ETH
Swiss Federal Institute of Technology Zurich

Seminar for Statistics

Department of Mathematics

Master Thesis Spring 2012

Markus Harlacher

Cointegration Based Statistical Arbitrage

Submission Date: July 13th 2012

Co-Adviser Dr. Markus Kalisch
Adviser: Prof. Dr. Sara van de Geer

To Natali

Preface

Studying at the Swiss Federal Institute of Technology (ETH) in Zurich is probably not always easy, but it is always a pleasure. That's my impression when I look back on the time I spent at the ETH. I had the opportunity to enjoy demanding lectures taught by didactically and technically brilliant people. To all those lecturers and researchers I would like to express my sincere admiration and deepest gratitude.

Special thanks are due to Markus Kalisch and Sara van de Geer for their willingness to support my master's thesis. Thank you Markus for your enriching comments and the excellent cooperation in general! Not forgetting the hardly ever beatable working environment provided by the SfS. A perfect LaTeX-template as well as any kind of guidance, manuals with useful tips and tricks for anything one could be in need of, everything was just there. In this respect I also want to mention the generous access I was granted to the HPC Cluster Brutus. Running a distributed memory program on sometimes up to 100 cores in parallel was indeed a very exciting and unique experience.

Finally, I would also like to thank Werner Stahel and Alois Gisler who considerably enriched my mind and broadened my intellectual horizon with their highly appreciated courses.

Abstract

This thesis analyses a cointegration based statistical arbitrage model. Starting with a brief overview of the topic, a simulation study is carried out that is intended to shed light on the mode of action of such a model and to highlight some potential flaws of the method. The study continues with a back-testing on the US equity market for the time period reaching from 1996 up to 2011. The results of all the different model versions that were tested look quite promising. "Traditional" mean-variance based performance measurements attest the employed cointegration based statistical arbitrage model very good results. The advanced dependence analysis with respect to the returns of the S&P 500 index and the returns obtained from the back-testing shows a very favourable structure and indicates that such a model can provide returns that are only very weekly related to the returns of the S&P 500 index.

Contents

List of Figures

List of Tables

Notation

Δ	Change in value
$e^1 = \exp\{1\}$	Euler's number
log	Logarithm with respect to base e
$\stackrel{\mathrm{d}}{=}$	Equal in distribution
$N(\mu, \sigma^2)$	Gaussian distribution
$E[X \mid W]$	Conditional expectation of X, given $W = w$
$\mathbb{R}_+$	Non-negative real numbers
$\mathbb{R}^n$	n-dimensional Euclidean space
i.i.d.	Independent and identically distributed
SDE	Stochastic differential equation
A^T	Transposed of matrix A
$tr(A)$	Trace of matrix A
$\boldsymbol{\nabla}$	Function gradient: $\boldsymbol{\nabla} f = \left(\frac{\partial f}{\partial x_1} \dots \frac{\partial f}{\partial x_d}\right)^T$
∇^2	Hessian matrix
HJB	Hamilton-Jacobi-Bellman equation
p.a.	per annum
$\overline{r}_i$	Average return of asset i
$\lvert R \rvert$	Determinant of matrix R
$\mathbb{I}_{\mathcal{M}}$	Characteristic function of the set $\mathcal{M}$
$N^-(\lambda, \chi, \psi)$	Generalised inverse Gaussian distribution
$GH(\lambda, \chi, \psi, \mu, \sigma^2, \gamma)$	Generalised hyperbolic distribution

Chapter 1

Introduction

In financial management the term *Statistical Arbitrage* encompasses a variety of investment strategies with, however, some particular common features. All the strategies are based on statistical modelling and typically have an almost market neutral trading book. A particular form of the strategy (sometimes also referred to as *Pairs Trading*) tries to determine pairs of assets where one of them is purchased while the other is sold without actually owning it. In this particular form, two assets with a close price relationship are traded. If the two prices move largely together, the general idea is to assume a mean reverting behaviour of the corresponding price spread. Thus, when the price of one asset increases relatively to the price of the pair-asset the strategy suggests to *sell* that asset *short* (i.e. borrowing that particular asset and selling it) and use the proceeds to purchase (i.e. *go long* in) the pair-asset. If the price spread approaches again its equilibrium level, the strategy leads to a gain. Such a strategy clearly belongs to the broad family of so-called *long/short* investments. A single pair position, however, does not necessarily have to be market neutral[1], although a portfolio of pairs may be. An explicit market neutral investing is a portfolio optimization exercise with the particular aim to achieve a negligible market exposure. Nevertheless, a statistical arbitrage portfolio should, if it really works, show a very distinct behaviour with respect to movements in the overall financial market, at least as opposed to "conventional" investments consisting of only long positions, and may, in principle, also lead to a profit when the overall financial market suffers heavy losses.

1.1 An Investment Strategy

One of the first scientific papers mentioning a typical statistical arbitrage strategy was the working paper written by Gatev, Goetzmann, and Rouwenhorst (1999), which was finally published in the Review of Financial Studies seven years later. In the years between, there were several scientific papers and even books published with respect to the topic. However, on *Wall Street* the strategy had been known already for many years before 1999 and has been practised vividly by hedge funds as well as investment banks ever since. Hence, the idea of statistical arbitrage as an investment strategy is definitely not a brand new concept.

[1]The term *market neutral* means that the returns of such a strategy are independent of the overall market returns.

So what is the justification for writing a thesis about this topic? First of all, an investment with an unconventional behaviour concerning the broad financial market is probably a very tempting one, especially with respect to portfolio diversification effects. This is presumably the reason why such investments have become so attractive in the past view years, not only for institutional investors but also for private ones.[2] The task to be done is, thus, not just to see whether a statistical arbitrage investment strategy leads to improved returns, as compared to a passive investment in the broad market[3], but also to shed light on the mode of action of the main concept behind it. The most interesting point is whether the strategy is actually able to hold what its particular construction concept promises. Analysing this is, however, not a trivial undertaking as the disclosure of relevant information is not compulsory for hedge funds and, thus, somewhat arbitrary. Hence, there is some reasonable suspicion that publicly available data on hedge funds suffer from a so-called *survivorship bias*. That is, bad performing funds do not publish. However, with the knowledge about the strategy pursued, we can easily calculate a historical time series of corresponding returns by ourselves. It is, of course, difficult to tell in every detail how the investment strategy of a particular hedge fund actually looks like. Even though it seems reasonable to assume that there must be some essential similarities among all the funds engaged in a quantitative statistical arbitrage strategy, the generated returns structure of particular funds may differ considerably from each other.

The analysis of this thesis is limited to *cointegration based* statistical arbitrage. A short description of the idea behind it is given in chapter 2 where we also briefly discuss some other popular models. Chapter 3 provides then a simulation study that is intended to discover some crucial details with respect to the implementation of a statistical arbitrage model. The basic question there is whether at all or in which particular cases the employed method really works in the way it is expected to. This also gives us the opportunity to examine some particular situations where it is not clear *ex ante* how well the chosen method is able to handle them. With the results of the simulation study in mind, we then implement in chapter 4 a simple cointegration based statistical arbitrage investment strategy and do a proper back-testing on the US stock market for the time period reaching from 1996 up to 2011.

1.2 Performance Measurement and Benchmark Dependence

Proposing an actively managed investment is one thing, evaluating its performance is another. At first sight, one might think that this is an easy task. An investment is rewarding if it leads to high profits. So, the higher the profit the better the strategy. Such a perception is, however, far from being adequate. Comparing the returns of two investments makes only sense if the two investments have similar risk profiles and are also bound by similar constraints. Comparing the returns of a portfolio that may only consist of so-called *investment-grade* bonds with an unrestrained portfolio is definitely not appropriate. This becomes clear when we ask the simple question of how high the profit of an investment will be in the next period. Of course, that is something we do not know ex ante. So, we have to estimate it. This is usually done by looking at the past. However, high

[2] Small investors are usually not allowed to engage directly in hedge funds, but they can participate indirectly through other investment funds.

[3] What this term actually means has to be discussed in some more detail later when we discuss about appropriate performance measures.

profits in the past may have just been realized by chance. High profits may be realized by a high risk exposure. So, it is possible that one is just lucky this period, gaining a large profit, but getting a bad break in the next period, loosing possibly everything one has. By using a highly leveraged version of a benchmark portfolio it is not difficult for a fund manager to achieve a higher return than a simple, non-leveraged investment in the same benchmark portfolio when the benchmark performs well in general. Of course, the same fund manager would also lose much more than the non-leveraged investor in case the benchmark moves down. So, the performance measurement of an investment should always consider the risks the investment is exposed to. How risk is appropriately measured is, however, not a question that can be generally answered. It depends quite a lot on the investor's situation and his particular preferences. An easy and (probably therefore) favoured concept is the standard deviation of the return time series.

Another critical issue with respect to a particular investment is its contribution to the investors overall portfolio diversification. Especially institutional investors aim for a well diversified financial portfolio as this is intended to protect them against losses from individual assets. The very popular portfolio concept proposed by Markowitz (1953) and Tobin (1958) illustrates the advantages of portfolio investments in a simple *mean-variance* framework. That is, one considers only the mean and the variance of the returns of the individual assets as well as their pairwise Pearson-correlations. This may be suitable, though still far from perfect, concerning classical investments, as passive stock and bond investments. With respect to hedge fund strategies there are, however, considerable doubts whether the dependence structure with respect to other financial assets can be properly captured by the concept of linear correlation. Market observations suggest that the bivariate return distribution of a "typical" statistical arbitrage portfolio and a broad stock market index is considerably non-elliptical.

In chapter 5 there is brief introduction to some "standard" performance measurement tools, which are then applied on the back-testing-results as obtained in chapter 4. The main part of chapter 5 is, however, devoted to some more flexible techniques aiming for an appropriate description of the dependence structure between the analysed statistical arbitrage strategies and a broad market portfolio. This is a very important aspect with respect to the performance measurement as well as the diversification analysis in this field, especially if one wants to avoid being fooled by misleading results of inappropriate analysis tools.

Chapter 2

Different Ideas behind Established Models

Before we start with the examination of the cointegration based statistical arbitrage, we first want to glance over the different ideas behind some "established" statistical arbitrage models. Basically, one can distinguish three broad approaches. They have been released in books of established publishers or have been published in well known peer-reviewed journals. Certainly, as the statistical arbitrage investment strategy has been established since many years, there are probably much more approaches, which are, however, either less well known, as they are strongly proprietary, or less successful and, thus, not used. So, the term *established methods* refers in this analysis to the main methods publicly accessible.

The first method to be reviewed is the one proposed by Gatev, Goetzmann, and Rouwenhorst (2006). It is based on the construction of a cumulative total returns index for each asset that is part of the considered investment universe. A matching partner for each asset is then found by minimising the sum of squared deviations between the two normalized price series. Thus, we shall refer to this method as *Minimum Distance* as it uses nothing else than the squared Euclidean norm. Its mode of action is quickly discussed in section 2.1.

The second method, which is suggested by Vidyamurthy (2004), uses the *Arbitrage Pricing Theory* (APT) initially proposed by Ross (1976). Unlike the the minimum distance method and the cointegration approach, it is not a pure statistical method. With the APT at its core, the strategy is based on a well established asset pricing theory, which leads to the assumption that it might be less prone to spurious results. How it exactly works is explained in section 2.2.

The third method, which is the method we analyse in some more detail in this thesis, is based on the well known time series concept of *cointegration*. It is mentioned in several books, as for example by Vidyamurthy (2004) or Wilmott (2006), and is, according to them, one of the most frequently applied method by practitioners to detect marketable pairs. Its basic idea is discussed in section 2.3

2.1 The Minimum Distance Method

The implementation of this strategy consists of two stages, i.e. the pair-formation period and the trading period. Gatev et al. (2006) choose rather arbitrarily a period of 12 months to form their pairs and then trade them during a period of 6 months. Taking only relatively liquid stocks, they construct a cumulative total return index based on daily observations for each stock over the 12-months formation period[1]. They then choose a matching partner for each stock by taking the security that minimises the sum of squared deviations between the two normalized price series. Once all stocks are paired up, they then take the top pairs with the smallest distance for the trading period.

Having determined marketable pairs, the next step is to choose a trading rule, i.e. a rule that tells us when to open a position and when to close it again. Gatev et al. (2006) take a rather simple rule by just opening a position in a pair, by going one dollar long in the lower-priced stock and one dollar short in the higher-priced stock, when the prices diverge by more than two historical standard deviations, as estimated during the formation stage. The position is unwinded at the next crossing of the prices. After 6 months a position is closed in any case, no matter whether any price convergence has taken place or not.

The empirical results obtained by the analysis of Gatev et al. (2006) with respect to the US equity market during 1962-2002 show average annualised excess[2] returns of up to 11% for portfolios of pairs. This is large and suggests that such a strategy is quite profitable. They also report some diversification benefits for portfolios consisting of several pairs. This is, as the number of pairs in a portfolio increases, the standard deviation of the portfolio is reduced. Interestingly, with an increased number of pairs they observe an increase in the minimum realised return, but a relatively stable maximum realised return.

Although a considerable excess return is clearly desirable from an investor's point of view, it is not the only crucial aspect. As one can consider such a strategy as an alternative investment, it has to be seen as just one asset class among many others. For diversification reasons, a typical investor would only hold a fraction of his total wealth in a particular class and is, thus, also interested in the co-movements of this class with other investment classes. Even though Gatev et al. (2006) provide some additional univariate statistics, the multivariate behaviour with regard to a well diversified portfolio, or components of it, is not part of their study.

Another potentially interesting aspect might be the sensitivity of the results with respect to changes in the distance measure. As already mentioned, their pair-formation method is just the squared Euclidean distance. As such it is sensitive to larger values. A less sensitive measure would be, for example, the Manhattan distance. So, an interesting question may also be, by how much and in which way the results would change if we used distinct distance measures. Basically, it is a very crucial point how potential pairs are determined. One principal characteristic of this method is that it makes no underlying model assumptions. The disadvantage of this is, however, that it basically restricts the estimation of the spread convergence time or the expected holding period to a simple historical average.

[1] Reinvested dividends are included in the index

[2] *Excess* means in this context just the return component that exceeds the return of a risk-free investment.

2.2 The Arbitrage Pricing Theory

Another way to detect marketable pairs, suggested by Vidyamurthy (2004), is to employ the arbitrage price theory (APT), initially proposed by Ross (1976) to determine asset prices. Assuming the *law of one price* to hold, and market participants having homogeneous expectations, it claims that the returns on any stock is linearly related to a set of so-called risk factors, i.e.

$$r_i = r_f + \sum_{j=1}^{k} \beta_{i,j} r_j^* + \epsilon_i \tag{2.2.1}$$

where r_j^* is the return contribution of risk factor j and $\beta_{i,j}$ is the risk exposure of asset i to risk factor j. r_f denotes the risk-free return. The remaining part, i.e. the ϵ_i, can be interpreted as the return component arising from the idiosyncratic or specific risk of asset i. Its expected value should be zero. Furthermore, it must hold that

$$E\left[\epsilon_i \epsilon_h\right] = 0 \text{ , for all } i \text{ and } h \text{ where } i \neq h \tag{2.2.2}$$

$$E\left[\epsilon_i \left(r_j^* - E\left[r_j^*\right]\right)\right] = 0 \text{ , for all assets and risk factors.} \tag{2.2.3}$$

Without repeating the whole APT, which is a very interesting theory on its own, the relevant statement in the context of a statistical arbitrage strategy is that assets with virtually identical risk exposures should yield very similar returns. In short, it may be possible to detect tradeable pairs according to the risk exposure of the assets to particular risk factors. How this would look like in practice is, however, not really explained by Vidyamurthy (2004). He proposes the method but without becoming more specific on the execution of it. So, which risk factors one has to take and how one should assess the different exposures concerning the pair detection remain unanswered questions.

Unlike the purely statistical methods of the sections 2.1 and 2.3, the strategy of this section uses a well known asset pricing theory with corresponding economic reasoning. This is certainly a very nice feature as it bases the strategy on a well founded theory, and should, therefore, be less prone to spurious results. It also reveals the spirit of statistical arbitrage strategies. In order to make up a tradeable pair, two assets are assumed to have very similar exposures to some relevant risk factors. Therefore, the long position in one asset is effectively hedged by an offsetting short position in another asset with very similar risk factor exposures, reducing the systematic risk of the pair position to a minimum. Nevertheless, this nice property with respect to the returns of the assets does not guarantee that there is a linear combination of them which shows a mean-reverting behaviour. It is therefore questionable whether the APT on its own, without any further restrictions, is really appropriate regarding a statistical arbitrage trading strategy. The method may be used in combination with some other methods, as the ones discussed in this thesis, for example. In this case one may use the APT-method to obtain a smaller subset of theoretical pair-candidates and then applies another appropriate method on these subsets. This reduces the total number of possible combinations to calculate through by the other method and is, thus, thought to be more efficient. In this case one could also extend the pairs-concept in the sense that one uses more than two assets to obtain a mean-reverting spread. That is, one looks for a linear combination of a subset of assets that is mean-reverting. In this case, with a given set of relevant risk factors, an obvious way to proceed would be to define a distance measure with respect to the β-values of relation

(2.2.1) and then determine a subset of similar assets accordingly. In order to do that, it is advisable to standardise the risk factors before estimating the β-values so that they are on the same scale. The assets that are in close distance to each other make a subset and one can try to find suitable linear combinations among them.

2.3 Cointegration for Pairs Detection

As described in section 2.1, the minimum distance method makes no model assumptions. The cointegration method is, in contrast, a model based parametric approach. The idea here is that if two asset prices follow a common stochastic trend, the spread between them may be weakly stationary. More precisely, if we have two time series that are both integrated of order d and there is a linear combination of the two that gives a time series which is integrated of order $d - b$, with $b > 0$, then the two series are said to be cointegrated, i.e. $CI\,(d, b)$. The relevant cases in the context of a statistical arbitrage investment strategy are the ons where we have $d - b = 0$, i.e. a stationary time series for the spread. As many prices have times series that are integrated of order one, $I(1)$, the focus will be on the cases where $d = b = 1$. The nice thing with such cointegrated series, X_t and Y_t, is that they can be represented in an *error correction model* (ECM) where the dynamics of one time series at the current point in time is a correction of last period's deviation from the equilibrium plus some possible lag dynamics, i.e.

$$\begin{aligned}\Delta y_t &= \psi_0 + \gamma_y\left(y_{t-1} - \alpha - \beta x_{t-1}\right) + \sum_{i=1}^{K}\psi_{x,i}\Delta x_{t-i} + \sum_{i=1}^{L}\psi_{y,i}\Delta y_{t-i} + \epsilon_{y,t}\\ \Delta x_t &= \xi_0 + \gamma_x\left(y_{t-1} - \alpha - \beta x_{t-1}\right) + \sum_{i=1}^{K}\xi_{y,i}\Delta y_{t-i} + \sum_{i=1}^{L}\xi_{x,i}\Delta x_{t-i} + \epsilon_{x,t}\end{aligned} \tag{2.3.1}$$

with $\epsilon_{y,t}$ and $\epsilon_{x,t}$ being white noise. If there is no deterministic trend in the series, the two constants ψ_0 and ξ_0 are zero. The big advantage of this model is that active forecasts can be carried out quite easily based on past information. From (2.3.1) it becomes clear that the part that represents the deviation from the long-run equilibrium[3], i.e. $(y_{t-1} - \alpha - \beta x_{t-1})$, must be weakly stationary and the two coefficients γ_y and γ_x must have opposite algebraic signs. Otherwise there is no error-correcting behaviour. To test this, and for cointegration in general, one can follow the procedure proposed by Engle and Granger (1987). It consists of two steps. First, one runs a linear regression of one series on the other, i.e.

$$y_t = \alpha + \beta x_t + \varepsilon_t\ ,\ \text{for } t = 1, \ldots, T \tag{2.3.2}$$

with ε_t being the error term. If the two variables Y and X are integrated of order one, the error term of the regression in equation (2.3.2) must be weakly stationary (i.e. $I(0)$) in order that X and Y are cointegrated. To check that, one usually employs a statistical test, as for example the augmented Dickey-Fuller test (ADF), or just any other test procedure designed to detect weak stationarity. There is, however, a word of cautious in this regard as the series to be tested is based on estimated coefficients with respect to the cointegration relation. Therefore, one should use the critical values provided by Phillips and Ouliaris (1990) instead. Another problem with the cointegration approach in the context of statistical arbitrage is that we do not know which asset to take on the left

[3] The long-run equilibrium is the value around which the spread values fluctuate and is denoted here by a.

hand side of equation (2.3.2) and which on the right hand side. Obviously, this matters in a test like the one proposed by Engle and Granger (1987). Thus, when considering only pairs, one would have to test in both ways. In case one wants to trade not just on the spread of a pair but on the spread of a whole portfolio the testing becomes cumbersome and expensive. Phillips and Ouliaris (1990) propose a test procedure that is independent of this choice[4]. It is based on a first order vector autoregressive model (VAR) of the form

$$\boldsymbol{z}_t = \hat{\Pi}\boldsymbol{z}_{t-1} + \hat{\boldsymbol{\xi}}_t \tag{2.3.3}$$

with $\boldsymbol{z}_t = (y_t \; x_t)^T$. The test statistic is calculated as

$$\hat{P}_z = Ttr\left(\hat{\Omega}M_{zz}^{-1}\right) \tag{2.3.4}$$

where

$$\hat{\Omega} = \frac{1}{n}\sum_{t=1}^{n}\hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_t^T + \frac{1}{n}\sum_{s=1}^{l}\omega\left(s,l\right)\sum_{t=s+1}^{n}\left(\hat{\boldsymbol{\xi}}_t\hat{\boldsymbol{\xi}}_{t-s}^T + \hat{\boldsymbol{\xi}}_{t-s}\hat{\boldsymbol{\xi}}_t^T\right) \tag{2.3.5}$$

with weighting function $\omega\left(s,l\right) = 1 - \frac{s}{l+1}$, which depends on a user chosen lag window l, and $M_{zz} = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{z}_t\boldsymbol{z}_t^T$.

Vidyamurthy (2004) uses log-prices[5] and does not necessarily premise on the cointegration condition. Instead, he just looks for evidence of mean reversion in the spread time series, defined as

$$m_t = \log(P_t^A) - \beta\log(P_t^B)\ . \tag{2.3.6}$$

If the two log-prices were cointegrated, this series would read

$$m_t = \alpha + \varepsilon_t \tag{2.3.7}$$

with respect to equation (2.3.2), meaning that the spread series moves around the equilibrium value α, as ε_t is assumed to be weakly stationary, though not necessarily i.i.d.

Basically, from the view of a pairs-trader, the spread series does not necessarily have to be weakly stationary in a statistical sense. Certainly, if it is, it is much simpler to find an effective trading rule. But in principle, the spread series just needs to be mean-reverting. There is, however, a problem with approaches that do not use the statistical cointegration concept properly. Just regressing two non-stationary time series on each other and then model the residuals as a mean reverting process is dangerous, as estimator consistency is not guaranteed in this case. So the estimated spread series could be far away from the true series, making any trading rule based on it unreliable or even completely corrupt.

In contrast to the distance method of section 2.1, the cointegration method is not limited to the detection of a mean-reverting spread between only two assets, but can, in principle, also be used to detect a mean-reverting equilibrium relationship among a larger pool of assets. The advantage of this is that there are many more combinations possible, increasing the likelihood of detecting more marketable combinations. A practical question is then, however, how one should proceed in such a case from a computational point of view. Calculating through all possible combinations might be only feasible for a very limited

[4] This version of the proposed test is based on a multivariate trace statistic.
[5] This is $\log\left(P_t\right)$.

asset universe. In this case, the APT, as already mentioned in section 2.2, might serve as a pre-selection method in order to reduce the set of possible combinations to subsets consisting of "similar" assets.

