Modeling of stock return correlation

Sander Sepp

August 29, 2011

Master’s Thesis

Supervisors:
Dr. Peter Spreij, Prof. Michel Vellekoop, MSc. Geeske Vlaming

Universiteit van Amsterdam & All Options
# 5.2 Common Factor Model

- **5.2.1 Definition of the Common Factor Model**  
- **5.2.2 Properties of the Common Factor Model**

# 5.3 Single factor model

- **5.3.1 Background: Multiplicative indexes**
- **5.3.2 Summarizing the single-factor model**

# 5.4 Conclusion

---

# 6 Correlation estimation using daily data

- **6.1 Overview of estimators**
  - **6.1.1 Classical estimator**
  - **6.1.2 Rogers’ estimator**
  - **6.1.3 Sign-based estimator**
- **6.2 Simulation study of correlation estimators**
- **6.3 Conclusions on estimators**

---

# 7 Study of correlation in the DAX and AEX index

- **7.1 Mean index correlation**

---

# 8 Fitting non-stochastic correlation models

- **8.1 Models**
- **8.2 Model fit**
- **8.3 Graphs of the fits**
- **8.4 Discussion of results**

---

# 9 Fitting a stochastic correlation model to daily price data

- **9.1 Introduction**
- **9.2 Model definition**
- **9.3 Parameter estimation**
9.3.1 Parameter estimation using sample correlation . . . . . 66
9.3.2 Parameter estimation using the EM-algorithm . . . . . 67
9.3.3 Parameter estimation using correlation as parameters 70
9.3.4 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . 70

10 Conclusion 71
Chapter 1

Introduction

To understand the behavior of a stock index, we must understand the relation between the price movements of its constituent stocks. One way to measure the relation between random variables is through their correlation, which measures linear dependence. We will study the correlation between the returns of stock prices. This should in turn lead to an improved understanding of the volatility of the stock index, which is highly relevant for index option pricing.

According to market experience, correlation between stock returns changes over time. In principle it could change continuously over time, but we can not estimate correlation continuously. We can overcome this problem by assuming correlation to be constant over certain time periods.

In each time period, there is a correlation between the returns of each pair of stocks in the index. If there are $n$ stocks in the index, there are $d = n(n-1)/2$ pairs of stocks, each with a potentially different correlation. For example, in the AEX index we have $n = 25$ and $d = 300$. When viewed this way, index correlation becomes a $d$-dimensional time series. Alternatively, we could consider it to be a $d$-dimensional continuous process.

However, not any $d$-dimensional time series is a valid candidate for a correlation model. At any time period, the different correlations need to be consistent with each other. Mathematically, the correlation matrix needs to be positive semidefinite. Defining a $d$-dimensional model in which at any time the corresponding correlation matrix is positive semidefinite is not an easy task, although several such models are known. However, the fact that a model is consistent alone does not imply that it is suitable for modeling stock return correlation.
Instead of modeling a $d$-dimensional time series, we can also simplify the situation by assuming that the correlations between all pairs of stock are equal to each other, at all times. This reduces the problem to a one-dimensional time series. This simplification is unrealistic, but might be acceptable for the purposes of predicting index volatility.

In this thesis we search for correlation models that are simple enough to be manageable, while accurate enough to be usable for formulating trading strategies.

In chapter two we look at some results concerning correlation between random variables. We start with basic knowledge, but also include some more advanced topics that are relevant to our subject. In chapter three we look at correlation between random processes. In particular we show how to define correlation between random processes, and how to construct random processes that are correlated. In chapter four we look at some of the known properties of correlation between stock returns, and some of the difficulties associated with modeling such processes. In chapter five we take a look at several models that can be used for modeling stock returns. In chapter six we compare different ways of estimating correlation between stock returns. In chapter seven we study correlation in the DAX and AEX index. In chapter eight we fit correlation models without a stochastic term to data from the DAX and AEX index. In chapter nine we study how to fit correlation models with a stochastic term. Chapter ten is the conclusion.
Chapter 2

Covariance and correlation between random variables

In this chapter we define covariance and correlation between random variables, and look at some results that will be useful later.

2.1 Basics

Let \( E \) denote expectation. The variance of a random variable \( X \) is defined as

\[
\text{var}(X) = E[(X - EX)^2],
\]

the covariance between random variables \( X \) and \( Y \) is defined as

\[
\text{cov}(X, Y) = E[(X - EX)(Y - EY)],
\]

and the correlation between \( X \) and \( Y \) is defined as

\[
\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}, \quad \text{if } \text{var}(X) > 0, \text{ var}(Y) > 0.
\]

For a random vector \( X \) we define the covariance matrix \( \text{var}(X) \) by \( \text{var}(X) = E[(X - EX)(X - EX)^\top] \), so \( \text{var}(X)_{i,j} = \text{cov}(X_i, X_j) \). Similarly to the covariance matrix, \( \text{corr}(X) \) denotes the correlation matrix.

Remark. It follows that if \( X \) is a random vector with \( \Sigma = \text{var}(X) \), \( C = \text{corr}(X) \), and \( D \) is the diagonal matrix given by \( D_{ii} = \sqrt{\text{var}(X_i)} \), then \( \Sigma = DC\bar{D} \) and \( C = D^{-1}\Sigma D^{-1} \).
We can generalize the preceding as follows. For random vectors $X$ and $Y$ the cross-covariance matrix $\text{cov}(X,Y)$ is defined by $\text{cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)^\top]$, so $\text{cov}(X,Y)_{ij} = \text{cov}(X_i,Y_j)$. The following theorem gives some important basic properties of (cross-)covariance matrices.

**Theorem 2.1.** If $X$ is an $n$-dimensional random vector, $a$ is an $n$-dimensional (non-random) vector, and $A$ is an $m \times n$ matrix, then

$$\text{var}(AX + a) = A\text{var}(X)A^\top.$$ 

If $X$ and $Y$ are random vectors of equal dimension, then

$$\text{var}(X + Y) = \text{var}(X) + \text{cov}(X,Y) + \text{cov}(Y,X) + \text{var}(Y).$$

**Proof.** These are basic results, see e.g. [van der Vaart, 2000].

**Corollary 2.2.** If $X$ and $Y$ are independent random vectors of equal dimension, then

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

In the following we assume that the matrices take their values in the reals. A symmetric matrix is a square matrix that is equal to its transpose.

**Definition 2.3.** A symmetric matrix is called positive definite if for all $n$-dimensional non-zero vectors $x$ it holds that $x^\top Cx > 0$. It is called positive semidefinite if for all such vectors $x$ it holds that $x^\top Cx \geq 0$.

There are many characterizations of positive (semi)definiteness. A list of these is given in [Bhatia, 2007]. We will use the following two characterizations.

**Lemma 2.4.** A symmetric matrix is positive definite if and only if all its eigenvalues are positive. It is positive semidefinite if and only if all its eigenvalues are non-negative.

**Lemma 2.5.** (Sylvester’s criterion) An $n$-by-$n$ symmetric matrix $M$ is positive definite if and only if $\det (M_{ij})^k_{i,j=1} > 0$ for $k = 1, \cdots, n$.

We call a matrix a covariance (resp. correlation) matrix if it is the covariance (resp. correlation) matrix of some random vector. We will investigate what matrices are covariance and correlation matrices.

**Lemma 2.6.** A covariance matrix is positive semi-definite.
Proof. Let \( X \) be a random vector with covariance matrix \( \text{var}(X) = V \), and let \( b \) be a vector. By Theorem 2.1 it follows that \( b^T V b = \text{var}(b^T \cdot X) \). Because a variance is always non-negative, it follows that \( b^T V b \geq 0 \). Because this holds for all vectors \( b \), it follows by definition that \( V \) is positive semidefinite.

The following theorem is useful for constructing random variables with a given covariance matrix.

**Theorem 2.7.** If \( A \) is a positive definite matrix, then there exists a unique lower triangular matrix \( L \) with strictly positive diagonal entries such that \( LL^T = A \). This is called the Cholesky decomposition of \( A \). If \( A \) is a positive semidefinite but not positive definite matrix, then there exists a (generally not unique) lower triangular matrix \( L \) such that \( LL^T = A \). This is also called a Cholesky decomposition.

Proof. The proof can be found in [Bau III and Trefethen, 1997], lecture 23.

From Lemma 2.6 and Theorem 2.7 it follows that every covariance matrix has a Cholesky decomposition. Using Theorem 2.1 it follows that if \( X \) is a random vector with \( \text{var}(X) = I_n \) (the \( n \)-dimensional identity matrix), and \( A \) is an \( n \)-by-\( n \) covariance matrix with Cholesky decomposition \( A = LL^T \), then \( \text{var}(LX) = A \). As a special case, if \( X \) is a vector of independent standard normal random variables, then \( LX \) is said to have a multivariate normal distribution with expectation zero and variance \( A \).

**Remark.** The marginal distributions and covariance matrix of a random vector do not uniquely determine the joint distribution of the vector. For example, let \( X \) be a standard normal random variable, and \( B \) a random variable (independent of \( X \)) with \( \mathbb{P}(B = 1) = \mathbb{P}(B = -1) = 1/2 \). We define a random variable \( Y \) by \( Y = BX \). By the symmetry of the normal distribution, \( Y \) also has a standard normal distribution. The random variables \( X \) and \( Y \) are clearly not independent, however they are uncorrelated:

\[
\text{cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X \mathbb{E}Y = \mathbb{E}XY - \mathbb{E}BX^2 = \mathbb{E}BEX^2 = 0 \cdot 1 = 0.
\]

Hence the vector \((X, Y)\) has the same marginal distributions and covariance matrix as a 2-dimensional vector of independent standard normal random variables, but has a different joint distribution.

**Corollary 2.8.** A matrix is positive semidefinite if and only if it is a covariance matrix.
Proof. We have seen in Lemma 2.6 that all covariance matrices are positive semidefinite. The converse statement follows from the above construction. 

**Theorem 2.9.** A matrix \( C \) is a correlation matrix if and only if the following two properties hold.

- \( C \) is positive semidefinite,
- \( C \) has a unit diagonal, i.e. \( C_{ii} = 1 \) for \( i = 1, \ldots, n \).

**Proof.** If \( C \) is positive semidefinite, then it is a covariance matrix. As shown above, it follows that there exists a random vector \( X \) such that \( \text{var}(X) = C \). Since \( \text{var}(X) \) has unit diagonal, it follows that \( \text{var}(X_i) = 1, i = \{1, \ldots, n\} \), hence \( \text{corr}(X) = \text{var}(X) = C \), so \( C \) is a correlation matrix.

To prove the converse, assume that \( C \) is the correlation matrix of some random vector \( X \). Then \( C \) is the covariance matrix of the random vector \( X' \) obtained by normalizing \( X \), hence \( C \) is positive semidefinite. By definition, every correlation matrix has a unit diagonal. 

### 2.2 Constant correlation matrices

The correlation matrices described in the following theorem can be useful in modeling, because of their simplicity.

**Theorem 2.10.** For \( \rho \in [0,1] \), let \( A \) be the \( n \times n \) matrix defined by \( A_{ii} = 1 \) for all \( i \), and \( A_{ij} = \rho \) for all \( i \neq j \). Then \( A \) is a correlation matrix.

**Proof.** By Theorem 2.9 it is sufficient to prove that such a matrix is positive definite. Let \( x \) be an \( n \)-dimensional non-zero vector. We will show that \( x^\top C x > 0 \). By a straightforward calculation it holds that

\[
x^\top C x = \sum_{i=1}^{n} x_i^2 + 2 \rho \sum_{i \neq j} x_i x_j.
\]

If the second term on the right is non-negative, we are done. If it is negative, then

\[
x^\top C x > \sum_{i=1}^{n} x_i^2 + 2 \sum_{i \neq j} x_i x_j = \left( \sum_{i=1}^{n} x_i \right)^2 > 0.
\]
This construction does not work for all $\rho \in [-1,0)$. In that case the matrix $A$ is not positive definite in general. For example, the 3-by-3 matrix with ones on the diagonal and -0.9 everywhere off the diagonal has a negative eigenvalue of -0.8, so this matrix is not positive definite.

### 2.3 Transforming a matrix to make it positive semidefinite

When estimating a correlation matrix, depending on the estimation procedure and available data, it can be the case that the resulting matrix is not positive semidefinite. In this scenario it might be desirable to find the ‘closest’ positive semidefinite matrix, relative to some error measure.

A general methodology to accomplish this is to use a parametrization of the set of all correlation matrices, and minimize the error measure over the parameters. A number of different parametrizations are described in [Pinheiro and Bates, 1996]. In [Rebonato and Jackel, 1999], this idea is applied to financial modeling. To explain the parametrization of Rebonato and Jackel we need the following.

**Definition 2.11.** The Gram matrix $G$ of a set of vectors $v_1, \cdots, v_n$ in a real inner product space is the matrix defined by $G_{ij} = \langle v_i, v_j \rangle$, where $\langle v_i, v_j \rangle$ denotes the inner product of $v_i$ and $v_j$.

Hence, in $n$-dimensional Euclidean space, the Gram matrix of $n$ vectors $x_1, x_2, \cdots, x_n \in \mathbb{R}^n$ is given by $G_{ij} = \sum_{k=1}^{n} x_{ik} x_{jk}$. Let $M$ be the $n$-by-$n$ matrix with columns $x_1, x_2, \cdots, x_n$, then $G = MM^\top$. If $X$ is a random vector with var($X$) = $I_n$ (the $n$-dimensional identity matrix), then var($MX$) = $MM^\top = G$, so $G$ is covariance matrix. We see that in Euclidean space, all Gram matrices are covariance matrices. The opposite also holds true, a matrix is a gram matrix if and only if it is a covariance matrix (see [Bhatia, 2007] for a proof).

If the vectors $x_1, \cdots, x_n$ all have norm 1, then the Gram matrix has a unit diagonal. From Theorem 2.9 it follows that this Gram matrix is a correlation matrix. This gives an onto mapping $(x_1, x_2, \cdots, x_n) \rightarrow C$ from tuples (ordered sets) of $n$ points on the unit sphere in $\mathbb{R}^n$ to $n$-dimensional correlation matrices.

