STATISTICAL METHODS WITH APPLICATION TO FINANCE

a.y. 2025-2026

Introduction to time series analysis

R. Pappadà

rpappada@units.it

Department of Economics, Business, Mathematics and Statistics "B. de Finetti"

University of Trieste

March 17, 2026

Table of Contents

Time series

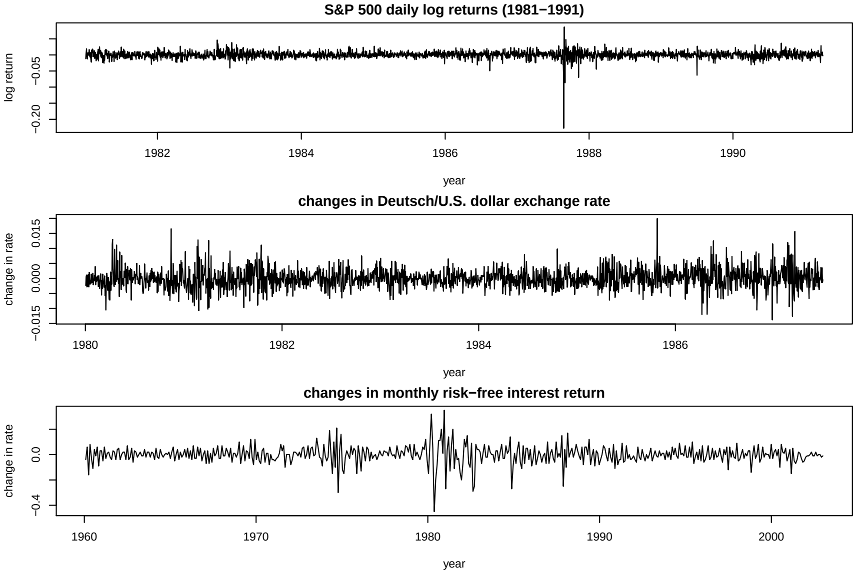
Let $\{x_t\}$ be a collection of certain financial measurements over time. The value x_t is observed at (roughly) equally spaced time intervals.

A **time series** can be defined as the sequence of such observations over time (e.g. prices, interest rates, log returns). The plot of the observations against time is the time series plot.

The *time series plot* allows to investigate whether there is

- a trend or a seasonal behavior
- a stationary behavior (the properties of one section of the data are much like those of any other section)
- any outlying observations

Examples: Visualization of Financial Data



Objectives

Statistical techniques for analysing time series range from descriptive methods to sophisticated models

The objectives of time series analysis depend on the particular field of application (economics, finance, medicine, social sciences, hydrology, etc.)

- A **time series model** is used simply to provide a compact and plausible description of the data, their main features and the random variation
- prediction of future values of a series is an important task in time series analysis

Types of variation

Stationary models assume some form of distributional invariance over time.

However, many time series data exhibit:

- a *trend*: a 'long-term change in the mean level'
- *cyclic variations*: e.g. seasonal patterns such as a repeating behaviour within each year (e.g., unemployment is typically 'high' in winter but lower in summer) or daily variation in temperature

Therefore we often work with **transformations** of the variables such as logs, square roots, or other power transformations.

Intuition of stationarity

A time series is said to be **stationary** if there is no systematic change in mean (no trend) and in variance, and no strictly periodic variations

Stationarity is a property defined for a time series model:

if the data exhibit features that indicate stationarity (e.g. the series shows random oscillation around some fixed level, a phenomenon called *mean-reversion*), then it is reasonable to assume a stationary model to describe the data

Many financial time series do not exhibit stationarity, but often the *changes* in them, perhaps after applying a log transformation, are approximately stationary.

Table of Contents

Transformations

The main reasons for making a transformation are as follows.

- To stabilize the variability of the series. If there is a *trend* in the series and the **variance appears to increase** with the mean, then a *logarithmic transformation* may be appropriate
- To attain stationarity. Differencing is a commonly used method to induce stationarity in time series analysis. For non-seasonal data, **first-order differencing** is usually sufficient.

  Example: It is common to take the log price series $x_t = \log(P_t)$, and obtain its first differenced series

  $$y_t = \log(P_t) - \log(P_{t-1}) = x_t - x_{t-1}$$

  y_t is the log return series, which is then a relatively stable process

Transformations / 2

Seasonal differencing can be used to remove seasonal variation. In general, for a time series x_t with periodicity s, seasonal differencing means

$$\nabla_s x_t = x_t - x_{t-s}$$

- $\rightarrow$ For the quarterly earnings the periodicity of the series is 4
- $\rightarrow$ If monthly data are considered then the periodicity is 12

- To make the data normally distributed. Many statistical methods work best when the data are normally distributed or at least symmetrically distributed. In some cases, a *Box-Cox power transformation* may be helpful

$$y_t = (x_t^\lambda - 1)/\lambda \quad \lambda \neq 0$$

- $\rightarrow$ A Box-Cox transformation with $\lambda < 1$ can remove right-skewness and stabilize the variance
- $\rightarrow$ For left-skewed data, a Box-Cox transformation with $\lambda > 1$ should be considered

Example 1

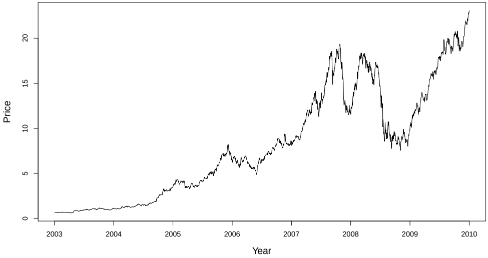


Figure 1: Daily closing prices of Apple stock from January 3, 2003 to April 5, 2010. The series exhibits certain degrees of variability and show an upward movement during the sample period.