One last remark concerns the question of whether one should use transformed prices, as suggested by Vidyamurthy (2004), or the price levels directly. The answer to this question depends on personal beliefs about the true data generating process. For example, using equation (2.3.6) with price levels directly, where the prices at time $t+1$ are given by

$$\begin{aligned} P_{t+1}^A &= P_t^A \left(1 + \tilde{r}_{t+1}^A\right) \\ P_{t+1}^B &= P_t^B \left(1 + \tilde{r}_{t+1}^B\right) \end{aligned}$$

with $\tilde{r}^i$, $i \in \{A, B\}$, denoting the simple return, the spread value at time $t+1$ is then

$$\begin{aligned} m_{t+1} &= P_t^A \left(1 + \tilde{r}_{t+1}^A\right) - \beta P_t^B \left(1 + \tilde{r}_{t+1}^B\right) \\ &= P_t^A - \beta P_t^B + P_t^A \tilde{r}_{t+1}^A - \beta P_t^B \tilde{r}_{t+1}^B \\ &= m_t + P_t^A \tilde{r}_{t+1}^A - \beta P_t^B \tilde{r}_{t+1}^B \end{aligned}$$

from which we can see that

$$m_{t+1} = m_t \quad \text{if and only if} \quad P_t^A \tilde{r}_{t+1}^A = \beta P_t^B \tilde{r}_{t+1}^B \ .$$

Now, it is important to see what happens to the spread value if both assets generate the same return in the observed period. Continuing with the expression above, we can write

$$\begin{aligned} m_{t+1} &= m_t + P_t^A \tilde{r}_{t+1} - \beta P_t^B \tilde{r}_{t+1} \\ &= m_t \left(P_t^A - \beta P_t^B\right) \tilde{r}_{t+1} \\ &= m_t + m_t \tilde{r}_{t+1} \\ &= m_t \left(1 + \tilde{r}_{t+1}\right) \end{aligned}$$

which shows that the spread value will not be constant in this case. With respect to the mean of the spread series this means that the mean value widens as the two asset prices go up and narrows when they go down, at least as long as the mean value is not zero. Using model (2.3.6) with untransformed price levels, the price of asset A at time t can be expressed as

$$P_t^A = \alpha + \beta P_t^B + \varepsilon_t$$

and so we see that with

$$\begin{aligned} 0 &= P_t^A \tilde{r}_{t+1}^A - \beta P_t^B \tilde{r}_{t+1}^B \\ &= \left(\alpha + \beta P_t^B + \varepsilon_t\right) \tilde{r}_{t+1}^A - \beta P_t^B \tilde{r}_{t+1}^B \\ &\Leftrightarrow \\ \tilde{r}_{t+1}^B &= \frac{\alpha + \beta P_t^B + \varepsilon_t}{\beta P_t^B} \tilde{r}_{t+1}^A \\ &= \left(1 + \frac{\alpha + \varepsilon_t}{\beta P_t^B}\right) \tilde{r}_{t+1}^A \end{aligned} \tag{2.3.8}$$

we obtain a relationship between the asset returns that must hold if we require $m_{t+1} = m_t$. If we assume for the moment that we are in the spread equilibrium at time t, i.e. $\varepsilon_t = 0$, we see from equation 2.3.8 that the required return $\tilde{r}^A_{t+1}$ in terms of $\tilde{r}^B_{t+1}$ depends still on the price level P^B_t. However, if the spread equilibrium is zero, i.e. $\alpha = 0$, then there is no such problem.

In case of $\alpha \neq 0$, a non-constant spread equilibrium can, however, also be avoided by taking the logarithm of the price levels. In this case, the returns on the two assets can differ by a multiplicative scalar value and the spread equilibrium remains constant. For example, using equation (2.3.6) and prices at time $t+1$, which we express this time as

$$P^A_{t+1} = P^A_t e^{r^A_{t+1}}$$
$$P^B_{t+1} = P^B_t e^{r^B_{t+1}}$$

with r^i, $i \in \{A, B\}$, denoting the continuously compounded return[6], the spread at time $t+1$ is given by

$$\begin{aligned} m_{t+1} &= \log\left(P^A_t e^{r^A_{t+1}}\right) - \beta \log\left(P^B_t e^{r^B_{t+1}}\right) \\ &= \log\left(P^A_t\right) + r^A_{t+1} - \beta \log\left(P^B_t\right) - \beta r^B_{t+1} \\ &= m_t + r^A_{t+1} - \beta r^B_{t+1} \end{aligned} \tag{2.3.9}$$

which means that

$$m_{t+1} = m_t \quad \text{if and only if} \quad r^A_{t+1} = \beta r^B_{t+1} \ .$$

This is independent of any price level, no matter what value α takes on. So, whether we should log-transform the prices or not depends on the chosen model and, thus, on our beliefs regarding the price relations.

[6] Continuously compounded returns are calculated as: $r = \log \frac{P_{t+1}}{P_t}$. See also Section 3.3.

Chapter 3

Application on Simulated Data

Before applying a statistical arbitrage model to real data, it may be worth analysing its behaviour on simulated data. By this we can get a better understanding of how well the model works under different circumstances. It is clear that there is a huge variety of circumstances under which we could analyse our model and there is no way to come up with an all-encompassing analysis. So which environment we generate certainly depends on the asset class and the market on which we want to apply the model. Where one should put the emphasis on is finally a question of beliefs with respect to the underlying data generating process.

3.1 The Price Process of a Common Stock

Established asset pricing theories usually consider the price movements of a traded financial asset as a stochastic process. According to Hull (2009) it makes sense to assume that the required return of a stock, or any other asset with a similar price process, is independent of its current price level. That is, the expected return of such an asset is assumed to be constant. For the return r of a particular stock with price S_t at time t we can write

$$\begin{aligned} r_{t+\Delta t} &= \frac{S_{t+\Delta t} - S_t}{S_t} \\ &= \frac{\Delta S_{t+\Delta t}}{S_t} \\ &= \mu \Delta t \end{aligned} \tag{3.1.1}$$

with $\mu > 0$. This can be rearranged to

$$\frac{\Delta S_{t+\Delta t}}{\Delta t} = \mu S_t \ . \tag{3.1.2}$$

If we have a deterministic function of the stock price with respect to time such that $S(t) = S_t$ and let $\Delta t \longrightarrow 0$, equation (3.1.2) becomes

$$\frac{dS(t)}{dt} = \mu S(t) \tag{3.1.3}$$

which is an ordinary (separable) differential equation with solution

$$\begin{aligned} S(t) &= S_0 + \int_0^t \mu S(u) du \\ &= S_0 e^{\mu t} \end{aligned} \tag{3.1.4}$$

where S_0 denotes the value of the function at time $t = 0$.

So, from a deterministic return process we can obtain a smooth function for the corresponding price process. In reality, however, we know that the price of a stock is not a smooth deterministic function of time. By adding an additive noise term, which depends on the time interval Δt, to the return process (3.1.1) we obtain

$$\frac{\Delta S_{t+\Delta t}}{S_t} = \mu \Delta t + \epsilon_t \ . \tag{3.1.5}$$

The "classical" asset path generating model in finance assumes ϵ_t to be normally distributed with $E[\epsilon_t] = 0$ and $Var(\epsilon_t) = \sigma^2 \Delta t$. In this case we can write

$$\epsilon_t \overset{\mathrm{d}}{=} \sigma \sqrt{\Delta t}\ \varepsilon_t$$

with $0 < \sigma < \infty$ and

$$\varepsilon_t \sim N(0, 1) \ .$$

The continuous time model, i.e. the model with $\Delta t \longrightarrow 0$, can then be expressed as

$$dS(t, \omega) = \mu S(t, \omega) dt + \sigma S(t, \omega) dW(t, \omega) \tag{3.1.6}$$

where we use the substitution

$$dW(t, \omega) = \varepsilon(t, \omega) \sqrt{dt}$$

which is also known as *Wiener-process*. The ω indicates that $S(t, \omega)$ is a random process defined on the appropriate filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t \geq 0}, P\right)$. For brevity we will, however, omit the argument ω in the further notation, i.e. we will use $S(t, \omega) = S(t)$.

The dynamics described by the stochastic differential equation (3.1.6) is known as *Geometric Brownian Motion* and is a special case of a so-called Itô-process.[1] One of its advantages is that it can be solved in closed-form by applying the corresponding Itô-calculus with the ansatz $f : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$, $(X_t, t) \mapsto \log(X_t)$. This is

$$\begin{aligned} S(t) &= S_0 + \int_0^t \mu S(u) du + \int_0^t \sigma S(u) dW(u) \\ &= S_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t)} \end{aligned} \tag{3.1.7}$$

with $S_0 = S(0) > 0$ and $W(t) \sim N(0, t)$. A nice property of this process is that it can only take on positive values, as one expects for asset prices without liabilities. The conditional expectation, given S_0, can easily be calculated from (3.1.7) as

$$\begin{aligned} E\left[S(t) | S_0\right] &= S_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right) t} E\left[e^{\sigma W(t)}\right] \\ &= S_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right) t} e^{\frac{1}{2} t \sigma^2} \\ &= S_0 e^{\mu t} \end{aligned} \tag{3.1.8}$$

[1] A sound introduction to stochastic differential equations and Itô-processes is given by Øksendal (2007).

which is the same as expression (3.1.4). The conditional variance is obtained by

$$Var\left(S\left(t\right)|S_0\right) = E\left[S^2\left(t\right)|S_0\right] - E\left[S\left(t\right)|S_0\right]^2 \tag{3.1.9}$$

with

$$\begin{aligned} E\left[S^2\left(t\right)|S_0\right] &= S_0^2 e^{2\mu t - \sigma^2 t} E\left[e^{2\sigma W(t)}\right] \\ &= S_0^2 e^{2\mu t - \sigma^2 t} e^{2\sigma^2 t} \\ &= S_0^2 e^{2\mu t + \sigma^2 t} \ . \end{aligned}$$

Plugging this together with (3.1.8) in equation (3.1.9) finally leads to

$$Var\left(S\left(t\right)|S_0\right) = S_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right) \ . \tag{3.1.10}$$

From the expressions (3.1.8) and (3.1.10) it becomes clear that this process is non-stationary because

$$\begin{aligned} \lim_{t\to\infty} E\left[S\left(t\right)|S_0\right] &= \infty \\ \lim_{t\to\infty} Var\left[S\left(t\right)|S_0\right] &= \infty \ . \end{aligned}$$

Figure 3.1 illustrates an example path with $S_0 = 30$, $\mu = 0.03$ and $\sigma = 0.2$.

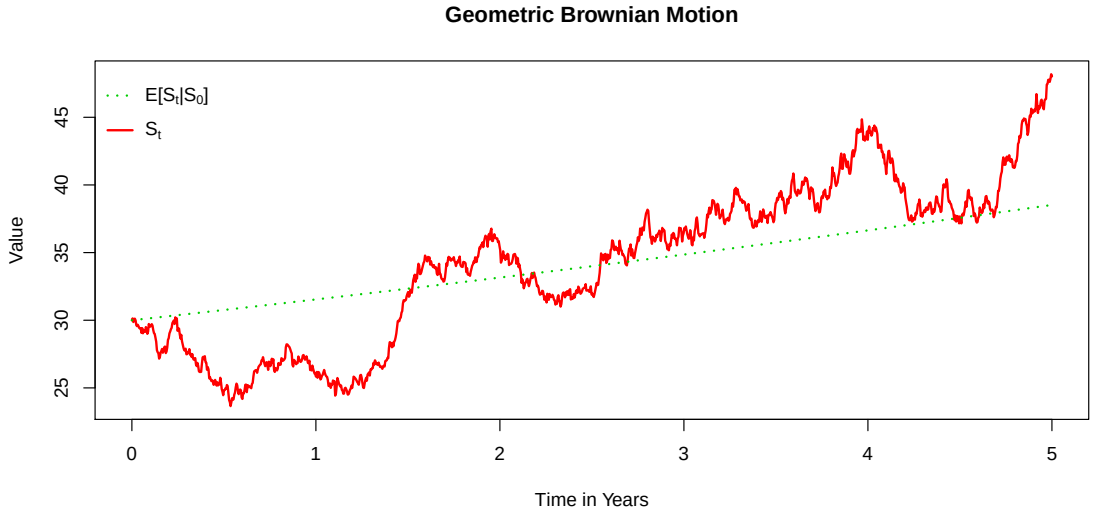

Figure 3.1: A typical sample path of a Geometric Brownian Motion.

The depicted trajectory really looks like the price path of a non-dividend-paying stock. Nevertheless, one has to be aware of the fact that the generated (daily) returns are in this case normally distributed which is actually not quite what we observe in the real stock market. There, daily returns usually show quite some fatter tails. One can, of course, adjust the disturbance distribution accordingly, however, only at the expanse of losing many of the nice analytical properties of the Itô-process.

3.2 A Mean-Reverting Process

Obviously, not all financial instruments show features as described in section 3.1. For example the nominal short rates of the term structure of interest rates are typically bounded. Their expected values are assumed to be finite also in the limit. Actually, the whole statistical arbitrage idea is based on a mean-reverting process. Thus, it makes sense to consider mean-reverting processes in more detail as well.

In principle, we can consider equation (3.1.6) as a base and look for appropriate modifications. We saw in section 3.1 that the conditional expectation of the process was determined by the first part of (3.1.6). By modifying this first part somewhat, we can write

$$dX(t) = \kappa\left(\theta - X(t)\right)dt + \sigma X(t)dW(t) \tag{3.2.1}$$

with $\kappa \geq 0$, $\theta \in \mathbb{R}$ and $dW(t)$ denoting again a Wiener-process. From this we can already see the mean-reversion behaviour as the process is dragged down if the value of the process lies above the parameter θ, i.e. the drift term becomes negative, and pushed up if $X_t < \theta$. However, instead of using $\sigma X(t)dW(t)$ as diffusion component, it is for many financial applications more realistic to use $\sigma\sqrt{X\,(t)}dW(t)$ instead, as in this case the hole process seems to correspond better to what is observed in the market. The solution of such a stochastic differential equation, i.e.

$$dX(t) = \kappa\left(\theta - X(t)\right)dt + \sigma\sqrt{X\,(t)}dW(t) \tag{3.2.2}$$

is called *square-root diffusion process*. Cox, Ingersoll, and Ross (1985) propose exactly this kind of stochastic process to model short term interest rates. It also plays an important role in stochastic volatility models such as the *Heston model* proposed by Heston (1993). Its main disadvantage is that expression (3.2.2) cannot be solved in closed-form.

To calculate an explicit expression for the conditional expectation of X_t we first take the integral form of the stochastic differential equation (SDE)

$$X(t) = \int_0^t \kappa\left(\theta - X(u)\right)du + \int_0^t \sigma\sqrt{X\,(u)}dW(u)\ . \tag{3.2.3}$$

Taking (conditional) expectations on both sides then leads to

$$E\left[X(t) \mid X_0\right] = E\left[\int_0^t \kappa\left(\theta - X(u)\right)du \mid X_0\right] + E\left[\int_0^t \sigma\sqrt{X\,(u)}dW(u) \mid X_0\right] \tag{3.2.4}$$

where the second term on the right hand side is the (conditional) expectation of an Itô-integral.

Theorem 3.2.1. *Let $dW(t)$ be a Winer-process. For any function $g : \mathbb{R}_+ \times \Omega \to \mathbb{R}$, $(t,\omega) \mapsto g(t,\omega)$ it holds that $E\left[\int_0^T g(t,\omega)dW(t,\omega) \mid g(0,\omega) = g_0\right] = 0$*

Proof.

$$\begin{aligned} E\left[\int_0^T g(t,\omega)dW(t,\omega) \mid g_0\right] &= E\left[\lim_{N\to\infty}\sum_{i=1}^{N} g(t_{i-1},\omega)\left(W(t_i,\omega) - W(t_{i-1},\omega)\right) \mid g_0\right] \\ &= \lim_{N\to\infty}\sum_{i=1}^{N} E\left[g(t_{i-1},\omega) \mid g_0\right] E\left[W(t_i,\omega) - W(t_{i-1},\omega)\right] \\ &= 0 \end{aligned}$$

□

According to Theorem 3.2.1 the (conditional) expectation of the Itô-integral is zero. In addition, by interchanging the integration and the expectation operation we obtain

$$\frac{dE\left[X(t) \mid X_0\right]}{dt} = \kappa\left(\theta - E\left[X(t) \mid X_0\right]\right) \tag{3.2.5}$$

which is a linear ordinary differential equation in the conditional expectation that can easily be solved by using the superposition principle. It follows that

$$E\left[X(t) \mid X_0\right] = \theta - (\theta - X_0)\, e^{-\kappa t} \tag{3.2.6}$$

with

$$\lim_{t\to\infty} E\left[X(t) \mid X_0\right] = \theta \tag{3.2.7}$$

for all $X_0 > 0$. The unconditional expectation is correspondingly

$$E\left[X(t)\right] = E\left[E\left[X(t) \mid X_0\right]\right] = \theta\ . \tag{3.2.8}$$

From these expressions we clearly see that a process as described by equation (3.2.2) is mean-reverting. The parameter θ acts thereby as the mean-reversion level of the whole process. The κ controls the speed of the mean-reversion behaviour. The parameter σ is still a measure of disturbance. With $\kappa > 0$ and any positive starting value X_0, both processes, as described by (3.2.1) and (3.2.2), cannot become negative as the diffusion term gets zero if X_t goes towards zero. So, while this process may be useful to model particular processes with bounded positive values, it is not appropriate for processes that fluctuate around zero. But here again, the process we just discussed, can easily be adjusted to suit to this situation too. One popular solution is given by

$$dX(t) = \kappa\left(\theta - X(t)\right)dt + \sigma dW(t) \tag{3.2.9}$$

where we just let the diffusion term be independent of the current function value of $X(t)$. The solution of the dynamics described by equation (3.2.9) (often with $\theta = 0$) is called *Ornstein-Uhlenbeck* process. The conditional and unconditional expectation of this process is still the same as for the square-root diffusion process. By using Itô's lemma together with the ansatz $f : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}\ ,\ (X_t, t) \mapsto X_t e^{\theta t}$ one obtains

$$\begin{aligned} df(X_t, t) &= \kappa X_t e^{\kappa t} dt + e^{\kappa t} dX_t \\ &= e^{\kappa t}\kappa\theta dt + e^{\kappa t}\sigma dW_t\ . \end{aligned}$$

Thus, we can write

$$\begin{aligned} X(t) &= X_0 e^{-\kappa t} + e^{-\kappa t}\int_0^t e^{\kappa u}\kappa\theta du + \int_0^t e^{\kappa(u-t)}\sigma dW(u) \\ &= X_0 e^{-\kappa t} + \theta\left(1 - e^{-\kappa t}\right) + \int_0^t e^{\kappa(u-t)}\sigma dW(u)\ . \end{aligned} \tag{3.2.10}$$

From this expression we can also see very well that the (conditional) expectation of the Ornstein-Uhlenbeck-process is indeed as stated in (3.2.6). The (conditional) variance of

X_t is also easy to derive, given (3.2.10). This is

$$\begin{aligned} Var(X_t \mid X_0) &= Var\left(X_0 e^{-\kappa t} + \theta\left(1 - e^{-\kappa t}\right) + \int_0^t e^{\kappa(u-t)}\sigma dW(u) \mid X_0\right) \\ &= Var\left(\int_0^t e^{\kappa(u-t)}\sigma dW(u)\right) \\ &= E\left[\left(\int_0^t e^{\kappa(u-t)}\sigma dW(u)\right)^2\right] \\ &= E\left[\int_0^t e^{2\kappa(u-t)}\sigma^2 du\right] \\ &= \sigma^2 e^{-2\kappa t}\int_0^t e^{2\kappa u} du \\ &= \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa t}\right) \end{aligned} \tag{3.2.11}$$

where we use the Itô-isometry[2] to get from line 3 to line 4. Thus, for $\kappa > 0$ the Ornstein-Uhlenbeck-process has finite variance. With t getting large it converges towards a constant value, i.e.

$$\lim_{t\to\infty} Var\left(X(t) \mid X_0\right) = \frac{\sigma^2}{2\kappa}\ . \tag{3.2.12}$$

Given $X_0 = x_0$, the distribution of X_t for any $t > 0$ is given by

$$X_t \sim N\left(\theta - (\theta - x_0)\, e^{-\kappa t}, \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa t}\right)\right)\ . \tag{3.2.13}$$

Figure 3.2 illustrates an example path with $X_0 = 0.7$, $\theta = 0$, $\kappa = 15$ and $\sigma = 2$.

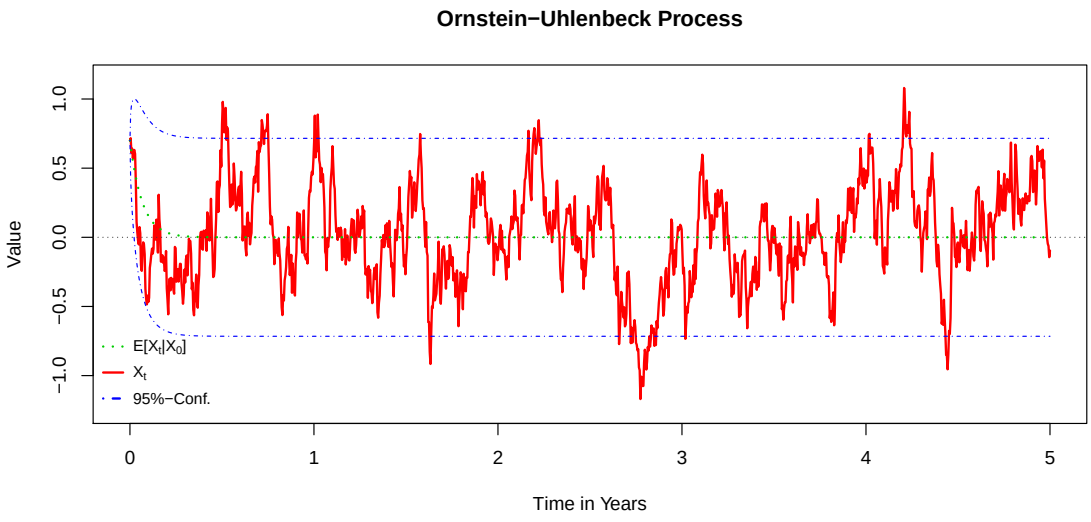

Figure 3.2: A typical sample path of an Ornstein-Uhlenbeck-process.

The possibly simplest way in order to obtain a discretised version of the Ornstein-Uhlenbeck process is to employ the so-called *Euler* scheme. In particular, the discrete process

[2] The proof can be found in Øksendal (2007) on page 26.

$\left\{\hat{X}_h, \hat{X}_{2h}, \ldots, \hat{X}_{mh}\right\}$, $h \in (0, +\infty)$, consisting of m time steps and generated by

$$\hat{X}_{th} = \kappa\theta h + (1 - \kappa h)\, \hat{X}_{(t-1)h}\sigma\sqrt{h}\zeta_t \tag{3.2.14}$$

with ζ_t being independent and standard normally distributed, follows an Ornstein-Uhlenbeck process for $h \to 0$. That is, The smaller the value of h, the closer our discretised path will be to the continuous-time path according to (3.2.10). For $h = 1$ we obtain

$$\hat{X}_t = \kappa\theta + (1 - \kappa)\, \hat{X}_{t-1}\sigma\zeta_t \tag{3.2.15}$$

which is a simple AR(1) process. Of course, with $h = 1$ we must also restrict the parameter κ, i.e. $0 < \kappa < 1$. In this case the AR(1) process is stationary due to $|1 - \kappa| < 1$, i.e. the root of the characteristic polynomial $z^2 - (1 - \kappa)z - \kappa\theta = 0$ lies inside the unit circle.

3.3 Simulation Results

In order to carry out a realistic simulation one needs to have an appropriate belief about the true data generating process. As already mentioned in chapter 2 the Asset Pricing Theory (APT) as proposed by Ross (1976) provides a well established theory about the development of stock prices. However, without further restrictions the APT-framework does not imply that two time series are cointegrated. Even though we can relax the necessary requirements on a linear combination of two asset price series in the context of a statistical arbitrage trading strategy in the sense that it does not necessarily have to be weekly stationary, the APT-approach does not even guarantee a mean-reverting behaviour of the spread series. Figure 3.3 illustrates this by means of a simulated example. As can be seen, the returns of the two series are chosen, on purpose, to be very strongly correlated. Nevertheless, from the upper graph it should become clear that the spread series can considerably diverge from a particular constant value. So, any mean-reversion based trading strategy on such a series would be extremely dangerous.

Thus, for two time series with a very close return-relationship according to the APT there is not necessarily a linear combination of them that is weakly stationary. This is, the APT does not imply cointegration. In the other direction, however, there is a special case where cointegration does imply the APT. In order to show this we have to employ continuously compounded returns and log-transformed prices.[3] This is, the simple one-period return $\tilde{r}$ on an asset with price S is defined as

$$S_1 = (1 + \tilde{r})\, S_0 \ .$$

If we split this period in n sub-periods and choose the simple return to be divided linearly, this results in

$$S_1^{(n)} = \left(1 + \frac{\tilde{r}}{n}\right)^n S_0$$

with the consequence that $S_1 < S_1^{(n)}$ due to the compounding effect. When n goes to infinity, we obtain

$$\lim_{n\to\infty} \left(1 + \frac{\tilde{r}}{n}\right)^n = e^{\tilde{r}} \ .$$

[3] Remember the short discussion about price level transformations in Section 2.3.