We can parametrize the points on the unit sphere by using spherical coordinates. A point $x$ in $\mathbb{R}^n$ is on the unit sphere if and only if it is of the
form

\[ x_i = \cos(\theta_i)\prod_{j=1}^{i-1} \sin(\theta_j), \quad i = 1, 2, \ldots, n - 1, \]
\[ x_n = \prod_{j=1}^{n-1} \sin(\theta_j), \]

where \(0 \leq \theta_i \leq \pi\) for \(i = 1, \ldots, n - 2\) and \(0 \leq \theta_{n-1} < 2\pi\).

Through the mapping described above, this is also a parametrization of the set of \(n\)-dimensional correlation matrices. According to [Pinheiro and Bates, 1996], if we further restrict the parameters to \(0 < \theta_i < \pi, i = 1, \ldots, n - 1\), then the mapping is one-to-one to the set of positive definite matrices.

An important reason why this is a useful parametrization of correlation matrices is that it is unconstrained, meaning that the parameters \(\theta_i\) can vary freely inside the given intervals. In contrast, if we would parametrize the set of correlation matrices \(C\) directly by their values \(C_{ij}, i < j\) above the diagonal, under the constraint that the resulting matrix is positive definite, this would be a parametrization with nonlinear constraints on the parameters. This kind of parametrization is much more difficult to work with, since it is not a priori obvious what parameter values are allowed.

To be able to use these ideas to find a 'closest' correlation matrix \(\hat{M}\) to a given non-positive semidefinite matrix \(M\), we need to define an error measure \(\epsilon(\cdot)\) on the set of matrices, so that we can minimize \(\epsilon(M - \hat{M})\). For instance, we could use \(\epsilon(N) = \text{trace}(N^\top N) = \sum_{i=1}^{n} \sum_{j=1}^{n} N_{ij}^2\).

Let \(M(\theta)\) denote the correlation matrix with parametrization \(\theta_1, \theta_2, \ldots, \theta_{n-1}\). We find \(\hat{M}\) by the following algorithm.

\[
\hat{\theta} = \arg\min_{\theta} \epsilon(M - M(\theta)),
\]
\[
\hat{M} = M(\hat{\theta}).
\]

The minimization can be sped up by a well-chosen initial guess for \(\theta\). [Rebonato and Jackel, 1999] describe how such a guess can be made.
Chapter 3

Correlation between stochastic processes

While in the last chapter we looked at correlation between random variables, in this chapter we will study correlation between stochastic processes. More precisely, we will study correlation between the returns of stochastic processes. This theory is relevant if we want to use continuous time models to model correlation. As we shall see, things get more complicated in continuous time. Much of this chapter is based on [Boortz, 2008].

3.1 Defining correlation between stochastic processes

Definition 3.1. Let $S_1$ and $S_2$ be two stochastic processes (e.g. stock price processes), and $0 < t_1 < \cdots < t_N$ a grid of time points on which $S_1$ and $S_2$ are observed. The realized pairwise correlation over period $[0, T]$ of $S_1$ and $S_2$ is defined as

$$\rho_R(S_1, S_2) = \frac{\sum_{i=1}^{N} \left( \log \frac{S_{1,t_i}}{S_{1,t_{i-1}}} - \hat{S}_1 \right) \left( \log \frac{S_{2,t_i}}{S_{2,t_{i-1}}} - \hat{S}_2 \right)}{\sqrt{\sum_{i=1}^{N} \left( \log \frac{S_{1,t_i}}{S_{1,t_{i-1}}} - \hat{S}_1 \right)^2 \sum_{i=1}^{N} \left( \log \frac{S_{2,t_i}}{S_{2,t_{i-1}}} - \hat{S}_2 \right)^2}}, \quad (3.1)$$

with

$$\hat{S}_i = \frac{1}{N} \sum_{i=1}^{N} \log \frac{S_{t_i}}{S_{t_{i-1}}}.$$
Note that $\rho_R$ measures correlation between the log-returns of the assets, rather than correlation between the prices themselves. This is similar to the usage of the word ‘volatility’ of geometric Brownian motions in the context of the Black-Scholes model.

If the stochastic processes $S_1, S_2$ are continuous, then we can let the amount of points $N$ that the processes are observed on go to infinity. We will investigate what happens when the stochastic processes are geometric Brownian motions. By doing this, we will find how to construct Brownian motions that are correlated in a precise sense. We assume that $S_1, S_2$ satisfy

$$
\begin{align*}
    dS_{1,t} &= \mu_1 S_{1,t} dt + \sigma_1 S_{1,t} dW_{1,t}, \\
    dS_{2,t} &= \mu_2 S_{2,t} dt + \sigma_2 S_{2,t} dW_{2,t},
\end{align*}
$$

(3.2)

where $\mu_1, \mu_2, \sigma_1, \sigma_2$ are constants and $W_1, W_2$ are Brownian motions. These Brownian motions have quadratic covariation $\langle W_1, W_2 \rangle_t$, which will become important later.

We fix an interval $[0, T]$, and use the grid of time points $[0, \Delta, 2\Delta, \cdots, N\Delta = T]$. So if $N \uparrow \infty$, then $\Delta = T/N \downarrow 0$.

**Lemma 3.2.** The random variables $\log(S_{j,n\Delta}/S_{j,(n-1)\Delta})$ for $n = 1, \cdots, N$ are independent and have an $\mathcal{N}(\Delta(\mu_j - \sigma_j^2/2), \Delta \sigma_j^2)$ distribution.

**Proof.** The solution of SDE 3.2 is $S_{j,t} = \exp((\mu_j - \sigma_j^2/2)t + \sigma_j W_{j,t})$. Hence,

$$
\begin{align*}
    \log \left( \frac{S_{j,n\Delta}}{S_{j,(n-1)\Delta}} \right) &= \log(S_{j,n\Delta}) - \log(S_{j,(n-1)\Delta}) \\
    &= \left( \mu_j - \frac{1}{2} \sigma_j^2 \right) n\Delta + \sigma_j W_{j,n\Delta} \\
    &\quad - \left[ \left( \mu_j - \frac{1}{2} \sigma_j^2 \right) (n-1)\Delta + \sigma_j W_{j,(n-1)\Delta} \right] \\
    &= \left( \mu_j - \frac{1}{2} \sigma_j^2 \right) \Delta + \sigma_j (W_{j,n\Delta} - W_{j,(n-1)\Delta}).
\end{align*}
$$

The result now follows from the defining properties of Brownian motion. \qed

**Theorem 3.3.** Consider $\rho_R(S_1, S_2)$ calculated on the grid of time points $[0, \Delta, 2\Delta, \cdots, N\Delta = T]$. If $N \uparrow \infty$ (with $T$ fixed, and $\Delta = T/N$), then $\rho_R(S_1, S_2) \rightarrow \langle \log(S_1), \log(S_2) \rangle_t/\sigma_1 \sigma_2 T$ in probability, where $\langle \log(S_1), \log(S_2) \rangle_t$ denotes the quadratic covariation process of $\log(S_1)$ and $\log(S_2)$. 

13
Proof. First we will consider the asymptotics of the denominator of formula (3.1). We use the notation \( N_i, i = 1, \cdots, N \) to denote independent standard normally distributed random variables. From Lemma 3.2 it follows that
\[
\sum_{i=1}^{N} \left( \log \frac{S_{j,i\Delta}}{S_{j,(i-1)\Delta}} - \hat{S}_j \right)^2
\]
\[
= \sum_{i=1}^{N} \left( (\mu_j - \frac{1}{2} \sigma_j^2) \frac{T}{N} + \sqrt{\frac{T}{N}} \sigma_j N_i - \frac{1}{N} \sum_{k=1}^{N} \left( (\mu_j - \frac{1}{2} \sigma_j^2) \frac{T}{N} + \sqrt{\frac{T}{N}} \sigma_j N_k \right) \right)^2.
\]
If we work out the square then each term except for one converges to zero almost surely as \( N \uparrow \infty \). The only term that does not converge to zero is
\[
\sum_{i=1}^{N} \frac{T}{N} \sigma_j^2 N_i^2,
\]
which converges to \( T \sigma_j^2 \) almost surely as \( N \uparrow \infty \), by the strong law of large numbers. It follows that the denominator of Lemma 3.2 converges almost surely to \( T \sigma_1 \sigma_2 \) as \( N \uparrow \infty \).

Next we will consider the asymptotics of the numerator of the Lemma 3.2
\[
\sum_{i=1}^{N} \left( \log \frac{S_{1,n\Delta}}{S_{1,(n-1)\Delta}} \hat{S}_1 \right) \left( \log \frac{S_{2,n\Delta}}{S_{2,(n-1)\Delta}} \hat{S}_2 \right).
\]
Working out the product, the first term is equal to
\[
\sum_{i=1}^{N} \left( \log S_{1,n\Delta} - \log S_{1,(n-1)\Delta} \right) \left( \log S_{2,n\Delta} - \log S_{2,(n-1)\Delta} \right),
\]
which converges in probability to \( \langle \log S_1, \log S_2 \rangle_T \) (Protter, 2004, Chapter 2, Theorem 23), the quadratic covariation between \( \log S_1 \) and \( \log S_2 \).

The other three terms can be shown to converge to zero, either in probability or almost surely. It follows that the numerator of the formula for \( \rho_R(S_1, S_2) \) converges to \( \langle \log S_1, \log S_2 \rangle_T \) in probability. Combining this with the earlier result on the convergence of the denominator, it follows that \( \rho_R(S_1, S_2) \) converges to \( \frac{\langle \log(S_1), \log(S_2) \rangle_T}{\sigma_1 \sigma_2 T} \) in probability.

So we have the following analogue of correlation for stochastic processes satisfying 3.2.

14
Definition 3.4. The expression $\Psi_{0,T}(S_1, S_2) := \frac{\langle \log(S_1), \log(S_2) \rangle_T}{\sigma_1 \sigma_2 T}$ is called the model-implied realized correlation between the processes $S_1$ and $S_2$.

Lemma 3.5. Let $S_1$, $S_2$ be the geometric Brownian motions defined in formula (3.2). If $d \langle W_1, W_2 \rangle_t = \rho_t dt$, then $d \langle \log S_1, \log S_2 \rangle_t = \sigma_1 \sigma_2 \rho_t dt$.

**Proof.** Because $\log S_j,t = (\mu_j - \sigma_j^2/2) t + \sigma_j W_j,t$, it follows that

\[
d \langle \log S_1, \log S_2 \rangle_t = d \langle \sigma_1 W_1, \sigma_2 W_2 \rangle_t = \sigma_1 \sigma_2 d \langle W_1, W_2 \rangle_t = \sigma_1 \sigma_2 \rho_t dt.
\]

□

Corollary 3.6. If $d \langle W_1, W_2 \rangle_t = \rho_t dt$, then

\[
\Psi_{0,T}(S_1, S_2) = \frac{1}{T} \int_0^T \rho_t dt.
\]

This corollary suggests the following definition.

Definition 3.7. The process $\rho_t(S_1, S_2) := d/dt \langle W_1, W_2 \rangle_t$ is called the instantaneous correlation between processes $S_1$ and $S_2$.

Corollary 3.8. Let $S_1$, $S_2$ be Itô processes with instantaneous correlation $\rho_t$, and let $V = V(S_1, S_2, t)$ be a function of $S_1, S_2, t$ (e.g. representing the value process of a portfolio), then

\[
dV = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S_1} dS_1 + \frac{\partial V}{\partial S_2} dS_2 + \frac{1}{2} \sigma_1^2 S_1^2 \frac{\partial^2 V}{\partial S_1^2} dt + \rho_t \sigma_1 \sigma_2 S_1 S_2 \frac{\partial^2 V}{\partial S_1 \partial S_2} dt + \frac{1}{2} \sigma_2^2 S_2^2 \frac{\partial^2 V}{\partial S_2^2} dt.
\]

**Proof.** By Lemma 3.5 this follows from Itô’s lemma. □

### 3.2 Constructing correlated processes

The following theorem gives us a method to construct Itô processes with a given instantaneous pairwise correlation. This method is mentioned in van Emmerich, 2006.
Lemma 3.9. Let $\rho_t$ be a suitable Itô process, and let $W_1, V$ be independent Brownian motions. The process $W_2$ defined as

$$W_{2,t} = \int_0^t \rho_s dW_{1,s} + \int_0^t \sqrt{1 - \rho_s^2} dV_s$$

is a Brownian motion, and $d\langle W_1, W_2 \rangle_t = \rho_t dt$.

Proof. First we will show that $W_2$ is a Brownian motion. Note that a stochastic integral w.r.t. Brownian motion is a continuous martingale, so $W_2$ is a continuous martingale. We calculate the quadratic variation of $W_2$ to be

$$\langle W_2 \rangle_t = \left\langle \int_0^t \rho_s dW_{1,s} + \int_0^t \sqrt{1 - \rho_s^2} dV_s \right\rangle_t$$

$$= \left\langle \int_0^t \rho_s dW_{1,s} \right\rangle_t + \left\langle \int_0^t \sqrt{1 - \rho_s^2} dV_s \right\rangle_t$$

$$+ 2 \left\langle \int_0^t \rho_s dW_{1,s}, \int_0^t \sqrt{1 - \rho_s^2} dV_s \right\rangle_t$$

$$= \int_0^t \rho_s^2 d\langle W_1 \rangle_s + \int_0^t (1 - \rho_s^2) d\langle V \rangle_s + 2 \int_0^t \rho_s \sqrt{1 - \rho_s^2} d\langle W_1, V \rangle_s ds$$

$$= \int_0^t \rho_s^2 ds + \int_0^t (1 - \rho_s^2) ds = t,$$

where we used that $\langle W_1 \rangle_t = \langle V \rangle_t = t$ and $\langle W_1, V \rangle_t \equiv 0$. According to Lévy’s characterization of Brownian motion, any martingale $X$ with quadratic variation process $\langle X \rangle_t = t$ is a Brownian motion. Hence, $W_2$ is a Brownian motion. We calculate the quadratic covariation between $W_1$ and $W_2$ to be

$$\langle W_1, W_2 \rangle_t = \left\langle W_1, \int_0^t \rho_s dW_{1,s} + \int_0^t \sqrt{1 - \rho_s^2} dV_s \right\rangle_t$$

$$= \left\langle W_1, \int_0^t \rho_s dW_{1,s} \right\rangle_t + \left\langle W_1, \int_0^t \sqrt{1 - \rho_s^2} dV_s \right\rangle_t$$

$$= \int_0^t \rho_s d\langle W_1 \rangle_s + \int_0^t \sqrt{1 - \rho_s^2} d\langle W_1, V \rangle_s$$

$$= \int_0^t \rho_s ds,$$

so $d\langle W_1, W_2 \rangle_t = \rho_t dt$. □
Chapter 4

Characteristics and problems of asset return correlation modeling

In this chapter we look at various aspects related to asset return correlation modeling.

