Financial Data Analysis

Introduction

Roman Liesenfeld
(University of Kiel)

October 6, 2008
Contents

1 Introduction
   1.1 Asset Returns and Their Properties .................................................. 4
   1.2 Quick Review of Linear Univariate Time Series Models .......................... 10
1 Introduction

In the introduction, I summarize fundamental definitions of returns and compounded returns, and review the basic theory of time series analysis that is useful for financial statistics.

1.1 Asset Returns and Their Properties

- One of the most important variable in financial statistics is the “return of financial assets.” Financial assets are, e.g., stocks, foreign currencies, treasury bills.

- I start with the various definitions of returns, and then I discuss some of their statistical properties.

- Let

\[ P_t = \text{price of an asset at date } t \]

and assume that the asset pays no dividends.

Then, the simple gross return is

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

the simple net return is

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

the multiperiod simple gross return, holding an asset \( k \) periods from \( t - k \) to \( t \), is

\[ 1 + R_{t[k]} = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \cdot \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-k+1}}{P_{t-k}} = (1 + R_t) (1 + R_{t+1}) \cdots (1 + R_{t-k+1}) \]

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

the multiperiod simple net return is

\[ R_{t[k]} = \frac{P_t - P_{t-k}}{P_{t-k}} . \]

- Returns are typically given in terms of annual returns, i.e., the relative price change over a one-year-horizon. In order to transform a multi-year return into an annual return, one can compute the annualized average return (geometric mean of \( k \) one-period returns):

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

\[ = \exp \left[ \frac{1}{k} \sum_{j=0}^{k-1} \ln (1 + R_{t-j}) \right] - 1 \]
- The annualized returns are often approximated by the simple arithmetic mean of \( k \) one-period returns

\[
\text{Ann} \{ R_t[k] \} \approx \frac{1}{k} \sum_{j=0}^{k-1} R_{t-j}.
\]

This approximation results from a first-order Taylor series expansion (TSE) of \( \text{Ann} \{ R_t[k] \} \) around \( R_{t-j} = 0 \) \( \forall j \). \(^1\) This TSE-approximation is only adequate if returns \( R_{t,j} \) are small.

- We now turn to the continuously compounded returns. Let

\[
r_t = \text{interest rate per annum} \\
P_t = \text{Asset price / value of the initial deposit}.
\]

The net value after one year, if interest is paid once a year is:

\[
P_{t+1} = P_t (1 + r_{t+1}),
\]

and the net value after one year if interest is paid \( m \) times a year:

\[
P_{t+1} = P_t \left(1 + \frac{r_{t+1}}{m}\right)^m.
\]

For \( m \to \infty \) (continuously compounding):

\[
P_{t+1} = \lim_{m \to \infty} P_t \left(1 + \frac{r_{t+1}}{m}\right)^m = P_t e^{r_{t+1}}.
\]

- Based on this relationship, simple returns under continuously compounding are obtained as

\[
R_{t+1} = \frac{P_{t+1}}{P_t} - 1 = e^{r_{t+1}} - 1
\]

and the natural logarithm of the gross return:

\[
\ln (R_{t+1} + 1) = r_{t+1} = \ln \left(\frac{P_{t+1}}{P_t}\right) = p_{t+1} - p_t.
\]

The quantity \( \ln(R_{t+1} + 1) \) is called “continuously compounded return” or “log return”.

\(^1\)The function to be approximated is

\[
f(R) = \exp \left\{ \frac{1}{k} \sum_{j=0}^{k-1} \ln \left(1 + R_{t-j}\right) \right\} - 1
\]

with \( R = (R_t, \ldots, R_{t-k+1})' \) and

\[
f(0) = \exp \left\{ \frac{1}{k} \sum_{j=1}^{k-1} \ln(1) \right\} - 1 = 0
\]

\[
\frac{\partial f(R)}{\partial R_{t-j}} = \frac{1}{k} \left( \frac{1}{1 + R_{t-j}} \right) \exp \left\{ \frac{1}{k} \sum_{j=0}^{k-1} \ln \left(1 + R_{t-j}\right) \right\} \bigg|_{R=0} = \frac{1}{k}.
\]

Hence, \( f(R) \) can be approximated by

\[
f(R) \approx f(0) + \frac{\partial f(0)}{\partial R} \cdot R = \left( \frac{1}{k}, \ldots, \frac{1}{k} \right) \cdot R = \frac{1}{k} \sum_{j=0}^{k-1} R_{t-j}.
\]
- The corresponding in multiperiod continuously compounded returns are

\[
\begin{align*}
    r_t[k] &= \ln (R_t[k] + 1) \\
    &= \ln \left( \prod_{j=0}^{k-1} (1 + R_{t-j}) \right) \\
    &= \sum_{j=0}^{k-1} \ln (1 + R_{t-j}) = \sum_{j=0}^{k-1} r_{t-j},
\end{align*}
\]

being the sum of continuously compounded one-period returns. This implies that compounding, which is a multiplicative operation, is transformed into an additive operation.

