http://www.springer.com/0-387-27959-8

Analysis of Integrated and Co-integrated Time Series with R

Pfaff, B.

2006, Approx. 150 p., Softcover
ISBN: 0-387-27959-8

2

Nonstationary Time Series

In this chapter, models for nonstationary time series are introduced. Before the characteristics of unit processes are presented, the differences between trend- and difference-stationary models are outlined. In the last section, long memory processes, i.e.*, fractionally integrated processes, are presented as a bridge between stationary and unit root processes.*

2.1 Trend- *versus* Difference-Stationary Series

In the last chapter, a model class for stationary time series has been introduced. For instance, it has been shown that a stable autoregressive (AR)(p) can be inverted to an infinite moving average (MA)–process with a constant mean. However, most macroeconomic time series seem not to adhere to such a data-generating process (see Figure 1.1). In this section, we will for examplary reasons consider a more encompassing data-generated process that was presented by Campbell and Perron [9].

Now, it is assumed that a time series $\{y_t\}$ is a realization of a deterministic trend and a stochastic component:

$$y_t = TD_t + z_t \ , \tag{2.1}$$

where TD_t assigns a deterministic trend: $TD_t = \beta_1 + \beta_2 t$ and z_t represents the stochastic component: $\phi(L)z_t = \theta(L)\varepsilon_t$ with $\varepsilon_t \sim$ i.i.d, *i.e.*, an autoregressive-moving average (ARMA)(p, q)–process. We distinguish two cases. First, if all roots of the autoregressive polynomial lie outside the unit circle (see Equation 1.14), then $\{y_t\}$ is stationary around a deterministic trend. In this instance, one could remove the trend from the original series $\{y_t\}$ and fit an ARMA(p, q) to the residuals.[1]

This *trend-stationary* model is also termed an integrated model of order zero or more compactly, the $I(0)$–model. Second, assume now that one root of the autoregressive polynomial lies on the unit circle and the remaining ones are all outside. Here, $\Delta z_t = (1 - L)z_t$ is stationary around a constant mean.

[1] A deterministic trend is most easily subtracted from a series, *i.e.*, a vector `y`, by issuing the following command: `detrended <- residuals(lm(y ~ seq(along=y)))`.

The series $\{y_t\}$ is *difference-stationary* because one has to apply the first difference filter with respect to time to obtain a stationary process. Likewise to the trend-stationary model this difference stationary model is referred to as an integrated model of order one or shortly the $I(1)$–model. The meaning of "intergrated" should now be obvious: Once the series has been differenced to obtain a stationary process, it must be integrated once, *i.e*, the reversal, to achieve the original series, hence the $I(1)$–model. An ARMA(p, q)–model could then be fitted to the differenced series. This model class is termed the *autoregressive integrated moving average* (ARIMA)(p, d, q), where d refers to the order of integration, *i.e.*, how many times the original series must be differenced until a stationary one is obtained. It should be noted that unit roots, *i.e.*, roots of the autoregressive polynomial that lie on the unit circle, are solely referring to the stochastic component in Equation 2.1.

The distinction between a trend- and a difference-stationary processes is examplified by the following two processes:

$$y_t = y_{t-1} + \mu = y_0 + \mu t \ , \tag{2.2a}$$

$$y_t = y_{t-1} + \varepsilon_t = y_0 + \sum_{s=1}^{t} \varepsilon_s \ , \tag{2.2b}$$

where μ is a fixed constant and ε_t is a white noise process. In Equation 2.2a, $\{y_t\}$ is represented by a deterministic trend, whereas in Equation 2.2b, the series is explained by its cumulated shocks, *i.e.*, a stochastic trend.

So far, the stochastic component z_t has been modeled as an ARIMA(p, d, q)–model. To foster the understanding of unit roots, we will decompose the stochastic component into a cyclical component c_t and a stochastic trend TS_t. It is assumed that the cyclical component is a mean-stationary process, whereas all random shocks are captured by the stochastic component. Now, the data-generating process for $\{y_t\}$ is decomposed into a *deterministic trend*, a *stochastic trend*, and a *cyclcical component*. For the trend stationary model, the stochastic trend is zero and the cyclical component is equal to the ARMA(p, q)–model: $\phi(L)z_t = \theta(l)\varepsilon_t$. In the case of a difference-stationary model, the autoregressive polynomial contains a unit root that can be factored out: $\phi(L) = (1 - L)\phi^*(L)$, whereby the roots of the polynomial $\phi^*(L)$ are outside the unit circle. It is then possible to express Δz_t as a moving average process (for comparison, see Equations 1.28a and 1.28b):

$$\phi^*(L)\Delta z_t = \theta(L)\varepsilon_t \ , \tag{2.3a}$$

$$\Delta z_t = \phi^*(L)\theta(L)\varepsilon_t \ , \tag{2.3b}$$

$$\Delta z_t = \psi(L)\varepsilon_t \ . \tag{2.3c}$$

Beveridge and Nelson have shown that Equation 2.3c can be transformed to (*Beveridge–Nelson decomposition*)

$$z_t = TS_t + c_t = \psi(1)S_t + \psi^*(L)\varepsilon_t \ , \tag{2.4}$$

where the sum of the moving average coefficients is denoted by $\psi(1)$, S_t is the sum of the past and present random shocks: $\sum_{s=1}^{t} \varepsilon_s$, and the polynomial $\psi^*(L)$ is equal to $(1-L)^{-1}[\psi(L) - \psi(1)]$ (see Beveridge and Nelson [7]).

The time series $\{y_t\}$ is now explained by a trend function that consists of a deterministic trend as well as a stochastic component, namely $TS_T = \psi(1)S_t$. The latter affects the absolute term in each period. Because the stochastic trend is defined as the sum of the moving average coefficients of Δz_t, it can be interpreted as the long-run impact of a shock to the level of z_t. In contrast, the cyclical component, $c_t = \psi^*(L)\varepsilon_t$ exerts no long-run impact on the level of z_t. Now, we can distinguish the following four cases: (1) $\psi(1) > 1$: The long-run impact of the shocks is greater than the intermediate ones, and hence the series is characterized by an explosive path; (2) $\psi(1) < 1$: The impact of the shocks diminishes as time passes by, (3) $\psi(1) = 0$: The time series $\{y_t\}$ is a trend-stationary process, and (4) $\psi(1) = 1$: The data-generated process is a random walk. The fourth case will be a subject in the next section.

