

Introduction to Time Series Analysis

Contents:

I. Basics of Time Series Analysis	4
I.1 Stationarity	5
I.2 Autocorrelation Function	9
I.3 Partial Autocorrelation Function (PACF)	14
I.4 Transformation of Data	18
II. Unit Roots	20
II.1 Integration Level	20
II.2 Random Walk	24
II.3 Nonstationarity and Unit Root Tests (Dickey-Fuller and Augmented Dickey-Fuller Test)	31
III.1 White Noise Property	37

Required Texts:

Tsay, R.S. (2002), *Analysis of Financial Time Series*, New York: Wiley.

Enders, W. (2004), *Applied Econometric Time Series*, 2nd ed., New York.: Wiley.

Greene, W. (2003), *Econometric Analysis*, 5th ed., Prentice Hall.

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

Mills, T.C. (1990), *Time Series Techniques for Economists*, New York: Cambridge University Press.

I. Basics of Time Series Analysis

Definition of Time Series

In time series analysis a time series is defined as a realisation of stochastic process where the time index takes on a finite or countable infinite set of values. Denoted, e.g. $\{Y_t \mid \text{for all integers } t\}$.

Time series models are all based on the assumption that the series to be forecasted has been generated by a stochastic process. Therefore, we assume that each observed value $Y_1, Y_2, \dots, Y_T$ in the series is drawn randomly from a probability distribution.

A stochastic process exhibits a random process, denoted as $\{Y_t\}$, which can take a value between $-\infty$ and $+\infty$. The observed value Y_t at time t describes one realisation of these stochastic processes.

I.1 Stationarity

Strict Stationary:

Joint distribution: $Y(t) = \{Y(1), Y(2), \dots, Y(T)\}$

→ invariant under time shift

The random variables $Y(t+1), \dots, Y(t+n)$ have the same joint distribution as $Y(t+1+c), \dots, Y(t+n+c)$, with c as a arbitrary positive integer. This is a very strong condition that is hard to verify empirically.

Weak Stationarity:

Weak stationarity exists, when expected value, variance and covariance of the distribution random variables are constant for all points of time.

- a) $E(Y_t) = \mu = \text{constant}, \forall t$, (mean stationarity)
- b) $\text{Var}(Y_t) = \sigma_t^2 = \sigma^2 = \text{constant}, \forall t$ (variance stationarity), and
- c) $\text{Cov}(Y_t, Y_{t-j}) = \sigma_{tj} = \sigma_j = \text{constant}, \forall t$ (covariance stationarity),
 $E[(y_t - \mu)(y_{t-j} - \mu)] = E[(y_s - \mu)(y_{s-j} - \mu)]$ for all $t \neq s$.

The data of the underlying process are time invariant and neither the shape nor the parameters of the distribution change over time. The covariance only depends on j , where j is an arbitrary integer.

Suppose that we have observed T data points $\{Y_t \mid t = 1, \dots, T\}$:

- weak stationarity implies that the time plot of the data would show that the T values fluctuate with constant variation around a constant level.
- we assume that the first two moments of Y_t are finite
- from definitions, if Y_t is strictly stationary and its first two moments are finite, then Y_t is also weakly stationary, but the converse is not true in general
- however, if the time series Y_t is normally distributed, then weak stationarity is equivalent to strict stationarity.

The covariance $\gamma_j = \text{Cov}(Y_t, Y_{t-j})$ is called the lag- j autocovariance of Y_t with the following properties:

(a) $\gamma_0 = \text{Var}(Y_t)$ and

(b) $\gamma_{-j} = \gamma_j$.

$$\text{Cov}(Y_t, Y_{t-(j)}) = \text{Cov}(Y_{t-(j)}, Y_t) = \text{Cov}(Y_{t+j}, Y_t) = \text{Cov}(Y_{t1}, Y_{t1-j}), \text{ where } t1 = t + j.$$

The statistical ratios to describe weakly stationary processes are

- a) the autocovariance function for the direction of interrelation,
- b) the autocorrelation function (ACF) for the strength and direction of interrelation, and
- c) the partial autocorrelation function (PACF) which measure the effect of one specific lag j holding all $t-j-1$ lags in between constant.