- If an asset pays dividends, the definition of asset returns has to be adjusted. Let

\[D_t = \text{dividend payment between date } t - 1 \text{ and } t.\]

Then the simple net return and the continuously compounded return are

\[
R_t = \frac{P_t + D_t}{P_{t-1}} - 1 \quad \text{and} \quad r_t = \ln (P_t + D_t) - \ln P_t.
\]

![Figure 1: Dividend Payment Timing Convention](image)

- We now turn to the discussion of some empirical properties of asset returns:
Figure 2: Time plots of monthly returns of IBM stock from January 1926 to December 1997. The upper panel is for simple net returns, and the lower panel is for log returns. (T02)

Figure 3: Time plots of monthly returns of the value-weighted index form January 1926 to December 1997. The upper panel is for simple net returns, and the lower panel is for log returns. (T02)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VW</td>
<td>62/7/3</td>
<td>8938</td>
<td>0.049</td>
<td>0.798</td>
<td>-1.23</td>
<td>30.06</td>
<td>-17.18</td>
<td>8.67</td>
</tr>
<tr>
<td>EW</td>
<td>62/7/3</td>
<td>8938</td>
<td>0.083</td>
<td>0.674</td>
<td>-1.09</td>
<td>18.09</td>
<td>-10.48</td>
<td>6.95</td>
</tr>
<tr>
<td>I.B.M.</td>
<td>62/7/3</td>
<td>8938</td>
<td>0.050</td>
<td>1.479</td>
<td>0.01</td>
<td>11.34</td>
<td>-22.96</td>
<td>12.94</td>
</tr>
<tr>
<td>Intel</td>
<td>72/12/15</td>
<td>6329</td>
<td>0.138</td>
<td>2.880</td>
<td>-0.17</td>
<td>6.76</td>
<td>-29.57</td>
<td>26.38</td>
</tr>
<tr>
<td>3M</td>
<td>62/7/3</td>
<td>8938</td>
<td>0.051</td>
<td>1.395</td>
<td>-0.55</td>
<td>16.92</td>
<td>-25.98</td>
<td>11.54</td>
</tr>
<tr>
<td>Microsoft</td>
<td>86/3/14</td>
<td>2985</td>
<td>0.201</td>
<td>2.422</td>
<td>-0.47</td>
<td>12.08</td>
<td>-30.13</td>
<td>17.97</td>
</tr>
<tr>
<td>Citi-Grp</td>
<td>86/10/30</td>
<td>2825</td>
<td>0.125</td>
<td>2.124</td>
<td>-0.06</td>
<td>9.16</td>
<td>-21.74</td>
<td>20.75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VW</td>
<td>62/7/3</td>
<td>8938</td>
<td>0.046</td>
<td>0.803</td>
<td>-1.66</td>
<td>40.06</td>
<td>-18.84</td>
<td>8.31</td>
</tr>
<tr>
<td>EW</td>
<td>62/7/3</td>
<td>8938</td>
<td>0.080</td>
<td>0.676</td>
<td>-1.29</td>
<td>19.98</td>
<td>-11.08</td>
<td>6.72</td>
</tr>
<tr>
<td>I.B.M.</td>
<td>62/7/3</td>
<td>8938</td>
<td>0.039</td>
<td>1.481</td>
<td>-0.33</td>
<td>15.21</td>
<td>-26.09</td>
<td>12.17</td>
</tr>
<tr>
<td>Intel</td>
<td>72/12/15</td>
<td>6329</td>
<td>0.096</td>
<td>2.894</td>
<td>-0.59</td>
<td>8.81</td>
<td>-35.06</td>
<td>23.41</td>
</tr>
<tr>
<td>3M</td>
<td>62/7/3</td>
<td>8938</td>
<td>0.041</td>
<td>1.403</td>
<td>-1.05</td>
<td>27.03</td>
<td>-30.08</td>
<td>10.92</td>
</tr>
<tr>
<td>Microsoft</td>
<td>86/3/14</td>
<td>2985</td>
<td>0.171</td>
<td>2.443</td>
<td>-1.10</td>
<td>19.65</td>
<td>-35.83</td>
<td>16.53</td>
</tr>
<tr>
<td>Citi-Grp</td>
<td>86/10/30</td>
<td>2825</td>
<td>0.102</td>
<td>2.128</td>
<td>-0.44</td>
<td>10.68</td>
<td>-24.51</td>
<td>18.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VW</td>
<td>26/1</td>
<td>864</td>
<td>0.99</td>
<td>5.49</td>
<td>0.23</td>
<td>8.13</td>
<td>-29.00</td>
<td>38.28</td>
</tr>
<tr>
<td>EW</td>
<td>26/1</td>
<td>864</td>
<td>1.32</td>
<td>7.54</td>
<td>1.65</td>
<td>15.24</td>
<td>-31.23</td>
<td>65.51</td>
</tr>
<tr>
<td>I.B.M.</td>
<td>26/1</td>
<td>864</td>
<td>1.42</td>
<td>6.70</td>
<td>0.17</td>
<td>1.94</td>
<td>-26.19</td>
<td>35.12</td>
</tr>
<tr>
<td>Intel</td>
<td>72/12</td>
<td>300</td>
<td>2.86</td>
<td>12.95</td>
<td>0.59</td>
<td>3.29</td>
<td>-44.87</td>
<td>62.50</td>
</tr>
<tr>
<td>3M</td>
<td>46/2</td>
<td>623</td>
<td>1.36</td>
<td>6.46</td>
<td>0.16</td>
<td>0.89</td>
<td>-27.83</td>
<td>25.77</td>
</tr>
<tr>
<td>Microsoft</td>
<td>86/4</td>
<td>141</td>
<td>4.26</td>
<td>10.96</td>
<td>0.81</td>
<td>2.32</td>
<td>-24.91</td>
<td>51.55</td>
</tr>
<tr>
<td>Citi-Grp</td>
<td>86/11</td>
<td>134</td>
<td>2.55</td>
<td>9.17</td>
<td>-0.14</td>
<td>0.47</td>
<td>-26.46</td>
<td>26.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>VW</td>
<td>26/1</td>
<td>864</td>
<td>0.83</td>
<td>5.48</td>
<td>-0.53</td>
<td>7.31</td>
<td>-34.25</td>
<td>32.41</td>
</tr>
<tr>
<td>EW</td>
<td>26/1</td>
<td>864</td>
<td>1.04</td>
<td>7.24</td>
<td>0.34</td>
<td>8.91</td>
<td>-37.44</td>
<td>50.38</td>
</tr>
<tr>
<td>I.B.M.</td>
<td>26/1</td>
<td>864</td>
<td>1.19</td>
<td>6.63</td>
<td>-0.22</td>
<td>2.05</td>
<td>-30.37</td>
<td>30.10</td>
</tr>
<tr>
<td>Intel</td>
<td>72/12</td>
<td>300</td>
<td>2.03</td>
<td>12.63</td>
<td>-0.32</td>
<td>3.20</td>
<td>-59.54</td>
<td>48.55</td>
</tr>
<tr>
<td>3M</td>
<td>46/2</td>
<td>623</td>
<td>1.15</td>
<td>6.39</td>
<td>-0.14</td>
<td>1.32</td>
<td>-32.61</td>
<td>22.92</td>
</tr>
<tr>
<td>Microsoft</td>
<td>86/4</td>
<td>141</td>
<td>3.64</td>
<td>10.29</td>
<td>0.29</td>
<td>1.32</td>
<td>-28.64</td>
<td>41.58</td>
</tr>
<tr>
<td>Citi-Grp</td>
<td>86/11</td>
<td>134</td>
<td>2.11</td>
<td>9.11</td>
<td>-0.50</td>
<td>1.14</td>
<td>-30.73</td>
<td>23.18</td>
</tr>
</tbody>
</table>