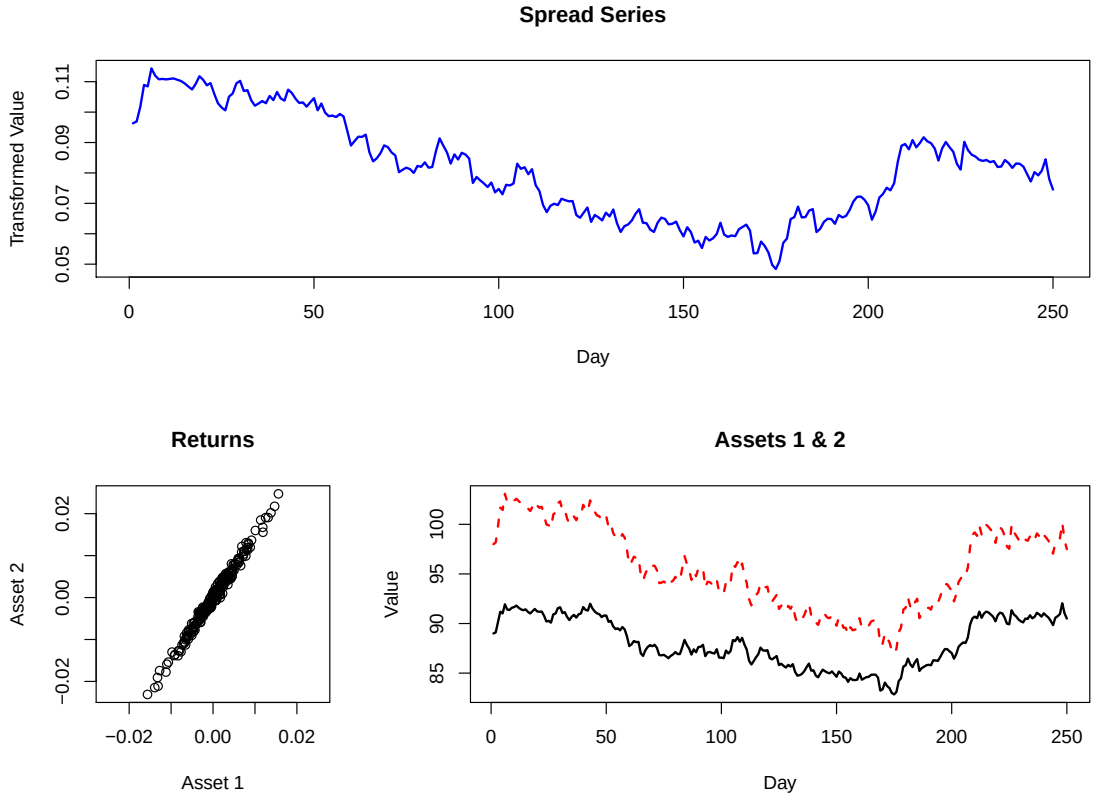

Figure 3.3: An example of an APT-generated pair.

The return r that implies $S_1 = S_1^{(n)}$ is called continuously compounded return. We can write

$$S_1 = S_0 e^r \Leftrightarrow r = \log\left(\frac{S_1}{S_0}\right) \ .$$

Obviously, the continuously compounded return is strictly smaller than the simple return, i.e. $r < \tilde{r}$. However, for small return values the difference between the two is small. The big advantage of the continuously compounded return is its additivity property, which is basically the key to make the relation from the cointegration approach in the context of statistical arbitrage to the APT. For a time span of m periods the price S of an asset j can be described as

$$\log\left(S_{m,j}\right) = \log\left(S_{0,j}\right) + \sum_{i=1}^{m} r_{i,j} \ . \tag{3.3.1}$$

If the APT holds true, the cumulative returns of two assets A and B are given as

$$\begin{aligned} \log\left(\frac{S_{m,A}}{S_{0,A}}\right) = \sum_{i=1}^{m} r_{i,A} &= \sum_{i=1}^{m}\left(r_f + \sum_{h=1}^{k} \beta_{h,A} r_{i,h} + \epsilon_{i,A}\right) \\ &= \sum_{i=1}^{m} r_f + \sum_{i=1}^{m}\sum_{h=1}^{k} \beta_{h,A} r_{i,h} + \sum_{i=1}^{m} \epsilon_{i,A} \end{aligned} \tag{3.3.2}$$

$$\log\left(\frac{S_{m,B}}{S_{0,B}}\right) = \sum_{i=1}^{m} r_{i,B} = \sum_{i=1}^{m}\left(r_f + \sum_{h=1}^{k}\beta_{h,B}r_{i,h} + \epsilon_{i,B}\right)$$
$$= \sum_{i=1}^{m} r_f + \sum_{i=1}^{m}\sum_{h=1}^{k}\beta_{h,B}r_{i,h} + \sum_{i=1}^{m}\epsilon_{i,B} \ . \tag{3.3.3}$$

If the factor exposures of the two assets are identical, i.e. $\beta_{h,A} = \beta_{h,B}$ for all $h \in \{1, 2, \ldots, k\}$, we obtain by subtracting equation 3.3.3 from equation 3.3.2

$$\log\left(\frac{S_{m,A}}{S_{0,A}}\right) - \log\left(\frac{S_{m,B}}{S_{0,B}}\right) = \sum_{i=1}^{m}\left(\epsilon_{i,A} - \epsilon_{i,B}\right) \ . \tag{3.3.4}$$

Using the definition of continuously compounded returns, this can be rewritten as

$$r_{m,A} = r_{m,B} + \tilde{\epsilon}_m \tag{3.3.5}$$

where the new error term $\tilde{\epsilon}_m$ depends on m. As this relation must hold for any m, the APT implies that for all $t > 0$ we have for the one period returns

$$r_{t+1,A} = r_{t+1,B} + \tilde{\epsilon}_{t+1} \ . \tag{3.3.6}$$

When we introduced the cointegration approach in section 2.3, we formulated with equation (2.3.9) kind of an equilibrium relation with respect to the return relationship. In general, it holds that for any two log-transformed time series, $\log(S_{t,A})$ and $\log(S_{t,B})$, that are cointegrated of order $CI(1,1)$ with relationship

$$\log(S_{t,A}) = \alpha + \beta \log(S_{t,B}) + \varepsilon_t$$

we obtain

$$\log(S_{t+1,A}) - \log(S_{t,A}) = \beta\left(\log(S_{t+1,B}) - \log(S_{t,B})\right) + \tilde{\varepsilon}_{t+1}$$
$$r_{t+1,A} = \beta r_{t+1,B} + \tilde{\varepsilon}_{t+1} \tag{3.3.7}$$

where we just use the definition of continuously compounded returns in the second equation. The return relationship the APT implies is, thus, just a special case of the return relationship in the cointegration framework. We can, therefore, continue as follows for the simulation: The price path of one asset of a particular pair can be obtained according to the APT. Having generated the price path of the asset we can then log-transform the series and generate the transformed series of the pair-asset by adding a weakly stationary spread series to it. The price paths of the two assets are then obtained by simply applying the exponential transformation on the log-transformed series. If we want to have different starting values for the two series, we can accomplish this by adding a constant value to the paired series before we apply the transformation.

As Gatev et al. (2006) use a one year asset fitting period based on daily observations, we can mimic this by generating 250 data points per cycle. For every cycle we generate 10 assets with a corresponding pair-asset for each of them. Furthermore, we generate 80 additional "noise"-assets with no explicit pair-asset. So, we end up with 100 asset price series in total for one cycle. As just mentioned above, one asset of each pair is generated by means of the APT. In order to do that we generate two non-stationary factor price series according to equation (3.1.7). One with a drift term and one without. For example, one of the factors could mimic the world oil price and the other an important currency

exchange rate. Of course, the choice of the factors is crucial in real applications. Here we limit ourselves to just two risk factors, as this seems appropriate in order to get a first feeling about the mode of action of the whole mechanism. The risk-free rate is generated according to a square-root diffusion process as described by equation (3.2.2). Figure 3.4 shows the realisation of the risk-free rate (denoted on a yearly basis) as well as the price series of the two risk factors for one cycle (i.e. one simulated year).

The theoretical values of the APT-coefficients used in the simulation are reported in table 3.1 (first three columns). As can be seen there, for the first ten pairs the signal stemming from the risk factors is chosen to be higher than for the last ten pairs. That is, β_1 and β_2 add up to 0.75 for the first ten pairs and to only 0.5 for the second ten pairs, while the variance of the error term is the same for all pairs. The columns 3 to 6 report the mean of the estimated APT-coefficients from a simulation of 1000 cycles. For every cycle we use an ordinary AR(1) process as the spread series, i.e.

$$X_t = \alpha X_{t-1} + \varepsilon_t \tag{3.3.8}$$

with $\alpha = 0.9$ and $Var(\varepsilon_t) = 0.0025\%$. As is reported in table 3.1, the average values of the estimated APT-coefficients are quite close to their theoretical values. The result also illustrates the theoretical considerations from above very nicely as the pair-assets that are generated by the summation of the spread series and the *APT-generated* pair-asset series, too, show on average estimated values of the APT-coefficients that hardly deviate form the values of their corresponding *APT-generated* pair-asset.

Figure 3.6 shows the price paths of the 20 pair-assets for one cycle. The paired assets of the first three pairs start more or less at the same price level. For the other pairs we chose different starting values, sometimes considerably different ones. For example, for the two assets 7 and 8 the starting value of asset 7 is approximately twice the value of asset 8. In this case the cointegration relation contains a constant. This is important to keep in mind when testing for cointegration according to Phillips and Ouliaris (1990).

Figure 3.7 depicts the scatterplots of the continuously compounded returns of the paired assets. The red dashed line is just the diagonal, i.e. a line with slope equal to one. As we can see, the line corresponds in most cases to the regression line we would expect to obtain according to the depicted data points. With the slope of the line being one, this also means that the variances of the two returns are approximately the same. Exceptions are the plots of the assets 13 and 14 as well as the one of the assets 19 and 20. There it seems that the asset returns, while still positively correlated, differ in variance. The variances of the assets 13 and 19 seem to be larger in relative terms compared to the variance of their corresponding pair-asset, i.e. the assets 14 and 20. This may be explained by the special way of constructing our paired series. When adding the values of one time series to the values of another time series, as we do in order to generate the pair-series, the resulting time series does not need to have the same variance as either of the series it is composed of. This is easy to see. If we denote the transformed prices of two assets at time t as $\log(S_{t,i}) = Y_{t,i}$ with $i \in \{1, 2\}$, we can write, in conjunction with (3.3.8), for the variance of any such APT and cointegration compliant transformed series

$$\begin{aligned} \mathrm{Var}\,(Y_{2,t}) &= \mathrm{Var}\,(Y_{1,t} + X_t) \\ &= \mathrm{Var}\,(Y_{1,t}) + \mathrm{Var}\,(X_t) + 2\,\mathrm{Cov}\,(Y_{1,t}, X_t) \quad . \end{aligned}$$

From this equation we clearly see that if $Y_{1,t}$ and X_t are uncorrelated, then it holds that $\mathrm{Var}(Y_{2,t}) > \mathrm{Var}(Y_{1,t})$. However, with the chosen parameters for the spread series in (3.3.8)

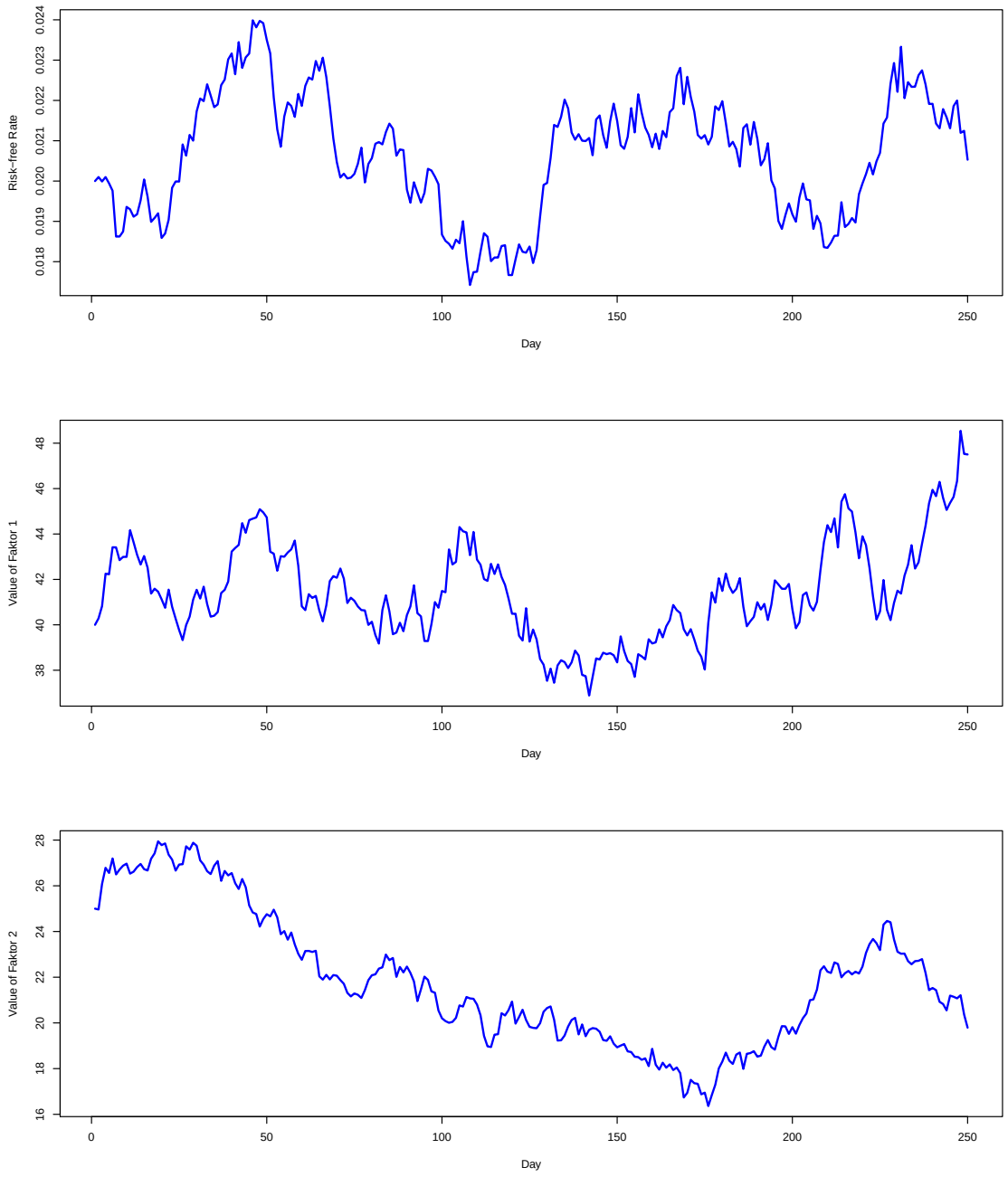

Figure 3.4: Three factor paths for one cycle.

we have $\mathrm{Var}(X_t) = \frac{0.0025\%}{1-0.81} = 0.013\%$, which means that the variance component stemming from the spread series is rather small. Thus, the relationship between the variances of the paired series depends quite a lot on the covariance between the spread series and the series that we generate by means of the APT. As we do not explicitly control for this covariance in the simulation study, it may be the case that for any two paired assets one has a larger variance than the other. However, this should not be crucial regarding our analysis. We just aim for a realistic scenario and this may occur in the real market as well.

	β_{r_f}	β_{r_1}	β_{r_2}	$\hat{\beta}^{MC}_{r_f}$	$\hat{\beta}^{MC}_{r_1}$	$\hat{\beta}^{MC}_{r_2}$
Asset 1	1.00	0.375	0.375	0.9991	0.3750	0.3750
Asset 2	1.00	0.375	0.375	1.0338	0.3742	0.3747
Asset 3	1.00	0.300	0.450	1.0026	0.2999	0.4500
Asset 4	1.00	0.300	0.450	1.0473	0.3006	0.4503
Asset 5	1.00	0.450	0.300	0.9974	0.4500	0.2999
Asset 6	1.00	0.450	0.300	0.9918	0.4513	0.2998
Asset 7	1.00	0.150	0.600	0.9933	0.1501	0.6000
Asset 8	1.00	0.150	0.600	0.9783	0.1500	0.5992
Asset 9	1.00	0.600	0.150	0.9867	0.6000	0.1501
Asset 10	1.00	0.600	0.150	0.9658	0.5998	0.1496
Asset 11	1.00	0.250	0.250	1.0023	0.2500	0.2500
Asset 12	1.00	0.250	0.250	1.0492	0.2501	0.2499
Asset 13	1.00	0.200	0.300	1.0177	0.2001	0.3000
Asset 14	1.00	0.200	0.300	1.0178	0.2005	0.2998
Asset 15	1.00	0.300	0.200	0.9848	0.3000	0.2000
Asset 16	1.00	0.300	0.200	0.9373	0.3013	0.2015
Asset 17	1.00	0.100	0.400	1.0008	0.1000	0.4000
Asset 18	1.00	0.100	0.400	0.9817	0.1004	0.3998
Asset 19	1.00	0.400	0.100	0.9984	0.4000	0.1000
Asset 20	1.00	0.400	0.100	0.9630	0.3998	0.1000

Table 3.1: Average APT-coefficients of 1000 simulations.

Given the existence of some paired series, the critical question is, of course, whether we are also able to identify them in a broader asset universe consisting of many assets. The idea is to apply the Phillips-Ouliaris test for cointegration, as introduced in section 2.3, on the corresponding log-transformed time series of each possible pair. The pairs are then detected by selecting the two assets with the highest multivariate trace statistic of the test. With an asset universe consisting of 100 assets, we have to compute the trace statistic for 4950 possible combinations in each cycle.

To test whether such a selection procedure is appropriate, we do a simulation of 1000 cycles according to the previously described setting. The results show that our pairing procedure identifies on average only 68.11% of the pairs correctly. When increasing the persistence in the spread time series, i.e. when the coefficient α in equation (3.3.8) is increased from 0.9 to 0.95, the average of correctly identified pairs even decrease to merely 24.90%. Figure 3.5 depicts the histograms of the detection rates with respect to the 1000 cycles of the spread series with $\alpha = 0.9$ and $\alpha = 0.95$, respectively.

As can be seen, the small increase in the coefficient of the AR(1) process of the spread series has quite a heavy impact. With $\alpha = 0.95$ in more than a quarter of the cycles no more than 10% of the pairs are correctly detected. In the best case only 70% of the pairs were found. With $\alpha = 0.9$ we could identify even all pairs in the best case. In the worst case no less than 20% were correctly found.

Also interesting to know is whether there are any differences in the pair detection with respect to the APT coefficient specifications or the starting values. Figure 3.8 shows the average rate of correctly identified pairs for each of the ten pairs. The dashed line marks

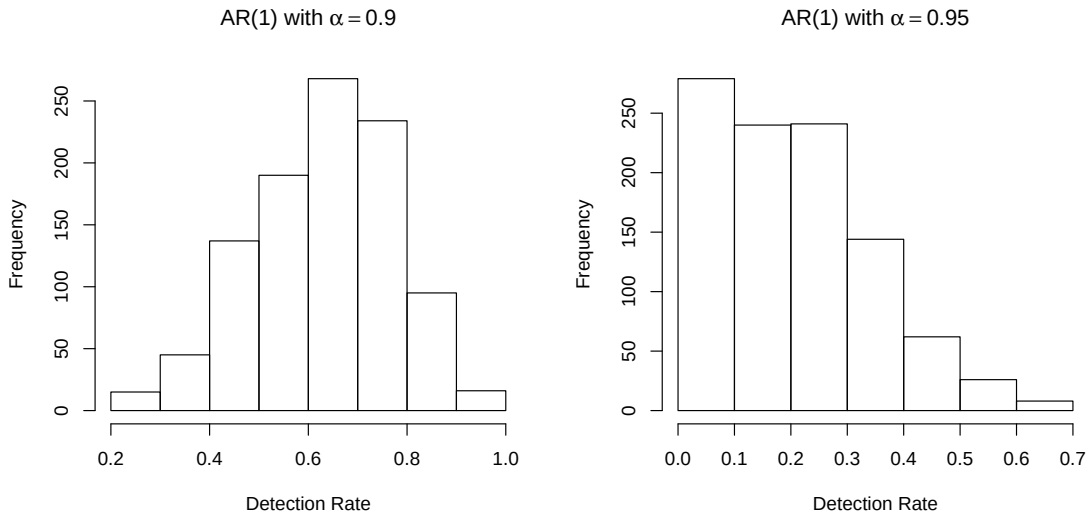

Figure 3.5: Frequency of detection rates in a simulation of 1000 cycles.

the average detection rate regarding all ten pairs. Interestingly, while the APT coefficient specification does not seem to have any impact on the pair detection, the different starting values obviously do. Pair number 4 which consists of the assets 7 and 8 seems to be difficult to be detected. Also the pairs 5 and 6 have rates below the average. This seems to correspond with the difference in starting values. This is, the starting value of asset 7 is about twice the value of asset 8. Also, asset 9 and 10 show quite a big difference in their starting values and a correspondingly low average pair detection rate. In general, it seems that the larger the difference in the starting values between the two pair-assets relative to their price levels, the worse the average rate of identified pairs becomes. It is clear that such a difference in the price level implies a constant in the cointegration vector that describes the cointegration relation. However, this was explicitly considered in our cointegration tests. Therefore, the result is indeed somewhat surprising.

Figure 3.9 shows a similar picture as figure 3.8 just on a lower level. Here as well, pair number 4 is the least identified and, in general, the pairs with a relatively large gap concerning the starting values of their assets are less often detected correctly.

Regarding the application of a statistical arbitrage strategy on real data, it may be good practice to normalise the price series in the sense that they start at the same value. In addition, we should be aware of the fact that the identification of potential pairs may be difficult when the spread series shows a high level of persistence.

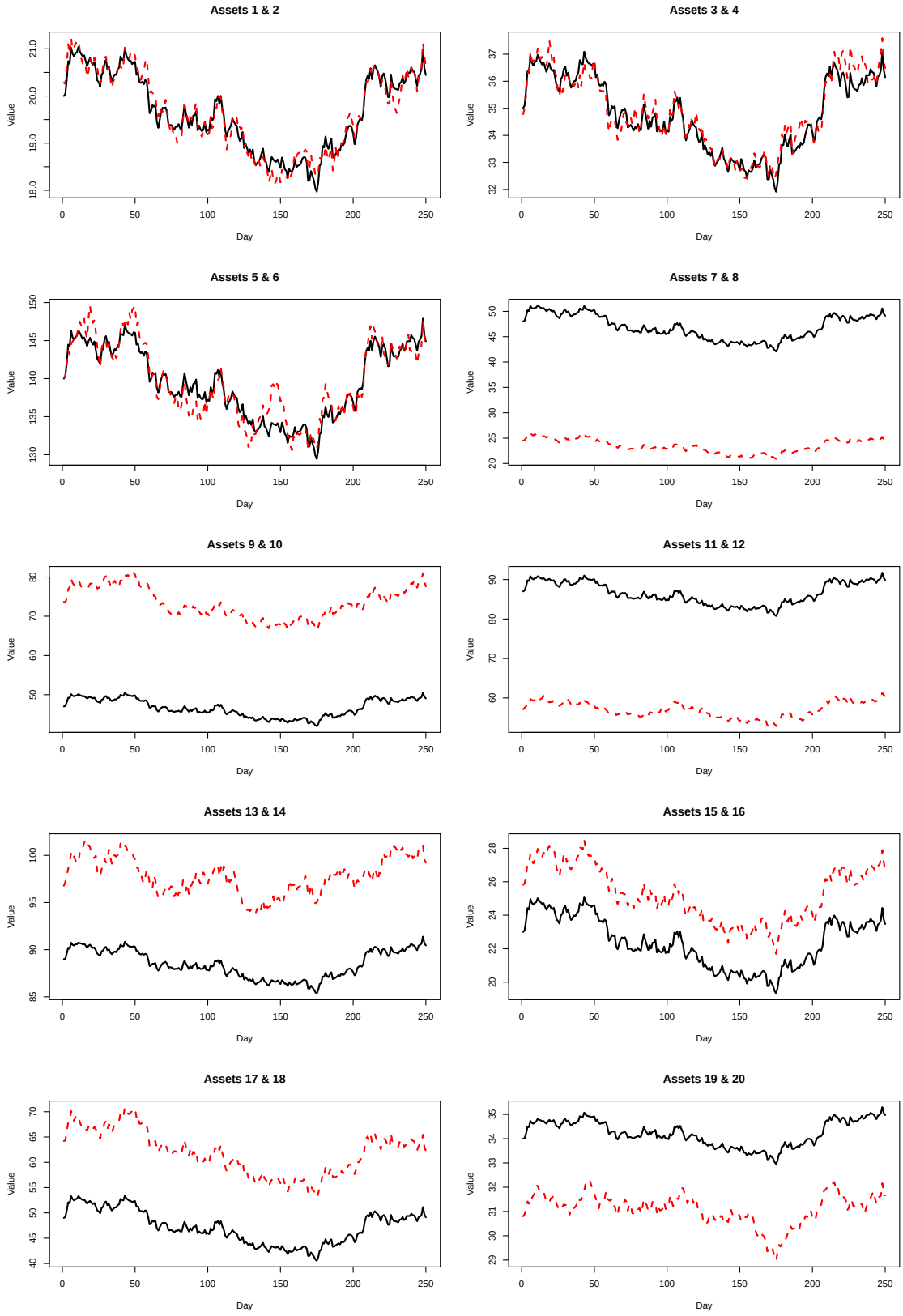

Figure 3.6: The paths of ten pairs for one cycle.