Example 1 (cont)

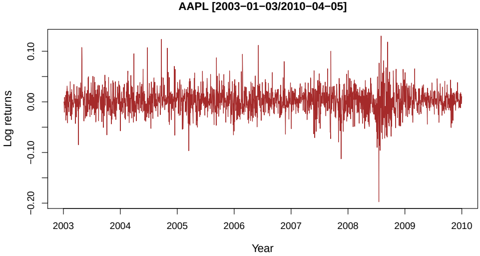


Figure 2: Daily log returns of Apple stock prices in Figure 1.

- The mean of the returns is constant over time and close to 0
- Extreme volatility indicates non-normality of data
- Periods of persistent low and high volatility arise when the variance is time-varying

Example

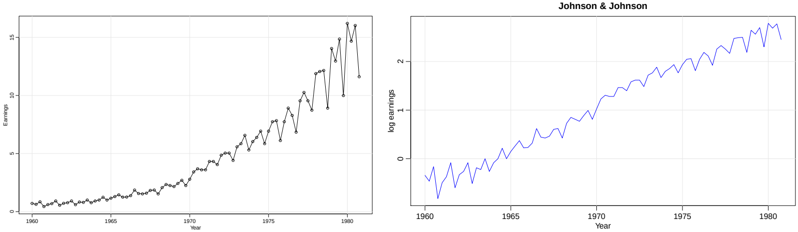


Figure 3: Time plot of quarterly earnings and log earnings per share of the Johnson & Johnson data.

Table of Contents

Stochastic processes

Let us now suppose that $y_1, y_2, \ldots, y_n$ is any observed discrete time series, so that y_t is a single number recorded at time t and observations are available for n consecutive times. We assume y_t to be the realized value of some random variable Y_t.

The sequence of random variables $Y_1, Y_2, \ldots$ is called a stochastic process, usually denoted by $\{Y_t\}$.

Remark We shall sometimes use the term *time series* to mean both the data and the process of which it is a realization.

Time series model

We are interested in the joint probability distribution of the r.v.s that form the process $\{Y_t\}$,

$$F(y_1, y_2, \ldots, y_n) = Pr(Y_1 \leq y_1, Y_2 \leq y_2, \ldots, Y_n \leq y_n)$$

$\rightarrow$ often very difficult!

Hence we could look at the *moments* of the process $\{Y_t\}$.

We will denote with $\{y_t\}$ a realization of the r.v. Y_t, observed at a time $t = 1, \ldots, n$, if time is discrete.

Table of Contents

The mean function

Assuming they exist, we define the *moments* of a time series, i.e. the **mean function** and **autocovariance function**. The **variance function** is a special case of the autocovariance function.
Such properties of $\{Y_t\}$ are referred to as the *second-order properties* of the process.

The mean function of a time series model is an average taken across the ensemble, that is the infinite set of time series that might have been observed. It is defined for all t by

$$\mu_t = \mathsf{E}(Y_t)$$

and, in general, is a function of t.

The ensemble

The set of all the possible time series that might have been produced by a time series model is called the **ensemble**. It constitutes the entire population.

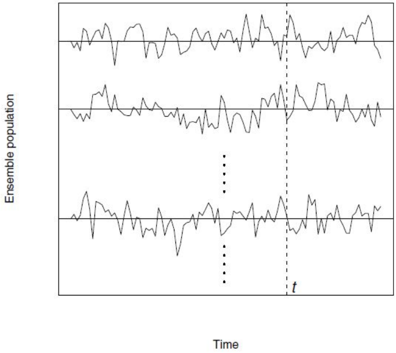


Figure 4: An ensemble of time series.

The variance and autocovariance function

The variance function of a time series model is defined for all t by

$$\sigma_t^2 = \mathsf{Var}(Y_t) = \mathsf{E}[(Y_t - \mu_t)^2]$$

The autocovariance function is the covariance of Y_t and Y_s

$$\gamma(t,s) := \mathsf{Cov}(Y_t, Y_s) = \mathsf{E}[(Y_t - \mu_t)(Y_s - \mu_s)]$$

which is a measure of linear association between Y_t and Y_s.

- $\gamma(t,t) = \sigma_t^2$
- $\gamma(t,s) = \gamma(s,t)$.

