it may be legitimate in some cases to assume that the time of the structural break is known, we cannot take this for granted. It is therefore important to generalize the test allowing for an unknown date for the occurrence of a structural break. The work of Zivot and Andrews (1992) has shown that the procedure proposed by Perron can be easily expanded in this direction. We keep the three alternative models presented in Table 7.3, but change the null hypothesis to a random walk with drift with no exogenous structural break. Under the null hypothesis, $\{X_t\}$ is therefore assumed to be generated by

$$X_t = \alpha + X_{t-1} + Z_t, \qquad Z_t \sim \text{WN}(0, \sigma^2).$$

The time of the structural T_B, respectively $\lambda = T_B/T$, is estimated in such a way that $\{X_t\}$ comes as close as possible to a trend-stationary process. Under the alternative hypothesis $\{X_t\}$ is viewed as a trend-stationary process with unknown break point. The goal of the estimation strategy is to chose T_B, respectively λ, in such a way that the trend-stationary alternative receives the highest weight. Zivot and Andrews (1992) propose to estimate λ by minimizing the value of the t-statistic $t_{\hat{\phi}}(\lambda)$ under the hypothesis $\phi = 1$:

$$t_{\hat{\phi}}(\hat{\lambda}_{\text{inf}}) = \inf_{\lambda \in \Lambda} t_{\hat{\phi}}(\lambda) \tag{7.2}$$

where Λ is a closed subinterval of $(0, 1)$.[15] The distribution of the test statistic under the null hypothesis for the three cases is tabulated in Zivot and Andrews (1992). This table then allows to determine the appropriate critical values for the test. In practice, one has to take the autocorrelation of the time series into account by one of the methods discussed previously.

This testing strategy can be adapted to determine the time of a structural break in the linear trend irrespective of whether the process is trend-stationary or integrated of order one. The distributions of the corresponding test statistics have been tabulated by Vogelsang (1997).[16]

[15] Taking the infimum over Λ instead over $(0, 1)$ is for theoretical reasons only. In practice, the choice of Λ plays no important role. For example, one may take $\Lambda = [0.01,\ 0.99]$.

[16] See also the survey by Perron (2006).

7.4.2 Testing for Stationarity (KPSS Test)

The unit-root tests we discussed so far tested the null hypothesis that the process is integrated of order one against the alternative hypothesis that the process is integrated of order zero (i.e. is stationary). However, one may be interested in reversing the null and the alternative hypothesis and test the hypothesis of stationarity against the alternative that the process is integrated of order one. Such a test has been proposed by Kwiatkowski et al. (1992), called the KPSS-Test. This test rests on the idea that according to the Beveridge-Nelson decomposition (see Sect. 7.1.4) each integrated process of order one can be seen as the sum of a linear time trend, a random walk and a stationary process:

$$X_t = \alpha + \delta t + d \sum_{j=1}^{t} Z_j + U_t,$$

where $\{U_t\}$ denotes a stationary process. If $d = 0$ then the process becomes trend-stationary, otherwise it is integrated of order one.[17] Thus, one can state the null and the alternative hypothesis as follows:

$$\mathrm{H}_0 : d = 0 \qquad \text{against} \qquad \mathrm{H}_1 : d \neq 0.$$

Denote by $\{S_t\}$ the process of partial sums obtained from the residuals $\{e_t\}$ of a regression of X_t against a constant and a linear time trend, i.e. $S_t = \sum_{j=1}^{t} e_j$.[18] Under the null hypothesis $d = 0$, $\{S_t\}$ is integrated of order one whereas under the alternative $\{S_t\}$ is integrated of order two. Based on this consideration Kwiatkowski et al. propose the following test statistic for a time series consisting of T observations:

$$\text{KPSS test statistic:} \qquad W_T = \frac{\sum_{t=1}^{T} S_t^2}{T^2 \widehat{J}_T} \tag{7.3}$$