I.2 Autocorrelation Function

Consider a weakly stationary series Y_t . When linear dependence between Y_t and its past values Y_{t-j} is of interest, the concept of correlation is generalized to autocorrelation. The correlation coefficient between Y_t and Y_{t-j} is called the lag- j autocorrelation of Y_t and is commonly denoted by ρ_j , which under the weak stationarity assumption is a function of j only. Specifically, we define

$$\rho_j = \frac{\text{Cov}(Y_t, Y_{t-j})}{\sqrt{\text{Var}(Y_t) \cdot \text{Var}(Y_{t-j})}} = \frac{\text{Cov}(Y_t, Y_{t-j})}{\text{Var}(Y_t)} = \frac{\gamma_j}{\gamma_0},$$

where the property $\text{Var}(Y_t) = \text{Var}(Y_{t-j})$ for a weakly stationary series is used. From the definition, we have $\rho_0 = 1$, $\rho_j = \rho_{-j}$, and $-1 \leq \rho_j \leq 1$. In addition, a weakly stationary series Y_t is not serially correlated if and only if $\rho_j = 0$ for all $j > 0$.

For a given sample of returns $\{r_t\}_{t=1}^T$, let $\bar{r}$ be the sample mean, i.e. $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$

Then the first-order autocorrelation coefficient of r_t is

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T (r_t - \bar{r})(r_{t-1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2} = \frac{\sum_{t=1}^{T-1} (r_t - \bar{r})(r_{t+1} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}$$

The autocorrelation coefficients of random data have a sampling distribution which is approximately normally distributed with a mean of zero and a standard deviation of $\frac{1}{\sqrt{N}}$. Under certain assumptions, $C_{95\%} = \pm 2 \cdot \frac{1}{\sqrt{N}}$ can be viewed as a 95% confidence interval, and empirical autocorrelation coefficients which lay outside of this interval, are significantly different from zero.

Box and Pierce (1970) propose the Portmanteau statistic:

$$Q(m) = T \cdot \sum_{j=1}^m \hat{\rho}_j^2 \sim \chi^2(m)$$

$$H_0: \rho_1 = \dots = \rho_m = 0$$

$$H_1: \rho_1 \neq 0 \quad \text{for } i \in \{1, \dots, m\}.$$

Ljung-Box (1978) modified the $Q(m)$ statistic to increase the power of the test in finite samples:

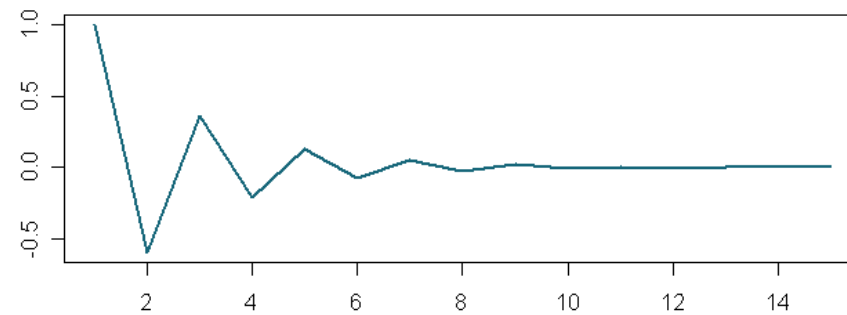
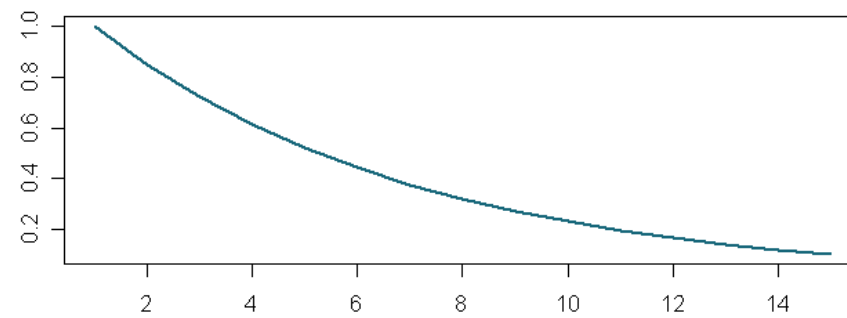
$$Q^*(m) = T(T+2) \cdot \sum_{j=1}^m \frac{\hat{\rho}_j^2}{T-j} \sim \chi^2(m)$$