Table of Contents

Definition: Weak Stationarity

A process is called **weakly stationary** if its mean, variance, and covariance are unchanged by time shifts.

More precisely, $\{Y_t\}$ is a weakly stationary process (or second-order stationary) if the first two moments exist and satisfy

$$\mathsf{E}(Y_t) = \mu$$

and, for all t

$$\mathsf{Cov}(Y_t, Y_{t+k}) = \mathsf{E}[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma_k$$

where k is called the **lag**. Thus the mean is constant and the autocovariance function depends only on the *time lag*. The definition also implies that both the mean and the variance must be finite. Note that

$$\gamma_0 = \mathsf{Cov}(Y_t, Y_t) = \mathsf{Var}(Y_t)$$

also denoted by σ^2.

Weak Stationarity: Remarks

- Weak stationarity is important because it provides the basic framework for prediction
- Weakly stationarity places no requirements on moments higher than second-order, compared to strict stationarity, that requires that the joint cdf of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_k})$ is the same as the joint cdf of $(Y_{t_1+\tau}, Y_{t_2+\tau}, \ldots, Y_{t_k+\tau})$, for all $t_1, \ldots, t_k, \tau$.
- Many of the properties of stationary models depend only on the structure of the process as specified by its first and second moments
- Second order stationarity implies strict stationarity for normal processes

Estimating the mean and the variance

If the mean function is constant, we say that the time series model is *stationary in the mean*.

In such case, all Y_t have a common expectation, μ, that can be estimated by the sample mean

$$\bar{y} = \sum_{t=1}^{n} y_t / n$$

If we assume that the population variance is constant over time, then σ^2 can be estimated from the sample variance:

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} (y_t - \bar{y})^2}{n-1}$$

Table of Contents

Autocorrelation

In the study of time series models, a key role is played by serial correlation.

Consider a time series model that is stationary in the mean and the variance. The variables related to different times may exhibit some correlation, and the model is weakly stationary if the correlation between variables depends only on the number of time steps–*lags*–separating them.

A correlation of a variable with itself at different times is known as *autocorrelation* or *serial correlation*.

Autocorrelation

The **lag-k autocorrelation function** (ACF), defined by

$$\rho_k = \frac{\text{Cov}(Y_t, Y_{t+k})}{\sqrt{\text{Var}(Y_t)\text{Var}(Y_{t+k})}} \tag{1}$$

for $k = 0, 1, 2, \ldots$.
Due to stationarity we can write $\text{Var}(Y_t) = \text{Var}(Y_{t+k})$ and express the autocovariance as a function of the *lag* k. Therefore Eq.(1) becomes

$$\rho_k = \frac{\text{Cov}(Y_t, Y_{t+k})}{\text{Var}(Y_t)} = \frac{\gamma_k}{\gamma_0} \tag{2}$$

A weakly stationary process $\{Y_t\}$ is serially uncorrelated if and only if $\rho_k = 0$ for all $k > 0$.

Properties of ACF

For a given k, ρ_k measures the extent to which Y_t and Y_{t+k} are linearly associated, in analogy with the standard correlation coefficient. If the process has mean μ and variance σ^2, then

$$\rho_k = \frac{\gamma_k}{\sigma^2}$$

From the definition, we have $\rho_0 = 1$.

- $\rho_k = \rho_{-k}$: the correlation between Y_t and Y_{t+k} is the same as that between Y_t and Y_{t-k}
- $-1 \leq \rho_k \leq 1$ ('usual' property of correlation)
- The ACF does not uniquely identify the underlying model

Table of Contents

Definition

The model $\{W_t\}$ is called a **weak white noise process** if it is a sequence of uncorrelated random variables, with zero mean and finite variance σ^2:

$$W_t \sim \text{WN}(0, \sigma^2)$$

- $\text{E}(W_t) = 0$
- $\text{Var}(W_t) = \sigma^2$ (finite)
- $\text{Cov}(W_t, W_s) = 0$, for all $t \neq s$

These are the simplest processes that are the building blocks for more complicated models.

Residuals

When we fit mathematical models to time series data, we refer to the discrepancies between the fitted values, calculated from the model, and the data as a **residual error series**

$$a_t = y_t - \hat{y}_t$$

where $\hat{y}_t$ is the value predicted by the model.

If a model has accounted for all the serial correlation in the data, the residual series would be serially uncorrelated, that is, the residuals should behave as a white noise series.

Iid White Noise

A stronger assumption is that the random variables in the process are *mutually independent*. Recall that

$$X, Y \text{ independent} \underset{\not\Leftarrow}{\Rightarrow} X, Y \text{ uncorrelated}$$

If $W_1, W_2, \ldots$ are independent and identically distributed with mean 0 and variance σ^2, then we refer to the sequence $\{W_t\}$ as an **iid white noise process**

$$W_t \sim iid\, \mathrm{WN}(0, \sigma^2)$$

This process is sometimes called a *purely random process*.

Remark. An iid white noise is **strictly stationary** in that shifting the time by k units has no effect on the joint CDF. An iid white noise process is also a weak white noise process, but not vice versa.