2.2 Unit Root Processes

As stated in the last section, if the sum of the moving average coefficients, $\psi(1)$ equals one, a *random walk* process results. This data-generating process has attracted much interest in the empirical literature, in particular in the field of financial econometrics. Hence, a random walk is not only a prototype for a unit root process, but it is implied by economic and financial hypotheses as well (*i.e.*, the efficient market hypothesis). Therefore, we will begin this section by analyzing random walk processes in more detail before statistical tests and strategies for detecting unit roots are presented.

A pure random walk without a drift is defined as

$$y_t = y_{t-1} + \varepsilon_t = y_0 + \sum_{s=1}^{t} \varepsilon_t \ , \tag{2.5}$$

where $\{\varepsilon_t\}$ is an i.i.d. process, *i.e.*, white noise. For the sake of simplicity, assume that the expected value of y_0 is zero and that the white noise process $\{\varepsilon_t\}$ is independent of y_0. Then it is trivial to show that (1) $E[y_t] = 0$ and $var(y_t) = t\sigma^2$. Clearly, a random walk is a *nonstationary time series* process because its variance grows with time. Second, the best forecast of random walk is its value one period earlier, *i.e.*, $\Delta y_t = \varepsilon_t$. Incidentally, it should be noted that the i.i.d. assumption for the error process $\{\varepsilon_t\}$ is important with respect to the conclusions drawn above. Suppose that the data-generated process for $\{y_t\}$ is

$$y_t = y_{t-1} + \varepsilon_t \ , \ \varepsilon_t = \rho\varepsilon_{t-1} + \xi_t \ , \tag{2.6}$$

where $|\rho| < 1$ and ξ_t is a white noise process instead. Then, $\{y_t\}$ is not a random walk process, but it still has a unit root and is a first-order nonstationary process.

Let us now consider the case of a random walk with drift:

$$y_t = \mu + y_{t-1} + \varepsilon_t = y_0 + \mu t + \sum_{s=1}^{t} \varepsilon_t \ , \tag{2.7}$$

where, likewise to the pure random walk process, $\{\varepsilon_t\}$ is white noise. For $\mu \neq 0$, $\{y_t\}$ contains a deterministic trend with drift parameter μ. The sign of this drift parameter causes the series to wander upward if positive and downward if negative, whereas the size of the absolute value affects the steepness.

In Rcode example 2.1, three time series have been generated. For a better comparability between those, all series have been calculated with the same sequence of random numbers drawn from a standard normal distribution. First, a pure random walk has been generated by calculating the cumulated sum of 500 random numbers stored in the vector object `e`. A deterministic trend has been set with the short form of the `seq()` function, *i.e.*, the colon operator. As a second time series model, a random walk with drift can now be easily created acording to Equation 2.7. Last, the deterministic trend has been overlayed with the stationary series of normally distributed errors. All three series are plotted in Figure 2.1. By ocular econometrics, it should be evident that the statistical discrimination between a deterministic trend contaminated with noise and a random walk with drift is not easy. Likewise, it is difficult to distinguish between a random walk process and a stable AR(1)–process in which the autoregressive coefficient is close to unity. The latter two time series processes are displayed in Exhibit 2.2.

Rcode 2.1 Stochastic and deterministic trends

```
set.seed(123456)
e <- rnorm(500)
# pure random walk
rw.nd <- cumsum(e)
# trend
trd <- 1:500
# random walk with drift
rw.wd <- 0.5*trd + cumsum(e)
# deterministic trend and noise
dt <- e + 0.5*trd
# plotting
par(mar=rep(5,4))
plot.ts(dt, lty=1, ylab='', xlab='')
lines(rw.wd, lty=2)
par(new=T)
plot.ts(rw.nd, lty=3, axes=FALSE)
axis(4, pretty(range(rw.nd)))
lines(rw.nd, lty=3)
legend(10, 18.7, legend=c('det. trend + noise (ls)', 'rw
    drift (ls)', 'rw (rs)'), lty=c(1, 2, 3))
```

Before a testing procedure for the underlying data-generating process is outlined, we will introduce a formal defintion of integrated series and briefly touch on the concept of *seasonal integration*, which will be presented in more detail in Section 5.2.

In the seminal paper by Engle and Granger [26] an integrated series is defined as follows.

Definition 2.1. *A series with no deterministic component that has a stationary, invertible, ARMA representation after differencing* d *times is said to be integrated of order* d*, which is denoted as* $x_t \sim I(d)$.

That is, a stationary series is simply written as an $I(0)$–process, whereas a random walk is said to follow an $I(1)$–process, because it has to be differenced once, before stationarity is achieved. It should be noted at this point that some macroeconomic series are already differenced. For example, the real net investment in an economy is the difference of its capital stock. If investment is an $I(1)$–process, then the capital stock must behave like an $I(2)$–process. Similarily, if the inflation rate, measured as the difference of the logarithmic price index, is integrated of order one, then the price index follows an $I(2)$–process. Therefore, stationarity of $y_t \sim I(2)$ is achieved by taking the first differences of the first differences:

$$\Delta\Delta y_t = \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}\,. \quad (2.8)$$