where $\widehat{J}_T$ is an estimate of the long-run variance of $\{U_t\}$ (see Sect. 4.4). As $\{S_t\}$ is an integrated process under the null hypothesis, the variance of $\{S_t\}$ grows linearly in t (see Sect. 1.4.4 or 7.2) so that the sum of squared S_t diverges at rate T^2. Thus, the test statistic remains bounded and can be shown to converge. Note that the test statistic is independent from further nuisance parameters. Under the alternative hypothesis, however, $\{S_t\}$ is integrated of order two. Thus, the null hypothesis will be rejected for large values of W_T. The corresponding asymptotic critical values of the test statistic are reported in Table 7.4.

---

[17] If the data exhibit no trend, one can set δ equal to zero.

[18] This auxiliary regression may include additional exogenous variables.

Table 7.4 Critical values of the KPSS test

	Regression without time trend		
Significance level	0.1	0.05	0.01
Critical value	0.347	0.463	0.739
	Regression with time trend		
Significance level	0.1	0.05	0.01
Critical value	0.119	0.146	0.216

See Kwiatkowski et al. (1992)

7.5 Regression with Integrated Variables

7.5.1 The Spurious Regression Problem

The discussion on the Dickey-Fuller and Phillips-Perron tests showed that in a regression of the integrated variables X_t on its past X_{t-1} the standard $\sqrt{T}$-asymptotics no longer apply. A similar conclusion also holds if we regress an integrated variable X_t against another integrated variable Y_t. Suppose that both processes $\{X_t\}$ and $\{Y_t\}$ are generated as a random walk:

$$X_t = X_{t-1} + U_t, \qquad U_t \sim \text{IID}(0, \sigma_U^2)$$
$$Y_t = Y_{t-1} + V_t, \qquad V_t \sim \text{IID}(0, \sigma_V^2)$$

where the processes $\{U_t\}$ and $\{V_t\}$ are uncorrelated with each other at all leads and lags. Thus,

$$\mathbb{E}(U_t V_s) = 0, \quad \text{for all } t, s \in \mathbb{Z}.$$

Consider now the regression of Y_t on X_t and a constant:

$$Y_t = \alpha + \beta X_t + \varepsilon_t.$$

As $\{X_t\}$ and $\{Y_t\}$ are two random walks which are uncorrelated with each other by construction, one would expect that the OLS-estimate of the coefficient of X_t, $\hat{\beta}$, should tend to zero as the sample size T goes to infinity. The same is expected for the coefficient of determination R^2. This is, however, not true as has already been remarked by Yule (1926) and, more recently, by Granger and Newbold (1974). The above regression will have a tendency to "discover" a relationship between Y_t and X_t despite the fact that there is none. This phenomenon is called *spurious correlation* or *spurious regression*. Similarly, unreliable results would be obtained by using a simple t-test for the null hypothesis $\beta = 0$ against the alternative hypothesis $\beta \neq 0$. The reason for these treacherous findings is that the model is incorrect under the null as well as under the alternative hypothesis. Under the null hypothesis $\{\varepsilon_t\}$ is an integrated process which violates the standard assumption for OLS. The alternative hypothesis is not true by construction. Thus, OLS-estimates should be interpreted with caution when a highly autocorrelated process $\{Y_t\}$ is regressed on another highly correlated process $\{X_t\}$. A detailed analysis of the spurious regression problem is provided by Phillips (1986).

The spurious regression problem can be illustrated by a simple Monte Carlo study. Specifying $U_t \sim \text{IIDN}(0,1)$ and $V_t \sim \text{IIDN}(0,1)$, we constructed $N = 1000$ samples for $\{Y_t\}$ and $\{X_t\}$ of size $T = 1000$ according to the specification above. The sample size was chosen especially large to demonstrate that this is not a small sample issue. As a contrast, we constructed two independent AR(1) processes with AR-coefficients $\phi_X = 0.8$ and $\phi_Y = -0.5$.