When we mention ‘index correlation’ in this chapter, this can be considered to mean the average correlation between all pairs of stocks in the index, weighted by their volatilities and index weights. In section 5.3.1 we explain why this is meaningful.

4.1 Stylized facts of correlation

There are certain patterns in asset return correlation that can be seen in many different markets. These are known as the stylized facts. These can be useful in correlation modeling, both as a source of inspiration for new models, and as a way to check how realistic a model is.

There is good overview of the stylized facts of correlation in [Boortz, 2008]. According to [Skintzi and Refenes, 2005], the most important features of asset correlation are

1. Correlation persistence and memory,

2. Correlation and volatility comovement,
3. Asymmetry in correlation.

Correlation persistence and memory means that asset return correlation significantly depends on its realizations in the last 20 to 30 trading days (e.g. if the correlation was high, it is more likely to stay high). This implies that correlation models should have positive autocorrelations for this duration.

Correlation and volatility comovement means that there is a positive relation between index correlation and index volatility. According to [Skintzi and Relenes, 2005], changes in index correlation impact future volatility, while changes in volatility do not impact future correlation.

Asymmetry in correlation means that there is a negative relation between stock returns and correlation. Index correlation strongly increases when the whole economy is doing poorly. When the economy is doing well, index correlation decreases, but this effect is not as strong.

It is interesting to note that the last two patterns imply that portfolio diversification is least effective when it is most needed. When markets are volatile or in a slump, correlation is higher, so diversification is less effective.

Another important property of correlation is its instability. According to [Wilmott, 2006] (pages 186, 195), correlations from financial time series data are notoriously unstable (highly volatile), and are even more unstable than volatilities.

4.2 Difficulties in asset return correlation modeling

4.2.1 Correlation is not directly observable

Correlation is not directly observable in the market, unlike e.g. stock returns. We have to estimate correlation using return data, and the estimation procedure introduces an error. This makes it more difficult to fit model parameters.

One approach is to pretend that the measured correlation is actually the real, underlying correlation. This is what we do in Chapter 8. This approach makes it easier to fit, but because it ignores a potentially significant distortion of the data, this could reduce the quality of resulting fit.
Another approach is to take the measuring error into account when fitting a model. This makes it much more complicated to perform the fit, but could result in better models. We take this approach in Chapter 9.

A workaround for this problem would appear to be to use high-frequency data. Using this type of data can give us hundreds (or thousands) of stock returns within a day, and a correlation estimator based on such a high amount of observations should have a low variance. However, this approach has its own problems.

One practical problem with intraday data is that such data can be costly to obtain. There are also be other, more fundamental, problems associated with the high-frequency approach. This subject has been studied in the case of volatility estimation, where it is found that such estimators can be biased by microstructure effects ([Alizadeh et al., 2002]), and lead to estimates that vary greatly in time ([Barndorff-Nielsen et al., 2010]). It seems reasonable that correlation estimates using high-frequency data has similar problems.

Another problem with the use of intraday data is caused by the fact that there is a difference between the closing price of a stock and the opening price on the next day. The opening price of a stock is determined by an auction held in the morning. This causes a jump in the stock price, which conflicts with the usual assumption that the stock price process is a continuous process (geometric Brownian motion). One interpretation of this jump is that the stock price is changing continuously but invisibly overnight. It is not obvious how to deal with this jump properly when using intraday data. One possibility is to disregard the jump and simply calculate correlation during the daytime, but this then introduces an unknown error.

4.2.2 High dimensionality

A stock index with $n$ stocks has $d = n(n - 1)/2$ pairs of stocks. For example, for the AEX index, $n = 25$ and $d = 300$, while for the S&P500, $n = 500$ and $d = 124750$. In principle, one could try to model the correlation between each pair of stocks in the index using a $d$-dimensional model, but the high dimensionality, combined with the fact that the correlations need to be consistent with each, makes this difficult.

One way to deal with this problem is to reduce the dimension of the model. In the Common factor model described in Section 5.2, a model is described that is $n$-dimensional (as opposed to $n(n - 1)/2$-dimensional). In Section 5.3.1, we describe a method that leads to a one-dimensional measure of index cor-
Another approach would be to first group stocks into baskets, for example grouped by industry, or by using mathematical techniques such as factor analysis. Then we can study return correlation between these baskets.

4.2.3 Requirement of positive semidefiniteness

As discussed in Section 2.1 all correlation (and covariance) matrices are positive-semidefinite. Although a model is not supposed to be true, but only an approximation of the truth, it seems likely that any good correlation model also implies positive semidefinite correlation matrices.

A big problem with matrices that are not-positive semidefinite is that, since they do not correspond to any actual process, is it not possible to simulate a process with such a matrix as 'correlation matrix'.

One way to sidestep this problem is to avoid modeling the correlations directly, instead modeling the stock price processes, and then taking the correlation model to be the correlation implied by the stock price model. In this case the correlation matrix will always be positive semidefinite, because it corresponds to an actual process.

Another solution is to use a model that has been designed to only produce positive definite matrices. An example of this is a Wishart process, described in e.g. [Gourieroux et al., 2009].

Yet another method would be to adjust a non-positive semidefinite matrix to make it positive semidefinite, for instance using the algorithm described in Section 2.3. A problem with this is that is not clear what kind of error the transformation introduces into the model. While this technique might be useful for a one-step prediction, it seems unlikely to be useful for a model spanning multiple time points.

4.2.4 Risk premium in stochastic correlation models

A stochastic correlation model is a model where an additional source of uncertainty (usually a Brownian motion) is introduced as part of the model. The introduction of the new source of uncertainty usually means that the market is no longer complete: not all derivative payoffs can be replicated by a dynamic self-financing portfolio consisting of stocks and bonds only.
This means that the principle of no arbitrage alone does not lead to unique prices. To make prices unique, one needs to define a risk premium for correlation, which could be called a market price of correlation risk. There is some discussion of these issues in the case of stochastic volatility models in [Boswijk, 2001].

These risk premiums pose a problem, because they have to be determined to be able to use the model for pricing. According to [Wilmott, 2006] (page 860), the risk premium for volatility risk is usually determined in practice by choosing the model parameters and the risk premium such that the model correctly prices certain exchange-traded options. Wilmott notes that it would be more theoretically sound to first calibrate a model of real (not implied) volatility based on the behavior of the stock price, and then to determine the volatility risk premium based on option prices.
Chapter 5

Asset return correlation models

5.1 A two-asset stochastic correlation model

The following model is described by [Ma, 2009]. It is a two-asset model with stochastic correlation. We model the assets under a martingale measure. We assume that, under the risk-neutral measure \( Q \), the assets \( S_1, S_2 \) are geometric Brownian motions with mean \( r \) (the risk-free interest rate) and constant volatilities \( \sigma_1 > 0, \sigma_2 > 0 \), with respect to Brownian motions \( W_1, W_2 \) satisfying

\[
dS_{1,t} = rS_{1,t}dt + \sigma_1 S_{1,t}dW_{1,t}, \\
dS_{2,t} = rS_{2,t}dt + \sigma_2 S_{2,t}dW_{2,t}.
\]

Let \( \rho_t \) be the instantaneous correlation between \( S_1 \) and \( S_2 \) at time \( t \). As discussed in Section 3.1, this means that \( d\langle S_1, S_2 \rangle_t = S_{1,t}\sigma_1 S_{2,t}\sigma_2 \rho_t dt \) and \( d\langle W_1, W_2 \rangle_t = \rho_t dt \). We assume that \( \rho_t \) satisfies the differential equation

\begin{equation}
\begin{aligned}
d\rho_t &= (\bar{\rho} - \beta \rho)dt + \sigma_3 \sqrt{(h - \rho)(\rho - l)}dW_{3,t},
\end{aligned}
\end{equation}

(5.1)

where \( W_3 \) is a Brownian motion, and \( \beta > 0, \sigma_3 > 0, h \in (-1, 1], l \in (-1, h), l < h, \, \bar{\rho} \in (l, h) \) are constants.

The process \( \rho \) is called a Jacobi Process. The term \((\bar{\rho} - \beta \rho)dt\) is a mean reversal term, and the factor \( \sqrt{(h - \rho)(\rho - l)} \) causes the process to stay within the bounds \([l, h]\), assuming that the parameters are appropriately chosen.

Let \( \varrho_1, \varrho_2 \in [-1, 1] \) be constants. We assume that \( W_1, W_2, W_3 \) are constructed
such that
\[ d \langle W_1, W_3 \rangle_t = \varrho_1 dt, \]
\[ d \langle W_2, W_3 \rangle_t = \varrho_2 dt, \]
\[ d \langle W_1, W_2 \rangle_t = \rho_t dt. \]

We will see later that not all choices for \( \varrho_1, \varrho_2 \) are allowed, because the equations of display (5.2) have to be satisfied.

**Lemma 5.1.** In the above model, if we choose the parameters such that
\[
(\bar{\rho} - \beta l) > \sigma_3^2(h - l)/2, \\
(\beta h - \bar{\rho}) > \sigma_3^2(h - l)/2,
\]
then the process \( \rho_t \) starting in the interval \((l, h)\) never hits the bounds of the interval.

**Proof.** Proof can be found in \[van Emmerich, 2006\].

The correlation matrix of the random vector \((W_{1,t}, W_{2,t}, W_{3,t})\) for some fixed \( t \geq 0 \) is
\[
V_t = \begin{pmatrix}
1 & \rho_t & \varrho_1 \\
\rho_t & 1 & \varrho_2 \\
\varrho_1 & \varrho_2 & 1
\end{pmatrix}.
\]

This matrix needs to be positive semidefinite for all values that \( \rho_t \) can assume, that is, for all \( l < \rho_t < h \). By Sylvester’s criterion (see Lemma 2.5), this matrix is positive definite if all of the following hold.

- \( 1 > 0 \),
- \( 1 - \rho^2 > 0 \),
- \( \det(V_t) > 0 \).

The second item follows from the requirement that \( l > -1 \) and \( h < 1 \). The determinant of \( V_t \) is
\[
\det(V_t) = 1 - \rho_t^2 - \varrho_1^2 - \varrho_2^2 + 2 \rho_t \varrho_1 \varrho_2.
\]
This is a quadratic function of $\rho_t$ which is positive if
\[
\begin{align*}
h &< \varrho_1\varrho_2 + \sqrt{(1 - \varrho_1)^2(1 - \varrho_2)^2}, \\
l &> \varrho_1\varrho_2 + \sqrt{(1 - \varrho_1)^2(1 - \varrho_2)^2}.
\end{align*}
\] (5.2)

So if $l, h$ satisfy these inequalities, then $V_t$ is positive definite.

Let $C$ be the value process (fair price) of a multi-asset derivative contingent on the stock price processes $S_1$ and $S_2$. We assume that the correlation process is visible to the market, so that $dC_t$ depends on $S_{1,t}, S_{2,t}$ and $\rho_t$.

We will derive a partial differential equation that $C$ must satisfy. We do this by constructing a riskless self-financing portfolio, which by the assumption of absence of arbitrage must have a return equal to the risk-free interest rate $r$. We can justify the assumption that there is no arbitrage on economic grounds. Alternatively, we can mathematically prove that arbitrage is not possible in our model.

**Theorem 5.2.** If $-1 < l < h < 1$, then there is no arbitrage in our model.

**Proof.** We can rewrite our model as
\[
\begin{align*}
\frac{dS_{1,t}}{S_{1,t}} &= r dt + \sigma_{1,t} dW_{1,t}, \\
\frac{dS_{2,t}}{S_{2,t}} &= r dt + \sigma_{2,t} \rho_t dW_{1,t} + \sigma_{2,t} \sqrt{1 - \rho_t^2} dW_{2,t},
\end{align*}
\]
where $W_{1,t}, W_{2,t}$ are independent Brownian motions. So
\[
\begin{pmatrix}
\frac{dS_{1,t}}{dS_{2,t}}
\end{pmatrix} = \mu dt + \sigma \begin{pmatrix}
\frac{dB_{1,t}}{dW_{2,t}}
\end{pmatrix},
\]
where
\[
\mu = \begin{pmatrix}
r S_{1,t} \\
r S_{2,t}
\end{pmatrix},
\]
and
\[
\sigma = \begin{pmatrix}
\sigma_{1,t} & 0 \\
\sigma_{2,t} \rho_t & \sigma_{2,t} \sqrt{1 - \rho_t^2}
\end{pmatrix}.
\]

According to theorem 12.1.8a from [Øksendal, 1998], there is no arbitrage if there is a solution $u(t) = (u_1(t), u_2(t))^T$ to the system $\sigma(t)u(t) = \mu(t)$ for which
\[
\mathbb{E}\left( \exp\left( \frac{1}{2} \int_0^T \left( u_1^2(t) + u_2^2(t) \right) dt \right) \right) < \infty. \tag{5.3}
\]
For our model we have \( u_1(t) = r/\sigma_1 \) and
\[
 u_2(t) = \frac{1}{\sqrt{1 - \rho^2}} \left( r - \frac{r}{\sigma_1} \rho \sigma_2 \right).
\]

If \(-1 < l < h < 1\) holds, then \( u_1(t) \) and \( u_2(t) \) are bounded processes, so then formula (5.3) holds and by the theorem from Øksendal there is no arbitrage.