→ number of autocorrelation coefficients should be around 20 % of the sample size

Example: Correlogram for: $Y_t = 0.85Y_{t-1} + \varepsilon_t$; $\varepsilon_t \sim N(0,1)$ (upper graph) and

$Y_t = -0.6Y_{t-1} + \varepsilon_t$; $\varepsilon_t \sim N(0,1)$ (lower graph), where $E(\varepsilon_t \cdot Y_{t-1}) = 0$

- if α is positive the autocorrelation function decreases exponentially
- if α is negative the ACF oscillates around zero. For both cases the ACF dies out over time



I.3 Partial Autocorrelation Function (PACF)

Assume an equation appear several variables, for example in the following model:

$$\text{AR}(1): \quad Y_t = \phi_{0,1} + \phi_{1,1}Y_{t-1} + u_{1t}$$

$$\text{AR}(2): \quad Y_t = \phi_{0,2} + \phi_{1,2}Y_{t-1} + \phi_{2,2}Y_{t-2} + u_{2t}$$

$$\text{AR}(3): \quad Y_t = \phi_{0,3} + \phi_{1,3}Y_{t-1} + \phi_{2,3}Y_{t-2} + \phi_{3,3}Y_{t-3} + u_{3t}$$

...

$$\text{AR}(p): \quad Y_t = \phi_{0,p} + \phi_{1,p}Y_{t-1} + \phi_{2,p}Y_{t-2} + \phi_{3,p}Y_{t-3} + \dots + \phi_{p,p}Y_{t-p} + u_{pt}$$

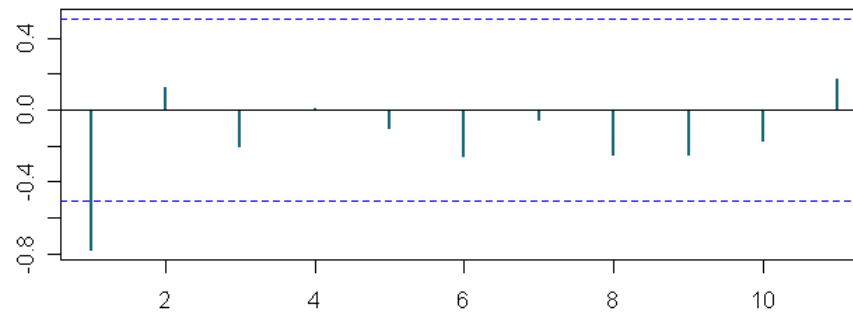
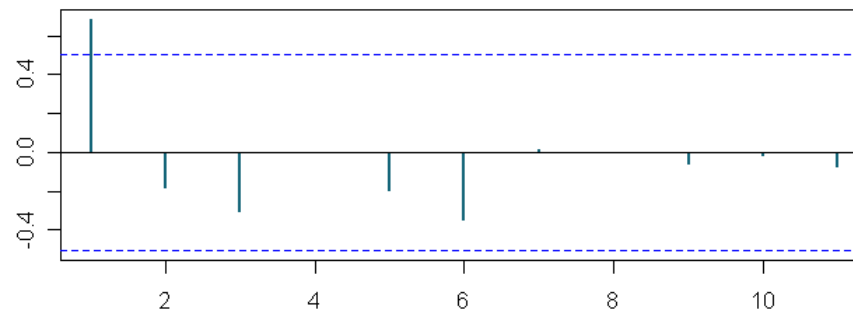
- Partial Autocorrelation Function (PACF), ϕ_τ : similar to the normal ACF, the PACF measures the correlation between Y_t and $Y_{t-\tau}$, but the lags in between $(t-j+1, t-j+2, \dots, t-1)$ are held constant so that the contribution of the individual lag can be assessed.
- The PACF for the same period is the variance: $\phi_0 = PACF(Y_t, Y_t) = 1$
- The PACF for the first lag is equal to the ACF since no lags in between exist: $\phi_1 = \rho_1$
- Generally the PACF can be described by:

$$\phi_{\tau} = \begin{cases} \rho_1 & \text{for } \tau = 1 \\ \frac{\rho_{\tau} - \sum_{j=1}^{\tau-1} \phi_{\tau-1,j} \cdot \rho_{\tau-j}}{1 - \sum_{j=1}^{\tau-1} \phi_{\tau-1,j} \cdot \rho_{\tau-j}} & \text{for } \tau > 1 \end{cases}$$

Example: PACF for: $Y_t = 0.85Y_{t-1} + \varepsilon_t$; $\varepsilon_t \sim N(0,1)$ (upper graph) and

$Y_t = -0.6Y_{t-1} + \varepsilon_t$; $\varepsilon_t \sim N(0,1)$ (lower graph)

- In contrast to the ACF which has the die-out property, the PACF is cut off after lag 1. This combination of properties for the ACF and PACF helps us to identify an AR process by looking at the correlogram.



I.4 Transformation of Data

Differencing to eliminate Trends (Mean-Stationarity)

Basically, a non-constant mean of a time series arises from two different characters:

- a) structural breaks with erratic changes of the means
- b) continuous increase or decrease of the mean over time

Difference operator Δ

If Y_t is the original series then it follows

$$\Delta Y_t = Y_t - Y_{t-1}$$

the first differences of the time series Y_t .

If a time series must be differenced twice we formulate

$$\Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = \Delta Y_t - \Delta Y_{t-1} = Y_t - Y_{t-1} - Y_{t-1} + Y_{t-2} = Y_t - 2Y_{t-1} + Y_{t-2}$$

Consequently, a twice differencing corresponds to a filter which is applied to a series with the weights of the filter (1,-2,1). If a time series is differenced d of times we can write $\Delta^d Y_t$.

II. Unit Roots

II.1 Integration Level

A time series is called integrated of order d ($= I(d)$), if after d differencing the series follows a stable and invertible ARMA process and thus an $I(0)$ process. The property of $I(0)$ implies stationarity whereas the reverse does not hold.

To highlight this issue we have to remind briefly, the properties of stability and invertibility of an ARMA process. In the simplest case of an ARMA(1,1) process without a constant and with a white noise error term ε we obtain:

$$Y_t = \gamma Y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t.$$

This process is stationary and invertible, if γ and θ in absolute values is lesser than 1.

Stability implies that we receive a MA term with coefficients tending against zero by repeating the substitution of the lagged Y variable.

$$Y_t = \varepsilon_t + \sum_{i=1}^{\infty} (\gamma + \theta) \gamma^i \varepsilon_{t-i}$$

Y is non-stationary if $\gamma = 1$ and we get a typical situation of a random walk where the error term exhibits a permanent influence on the time series. Conversely, by repeating substitution of lagged error terms ε , we obtain an AR representation with against zero tending coefficients:

$$Y_t = \sum_{i=1}^{\infty} (\gamma + \theta)(-\theta)^i Y_{t-i} + \varepsilon_t$$

Thus it appears that if $\theta = -1$ the coefficients of the AR representation independent of the lag are always $(\theta + \gamma)$ and therefore the time series cannot be approximated by an AR process.

The difference of Y is under the assumption $|\theta| < 1$ a $I(0)$ series. But differencing does not always result in a $I(0)$ time series. For example, let us consider a trend stationary (or for $a = 0$ stationary) process:

$$Y_t = c + at + \varepsilon_t.$$

By differencing we obtain:

$$Y_t - Y_{t-1} = \Delta Y_t = a + \varepsilon_t - \varepsilon_{t-1}.$$

For this reason the trend of the time series is indeed eliminated and the series is stationary with mean a . But, at the same time a noninvertible MA(1) process was generated with $\theta = -1$ so that we obtain no $I(0)$ series. In this context we speak about overdifferencing.