Figure 4: Table with Descriptive Statistics for Daily and Monthly Simple and Log Returns of Selected Indexes and Stocks. Returns Are in Percentages, and the Sample Period Ends on December 31, 1997. The Statistics Are Defined in Equations (1.10) to (1.13), and VW and EW Denote Value-Weighted and Equal-Weighted Indexes. (T02)
Figures 2 and 3 plot simple net returns and log returns. The table in Figure 4 provides the following quantities characterizing the empirical distribution:

sample mean: $\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} x_t$

sample standard deviation: $\hat{\sigma} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (x_t - \hat{\mu})^2}$

sample skewness: $\hat{S} = \frac{1}{T-1} \sum_{t=1}^{T} \frac{(x_t - \hat{\mu})^3}{\hat{\sigma}^3}$

sample kurtosis: $\hat{K} = \frac{1}{T-1} \sum_{t=1}^{T} \frac{(x_t - \hat{\mu})^4}{\hat{\sigma}^4}$

The skewness $\hat{S}$ contains information about the symmetry of the empirical distribution and the kurtosis $\hat{K}$ about the tail thickness and peakedness. $\hat{S}$ and $\hat{K}$ are estimates of the corresponding standardized third and fourth-order moments:

$$S = \frac{\text{E} \left[ (\bar{x} - \text{E} [\bar{x}])^3 \right]}{\left( \text{E} \left[ (\bar{x} - \text{E} [\bar{x}])^2 \right] \right)^{3/2}}$$

and

$$K = \frac{\text{E} \left[ (\bar{x} - \text{E} [\bar{x}])^4 \right]}{\left( \text{E} \left[ (\bar{x} - \text{E} [\bar{x}])^2 \right] \right)^2}.$$
1.2 Quick Review of Linear Univariate Time Series Models

- In the following I review the basic theory of linear time series models, which is useful for the empirical analysis of financial time series such as the time series of asset returns.