First we use Itô’s lemma to derive an expression for \( dC \).
\[
dC = \frac{\partial C}{\partial t} \, dt + \frac{\partial C}{\partial S_1} \, dS_1 + \frac{\partial C}{\partial S_2} \, dS_2 + \frac{\partial C}{\partial \rho} \, d\rho \\
+ \frac{1}{2} \frac{\partial^2 C}{\partial S_1^2} \, d\langle S_1 \rangle_t + \frac{1}{2} \frac{\partial^2 C}{\partial S_2^2} \, d\langle S_2 \rangle_t + \frac{1}{2} \frac{\partial^2 C}{\partial \rho^2} \, d\langle \rho \rangle_t \\
+ \frac{\partial^2 C}{\partial S_1 \partial \rho} \, d\langle S_1, \rho \rangle_t + \frac{\partial^2 C}{\partial S_2 \partial \rho} \, d\langle S_2, \rho \rangle_t \\
= \left[ \frac{\partial C}{\partial t} + \frac{\partial C}{\partial S_1} r S_1 + \frac{\partial C}{\partial S_2} r S_2 + \frac{\partial C}{\partial \rho} (\overline{\rho} - \beta \rho) \\
+ \frac{1}{2} \frac{\partial^2 C}{\partial \rho^2} \sigma_3^2 (h - \rho)(\rho - l) + \frac{1}{2} \frac{\partial^2 C}{\partial S_1^2} S_1^2 \sigma_1^2 + \frac{1}{2} \frac{\partial^2 C}{\partial S_2^2} S_2^2 \sigma_2^2 \\
+ \frac{\partial^2 C}{\partial S_1 \partial S_2} \sigma_1 \sigma_2 S_1 S_2 \rho + \frac{\partial^2 C}{\partial S_1 \partial \rho} \sigma_1 S_1 \sigma_3 \sqrt{(h - \rho)(\rho - l)} \\
+ \frac{\partial^2 C}{\partial S_2 \partial \rho} \sigma_2 S_2 \sigma_3 \sqrt{(h - \rho)(\rho - l)} \right] \, dt \\
+ \frac{\partial C}{\partial S_1} \sigma_1 S_1 dW_1 + \frac{\partial C}{\partial S_2} \sigma_2 S_2 dW_2 + \frac{\partial C}{\partial \rho} \sigma_3 \sqrt{(h - \rho)(\rho - l)} dW_3.
\]

We define a portfolio \( \Pi \) by
\[
\Pi = C_1 + \Gamma_1 C_2 + \Gamma_2 S_1 + \Gamma_3 S_2,
\]
where \( S_1, S_2 \) are stocks, \( C_1, C_2 \) are derivatives contingent on \( S_1, S_2, \rho \), and the amounts \( \Gamma_1, \Gamma_2, \Gamma_3 \) are yet to be determined. We assume that \( \Pi \) is self-
financing. It follows that

\[ d\Pi = [\cdots]_1 dt + \Gamma_1[\cdots]_2 dt + \Gamma_2 rS_1 dt + \Gamma_3 r S_2 dt \]
\[ + \left( \frac{\partial C_1}{\partial \rho} \sigma_1 S_1 + \frac{\partial C_2}{\partial S_1} \Gamma_1 \sigma_1 S_1 + \Gamma_2 \sigma_1 S_1 \right) dW_1 \]
\[ + \left( \frac{\partial C_1}{\partial \rho} \sigma_2 S_2 + \frac{\partial C_2}{\partial S_2} \Gamma_1 \sigma_2 S_2 + \Gamma_3 \sigma_2 S_2 \right) dW_2 \]
\[ + \left( \frac{\partial C_1}{\partial \rho} + \Gamma_1 \frac{\partial C_2}{\partial \rho} \right) \left( \sigma_3 \sqrt{(h - \rho)(\rho - l)} \right) dW_3, \]

where \([\cdots]_1, [\cdots]_2\) refer to the \(dt\) terms of \(dC_1\) respectively \(dC_2\), as calculated above.

For the portfolio \(\Pi\) to be risk neutral, the factors in front of \(dW_1, dW_2, dW_3\) need to be zero. We can achieve this by choosing \(\Gamma_1, \Gamma_2, \Gamma_3\) to be

\[ \Gamma_1 = -\frac{\partial C_1}{\partial \rho}, \]
\[ \Gamma_2 = \frac{\partial C_1}{\partial \rho} \frac{\partial C_2}{\partial \rho} \sigma_1 S_1 - \frac{\partial C_1 \sigma_1 S_1}{\partial S_1}, \]
\[ \Gamma_3 = \frac{\partial C_1}{\partial \rho} \frac{\partial C_2}{\partial \rho} \sigma_2 S_2 - \frac{\partial C_1 \sigma_2 S_2}{\partial S_2}. \]

With these choices of \(\Gamma_1, \Gamma_2, \Gamma_3\), the portfolio is risk neutral, so by absence of arbitrage it must hold that \(d\Pi = r\Pi dt\). This means that

\[ [\cdots]_1 dt - \frac{\partial C_1}{\partial \rho} [\cdots]_2 dt + rS_1 \left( \frac{\partial C_1}{\partial \rho} \frac{\partial C_2}{\partial \rho} \frac{\partial C_2}{\partial S_1} - \frac{\partial C_1}{\partial S_1} \right) dt \]
\[ + rS_2 \left( \frac{\partial C_1}{\partial \rho} \frac{\partial C_2}{\partial \rho} \frac{\partial C_2}{\partial S_2} - \frac{\partial C_1}{\partial S_2} \right) dt \]
\[ = r \left[ C_1 - \frac{\partial C_1}{\partial \rho} C_2 + \left( \frac{\partial C_1}{\partial \rho} \frac{\partial C_2}{\partial \rho} \frac{\partial C_2}{\partial S_1} - \frac{\partial C_1 \sigma_1 S_1}{\partial S_1} \right) S_1 \right. \]
\[ + \left. \frac{\partial C_1}{\partial \rho} \frac{\partial C_2}{\partial \rho} \frac{\partial C_2}{\partial S_2} - \frac{\partial C_1 \sigma_2 S_2}{\partial S_2} \right] dt. \]

We can rewrite this as

\[ \frac{[\cdots]_1}{\partial C_1/\partial \rho} C_1 dt = \frac{[\cdots]_2}{\partial C_2/\partial \rho} C_2 dt. \]
The left-hand side of the above does not depend on $C_2$, and the right-hand side does not depend on $C_1$, so both sides of the equation do not depend on $C_1$ and $C_2$, so are equal to a function $g(S_1, S_2, \rho, t)$, which can be considered a premium for correlation risk. This tells us that the price process of a derivative $C$ is a solution of the PDE

$$[\cdots]_1 - rC - g \frac{\partial C}{\partial \rho} = 0.$$ 

Writing this out fully gives us

$$\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S_1} rS_1 + \frac{\partial C}{\partial S_2} rS_2 + \frac{1}{2} \frac{\partial^2 C}{\partial \rho^2} \sigma_1^2 (h - \rho)(\rho - l) + \frac{1}{2} \frac{\partial^2 C}{\partial S_1^2} S_1^2 \sigma_1^2$$

$$+ \frac{1}{2} \frac{\partial^2 C}{\partial S_2^2} S_2^2 \sigma_2^2 + \frac{\partial^2 C}{\partial S_1 \partial S_2} \sigma_1 \sigma_2 S_1 S_2 \rho + \frac{\partial^2 C}{\partial S_1 \partial \rho} \sigma_1 \sigma_1 S_1 \sigma_3 \sqrt{(h - \rho)(\rho - l)}$$

$$+ \frac{\partial^2 C}{\partial S_2 \partial \rho} \sigma_2 \sigma_2 S_2 \sigma_3 \sqrt{(h - \rho)(\rho - l)} + (\bar{p} - \beta \rho - g) \frac{\partial C}{\partial \rho} - rC$$

$$= 0.$$ 

This PDE is analogous to the Black-Scholes PDE, but where Black-Scholes is about options on single stocks, this PDE applies to options on two stocks whose correlation is a Jacobi process. This PDE can be used to determine the fair price of such derivatives under the assumptions of this model. In certain special cases it is possible to find a explicit solution by solving the PDE. More generally, one can solve the PDE numerically using the method of finite differences.

One problem with this model is that it is not clear if the choice of a Jacobi process to model the correlation is a good one. Another problem is that it only models the correlation between one pair of stocks, so it can not be used directly to price options that depends on more than two stock values, such as index options. Extending the model to encompass more stocks would not be trivial, since positive definiteness of the resulting correlation matrices would need to be somehow guaranteed.

### 5.2 Common Factor Model

In the PhD thesis [Boortz, 2008](#), the author describes a model called the Common factor model (CFM).
5.2.1 Definition of the Common Factor Model

We assume that the stock prices $S_i$ are geometric Brownian motions satisfying
\[ dS_{i,t} = \mu_i dt + \sigma_i dW_{i,t}, \quad i = \{1, \cdots, n\}. \]

We decompose the Brownian motions $W_i$ as
\[ dW_{i,t} = \sqrt{1 - \varrho_{i,t}^2} dB_{i,t} + \varrho_{i,t} dB_{t} \quad i = \{1, \cdots, n\}, \]
where $B_{1}, B_{2}, \cdots, B_{n}, B_t$ are independent Brownian motions, and $\varrho_{i,t}, \cdots, \varrho_{n,t}$ are stochastic processes.

By this construction, the stock prices are related through the ‘common factor’ $B_t$. The processes $\varrho_i$ can be interpreted as the stock price sensitivity to the common market factor, and $\varrho_i$ is called the ‘asset to market correlation’ of stock $i$. Because all stocks are related to the same market factor, the stocks are correlated. The instantaneous correlation between stock $i$ and stock $j$ (as defined in Section 3.1) in this model is given by $\rho_{i,j,t} = \varrho_{i,t} \varrho_{j,t}$, so the correlation matrix is
\[
\begin{pmatrix}
1 & \varrho_{1,t} \varrho_{2,t} & \cdots & \varrho_{1,t} \varrho_{n,t} \\
\varrho_{2,t} \varrho_{1,t} & 1 & \cdots & \varrho_{2,t} \varrho_{n,t} \\
\vdots & \ddots & \ddots & \vdots \\
\varrho_{n,t} \varrho_{1,t} & \cdots & \varrho_{n,t} \varrho_{n-1,t} & 1
\end{pmatrix}.
\]

The asset to market correlations are modeled as Jacobi processes.
\[ d\varrho_{i,t} = \kappa(m_i - \varrho_{i,t}) dt + \eta \sqrt{1 - \varrho_{i,t}^2} d\tilde{W}_{i,t}, \]
where $\kappa$ is the speed of mean reversion, $m_i$ is the level of mean reversion, $\eta$ is a scaling constant, and the term $\sqrt{1 - \varrho_{i,t}^2}$ causes the processes to stay within the interval $[-1, 1]$.

Furthermore, we assume that the Brownian motions driving the correlations $\tilde{W}_i$ and the market factor $B_t$ have quadratic covariation given by
\[ \frac{d}{dt} \langle \tilde{W}_i, B_t \rangle_t = \xi_t, \quad \text{with } \xi_t \in [-1, 1] \text{ constant.} \]
5.2.2 Properties of the Common Factor Model

The matrix (5.4) is always positive definite. This follows from the fact that it is the covariance matrix of the random vector

\[
\left( \sqrt{1 - \varrho_1^2} X_1 + \varrho_1 Y, \sqrt{1 - \varrho_2^2} X_2 + \varrho_2 Y, \ldots, \sqrt{1 - \varrho_n^2} X_n + \varrho_n Y \right)^	op,
\]

where \( X_1, X_2, \ldots, X_n, Y \) are independent random variables with variance 1.

In this model, the \( n(n-1)/2 \) correlations between \( n \) stocks at time \( t \) are represented by only \( n \) numbers \( (\varrho_{1,t}, \ldots, \varrho_{n,t}) \). As a consequence, not all possible correlation matrices can be represented. When fitting the model to real world data (at time \( t \)), we should find numbers \( \varrho_{1,t}, \varrho_{2,t}, \ldots, \varrho_{n,t} \) so that the correlation matrix implied by these numbers matches the real correlation matrix as close as possible. According to Boortz (the author of the model) the goodness of fit is higher when the correlation data comes from stocks from the same (geographical) region, as opposed to correlation between stocks from different regions. This could be explained by stocks from the same region having exposure to the same 'common factor' market. Finding the best fit is computationally expensive. Boortz describes an algorithm for finding \( \varrho_1, \ldots, \varrho_n \) that generate a correlation matrix that resembles a given correlation matrix, but the algorithm only applies when all correlations (in a market) are positive, which is often not the case in correlation matrices derived from real stock markets.

If the parameters \( \xi_i \) are negative, then simulations show that the model exhibits the stylized fact of asymmetry in average correlation and market returns: If the common factor \( \bar{B}_t \) has negative returns, then the average correlation \( \bar{\rho}_t \) (see definition 5.7) increases, and vice versa. To explain this behavior intuitively, let us look at the differential equation for \( \bar{\rho}_t \).

\[
d\bar{\rho}_t = [\ldots] dt + \eta \left( \sum_{i=1}^{n} \left( \sum_{j=1, j \neq i}^{n} \varrho_{i,t} \xi_i \sqrt{1 - \varrho_{i,t}^2} \right) d\bar{B}_t \right) + \eta \left( \sum_{i=1}^{n} \left( \sum_{j=1, j \neq i}^{n} \varrho_{i,t} \sqrt{1 - \xi_i^2} \right) d\tilde{B}_{i,t} \right).
\]

For negative constants \( a_1, a_2, \ldots, a_n \) and i.i.d. standard normally distributed random variables \( X_1, X_2, \ldots, X_n, Y \), it holds that

\[
\text{var} \left( \sum_{i=1}^{n} a_i X_i \right) = \sum_{i=1}^{n} a_i^2 < \left( \sum_{i=1}^{n} a_i \right)^2 = \text{var} \left( \sum_{i=1}^{n} a_i Y \right).
\]
This same mechanism leads to the randomness of $\rho$ to be mostly caused by $B_t$, and if the $\xi_i$ are negative, this relation is negative.

The Common Factor Model uses a reasonably simple mathematical structure to obtain a multi-asset correlation model with guaranteed positive definite-ness. The model honors some of the stylized facts of correlation. The individual correlation processes are modeled as products of dependent Jacobi processes. It is not clear how realistic this is.

5.3 Single factor model

5.3.1 Background: Multiplicative indexes

Most stock indices are of the form

$$I_t = \sum_{i=1}^{N} a_i S_{i,t},$$

where $S_{i,t}$ is the price of stock $i$ at time $t$, and the $a_i$ are constant, positive numbers. The standard assumption is that the processes $S_{i,t}$ are geometric Brownian motions satisfying

$$dS_{i,t} = \mu_i S_{i,t} dt + \sigma_i S_{i,t} dW_{i,t},$$

where $\mu_i$ and $\sigma_i$ are constants, and $W_{i,t}$ are Brownian motions.