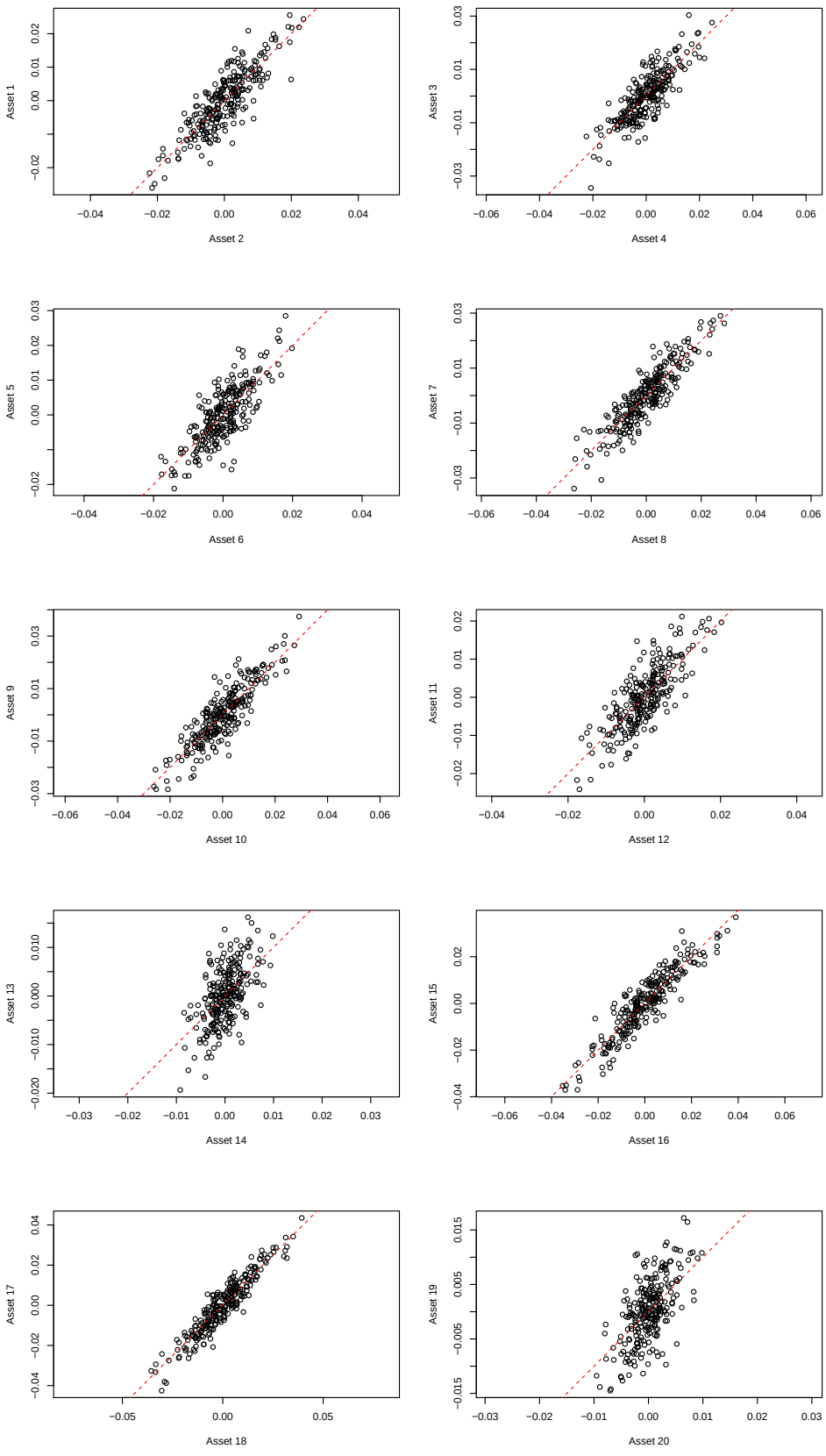

Figure 3.7: The scatterplots of the ten pairs.

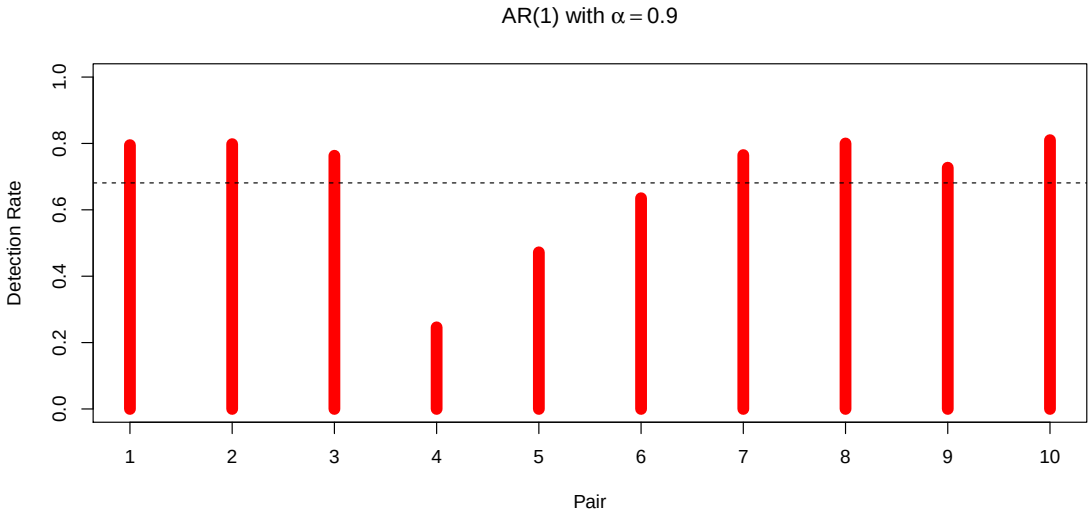

Figure 3.8: Detecting cointegrated pairs if the spread follows an AR(1) with $\alpha = 0.9$.

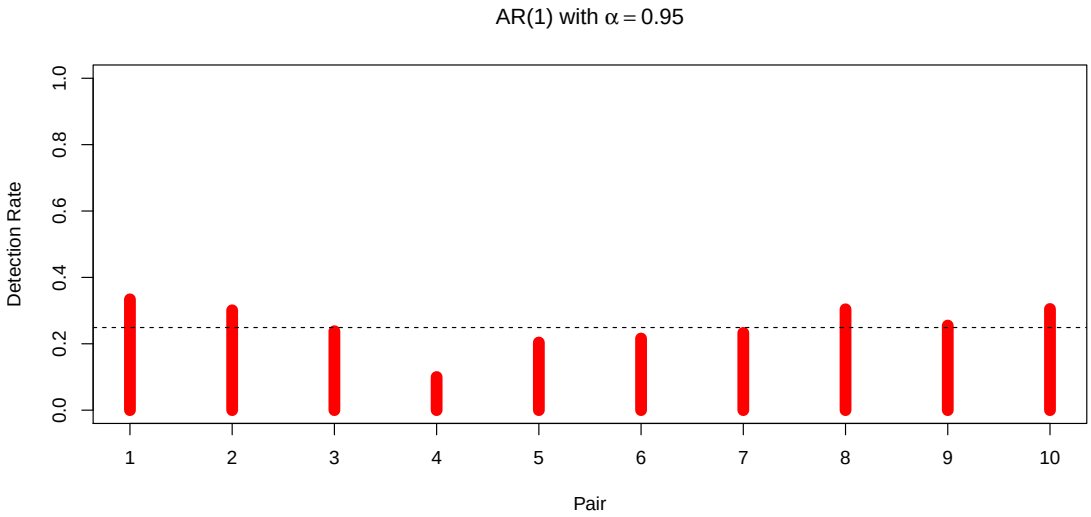

Figure 3.9: Detecting cointegrated pairs if the spread follows an AR(1) with $\alpha = 0.95$.

Chapter 4

Back-testing on the S&P 500

After the theoretical and simulation based analysis of section 3 we now want to see how such a statistical arbitrage model would have performed in the past if we had applied it to a real security market. Thus, in this chapter we do a back-testing on assets belonging to the S&P 500 index for the time period reaching from the beginning of 1996 up to the end of 2011.

4.1 The S&P 500

Why the *Standard and Poor's 500* (S&P 500) index? The S&P 500 is a free-float capitalisation weighted equity index that consists of the 500 most capitalised companies traded on either the *New York Stock Exchange* (NYSE) or the NASDAQ. Thus, if we limit our investment universe to those securities that are part of the S&P 500 index, we limit ourselves to very liquid large-cap common stocks held by a broad public. This is important for our strategy as we need to have liquid assets with only tiny bid-ask spreads and prices that are hardly ever influenced when buying or selling large chunks of an asset. With an investment universe of 500 assets we obtain in total 124750 different combinations of two assets, which is a sufficiently large number.

Another crucial issue in back-testing is the so-called *survivorship bias*. That is the tendency for failed companies to be excluded from performance studies because they no longer exist. Considering only those companies that have been successful enough to survive until the end of a particular period usually causes the obtained performance to be higher then it would have been in reality. Companies that go bankrupt during the analysed period can cause considerable losses in an investment strategy. This is also the case in the context of a statistical arbitrage strategy. For example, if we have identified a pair of two assets, A and B, in the previous period, which we then trade in the subsequent period, it may happen that company A faces some company specific problems which translate in a decreasing share price. According to our strategy, we would very likely buy the shares of company A and sell-short the shares of company B. If the situation of company A worsens so that it finally goes bankrupt, its share price would fall sharply. Holding a long position in asset A and a short position in asset B, this would mean that the value of our long-short portfolio also decreases, resulting in a potentially heavy loss. Therefore, every serious back-testing has to account for that fact.

However, every sincere back-testing is expensive. Very often it is already difficult to acquire the relevant information about the historical asset universe of a particular market. Finally, having the required information, the data collection and preparation process is usually time consuming and sometimes tricky.

In this study only stocks that belong to the S&P 500 index are considered. This means for every selection period we only choose pairs out of the stock universe that make up the components of the S&P 500 index at that particular point in time. The composition of the asset universe, therefore, changes over time. Some companies fall out of the index at some point in time and new companies enter instead of them. For the time span from 1995 up to 2011 we obtain from *Wharton Research Data Services* (WRDS) a list of companies that were part of the index for at least one month during the whole considered time span. In total we obtain by this definition a list that consists of more than 1000 companies.

4.2 Pairs Formation and Trading Rule

As in the simulation study of section 3.3, the matching of the pairs is done by allocating each asset to its pair-asset according to the highest cointegration test statistic. A particular asset is also allowed to be part of several pairs. For the trading we then take the n best pairs with respect to the test statistic. The only criteria regarding the value of the test statistic is its magnitude. Whether it is significant in a Neyman-Pearson sense is not really informative and may be of little relevance in our case. Why this is the case becomes clear if we consider, for example, the case where actually non of our assets were cointegrated. In this case we would nevertheless expect to observe about 1247 significant test results on the 1%-significance level. Obviously, with 124750 possible pair-combinations the number of 1247 corresponds to the amount of cases where we mistakenly reject the null hypothesis of the test (i.e. H_0: *there is no cointegration relation*). Thus, relying on a significant test result would be misleading. Certainly, we could try to tackle this problem by applying statistical methods designed to control for multiple hypothesis testing, such as the concept of *false discovery rate*[1] (FDR) as discussed by Benjamini and Hochberg (1995). The advantage of the FDR is that it is equivalent to the popular *familywise error rate*[2] (FWER) when all hypotheses are true, but is smaller otherwise. While a correction according to the FWER would be too conservative in our case, employing the FDR procedure may indeed be beneficial. As the FDR procedure relies, however, on appropriately calculated p-values, it is a bit tricky to exert it in our case as the true distribution of our cointegration test statistic is unknown. The critical values in the cointegration test according to Phillips and Ouliaris (1990) are based on simulation results. We, therefore, abandon the FDR approach at this point.

The pairs formation period is chosen to be one year. This means, at a particular point in time we take the data of the period reaching back one year from this point, which usually consists of about 260 trading days, and use this data to match the pairs as described. This time horizon is chosen arbitrarily. The guiding idea behind this choice is that the formation period should be long enough in order to obtain a reasonable amount of data points for the estimation of the model parameters, but should neither reach to far into the past in order to avoid any structural breaks in the data. Having determined n pairs, we obtain

[1] The FDR controls the expected proportion of type I errors when performing multiple pairwise tests.

[2] The FWER controls the probability of making at least one type I error when performing multiple pairwise tests.

the equilibrium relation between the two stocks of one pair by means of linear regression. By this we can make some judgements about the relative values of two pair-assets. This actually means that if we have two stocks, A and B, that form a pair and we know the value $S^{(\cdot)}$ of one of the two stocks, we can say something about the expected value of the pair-stock. So, if we have, for example, the relationship $E\left[S^{(A)} \mid S^{(B)}\right] = \alpha + \beta S^{(B)}$ and observe at time t that $E\left[S_t^{(A)} \mid S_t^{(B)}\right] < S_t^{(A)}$, we would conclude that stock A is overvalued relatively to stock B. In this case we would short-sell stock A and purchase with the proceeds stock B instead. This obviously corresponds to the situation where the spread $m_t = S_t^{(A)} - \alpha - \beta S_t^{(B)}$ is positive. Similarly, if $m_t < 0$, it would imply that stock A is undervalued relatively to stock B, in which case we would short-sell stock B and purchase stock A. The assumption of a mean-reverting spread series m_t is, thus, at the core of any cointegration based statistical arbitrage model. The mean-reversion of the spread series just means that the relative miss-pricing of the two assets is corrected in the future. A spread value of α can be interpreted as an equilibrium relation around which the actual relative value of the two assets fluctuates. Thus, having the pairs, the trading rule is based on the spread value m_t. It is also important to stress that we only make relative statements. In general, we cannot say whether asset A or B is over- or undervalued in absolute terms. In addition, cases with a negative β in the regression of one asset on its pair-asset, cannot be exploited in our setting. This is easy to see if we consider, for instance, an observation where $E\left[S_t^{(A)} \mid S_t^{(B)}\right] < S_t^{(A)}$. If $\beta > 0$ in the pair-relationship, then we can conclude that either asset A is overvalued or asset B is undervalued, which leads in both cases to the same trading action, namely short-selling asset A and buying asset B. There is no need to decide whether asset A is overvalued or asset B is undervalued. By contrast, if $\beta < 0$ in the pair-relationship, then we would need to decide whether asset A or asset B is overvalued in order to exploit this situation by opening a pair-position on the correct side. As this is not a trivial issue, we will exclude cases where β is negative.

Given the correct relationship between the prices of two pair-assets, the spread value m_t is observable at time t as $S_t^{(A)}$ and $S_t^{(B)}$ are observable. Certainly, the relationship, if there is one at all, is not just given, but has to be estimated by statistical methods. It is also clear that any statistical modelling is subject to model and parameter risks. With this in mind we can now go one step further and discuss appropriate trading rules. Thereby, it makes sense to consider the whole relative value arbitrage investment as a portfolio of spreads. This allows us to find ways to optimise the overall investment and to control its risks. In the following, two rather extreme cases are stated. On the one hand, a simple static trading approach, which is explained in section 4.2.1. On the other hand, a much more complex, dynamic trading strategy, treated in some detail in section 4.2.2.

4.2.1 A Simple Trading Rule

First of all, a relative value arbitrage investment needs proper capital. It is wrong to believe that one could use the proceeds of the assets sold short to finance the long-position without the requirement of own capital. This is the case because short positions always have to be backed by some appropriate collateral and there must be enough capital available to ensure the so-called margin requirements. Hence, the proper capital that is needed in order to fulfil the margin requirements is usually assumed to earn returns equal to the risk-free rate as it is usually invested in the money market. This is generally important to keep in mind when designing a profitable trading rule, at least for the more sophisticated ones,

as, for example, the dynamic rule analysed in section 4.2.2. In the case of simple trading rules, this may be neglected. Here we just take the spread series m_t as mentioned above into consideration. For example, let's take the two companies *Verizon Communications Inc.* and *Bellsouth Corp.* These two companies are chosen because they get the highest cointegration test statistic for the first matching period[3], i.e. the period reaching from January 4th, 1995 up to December 31st, 1995. Regressing the log-prices of *Verizon* on the log-prices of *Bellsouth* leads to the estimated coefficients $\widehat{\alpha} = 1.45$ and $\widehat{\beta} = 0.64$ (where both values are highly significant). The corresponding spread series m_t is depicted in Figure 4.1. As can be seen there, the spread series in the fitting period looks pretty much

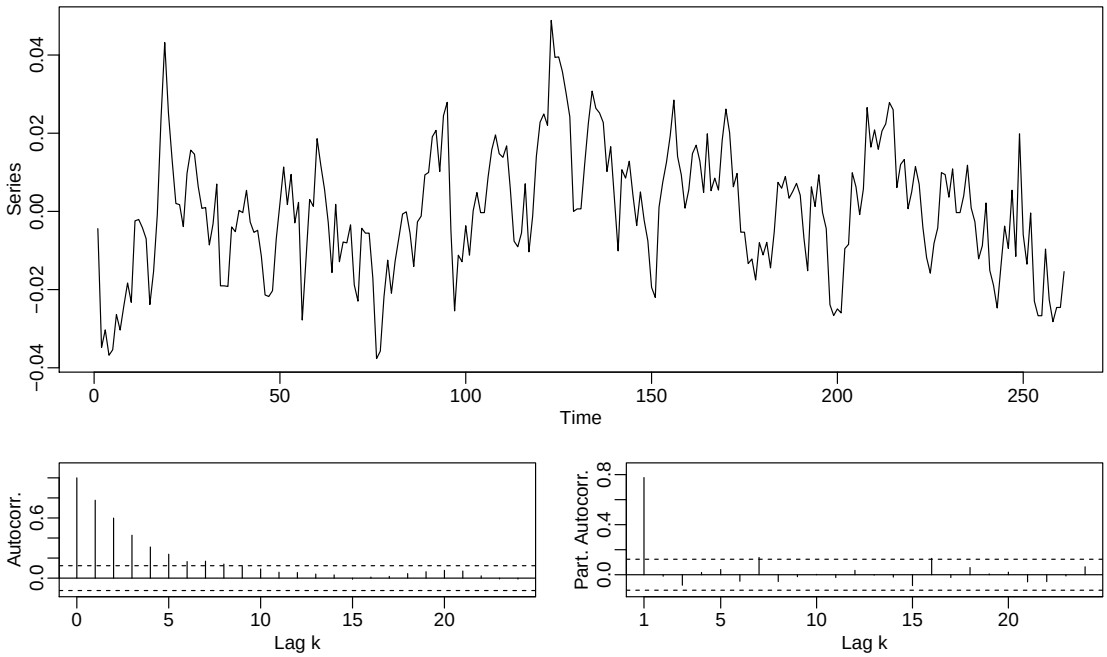

Figure 4.1: Spread series of a potential pair.

like an AR(1) process and is clearly mean-reverting. Now, the idea is to always open a pair-position when the spread series shows a peak, i.e. a relative maximum or minimum, and closing it again when it crosses the zero-line. Of course, ex-ante it is far from clear when the process reaches such a peak. In this particular case here, we could try to model the spread as an AR(1) process and estimate its coefficient, which would be somewhere around 0.8. With that, we could make a forecast for the next value or the next few values of the spread and then decide on this basis whether we should open a pair now or wait until a later point in time as we expect the spread to widen further. Unfortunately, the assumption of m_t following an AR-process may be wrong[4], as a closer look on Figure 4.2 suggests. The figure shows how the spread series, as determined during the matching period, would have developed in the subsequent trading period, i.e. in the period reaching from January 2nd, 1996 up to July 3rd, 1996. Even though the spread seems to be mean-reverting to some extent, it is unlikely to be weakly stationary. Especially towards the end of the trading period the spread departs considerably from zero. This can also be

[3]There is actually a pair with a higher test statistic, but it turns out that those two series refer to the same company.

[4]This may be due to several reasons. One major source may be the general parameter risk involved in every statistical model.

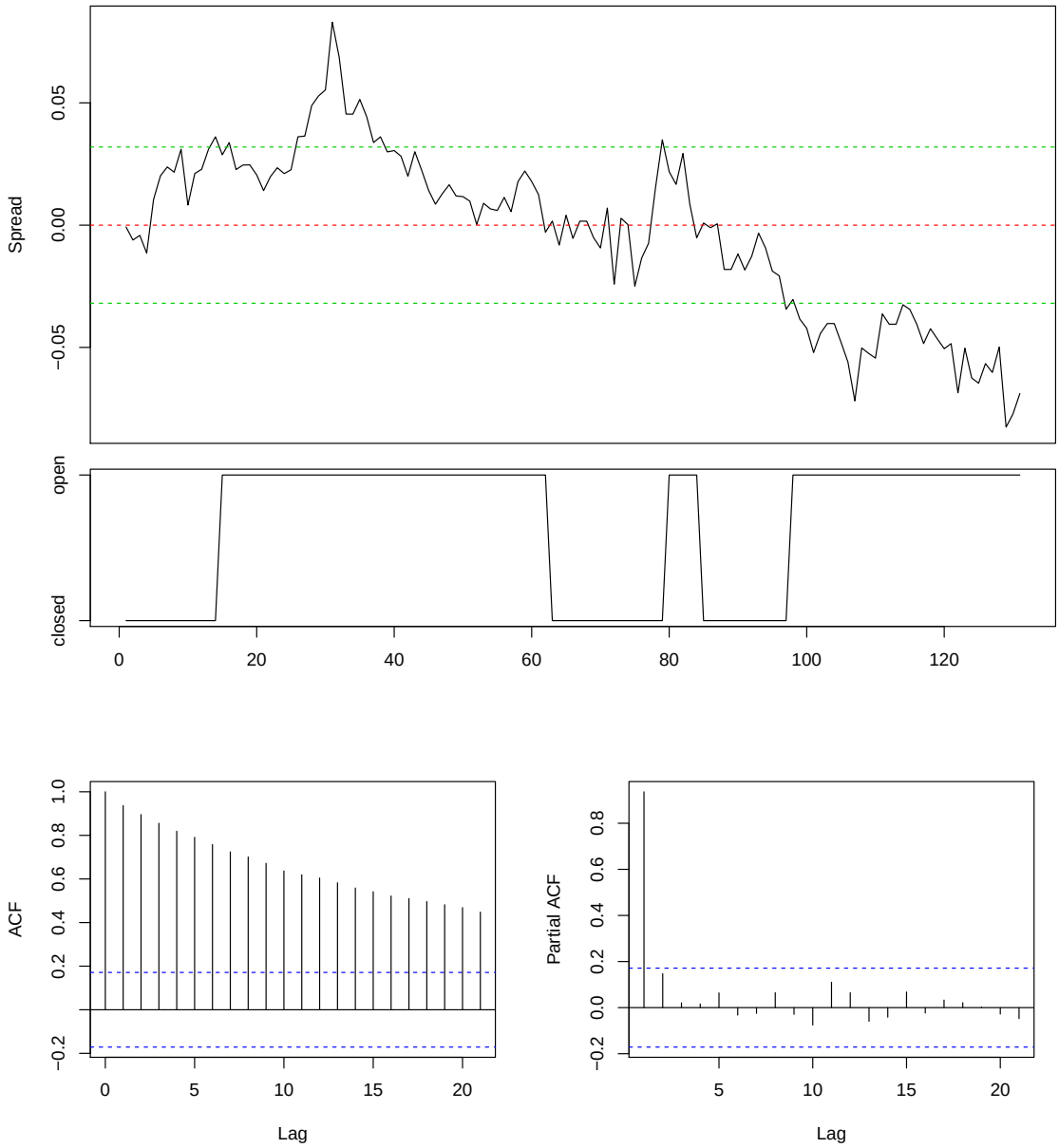

Figure 4.2: Spread series of a fitted pair in the trading period.

seen in Figure 4.3. The distance between the log-prices of the two stocks becomes smaller towards the end of the trading period. Nevertheless, a simple rule that is still promising to work under such circumstances is one that bases the decision to open a pair-position on a simple distance measure. For example, the green dashed lines in Figure 4.2 mark the deviation from the equilibrium relation by two the standard deviations of the spread series as calculated during the matching period. This means, the pair is opened when the spread series deviates by more than two historical standard deviations from its assumed equilibrium. The pair is closed again when the spread series is in equilibrium, i.e. when it crosses the zero-line or, as we limit the trading period of each run, at the end of the trading period, independent of the actual spread value. Such a rule does not condition on the use of log-price series, but can also be applied on the untransformed price series

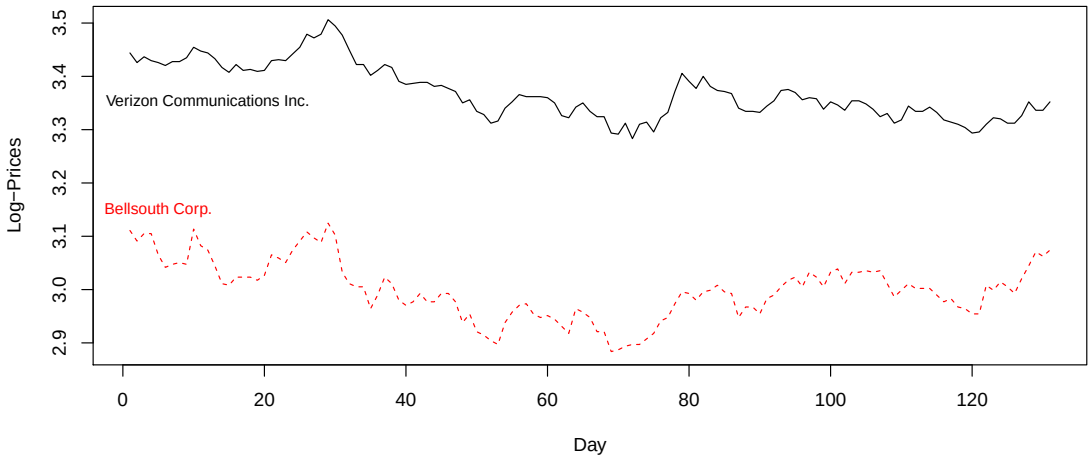

Figure 4.3: Log-price series of a potential pair in the trading period.

directly.