Gaussian White Noise

If the variables also follow a Normal distribution the process is called a **Gaussian white noise**

$$W_t \sim \text{iid } \mathcal{N}(0, \sigma^2)$$

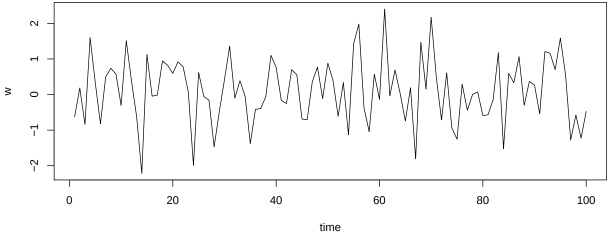


Figure 5: Time plot of simulated Gaussian white noise series.

Properties

A white noise process $w_t \sim \text{WN}(0, \sigma^2)$ is weakly stationary:

$$\mu_w = 0,$$

(by definition, the w_t are uncorrelated)

$$\gamma_k = \text{Cov}(w_t, w_{t+k}) = \begin{cases} 0 & \text{if } k \neq 0 \\ \text{Var}(w_t) = \sigma^2 & \text{if } k = 0 \end{cases}$$

The autocorrelation function follows as

$$\rho_k = \frac{\gamma_k}{\sigma^2} = \begin{cases} 0/\sigma^2 = 0 & \text{if } k \neq 0 \\ \sigma^2/\sigma^2 = 1 & \text{if } k = 0 \end{cases}$$

Therefore, the autocorrelation coefficients of a white noise are null at all lags except lag 0.

Table of Contents

Correlation between adjacent observations

Suppose we have n observations on a stationary process: $(y_1, y_2, \ldots, y_n)$.
We can form $n-1$ pairs of observations

$$\begin{gathered}(y_1, y_2)\\(y_2, y_3)\\\vdots\\(y_{n-1}, y_n)\end{gathered}$$

each separated by one time interval (*lag*).
The **sample autocorrelation coefficient at lag one** is computed as the correlation between the observations y_{t+1} and y_t, for $t = 1, \ldots, n-1$:

$$\hat{\rho}_1 = \frac{\sum_{t=1}^{n-1}(y_t - \bar{y})(y_{t+1} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{3}$$

Sample Autocorrelation coefficients

In a similar way, we can find the correlation between the $n-k$ pairs of observations that are k steps apart:

$$(y_1, y_{k+1}), (y_2, y_{k+2}), \ldots, (y_{n-k}, y_n)$$

The **lag-k autocorrelation coefficient** is

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{4}$$

where $\bar{y} = \sum_{t=1}^{n} y_t/n$ is the overall mean.

Sample Autocorrelation coefficients / 2

The autocorrelation coefficients up to a given lag M can be calculated by estimating the autocovariance coefficients as

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})$$

We then compute the ratios

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \tag{5}$$

for $k = 1, \ldots, M$, $M < n$, where, for $k = 0$, we obtain the estimate of variance

$$\hat{\gamma}_0 = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})^2$$

Table of Contents

Example 1: A Simulated Time Series

Consider a time series generated by tossing a fair coin repeatedly, where, at each time t, we let

$$\begin{cases} x_t = 1 & \text{a head is obtained} \\ x_t = -1 & \text{a tail is obtained} \end{cases}$$

Construct y_t as

$$y_t = 5 + x_t - 0.7x_{t-1} \tag{6}$$

Suppose we obtain the following result for 10 tosses

$$x_t : \quad 1 \quad 1 \quad -1 \quad 1 \quad -1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1$$

A Simulated Time Series / 2

t	1	2	3	4	5	6	7	8	9	10
Coin	H	H	T	H	T	T	T	H	T	H
x_t	1	1	-1	1	-1	-1	-1	1	-1	1
y_t	6.7	5.3	3.3	6.7	3.3	4.7	4.7	6.7	3.3	6.7
$y_t - \bar{y}$	1.56	0.16	-1.84	1.56	-1.84	-0.44	-0.44	1.56	-1.84	1.56

Table 1: The values of y_t according to Eq.(6), and deviations $y_t - \bar{y}$, for $x_0 = -1$ and $N = 10$.

The sample autocorrelation for the series y_t can be calculated using Eq.(5), for lags $k = 0, 1, 2, \ldots 10$. For $k = 0$ we have

$$\hat{\gamma}_0 = \frac{1}{10}\sum_{t=1}^{10}(y_t - \bar{y})^2 = \frac{1}{10}[1.56^2 + 0.16^2 + (-1.84)^2 + 1.56^2 + (-1.84)^2 + (-0.44)^2 + (-0.44)^2 + 1.56^2 + (-1.84)^2 + 1.56^2] = 2.030$$