- A time series is defined as:

  a collection of random variables over time \( \{r_t, t : 1 \rightarrow T\} \). For each time period \( t \) we have one observation of the random variable.

Stationarity

- Let’s start with an important definition of time series analysis, the definition of stationarity:

  A time series is said to be strictly stationary \( \Leftrightarrow \) joint pdf is invariant under time shifts, i.e.,

  \[
  f(r_{t_1}, \ldots, r_{t_k}) = f(r_{t_1+\ell}, \ldots, r_{t_k+\ell}) \quad \forall \ell, k \in \mathbb{N}.
  \]

- This is a very strong requirement, which is hard to verify empirically. A weaker form of stationarity referring to the first and second order moments of the pdf is:

  A time series is said to be weakly stationary \( \Leftrightarrow \)

  1. \( \text{E}[r_t] = \mu < \infty \)
  2. \( \text{E}[(r_t - \mu)^2] = \sigma^2 < \infty \)
  3. \( \text{Cov}(r_t, r_{t-\ell}) = \gamma_\ell < \infty \).

This definition implies:

- First and second order moments are time invariant.

- A plot of weakly stationary time series \( \{r_t\}_{t=1}^T \) fluctuates with constant variation around a constant level:
- If the first two moments are finite strict stationarity leads to weak stationarity. However, the converse is not true.

- Here we are only concerned with weak stationarity.

- In empirical finance it is common to assume that a times series of asset returns is weakly stationary.

**Autocorrelation function (ACF)**

- The autocovariance and autocorrelation of a time series is defined as follows:

  For a stationary time series \{r_t\} the lag-ℓ autocovariance is

  \[
  \gamma_\ell = \text{E} \left[ (r_t - \text{E}[r_t]) (r_{t-\ell} - \text{E}[r_t]) \right], \quad \ell = 0, 1, 2, \ldots
  \]

  with \(\gamma_0 = \text{var}[r_t]\). The lag-ℓ autocorrelation is

  \[
  \rho_\ell = \frac{\gamma_\ell}{\gamma_0} \in [-1, 1], \quad l = 0, 1, 2, \ldots
  \]

  \{\rho_\ell\} is called ACF.

- The ACF measures linear (serial) dependence between \(r_t\) and its past values \(r_{t-1}, r_{t-2}\). For \(\rho_\ell = 0\), the time series variables \{\(r_t\)\} are serially uncorrelated.

- \(\rho_\ell\) is a population moment which is typically unknown, but can be consistently estimated by its sample counterpart. The lag-ℓ sample ACF is

  \[
  \hat{\rho}_\ell = \frac{\hat{\gamma}_\ell}{\gamma_0}, \quad \text{with} \quad \hat{\gamma}_\ell = \frac{1}{T} \sum_{t=\ell+1}^{T} (r_t - \bar{r}) (r_{t-\ell} - \bar{r})
  \]

  \[
  \bar{r} = \frac{1}{T} \sum_{t=1}^{T} r_t.
  \]

- The sample ACF is an important instrument to analyze time series:

  - useful to find an appropriate model
  
  - useful to estimate model parameters (Yule-Walker equations).

- Figure 5 and 6 show the sample ACF of monthly simple and Log-returns of IBM stock and of the value weighted index of U.S. Markets form January 1926 to December 1997.
Figure 5: Sample autocorrelation functions of monthly simple and log returns of IBM stock from January 1926 to December 1997. In each plot, the two horizontal lines denote two standard-error limits of the sample ACF. (T02)

Figure 6: Sample autocorrelation functions of monthly simple and log returns of the value-weighted index of U.S. Markets from January 1926 to December 1997. In each plot, the two horizontal lines denote two standard-error limits of the sample ACF. (T02)
- We will come back to the ACF in the next chapter when discussing market efficiency. Under the assumption of market efficiency asset returns should not be predictable and should not exhibit a significant autocorrelation.

- Now, let’s turn to a brief review to the most prominent linear time series models, used to capture the dynamic properties of a time series.

**White Noise (WN)**

- The white noise process is an essential ingredient of linear time series models and is defined as follows:

{\{\varepsilon_t\} is called a \textbf{white noise} \iff \varepsilon_t \sim \text{iid} with E[\varepsilon_t] = \mu_\varepsilon < \infty and \text{var}[\varepsilon_t] = \sigma_\varepsilon^2 < \infty.}

Hence, a WN-process is supposed to have no dynamic structure in its behavior.