As this example illustrates with its behaviour during the last few days of the trading period, the deviation of the spread vale from its assumed equilibrium value may become large, translating in a heavy trading loss. The loss could even be so high that the portfolio is completely wiped out. One should, therefore, consider a stop-loss rule in order to limit the loss.

4.2.2 A Dynamic Trading Rule

The proposed dynamic trading rule considered in the sequel follows closely the idea of Kim, Primbs, and Boyd (2008) and uses results discussed by Mudchanatongsuk, Primbs, and Wong (2008). It leads to an actively managed investment portfolio. As opposed to the simple trading rule, the dynamic trading rule considered here does not only decide on whether to open or close a particular pair at a particular point in time, but decides for each point in time during the trading period by how much a pair-position should be opened, if at all. Hence, it decides on how much weight we should allocate to an individual position at a certain point in time, thereby considering the development of the whole portfolio of pairs (and not just the single position under consideration). As already mentioned in the previous section, one usually assumes that proper capital that is needed as a collateral is invested in the money market and earns a corresponding interest rate. Kim et al. (2008), however, assume that only the proper capital that is not explicitly used for financing any trades is invested in the money market and evolves (in a continuous time view) according to

$$dB_t = rB_t dt \tag{4.2.1}$$

with r denoting the continuously compounded risk-free rate. In order to capture the mean-reverting behaviour of the spread series m_t, we take the presumably simplest continuous time process suitable for this purpose. This is, we take an OU process as discussed

in section 3.2. In particular, for each of the n pairs in our portfolio, we assume that their spread time series $m_{i,t}$, with $i = 1, 2, \ldots, n$, follow a mean-reverting OU process, where the corresponding parameters are allowed to differ between the different processes. Combining the values of the spread processes for every point in time t into a vector $\boldsymbol{m}_t = (m_{1,t}, m_{2,t}, \ldots, m_{n,t})^T$, we can write all processes together in a compact form, as

$$\boldsymbol{dm}_t = \Theta\left(\boldsymbol{\mu} - \boldsymbol{m}_t\right) dt + S\boldsymbol{dW}_t \tag{4.2.2}$$

with

$$\Theta = \begin{pmatrix} \theta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_n \end{pmatrix} \ , \ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \ , \ S = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{np} \end{pmatrix}$$

and $\boldsymbol{dW}_t = (dW_{1,t}, dW_{2,t}, \ldots, dW_{n,t})^T$ being the vector containing the standard Brownian motions of the n spread series. In the following, the vector $\boldsymbol{h}_t = (h_{1,t}, h_{2,t}, \ldots, h_{n,t})^T$ contains the amount of spread units held at time t for each spread m_i, with $i = 1, 2, \ldots, n$. It tells us how much weight or capital we should allocate to the different spreads at a particular point in time. With I_t denoting the total value of our whole investment at time t, $\boldsymbol{h}_t^T \boldsymbol{m}_t$ is the total value of the positions in the underlying securities of the spreads, and $I_t - \boldsymbol{h}_t^T \boldsymbol{m}_t$ is the amount of money invested in the money market. Hence, the dynamics of the total investment is given by

$$dI_t = \boldsymbol{h}_t^T \boldsymbol{dm}_t + \left(I_t - \boldsymbol{h}_t^T \boldsymbol{m}_t\right) r dt \ . \tag{4.2.3}$$

Now plugging the spread dynamics of equation (4.2.2) into equation (4.2.3) leads to

$$dI_t = \left(r\left(I_t - \boldsymbol{h}_t^T \boldsymbol{m}_t\right) + \boldsymbol{h}_t^T \Theta\left(\boldsymbol{\mu} - \boldsymbol{m}_t\right)\right) dt + \boldsymbol{h}_t^T S\boldsymbol{dW}_t \ . \tag{4.2.4}$$

With this investment value dynamics one can now, in conjunction with an appropriate utility function $U(I_t)$, formulate a stochastic optimal control problem in order to maximize the expected utility for a given investment expiration time T. The utility function is needed as we have to define what we mean by the term *optimal*. It allows us to quantify our preferences. If we neglect all kinds of transaction costs for the moment, the control problem reads then

$$\begin{aligned} &\max_{\boldsymbol{h}} \ E\left[U\left(I_T\right)\right] \\ &\text{subject to} \\ &\boldsymbol{dm}_t = \Theta\left(\boldsymbol{\mu} - \boldsymbol{m}_t\right) dt + S\boldsymbol{dW}_t \\ &dI_t = \left(r\left(I_t - \boldsymbol{h}_t^T \boldsymbol{m}_t\right) + \boldsymbol{h}_t^T \Theta\left(\boldsymbol{\mu} - \boldsymbol{m}_t\right)\right) dt + \boldsymbol{h}_t^T S\boldsymbol{dW}_t \end{aligned} \tag{4.2.5}$$

with given starting values $\boldsymbol{m}_0 = (m_{1,0}, m_{2,0}, \ldots, m_{n,0})^T$ and I_0. $\boldsymbol{h}$ is assumed to be adapted to the filtration $\mathcal{F}_t$ that is associated with the standard Brownian motions under consideration.

The value function is correspondingly defined as

$$J\left(t, I, \boldsymbol{m}\right) = \sup_{\boldsymbol{h}} \ E\left[U\left(I_T\right)\right] \tag{4.2.6}$$

where the supremum is taken over all $\mathcal{F}_t$-adapted stochastic processes. Assuming that the value function is sufficiently regular, we can now establish the *Hamilton-Jacobi-Bellman* equation (HJB). In order to keep the following expressions readable, the function arguments are omitted in the notation below. The HJB equation is, thus, given by

$$-\frac{\partial J}{\partial t} = \sup_{\boldsymbol{h}} \left[\frac{\partial J}{\partial I} \left(r \left(I - \boldsymbol{h}^T \boldsymbol{m} \right) + \boldsymbol{h}^T \Theta \left(\boldsymbol{\mu} - \boldsymbol{m} \right) \right) + \boldsymbol{\nabla}_{\boldsymbol{m}}^T \Theta \left(\boldsymbol{\mu} - \boldsymbol{m} \right) + \frac{1}{2} \frac{\partial^2 J}{\partial I^2} \boldsymbol{h}^T \Sigma \boldsymbol{h} + \frac{1}{2} tr \left(S^T \nabla_{\boldsymbol{m}}^2 S \right) + \boldsymbol{h}^T \Sigma \boldsymbol{\nabla}_{\boldsymbol{m}I} \right] \tag{4.2.7}$$

with $\Sigma = SS^T$, $\boldsymbol{\nabla}_{\boldsymbol{m}}$ denoting the gradient of J with respect to all spread-variables m_i, $\nabla_{\boldsymbol{m}}^2$ being the corresponding Hessian matrix and $\boldsymbol{\nabla}_{\boldsymbol{m}I}$ standing for the partial derivative of the gradient with respect to I. As Σ is the variance-covariance matrix with respect to the diffusion terms of the spreads, it should be positive definite. We can, therefore, say that the expression in the square brackets of equation (4.2.7) multiplied by -1 is a concave function in $\boldsymbol{h}$. As a consequence, we just need to consider the first order conditions

$$\frac{\partial J}{\partial I} r \boldsymbol{m} + \Theta \left(\boldsymbol{\mu} - \boldsymbol{m} \right) + \frac{\partial^2 J}{\partial I^2} \Sigma \boldsymbol{h} + \Sigma \boldsymbol{\nabla}_{\boldsymbol{m}I} = \boldsymbol{0} \tag{4.2.8}$$

in order to express the control vector as

$$\boldsymbol{h} = -\frac{1}{J_{II}} \Sigma^{-1} \left[J_I \left(\Theta \left(\boldsymbol{\mu} - \boldsymbol{m} \right) - r \boldsymbol{m} \right) + \Sigma \boldsymbol{\nabla}_{\boldsymbol{m}I} \right] \tag{4.2.9}$$

with $\frac{\partial J}{\partial I} = J_I$ and $\frac{\partial^2 J}{\partial I^2} = J_{II}$. Substituting this expression back into equation (4.2.7) leads to the following partial differential equation for the value function

$$\begin{aligned} 0 =& J_t + J_I r J + \boldsymbol{\nabla}_{\boldsymbol{m}}^T \Theta \left(\boldsymbol{\mu} - \boldsymbol{m} \right) + \frac{1}{2} tr \left(S^T \nabla_{\boldsymbol{m}}^2 S \right) \\ & - \frac{1}{2 J_{II}} \left[J_I \left(\Theta \left(\boldsymbol{\mu} - \boldsymbol{m} \right) - r \boldsymbol{m} \right) + \Sigma \boldsymbol{\nabla}_{\boldsymbol{m}I} \right]^T \\ & \Sigma^{-1} \left[J_I \left(\Theta \left(\boldsymbol{\mu} - \boldsymbol{m} \right) - r \boldsymbol{m} \right) + \Sigma \boldsymbol{\nabla}_{\boldsymbol{m}I} \right] \end{aligned} \tag{4.2.10}$$

with terminal condition $J \left(T, I, \boldsymbol{m} \right) = U \left(I \right)$.

According to Kim et al. (2008) there exist explicit analytical solutions for such partial differential equations if the utility functions are of the type

$$U \left(I \right) = \begin{cases} \frac{I^{1-\gamma}}{1-\gamma} & , \ \gamma \neq 1 \ , \ \gamma > 0 \\ \log I & , \ \gamma = 1 \end{cases} \tag{4.2.11}$$

or

$$U \left(I \right) = -\frac{1}{\gamma} e^{-\gamma I} \ , \ \gamma > 0 \ . \tag{4.2.12}$$

If the problem is not analytically solvable, we could still evaluate equation (4.2.10) numerically and use the resulting values to calculate the control vector.

Obviously, this approach is computationally more expensive than the simple trading rule of section 4.2.1. Its dynamic nature may lead to a better result in the sense that it leads to a higher utility, which finally means something like more profit and / or a less risky price path of the whole pairs-portfolio. It relies, however, on the very crucial assumption that

the spread series are indeed mean-reverting. The approach outlined above assumes even a weekly stationary behaviour of the single spread series as they are assumed to follow an OU process. So, one would need to think quite a bit about the implications of not so well behaved spread series on the optimisation results. In addition, the rule involves quite some economic theory, i.e. utility theory, which should also be captured properly before applying it. The rule as sketched here, would probably give enough material to cover a complete thesis. It is, of course, a very interesting approach, but it is also clearly beyond the scope of this thesis. It is, therefore, thought as an outlook but will not be considered in the sequel.

4.3 Return Calculation

Calculating returns is usually[5] not a big issue. As briefly discussed at the beginning of section 3.3, there are two main concepts, i.e simple returns and continuously compounded returns. Both are relative values, this is, they measure one value relative to another value. The return with respect to a particular period in time of an asset held *long* is usually calculated as the value of the asset at the end of the period relative to its value at the beginning of the period. The simple return on a traditional investment is, thus, just the money earned on one monetary unit of the whole investment, i.e. the amount an investor gets for investing one Dollar.

Regarding a statistical arbitrage investment, the return calculation may not be so obvious at the first sight. One could argue that an investor does actually not need to invest any own money at all as the corresponding portfolio is self-financing. This is, one sells-short some particular assets and buys with the proceeds some other assets (the long-position). Thus, the long-position is financed by the short-position. This is also the reason why the term *arbitage* appears in the strategy's name, which actually means in its strong form that there is a risk-free profit without the need of investing any own money. In reality things look, however, a bit different. It should have become clear from the discussion on the topic so far that a statistical arbitrage investment is not risk-free at all. In addition, as we already stated above, such a strategy does indeed require some proper capital. This is because in order to short-sell assets, one needs to borrow them first. Usually, this can only be done by depositing other assets (or just money) of the same value (or at least a particular fraction of the value), a so-called *collateral*. An investor usually has to pay an *initial margin*. This is a particular amount of money paid at the beginning, which serves as a buffer with respect to the fluctuations in the net value of the investor's positions. If the value of the portfolio falls short of a previously specified lower boundary, the investor gets a so-called *margin-call*, meaning that he has to provide additional funds, otherwise his positions will be closed. The amount put aside to serve as collateral typically earns a short-term interest rate, at least as long as the margin requirements are fulfilled. In this situation, the change in value of the pair-positions relative to the amount of the collateral can, therefore, be interpreted as excess return, i.e. the return earned above the risk-free rate.

It is clear that the value of a pair-position and the cash flows it generates are two different concepts. While the later is important concerning liquidity issues of the whole fund, it is only the value of the fund that matters with respect to the return calculation. We can,

[5] That is for investments consisting of only long-positions.

therefore, first calculate the value of the fund for each point in time and then calculate the simple returns as described in section 3.3. For a particular investment period, the value of the fund, denoted as V, can be determined recursively as

$$V_{t+1} = V_t + C_t r_{f,t} + \sum_{i=1}^{n} h_{i,t} \left(\upsilon_{\tau_i,t+1}^{(i,long)} - \upsilon_{\tau_i,t+1}^{(i,short)} \right) \tag{4.3.1}$$

where i indicates the pair-position, $r_{f,t}$ denotes the return of the risk-free rate and C_t is the value of the capital used as collateral at time t. By this, the initial margin at the beginning of a trading period is $C_0 = V_0$. As this capital is interest bearing, it simply evolves according to $C_{t+1} = C_t(1 + r_{f,t})$. $h_{i,t}$ is the weight we put on the pair-position i at time t. Clearly, it must hold that $\sum_{i=1}^{n} h_{i,t} = C_0$ if we require an initial margin of 100%. In the context of the simple trading rule, as explained in 4.2.1, it is then just $\frac{C_0}{n}$ if the corresponding position is open, and 0 if it is closed. The $\upsilon_{\tau_i,t+1}^{(i,long)}$ refers to the value of an investment of one Dollar in the long position of pair i with transaction time τ_i, where $0 \leq \tau_i \leq t$. $\upsilon_{\tau_i,t+1}^{(i,short)}$ denotes the corresponding short position, i.e. the position that is sold short in order to receive one Dollar at time τ_i. The term $\upsilon_{\tau_i,t+1}^{(i,long)} - \upsilon_{\tau_i,t+1}^{(i,short)}$ represents, thus, just the net-value of the long and the short position of a pair. For example, if we open a pair-position at time τ, we lend as many shares of the relatively overvalued asset so that we obtain one Dollar when we sell this position. With the received Dollar we then buy as many shares of the relatively undervalued asset as we can get for that amount of money. So, at time τ the net value of this pair-position is zero. If we observe in the next period, i.e. $\tau+1$, that the asset of our long position increases by 10% while the asset of our short position increased by just 5%, the value of the long position is then \$ 1.10 and the value of the short position \$1.05. The net-value, i.e. the value that remains after reversing the positions and returning the amount of shares lent, is in this case $1.10 - 1.05 = 0.05$.

4.4 Results

Even though the performance of the past does not necessarily have to be the same in the future, analysing the past may, nevertheless, tell us whether the strategy is at least promising for future investments. Using the past 16 years, we can analyse how an investment of one Dollar would have performed up to today. The time span from the beginning of 1996 up to the end of 2011 is certainly an interesting time period for any back-testing as it comprises *bullish* as well as *bearish* market environments, i.e. good times as well as crises in the financial market. As an appropriate benchmark we choose the S&P 500 index. First of all, investing in the S&P 500 index[6] can be seen as a passive investment in a large portfolio. Secondly, the S&P 500 is also the market from which we take the assets involved in our strategy.

The back-testing is done for four different strategy specifications. This is, two funds base their selection and trading criteria on log-price series. We call them *LPB* or *log price based* funds. The other two funds base the corresponding criteria on the price level series directly. Remember, there is no *right* or *wrong* with respect to price level transformations. It is a question of believes about the true data generating model[7]. We call them *PLB* or *price level based* funds. For each fund type we select one to consist of a portfolio of 20 pairs

[6] This is usually done by investing in so-called tracker-funds or by future contracts.

[7] See again the short discussion in section 2.3.

(i.e. 40 assets) and the other to consist of a portfolio of 50 pairs (i.e. 100 assets). These numbers are chosen arbitrarily. As a broad guidance we can argue as follows: On the one hand, the more assets we choose, the higher the risk becomes to have *error-I-pairs*, i.e. pairs that are spuriously taken as cointegrated, in the trading portfolio. It also increases the transaction costs. On the other hand, the less assets we take, the less diversified the portfolio becomes, resulting in a higher dependence on the single pairs in the portfolio.

Figure 4.4 and Figure 4.5 show the development of the 20 pair-positions of the LPB 20 fund during the first trading period, i.e. from January 1996 up to June 1996, with an implemented stop-loss level at -20%. Similarly, Figure 4.6 and Figure 4.7 correspond to the development of the PLB 20 fund with respect to the same period. The plateaus in the graphs correspond to trading days where the pair-position is closed. As can be seen in Figure 4.4, the pair-position with the highest cointegration statistic, *Best-Pair 1*, does never open during the first trading period. *Best-Pair 2* in this figure corresponds to the spread series depicted in Figure 4.1. When we look at this position a bit more closely, we might wonder why the plateau in the net-value of a pair-position is not necessarily a local maximum. This is because of the fact that the net-value of a pair-position is simply the difference between the cumulative gross-return of the long and the short position. As just explained above, a pair-position is opened by selling one asset short of the amount of \$ 1.00 and purchasing the pair-asset of the amount of \$ 1.00. As long as both of these assets move up or down by the same percentage value, the net-value of the pair-position is unchanged. We can say that the trading portfolio is *money based*, as we use the proceeds of the short-sale in order to purchase the pair-asset, not in the same amount of stocks, but in the same amount of money. The estimated equilibrium relationships are, however, *unit based*. This is, we use in the relationship estimation the prices or log-prices of one share of the assets, i.e. one unit of them. As a consequence, an increase or decrease of the asset values by one and the same percentage amount does not leave the unit based relationships unchanged when we have $\beta \neq 1$, i.e. when the slope coefficient is different from one, in the context of the log-price based rule, or $\alpha \neq 0$, i.e. when there is a constant in the cointegration relationship, in the context of the price level based rule.

Table 4.1 summarises the results of a one Dollar investment in the two fund types, each consisting of 20 and 50 pairs, respectively, as well as in the S&P 500 index. The first two lines report the means of the daily and the yearly returns, respectively. As can be seen, the price-level based fund consisting of 20 pairs (PLB 20) yielded an average[8] yearly return of 9.1% over the whole 16 years. By comparison, an investment in the S&P 500 index yielded only 4.1% p.a. on average over the same period. An investment in the S&P 500 index performed even worst over the whole investment period. However, as can be seen in Figure 4.8 and Figure 4.9, during particular sub-periods the index investment developed better than any of the four statistical arbitrage funds. The mean excess return is just the return in excess of the risk-free rate. The corresponding excess returns are stated in line 3 and 4 of the result table.

Clearly, as investors we are not only interested in the average returns but also in the fluctuations of the returns. Line 5 and 6 display the standard deviations of the daily and yearly returns. Here again, the S&P 500 index shows the worst result with a yearly standard deviation of 20.93%. The best value here is shown by the LPB 50 fund with only 5.1%. The diversification effect is clearly visible as the portfolios consisting of 50 pairs both dominate the portfolios consisting of 20 pairs. In this respect Figure 4.10

[8] This is the geometric mean as we work with simple returns here.

	LPB 20	LPB 50	PLB 20	PLB 50	S&P 500
Daily Mean	0.00030	0.00025	0.00033	0.00030	0.00015
Yearly Mean	0.08048	0.06791	0.09104	0.08054	0.04101
Daily Mean Excess	0.00016	0.00012	0.00020	0.00016	0.00008
Yearly Mean Excess	0.04389	0.03175	0.05409	0.04394	0.02176
Daily Stand. Dev.	0.00432	0.00315	0.00479	0.00339	0.01295
Yearly Stand. Dev.	0.06977	0.05097	0.07739	0.05483	0.20928
Daily Min.	-0.03789	-0.03319	-0.03670	-0.03556	-0.09035
Daily Max.	0.04422	0.03759	0.03451	0.02797	0.11580
Skewness (daily)	0.81558	0.56374	0.22405	0.19655	-0.03013
Kurtosis (daily)	16.00432	16.40852	7.48785	9.60749	10.55648
Value at the end	3.41496	2.83631	3.98450	3.41756	1.89210

Table 4.1: Summary statistics of the four prototype funds

shows a very nice picture. It shows the returns of all the funds and the index over time. By this way, one can see very well, when and how strongly the returns fluctuated. An investment in the S&P 500 index is, thus, from a risk avoiding perspective clearly less desirable. In this respect we may also want to know what the worst return on any day during the whole investment period was. This number is stated in line 7 of Table 4.1. While non of the statistical arbitrage funds ever had a daily return smaller than -3.8%, the S&P 500 index shows a worst return that is smaller than -9%. Line 8 and 9 give some additional information about the univariate distribution of the returns, the skewness and the kurtosis. Interestingly, all the fund returns are positively skewed, while the returns of the S&P 500 index are skewed to the left. With a kurtosis of clearly larger than 3, all the funds, as well as the index, show so-called excess kurtosis, meaning that their distributions have considerably more mass in the tails as compared to a normal distribution. This fact is, however, nothing really special. It is widely acknowledged in the financial literature that daily returns show considerable excess kurtosis as compared to a normally distributed variate. The last line of Table 4.1 shows what a $ 1.00 investment executed at the beginning of 1996 would have been at the end of 2011. For example, an investment in the PLB 20 fund at the beginning of 1996 would have been almost four times as much at the end of 2011.

Clearly, an investment that yields high returns with limited fluctuations is desirable. However, having a closer look at the fund developments, as depicted in Figure 4.8 and Figure 4.9, reveals another very interesting aspect. The S&P 500 index shows quite neatly the occurrence of two major financial market incidences. The first one is the burst of the so-called *dot-com bubble* in mid 2000. The second one is the so-called *sub-prime crisis* starting towards the end of 2007. While the S&P 500 index lost pretty much in value during those incidences and the following transition, the statistical arbitrage funds were hardly affected by the *dot-com bubble burst* and even positively affected by the *sub-prime crisis*. This suggests an interesting dependence structure between the statistical arbitrage funds and the S&P 500 index, which reflects the behaviour of a broad well diversified market portfolio to some extend.

One last remark concerns transaction costs. One could argue that an investment in the S&P 500 index is a passive one and requires, thus, only one transaction at the beginning of the investment in 1996 and one at the end, i.e. in 2011. The statistical arbitrage

funds, by contrast, open and close several positions during the year, which finally means that the transaction costs of the funds are much higher than those of passive investments. This is partially true. However, it is wrong to believe that one could buy a share in an index as one can do in a common stock. In order to replicate the S&P 500 index one has either to buy shares according to the index weighting of the 500 stocks, where we have to keep in mind that the composition of the index also changes over time, or to buy corresponding future contracts with limited maturity, which means that the contracts have to be renewed several times over a longer investment period. Obviously, one can also invest in an index-replicating fund. However this fund faces the same transaction costs as just explained and will certainly allocate these costs to the investor. Nevertheless, with 17.3 and 16.3 transactions per year and pair-position on average for the LPB 20 fund and the PLB 20 fund, respectively, the transaction costs are still about an estimated 2.5 times higher than the transaction costs of holding a replicating portfolio of the S&P 500 index without futures. The replication by using future contracts is presumably even cheaper. In this regard, when taking all transaction costs into account, the return difference between the funds and the index becomes smaller. The favourable volatilities of the funds do, however, not change. An argument in favour of the fund can also be added. This is, in the back-testing, as presented here, we chose an initial margin requirement of 100%. This is a rather conservative assumption. Depending on the kind and size of the investment fund, the initial margin requirement can be considerably lower. This, in turn, means that the fund returns could probably even be inflated by using a higher leverage.