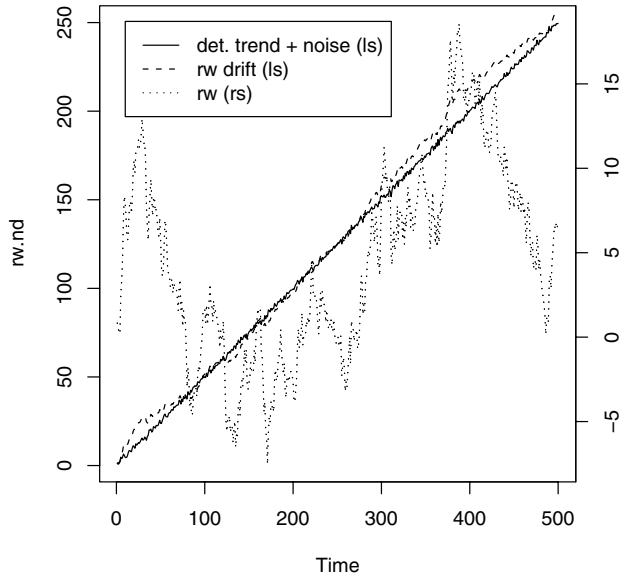

Fig. 2.1. Time series plot of deterministic and stochastic trends

If a series is already stationary $I(0)$, then no further differencing is necessary. When a series $\{y_t\}$ is a linear combination of $x_{1t} \sim I(0)$ and $x_{2t} \sim I(1)$, then $\{y_t\}$ will be an $I(1)$–process. Likewise, a linear transformation of an $I(d)$–process conserves the order of integration: $y_t \sim I(d)$, so it will be $\alpha + \beta y_t \sim I(d)$, where α and β are constants.

By now, we have only considered data-generating processes in which the unit root occurs for its own values lagged by one period. One can generalize these processes to

$$y_t = y_{t-s} + \varepsilon_t \ , \tag{2.9}$$

where $s \geq 1$. If s equals a seasonal frequency of the series, then $\{y_t\}$ is determined by its prior seasonal values plus noise. Likewise to the concept of a stochastic trend, this data-generating process is termed *stochastic seasonality*. In practice, seasonality is often accounted for by the inclusion of seasonal dummy variables or the use of seasonally adjusted data. However, there might be instances where the allowance of a seasonal component to drift over time is necessary. Analogously to the presentation of the unit root processes at the zero frequency, we can define the lag operator for seasonal unit roots as

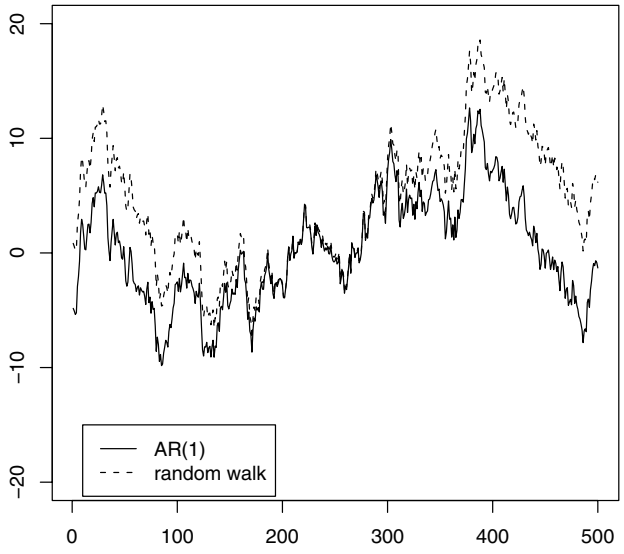

Fig. 2.2. Plot of a random walk and a stable AR(1)–process, $\phi = 0.99$

$$\Delta_s = (1 - L^s) \tag{2.10a}$$
$$= (1 - L)(1 + L + L^2 + \ldots + L^{s-1}) = \Delta S(L) \; . \tag{2.10b}$$

In Equation 2.10b the unit root at the zero frequency has been factored out. Hence, a seasonally integrated series can be represented as the product of the first difference operator and the moving-average seasonal filter $S(L)$. According to Engle *et al.* [27] a seasonally integrated series can be defined as follows.

Definition 2.2. *A variable* $\{y_t\}$ *is said to be seasonally integrated of orders* d *and* D*, which are denoted as SI(*d, D*), if* $\Delta^d S(L)^D y_t$ *is stationary.*

Therefore, if a quarterly series $\Delta_4 y_t$ is stationary, then $\{y_t\}$ is $SI(1, 1)$. Testing for seasonal unit roots is similar although a bit more complicated to testing for unit roots at the zero freqeuncy, which will be presented in the next paragraphs. The probably simplest test has been proposed by Hasza and Fuller [40], Dickey *et al.* [18], and a modification of it by Osborn *et al.* [71]. However, more complicated testing procedure that allows for cyclical movements at different frequencies has been introduced into the literature by Hylleberg *et al.* [49]. In `R`, seasonal unit root tests are implemented in the `CRAN`–package

`uroot`.

Recall the decomposition of a time series $\{y_t\}$ as in Equation 2.1. Now we want to investigate if the process $\{z_t\}$ does contain a unit root:

$$z_t = y_t - TD_t \ . \tag{2.11}$$

Hence, a deterministic trend is removed from the original series first and the residuals are tested for a unit root. Dickey and Fuller [16] proposed the following test regression that is delineated from an assumed AR(1)–process of $\{z_t\}$ (henceforth: DF–test):

$$z_t = \theta z_{t-1} + \varepsilon_t \ , \tag{2.12a}$$
$$z_t - z_{t-1} = \theta z_{t-1} - z_{t-1} + \varepsilon_t \ , \tag{2.12b}$$
$$\Delta z_t = (\theta - 1) z_{t-1} + \varepsilon_t \ , \tag{2.12c}$$
$$\Delta z_t = \pi z_{t-1} + \varepsilon_t \ . \tag{2.12d}$$

Under the null hypothesis of a unit root $\pi = 0$, which is equivalent to $\theta = 1$ and the alternative is a trend stationary process, *i.e.*, $\pi < 0$ or $\theta < 1$. Please note, that an explosive path for $\{z_t\}$, $\pi > 0$, is excluded. Equation 2.12d can be estimated by the ordinary least-squares method. The significance of π can be tested by a usual Student t ratio. However, this test statistic does not have the familiar Student t distribution. Under the null hypothesis, an $I(0)$–variable is regressed on an $I(1)$–variable in Equation 2.12d. In this case, the limiting distribution of the Student t ratio is not normal. Fortunately, critical values have been calculated by simulation and are publicized in Fuller [32], Dickey and Fuller [17], and MacKinnon [62], for instance.