From this it follows that the index process is a weighted sum of geometric Brownian motions. This sum is not itself a geometric Brownian motion. (If it was, this would imply that a sum of lognormal distributions is lognormal, making the lognormal distribution a *stable* distribution. This can not be correct, because the only stable distribution with finite variance is the normal distribution.)

Sums of geometric Brownian motions do not have a convenient distribution. To avoid having to work with such sums, we will make a change in the definition of the index.
From the above it follows that
\[ \frac{dI_t}{I_t} = \frac{\sum_{i=1}^{N} a_i dS_{i,t}}{\sum_{j=1}^{N} a_j S_{j,t}} \]
\[ = \sum_{i=1}^{N} \frac{a_i S_{i,t}}{\left( \sum_{j=1}^{N} a_j S_{j,t} \right)} \frac{dS_{i,t}}{S_{i,t}} \]
\[ = \sum_{i=1}^{N} w_{i,t} \frac{dS_{i,t}}{S_{i,t}}, \]
where
\[ w_{i,t} = \frac{a_i S_{i,t}}{\sum_{j=1}^{N} a_j S_{j,t}} \]
is the index weight of stock \( i \). The index weights change over time since they depend on the stock prices. To make the index easier to work with, we will assume that the index weights are constant, so
\[ \frac{dI_t}{I_t} = \sum_{i=1}^{N} w_{i,t} \frac{dS_{i,t}}{S_{i,t}}, \]
for \( w_{i,t} \) constants. This is an assumption that is often (implicitly) made, such as in the article [Skintzi and Refenes, 2005], and in the calculation of the CBOE S&P 500 Implied Correlation Index.

The assumption changes the index process, since in reality the weights are not constant. This loss of accuracy is the price we pay for making the index easier to work with.

It follows from the assumption of constant weights that
\[ dI_t = \sum_{i=1}^{N} \mu_i w_{i,t} dt + \sigma_i w_{i,t} dW_{i,t} = I_t \left( \sum_{i=1}^{N} \mu_i w_{i,t} dt + \sigma_i w_{i,t} dW_{i,t} \right), \]
so the quadratic variation process \( \langle I \rangle_t \) satisfies
\[ d \langle I \rangle_t = \sum_{i=1}^{N} w_{i,t}^2 \sigma_i^2 I_t^2 dt + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} 2w_{i,t} \sigma_i \sigma_j I_t^2 d \langle W_i, W_j \rangle_t. \]
With Itô’s lemma we calculate
\[ d \log(I_t) = \frac{1}{I_t} dI_t - \frac{1}{2} \frac{1}{I_t^2} d \langle I \rangle_t \]
\[ = \sum_{i=1}^{N} \left( \mu_i dt + \sigma_i dW_{i,t} \right) - \frac{1}{2} w_{i,t}^2 \sigma_i^2 dt \]
\[ - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_{i,t} \sigma_i \sigma_j d \langle W_i, W_j \rangle_t. \]
Hence

$$\log(I_t) = \log(I_0) + \sum_{i=1}^{N} \left( w_i(\mu_i t + \sigma_i W_{i,t}) - \frac{1}{2} w_i^2 \sigma_i^2 t \right) - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \langle W_i, W_j \rangle_t. $$

So the index process under the assumption of constant weights is

$$I_t = I_0 \exp \left( \sum_{i=1}^{N} \left( w_i(\mu_i t + \sigma_i W_{i,t}) - \frac{1}{2} w_i^2 \sigma_i^2 t \right) - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \langle W_i, W_j \rangle_t \right)$$

$$= I_0 \exp \left( \sum_{i=1}^{N} \left( w_i \mu_i - \frac{1}{2} w_i^2 \sigma_i^2 t + w_i \sigma_i W_{i,t} \right) \right) \exp \left( - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \langle W_i, W_j \rangle_t \right)$$

$$= I_0 \exp \left( \sum_{i=1}^{N} \left( w_i \mu_i - \frac{1}{2} w_i^2 \sigma_i^2 t + w_i \sigma_i W_{i,t} \right) \right) \exp \left( - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \langle W_i, W_j \rangle_t \right)$$

$$\cdot \exp \left( \sum_{i=1}^{N} \frac{1}{2} (w_i - w_i^2) \sigma_i^2 t \right)$$

$$= I_0 (\Pi_{i=1}^{N} S_{i,t}^w) \exp \left( - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \langle W_i, W_j \rangle_t \right) \exp \left( \sum_{i=1}^{N} \frac{1}{2} (w_i - w_i^2) \sigma_i^2 t \right).$$

By the last formula we see that the index, which was originally a weighted arithmetic mean, has become a weighted geometric mean (multiplied by a factor), by the assumption of constant weights.

Assuming that $\rho_{ij}(t)$ is deterministic, it follows that

$$\log I_t = \sum_{i=1}^{N} w_i \sigma_i W_{i,t} + \text{deterministic terms},$$

hence the volatility of (the log-returns of) the index is

$$\sigma_i^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \rho_{ij}(t), \quad (5.5)$$

where $\rho_{ij}(t)$ is the pair-wise correlation between $S_i$ and $S_j$ at time $t$.

Under the assumption that all correlations are equal ($\rho_{ij}(t) = \rho(t)$ for all $i \neq j$), this becomes

$$\sigma_i^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \rho(t). \quad (5.6)$$
As we make the switch from correlations that differ per pair of stocks \( \rho_{ij} \) to correlation that is the same for each pair of stocks \( \rho \), we want the implied index volatility (as calculated in formulas (5.5) and (5.6) to remain unchanged. This implies that we must choose

\[
\rho(t) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \rho_{ij}(t)}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j}.
\]

(5.7)

We can use this formula as a measure of ‘mean correlation’ in the index.

If we define \( \rho(t) \) to be some stochastic process or time series, this gives us a model for correlation.

### 5.3.2 Summarizing the single-factor model

In this section we summarize the model that we derived in the previous section. We have derived a model under the following two assumptions.

- Index weights are constant.
- The stock return correlation \( \rho_{ij}(t) \) of stock \( i \) and stock \( j \) in time period \( t \) is independent of \( i \) and \( j \): \( \rho_{ij}(t) = \rho(t), \text{ for } i \neq j. \)

As seen in Theorem 2.10 the above assumptions will lead to a positive semidefinite correlation matrix if \( 0 \leq \rho(t) \leq 1. \)

If we know (or predict) \( \rho(t) \) then (by the assumption of constant index weights) we can calculate index volatility by

\[
\sigma_I^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \rho(t).
\]

If we have market data consisting of asset return correlation \( \rho_{ij}(t) \) for all \( i, j \), for a fixed time \( t \), then we can calibrate the value of \( \rho(t) \) as

\[
\rho(t) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \rho_{ij}(t)}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j}.
\]

33
The point of this calibration is that the value of \( \rho(t) \) thus obtained implies the same index volatility as the full set of asset correlations \( \rho_{ij}(t) \) does, i.e.

\[
\sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \rho_{ij}(t) = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \rho(t).
\]

We can create a correlation model by specifying how \( \rho(t) \) behaves as a process. The only technical constraint is that \( 0 \leq \rho(t) \leq 1 \) should hold. In fact, a value of \( \rho(t) \) which is slightly below zero is also allowed from a theoretical perspective, but as we will see in Section 7.1, a negative value of \( \rho(t) \) does not conform to reality.

This model is easy to work with because there is only one dimension, and the only restriction is the straightforward constraint that \( 0 \leq \rho(t) \leq 1 \) should hold. Despite the simplicity, the model can be used to calculate an estimate of the index volatility, which makes it interesting for option pricing.

5.4 Conclusion

In this section we studied the two-asset model by Jun Ma, the Common Factor Model by Boortz, and single-factor models. Note that the two-asset model and the Common Factor Model are both ‘complete’ models, with only the parameters to be determined, while a single-factor still needs the dynamics of \( \rho(t) \) to be defined. This means that the single-factor models as we defined them in this chapter should be considered a family of models rather than a specific model.

The two-asset model is only usable for modeling the correlation between the returns of two assets, while the Common Factor Model and single-factor models can be used for modeling the correlations of all the assets in an index together. This means that the two-asset model is not suitable for index option pricing, because indices contain more than two assets. In the coming chapters we will use the single-factor models because of their simplicity (only one dimension) and flexibility (we can define different models for \( \rho(t) \)).
Chapter 6

Correlation estimation using daily data

As discussed in Section 4.2.1, there are difficulties with estimating correlation based on high-frequency data. To avoid these problems, we can instead estimate correlation using start-of-day and end-of-day stock prices, and possibly also using daily low (lowest price of the day) and high (highest price of the day) prices of the stocks. It is very important to use good estimators, because high estimation error makes it very difficult to fit models. In this chapter we compare the quality of different correlation estimators based on daily data.

6.1 Overview of estimators

6.1.1 Classical estimator

The most straightforward approach to calculating correlation based on daily data is to use the daily returns $r_X(i), r_Y(i)$ of stock $X$ and stock $Y$ on day $i$ to calculate an estimator using the standard sample correlation formula

$$
\hat{\rho}_{X,Y} = \frac{\sum_{i=1}^{n} (r_X(i) - \bar{r}_X)(r_Y(i) - \bar{r}_Y)}{\sqrt{\sum_{i=1}^{n} (r_X(i) - \bar{r}_X)^2 \sum_{i=1}^{n} (r_Y(i) - \bar{r}_Y)^2}},
$$

where $\bar{r}_X, \bar{r}_Y$ are the sample means of $r_X$ and $r_Y$. This estimator is biased (Zimmerman et al., 2003), with the bias depending on the underlying dis-
distribution. It might seem surprising that the standard sample correlation is a biased estimator, but the bias is generally very small.

6.1.2 Rogers’ estimator

In the article [Rogers and Zhou, 2008], the authors derive a covariance estimator with a low variance by using not only the daily returns, but also daily high and daily low prices. We need to introduce some notation. Let $O_X(i), C_X(i), H_X(i), L_X(i)$ be the opening price, closing price, high price and low price of stock $X$ during day $i$. Define $c_X(i) = \log(C_X(i)) - \log(O_X(i))$, $h_X(i) = \log(H_X(i)) - \log(O_X(i))$, $l_X(i) = \log(L_X(i)) - \log(O_X(i))$. The potential estimators that Rogers and Zhou study are all linear combinations of the nine crossproducts

$$
\begin{align*}
&c_X(i)c_Y(i), \ c_X(i)h_Y(i), \ c_X(i)l_Y(i), \\
&h_X(i)c_Y(i), \ h_X(i)h_Y(i), \ h_X(i)l_Y(i), \\
&l_X(i)c_Y(i), \ l_X(i)h_Y(i), \ l_X(i)l_Y(i).
\end{align*}
$$

They assume that the underlying stock price process is a geometric Brownian motion without drift. Based on this assumption, they calculate the expectations of all nine crossproducts for a real underlying correlation $\rho \in \{-1, 0, 1\}$, as well as the (9-by-9) covariance matrix of the crossproducts for $\rho = 0$. Then they determine which linear combinations of the nine crossproducts has minimum variance for $\rho = 0$ under the condition that it is unbiased for $\rho \in \{-1, 0, 1\}$. The outcome is the following estimator of return covariance

$$
\hat{\text{cov}}_{X,Y} = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{1}{2} c_X(i)c_Y(i) \right. \\
\left. + \frac{1}{6 - 8 \log(2)} \left( h_X(i) + l_X(i) - c_X(i) \right) \left( h_Y(i) + l_Y(i) - c_Y(i) \right) \right].
$$

To estimate correlation, we need to combine this covariance estimator with volatility estimators, for example the estimator (also using daily high and low) described in [Rogers et al., 1994]

$$
\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^{n} h_X(i)(h_X(i) - c_X(i)) + l_X(i)(l_X(i) - c_X(i)).
$$

This leads to the following correlation estimator (we omit dependence on $i$}
in the notation)

\[ \hat{\rho}_{X,Y} = \frac{\sum_{i=1}^{n} \frac{1}{2} c_X c_Y + \frac{1}{6 - 8 \log(2)} (h_X + l_X - c_X)(h_Y + l_Y - c_Y)}{\sqrt{\sum_{i=1}^{n} (h_X (h_X - c_X) + l_X (l_X - c_X)) \sum_{i=1}^{n} (h_Y (h_Y - c_Y) + l_Y (l_Y - c_Y))}}. \]  

\[ (6.2) \]

### 6.1.3 Sign-based estimator

In this section we derive an estimator which only depends on the signs of the returns. An advantage of this estimator is that it is not necessary to know or determine volatilities in order to use it. This estimator can also be used in the scenario where have multivariate normal pairs of data with equal correlation but different covariances, and we want to estimate the correlation. This estimator can be used in the situation where we want to calculate average correlation using only one day’s return data.

We start in the simple case where \( \mathbf{r} = (x, y) \) is a two-dimensional random variable representing the returns of two stocks over a certain time period, for instance end-of-day to end-of-day. We will assume that \( \mathbf{r} \) is distributed according to a multivariate normal distribution with mean \((0, 0)\) and unknown variance matrix.

Let \( \mathbf{r}_1 = (x_1, y_1) \) be a realization of \( \mathbf{r} \). Without knowing or estimating the variance of \( x \) and \( y \), the absolute values \( |x_1|, |y_1| \), do not help in determining the correlation. However, the sign of the product \( x_1 y_1 \) does hold information. For example, if the correlation between \( x \) and \( y \) is one, we would expect \( \text{sign}(x_1 y_1) = 1 \) to be true. Note that \( \text{var}(x) \) and \( \text{var}(y) \) do not affect the distribution of \( \text{sign}(xy) \), since \( \text{sign}(axby) = \text{sign}(xy) \) for \( ab > 0 \). This means that the distribution of \( \text{sign}(xy) \) is only a function of \( \rho := \text{corr}(x, y) \), since \( \text{var}(x), \text{var}(y) \) and \( \rho \) uniquely determine the variance matrix of \( (x, y) \).