A Simulated Time Series / 3

Consider now the observations that are separated by 3 lags, e.g., y_1 and y_4, y_2 and y_5, and so on. For these observations, the sample correlation is given by

$$\begin{aligned}\hat{\gamma}_3 &= \frac{1}{10}\sum_{t=1}^{7}(y_{t+3}-\bar{y})(y_t-\bar{y}) \\ &= \frac{1}{10}[(1.56)(1.56)+(-1.84)(.16)+(-.44)(-1.84)+(-.44)(1.56) \\ &\quad +(1.56)(-1.84)+(-1.84)(-.44)+(1.56)(-.44)] = -0.48\end{aligned}$$

Thus we obtain the sample autocorrelation coefficient at lag $k=3$

$$\hat{\rho}_3 = \frac{\hat{\gamma}_3}{\hat{\gamma}_0} = \frac{-0.48}{2.030} = -0.024$$

A Simulated Time Series / 4

The *theoretical ACF* can be obtained from Eq.(2) using model (6). We have that, fol all t,

$$\mathsf{E}(X_t) = 1\left(\frac{1}{2}\right) + (-1)\frac{1}{2} = 0$$
$$\mathsf{Var}(X_t) = \mathsf{E}(X_t^2) - E(X_t)^2 = 1 - 0 = 1$$

Hence

$$\begin{aligned} \mathsf{E}(Y_t) =& 5, \\ \mathsf{Var}(Y_t) =& \mathsf{Var}(5 + X_t - 0.7X_{t-1}) = \mathsf{Var}(X_t - 0.7X_{t-1}) \\ =& \mathsf{Var}(X_t) + 0.7^2\mathsf{Var}(X_{t-1}) = 1 + 0.7^2 = 1.49 \end{aligned}$$

where we used the properties of the variance for a linear combination of uncorrelated random variables: $\mathsf{Var}(aX + bY) = a^2\mathsf{Var}(X) + b^2\mathsf{Var}(Y)$.

A Simulated Time Series / 5

Moreover

$$\begin{aligned}\mathrm{Cov}(Y_t, Y_{t-1}) &= \mathrm{Cov}(5 + X_t - 0.7X_{t-1},\, 5 + X_{t-1} - 0.7X_{t-2}) \\ &= \mathrm{Cov}(X_t, X_{t-1}) - 0.7\,\mathrm{Cov}(X_t, X_{t-2}) - 0.7\,\mathrm{Cov}(X_{t-1}, X_{t-1}) \\ &\quad + (-0.7)(-0.7)\mathrm{Cov}(X_{t-1}, X_{t-2}) \\ &= -0.7\,\mathrm{Cov}(X_{t-1}, X_{t-1}) = -0.7\,\mathrm{Var}(X_{t-1}) = -0.7\,\mathrm{Var}(X_t)\end{aligned}$$

where we used the fact that the X_t are uncorrelated

A Simulated Time Series / 6

Therefore

$$\rho_1 = \frac{\text{Cov}(Y_t, Y_{t-1})}{\text{Var}(Y_t)} = \frac{-0.7}{1.49} = -0.47$$

Moreover, it can be shown that for $k \geq 2$

$$\rho_k = 0$$

(**exercise**).

The same values can be obtained for negative values of k, yielding $\rho_{-k} = \rho_k$.

Remark. The binary process $\{X_t\}$ is an example of iid noise.

Theoretical and Sample ACFs

The Table compares the theoretical ACF with sample ACFs for a realization of Y_t where $N = 10$ and another realization where $N = 100$.

		$N = 10$	$N = 100$
k	ρ_k	$\hat{\rho}_k$	$\hat{\rho}_k$
0	1.00	1.00	1.00
± 1	-0.47	-0.55	-0.45
± 2	0.00	0.17	-0.12
± 3	0.00	-0.02	0.14
± 4	0.00	0.15	0.01
± 5	0.00	-0.46	-0.01

Table 2: Theoretical and Sample ACFs for $N = 10$ and $N = 100$ for model (6). Note the increased variability in the smaller size sample.

Large-Sample Distribution of the ACF

Suppose $y_1, y_2, \ldots y_n$ are observations on independent and identically distributed random variables with arbitrary mean (*white independent noise*). It has been shown that

$$E(\hat{\rho}_k) \approx -1/n$$
$$V(\hat{\rho}_k) \approx 1/n$$

and $\hat{\rho}_k$ is asymptotically normally distributed, that is, for any fixed positive integer k and large n

$$\hat{\rho}_k \dot{\sim} \mathcal{N}\left(-\frac{1}{n}, \frac{1}{n}\right)$$

We focus on the properties of $\hat{\rho}_k$ when sampling from a purely random process (all theoretical autocorrelation coefficient are 0 except at lag 0), to decide if the observed values of $\hat{\rho}_k$ from a given time series are *significantly different from zero*.

Example 2: Moving Averages

Consider a white noise process $w_t \sim \text{iid}\, \mathcal{N}(0,1)$. A collection of 500 such rv's is in figure(6)-top. The plot tends to show visually a mixture of many different kinds of oscillations.

We might replace the white noise series w_t by a **moving average** that *smooths the series*. For example, consider replacing w_t by an average of its current value and its immediate neighbours in the past and future:

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1}) \tag{7}$$

where each term is weighted $1/3$.