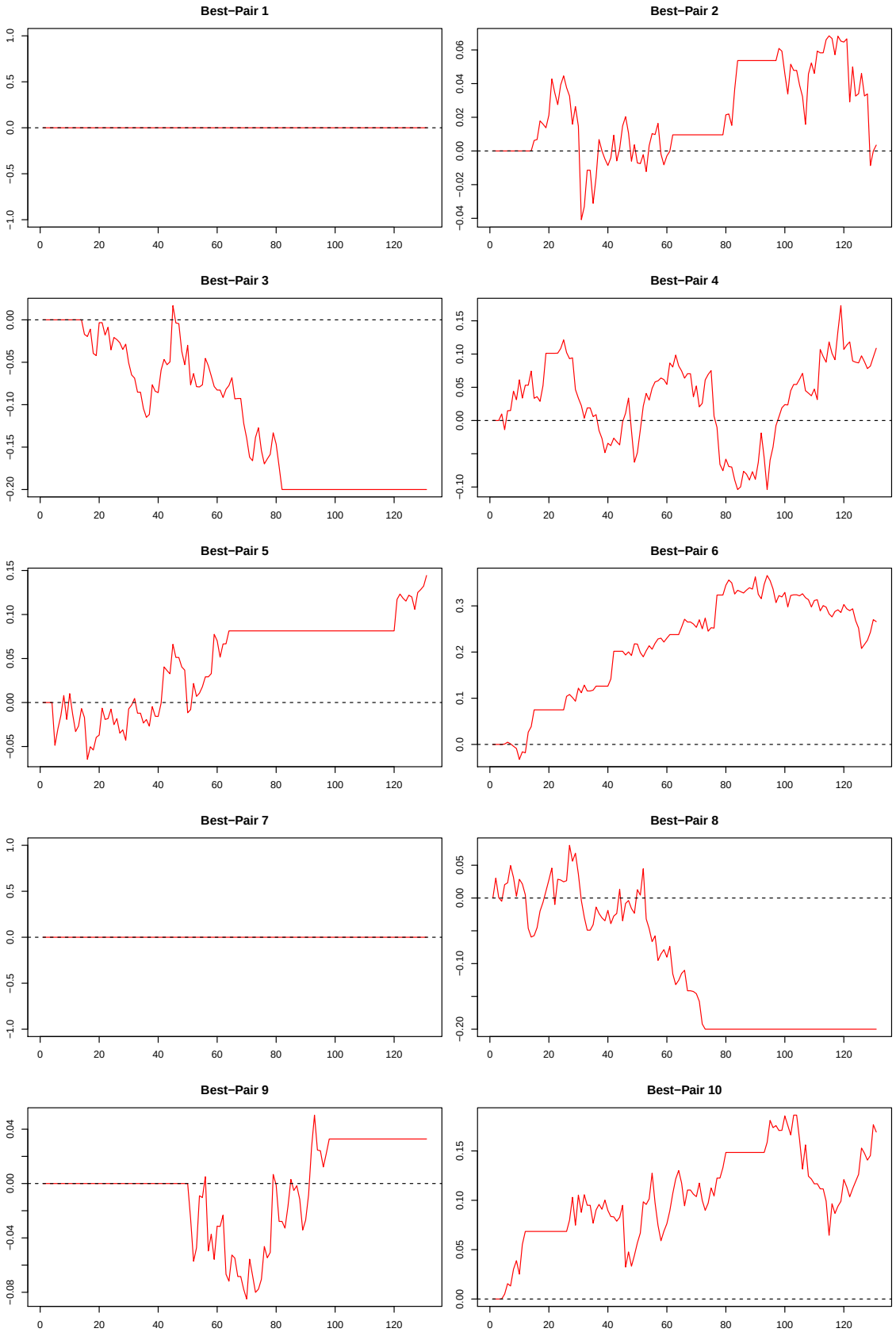

Figure 4.4: LPB20: Development in value of the first ten positions of period one.

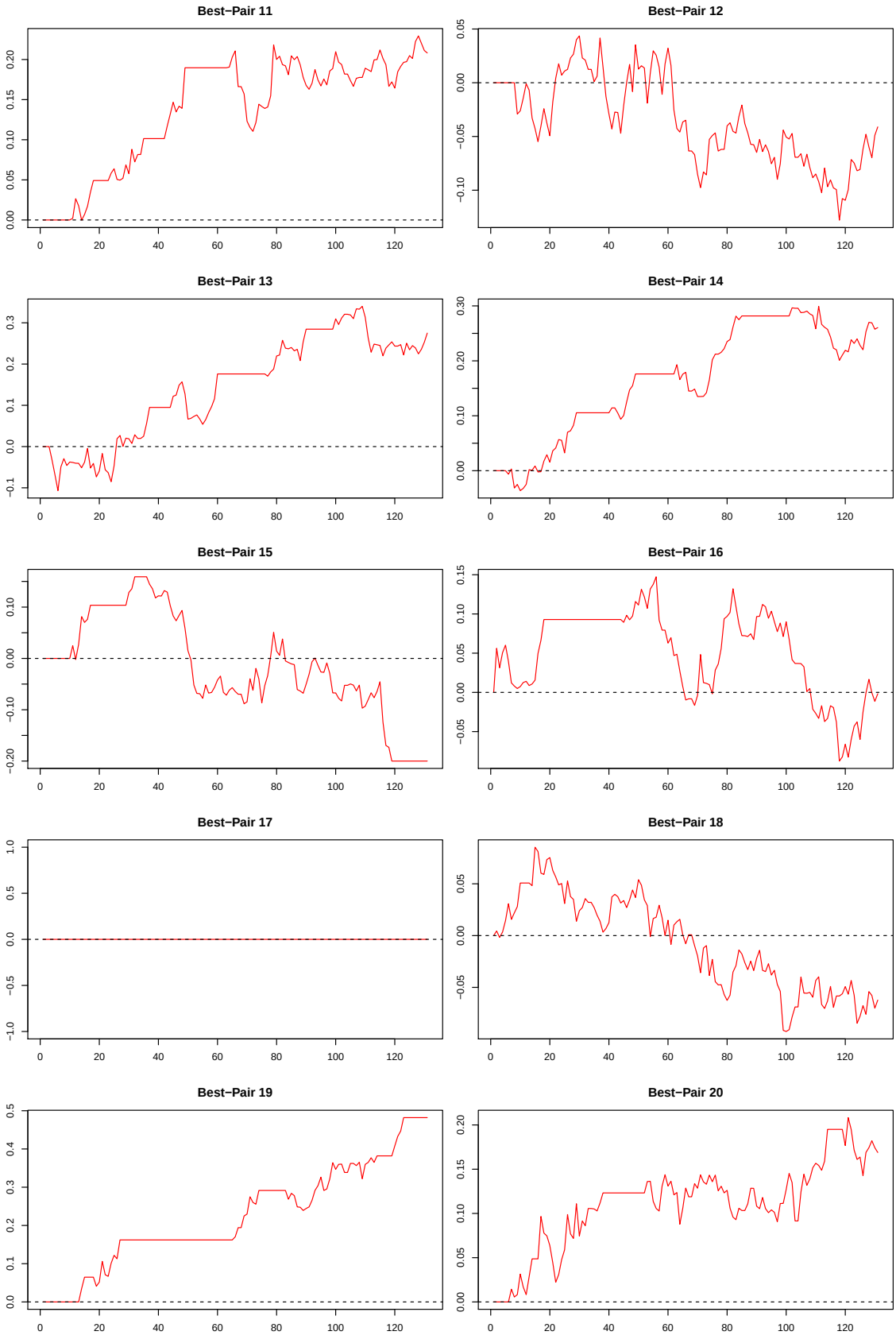

Figure 4.5: LPB20: Development in value of the second ten positions of period one.

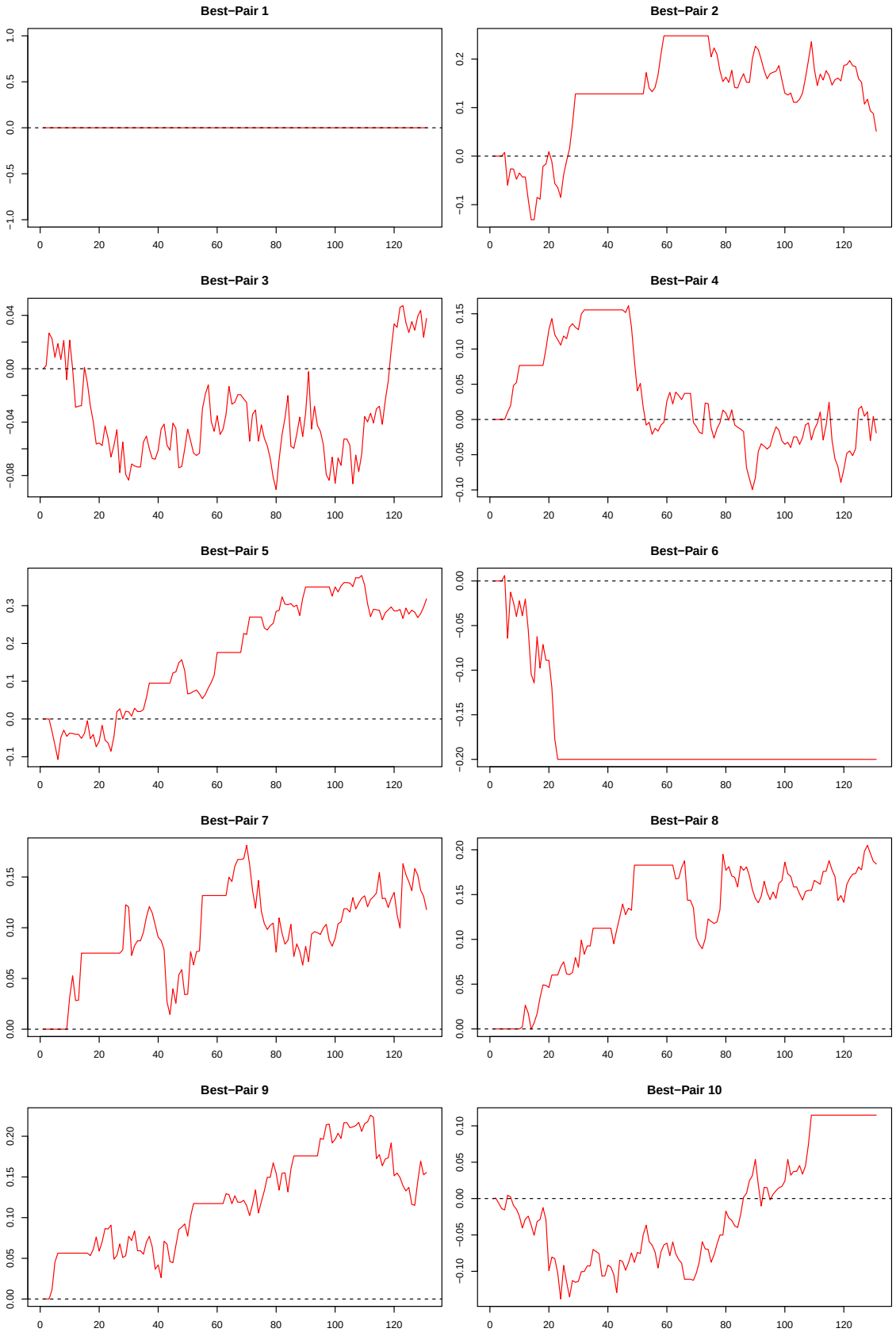

Figure 4.6: PLB20: Development in value of the first ten positions of period one.

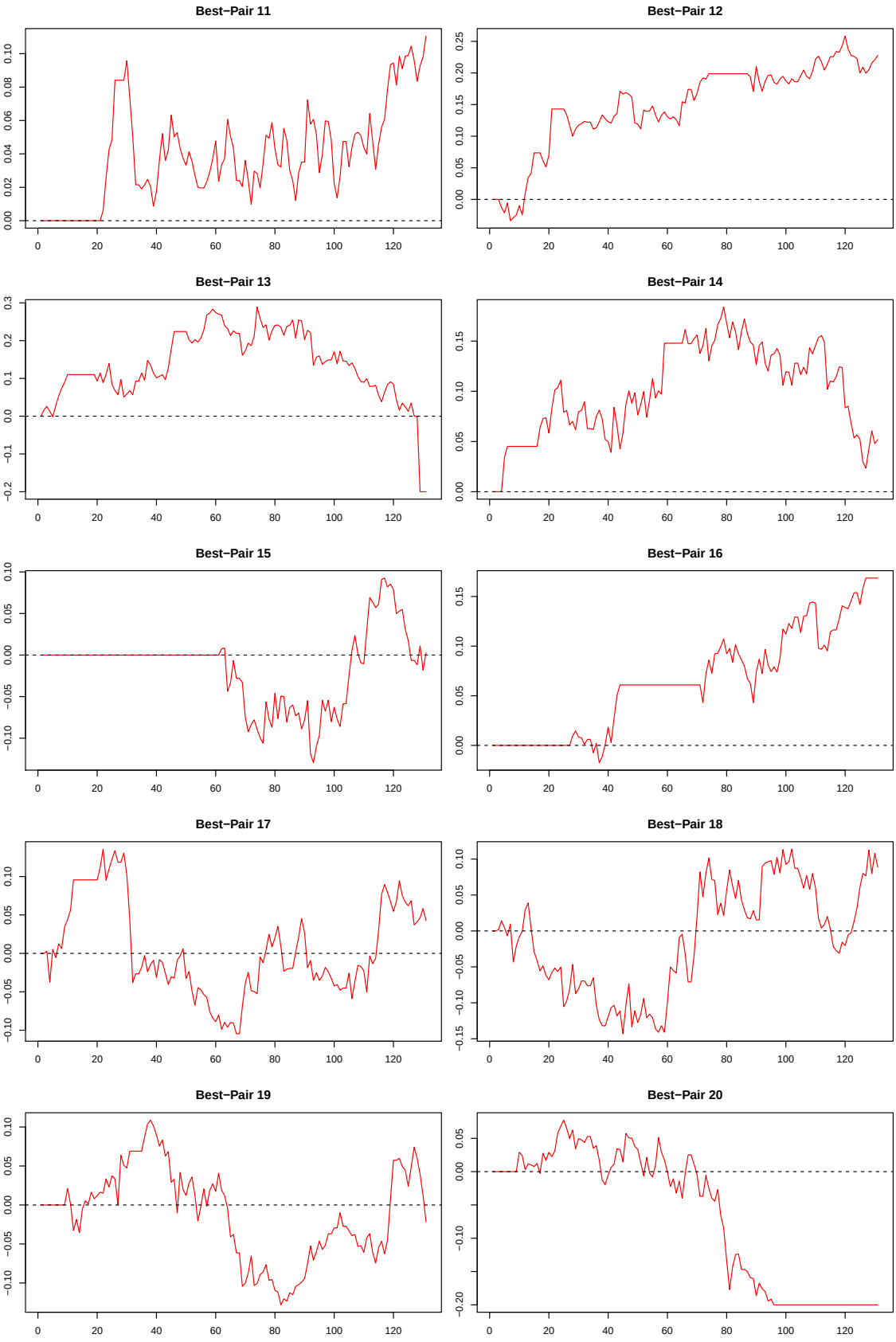

Figure 4.7: PLB20: Development in value of the second ten positions of period one.

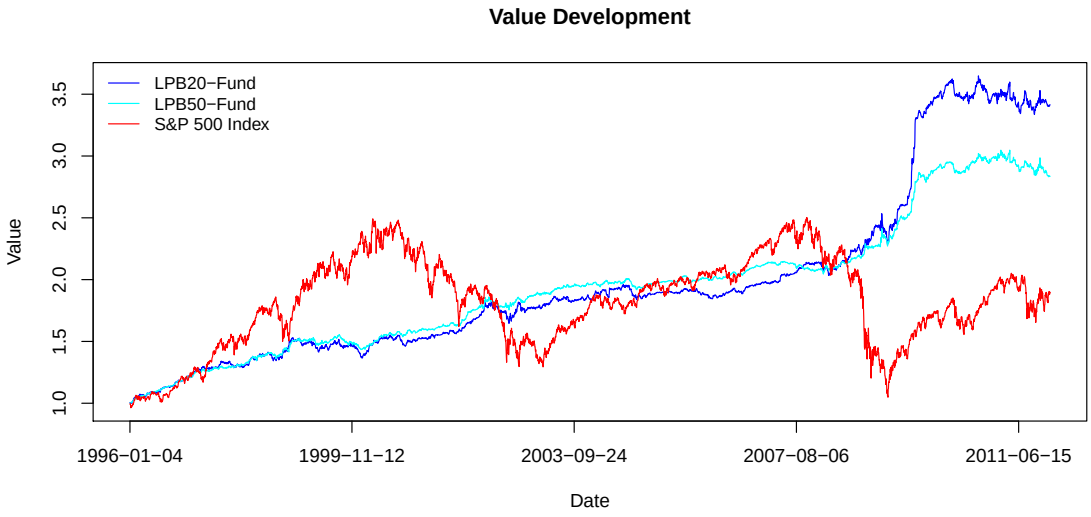

Figure 4.8: Value of the log-price based statistical arbitrage funds.

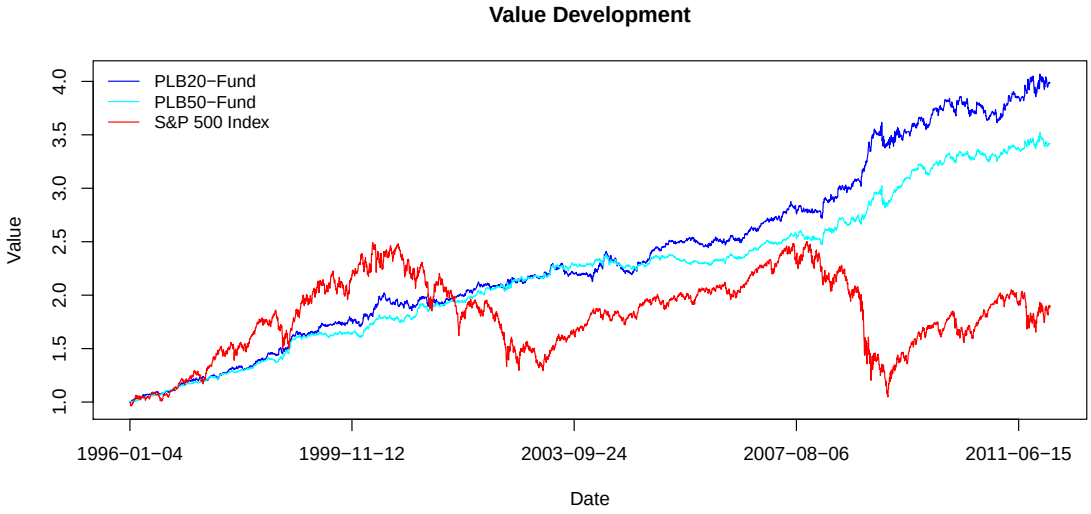

Figure 4.9: Value of the price-level based statistical arbitrage funds.

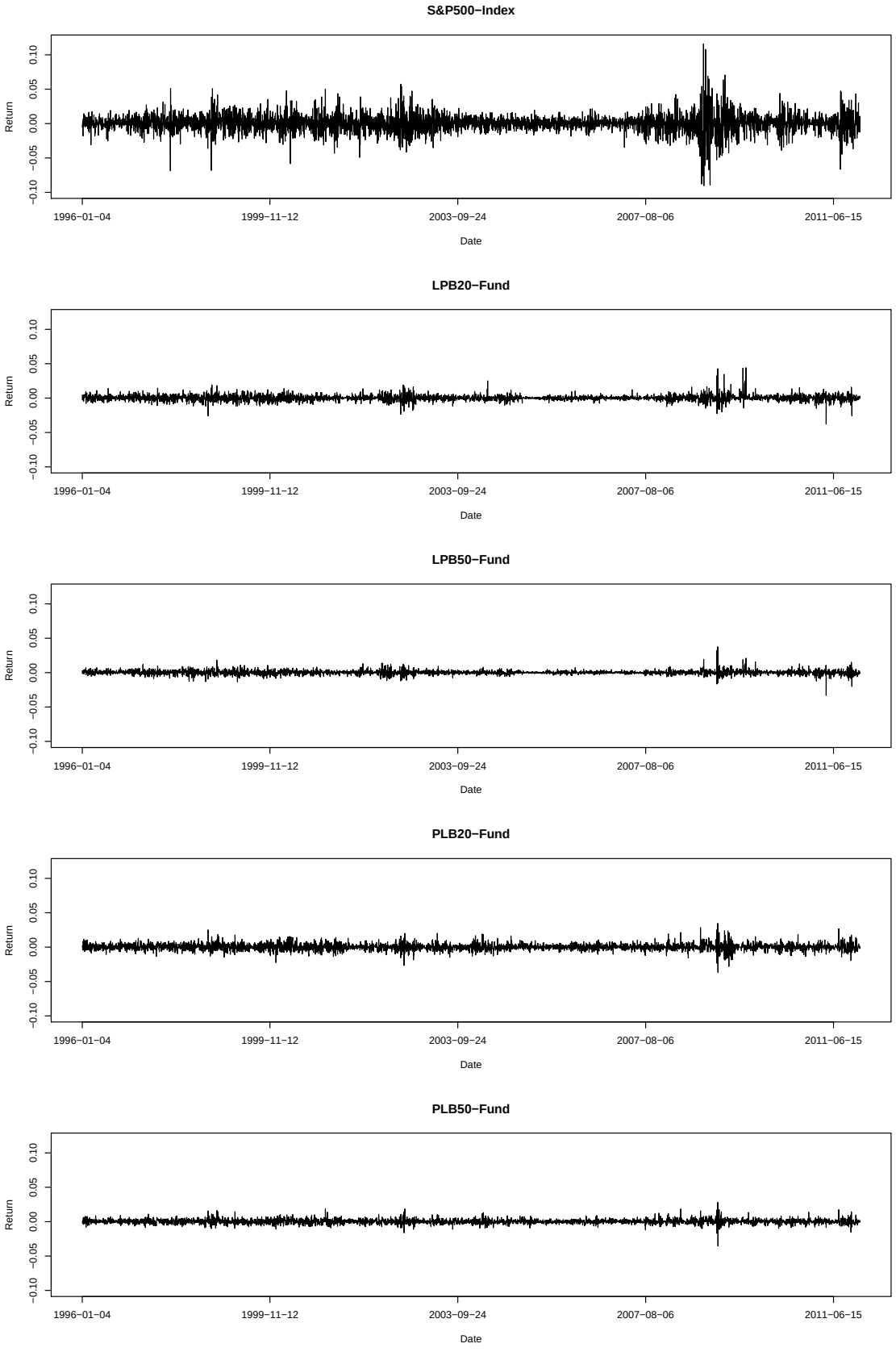

Figure 4.10: Returns over the whole period.

Chapter 5

Measuring Performance and Market Dependence

After the back-testing of chapter 4 we now want to analyse the obtained returns of the statistical arbitrage funds more closely. In "classical" finance it is usually assumed that an investor is only rewarded for the bearing of systematic risk, i.e. the risk that cannot be diversified away. The idea behind this is that the so-called unsystematic risk, or idiosyncratic risk, can be avoided by holding a well diversified portfolio. The *market portfolio*, as it is understood by Tobin (1958), is the (theoretically) most diversified and mean-variance optimised portfolio available. The risk of the *market portfolio* is, thus, purely systematic. So, if we want to measure the performance of a particular investment in this context, we must be able to appropriately capture its systematic risk component. The most famous model that aims to do this is the *Capital Asset Pricing Model* (CAPM). The "market" in this context is, however, a theoretical construct. In practice one usually takes any broad and well diversified portfolio that is of particular relevance for an investor, like that of the S&P 500 index, for instance. Another shortcoming of the CAPM is that it is based of a linear dependence concept only. This can be problematic, as we will illustrate later in this part. Therefore, we want to go one step further by focusing more closely on the dependence structure between the proposed actively managed statistical arbitrage portfolios and a *market* substitute like the S&P 500 index. In order to account for unusual multivariate interactions, we use best a very flexible multivariate modelling technique, such as copulas. Before we start with this task, we first want to highlight the need of it by a brief introduction to the topic of traditional financial performance evaluation which is then followed by a counterintuitive example concerning the concept of linear correlation and the view on marginal distributions only.

5.1 Traditional Performance Evaluation

Investments yielding high returns are clearly desirable. However, a high return is usually not the only criteria. High returns can be bought to some extent by investing in more risky financial products. For example, an investment in stocks can generate much higher returns than an investment in bonds. However, a stock can also loose a considerable amount of its value over a rather short period of time. Hence, the risk of an investment must always be taken into account when measuring its financial performance. Otherwise we just compare

apples and oranges. The remaining question is then, of course, what we exactly mean by the term *risk*. Here one can certainly debate long and extensively. In "conventional" finance one often just takes the standard deviation of the returns, the so-called volatility, as a measure of risk. The higher the volatility, the riskier the investment is. Having said this, a very natural and simple performance measure is the so-called *Sharpe ratio*, defined as

$$SR_i := \frac{\overline{r}_i - \overline{r}_f}{\sigma_i} \tag{5.1.1}$$

where $\overline{r}_i$ stands for the average return of asset i and $\overline{r}_f$ for the average risk-free return. The σ_i denotes the standard deviation of the returns of asset i. So, the Sharpe ratio expresses the excess return of an investment in units of standard deviations. In other words, it just measures how much risk premium an investment generates per unit of risk it is exposed to. The *Sharpe ratio* does, however, not distinguish between the systematic and the unsystematic risk of an asset. A concept that makes this distinction explicitly is the *Capital Asset Pricing Model* (CAPM) as proposed by Sharpe (1964), Lintner (1965) as well as Mossin (1966). By only considering the systematic risk component, it is, thus, a theory compliant pricing model for basically any kind of financial asset. In theory, it gives a precise prediction of the relationship that should be observed between the overall volatility of an asset and its expected return. If the theory holds true, then the model should provide a benchmark rate of return for evaluating any investment project. The relationship the CAPM proposes for the return of an asset i is

$$E[r_i] - r_f = \beta_i \left(E[r_M] - r_f\right) \tag{5.1.2}$$

where $E[r_M]$ denotes the expected return of the market portfolio, which comprises any imaginable kind of asset, and $\beta_i = \frac{\mathrm{Cov}(r_i, r_M)}{\mathrm{Var}(r_M)}$. So, on the left hand side of equation (5.1.2) we have the expected excess return, or so-called risk premium, of asset i, and on the right hand side the expected excess return of the market portfolio multiplied by some factor β_i. Obviously, if we denote again the standard deviation by σ, we can write $\beta_i = \frac{\rho_{i,M}\sigma_i}{\sigma_M}$. The $\rho_{i,M}$ stands in this case for the linear correlation according to Bravais-Pearson between the market return and the return of asset i. From this, it becomes clear that the CAPM demands a higher risk premium for assets that are more volatile and have at the same time returns that are higher correlated with the market returns. If we now observe that the actual excess returns of an asset is systematically higher than the one predicted by the CAPM, we say that the asset has so-called α. The α is usually interpreted by finance people as the part of the risk premium that does not depend on the market, meaning that the asset shows, at least partially, *absolute* returns. The idea stems from the parameter estimation in the CAPM which is usually carried out by estimating a simple linear model, including a constant α and a slope coefficient β. So we see, the CAPM, too, uses the standard deviation as a measure of risk. However, the relationship that the model states is based on the linear correlation between the returns of a particular asset and the market returns. That this concept may be misleading when we leave the world of elliptical distributions, is demonstrated in the next section.