Please note that:

- **trend stationary time series** are not mean stationary but include a trend. This trend can be eliminated by including a trend component into the regression model

$$Y_t = a + bt + \beta X_t + \varepsilon_t.$$

As we can see above differencing is not appropriate to eliminate trends because the variance of the error term would increase.

- **difference stationary time series** (which are most of economic time series) contain a stochastic trend, i.e. a non stationarity in the variance component so that with the length of forecasting horizon the uncertainty increases to endless.

In this context only differencing results in a stationary time series.

For stationarity of the error terms of the estimation equation $Y_t = a + \beta X_t + \varepsilon_t$ the following rules are observed:

$$Y_t \sim I(0) \text{ and } X_t \sim I(0) \quad \Rightarrow \varepsilon_t \sim I(0),$$

$$Y_t \sim I(1) \text{ and } X_t \sim I(0) \quad \Rightarrow \varepsilon_t \sim I(1),$$

$$Y_t \sim I(1) \text{ and } X_t \sim I(1) \quad \Rightarrow \varepsilon_t \sim I(1), \text{ if } Y \text{ and } X \text{ are not cointegrated,}$$

$$Y_t \sim I(1) \text{ and } X_t \sim I(1) \quad \Rightarrow \varepsilon_t \sim I(0), \text{ if } Y \text{ and } X \text{ are cointegrated.}$$

The residuals are only then $I(0)$ if both variables Y and X either are $I(0)$ or $I(1)$ and cointegrated. The simplest case of cointegration is given when Y and X are $I(1)$ and the linear combination of both variables is $I(0)$, i.e. the residuals are stationary.

II.2 Random Walk

A simple example for a stochastic (non-stationary) time series is a random walk:

$$Y_t = Y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma)$$

with a white noise error term ε_t and the properties as follows:

the random walk (RW1) is also a fair game but in contrast to the martingale also implies not only zero autocovariance, $Cov(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$, but also any non-linear transformation, e.g. $Cov(\varepsilon_t^2, \varepsilon_s^2) = 0$ for $t \neq s$ is uncorrelated.

$$E[Y_{t+n}] = Y_t + \sum_{i=1}^n E[\varepsilon_{t-i}] = Y_t$$

process is *mean stationary*

$$\text{Var}(Y_{t+n}) = \sum_{i=1}^n \text{Var}(\varepsilon_{t-i}) = \sum_{i=1}^n \sigma^2 = n \cdot \sigma^2$$

process is *not variance stationary*

(the variance changes over time)

$$\text{Cov}(Y_{t+n}, Y_{t+n-s}) = \sum_{i=1}^n \sigma^2 (t+n-s)$$

process is *not covariance stationary*

ε_t is a special stationary stochastic process where the autocovariance equals zero.

These properties of a distribution of ε_t are called white noise.

Y is also not stationary but the first difference of Y is a stationary random variable:

$$Y_t - Y_{t-1} = \Delta Y_t = \varepsilon_t$$

Time series that follow a random walk are integrated of order one $I(1)$. By adding a constant to the random walk equation, we obtain a **random walk with drift**:

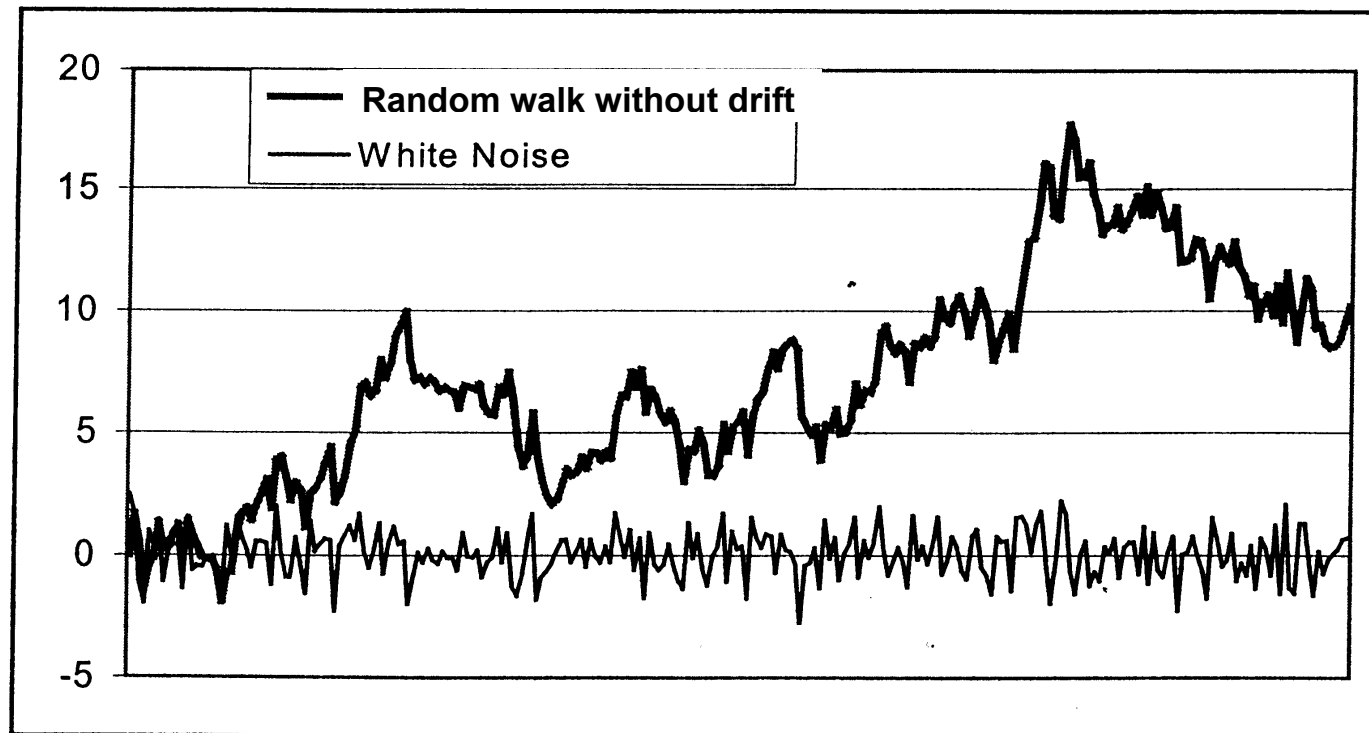
$$Y_t = a + Y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma)$$