- Based on a WN, a \textbf{linear time series} can be written as

\[ r_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}; \quad \varepsilon_t \sim \text{WN} \left(0, \sigma_\varepsilon^2\right), \]

\[ \psi_0 \equiv 1 \]

\{\psi_i\} = “\psi\text{-weights of } r_t”

This representation of a linear time series implies the following:

\[ \text{E} [r_t] = \mu \]

\[ \text{var} [r_t] = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2 \]

\[ \gamma_\ell = \text{E} \left[ (r_t - \mu) (r_{t-\ell} - \mu) \right] \]

\[ = \text{E} \left[ \left( \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \right) \left( \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-\ell-i} \right) \right] \]

\[ = \text{E} \left[ \sum_{i,j=0}^{\infty} \psi_i \psi_j \varepsilon_{t-i} \varepsilon_{t-\ell-j} \right] \]

\[ = \sum_{j=0}^{\infty} \psi_{j+\ell} \psi_j \text{E} \left[ \varepsilon_{t-\ell-j}^2 \right] \quad \text{(since expectations of cross terms are 0)} \]

\[ = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+\ell} \]

\[ \rho_\ell = \frac{\gamma_\ell}{\gamma_0} = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+\ell}}{1 + \sum_{j=1}^{\infty} \psi_j^2} \quad \ell \geq 0. \]
Hence, the linear dynamic structure is governed by the weights $\psi_j$ and linear time series models are structural models used to describe the pattern of the $\psi$-weights.

### AR–models

The simplest version of an autoregressive (AR) model is the AR(1) model:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim \text{WN} \left(0, \sigma^2_{\varepsilon}\right).$$

- It’s a stochastic difference equation of order 1;
- has the form of a linear regression model with a linear regression function:
  $$E \left[ r_t \mid r_{t-1} \right] = \phi_0 + \phi_1 r_{t-1}$$
  
  and
  $$\text{var} \left[ r_t \mid r_{t-1} \right] = \sigma^2_{\varepsilon};$$

- can be written as (for $|\phi_1| < 1$)

$$\begin{align*}
(1 - \phi_1 L) r_t &= \phi_0 + \varepsilon_t \\
  r_t &= \frac{\phi_0}{1 - \phi_1} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i},
\end{align*}$$

where $L$ is the lag operator defined as $L^\ell x_t \equiv x_{t-\ell}$.

#### Properties of the AR(1)

Suppose that $\{r_t\}$ under the AR(1) is weakly stationary with $E[r_t] = \mu$, $\text{var}[r_t] = \gamma_0$ and $\text{cov}(r_t, r_{t-1}) = \gamma_\ell$, then:

- the unconditional mean is obtained as

$$E \left[ r_t \right] = \phi_0 + \phi_1 E \left[ r_{t-1} \right], \quad \text{with} \quad E \left[ r_t \right] = E \left[ r_{t-1} \right].$$

Hence:

$$E \left[ r_t \right] = \mu = \frac{\phi_0}{1 - \phi_1}, \quad \phi_1 \neq 1.$$

- Furthermore, using $\phi_0 = \mu \left(1 - \phi_1\right)$, the AR(1) can be written as

$$r_t - \mu = \phi_1 \left(r_{t-1} - \mu\right) + \varepsilon_t$$

$$= \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}. $$
- The unconditional variance is obtained by taking expectation of the squares on both sides:

\[
E \left[ (r_t - \mu)^2 \right] = \phi_1^2 E \left[ (r_{t-1} - \mu)^2 \right] + E \left[ \varepsilon_t^2 \right] + 2E \varepsilon_t (r_{t-1} - \mu) = 0
\]

\[
\gamma_0 = \phi_1^2 \gamma_0 + \sigma^2 \varepsilon
\]

\[
\gamma_0 = \frac{\sigma^2 \varepsilon}{1 - \phi_1^2}, \quad \phi_1^2 < 1.
\]

- Hence, for weak stationarity the AR(1) has to satisfy:

\[
|\phi_1| < 1 \Leftrightarrow m - \phi_1 = 0,
\]

where \( m \) is the solution of the characteristic equation of the AR(1).

- A special AR(1) is the random walk:

\[
r_t = \phi_0 + 1 \cdot r_{t-1} + \varepsilon_t,
\]

with \( m = 1 \) (unit root). Note that in this specification the stationarity conditions are violated.

- The ACF of a stationary AR(1) can be obtained as follows:

\[
\gamma_\ell = E [(r_t - \mu) (r_{t-\ell} - \mu)]
\]

\[
= E \left\{ \phi_1 (r_{t-1} - \mu) + \varepsilon_t \right\} (r_{t-\ell} - \mu)
\]

\[
= \phi_1 E [(r_{t-1} - \mu) (r_{t-\ell} - \mu)] + 0
\]

\[
= \phi_1 \gamma_{\ell-1}.
\]

This is a non-homogenous first-order difference equation with

\[
\gamma_1 = \phi_1 \gamma_0 = \frac{\phi_1 \sigma^2 \varepsilon}{1 - \phi_1^2}.
\]

Repeated substitution yields

\[
\gamma_\ell = \phi_1^\ell \gamma_0.
\]

Hence, the ACF is obtained as

\[
\rho_\ell = \frac{\gamma_\ell}{\gamma_0} = \phi_1^\ell, \quad \ell \geq 0.
\]

Note that this implies an exponential decay of the ACF.