5.2 The Problem with Linear Correlation and Marginal Distributions

A nice way to illustrate the problems arising from only considering the pairwise linear correlation and the univariate marginal distributions of a multivariate problem, is to give a rather counterintuitive example following McNeil, Frey, and Embrechts (2005) (chapter 5). Figure 5.1 shows the scatter plots of four different bivariate problems.

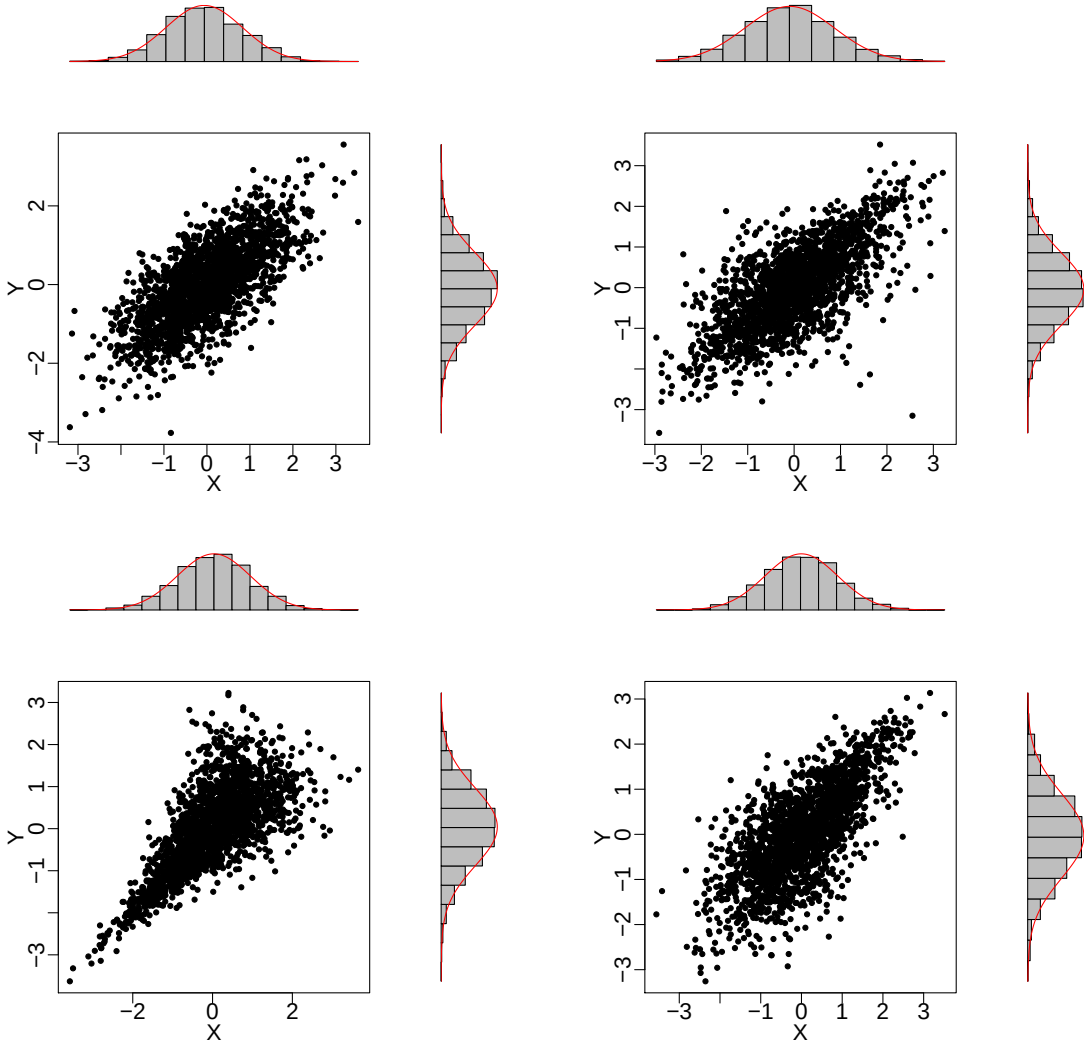

Figure 5.1: Four different bivariate problems.

If we consider for the moment just the scatter plots without the marginal distributions, the plots look obviously very different, even though all of them certainly show a positive correlation between the two variables X and Y. The plot on the upper left looks very much like the one generated by a bivariate standard normal distribution. The one on its right hand side looks similar, though with more dispersion and generally more joint extreme values on the upper as well as on the lower tail. The two plots on the lower left and lower right also show some joint extreme behaviour, although only with respect to one tail. A distribution as the one depicted on the lower left in Figure 5.1 is obviously the least liked from a portfolio manager's point of view with respect to asset returns. This would mean

that in normal and good times the diversification effect is well present, but unfortunately not in bad times when it would be needed most.

Copula	Linear Correlation	Rank-Correlation	
	Bravais-Pearson	Spearman's rho	Kendall's tau
Gaussian	0.70	0.68	0.50
Student t	0.68	0.66	0.49
Clayton	0.70	0.71	0.53
Gumbel	0.69	0.68	0.49

Table 5.1: The different correlations.

We might be surprised when we examine the marginal distributions of the four scatter plots, which are also depicted in Figure 5.1. The red line in each of the histograms presents the theoretical standard normal distribution. As can be seen, there is not much distinction between the histograms of the different marginal distributions. In fact, all margins have a standard normal distribution. In addition, when we now have a look at the popular Bravais-Pearson coefficient of linear correlation, as given in Table 5.1, we might be surprised a second time. As can be seen, the coefficients with respect to the four different bivariate distributions are almost identical. The example here illustrates very well that the marginal distributions together with the pairwise linear correlations of a random vector do not determine its joint distribution. To make this point clear Figure 5.2 illustrates the four joint density functions from which the scatter plots in Figure 5.1 have been generated. Figure 5.3 shows the contour plots of the corresponding densities. No doubt, they are distinct, even if they have the same margins and virtually the same linear correlation. The problem with the concept of linear correlation, and in general with a picture given by the plots in Figure 5.1, is that it does not only incorporate information about the dependence between two random variables but also about their marginal behaviour. This becomes intuitively clear when we transform each of the individual variables so that they obtain a different univariate distribution[1]. Clearly, the picture of the corresponding scatter plots, as well as the linear correlation coefficients, would look very different after the transformation. If one considered the copula of the four joint distributions instead, one would realize that they were still the same as before the transformation. This is because of the fact that copulas are invariant under strictly increasing transformations of the margins. Figure 5.4 shows the copula generated points of the four example scatter plots in Figure 5.1.

5.3 Theoretical Considerations

The joint modelling of random variables is a somewhat tricky thing. One typically needs to estimate the joint distribution of the random variables. Traditional methods are somewhat limited in the sense that in a parametric framework the individual behaviour of the variables must be characterized by the same parametric family of univariate distributions. In a non-parametric framework we face the so-called *course of dimensionality*, which causes already difficulties in estimating the joint density for two dimensional problems and makes the estimation of issues with more than two dimensions virtually impossible. Copula models avoid these problems. They provide a nice way to analyse the univariate marginal

[1] This means the marginal distributions are changed.

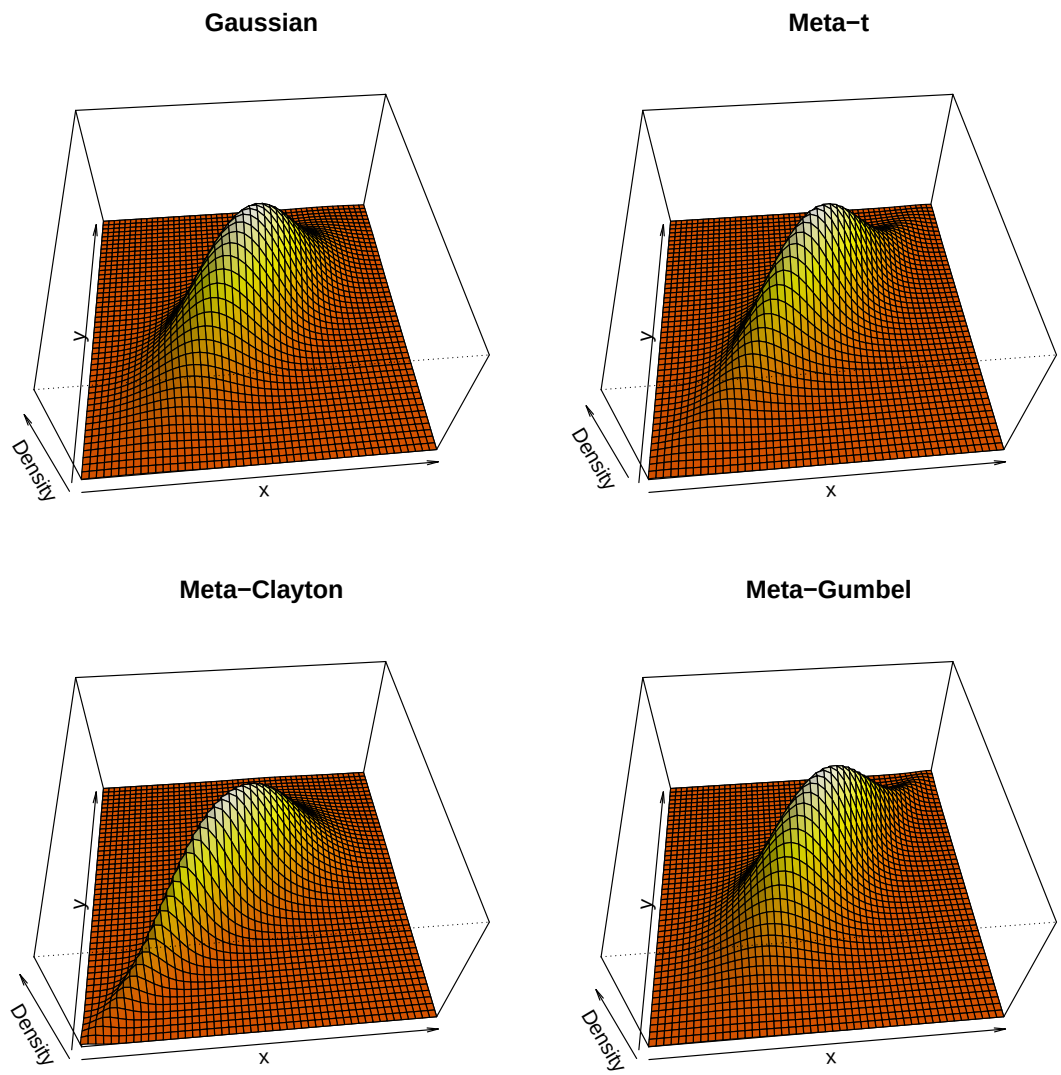

Figure 5.2: The joint densities of the corresponding meta-distributions.

distributions and the dependence structure separately. It is probably therefore why they have become so popular in recent years. According to Mikosch (2006) the number of *Google* hits for the word *copula* jumped from 10000 to 650000 between the years 2003 and 2005[2]. In 2006 Genest, Gendron, and Bourdeau-Brien (2009) surveyed the academic literature on copulas by using 26 bibliographic databases. They found that before 1986 the literature on copulas was very sparse and mostly mathematical. Between 1986 and 1999 they could observe a slow but systematic rise in the number of publications, which was largely due to the emergence of the concept in statistical applications. From the year 1999 on, however, the number of contributions increased largely. In this context it is certainly worth mentioning the influential contributions on the theory of copulas and dependence in general at that time by Joe (1997), Nelsen (1999), Drouet-Mari and Kotz (2001), Frees and Valdez (1998) as well as Embrechts, McNeil, and Straumann (1999). Especially the last two mentioned publications fuelled the actuarial and financial applications of copulas

[2]Of course, the term *copula* also appears in non-scientific contexts.

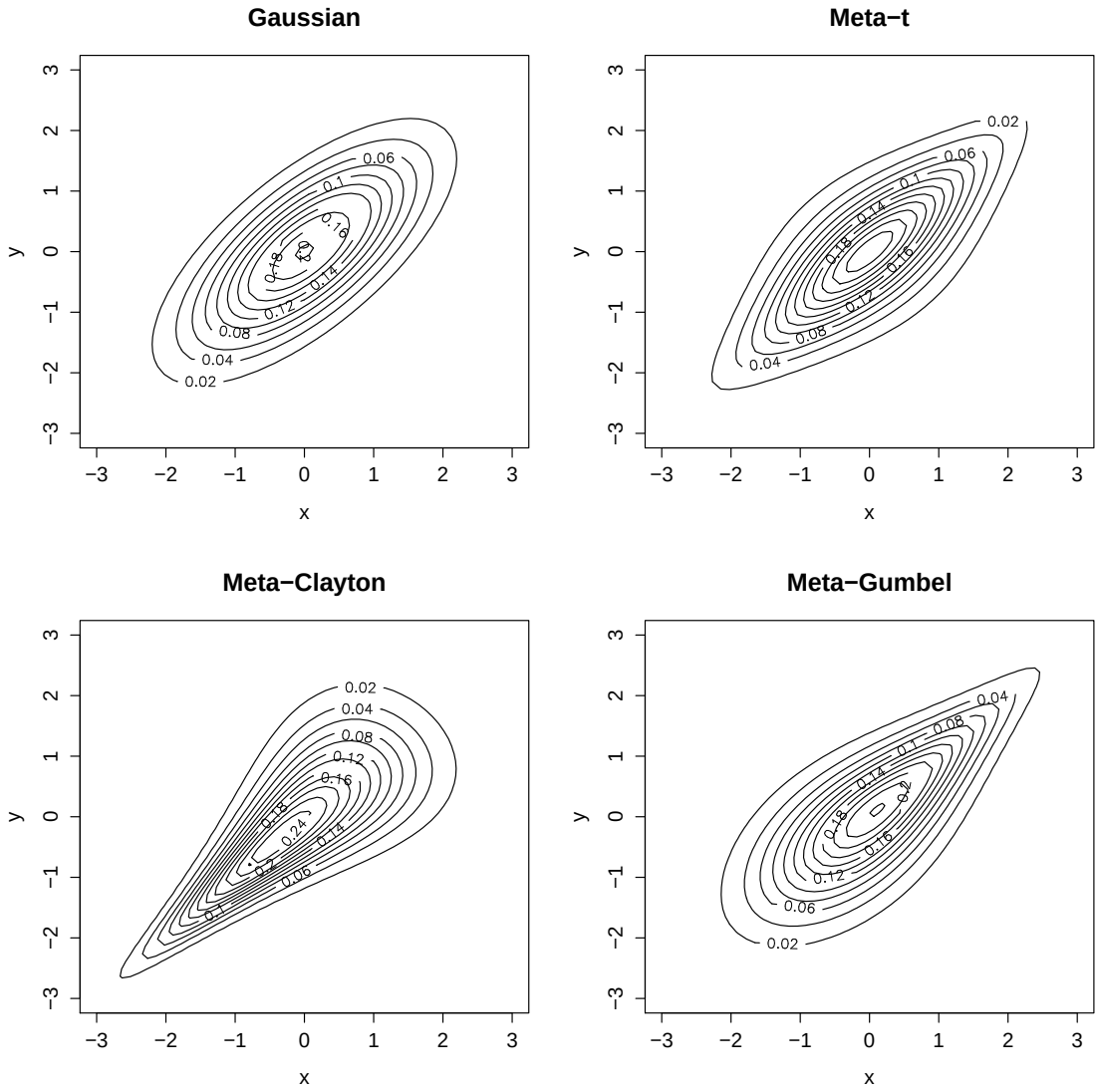

Figure 5.3: The contour plots of the joint densities.

and showed its potential there. It is also noteworthy that even though people in finance have been interested in copulas basically only since the year 2000, they provided with approximately 41% of the documents screened by Genest et al. (2009) by far the largest portion.

5.3.1 Definition and Basic Properties of Copulas

Every joint distribution function of random variables implicitly contains a description of the marginal behaviour of each variable and a description of their dependence structure. From a statistical point of view, there are well known and powerful methods in order to obtain a decent description of univariate distributions and, thus, also of the univariate marginal distributions of a multidimensional problem. The idea of the copula is now to address the rest that is needed to describe the joint distribution function, i.e. the

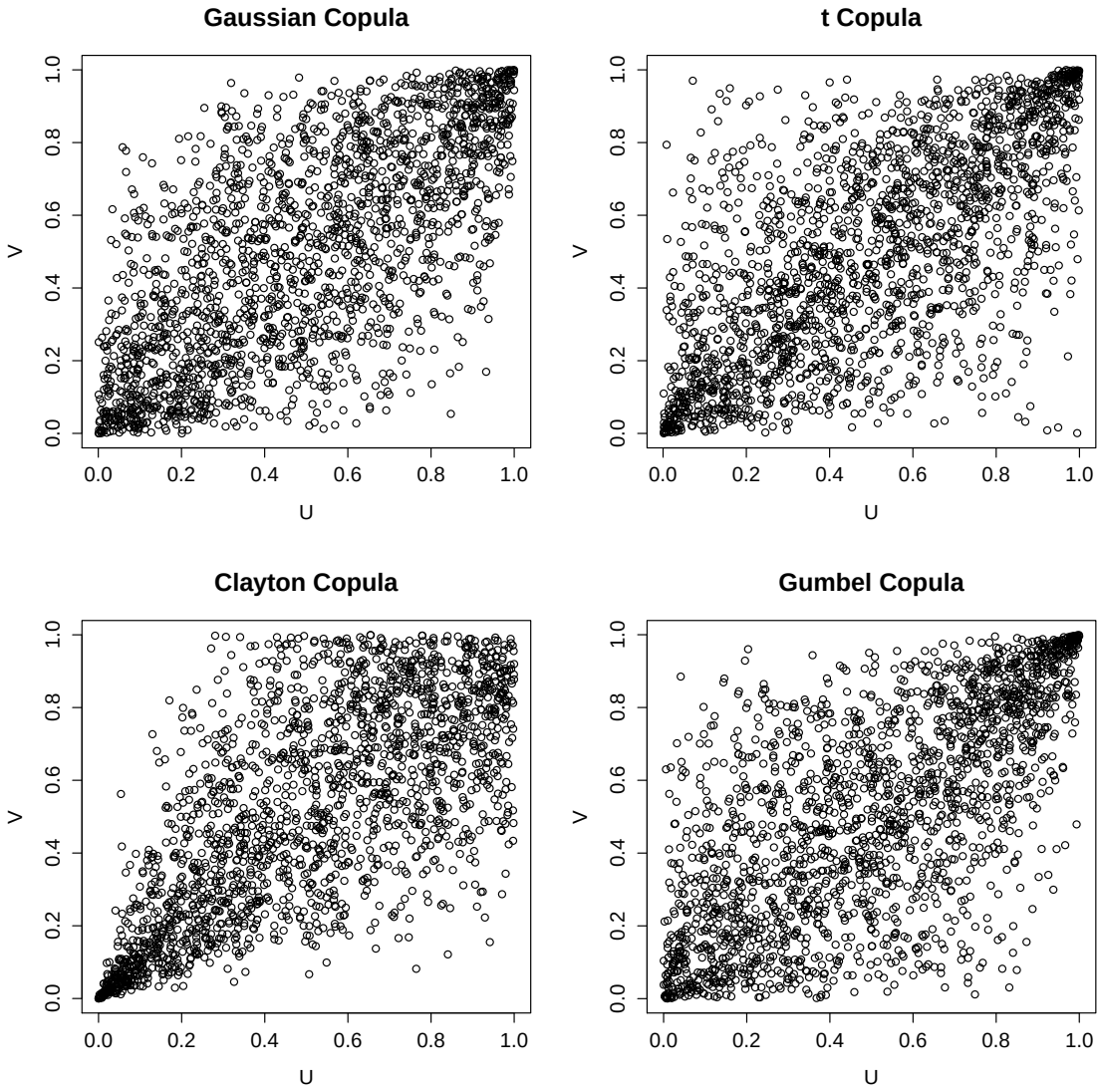

Figure 5.4: The copulas.

dependence structure. So, a copula is defined as follows:

Definition 5.3.1. *A continuous function* $C : [0,1]^d \to [0,1]$ *is a copula if and only if*

i.) $C(u_1, \ldots, u_d)$ *is increasing in each component* u_i.

ii.) $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$

iii.) $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0$

iv.) For all $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0,1]^d$ *with* $a_i \leq b_i$ *we have*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\ldots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0$$

where $u_{j1} = a_j$ *and* $u_{j2} = b_j$ *for all* $j \in \{1, \ldots, d\}$.

Hence, copulas are multivariate distribution functions with uniformly distributed marginals. In order to interpret the four properties as given in Definition 5.3.1 it makes sense to introduce *Sklar's Theorem* first, which constitutes one of the most relevant results in the whole copula methodology.

Theorem 5.3.2. *(Sklar, 1959) Let $X_1, \ldots, X_d$ be random variables with distribution functions $F_1, \ldots, F_d$ and joint distribution function F. Then there exists at least one copula C, such that for all $\mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)) \quad . \tag{5.3.1}$$

In particular, if the univariate distribution functions $F_1, \ldots, F_d$ are continuous, then the copula C is unique. Equation (5.3.1) can also be written for $\mathbf{u} = (u_1, \ldots, u_d)^T \in [0,1]^d$ as

$$C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right) \quad . \tag{5.3.2}$$

Theorem 5.3.2 expresses the basic idea of dependence modelling via copula functions by stating exactly what was just mentioned at the beginning of this section, i.e. that for any multivariate distribution function the univariate marginal distribution functions and the dependence structure can be separated, with the later being completely described by a copula function. Equation 5.3.1 allows, thus, a representation of the copula function in terms of a multivariate distribution function (with continuous marginal distributions). This allows now to further re-establish the properties stated in Definition 5.3.1. The first property is clearly required for any multivariate distribution. The second and third properties follow from the fact that if the random variables $X_1, X_2, \ldots$ have continuous distribution functions $F_1, F_2, \ldots$, then the random variables $U_1 = F_1(X_1), U_2 = F_2(X_2), \ldots$ are uniformly distributed on $[0,1]$. The forth property is a bit less obvious. However, it ensures that the copula function respects the defining characteristics of a proper multivariate distribution function, assigning non-negative weights to all rectangular subsets in $[0,1]^n$, and thus, returning always non-negative probabilities.

With Theorem 5.3.2 in mind, we can now also easily derive the multivariate copula density function, which is needed if we want to estimate the copula parameters by the maximum likelihood method. Using the chain rule, we obtain

$$\begin{aligned} f(x_1, \ldots, x_n) &= \frac{\partial^n C(F_1(x_1), \ldots, F_n(x_n))}{\partial F_1(x_1) \ldots \partial F_n(x_n)} \prod_{i=1}^{n} f_i(x_i) \\ &= c(F_1(x_1), \ldots, F_n(x_n)) \prod_{i=1}^{n} f_i(x_i) \end{aligned} \tag{5.3.3}$$

or equivalently

$$c(F_1(x_1), \ldots, F_n(x_n)) = \frac{f(x_1, \ldots, x_n)}{\prod_{i=1}^{n} f_i(x_i)} \quad . \tag{5.3.4}$$

5.3.2 Some Commonly Used Copulas

In general, every function C that fulfils the four properties as stated in Definition 5.3.1 is a copula. This allows us to explicitly formulate copulas. In addition, using Theorem 5.3.2 we can extract the copula of any known multivariate distribution function. Accordingly, there are many copula functions with very different properties. This is probably the

most challenging part concerning dependence modelling with copulas. Coming up with an appropriate copula function for the problem at hand is, especially in high-dimensional cases, not so trivial. As empirical results with regard to hedge fund returns emphasise to monitor especially the behaviour in the tails, we limit ourselves to four well known showcase copulas that show different kinds of tail dependence, i.e. the tendency to generate simultaneous extreme values. These are the *Gaussian*, the *t*, the *Clayton* and the *Gumbel* copula. We have already used them in the example of section 5.2. Employing Theorem 5.3.2, the *Gaussian* copula is extracted from the multivariate normal or Gaussian distribution. Similarly, the *t* copula is obtained from the multivariate *t* distribution.