We can calculate \( g(\rho) := \mathbb{P}(\text{sign}(xy) > 0) = \mathbb{E} \mathbf{1}(xy > 0) \) by

\[
g(\rho) = \int_{0}^{\infty} \int_{0}^{\infty} f(s_1, s_2) ds_1 ds_2 \\
+ \int_{-\infty}^{0} \int_{-\infty}^{0} f(s_1, s_2) ds_1 ds_2, \text{ with } f \text{ the density of } \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1, \rho \\ \rho, 1 \end{pmatrix} \right). \]

Alternatively we can write this in terms of the cumulative distribution func-
tion $F_\rho$ of this distribution.

$$
g(\rho) = 1 - \mathbb{P}(x > 0, y < 0) - \mathbb{P}(x < 0, y > 0)
= 1 - (F_\rho(\infty, 0) - F_\rho(0, 0)) - (F_\rho(0, \infty) - F_\rho(0, 0))
= 1 + 2F_\rho(0, 0) - F_\rho(\infty, 0) - F_\rho(0, \infty).
$$

We show a plot of function $g$ in figure 6.1. We see that the function is one-to-one, which means that it has an inverse $g^{-1}$, which we can approximate numerically. We also see that the function is close to being linear in a large part of its domain.

![Figure 6.1: The function $g(\rho)$.](image)

Let $\mathbf{r}_1 = (x_1, y_1), \mathbf{r}_2 = (x_2, y_2), \cdots, \mathbf{r}_n = (x_n, y_n)$ be two-dimensional multivariate normal random variables with mean $(0, 0)^T$, unknown covariance matrix (possibly depending on $n$), and unknown correlation $\rho = \text{corr}(x_1, y_1) = \text{corr}(x_2, y_2) = \cdots = \text{corr}(x_n, y_n)$. By linearity of expectations it follows that

$$
\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} 1(x_iy_i > 0) \right] = g(\rho),
$$

so

$$
g^{-1} \left( \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} 1(x_iy_i > 0) \right] \right) = \rho.
$$

Because the function $g^{-1}$ is not linear, it is not mathematically correct to invert the order of applying $g^{-1}$ and taking the expectation. However, as
we see in figure 6.1, \( g^{-1} \) does come close to being linear in a large (and important) part of its domain. This suggests the following estimator of \( \rho \)

\[
\hat{\rho} := g^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} 1(x_i y_i > 0) \right). \tag{6.3}
\]

**Theorem 6.1.** The estimator of formula (6.3) is consistent, i.e. \( \hat{\rho}(n) \) converges to \( \rho \) in probability as \( n \to \infty \) for all \( \rho \in [-1, 1] \).

**Proof.** By the weak law of large numbers, \( \frac{1}{n} \sum_{i=1}^{n} 1(x_i y_i > 0) \) converges to \( \mathbb{P}(\text{sign}(x_1 y_1) > 0) \) in probability, as \( n \to \infty \). Applying the continuous mapping theorem, it follows that \( \hat{\rho}(n) = g^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} 1(x_i y_i > 0) \right) \) converges to \( g^{-1}(\mathbb{P}(\text{sign}(x_1 y_1) > 0)) = \rho \) (this equality holds by definition of \( g \)), in probability, as \( n \to \infty \).

### 6.2 Simulation study of correlation estimators

We will study the bias and standard deviation of these estimators using simulation. For this we use a simulated set of 30 stocks, all of which follow geometric Brownian motions with daily volatility 0.25 and mean 0. For the correlation matrix of the simulation we use a matrix derived from DAX stock data using end-of-day data, using the traditional correlation estimator (6.1) with 35 days’ data. We assume that the correlations do not change over time. Under this setting we generate 10 days’ data (start-of-day, end-of-day, daily high, daily low) prices of all stocks, then calculate the correlation estimators for each pair of stocks. We repeat this a large number of times to derive the properties of the estimators, as a function of the real correlation.

The bias and standard deviation of the estimators of formulas (6.1), (6.2) and (6.3) (estimated by simulation) are shown in Figures 6.2, 6.3 and 6.4. Note that, while these properties are displayed for the case when we have 10 days worth of data, we can generalize to the case of \( n \) days using these graphs. By the simulation assumption that the returns of each day are i.i.d., the estimator bias does not depend on the number of days on which we have data. For Rogers’ estimator and the sign-based estimator it holds that, if \( \hat{\rho}_n \) is an estimator based on \( n \) days’ data, then \( \text{stdev}(\hat{\rho}_n) = (\sqrt{m/n})\text{stdev}(\hat{\rho}_m) \).

To create Figure 6.2 we used 50000 simulations. For Figure 6.3 we discretized the daily stock price paths into 1000 steps to determine daily low and high
prices, and simulated the 10 days’ data 50000 times. To produce Figure 6.4 we used 50000 simulations.

The above estimates are for correlation between the returns of two stocks. It might also be the case that we are interested in mean correlation (weighted with stock index weight and volatility), as discussed in Section 5.3.2. We will investigate these estimators using simulation. As before we assume all volatilities to be equal, and for simplicity we assume that all stock index weights are equal as well. This means we give all correlations equal weights when taking the mean. We estimate mean correlation \((\rho(m))\) by first calculating all correlations independently using one of the discussed estimators, and then averaging the result, leading to an estimator \(\bar{\rho}(m)\). In Figures 6.5, 6.6 and 6.7 we show bias and variance of this estimator for the estimators of Formulas 6.1, 6.2 and 6.3 as a function of \(n\), the number of days’ worth of data that is used. In section 6.3 we use these figures to draw conclusions about the estimators.

It is an interesting question if we really need multiple days’ data to determine mean correlation, or if one days’ data is enough. If all stocks move in the same direction during a day, we would expect mean correlation to be higher than in the case where half the stocks move up, and half move down, regardless of volatilities. Based on this intuition, we define a sign-based estimator of mean correlation (similar to the estimator of formula 6.3) using only one day’s data as

\[
\hat{\rho} := g^{-1} \left( \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1(r_i r_j > 0) \right),
\]

where \(r_i\) is the return of stock \(i\). We test this estimator on a simulated stock market with 30 stocks whose returns have a multivariate normal distribution with mean \((0, \cdots, 0)^\top\) and covariance matrix \(C\) with \(C_{ii} = 1\) and \(C_{ij} = \rho\) for \(i \neq j\), for a fixed \(\rho\). For this simulation we calculated 25000 realizations of this estimator, for each value of \(\rho \in \{0.1, 0.2, \cdots, 0.99\}\), for \(n = 30\). In Figure 6.8 we plot expectation and standard deviation of this estimator, as a function of the real parameter value.
Figure 6.2: Bias and standard deviation of estimator 6.1 (standard sample correlation), calculated for 10 days’ data.

Figure 6.3: Bias and standard deviation of estimator 6.2 (Rogers’ estimator), calculated for 10 days’ data.
Figure 6.4: Bias and standard deviation of estimator 6.3 (sign-based estimator), calculated for 10 days’ data.

Figure 6.5: Bias and standard deviation when estimating mean correlation by averaging the results of estimator 6.1 (standard sample correlation).
Figure 6.6: Bias and standard deviation when estimating mean correlation by averaging the results of estimator 6.2 (Rogers’ estimator).

Figure 6.7: Bias and standard deviation when estimating mean correlation by averaging the results of estimator 6.3 (sign-based estimator).
Figure 6.8: Bias and standard deviation when estimating mean correlation by using the estimator of formula (6.4) (sign-based estimator using one day of data).
6.3 Conclusions on estimators

We have examined the performance of three estimators: the classical correlation estimator, Roger’s correlation estimator, and a sign-based correlation estimator. The classical estimator and Roger’s estimator both have very low bias. The classical estimator has lower variance when the real correlation is high, and Roger’s estimator has lower variance when the real correlation is low (Figures 6.2, 6.3).

There are however some problems with Roger’s estimator. The first disadvantage is that Roger’s estimator is dependent on the assumption that the underlying stock price process is really a geometric Brownian motion, since the estimator is derived under this assumption. This means that for a stock price process that is not a perfect geometric Brownian motion, we should expect the performance of Roger’s estimator to worsen, leading to a higher bias and variance. The classical estimator does not have this problem, since it is not derived under any assumption on the underlying process. This does not mean that the performance of the classical estimator does not depend on the stochastic nature of the stock price process, but it does mean that we do not expect the performance of the classical estimator to significantly worsen if the stock process is not a geometric Brownian motion. The second disadvantage with Roger’s estimator compared to the classical estimator is that we need additional information to be able to use it, namely the high and low prices. In particular this means that we cannot use Roger’s estimator for estimating overnight correlations. In the next chapters we will use the classical estimator, as we want to take into account overnight returns.

We should use the classical estimator if we want to use overnight returns, or if we believe that the stock process significantly deviates from a geometric Brownian motion, or if we expect that the real correlation is high. We should use Roger’s estimator if we do not want to use overnight returns, and we believe the stock process does not significantly deviate from a geometric Brownian motion, and we expect the real correlation to be low. Another question is how many days’ data we need to use for the estimation. This is a question that can not be answered in general, since the acceptable level of variance of the estimator depends on the purpose of the estimation. While a higher number of days will give a lower (theoretical) volatility of the estimator, it can also introduce problems as it smears out correlation over more days, which might make it harder to e.g. study the effect of daily index returns on correlation. In the next chapters we will use ten days’ data, since this seems to be a reasonable compromise.
The sign-based estimator has a higher bias and variance than the classical estimator and Roger’s estimator. The advantage of the sign-based estimator is that it is the only estimator (of the three that we are comparing) that can be used to estimate mean correlation using one day’s return data. In our test using the DAX correlation matrix, the sign-based estimator had a high bias and variance when estimating mean correlation using one day of data (Figure 6.8). It seems that the 30 stocks in the DAX index is not enough for this estimator to be usable for this index. For larger indices however, such as the S&P500, this estimator could be usable, especially considering the consistency result of Theorem 6.1.
Chapter 7

Study of correlation in the DAX and AEX index

In this chapter we study sample correlations of stock returns, based on historical data. The stocks we use are the constituents of the AEX and DAX indices. We have two data sets. The first data set contains opening and closing prices of these stocks from January 2002 until October 2010. This period contains 2260 trading days. The second data set contains intraday price data on these stocks from December 15 2010 until April 1 2011, with one price quote per two minutes for each stock. This period contains 78 trading days.

With the first data set containing opening and closing prices, we calculate correlation between two stocks in the following way. We calculate the stocks return from open to close (daytime returns) and the stock returns from close to open (overnight return) over a period of 10 days for both stocks. This gives us a series of 20 returns for each stock, for which we calculate the standard sample correlation. The datasets spans about 226 10-day periods.

The second data set (containing intraday data) allows us to calculate sample correlation on a daily basis, using all the intraday returns for each stock. This gives us 78 daily sample correlations for each pair of stocks which were part of the AEX or DAX during the whole 78 days period.

Using these correlations between stock returns then allows us to calculate the measure for ‘mean’ index correlation which we derived in Section 5.3.1:

\[
\rho_t = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j \rho_{ij,t}}{\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_i \sigma_j }.
\]  

(7.1)
7.1 Mean index correlation

We will check whether the ‘stylized facts’ from Section 4.1 apply to our data. Figure 7.1 and Figure 7.2 show mean AEX and DAX index correlations (calculated as described above, using end-of-day prices for ten consecutive days) plotted together with index level. In Figure 7.3 and Figure 7.4 we show mean AEX and DAX correlation calculated daily using intraday data, plotted together with the index level.

According to the stylized fact of asymmetry in correlation, there should be a negative relation between stock return and correlation. In these graphs we do see hints of this phenomenon, but it is overall not very convincing.

In Figure 7.5 and Figure 7.6 we show average correlation and average index volatility, calculated over 10-day periods, for the DAX index. In the scatter plot, a positive relation between correlation and volatility seems to be visible, which is in agreement with the stylized fact of comovement of correlation and volatility.

In Figure 7.7 we show a scatter plot of the change of DAX correlation as calculated over two consecutive ten-day periods \((\rho(t+11, \cdots, t+20) - \rho(t+1, \cdots, t+10))\) against the index return on the day preceding the second 10-day period on which the correlation was calculated \((r(t+10))\). In Figure 7.8 we show a scatter plotter of the change of DAX correlation calculated on two consecutive days using intraday data \((\rho(t+1) - \rho(t))\) against the index return on the day before \((r(t))\). These graphs do not appear to show any regularity. This is unfortunate, because we were hoping to be able to use index return to predict future correlation. These graphs do not imply that such a prediction is impossible, but they do show that it is not straightforward.

To make a time series \(\rho_t\) of correlations using end-of-day data, we could choose \(\rho_1\) to be the sample correlation of the returns of two stocks on day 2, \(\cdots\), 11, and \(\rho_2\) to be the sample correlation of returns of the two stocks on day 3, \(\cdots\), 12, etcetera. Alternatively we could choose \(\rho_1\) to be the sample correlation of the returns of the two stocks on day 2, \(\cdots\), 11, and \(\rho_2\) to be the sample correlation of the returns of the two stocks on day 12, \(\cdots\), 21, etcetera. With the first method we get a longer time series, which can be useful, but there is much overlap in the data points used. In particular, in the first method, the difference between \(\rho_1\) and \(\rho_2\) depends as much on the returns on day 2, as it does on the returns of day 12, which is counterintuitive. That is why we use the latter method: we calculate a time series of sample correlations using non-overlapping return data.
The autocorrelation structure of the time series of mean stock correlations on AEX and DAX (using non-overlapping data) using end-of-day data is shown in Figure 7.9 and Figure 7.10. If the time series (of mean stock correlations) was Gaussian white noise, we would expect the sample correlation to lie within the horizontal bars for 95% (19 out of 20) of the lags. In these graphs we see that the correlation is outside of the bounds for 3 of the 20 displayed lags, so it is unlikely that the time series is Gaussian white noise.