Figures 7.5a,b show the drastic difference between a regression with stationary variables and integrated variables. Whereas the distribution of the OLS-estimates of β is highly concentrated around the true value $\beta = 0$ in the stationary case, the distribution is very flat in the case of integrated variables. A similar conclusion holds for the corresponding t-value. The probability of obtaining a t-value greater than 1.96 is bigger than 0.9. This means that in more than 90 % of the time the t-statistic leads to a rejection of the null hypothesis and therefore suggests a relationship between Y_t and X_t despite their independence. In the stationary case, this probability turns out to be smaller than 0.05. These results are also reflected in the coefficient of determination R^2. The median R^2 is approximately 0.17 in the case of the random walks, but only 0.0002 in the case of AR(1) processes.

The problem remains the same if $\{X_t\}$ and $\{Y_t\}$ are specified as random walks with drift:

$$X_t = \delta_X + X_{t-1} + U_t, \qquad U_t \sim \text{IID}(0, \sigma_U^2)$$
$$Y_t = \delta_Y + Y_{t-1} + V_t, \qquad V_t \sim \text{IID}(0, \sigma_V^2)$$

where $\{U_t\}$ and $\{V_t\}$ are again independent from each other at all leads and lags. The regression would be same as above:

$$Y_t = \alpha + \beta X_t + \varepsilon_t.$$

7.5.2 Bivariate Cointegration

The spurious regression problem cannot be circumvented by first testing for a unit root in Y_t and X_t and then running the regression in first differences in case of no rejection of the null hypothesis. The reason being that a regression in the levels of Y_t and X_t may be sensible even when both variables are integrated. This is the case when both variables are cointegrated. The concept of *cointegration* goes back to Engle and Granger (1987) and initiated a literal research boom. We will give a more general definition in Chap. 16 when we deal with multivariate time series. Here we stick to the case of two variables and present the following definition.

Definition 7.2 (Cointegration, Bivariate)**.** *Two stochastic processes* $\{X_t\}$ *and* $\{Y_t\}$ *are called* cointegrated *if the following two conditions are fulfilled:*

