Chapter 1:

Introduction
Contents:

I. Introduction ................................................................................................................................. 3

I.1 Assets Returns ............................................................................................................................ 4
   I.1.1 Definition of Asset Returns ............................................................................................... 5
   I.1.2 Statistical Properties of Returns .................................................................................... 11
   I.1.3 Empirical Properties of Returns .................................................................................... 20

I.2 Basics of Time Series Analysis ............................................................................................... 24
   I.2.1 Stationarity ....................................................................................................................... 24
   I.2.2 Autocorrelation Function ............................................................................................... 28
   I.2.3 Partial Autocorrelation Function (PACF) ....................................................................... 32
   I.2.3 Transformation of Data .................................................................................................... 34
I. Introduction

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 \( Y_1, Y_2, \ldots, Y_T \) 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 Assets Returns

- return of an asset is a complete and scale-free summary of the investment opportunity

- return series are easier to handle than price series because former have more attractive statistical properties e.g. stationarity
I.1.1 Definition of Asset Returns

a) One-Period Simple Return

Holding the asset for one period from date \( t-1 \) to date \( t \) would result in a *simple gross return*:

\[
1 + R_t = \frac{P_t}{P_{t-1}}
\]

\[P_t = P_{t-1}(1 + R_t)\]

The corresponding *one-period simple net return* or *simple return* is

\[
R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}
\]
b) Multi-period Simple Return

Holding the asset for \( k \) periods between dates \( t - k \) and \( t \) gives a \( k \)-period simple gross return

\[
1 + R_t[k] = \frac{P_t}{P_{t-k}} \cdot \frac{P_{t-1}}{P_{t-2}} \cdot \ldots \cdot \frac{P_{t-k+1}}{P_{t-k}}
\]

\[
1 + R_t[k] = \frac{P_t}{P_{t-k}}
\]

\[
1 + R_t[k] = (1 + R_t) \cdot (1 + R_{t-1}) \cdot \ldots \cdot (1 + R_{t-k+1})
\]

\[
1 + R_t[k] = \prod_{j=0}^{k-1} (1 + R_{t-j})
\]

Thus, the \( k \)-period simple gross return is just the product of the \( k \) one-period simple gross return, which is called compound return.

\[
R_t[k] = \frac{P_t - P_{t-k}}{P_{t-k}}
\]
Mostly, when we present descriptive statistics returns are annualized. If the asset was held for $k$ years, then the annualized (average) return is defined as geometric mean of the $k$ one-period simple gross returns:

\[
\{R_t[k]\} = \left[ \prod_{j=0}^{k-1} (1 + R_{t-j}) \right]^{\frac{1}{k}} - 1
\]

We also can annualized the arithmetic mean of monthly average return by:

\[
\{R_t[k]\}_{\text{annulized}} = \left[ 1 + \left( \frac{1}{k} \sum_{j=1}^{k} R_j \right) \right]^{12}
\]
c) Continuously Compounded Return

\[ r_t = \ln(1 + R_t) \]
\[ r_t = \ln \frac{P_t}{P_{t-1}} \]
\[ r_t = \ln P_t - \ln P_{t-1} \]
\[ r_t = p_t - p_{t-1} \]

where \( p_t = \ln(P_t) \).
Advantages of continuously compounding:

- Continuously multi-period return:

\[ r_t[k] = \ln(1 + R_t[k]) \]
\[ r_t[k] = \ln[(1 + R_t) \cdot (1 + R_{t-1}) \cdot \ldots \cdot (1 + R_{t-k+1})] \]
\[ r_t[k] = \ln(1 + R_t) + \ln(1 + R_{t-1}) + \ldots + \ln(1 + R_{t-k+1}) \]
\[ r_t[k] = r_t + r_{t-1} + \ldots + r_{t-k+1} \]

where \( r_t = \ln(1 + R_t) \).

- Statistical properties of log returns are more tractable
d) Relationship between Simple and Continuously Compounded Returns

The relationships between simple return $R_t$ and continuously compounded return $r_t$ are:

$$r_t = \ln(1 + R_t)$$

$$R_t = e^{r_t} - 1$$
I.1.2 Statistical Properties of Returns

a) Moments of a Random Variable (RV)

The n-th moment of a continuous random variable Y is defined as

\[ M_n = E[Y^n] = \int_{-\infty}^{+\infty} y^n f(y) \cdot dy, \]

where \( E \) stands for expectation and \( f(y) \) is the probability function of Y.