In Figure 7.11 and Figure 7.12 we show the autocorrelation of mean index correlation (as a time series where one point is the correlation of one day) calculated using intraday data. In these plots we see that there is a positive autocorrelation for a few lags, which means that the stylized fact of correlation memory and persistence seems to hold true for this data.
Figure 7.1: AEX level and mean index correlation (calculated over 10 days)

Figure 7.2: DAX level and mean index correlation (calculated over 10 days)
Figure 7.3: AEX level and mean index correlation (calculated daily using intraday data), 2010/2011

Figure 7.4: DAX level and mean index correlation (calculated daily using intraday data), 2010/2011
Figure 7.5: DAX volatility and DAX mean correlation (calculated over 10 days)

Figure 7.6: Scatter plot of DAX volatility and DAX mean correlation (calculated over 10 days)
Figure 7.7: The change of DAX correlation as calculated over two consecutive ten-day periods \((\rho(t + 11, \cdots, t + 20) - \rho(t + 1, \cdots, t + 10))\) against the index return on the day preceding the second 10-day period on which the correlation was calculated \((r(t + 10))\)

Figure 7.8: The change of DAX correlation calculated on two consecutive days using intraday data \((\rho(t + 1) - \rho(t))\) against the index return on the day before \((r(t))\)
Figure 7.9: Autocorrelation of AEX mean correlation (as calculated over 10 days). If the time series (of mean stock correlations) was Gaussian white noise, we would expect the sample correlation to lie within the horizontal bars for 95% (19 out of 20) of the lags.

Figure 7.10: Autocorrelation of DAX mean correlation (as calculated over 10 days)
Figure 7.11: Autocorrelation of AEX mean correlation (as calculated daily using intraday data)

Figure 7.12: Autocorrelation of DAX mean correlation (as calculated daily using intraday data)
Chapter 8

Fitting non-stochastic correlation models

In this chapter we will fit time series of estimated correlation to non-stochastic correlation models (i.e. models without a stochastic term) using the method of least squares. We will use four different data sets of correlation time series. We use two correlation time series calculated daily using intraday data, one for the DAX index and one for the AEX index, that we will refer to as DAX-intra and AEX-intra. We also use two correlation time series calculated using daily opening and closing data using daytime returns and overnight returns (as described at the start of Chapter 7) over periods of 10 days, that we will refer to as DAX-10 and AEX-10. For fitting DAX-10 and AEX-10 we use non-overlapping, consecutive 10-day periods. These are the same datasets that we used in Chapter 7.

8.1 Models

We have fit these four time series to a number of different models. These models use one or two autoregressive terms, possibly a mean-reversion term, possibly a return-dependent term, and a noise term. We use the following models.
Model 1: $\rho(t) = \alpha \rho(t - 1) + \epsilon(t)$
Model 2: $\rho(t) = \alpha \rho(t - 1) + \kappa (\bar{p} - \rho(t - 1)) + \epsilon(t)$
Model 3a: $\rho(t) = \alpha \rho(t - 1) + \beta r(t - 1) \mathbf{1}(r(t - 1) < 0) + \epsilon(t)$
Model 3b: $\rho(t) = \alpha \rho(t - 1) + \beta \mathbf{1}(r(t - 1) < 0) + \epsilon(t)$
Model 4a: $\rho(t) = \alpha \rho(t - 1) + \kappa (\bar{p} - \rho(t - 1)) + \beta r(t - 1) \mathbf{1}(r(t - 1) < 0) + \epsilon(t)$
Model 4b: $\rho(t) = \alpha \rho(t - 1) + \kappa (\bar{p} - \rho(t - 1)) + \beta \mathbf{1}(r(t - 1) < 0) + \epsilon(t)$
Model 5a: $\rho(t) = \alpha \rho(t - 1) + \kappa (\bar{p} - \rho(t - 1)) + \beta (r(t - 1) - \gamma) \mathbf{1}(r(t - 1) < \gamma) + \epsilon(t)$
Model 5b: $\rho(t) = \alpha \rho(t - 1) + \kappa (\bar{p} - \rho(t - 1)) + \beta \mathbf{1}(r(t - 1) < \gamma) + \epsilon(t)$
Model 6: $\rho(t) = \alpha_1 \rho(t - 1) + \alpha_2 \rho(t - 2) + \kappa (\bar{p} - \rho(t - 1)) + \epsilon(t)$

The expression $\epsilon(t)$ is the noise term. For datasets DAX-intra and AEX-intra, the notation $r(t - 1)$ stands for the index-return close-to-close on day $t - 1$. For datasets DAX-10 and AEX-10 the notation $r(t - 1)$ stands for the index-return on the day before the 10-day period on which $\rho(t)$ was calculated.

Model 1 consists of only an autoregressive (AR(1)) term. Model 2 consists of an AR(1) term and a mean reversion term. Model 3a consists of an AR(1) term, and a term dependent on the index return, which is proportional to the index return and equal to zero if the index return was greater than zero. Model 3b consists of an AR(1), and a term dependent dependent on the index, which is not proportional to the index return and equal to zero if the index return was greater than zero. Model 4a (4b) has all the terms of model 3a (3b) plus a mean reversion term. Model 5a (5b) is similar to model 4a (4b), but with the index return offset by $\gamma$. Model 6 has two autoregressive terms, and a mean reversion term.

8.2 Model fit

To perform the least squares minimization we used the ‘Global Optimization’ toolkit of Matlab. With this toolkit we estimated a global mininum via estimating local minima w.r.t. thousands of starting points. In the minimiza-
tion we restrained the parameter values to \( \alpha \in [0, 1] \), \( \alpha_1 \in [0, 1] \), \( \alpha_2 \in [0, 1] \), \( \kappa \in [0, 0.5], \kappa_1 \in [0, 1], \beta \in [0, 100], \gamma \in [-1, 1] \).

The following tables display the results of the least squares fits of the four data sets for all the models.

**Table 1: Mean squared fit error**

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.0260</td>
<td>0.0257</td>
<td>0.0127</td>
<td>0.0072</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.0191</td>
<td>0.0197</td>
<td>0.0100</td>
<td>0.0061</td>
</tr>
<tr>
<td>Model 3a</td>
<td>0.0260</td>
<td>0.0257</td>
<td>0.0127</td>
<td>0.0072</td>
</tr>
<tr>
<td>Model 3b</td>
<td>0.0260</td>
<td>0.0251</td>
<td>0.0124</td>
<td>0.0071</td>
</tr>
<tr>
<td>Model 4a</td>
<td>0.0191</td>
<td>0.0197</td>
<td>0.0100</td>
<td>0.0061</td>
</tr>
<tr>
<td>Model 4b</td>
<td>0.0191</td>
<td>0.0197</td>
<td>0.0100</td>
<td>0.0060</td>
</tr>
<tr>
<td>Model 5a</td>
<td>0.0188</td>
<td>0.0196</td>
<td>0.0100</td>
<td>0.0061</td>
</tr>
<tr>
<td>Model 5b</td>
<td>0.0188</td>
<td>0.0195</td>
<td>0.0091</td>
<td>0.0055</td>
</tr>
<tr>
<td>Model 6</td>
<td>0.0186</td>
<td>0.0194</td>
<td>0.0100</td>
<td>0.0057</td>
</tr>
</tbody>
</table>

**Table 2: Model 1 fitted parameters**

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.947</td>
<td>0.938</td>
<td>0.924</td>
<td>0.926</td>
</tr>
</tbody>
</table>

**Table 3: Model 2 fitted parameters**

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.925</td>
<td>0.983</td>
<td>0.658</td>
<td>0.691</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.500</td>
<td>0.500</td>
<td>0.113</td>
<td>0.082</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.546</td>
<td>0.449</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Table 4: Model 3a fitted parameters**

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.955</td>
<td>0.943</td>
<td>0.924</td>
<td>0.926</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.866</td>
<td>0.569</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Table 5: Model 3b fitted parameters**

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.936</td>
<td>0.894</td>
<td>0.878</td>
<td>0.872</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.012</td>
<td>0.044</td>
<td>0.028</td>
<td>0.027</td>
</tr>
</tbody>
</table>

**Table 6: Model 4a fitted parameters**

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>0.702</td>
<td>0.708</td>
<td>0.658</td>
<td>0.784</td>
</tr>
<tr>
<td>( \kappa )</td>
<td>0.273</td>
<td>0.225</td>
<td>0.113</td>
<td>0.174</td>
</tr>
<tr>
<td>( \rho )</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
<td>0.472</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.427</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

58
### Table 7: Model 4b fitted parameters

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.735</td>
<td>0.983</td>
<td>0.979</td>
<td>0.761</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.310</td>
<td>0.500</td>
<td>0.434</td>
<td>0.162</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.880</td>
<td>0.449</td>
<td>0.260</td>
<td>0.489</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.011</td>
</tr>
</tbody>
</table>

### Table 8: Model 5a fitted parameters

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.924</td>
<td>0.812</td>
<td>0.718</td>
<td>0.822</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.500</td>
<td>0.330</td>
<td>0.173</td>
<td>0.213</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.684</td>
<td>1.000</td>
<td>0.623</td>
<td>0.387</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.022</td>
<td>0.137</td>
<td>0.031</td>
<td>0.009</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.067</td>
<td>0.767</td>
<td>-0.903</td>
<td>-0.363</td>
</tr>
</tbody>
</table>

### Table 9: Model 5b fitted parameters

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.911</td>
<td>0.737</td>
<td>0.742</td>
<td>0.878</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.500</td>
<td>0.268</td>
<td>0.237</td>
<td>0.305</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.556</td>
<td>0.856</td>
<td>0.500</td>
<td>0.269</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.162</td>
<td>0.136</td>
<td>0.264</td>
<td>0.073</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.029</td>
<td>-0.032</td>
<td>-0.024</td>
<td>-0.008</td>
</tr>
</tbody>
</table>

### Table 10: Model 6 fitted parameters

<table>
<thead>
<tr>
<th></th>
<th>DAX-10</th>
<th>AEX-10</th>
<th>DAX-intra</th>
<th>AEX-intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>0.669</td>
<td>0.610</td>
<td>0.737</td>
<td>0.998</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.156</td>
<td>0.138</td>
<td>0.033</td>
<td>0.000</td>
</tr>
<tr>
<td>$\bar{\rho}$</td>
<td>0.748</td>
<td>1.000</td>
<td>0.510</td>
<td>0.212</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.307</td>
<td>0.193</td>
<td>0.220</td>
<td>0.376</td>
</tr>
</tbody>
</table>

### 8.3 Graphs of the fits
Figure 8.1: This figure shows the DAX-10 time series of correlations as black dots, and the predicted correlations according to model 5b as red dots. The blue lines between the dots indicate which predicted correlation corresponds with which measured correlation. A shorter line means a better fit.
Figure 8.2: This figure shows the AEX-10 time series of correlations as black dots, and the predicted correlations according to model 5b as red dots. The blue lines between the dots indicate which predicted correlation corresponds with which measured correlation. A shorter line means a better fit.
Figure 8.3: This figure shows the DAX-intra time series of correlations as black dots, and the predicted correlations according to model 5b as red dots. The blue lines between the dots indicate which predicted correlation corresponds with which measured correlation. A shorter line means a better fit. The months are in 2010/2011.
Figure 8.4: This figure shows the AEX-intra time series of correlations as black dots, and the predicted correlations according to model 5b as red dots. The blue lines between the dots indicate which predicted correlation corresponds with which measured correlation. A shorter line means a better fit. The months are in 2010/2011.
8.4 Discussion of results

We have fit four datasets to nine models. Table 1 shows the mean squared prediction error for each fit. We will compare the quality of the models using this table. Comparing the performance between model 1 and 2, we see that adding a mean reversion term \((\kappa(\bar{\rho} - \rho(t - 1)))\) significantly reduces the prediction error for all datasets. Comparing model 1 and 3a/b, we see that adding a return-dependent term of the form \((\beta r(t - 1)1(r(t - 1) < 0))\) or \((\beta 1(r(t - 1) < 0))\) does not seem to help much in reducing the prediction error, although the second form seems to perhaps slightly better than the first form. Comparing models 2 and 4a/b, we see that adding a return-dependent term of the above forms and a mean reversion term does not give better performance than only using a mean reversion term. Comparing models 5a/b and 4a/b we see that changing the form of the return-dependent terms to \(\beta(r(t - 1) - \gamma)1(r(t - 1) < \gamma)\) resp. \(\beta 1(r(t - 1) < \gamma)\) makes the fit slightly better. Comparing models 6 and 2, we see that adding a second autoregressive term only makes the fit slightly better. From these comparisons it follows that only term that significantly improved on the AR(1) model is the mean-reversion term, while none of the return-dependent terms seemed to have much impact.

Now let us look at figures 8.1, 8.2, 8.3, 8.4. The most notable aspect of these figures is that the predicated correlations (red dots) tend to be in the center, while the measured correlation (blue dots) have a lot of outlying values. This is possibly caused by the variance of the estimator: correlation estimators tend to have high variance, and these random errors are (by definition) something that we can not predict.

Another thing we see is that the fit errors of datasets AEX-intra and DAX-intra are lower than the fit errors of datasets AEX-10 and DAX-10. This could be caused by the fact that the correlations in AEX-intra and DAX-intra are calculated using a larger number of datapoints than the correlations in AEX-10 and DAX-10, and hence have a lower estimation error.
Chapter 9

Fitting a stochastic correlation model to daily price data

9.1 Introduction

In this chapter we try to fit a stochastic correlation model to daily price data. Unlike the previous chapter, in this chapter we take into account that the measured correlation is not equal to the underlying (‘real’) correlation, because there is an error introduced by estimating the correlation. This makes it more difficult to perform a model fit, but could lead to better models, because it is based on more realistic assumptions.

9.2 Model definition

We define our model as follows. Let $W_{1,t}, W_{2,t}, W_{3,t}$ be independent Brownian motions. We define

$$
\Delta \tilde{W}_{1,t} = \Delta W_{1,t},
$$

$$
\Delta \tilde{W}_{2,t} = \rho_t \Delta W_{1,t} + \sqrt{1 - \rho_t^2} \Delta W_{2,t},
$$

$$
\rho_{t+1} = \rho_t + \kappa (\bar{\rho} - \rho_t) + c \Delta W_{3,t+1},
$$

where $\kappa, \bar{\rho}$ and $c$ are constants satisfying $\kappa > 0$, $-1 \leq \bar{\rho} \leq 1$ and $c > 0$, and where $\Delta X_t = X_t - X_{t-1}$. Like before, we might want to add a term depending on index return to the definition of $\rho_t$, but for now we will test the model without such a term.
The interpretation of this model is that \( \tilde{W}_{1,t} \) and \( \tilde{W}_{2,t} \) are the returns of two stocks at time \( t \), and \( \rho_{t-1} \) is the correlation between the stock returns at time \( t \). In our analysis we consider the stock returns \( \Delta \tilde{W}_{1,t} \), \( \Delta \tilde{W}_{2,t} \) to be observable, and the processes \( \Delta W_{2,t}, \Delta W_{3,t}, \rho_{t} \) to be unobservable.