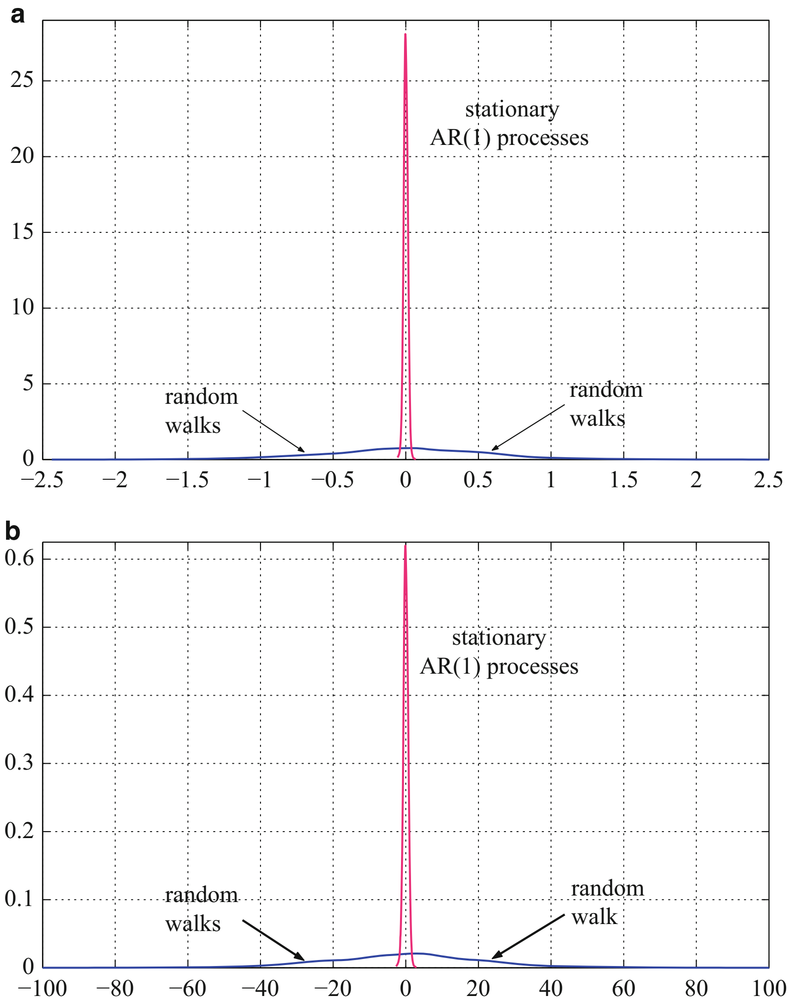

Fig. 7.5 Distribution of OLS-estimate $\hat{\beta}$ and t-statistic $t_{\hat{\beta}}$ for two independent random walks and two independent AR(1) processes. (**a**) Distribution of $\hat{\beta}$. (**b**) Distribution of $t_{\hat{\beta}}$. (**c**) Distribution of $\hat{\beta}$ and t-statistic $t_{\hat{\beta}}$

(i) $\{X_t\}$ *and* $\{Y_t\}$ *are both integrated processes of order one, i.e.* $X_t \sim \mathrm{I}(1)$ *and* $Y_t \sim \mathrm{I}(1)$*;*
(ii) there exists a constant $\beta \neq 0$ *such that* $\{Y_t - \beta X_t\}$ *is a stationary process, i.e.* $\{Y_t - \beta X_t\} \sim \mathrm{I}(0)$.

The issue whether two integrated processes are cointegrated can be decided on the basis of a unit root test. Two cases can be distinguished. In the first one, β is assumed to be known. Thus, one can immediately apply the augmented Dickey-Fuller (ADF) or the Phillips-Perron (PP) test to the process $\{Y_t - \beta X_t\}$. Thereby the same issue regarding the specification of the deterministic part arises. The critical values can be retrieved from the usual tables (for example from MacKinnon 1991). In the second case, β is not known and must be estimated from the data. This can be done running, as a first step, a simple (cointegrating) regression of Y_t on X_t including a constant and/or a time trend.[19] Thereby the specification of the deterministic part follows the same rules as before. The unit root test is then applied, in the second step, to the residuals from this regression. As the residuals have been obtained from a preceding regression, we are faced with the so-called "generated regressor problem".[20] This implies that the usual Dickey-Fuller tables can no longer be used, instead the tables provided by Phillips and Ouliaris (1990) become the relevant ones. As before, the corresponding asymptotic distribution depends on the specification of the deterministic part in the cointegrating regression. If this regression included a constant, the residuals have necessary a mean of zero so that the Dickey-Fuller regression should include no constant (case 1 in Table 7.1):

$$e_t = \phi e_{t-1} + \xi_t$$

where e_t and ξ_t denote the residuals from the cointegrating and the residuals of the Dickey-Fuller regression, respectively. In most applications it is necessary to correct for autocorrelation which can be done by including additional lagged differences $\Delta\hat{\varepsilon}_{t-1}, \ldots, \Delta\hat{\varepsilon}_{t-p+1}$ as in the ADF-test or by adjusting the t-statistic as in the PP-test. The test where β is estimated from a regression is called the *regression test for cointegration*. Note that if the two series are cointegrated then the OLS estimate of β is (super) consistent.

In principle it is possible the generalize this single equation approach to more than two variables. This encounters, however, some conceptual problems. First, there is the possibility of more than one linearly independent cointegrating relationships which cannot be detected by a single regression. Second, the dependent variable in the regression may not be part of the cointegrating relation which might involves only the other variables. In such a situation the cointegrating regression is again subject to the spurious regression problem. These issues turned the interest of the profession towards multivariate approaches. Chapter 16 presents alternative procedures and discusses the testing, estimation, and interpretation of cointegrating relationships in detail.