So far we have only stated that a deterministic trend is removed before testing for a unit root. In reality neither the existence nor the form of the deterministic component is known *a priori*. Hence, we have to choose from the set of deterministic variables DV_t the one that best suites the data-generating process. The most obvious candidates as DV_t are simply a constant, a linear trend, or higher polynomials in the trend function, *i.e.*, square or cubic. In practice, only the first two are considered. The aim of characterizing the noise-function $\{z_t\}$ is still the same, but now we have to take the various DV_t as deterministic regressors DR_t into account too. The above-described two-step procedure (Equations 2.11 and 2.12) can be carried out in one equation:

$$\Delta y_t = \boldsymbol{\tau}' DR_t + \pi y_{t-1} + u_t \ , \tag{2.13}$$

where $\boldsymbol{\tau}$ is the coefficient vector of the deterministic part and $\{u_t\}$ assigns an error term. For the one-step procedure, a difficulty now arises, because under the validity of the null hypothesis, the deterministic trend coefficient $\boldsymbol{\tau}$ is null, whereas under the alternative it is not. Hence, the distribution of the Student t ratio of π depends now on these nuisance parameters too. The

reason for this is that the true deterministic component is unknown and must be estimated. Critical values for different deterministic components can be found in the above-cited literature as well as in Ouliaris *et al.* [72].

A weakness of the original DF–test is that it does not take a possible serial correlation of the error process $\{u_t\}$ into account. Dickey and Fuller [17] have suggested replacing the AR(1)–process for $\{z_t\}$ in Equation 2.12a with an ARMA(p, q)–process, $\phi(L)z_t = \theta(L)\varepsilon_t$. If the noise component is an AR(p)–process, it can be shown that the following test regression:

$$\Delta y_t = \boldsymbol{\tau}' DR_t + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + u_t \text{ with } k = p - 1 \tag{2.14}$$

ensures that the serial correlation in the error is removed. This test regression is called the *augmented Dickey–Fuller* test, henceforth, the ADF–test. Several methods for selecting k have been suggested in the literature. The most prominent one is the *general-to-specific* method. Here, one starts with an *a priori* choosen upper bound $k_{\max}$ and then drops the last lagged regressor if it is insignificant. In this case, the Student t distribution is applicable. You repeat these steps until the last lagged regressors is significant, otherwise you drop it each time the equation is reestimated. If no endogenously lagged regressor turns out to be significant, you choose $k = 0$, hence the DF–test results. This procedure will asymptotically yield the correct lag order or greater to the true order with probability one. Other methods for selecting an appropriate order k are based on information criteria, like Akaike (AIC) [1] or Schwarz (SC) [90]. Alternatively, the lag order can be determined by testing the residuals for a lack of serial correlation as can be tested *via* the Ljung–Box Portmanteau test (LB) or a Lagrange multiplier test (LM). In general, the SC, LB, or LM tests coincide with respect to selecting an optimal lag length k. Whereas the AIC and the general-to-specific method will mostly imply a lag length at least as large as the one by the former methods.

Once the lag order k is empirically determined, the next steps involve a *testing procedure* as graphically illustrated in Figure 2.3. First, the encompassing ADF–test equation:

$$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + u_t \tag{2.15}$$

is estimated. Dependent on this result are the further steps to be taken, until one can conclude that the series is

i) stationary around a zero mean,
ii) stationary around a nonzero mean,
iii) stationary around a linear trend,

iv) contains a unit root with zero drift,
v) contains a unit root with nonzero drift.

To be more concrete, the testing strategy starts by testing if $\pi = 0$ using the t statistic τ_τ. This statistic is not standard Student t distributed, but critical values can be found in Fuller [32]. If this test is rejected, then there is no need to proceed further. The testing sequence is continued by an F type test Φ_3 with $H_0 : \beta_2 = \pi = 0$ using the critical values tabulated in Dickey and Fuller [17]. If it is significant, then test again for a unit root using the standardized normal. Otherwise if the hypothesis $\beta_2 = 0$ cannot be rejected, reestimate the Equation 2.15 but without a trend. The corresponding t and F statistics for testing if $H_0 : \pi = 0$ and $H_0 : \beta_1 = \pi = 0$ are denoted by $\tau_\mu(\tau)$ and Φ_1. Again, the critical values for these test statistics are provided in the above-cited literature. If the null hypothesis of $\tau_\mu(\tau)$ is rejected, then there is again no need to go further. If it is not, then employ the F statistic Φ_1 for testing of the presence of a constant and a unit root.

However, the testing procedure does not end here. If the hypothesis $\pi = 0$ cannot be rejected in Equation 2.15, then the series might be integrated of order higher than zero. Therefore, one has to test whether the series is $I(1)$ or possibly $I(2)$ or even integrated to a higher degree. A natural approach would be to apply the DF– or ADF–test to

$$\Delta\Delta y_t = \pi \Delta y_{t-1} + u_t \ . \tag{2.16}$$

If the null hypothesis $\pi = 0$ is rejected, then $\Delta y_t \sim I(0)$ and $y_t \sim I(1)$, otherwise one must test subsequently whether $y_t \sim I(2)$. This testing procedure is termed *bottom-up*. However, two possibilities arise by using this bottom-up approach. First, the series cannot be transformed to stationarity regardless of how many times the difference operator is applied. Second, the danger of overdifferencing exists; that is, one falsly concludes an integration order higher than the true one. This can be detected by high positive values of the DF–test statistic. This risk can be circumvented by a general-to-specific testing strategy proposed by Dickey and Pantula [19]. They recommend by starting from the highest sensible order of integration, say $I(2)$, and then test downward to the stationary case.