Note that \( \text{var}(\Delta \tilde{W}_{1,t}) = \text{var}(\Delta \tilde{W}_{2,t}) = 1, \ E(\Delta \tilde{W}_{1,t} \Delta \tilde{W}_{2,t}) = \text{E}\rho_{t-1} \) and \( E(\Delta \tilde{W}_{1,t} \Delta \tilde{W}_{2,t} | \rho_{t-1}) = \rho_{t-1} \). The random variables \( (\Delta \tilde{W}_{1,t}) \) are i.i.d. standard normal, as are \( (\Delta \tilde{W}_{2,t} | \rho_{t-1}) \). It is less clear what the unconditional distributions of \( (\Delta W_{2,t}) \) are, but numerical testing using the Kolmogorov-Smirnov test does not show a notable deviation from standard normality.

9.3 Parameter estimation

Estimating the parameters \((\kappa, \bar{\rho}, c)\) of the model is difficult, because the process \( \rho_{t} \) is not directly observable. We have tried different methods for performing the estimation. In the following subsections we describe the different approaches and their results.

9.3.1 Parameter estimation using sample correlation

It seems tempting to use sample correlation calculated using a number of days’ return data as a starting point to estimate the parameters of our model. However, there seems to be a problem with this approach.

If we calculate sample correlation based on data with a constant real correlation, then the sample correlation has a variance that depends on the underlying real correlation, as we saw in Chapter 6. If we calculate sample correlation based on data where the real correlation itself also has a variance, then the sample correlation has two sources of variance: the variance introduced by the estimating process, and the variance introduced by the variance of the real correlation. If we know what the variance introduced by the estimator is, then we can untangle these two sources of variance. However, the variance introduced by the estimator depends on the real correlation, which is unknown. This makes it difficult to determine the variance of correlation when using sample correlation as a starting point.
9.3.2 Parameter estimation using the EM-algorithm

The Expectation-Maximization algorithm (EM-algorithm) is a method for finding estimates of parameters in models with unobserved variables. The classical paper on this algorithm is [Dempster et al., 1977], and a recent survey article is [Gupta and Chen, 2010].

First we describe the algorithm. Let $\theta$ be the set of the parameters that are to be estimated, $D$ the observable data, and $H$ some unobservable data (not necessarily all unobservable data). Let $\theta^{(0)}$ be an initial guess for $\theta$. The steps of the standard EM-algorithm are

1. Calculate the conditional density $p(H|D, \theta^{(0)})$.
2. Calculate $\theta^{(1)} = \arg\max_{\theta} \int \log(p(D,H|\theta))p(H|D,\theta^{(0)})dH$.
3. Set $\theta^{(0)} = \theta^{(1)}$ and go back to Step 1.

The steps where a conditional density is calculated are called ‘expectation steps’, and the steps where a new $\theta^{(1)}$ is calculated are called ‘maximization steps’, hence the name EM-algorithm.

There is also a simplified version of the EM-algorithm, called the point-estimate variant, of which the steps are

1. Calculate $H^{(0)} = \arg\max_{H} p(H|D, \theta^{(0)})$.
2. Calculate $\theta^{(1)} = \arg\max_{\theta} p(H^{(0)}, D|\theta)$.
3. Set $\theta^{(0)} = \theta^{(1)}$ and go back to Step 1.

We want to apply these algorithms to our model. We use parameters $\theta = \{\kappa, \bar{p}, c\}$, observable data $D = \{\Delta \bar{W}_{1,t}, \Delta \bar{W}_{2,t}\}_{t=1, \ldots, T}$, and unobservable data $H = \{\rho_t\}_{t=0, \ldots, T-1}$. This gives us

$$p(H|D, \theta) = p(\{\rho_t\}_{t=0, \ldots, T-1}|\{\Delta \bar{W}_{1,t}, \Delta \bar{W}_{2,t}\}_{t=1, \ldots, T}; \kappa, \bar{p}, c)$$

$$= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi(1-\rho_t^2)}} \exp \left( -\frac{1}{2} \left( \frac{\Delta \bar{W}_{2,t} - \rho_{t-1} \Delta \bar{W}_{1,t}}{\sqrt{1-\rho_t^2}} \right)^2 \right)$$

$$\cdot \prod_{t=0}^{T-2} \frac{1}{\sqrt{2\pi c^2}} \exp \left( -\frac{1}{2} \left( \frac{\rho_{t+1} - \rho_t - \kappa(\bar{p} - \rho(t))}{c} \right)^2 \right),$$
and
\[ p(D, H|\theta) = p(\{\rho_t\}_{t=0,T-1}, \{\Delta \tilde{W}_{1,t}, \Delta \tilde{W}_{2,t}\}_{t=1,T}|\kappa, \bar{\rho}, c) \]
\[ = p(\{\rho_t\}_{t=0,T-1}|\kappa, \bar{\rho}, c) \]
\[ = \prod_{t=0}^{T-2} \frac{1}{\sqrt{2\pi c^2}} \exp \left( -\frac{1}{2} \left( \frac{\rho_{t+1} - \rho_t - \kappa(\bar{\rho} - \rho(t))}{c} \right)^2 \right), \]
so
\[ \log(p(D, H|\theta)) = \sum_{t=0}^{T-2} -\frac{1}{2} \log(2\pi) - \log(c) - \frac{1}{2} \left( \frac{\rho_{t+1} - \rho_t - \kappa(\bar{\rho} - \rho(t))}{c} \right)^2 \]
So, to apply the standard EM-algorithm to our model, we need be able to calculate
\[ \arg\max_{\theta} \int \log(p(D, H|\theta))p(H|D, \theta) dH \]
\[ = \arg\max_{\kappa, \bar{\rho}, c} \int_{[-1,1]^T} \left( \sum_{t=0}^{T-2} -\frac{1}{2} \log(2\pi) - \log(c) - \frac{1}{2} \left( \frac{\rho_{t+1} - \rho_t - \kappa(\bar{\rho} - \rho(t))}{c} \right)^2 \right) \]
\[ \cdot \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi(1-\rho_t^2)}} \exp \left( -\frac{1}{2} \left( \frac{\Delta \tilde{W}_{2,t} - \rho_{t-1}\Delta \tilde{W}_{1,t}}{\sqrt{1-\rho_{t-1}^2}} \right)^2 \right) \]
\[ \cdot \prod_{t=0}^{T-2} \frac{1}{\sqrt{2\pi(c)^2}} \exp \left( -\frac{1}{2} \left( \frac{\rho_{t+1} - \rho_t - \kappa_i(\bar{\rho} - \rho(t))}{c} \right)^2 \right) \]
\[ \cdot \prod_{i=0}^{T-1} \rho_0 \cdot \rho_1 \cdot \cdots \cdot \rho_{T-1} \]
We have not been able to find an explicit analytic solution to this problem. It is not not allowed to invert the order of argmax and integration, and the integral does not seem to have an analytic solution. We have tried to find a solution to the subproblem of integrating the integrand with respect to \( \rho_i \) (for some \( i \)) using Matlab (symbolic toolkit), and the software was not able to find an explicit solution.

To apply the point-estimate variant of the EM-algorithm to our model, we need to evaluate the expressions
\[ \arg\max_{\theta} p(H|D, \theta) \]
\[ = \arg\max_{\rho_0, \cdots, \rho_{T-1}} \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi(1-\rho_t^2)}} \exp \left( -\frac{1}{2} \left( \frac{\Delta \tilde{W}_{2,t} - \rho_{t-1}\Delta \tilde{W}_{1,t}}{\sqrt{1-\rho_{t-1}^2}} \right)^2 \right) \]
\[ \cdot \prod_{t=0}^{T-2} \frac{1}{\sqrt{2\pi c^2}} \exp \left( -\frac{1}{2} \left( \frac{\rho_{t+1} - \rho_t - \kappa(\bar{\rho} - \rho(t))}{c} \right)^2 \right), \]
and

$$\operatorname{argmax}_\theta p(H, D|\theta) = \operatorname{argmax}_{\kappa, \rho, c} T^{-2} \prod_{t=0}^{T-2} \frac{1}{\sqrt{2\pi c^2}} \exp \left( -\frac{1}{2} \left( \frac{\rho_{t+1} - \rho_t - \kappa (\overline{\rho} - \rho(t))}{c} \right)^2 \right).$$

We have implemented the point-variant of the EM-algorithm in matlab using numerical estimation for both argmax steps. A problem with this approach is that one can never know for sure if a numerically estimated maximum is (close to) the real maximum, or if it is only (close to) a local maximum. You can try find the maximum by calculating the local maximum from different starting location, but you can never be sure if you have found the real maximum (using numerical methods).

To test the quality of this estimation method we generated return data based on our model, to see if the estimation method can recover the parameters used to generate the data.

The test results are that under this method, the estimated correlations $\hat{\rho}_0, \cdots, \hat{\rho}_{T-1}$ converge to values that are too close to each other (close to $\overline{\rho}$), the estimated volatility of correlation $\hat{c}$ converges to 0, the estimated speed of mean reversion $\bar{k}$ converges to a value lower than the value of $k$ used to generate the data, and the estimated level of mean reversion $\overline{\rho}$ converges to a value close to the value of $\overline{\rho}$ used to generate the data. The fundamental problem seems to be the apparent inability of the algorithm to make a reasonable estimate of the volatility of correlation $c$, which in turn leads to estimates of the correlation that are too close together, which then leads to the estimate of the speed of mean reversion $k$ being too low.

Possibly we can sidestep this estimation problem by estimating the volatility of correlation of $c$ separately, for instance by using the volatility of sample correlation, and then using the EM algorithm to estimate the remaining parameters. This is also not an easy problem. Firstly, there is the problem mentioned in Section 9.3.1. Another problem is that the volatility of $\sum_{i=1}^n \rho_{t-i}/n$ under our model not only depends on $c$, but also on $k$, the speed of mean reversion (a higher speed of mean reversion makes it less volatile), making it difficult to estimate $c$ first and $k$ later.
9.3.3 Parameter estimation using correlation as parameters

Another approach at parameter estimation is to consider the set \( \{ \kappa, \bar{p}, c, \rho_0, \rho_1, \cdots, \rho_{T-1} \} \) to be our parameters, and estimate them all together using maximum-likelihood estimation. This means we no longer make a distinction between parameters and missing data. The corresponding likelihood function is

\[
\mathcal{L}(\kappa, \bar{p}, c, \rho_0, \rho_1, \cdots, \rho_{T-1} | (\Delta \tilde{W}_{1,t}, \Delta \tilde{W}_{2,t})_{t=1,\cdots,T}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi(1-\rho_t^2)}} \exp \left( -\frac{1}{2} \left( \frac{\Delta \tilde{W}_{2,t} - \rho_{t-1}\Delta \tilde{W}_{1,t}}{\sqrt{1-\rho_t^2}} \right)^2 \right) \cdot \prod_{t=0}^{T-2} \frac{1}{\sqrt{2\pi c^2}} \exp \left( -\frac{1}{2} \left( \frac{\rho_{t+1} - \rho_t - \kappa(\bar{p} - \rho(t))}{c} \right)^2 \right).
\]

We implemented this estimation method in Matlab using numerical approximation. On testing the quality of the estimator using return data generated based on our model, we find the same results as in the case of the point-variant version of the EM estimator: the (numerical approximation of) the value of \( c \) that maximizes the likelihood function is close to zero, the estimated correlation are too close together, and the estimated variance of correlation is too low.

9.3.4 Conclusion

The approach we outlined in this chapter is promising, because it takes into account the fact that correlation is not directly observable. However, it is complex to fit such models, and we were not successful at finding an approach that works in practice. If such an approach was found, it might become the preferred way to model correlation.
Chapter 10

Conclusion

In this thesis we researched the modeling of correlations between stock returns. The reason we wanted to understand this subject is because it is directly linked to understanding the volatility of index options, which is in turn relevant for index option pricing. To use correlations to find the volatility of the index, we need to know the correlation between all pairs of stocks that are part of the index. That is why we chose to investigate a model that accounts for all correlations in a stock market, namely the single factor model, where all correlations are considered to be equal. This is not a realistic assumption, but it might be acceptable, since in principle this model could still generate a correct index volatility, and that is what we are really interested in.

According to experience by traders, correlation as a time series is a mean reverting process, and is related to index return. That is why we have constructed single factor models with mean reversion, and a term related to index return. The hope was that using index returns can lead to a correlation model with better predictive ability. We have tried to fit such models to market data, to test their predictive ability.

First we have fit time series of estimated correlation to models without stochastic terms in chapter eight. We found that a mean reversion term can be used to improve the predictive quality of the model, while adding a return-dependent term did not seem to have much effect. A possible reason is that the error introduced by estimation of the correlation (we can not observe correlation directly) overshadows the effect of the index return, making it difficult to notice how the correlation changes due to index returns because of this estimation noise.
A possible solution to this problem would be to take this estimation error into account when we perform the fit. This is something we have studied in chapter nine. Here we have outlined some methods by which this kind of fit could be performed. This is however quite complex, and we were not able to turn it into a practically usable model.

It turns out that the modeling/predicting of correlation is a very difficult problem. Time series of estimated correlation always fluctuate strongly. Presumably this is partially due to estimation error, and partially due to the underlying ‘real’ correlation fluctuating strongly itself.

As for possible solutions ways to overcome this, one approach could be to find better methods to estimate correlation, to reduce estimation error, making it easier to fit the model properly. Another approach could be to find ways to fit correlation model while taking into account estimation error, as we attempted in chapter nine. Yet another approach would be to formulate stochastic correlation models that imply the same probability distributions of correlation as we see in the markets, without worrying about prediction. However, dropping the requirement that the model be predictive would make it less useful in practice.

Hopefully this thesis led to an increased understanding of the modeling of stock return correlation, in particular about the difficulties involved with this.
Bibliography