[19]Thereby, in contrast to ordinary OLS regressions, it is irrelevant which variable is treated as the left hand, respectively right hand variable.

[20]This problem was first analyzed by Nicholls and Pagan (1984) and Pagan (1984) in a stationary context.

An Example for Bivariate Cointegration

As an example, we consider the relation between the short-term interest rate, $\{R3M_t\}$, and inflation, $\{INFL_t\}$, in Switzerland over the period January 1989 to February 2012. As the short-term interest rate we take the three month LIBOR. Both time series are plotted in Fig. 7.6a. As they are integrated according to the unit root tests (not shown here), we can look for cointegration. The cointegrating regression delivers:

$$\mathrm{INFL}_t = -0.0088 + 0.5535\ \mathrm{R3M}_t + e_t, \qquad \mathrm{R}^2 = 0.7798. \tag{7.4}$$

The residuals from this regression, denoted by e_t, are represented in Fig. 7.6b. The ADF unit root test of these residuals leads to a value of -3.617 for the t-statistic. Thereby an autoregressive correction of 13 lags was necessary according to the AIC criterion. The corresponding value of the t-statistic resulting from the PP unit root test using a Bartlett window with band width 7 is -4.294. Taking a significance level of 5 %, the critical value according to Phillips and Ouliaris (1990, Table IIb) is -3.365.[21] Thus, the ADF as well as the PP test reject the null hypothesis of a unit root in the residuals. This implies that inflation and the short-term interest rate are cointegrated.

7.5.3 Rules to Deal with Integrated Times Series

The previous sections demonstrated that the handling of integrated variables has to be done with care. We will therefore in this section examine some rules of thumb which should serve as a guideline in practical empirical work. These rules are summarized in Table 7.5. In that this section follows very closely the paper by Stock and Watson (1988b) (see also Campbell and Perron 1991).[22] Consider the linear regression model:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \ldots + \beta_K X_{K,t} + \varepsilon_t. \tag{7.5}$$

This model is usually based on two assumptions:

(1) The disturbance term ε_t is white noise and is uncorrelated with any regressor. This is, for example, the case if the regressors are deterministic or exogenous.
(2) All regressors are either deterministic or stationary processes.

If Eq. (7.5) represents the true data generating process, $\{Y_t\}$ must be a stationary process. Under the above assumptions, the OLS-estimator is consistent and the OLS-estimates are asymptotically normally distributed so that the corresponding t- and F-statistics will be approximately distributed as t- and F- distributions.

---

[21] For comparison, the corresponding critical value according to MacKinnon (1991) is -2.872.

[22] For a thorough analysis the interested reader is referred to Sims et al. (1990).