- Figure 7 shows the ACF of an AR(1)-process with \( \phi = 0.8 \) and \( \phi_1 = -0.8 \).
Figure 7: The autocorrelation function of an AR(1) model: (a) for $\phi_1 = 0.8$ and (b) for $\phi_1 = -0.8$. (T02)
**AR(2) Model**

The AR(1) can be generalized by including a further lag:

\[ r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN} \left( 0, \sigma^2 \right). \]

Based on the same considerations as for the AR(1) one can obtain under the assumption of weak stationarity:

- **Unconditional mean:**
  \[ E[r_t] = \mu = \frac{\phi_0}{1 - \phi_1 - \phi_2}, \quad \phi_1 + \phi_2 \neq 1; \]

- based on \( \phi_0 = \mu (1 - \phi_1 - \phi_2) \) the AR(2) can be “centered”:
  \[ r_t = \phi_1 (r_{t-1} - \mu) + \phi_2 (r_{t-2} - \mu) + \varepsilon_t; \]

- Multiplying the last equation by \((r_t - \mu)\) and taking the expectation, yields the following form for the autocovariance:
  \[ E[(r_t - \mu)^2] = \gamma_\ell = \phi_1 \gamma_{\ell-1} + \phi_2 \gamma_{\ell-2}, \quad \ell > 0. \]

- Hence, the **ACF** \( \rho_\ell = \gamma_\ell / \gamma_0 \) is the following second-order difference equation:
  \[ \rho_\ell = \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-2}, \]

with starting values:

\[ \rho_0 = 1 \quad \text{and} \quad \rho_1 = \phi_1 \rho_0 + \phi_2 \rho_{-1} \Leftrightarrow \rho_1 = \frac{\phi_1}{1 - \phi_2}, \]

since \( \rho_1 = \rho_{-1} \) by construction.

- The characteristic equation of this difference equation for \( \rho_\ell \) (and hence for the AR(2)) is
  \[ b^2 - \phi_1 b - \phi_2 = 0, \]

with solutions (characteristic roots):

\[ m_{1/2} = \frac{\phi_1}{2} \pm \frac{1}{2} \sqrt{\phi_1^2 + 4 \phi_2}. \]

Note that:

\[ \phi_1^2 + 4 \phi_2 \geq 0 \quad \Rightarrow \quad \text{"real roots"} \]

\[ \phi_1^2 + 4 \phi_2 < 0 \quad \Rightarrow \quad \text{"complex roots"}. \]
- The general form of the solution of the difference equation is:

\[ \rho_\ell = A_1 m_1^\ell + A_2 m_2^\ell, \quad \ell = 0, 1, \ldots, \]

with \( A_1, A_2 \) being coefficients depending on \( \rho_0 \) and \( \rho_1 \). Hence, stationarity for the AR(2) requires that the absolute values of the characteristic roots \( m_{1/2} \) are less than one. This ensures that

\[ \lim_{\ell \to \infty} \rho_\ell = 0. \]

- Figure 8 shows the ACF of four stationary AR(2) models. Panel (b) shows the ACF of a model with complex roots. Hence, its ACF exhibits damping sine and cosine waves.
Figure 8: The autocorrelation function of an AR(2) model: (1) \( \phi_1 = 1.2 \) and \( \phi_2 = 0.35 \), (b) \( \phi_1 = 0.6 \) and \( \phi_2 = -0.4 \), (c) \( \phi_1 = 0.2 \) and \( \phi_2 = 0.35 \), (d) \( \phi_1 = -0.2 \) and \( \phi_2 = 0.35 \). (T02)
**AR(p) Model**

The results for the AR(1) and AR(2) can be generalized to a general AR(p) model:

\[ r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim WN \left( 0, \sigma^2 \right). \]

**Identifying AR–Models**

In applications, the order \( p \) of the AR(p) is typically unknown. However, \( p \) must be known to estimate the model.

Two general approaches are available for determining the order \( p \).

- The selection of \( p \) based on information criteria such as the **Akaike Information Criterion**:  

\[ AIC = -\frac{2}{T} \cdot \ln (\text{likelihood}) + \frac{2}{T} \cdot (\# \text{ of parameters}), \]

where the likelihood is evaluated at the ML-estimates. \( p \) is selected such that AIC is minimized. Note, that AIC penalizes models with a large number of parameters (second term of AIC) and rewards a good fit (first term of AIC).

- Another possibility to select \( p \) is to use the **partial ACF (PACF)**:

This is a kind of F-Test evaluating the additional explanatory power obtained by including a further lag (for details, see T02, p. 36f).

**Estimation of AR–Models**

- Note that the AR(p) has a form of a linear regression model. Hence, it can be estimated by the **least squares (LS)** method.