The resulting series is shown in figure(6)-bottom, and reflects the fact that the slower oscillations are more apparent and some of the faster oscillations are taken out.

Moving Averages / 2

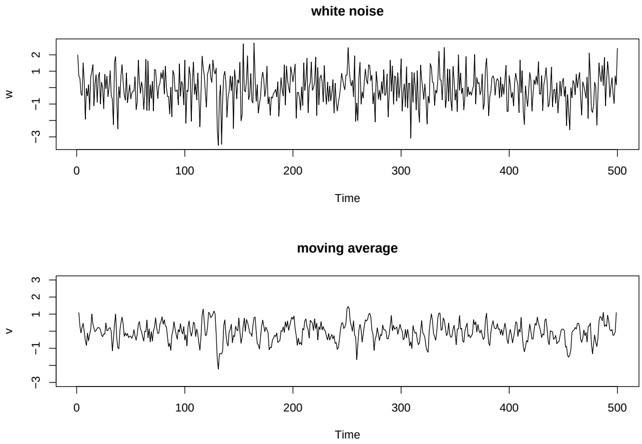


Figure 6: Gaussian white noise series (top) and three-point moving average of the Gaussian white noise series (bottom).

Filtering

A linear combination of values in a time series such as in (7) is referred to, generically, as a **filtered series**, given by the linear operation

$$y_t = \sum_{r=-q}^{s} b_r\, x_{t+r}$$

where $b_{-q}, b_{-q+1}, \ldots, b_s$ are the weights. In the three-point moving average example

$$v_t = \sum_{r=-1}^{1} \frac{1}{3}\, w_{t+r}$$

- Moving averages are often symmetric with $s = q$ and $b_j = b_{-j}$
- moving averages can be useful for removing seasonal variation
- a special type of filtering is first differencing

Moving Averages: autocovariance function

We now compute the autocovariance function for the three-point moving average

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1})$$

where $w_t \sim \text{WN}(0, \sigma_w^2)$.

$$\gamma(s,t) = \text{Cov}(v_s, v_t) = \text{Cov}\left\{\frac{1}{3}(w_{s-1} + w_s + w_{s+1}), \frac{1}{3}(w_{t-1} + w_t + w_{t+1})\right\}$$

When $s = t$ we have

$$\begin{aligned}\gamma(t,t) &= \frac{1}{9}\text{Cov}\left\{(w_{t-1} + w_t + w_{t+1}), (w_{t-1} + w_t + w_{t+1})\right\} \\ &= \frac{1}{9}[\text{Cov}(w_{t-1}, w_{t-1}) + \text{Cov}(w_t, w_t) + \text{Cov}(w_{t+1}, w_{t+1})] = \frac{3}{9}\sigma_w^2\end{aligned}$$

Moving Averages: autocovariance function

When $s = t + 1$ we have

$$\begin{aligned}\gamma(t+1,t) &= \frac{1}{9}\text{Cov}\left\{(w_t + w_{t+1} + w_{t+2}), (w_{t-1} + w_t + w_{t+1})\right\} \\ &= \frac{1}{9}[\text{Cov}(w_t, w_t) + \text{Cov}(w_{t+1}, w_{t+1})] = \frac{2}{9}\sigma_w^2\end{aligned}$$

Similar computations give

$$\gamma(t-1,t) = 2\sigma_w^2/9, \quad \gamma(t+2,t) = \gamma(t-2,t) = \sigma_w^2/9$$

Thus $E(v_t) = 0$ and

$$\gamma(s,t) = \begin{cases} \frac{3}{9}\sigma_w^2 & s = t \\ \frac{2}{9}\sigma_w^2 & |s-t| = 1 \\ \frac{1}{9}\sigma_w^2 & |s-t| = 2 \\ 0 & |s-t| > 2 \end{cases}$$

Note that $\gamma(s,t)$ only depends on the separation of v_s and v_t, say, $k = |s-t| \Rightarrow$ The process is weakly stationary.

Table of Contents

The correlogram

For an observed time series $y_1, y_2, \ldots, y_n$ the values $\hat{\rho}_1, \hat{\rho}_2, \ldots$ defined in Equation(4) represent the *sample ACF* of the series.

The sample ACF allows us to assess whether the data comes from a completely random series or whether correlations are *statistically significant* at some lags.

Correlogram A useful aid in interpreting a set of autocorrelation values is the plot of $\hat{\rho}_k$ against the lag k, for $k = 0, 1, \ldots, M$, where M is usually much less than the series length n.

Random series

Simulated white noise data will not have autocorrelations that are exactly zero (when $k \neq 0$) because of sampling variation.

For large n, the sample autocorrelations of an iid sequence $Y_1, \ldots, Y_n$ with finite variance are **approximately** iid with distribution $\mathcal{N}(-1/n, 1/n)$. Hence, if $y_1, y_2, \ldots, y_n$ is a realization of such an iid sequence, about 95% of the sample autocorrelations should fall between the bounds

$$-1/n \pm 1.96/\sqrt{n}$$

which are often further approximated to $\pm 2/\sqrt{n}$ and usually plotted when the ACF plot is produced.