The first moment is called the mean or expectation of Y. It measures the central location of the distribution. We denote the mean of Y by \( \mu_Y \). The n-th central moment of Y is defined as

\[ M_n = E[(Y - \mu_Y)^n] = \int_{-\infty}^{+\infty} (y - \mu_Y)^n f(y) \cdot dy, \]

provided that the integral exists.
Skewness:

\[ S(y) = E \left( \frac{(y - \mu_y)^3}{\sigma_y^3} \right) \]

Kurtosis:

\[ K(y) = E \left( \frac{(y - \mu_y)^4}{\sigma_y^4} \right) \]

\( K(y) - 3 \) is called the excess kurtosis because \( K(y) = 3 \) for a normal distribution.
The sample mean is

\[ \hat{\mu}_y = \frac{1}{T} \sum_{t=1}^{T} y_t, \]

the sample variance is

\[ \hat{\sigma}^2_y = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\mu}_y)^2, \]

the sample skewness is

\[ \hat{s}(y) = \frac{1}{T} \sum_{t=1}^{T} \frac{(y_t - \hat{\mu}_y)^3}{\hat{\sigma}^3_y}, \]
the sample kurtosis is

\[ \hat{k}(y) = \frac{1}{T} \sum_{t=1}^{T} \frac{(y_t - \hat{\mu}_y)^4}{\hat{\sigma}_y^4}. \]

Under normality assumption, \( \hat{s}(y) \) and \( \hat{k}(y) \) are distributed asymptotically as normal with zero mean and variances \( 6/T \) and \( 24/T \), respectively.

Financial data often exhibit leptokurtosis, i.e. a kurtosis higher than 3 or an excess kurtosis higher than 0. We consider such return pattern especially for high frequency data, for example daily data. For monthly, quarterly or yearly aggregated data the distribution turns more towards a normal distribution.
Figure 1: Skewness and Excess Kurtosis
Test for Skewness and Kurtosis:

Distribution:

\[
\hat{s}(y) = \frac{1}{T} \sum_{t=1}^{T} \frac{(y_t - \hat{\mu}_y)^3}{\hat{\sigma}_y^3} \sim N \left( 0, \frac{6}{T} \right)
\]

\[
\hat{k}(y) = \frac{1}{T} \sum_{t=1}^{T} \frac{(y_t - \hat{\mu}_y)^4}{\hat{\sigma}_y^4} \sim N \left( 3, \frac{24}{T} \right)
\]
For the coefficient of skewness the exact test of Urzua (1996) is defined as:

\[ \left| \frac{\hat{s}}{\sqrt{c}} \right| > z_\alpha \]

\[ c = 6 \cdot \frac{(T - 2)}{(T + 1)(T + 3)} \]

The coefficient of kurtosis is significant different from 3, if

\[ \left| \frac{\hat{k} - 3}{\sqrt{24/T}} \right| > z_\alpha \]
This test only works approximately for very large sample sizes due to the fact that the estimator for the kurtosis coefficient is biased. However, a test based on exact moments can be formulated as (Urzua 1996):

\[
\left| \frac{\hat{k} - a}{\sqrt{b}} \right| > z_\alpha
\]

\[
a = 3 \cdot \frac{T - 1}{T + 1}
\]

\[
b = 24 \cdot \frac{T(T - 2)(T - 3)}{(T + 1)^2(T + 3)(T + 5)}
\]
b) Test of Normality

Jarque-Bera test statistic:

The test statistic measures the difference of the skewness and kurtosis of the series with those from the normal distribution. The statistic is computed as:

\[ JB = \frac{T}{6} \cdot \left( \frac{s^2}{4} + \frac{1}{4} (k - 3)^2 \right) \sim \chi^2(2) \]

Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is distributed as \( \chi^2 \) with 2 degrees of freedom.

1% \( \approx 9.21 \)

5% \( \approx 5.99 \)

The test is only adequate for large samples, whereas for small samples you have to interpret it cautiously.
I.1.3 Empirical Properties of Returns

Figure 2: DAX Total Return Index
Figure 3: Simple discrete (monthly) returns of the DAX Total Return Index
Figure 4: Continously (monthly) returns of the DAX Total Return Index
Figure 5: Descriptive Statistics of DAX Returns
I.2 Basics of Time Series Analysis

I.2.1 Stationarity

Strict Stationary:

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

\( \rightarrow \) invariant under time shift

The random variables \( Y(t+1), \ldots, Y(t+n) \) have the same joint distribution as \( Y(t+1+c), \ldots, Y(t+n+c) \), with \( c \) as an 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.