- In particular, conditioning on the first \( p \) observations, we have:

\[ r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \varepsilon_t, \quad t : p + 1 \rightarrow T. \]

Then, the LS–estimator is

\[ \hat{\phi} = \left( \hat{\phi}_0, \ldots, \hat{\phi}_p \right)' = \left( X'X \right)^{-1} X'y \]

with

\[ X = \begin{pmatrix} 1 & r_p & \cdots & r_1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & r_{T-1} & \cdots & r_{T-p} \end{pmatrix}, \quad y = \left( r_{p+1}, \ldots, r_T \right)'. \]
- Note that under a Gaussian error $\varepsilon_t$, the LS-estimator $\hat{\phi}$ corresponds to the conditional ML-estimator based on the conditional joint pdf of the $r_t$'s given their first $p$ observations:

\[
f (r_{p+1}, \ldots, r_T \mid r_1, \ldots, r_p, \theta),
\]

where $\theta$ denotes the parameter vector.

- Alternatively, one can use the unconditional ML approach based on the unconditional joint pdf:

\[
f (r_1, \ldots, r_T; \theta) = f (r_{p+1}, \ldots, r_T \mid r_1, \ldots, r_p; \theta) \cdot f (r_1, \ldots, r_p; \theta).
\]

Note that is a non-linear estimation problem.

Forecasts using AR-Models

An important application of time series analysis is forecasting.

- Denote the future value to be predicted by

\[
r_{h+\ell}; \quad h = \text{current period}
\]

\[
\ell = \text{forecast horizon}.
\]

- A forecast is some transformation $g(\cdot)$ of the information $I_h$ available at time $h$.

- To evaluate the quality of the forecast a quadratic loss function of the form

\[
E \left[ (r_{h+\ell} - g (I_h))^2 \right]
\]

is often used.

Hence, the best forecast $g$ based on this loss function minimizes the expected quadratic difference (MSE, mean squared error) between the predicted value based on a function $g$ and the true value.

- It is easy to show that the best (MSE-minimizing) forecast $\hat{r}_{h+\ell}$ is

\[
\hat{r}_{h+\ell} = E [r_{h+\ell} \mid I_h] = \arg \min_{g(\cdot)} E \left[ (r_{h+\ell} - g (I_h))^2 \right].
\]

- For $\ell = 1$ and an AR($p$), we obtain, e.g.:

\[
\hat{r}_{h+1} = E [r_{h+1} \mid r_h, r_{h-1}, \ldots] = \phi_0 + \phi_1 r_h + \cdots + \phi_p r_{h-p}.
\]

The corresponding forecast error is:

\[
\hat{\varepsilon}_{h+1} = r_{h+\ell} - \hat{r}_{h+\ell}, \quad \text{with} \quad \text{var} [\hat{\varepsilon}_{h+1}] = \sigma_\varepsilon^2,
\]

where $\sigma_\varepsilon^2$ measures the uncertainty associated with the forecast.
- In practice, the parameters $\phi_i$ are unknown and are typically substituted by their estimates. Note that this is a further source of uncertainty associated with forecasts.

- For a detailed discussion of multistep-ahead forecasts ($\ell > 1$), see T02 (page 40). Here, I only note that multistep-ahead forecast for stationary time series have the property:

$$\lim_{\ell \to \infty} \hat{r}_{h+\ell} = E [r_t] = \mu. $$

### MA–models

We now turn to the class of moving-average (MA) models. The simplest version is the MA(1):

$$r_t = c_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1}; \quad \varepsilon_t \sim \text{WN} \left(0, \sigma^2_\varepsilon \right)$$

- It's a weighted average of shocks $\varepsilon_t$ and $\varepsilon_{t-1}$.

- It can be written as an AR($\infty$):

$$r_t = c_0 + \varepsilon_t \left(1 - \theta_1 L\right)$$

$$\frac{r_t}{1 - \theta_1 L} = \frac{c_0}{1 - \theta_1} + \varepsilon_t, \quad |\theta_1| < 1$$

$$\sum_{i=0}^{\infty} \theta^i r_{t-i} = \frac{c_0}{1 - \theta_1} + \varepsilon_t$$

$$r_t = \frac{c_0}{1 - \theta_1} - \theta_1 r_{t-1} - \theta^2_1 r_{t-2} - \cdots + \varepsilon_t.$$ 

- The MA(1) has the following stochastic properties

$$E [r_t] = c_0$$

$$\text{var} [r_t] = \sigma^2_\varepsilon \left(1 + \theta^2_1 \right)$$

$$\text{cov} (r_t, r_{t-\ell}) = E \left[ (\varepsilon_t - \theta_1 \varepsilon_{t-1}) (\varepsilon_{t-\ell} - \theta_1 \varepsilon_{t-\ell-1}) \right]$$

$$= \begin{cases} -\theta_1 \sigma^2_\varepsilon, & \text{if } \ell = 1 \\ 0, & \text{else} \end{cases}$$

$$p(\ell) = \begin{cases} -\frac{\theta_1}{1+\theta^2_1}, & \text{if } \ell = 1 \\ 0, & \text{else} \end{cases}.$$  

Hence, the ACF cuts off at lag 1. (This property generalizes to higher–order MA models.) Consequently, an MA(1) series is only linearly related to its first lagged value.
- Furthermore, note that the MA(1) is always stationary - independently of the \( \theta_1 \)-value - because it's a finite linear combinations of a WN-sequence.