So far, we have only considered the DF– and the ADF–test as a means to detect the presence of unit roots. Since the early 1980s numerous other statistical tests have been proposed in the literature. The most important and widely used ones will be presented in the second part of the book.

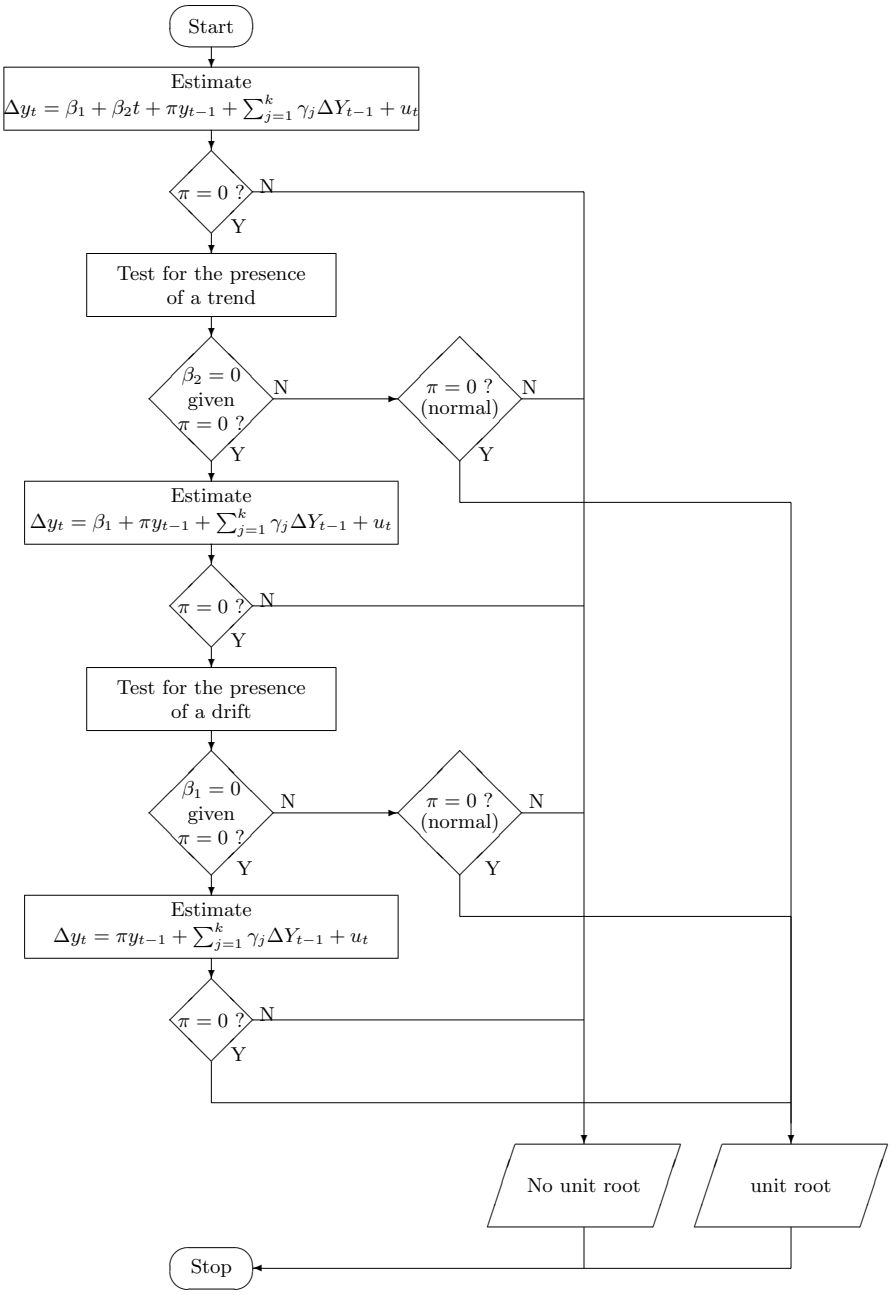

Fig. 2.3. Testing sequence for unit roots

2.3 Long Memory Processes

So far, we have considered data-generating processes that are either stationary or integrated of an integer order higher than zero (for example, the random walk as a prototype of an $I(1)$ series). Hence, it is a knife-edge decision if a series is $I(1)$ or $I(0)$ or is integrated at an even higher integer order. Furthermore, it has been shown that for a $y_t \sim I(1)$ series, the ACF declines linearily, and for a stationary $y_t \sim I(0)$ process, the ACF declines exponentially so that observations separated by a long time span may be regarded as independent. However, some empirically observed time series do not share neither one of these characteristics even though they are transformed to stationarity by suitable differencing. These time series still exhibit a dependency between distant observations. Their occurences are encountered in many disciplines such as finance, geophysical sciences, hydrology, and marcoeconomics. Although having argued heuristically, Granger [34] provides a theoretical justification for these processes. To cope with such time series, our current model class has to be enlarged by so-called *fractionally integrated* processes, *i.e.*, *long memory processes*. The literature about fractionally integrated processes has grown steadily since its detection in the early 1950s of the last century. Baillie [4] cites in his survey about these processes 138 articles and 38 background references.

Before the more encompassing class of *autoregressive fractionally integrated moving average* processes (henceforth: ARFIMA) is introduced, it is noteworthy to define a long memory process and the filter for transforming fractionally integrated series.
First, we draw on the definition of McLeod and Hipel [67].

Definition 2.3. *A process is said to posses a long memory if*

$$\lim_{T\to\infty} \sum_{j=-T}^{T} |\rho_j| \tag{2.17}$$

is nonfinite.