1. \( E(Y_t) = \mu = \text{constant}, \ \forall \ t, \) (mean stationarity)
2. \( \text{Var}(Y_t) = \sigma_t^2 = \sigma^2 = \text{constant}, \ \forall \ t \) (variance stationarity), and
3. \( \text{Cov}(Y_t, Y_{t-j}) = \sigma_{tj} = \sigma_j = \text{constant}, \ \forall \ t \) (covariance stationarity),

\[
E[(y_t - \mu)(y_{t-j} - \mu)] = E[(y_s - \mu)(y_{s-j} - \mu)] \text{ for all } t \neq s.
\]

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 \( \{Y_t | t = 1, \ldots, T\} \):

- 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 for the strength and direction of interrelation, and

c) the partial autocorrelation function for the contribution of adding a new regression coefficient.
I.2.2 Autocorrelation Function
Consider a weakly stationary series $Y_t$. When the 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 $\rho_j$, 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 standard error of the correlation coefficient is calculated as

\[
SE_r = \pm \frac{2}{\sqrt{n}}
\]

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 \( \frac{2}{\sqrt{n}} \). Under certain assumptions this term can be viewed as a 95% confidence interval and empirical autocorrelation coefficients, which lay outside of this interval, are significant different from null.
Box and Pierce (1970) propose the Portmanteau statistic:

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

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

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

$\rightarrow r_t \sim \text{i.i.d. } \chi^2(m)$

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

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

$\rightarrow m \approx \ln(T)$

$\rightarrow$ number of autocorrelation coefficients should be around 20% of the sample size
Figure 6: Autocorrelation, Partial Autocorrelation and Ljung-Box (LB) Test
I.2.3 Partial Autocorrelation Function (PACF)

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

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

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

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

\[\ldots\]

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

What additional contribution is delivered from \(Y_{t-p}\), if the explanation is controlled by \(Y_{t-1} \ldots Y_{t-p-1}\)?
Given \( \{Y_t\} \) is a stationary process and the partial autocorrelation \( \hat{\phi}_{j,j} = \hat{\phi}_j \) (for \( j \geq 2 \)) is the partial correlation of \( Y_{t-j} \) and \( Y_t \) under holding constant all random variables \( Y_i \), which lay between \( t-j < i < t \):

- \( \hat{\phi}_{j,i} \) describes the partial correlation between \( Y_{t-j} \) and \( Y_{t-j+i} \)
- \( \hat{\phi}_0 = 1 \)
- \( \hat{\phi}_1 = \hat{\rho}_1 \) → autocorrelation coefficient of lag 1
- \( \hat{\phi}_{-j} = \hat{\phi}_j \) → the partial autocorrelation function is symmetric
- Generally the equation for the partial autocorrelation coefficients is defined as:

\[
\phi_j = \begin{cases} 
\rho_1 & \text{for } j = 1 \\
\rho_j - \sum_{i=1}^{j-1} \hat{\phi}_{j-1,i} \rho_{t-i} & \text{for } j > 1 \\
1 - \sum_{i=1}^{j-1} \hat{\phi}_{j-1,i} \rho_{t-i} & 
\end{cases}
\]
I.2.3 Transformation of Data

I.2.3.1 Differencing for Trends Elimination (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 to a greater or lesser extent 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 \( \Delta^d Y_t \).
I.2.3.2 Box-Cox-Transformation for Stabilization of the Variance

If $Y$ is the original and $X$ the transformed series, then we can denote the approach in general as:

$$X_t = \begin{cases} \frac{Y_t^\theta - 1}{\theta} & \text{for } 0 < \theta < 1 \\ \log (Y_t) & \text{for } \theta = 0 \end{cases}$$

To apply the transformation procedure we have to determine $\theta$, which can approximately be chosen so that the variance of the series $X$ is constant:

$$\sigma_t = c \mu_t^{1-\theta}$$

where $\sigma$ is the standard deviation and $\mu$ is the mean of the time series $Y_t$. 
Divide the series into \( K \) sub-periods:

\[
s_i = cm_i^{1-\theta} \quad \text{with} \quad i = 1, \ldots, K
\]

We also can determine \( \theta \) on the basis of the logarithmized equation:

\[
\ln(s_i) = \ln c + (1 - \theta) \ln(m_i)
\]

As a special case of this transformation, for \( \theta = 0 \), we receive the logarithmized transformation of the time series:

\[
Y_t = \log(X_t)
\]