**MA(1) Model**

The MA(1) is obtained by including \( q \) lagged WN-terms:

\[
    r_t = c_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}; \quad \varepsilon_t \sim \text{WN} \left(0, \sigma^2_\varepsilon\right).
\]

- The moments are obtained analogously to the MA(1):

\[
    \mathbb{E}[r_t] = c_0 \\
    \text{var}[r_t] = \sigma^2_\varepsilon \left(1 + \theta^2_1 + \cdots + \theta^2_q\right) \\
    \text{cov}(r_t, r_{t-\ell}) = \left\{
        \begin{array}{ll}
            \sigma^2_\varepsilon \left[-\theta_\ell + \theta_{\ell+1}\theta_1 + \cdots + \theta_q\theta_{q-\ell}\right], & \text{if } \ell \leq q \\
            0, & \text{else}
        \end{array}
    \right.
\]

Hence, the MA(1) has a “memory” of \( q \) lags and is always stationary.

**Estimation of MA-Models**

MA-models can be estimated by ML-estimation. Such an estimation is based on a recursive evaluation of the shock for a given parameter value \( \theta \) and starting values \( \varepsilon_0, \varepsilon_{-1}, \ldots \).

- For an MA(1) with \( \varepsilon_0 \equiv 0 \), e.g., we obtain

\[
    r_1 = c_0 + \varepsilon_1 \iff \varepsilon_1 = r_1 - c_0 \\
    r_2 = c_0 + \varepsilon_2 - \theta_1 \varepsilon_1 \iff \varepsilon_2 = r_2 - c_0 + \theta_1 \varepsilon_1, \quad \text{and so on.}
\]

(for further details, see T02, p.46)

**Forecasting using MA-Models**

Lets focus on the best 1-step-ahead forecast based on a MA(\( q \)) (best in the sense of minimizing a quadratic loss function).

- The random variable to be forecasted is:

\[
    r_{h+1} = c_0 + \varepsilon_{h+1} - \theta_1 \varepsilon_h - \cdots - \theta_p \varepsilon_{h+1-q}
\]

The best forecast is:

\[
    \hat{r}_{h+1} = \mathbb{E}[r_{h+1} | r_h, r_{h-1}, \ldots] = c_0 + 0 - \theta_1 \varepsilon_h - \cdots - \theta_p \varepsilon_{h+1-q},
\]

where the sequence \( \{\varepsilon_t, t \leq h\} \) can be calculated recursively.
- For details of the multi-step-ahead forecast, see T02, (page 46). Finally, note that

\[ \hat{r}_{h+\ell} = E[r_t] = c_0 \forall \ell > q, \]

which reflects the fact that the MA(q) has a memory of only q lags.

**ARMA–Models**

We now turn to the combination of AR– and MA–components leading to the ARMA–models. The general form of an ARMA\((p,q)\) is given by

\[ r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \]

or more compact

\[ (1 - \phi_1 L - \cdots - \phi_p L^p) r_t = \phi_0 + (1 - \theta_1 L - \cdots - \theta_q L^q) \varepsilon_t \]

\[ \phi(L)r_t = \phi_0 + \theta(L)\varepsilon_t, \]

where \(\phi(L)\) and \(\theta(L)\) are the so called lag–polynomials.

- The ARMA–model is stationary if the AR–part defines a stationary AR–model (i.e., if the absolute values of the solutions of the characteristic equation are less than 1), whereas the MA–part is irrelevant. For a discussion of the moments, the identification of the order \((p,q)\) and forecasting, see T02 (page 49).

- We now turn to the AR and MA–representation of ARMA–models.

The **AR**–representation is obtained as follows (provided that \(1/\theta(L)\) exists):

\[ \phi(L)r_t = \phi_0 + \theta(L)\varepsilon_t \]

\[ \frac{\phi(L)}{\theta(L)} r_t = \frac{\phi_0}{1 - \theta_1 - \cdots - \theta_q} + \varepsilon_t \]

\[ r_t = \frac{\phi_0}{1 - \theta_1 - \cdots - \theta_q} + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \cdots + \varepsilon_t, \]

where the \(\pi_i\)'s are obtained by the long division of the lag–polynomials \(\phi(L)/\theta(L)\). This representation shows the dependence of current \(r_t\) on its past values.

The **MA**–representation is (provided that \(1/\phi(L)\) exists):

\[ r_t = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p} + \frac{\theta(L)}{\phi(L)} \varepsilon_t \]

\[ = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p} + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots, \]

where the \(\psi_i\)'s are obtained by the long division \(\theta(L)/\phi(L)\). This representation shows the impact of past shocks on current \(r_t\). The sequence \(\{\psi_i\}\) is called as the impulse response function of the ARMA.