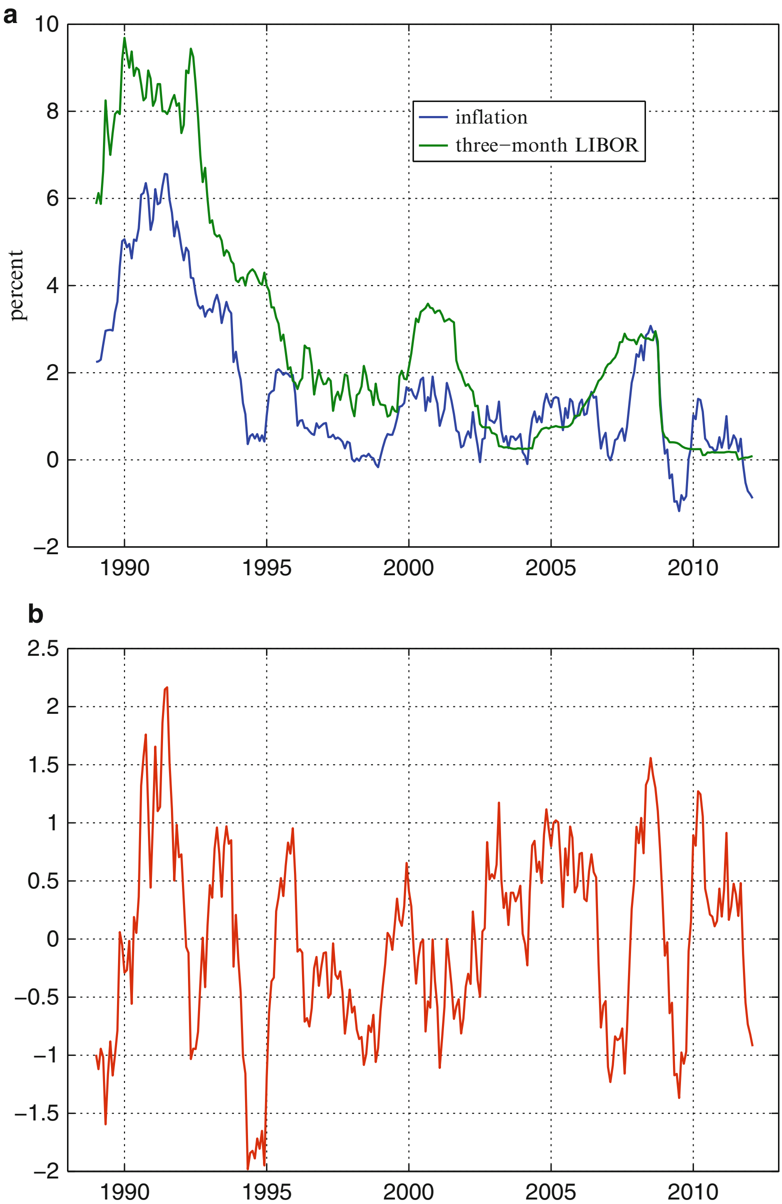

Fig. 7.6 Cointegration of inflation and three-month LIBOR. (**a**) Inflation and three-month LIBOR. (**b**) Residuals from cointegrating regression

Consider now the case that assumption 2 is violated and that some or all regressors are integrated, but that instead one of the two following assumptions holds:

(2.a) The relevant coefficients are coefficients of mean-zero stationary variables.
(2.b) Although the relevant coefficients are those of integrated variables, the regression can be rearranged in such a way that the relevant coefficients become coefficients of mean-zero stationary variables.

Under assumptions 1 and 2.a or 2.b the OLS-estimator remains consistent. Also the corresponding t- and F-statistics remain valid so that the appropriate critical values can be retrieved from the t-, respectively F-distribution. If neither assumption 2.a nor 2.b holds, but the following assumption:

(2.c) The relevant coefficients are coefficients of integrated variables and the regression cannot be rewritten in a way that they become coefficients of stationary variables.

If assumption 1 remains valid, but assumption 2.c holds instead of 2.a and 2.b, the OLS-estimator is still consistent. However, the standard asymptotic theory for the t- and the F-statistic fails so that they become useless for normal statistical inferences.

If we simply regress one variable on another in levels, the error term ε_t is likely not to follow a white noise process. In addition, it may even be correlated with some regressors. Suppose that we replace assumption 1 by:

(1.a) The integrated dependent variable is cointegrated with at least one integrated regressor such that the error term is stationary, but may remain autocorrelated or correlated with the regressors.

Under assumptions 1.a and 2.a, respectively 2.b, the regressors are stationary, but correlated with the disturbance term, in this case the OLS-estimator becomes inconsistent. This situation is known as the classic omitted variable bias, simultaneous equation bias or errors-in-variable bias. However, under assumptions 1.a and 2.c, the OLS-estimator is consistent for the coefficients of interest. However, the standard asymptotic theory fails. Finally, if both the dependent variable and the regressors are integrated without being cointegrated, then the disturbance term is integrated and the OLS-estimator becomes inconsistent. This is the spurious regression problem treated in Sect. 7.5.1.