Using the correlogram

Given that the data are really random

- We expect that only the 5% of the autocorrelations will be significantly different from zero (e.g. 1 out of 20)
- The overall probability of getting at least one coefficient outside these limits, given that the data come from a white noise, increases with the number of coefficients plotted

In practice, a single coefficient just outside the ‘null’ 95% confidence limits may be ignored, but several values well outside the bounds will provide plausible evidence of non-randomness.

Simulated white noise data

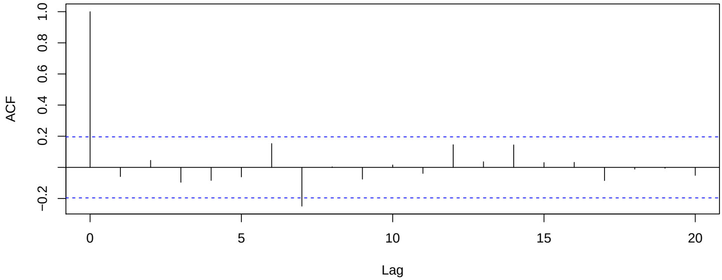


Figure 7: Correlogram for 100 observations supposed to be i.i.d. normally distributed. The ‘null’ 95% confidence limits (dotted lines) are approximately $\pm 2/\sqrt{100} = \pm 0.2$. The statistically significant value at lag 7 is due to sampling variation.

Ljung-Box Test

A *simultaneous test* is commonly used to test the *joint hypothesis that all m of the correlation coefficients are simultaneously equal to zero*

The null hypothesis of the **Ljung-Box test** is

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0 \quad \text{for some } m$$

against the alternative hypothesis

$$H_1 : \rho_i \neq 0 \text{ for some } i \in \{1, \ldots, m\}$$

The test statistic (Ljung and Box (1978)) is

$$Q(m) = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k}$$

where n = sample size, m = maximum lag length (different values of m are often used).

Ljung-Box Test

Under the null hypothesis H_0 that $\{x_t\}$ is a sequence of iid random variables, Q is asymptotically a **chi-squared r.v. with m degrees of freedom**:

$$Q(m) \sim \chi^2_m$$

Rejection rule: Reject H_0 of iid data (at level α) if $Q(m) > \chi^2_{1-\alpha,m}$, where $\chi^2_{1-\alpha,m}$ is the $(1-\alpha)$-quantile of the chi-squared distribution with m degrees of freedom

For the same problem, we can compute the p-value: a small p-value is evidence against the null hypothesis, meaning that the data does not support our original beliefs

ACF Plots and L-B Test: Example 1

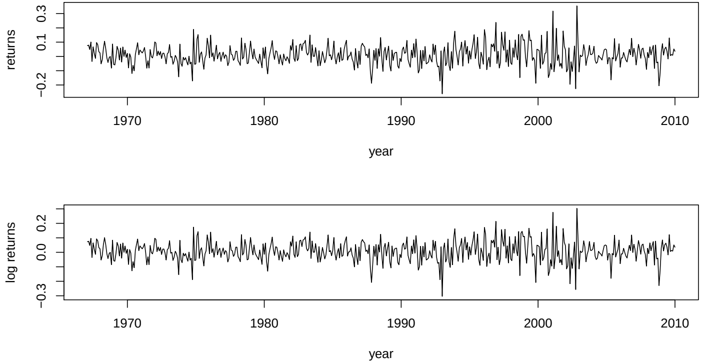


Figure 8: Monthly net and log returns of IBM stock from January 1967 to December 2009.

ACF Plots and L-B Test: Example 1

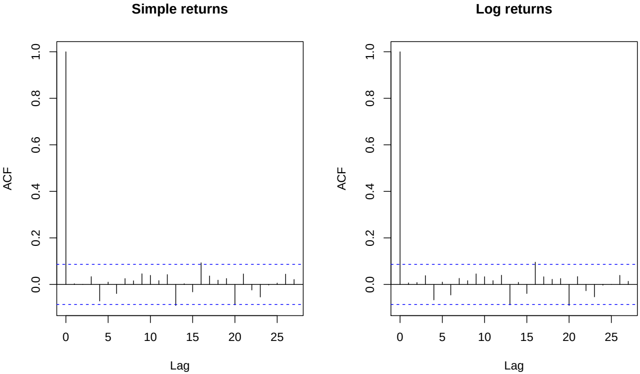


Figure 9: Sample ACF of the monthly simple and log returns of IBM stock from January 1967 to December 2009.