This is equivalent by stating that the spectral density of a long memory process becomes unbounded at low frequencies.[2]
Second, recall that an integrated process of order d can be written as

[2] For an exposition of frequency domain analysis, the reader is referred to Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T. Lee, The Theory and Practice of Econometrics, John Wiley and Sons, New York, 2nd edition, 1985, and Bloomfield, P., Fourier Analysis of Time Series: An Introduction, John Wiley and Sons, New York, 2nd edition, 2000. The spectral density of a series can be estimated by the function `spectrum()` in R. For more information on how this is implemented, the reader is referred to Venables and Ripley [98] as well as the function's documentation.

$$(1-L)^d y_t = \psi(L)\varepsilon_t \ , \tag{2.18}$$

where absolute or square summabilty of ψ_j is given; *i.e.*, $\sum_{j=0}^{\infty} |\psi_j| < \infty$ or $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Premultiplying Equation 2.18 by $(1-L)^{-d}$ yields

$$y_t = (1-L)^{-d}\psi(L)\varepsilon_t \ . \tag{2.19}$$

Now, define the function $f(z) = (1-z)^{-d}$ for the scalar z. The derivatives of this function are

$$\frac{\mathrm{d}f}{\mathrm{d}z} = d(1-z)^{-d-1} \ , \tag{2.20a}$$

$$\frac{\mathrm{d}^2 f}{\mathrm{d}z^2} = (d+1)d(1-z)^{-d-2} \ , \tag{2.20b}$$

$$\vdots$$

$$\frac{\mathrm{d}^j f}{\mathrm{d}z^j} = (d+j-1)(d+j-2)\cdots(d+1)d(1-z)^{-d-j} \ . \tag{2.20c}$$

Therefore, the fractional difference operator for $d \in (-\frac{1}{2}, \frac{1}{2}]$ can be expressed as

$$(1-L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^j \ , \tag{2.21}$$

by making use of a power series expansion around $z = 0$ and the binomial theorem. The coefficient sequence $\binom{d}{j}(-1)^j$ is square summable and can be expressed in terms of the gamma function $\Gamma()$ as

$$\binom{d}{j}(-1)^j = \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} \ . \tag{2.22}$$

Two points are worthy to note. When $d > \frac{1}{2}$, an integer difference operator can be applied first. Incidentally, in this case, the process becomes nonstationary with unbounded variance. Robinson [85] calls such a process "less nonstationary" than a unit root process, smoothly bridging the gulf between $I(0)$– and $I(1)$–processes. Second, because in pratice no series with infinite observations are at hand, one truncates the expression in Equation 2.21 for values y_{t-j} outside the sample range and sets $y_{t-j} = 0$:

$$y_t^* = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} \, y_{t-j} \ , \tag{2.23}$$

where y_t^* assigns the fractional differenced series.

The now to be introduced ARFIMA(p, d, q) class has been developed independently by Granger and Joyeux [36] and Hosking [47]. The estimation and simulation of these models is implemented in `R` within the contributed package `fracdiff`. Formally, an ARFIMA(p, d, q)–model is defined as follows.

Definition 2.4. *The series* $\{y_t\}$ *is an invertible and stationary ARFIMA(p, d, q)–process if it can be written as*

$$\Delta^d y_t = z_t \ , \tag{2.24}$$

where $\{z_t\}_{t=-\infty}^{\infty}$ *is an ARMA(p, q)–process such that* $z_t = \phi_p(L)^{-1}\theta_q(L)\varepsilon_t$ *and both lag polynomials have their roots outside the unit circle, where* ε_t *is a zero-mean i.i.d. random variable with variance* σ^2 *and* $d \in (-0.5, 0.5]$.

For parameter values $0 < d < 0.5$, the process is long memory, and for the range $-0.5 < d < 0$, the sum of absolute values of its autocorrelations tends to a constant. In this case, the process exhibits negative dependency between distant observations and is therefore termed "*antipersistent*" or to have "*intermediate memory*." Regardless, whether the process $\{y_t\}$ is long memory or intermediate memory, as long as $d > -0.5$, it has an invertible moving average representation. How is the long memory behavior incorporated in such a process? It can be shown that the autocorrelation finction (ACF) of long memory processes declines hyperbolically instead of exponentially as would be the case for stable ARMA(p, q)–models. The speed of the decay depends on the parameter value d. For instance, given a fractional white noise process ARFIMA(0, d, 0), Granger and Joyeux [36] and Hosking [47] have proven that the autocorrelations are given by

$$\rho_j = \frac{\Gamma(j+d)\Gamma(1-d)}{\Gamma(j-d+1)\Gamma(d)} \ . \tag{2.25}$$

The counterpart of this behavior in the frequency domain analysis is an unbounded spectral density as the frequency ω tends to zero. In the Rcode example 2.2, an ARIMA(0.4, 0.0, 0.0) and an ARFIMA(0.4, 0.4, 0.0) have been generated and their ACFs as well as spectral densities are displayed in Figure 2.4.

Rcode 2.2 ARMA versus ARFIMA model

```
library(fracdiff)                                                    1
set.seed(123456)                                                     2
# ARFIMA(0.4,0.4,0.0)                                                3
y1 <- fracdiff.sim(n=1000, ar=0.4, ma=0.0, d=0.4)                    4
# ARIMA(0.4,0.0,0.0)                                                 5
y2 <- arima.sim(model=list(ar=0.4), n=1000)                          6
# Graphics                                                           7
op <- par(no.readonly=TRUE)                                          8
layout(matrix(1:6, 3, 2, byrow=FALSE))                               9
plot.ts(y1$series, main='Time series plot of long memory',           10
    ylab='')
acf(y1$series, lag.max=100, main='Autocorrelations of long           11
    memory')
spectrum(y1$series, main='Spectral density of long memory')          12
plot.ts(y2, main='Time series plot of short memory', ylab=''         13
    )
acf(y2, lag.max=100, main='Autocorrelations of short memory'         14
    )
spectrum(y2, main='Spectral density of short memory')                15
par(op)                                                              16
```

A long memory series with 1000 observations has been generated with the function `fracdiff.sim()` contained in the package `fracdiff`, whereas the short memory series has been calculated with the function `arima.sim()` (see command lines 4 and 6).[3] As can be clearly seen in Figure 2.4, the autocorrelations decline much more slowly compared with the stationary AR(1)–model and its spectral density is as $\omega \to 0$ higher about a factor of 100.