Example: Term Structure of Interest

We illustrate the above rules of thumb by investigating again the relation between inflation ($\{\text{INFL}_t\}$) and the short-term interest rate ($\{\text{R3M}_t\}$). In Sect. 7.5.2 we found that the two variables are cointegrated with coefficient $\hat{\beta} = 0.5535$ (see Eq. (7.4)). In a further step we want to investigate a dynamic relation between the two variables and estimate the following equation:

Table 7.5 Rules of thumb in regressions with integrated processes

		Remarks		
Assumptions	OLS-estimator	Consistency	Standard asymptotics	
(1)	(2)	Yes	Yes	Classic results for OLS
(1)	(2.a)	Yes	Yes	
(1)	(2.b)	Yes	Yes	
(1)	(2.c)	Yes	No	
(1.a)	(2.a)	No	No	Omitted variable bias
(1.a)	(2.b)	No	No	Omitted variable bias
(1.a)	(2.c)	Yes	No	
Neither (1) nor (1.a)	(2.c)	No	No	Spurious regression

Source: Stock and Watson (1988b); results for the coefficients of interest

$$\mathrm{R3M}_t = c + \phi_1 \mathrm{R3M}_{t-1} + \phi_2 \mathrm{R3M}_{t-2} + \phi_3 \mathrm{R3M}_{t-3} + \delta_1 \mathrm{INFL}_{t-1} + \delta_2 \mathrm{INFL}_{t-2} + \varepsilon_t$$

where $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$. In this regression we want to test the hypotheses $\phi_3 = 0$ against $\phi_3 \neq 0$ and $\delta_1 = 0$ against $\delta_1 \neq 0$ by examining the corresponding simple t-statistics. Note that we are in the context of integrated variables so that the rules of thumb summarized in Table 7.5 apply. We can rearrange the above equation as

$$\begin{aligned}\Delta\mathrm{R3M}_t = c \\ &+ (\phi_1 + \phi_2 + \phi_3 - 1)\mathrm{R3M}_{t-1} - (\phi_2 + \phi_3)\Delta\mathrm{R3M}_{t-1} - \phi_3 \Delta\mathrm{R3M}_{t-2} \\ &+ \delta_1 \mathrm{INFL}_{t-1} + \delta_2 \mathrm{INFL}_{t-2} + \varepsilon_t.\end{aligned}$$

ϕ_3 is now a coefficient of a stationary variable in a regression with a stationary dependent variables. In addition $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$ so that assumptions (1) und (2.b) are satisfied. We can therefore use the ordinary t-statistic to test the hypothesis $\phi_3 = 0$ against $\phi_3 \neq 0$. Note that it is not necessary to actually carry out the rearrangement of the equation. All relevant item can be retrieved from the original equation.

To test the hypothesis $\delta_1 = 0$, we rearrange the equation to yield:

$$\begin{aligned}\Delta\mathrm{R3M}_t = c \\ &+ (\phi_1 + \delta_1 \hat{\beta} - 1)\mathrm{R3M}_{t-1} + \phi_2 \mathrm{R3M}_{t-2} + \phi_3 \mathrm{R3M}_{t-3} \\ &+ \delta_1(\mathrm{INFL}_{t-1} - \hat{\beta}\mathrm{R3M}_{t-1}) + \delta_2 \mathrm{INFL}_{t-2} + \varepsilon_t.\end{aligned}$$

As $\{\mathrm{R3M}_t\}$ and $\{\mathrm{INFL}_t\}$ are cointegrated, $\mathrm{INFL}_{t-1} - \hat{\beta}\mathrm{R3M}_{t-1}$ is stationary. Thus assumptions (1) and (2.b) hold again and we use once more the simple t-test. As before it is not necessary to actually carry out the transformation.