ACF Plots and L-B Test: Example 1

- The ACF plot starts with $\rho_0 = 1$, then the plots show that serial correlations of monthly IBM stock returns are very small
- To verify that the returns have no serial correlations, we test

$$H_0 : \rho_1 = \cdots = \rho_m = 0 \quad \text{versus} \quad H_1 : \rho_i \neq 0 \text{ for some } i \in \{1, \ldots, m\}$$

with $\alpha = 0.05$, for $m = 12$ and $m = 24$ (lags at the multiples of the seasonality are more important):

	$Q(12)$	p-value ($m = 12$)	$Q(24)$	p-value ($m = 24$)
net returns	7.57	0.82	25.49	0.38
log returns	7.40	0.83	25.39	0.38

- The 0.95-quantiles of the χ^2 distribution with 12 and 24 degrees of freedom are 21.03 and 36.42, respectively
- We cannot reject the null hypothesis of no serial correlations in the IBM stock returns

ACF Plots and L-B Test: Example 2

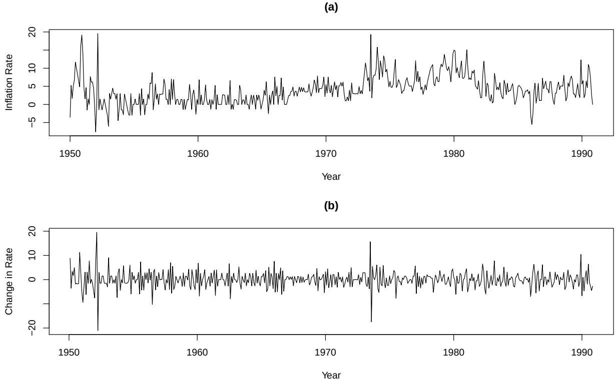


Figure 10: (a) One-month US inflation rate (annual rate, in percent) and (b) first differences (changes in rate).

ACF Plots and L-B Test: Example 2

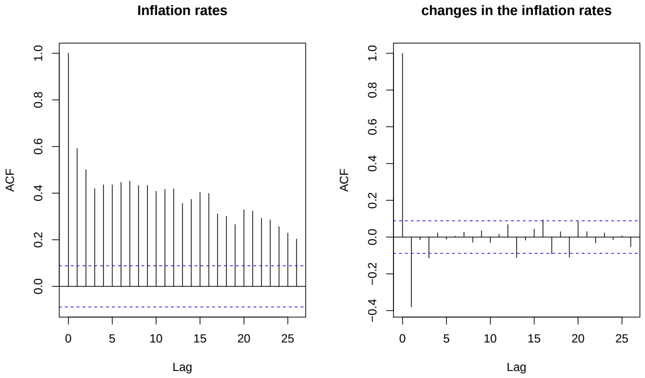
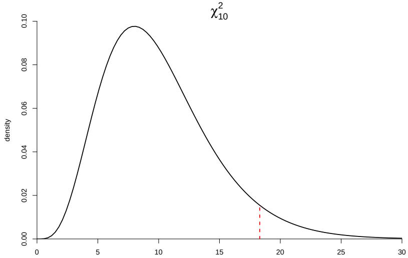


Figure 11: Sample ACF of the one-month inflation rate and changes in the inflation rate. The first series has a slowly decaying ACF typical of a nonstationary or possibly long-memory process. The sample ACF on the right decays to zero quickly, which is a sign of stationarity.

ACF Plots and L-B Test: Example 2

- Several of the autocorrelations of the rate change series fall outside the test bounds, which suggests to perform the Ljung-Box test
- For $m = 10$ the value of the test statistic is $Q(m) = 79.92$.
  At level $\alpha = 0.05$, the region for rejection of the hypothesis of randomness is:

$$Q > \chi^2_{0.95,10} = 18.31$$



The p-value is
$Pr(\chi^2_{10} > 79.92) = 5.2 \times 10^{-13}$

The null hypothesis of no serial correlation is strongly rejected.

Simple symmetric random walk

The **random walk** $\{S_t\}$, $t = 0, 1, 2, \dots$ (starting at zero) is defined by cumulatively summing iid random variables. Thus a random walk model with zero mean is given by

$$\begin{aligned} S_0 =& 0 \\ S_t =& W_1 + W_2 + \cdots + W_t, \quad t = 1, 2, \dots \end{aligned}$$

where $\{W_t\}$ is iid white noise. If $\{W_t\}$ is the binary process of Example 1 (coin tosses), then $\{S_t\}$ is a **simple symmetric random walk**.

The outcomes of the coin tosses can be recovered from $\{S_t\}$ by differencing:

$$S_t - S_{t-1} = W_t$$

As a result, first-order differencing of a simple symmetric random walk yields a white noise process, that is a stationary process.

Random walks: non stationarity

If $\{S_t\}$ is a simple symmetric random walk defined in the previous example, then

$$E(S_t) = 0, \quad \text{Var}(S_t) = t\sigma^2$$

for all t, and, for $k \geq 0$,

$$\begin{aligned}\gamma(t+k,t) =&\text{Cov}(S_{t+k}, S_t)\\ =&\text{Cov}(S_t + W_{t+1} + \cdots + W_{t+k}, S_t)\\ =&\text{Cov}(S_t, S_t)\\ =&t\sigma^2\end{aligned}$$

Since $\gamma(t+k,t)$ depends t, the series $\{S_t\}$ is *not* stationary.