By now, the question of how to estimate the fractional difference parameter d or to detect the presence of long memory behavior in a time series is unanswered. We will now present three approaches to do so, whereas the last one deals with the simultaneous estimation of all parameters in an ARFIMA(p, d, q)–model.

The classic approach for detecting the presence of long-term memory can be found in Hurst [48]. He proposed the rescaled range statistic, or shortly the *R/S statistic*. This descriptive measure is defined as

$$R/S = \frac{1}{s_T}\left[\max_{1\le k\le T}\sum_{j=1}^{k}(y_j-\bar{y}) - \min_{1\le k\le T}\sum_{j=1}^{k}(y_j-\bar{y})\right], \tag{2.26}$$

[3] Functions for generating and modeling long memory series can also be found in the contributed CRAN package `fSeries` [102].

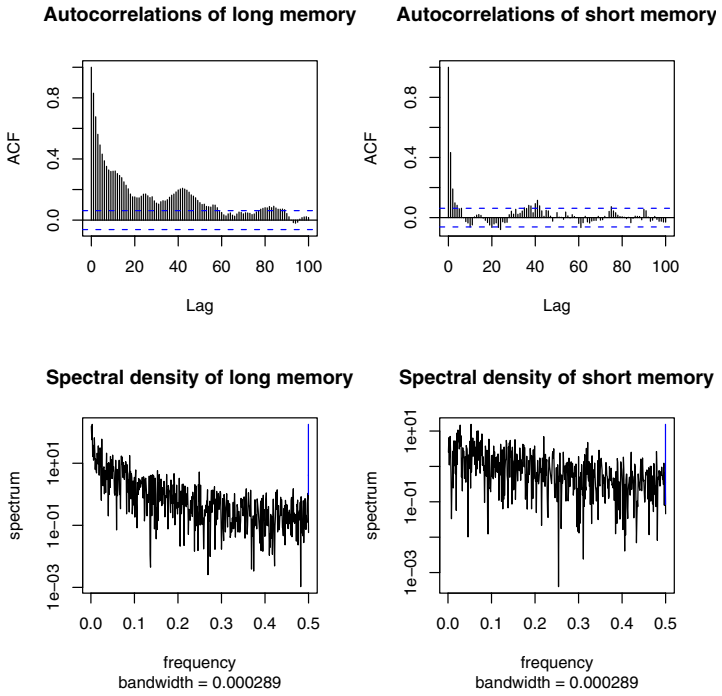

Fig. 2.4. Graphical display: ARIMA versus ARFIMA

where s_T is the usual maximum likelihood standard deviation estimator: $s_T = [\frac{1}{T}\sum_{j=1}^{T}(y_j - \bar{y})^2]^{\frac{1}{2}}$. This measure is always nonnegative because the deviations from the sample mean $\bar{y}$ sum up to zero. Hence, the maximum of the partial sums will always be positive and likewise the minimum will always be negative. Hurst [48] showed the probability limit of

$$\operatorname*{plim}_{T\to\infty}\left\{T^{-H}(\frac{R/S}{s_t})\right\} = \text{const}\ , \tag{2.27}$$

where H assigns the *Hurst coefficient*. The Hurst coefficient is then estimated as

$$\hat{H} = \frac{\log(R/S)}{\log(T)}\ . \tag{2.28}$$

A short memory process is associated with a value of $H = \frac{1}{2}$, and estimated values greater than $\frac{1}{2}$ are taken as hindsight for long memory behavior. Therefore, the differencing parameter d can be estimated as $\hat{d} = \hat{H} - \frac{1}{2}$. The R/S-statistic can fairly easily be calculated in `R`, as shown in Rcode example 2.3.

Rcode 2.3 R/S statistic

```
library(fracdiff)
set.seed(123456)
# ARFIMA(0.0,0.3,0.0)
y <- fracdiff.sim(n=1000, ar=0.0, ma=0.0, d=0.3)
# Demean the series
y.dm <- y$series
max.y <- max(cumsum(y.dm))
min.y <- min(cumsum(y.dm))
sd.y <- sd(y$series)
RS <- (max.y - min.y)/sd.y
H <- log(RS)/log(1000)
d <- H - 0.5
```

Because the default mean in the function `fracdiff` is zero, no demeaning has to be done. The estimated Hurst coefficient is 0.7843, which implies an estimated value for d of 0.2843 that is close to its simulated value of 0.3.

Since the seminal paper of Hurst, the rescaled range statistic has received intensive further research.[4] Although it has been long established that the R/S statistic has the ability to detect long-range dependence, it is however sensitive to short-range dependence and heteroskedasticity.[5] Hence, any incompatibility between the data and the predicted behavior of the R/S statistic under the null hypothesis of no long-run dependence need not come from long-term memory, but it may be merely a symptom of short-term autocorrelation. Lo [58] proposes a modified rescaled range statistic to cope with this deficiency. The modified R/S_{mod} is defined as

$$R/S_{mod} = \frac{1}{s_T(q)} \left[\max_{1 \le k \le T} \sum_{j=1}^{k} (y_j - \bar{y}) - \min_{1 \le k \le T} \sum_{j=1}^{k} (y_j - \bar{y}) \right] , \tag{2.29}$$

where

$$s_T(q) = s_T + 2 \sum_{j=1}^{q} \omega_j(q) \hat{\gamma_j} \ , \omega_j(q) = 1 - \frac{j}{q+1} \text{ with } q < T \ . \tag{2.30}$$

The maximum likehood standard deviation estimator is assigned by s_T and the jth-order sample autocorrelation by $\hat{\gamma}_j$. The sample autocorrelations are

[4] For instance, see Mandelbrot and Wallis [65][66] and Davies and Harte [15] who discuss alternative methods for estimating H. Anis and Loyd [2] determine the small sample bias.