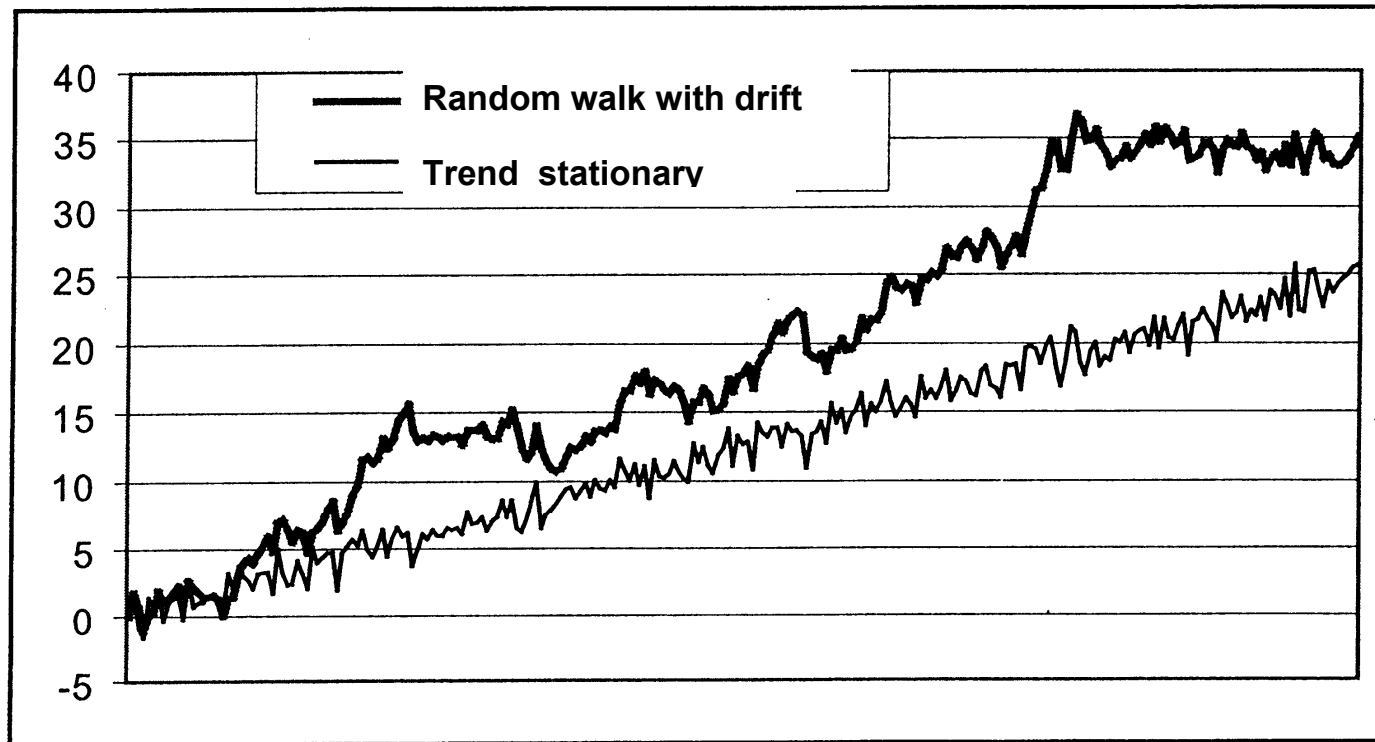
If the constant is positive then Y exhibits an upward tendency and when the constant is negative, Y has a downward tendency. The first and second moments of a random walk with drift are:

$$E(Y_{t+n}) = Y_t + n \cdot a + \sum_{i=1}^n E(\varepsilon_{t-i}) = Y_t + n \cdot a \quad \Rightarrow \text{process is *not mean stationary*}$$

$$Var(Y_{t+n}) = \sum_{i=1}^n Var(\varepsilon_{t-i}) = \sum \sigma^2 = n \cdot \sigma^2 \quad \Rightarrow \text{process is *not variance stationary* (the variance increases linearly over time)}$$

Random Walk without Drift





Accordingly, a random walk with drift defines a non-stationary random variable. The variance is infinite as n gets very large (increases linearly with slope a). The first difference yields,

$$Y_t - Y_{t-1} = a + \varepsilon_t,$$

which is a stationary sequence. In comparison to a trend stationary process,

$$Y_t = c + at + \varepsilon_t \text{ (with } \varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)\text{)},$$

which changes deterministically in time, a random walk with drift exhibits a stochastic trend. For both processes the expected change over t periods are indeed the same ($t \cdot a$). Whereas for a random walk with drift the first difference is stationary, for a trend stationary process the deviations from the trend ($Y_t - at = c + \varepsilon_t$) are stationary.

II.3 Nonstationarity and Unit Root Tests (Dickey-Fuller and Augmented Dickey-Fuller Test)

The most frequently used test of the null hypothesis of a $I(1)$ series against a $I(0)$ alternative hypothesis is the Dickey-Fuller t -test. First, let us consider this test in the simplest AR(1) case. The initial equation is:

$$Y_t = \gamma Y_{t-1} + \varepsilon_t$$

$H_0: \gamma = 1 \Rightarrow$ random walk without drift

$H_1: \gamma < 1 \Rightarrow$ stationary AR(1) process

This is a one-sided test with the null hypothesis of non-stationarity and the alternative hypothesis of stationarity. For an AR(1) process the t -statistic of the OLS estimation is biased, i.e. it does not follow

a t -distribution under the null hypothesis. Hence, Dickey and Fuller derive the distribution of this test statistic and determine the critical values.