[5] For instance, see Mandelbrot [63][64], Mandelbrot and Wallis [65], Davies and Harte [15], Aydogan and Booth [3], and Lo [58].

weighted by the function $\omega_j(q)$ proposed in Newey and West [70]. However, the choice of an appropriate order q is an unresolved issue.

A popular method for estimating d has been proposed by Geweke and Porter-Hudak [33]. They suggested a semiparametric estimator of d in the frequency domain. They consider as a data-generating process $(1-L)^d y_t = z_t$, where $z_t \sim I(0)$. This process can be represented in the frequency domain

$$f_y(\omega) = 1 - \exp(-i\omega)|^{-2d} f_z(\omega) \ , \tag{2.31}$$

where $f(\omega)_y$ and $f(\omega)_z$ assign the spectral densities of y_t and z_t, respectively. Equation 2.31 can be transformed to

$$\log\{f_y(\omega)\} = \{4\sin^2(\frac{\omega}{2})\}^{-d} + \log\{f_z(\omega)\} \ , \tag{2.32a}$$

$$\log\{f_y(\omega_j)\} = \log\{f_z(0)\} - d\log\{4\sin^2(\frac{\omega_j}{2})\} + \log\{\frac{f_u(\omega_j)}{f_z(0)}\} \ . \tag{2.32b}$$

The test regression is then a regression of the ordinates of the log spectral density on a trigonemetric function of frequencies:

$$\log\{I_y(\omega_j)\} = \beta_1 + \beta_2 \log\{4\sin^2(\frac{\omega_j}{2})\} + \nu_j \ , \tag{2.33}$$

where $\nu_j = \log\{\frac{f_z(\omega_j)}{f_z(0)}\}$ and $j = 1, \ldots, m$. The error term is assumed to be i.i.d. with zero mean and variance $\frac{\pi}{6}$. The estimated order of fractional differencing is equal to $\hat{d} = -\hat{\beta}_2$. Its significance can be tested with either the usual t ratio distributed as Student t, or one can set the residual variance equal to $\frac{\pi}{6}$. An example of this method is Rcode example 2.4, where a fractionally differenced series has been generated first with $d = 0.3$.

Rcode 2.4 Geweke and Porter-Hudak method

```
library(fracdiff)
set.seed(123456)
y <- fracdiff.sim(n=1000, ar=0.0, ma=0.0, d=0.3)
y.spec <- spectrum(y$series, plot=FALSE)
lhs <- log(y.spec$spec)
rhs <- log(4*(sin(y.spec$freq/2))^2)
gph.reg <- lm(lhs ~ rhs)
gph.sum <- summary(gph.reg)
sqrt(gph.sum$cov.unscaled*pi/6)[2,2]
```

The results for the simulated fractionally differenced series are given in Table 2.1. The negative of the estimated coefficient $\hat{\beta}_2$ is 0.2968, which is close to its true value of $d = 0.3$ and higly significant on both accounts, *i.e.*, its

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	−1.6173	0.1144	−14.14	0.0000
rhs	−0.2968	0.0294	−10.11	0.0000

Table 2.1. Results of Geweke and Porter-Hudak method

t ratio as well as the computed standard error with residual variance equal to $\frac{\pi}{6}$. Please note, a major issue with this approach is the selection of the range of frequencies to include in the regression. In Rcode example 2.4, all frequencies have been included, *i.e.*, 500. Diebold and Rudebusch [20] have set $m = \sqrt{T}$, Sowell [93] has suggested setting m to the shortest cycle associated with long-run behavior. A third possibility would be to choose m such that the estimated standard error of the regression is approximately equal to $\sqrt{\pi/6}$.

Finally, the estimation of an ARFIMA(p, d, q)–model is implemented in the contributed package `fracdiff` as function `fracdiff()`. The parameters are estimated by an approximated maximum likelihood using the method of Haslett and Raftery [39]. To lessen the computational burden, a range for the parameter d can be supplied as functional argument. In the case of a "less nonstationary" series, *i.e.*, $d > \frac{1}{2}$, the estimation fails and the series must be integer differenced first. In this case, the fractional differencing filter $(1 - L)^d$ is a combination of Equation 2.21 and integer differencing.

Summary

In this chapter, a more encompassing data-generating process that was introduced into the literature by Campbell and Perron [9] has been presented. You should now be familiar with the concepts of trend- *versus* difference-stationary and the decomposition of a time series into a deterministic trend, a stochastic trend, and a cyclical component. Furthermore, unit root processes have been introduced as a subclass of random walk processes. How one applies a sequential testing strategy to detect the underlying data-generating process of a possible nonstationary time series was discussed. The important definitions of integrated, seasonally integrated, and fractionally integrated time series processes have been presented too, whereas the latter can be viewed as a bridge between stationary and unit root processes, thereby closing the circle of the exposition in the first two chapters.

So far we have adressed univariate time series analysis only. The obstacles and solutions in a multivariate context are the subject of the next and last chapter of Part I.

Exercises

1. Write a function in `R` that returns the critical values given in Fuller [32]. As functional arguments should the test type, the level of significance and the sample size should be supplied.
2. Write a function in `R` that implements the ADF–test regression as shown in Equation 2.14. The series, the inclusion of a constant, trend, both or none, and the order of lagged differenced series should be included as functional arguments. The function should return a summary object of class `lm`.
3. Now include the function of Exercise 1 in the function of 2 such that the relevant critical values are returned beside a summary object of class `lm`.
4. Generate various long and intermediate processes for different values of d and AR(p) and MA(q) orders, and analyze their autocorrelation functions.
5. Write a function that estimates the Hurst coefficient, *i.e.*, the R/S statistic as well as its modified version by Lo [58] and the order of the difference operator d.
6. Write a function for the single equation estimation of d as proposed by Geweke and Porter-Hudak [33].
7. Apply the functions of Exercises 5 and 6 to the absolute logarithmic returns of the stock indices contained in the data set `EuStockMarkets`. Can you detect long memory behavior in any of these series?