If we want to test a random walk with drift against the alternative of a stationary AR(1) process with a mean that does not equal zero, we have to insert a constant term into the regression equation:

$$Y_t = \mu + \gamma Y_{t-1} + \varepsilon_t$$

$H_0: \gamma = 1 \Rightarrow$ random walk with drift

$H_1: \gamma < 1 \Rightarrow$ stationary AR(1) process with mean $\mu \neq 0$

Finally, for the alternative hypothesis of a trend stationary process we insert an additional trend variable into the regression equation:

$$Y_t = \mu + \beta t + \gamma Y_{t-1} + \varepsilon_t$$

$H_0: \gamma = 1 \Rightarrow$ random walk with drift and trend

$H_1: \gamma < 1 \Rightarrow$ trend stationary

The Dickey-Fuller test, however, may be biased if lagged differences $\Delta Y_{t-2}, \Delta Y_{t-3}, \dots, \Delta Y_{t-p-1}$ effect ΔY_t . For this, we have to expand the test equation by subtracting the lagged variable on both sides.

Subsequently, we complement the equation by $p-1$ lagged differences:

$$Y_t - Y_{t-1} = \Delta Y_t = \mu + \beta t + \gamma^* Y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \varepsilon_t \quad \text{with} \quad \gamma^* = \gamma - 1$$

$H_0: \gamma^* = 0 \Rightarrow$ stationary AR(p-1) process in differences, i.e. nonstationary in levels

$H_1: \gamma^* < 0 \Rightarrow$ trend stationary AR(p) process

For this so-called **Augmented Dickey-Fuller test**, the same critical values are valid as for the simple model of this test. This means, if we do not consider any lag we obtain the simple Dickey-Fuller test.

Let us turn to an AR(p) process:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \dots + a_{p-2} y_{t-p+2} + a_{p-1} y_{t-p+1} + a_p y_{t-p} + \varepsilon_t$$

add and subtract: $a_p y_{t-p+1}$

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \dots + a_{p-2} y_{t-p+2} + (a_{p-1} + a_p) y_{t-p+1} - a_p \Delta y_{t-p+1} + \varepsilon_t$$

add and subtract: $(a_{p-1} + a_p) y_{t-p+2}$

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \dots + (a_{p-1} + a_p) \Delta y_{t-p+2} - a_p \Delta y_{t-p+1} + \varepsilon_t$$

We can take the difference and receive:

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t$$

$$\gamma = -\left(1 - \sum_{i=1}^p a_i\right)$$

$$\beta_i = \sum_{j=1}^p a_j$$

$\gamma = 0$ in the first difference has a unit root when $\sum a_i = 1$.

Sequential test procedure

1. Starting with a relatively high number of 10 lags,
2. Subsequently, reduce the number of lags until the last coefficient is significantly different from zero on the 10 % level.
3. Compare the three different models (without drift and trend, with drift, and with drift and trend) by looking at the Akaike criterion. Then choose the model with the lowest Akaike criterion.
4. If the value of the test statistic is bigger (or in absolute values smaller) than the critical value, you cannot reject the $I(1)$ null hypothesis on conventional significance levels.

III.1 White Noise Property

A time series is called white noise if $\{Y_t\}$ is a sequence of independently and identically distributed random variables with finite mean and variance. That means $Y_t = u_t$ describes a pure random process with:

$$E(u_t) = 0$$

$$\text{Var}(u_t) = \sigma^2$$

$$E(u_t u_s) = 0 \quad \forall t \neq s$$

and in addition u_t is normal distributed. This assumption is expressed by writing:

$$u_t \sim \text{i.i.d. } N(0, \sigma^2),$$

where i.i.d. stands for independently and identically distributed random variables. Time series processes with such properties are called white noise, whereas the assumption of normality is not obligatory. If normality is existent the process is refer to as Gaussian white noise.