The Gaussian Copula: With $\mathbf{\Phi}_R$ standing for the multivariate standard normal distribution and R being a symmetric, positive definite matrix with only ones on the diagonal and components denoted by ϱ_{ij}, the d-dimensional Gaussian copula is given by

$$C_R^{Ga}(\mathbf{u}) = \mathbf{\Phi}_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right) = \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{(2\pi)^{\frac{n}{2}} |R|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{x}^T R^{-1}\mathbf{x}\right) dx_d \ldots dx_1 \tag{5.3.5}$$

with $\mathbf{u} = (u_1, \ldots, u_d)^T \in [0,1]^d$ and $\Phi^{-1}(u)$ denoting the quantile function of the univariate standard normal distribution.

The *t* Copula: As in the case of the *Gaussian* copula, we also need a dispersion matrix R, which is again positive definite with only ones on the diagonal. In addition, defining $\mathbf{t}_{\nu,R}$ to be the multivariate standard t distribution with corresponding degrees of freedom ν, the t copula is given by

$$C_{\nu,R}^{t}(\mathbf{u}) = \mathbf{t}_{\nu,R}\left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\right) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_d)} \frac{\Gamma\left(\frac{\nu+d}{2}\right)|R|^{-\frac{1}{2}}}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{\frac{d}{2}}} \left(1 + \frac{1}{\nu}\mathbf{x}^T R^{-1}\mathbf{x}\right)^{-\frac{\nu+d}{2}} dx_d \ldots dx_1 \tag{5.3.6}$$

with t_ν^{-1} denoting the quantile function of the univariate standard t distribution and $\Gamma(\cdot)$ standing for the gamma function.

As the *Gaussian* and the *t* copula are extracted form elliptical distributions[3], they are said to belong to the family of elliptical copulas. On the other hand, the *Clayton* and the *Gumbel* copula are said to belong to the family of so-called *Archimedean* copulas. Their construction follow general mathematical rules. Genest and MacKay (1986) define the *Archimedean* type of copulas as follows:

$$C(\mathbf{u}) = \begin{cases} \varphi^{-1}\left(\sum_{i=1}^d \varphi(u_i)\right) & \text{, if } \sum_{i=1}^d \varphi(u_i) \leq \varphi(0) \\ 0 & \text{, otherwise} \end{cases} \tag{5.3.7}$$

with $\mathbf{u} = (u_1, \ldots, u_d)^T \in [0,1]^d$ and φ being the so-called copula generator. For more information on the construction procedure, the interested reader is referred to Nelsen (1999) and Joe (1997) where the properties and requirements of generator functions, and *Archimedean* copulas in general, are discussed in detail. Here, we only consider the *Clayton* and the *Gumbel* copula.

[3] This is because of the fact that points with equal density lie on ellipsoids.

The Clayton Copula: Again, with $\mathbf{u} = (u_1, \ldots, u_d)^T \in [0,1]^d$ and $\theta \geq 0$ the *Clayton* copula is given as follows[4]:

$$C_\theta^{Cl}(\mathbf{u}) = \left(\sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right)^{-\frac{1}{\theta}} \tag{5.3.8}$$

with generator function $\varphi(u) = \left(u^{-\theta} - 1\right)/\theta$.

The Gumbel Copula: As with the *Clayton* but with the parameter $\theta \geq 1$, the *Gumbel* copula is obtained by

$$C_\theta^{Gu}(\mathbf{u}) = \exp\left(- \left[\sum_{i=1}^{d} (-\ln u_i)^\theta \right]^{\frac{1}{\theta}} \right) \tag{5.3.9}$$

with generator function $\varphi(u) = (-\ln u)^\theta$.

In order to get an approximate idea of what happens when the parameter θ of the *Archimedean* copulas or the degrees of freedom ν of the t copula is altered, Figure 5.5, 5.6 and 5.7 show the density contour plots of the corresponding bivariate meta distributions with standard normal margins. As can be seen in Figure 5.5, by increasing the degrees of freedom of the meta-t distribution, the distribution converges toward the *Gaussian* distribution with the same dispersion matrix R, as it is depicted on the upper left in Figure 5.3. For finite values of ν, however, the distribution shows both upper as well as lower tail dependence[5]. This holds true in general. A t copula can, thus, be used to model tail dependence on both tails. In contrast to the t copula, the *Clayton* and the *Gumbel* copula have only one parameter to determine. There, the dependency between the variables increase as the parameter θ is increased. For both, the range of dependency reaches from independence to perfect positive dependence. They show, however, quite a different behaviour with respect to tails. The *Clayton* copula models lower tail dependence whereas the *Gumbel* copula models upper tail dependence.

5.3.3 Fitting Copulas to Data

Depending on the problem at hand, there are different ways to estimate the corresponding parameters, such as the *method of moments*, the *exact maximum likelihood*, the *inference functions for margins* (IFM) or the *canonical maximum likelihood* (CML).

Exact Maximum Likelihood: Using equation (5.3.3) with continuous marginal distributions $f_i(x_{t,i}; \theta_i)$, the log-likelihood function to estimate all parameters at once is given by

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \ln c\left(F_1(x_{t,1}; \theta_1), \ldots, F_d(x_{t,d}; \theta_d); \boldsymbol{\theta}_c\right) + \sum_{t=1}^{T} \sum_{i=1}^{d} \ln f_i(x_{t,i}; \theta_i) \tag{5.3.10}$$

[4] In the special case of $d = 2$ the parameter θ is defined on a slightly broader range.

[5] By contrast, the Gaussian copula shows no tail dependence at all.

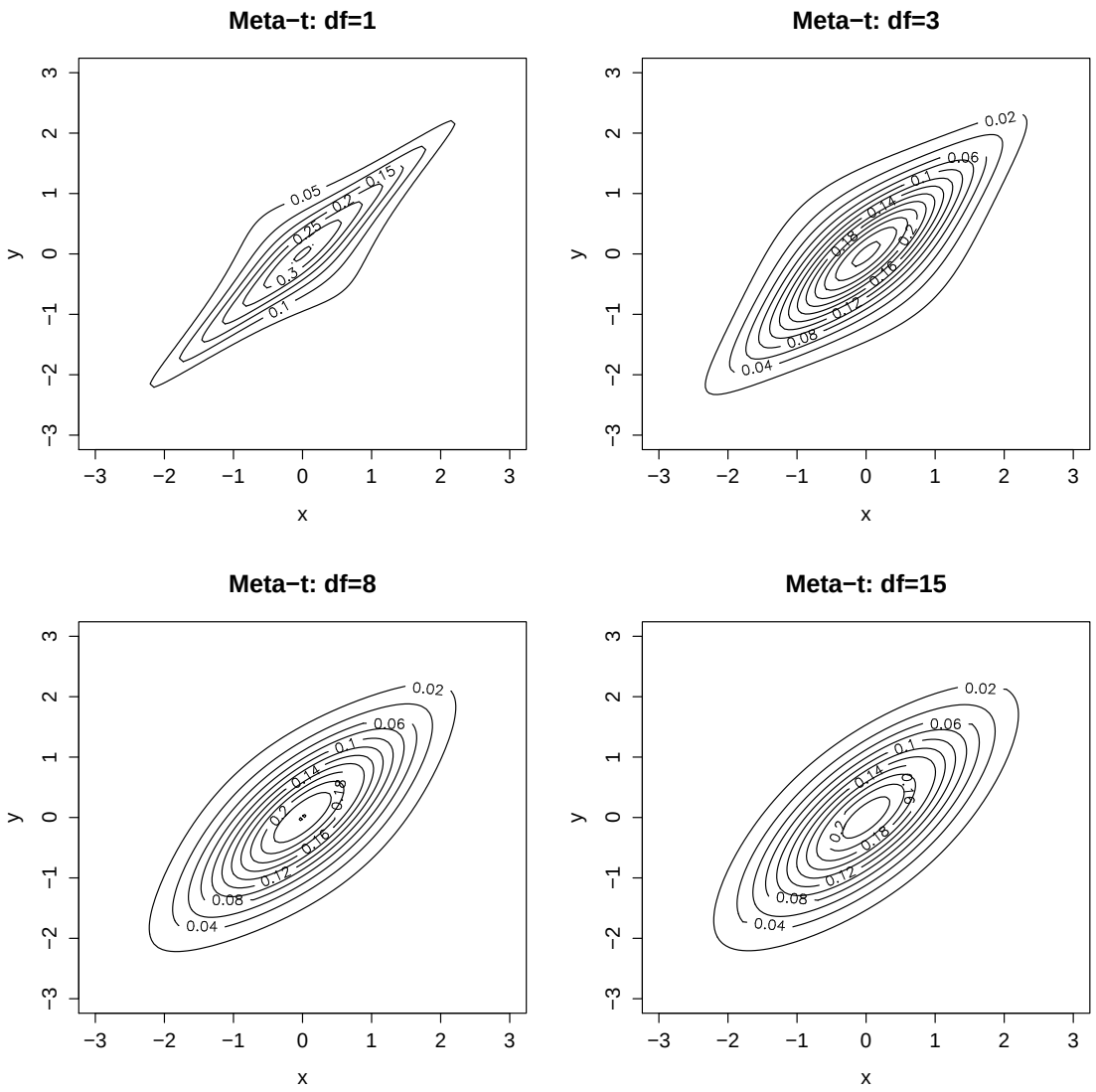

Figure 5.5: The meta-t density for different values of ν.

where $\boldsymbol{\theta}_c$ stands for the parameter vector of the copula. This approach is feasible if we have not too many parameters to estimate and there is an appropriate amount of data points available. From a practical point of view, there is the important question of how to choose the appropriate copula and the most suitable marginal distributions. As this usually means to analyse the margins and the dependence structure separately, it is in most cases convenient to pursue also with respect to the parameter estimation a two-step procedure like the *method of moments*, the *inference functions for margins* (IFM) or the *canonical maximum likelihood* (CML).

Method of Moments: In this approach we estimate the parameters of the marginal distributions and the copula parameters separably. The method of moments estimation is usually more efficient then the maximum likelihood based methods when there are many parameters to be estimated. Clearly, there are issues where this method is not applicable, or just too cumbersome. Sometimes it may be useful to estimate some of the overall

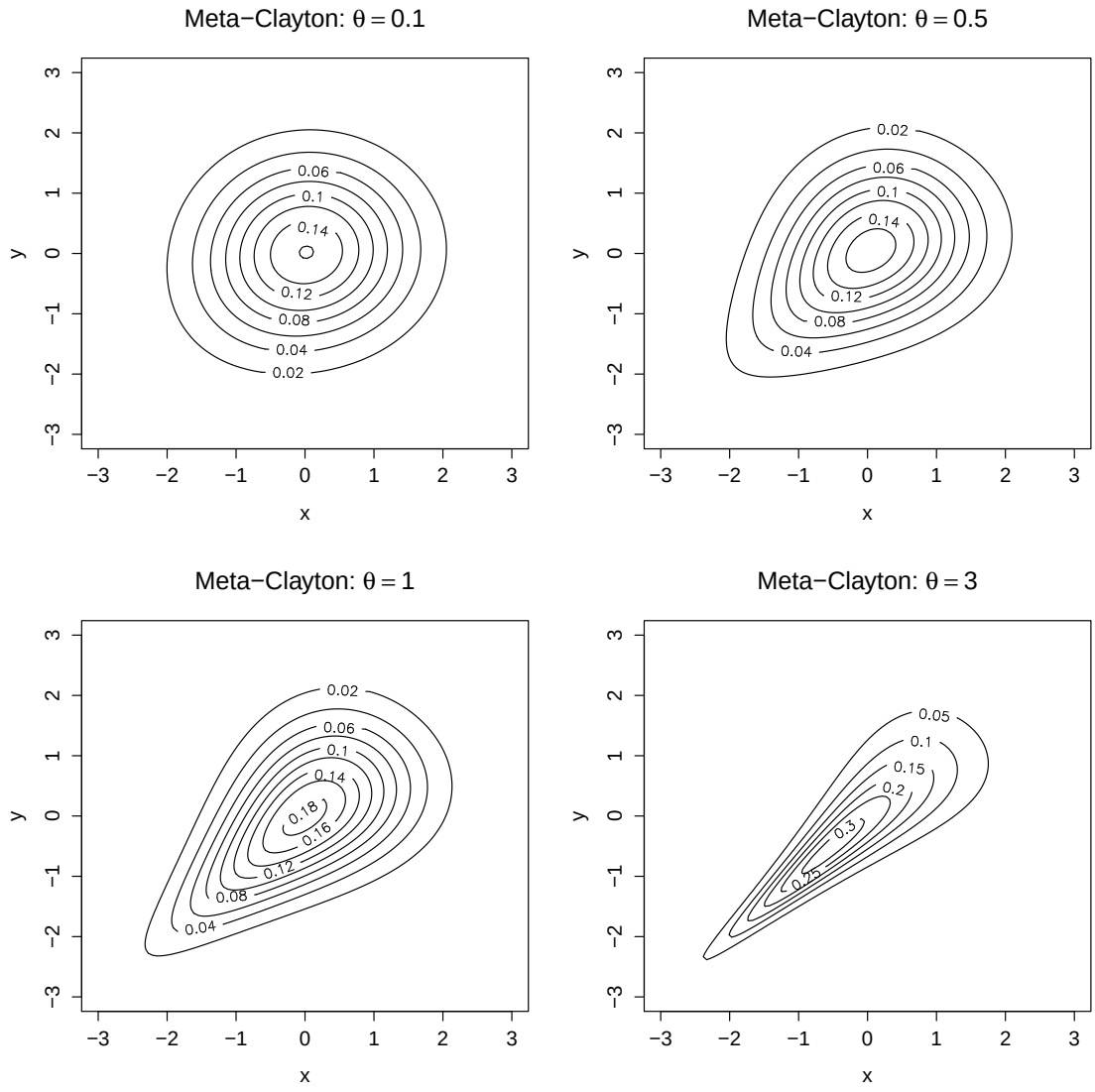

Figure 5.6: The meta-Clayton density for different values of θ.

parameters by the method of moments and then using the maximum likelihood method in a second stage in order to estimate the remaining parameters. Important tools here are *Spearman's* rho, which can be written as

$$\rho_S(X,Y) = \rho\left(F_X(X), F_Y(Y)\right) \tag{5.3.11}$$

and is, thus, just the *Bravais-Pearson* correlation on the (truly) transformed input data, and *Kendall's* tau, which is given by

$$\tau(X,Y) = P\left(\left(X-\tilde{X}\right)\left(Y-\tilde{Y}\right)>0\right) - P\left(\left(X-\tilde{X}\right)\left(Y-\tilde{Y}\right)<0\right) \tag{5.3.12}$$

where the probabilities P are usually estimated by calculating the relative values of the concording pairs and discording pairs, respectively. It then remains to calculate the corresponding parameter θ by solving a usually simple functional relationship of the form $\rho_S(\mathbf{X}) = f(\theta)$ or $\tau(\mathbf{X}) = f(\theta)$, respectively. For example, in the bivariate case with a

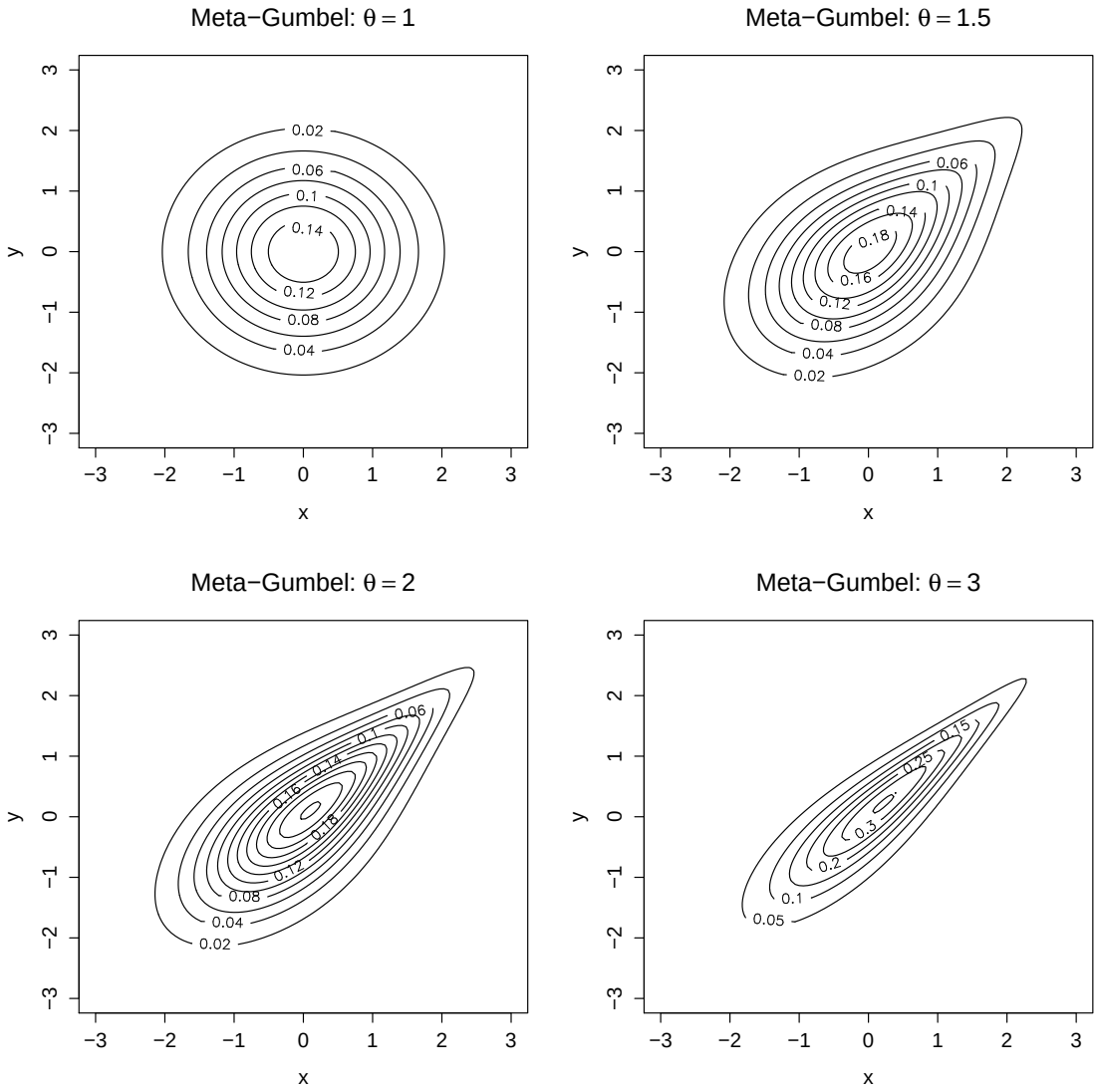

Figure 5.7: The meta-Gumbel density for different values of θ.

Gaussian copula C_R^{Ga}, the relationships are given as follows:

$$\rho_S(X,Y) = \frac{6 \arcsin \frac{\varrho}{2}}{\pi} \tag{5.3.13}$$

$$\tau(X,Y) = \frac{2 \arcsin \varrho}{\pi} \tag{5.3.14}$$

Relation (5.3.14) holds more generally for the copulas of essentially all normal variance mixture distributions, such as the t copula. In the case of two dimensional *Archimedean* copulas with generator φ, the relationship between the parameter θ and *Kendall's* τ is

$$\tau(X,Y) = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} du \ . \tag{5.3.15}$$

For a more detailed discussion on this topic, also with respect to other copulas, and the corresponding proofs, see McNeil et al. (2005) and Nelsen (1999).

The Inference Functions for Margins (IFM): In contrast to the exact maximum likelihood estimation, the IFM proposed by Joe and Xu (1996) exploits the fundamental idea of the copula theory by separating the estimation of the univariate margins and the dependence structure. It is, thus, a two-stage procedure. This is, first the parameters of the univariate margins $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ are estimated by maximum likelihood

$$\widehat{\theta}_i = \arg\max_{\theta_t} \sum_{t=1}^{T} \ln f_i\left(x_{t,i}; \theta_i\right) \ . \tag{5.3.16}$$

Then the copula parameters $\boldsymbol{\theta}_c$ are estimated in a second stage, also by maximum likelihood, i.e.

$$\widehat{\boldsymbol{\theta}}_c^{IFM} = \arg\max_{\boldsymbol{\theta}_c} \sum_{t=1}^{T} \ln c\left(F_1\left(x_{t,1}; \widehat{\theta}_1\right), \ldots, F_d\left(x_{t,d}; \widehat{\theta}_d\right); \boldsymbol{\theta}_c\right) \ . \tag{5.3.17}$$

The Canonical Maximum Likelihood (CML): The IFM method assumes a parametric form of the univariate marginals. The CML[6] method, by contrast, does not imply any a priori assumptions on the distributional form of the marginals but uses the empirical cumulative distribution functions $\widehat{F}_i\left(x_i\right)$ to transform the data from $\mathbb{R}$ to $[0, 1]$. This is

$$\widehat{F}_i\left(x_i\right) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\{X_{t,i} \leq x_i\}} \ , \text{ for } i = 1, \ldots, d \tag{5.3.18}$$

with $\mathbb{I}_{\{X_{t,i} \leq x_i\}}$ denoting the characteristic function. With regard to the fitting of the copula, it is, however, advisable to use $T + 1$ instead of T in the denominator of equation (5.3.18) as this ensures that the transformed data points lie strictly in the interior of the unit cube.

5.4 Application to the Fund Data

With the discussed methods we can now analyse our return data in detail. Firstly, we have a look at the conventional performance measurements as the *Sharpe ratio* and the CAPM. Secondly, we model the bivariate dependence structure by a parametric, though very flexible approach.

5.4.1 Fund Performance

Table 5.2 summarises the results of the fund performance measurements. Where meaningful, we compare the results of the funds with them of the S&P 500 index. The *Sharpe ratios* of all four fund types are quite impressive. They show in deed a very favourable risk-return profile. Compared to the ratio of the S&P 500 index, the ratios of the funds are all more than six times higher.

In order to apply the CAPM according to its theory, one would actually need to have the returns of the overall market, i.e. the market that consists of any kind of marketable assets.

[6] McNeil et al. (2005) as well as Genest and Favre (2007) refer to this method as *Maximum Pseudolikelihood Estimator*.

	LPB 20	LPB 50	PLB 20	PLB 50	S&P 500
Sharpe Ratio	0.6291	0.6229	0.6989	0.8014	0.1040
$\widehat{\alpha}$	1.72e-4	1.22e-4	2.11e-4	1.67e-4	0.0000
t_α	2.5612	2.5015	2.8386	3.1916	-
$\widehat{\beta}$	0.0218	0.0261	0.0230	0.0375	1.0000
t_β	4.2132	6.9208	4.0124	9.2998	-
$\widehat{\rho}$	0.0654	0.1070	0.0624	0.1432	1.0000

Table 5.2: Performance summary.

This is however only a theoretical quantity, as it is just impossible to capture every such asset of the world. So, one usually takes the returns of a large index as a substitute of the theoretical market returns. Clearly, the choice of the substitute depends on the feasible investment universe of an individual investor. If we reduce this investment universe to very liquid US stocks, then the S&P 500 index may qualify very well as market substitute. This makes sense as our statistical arbitrage funds deal exclusively with stocks that are part of the S&P 500 index. However, with the results of the *Sharpe ratio* in mind, we can already say that an estimation of the CAPM will suggest that our funds show positive *absolute returns*. Adding a constant parameter α to equation (5.1.2) and estimating this simple linear model for each fund by *ordinary least square* (OLS) gives the results as reported in Table 5.2. As can be seen, the estimated β values are small in absolute terms, though still significantly different from zero. Also the estimated α values are significantly different from zero, which is basically what we expected. Hence, according to the CAPM, only a small part of the fund's excess returns can be explained by the excess returns of the market. A significant part of the daily fund returns are linearly independent of the S&P 500 index returns. Clearly, as the β values consist of the corresponding linear correlations, the effect is also visible in the estimates of the linear correlation coefficients ρ between the funds and the S&P 500 index returns.

However, as illustrated in section 5.2, any measure based on linear correlation may be misleading as the dependence structure may be more complex than that captured by a linear dependence measure. In order to study the general dependence structure between the returns of our statistical arbitrage funds and the S&P 500 index returns we now apply the methods we introduced in section 5.3.

5.4.2 General Dependence Structure

Starting with the margins, we have already seen from the results reported in section 4.4 that the normal distribution does not suit really well in order to describe the return distributions of the funds and the S&P 500 index. We have seen that all the return data show considerable excess kurtosis. The main problem with a distribution of the *Gaussian* type lies, thus, in its tails. Using the empirical distribution function $\widehat{F}_i(x_i)$, as proposed by the CML approach, does not really help to cope with this problem as empirical distribution functions are known to be poor estimators of the tails of an underlying distribution. Clearly, we can employ a parametric type of distribution that is known to be able to capture heavy tails. The t-distribution, for example, would be a candidate. A very flexible parametric distribution type, which contains the t distribution as a special case, is the *generalised hyperbolic* (GH) distribution as discussed by Barndorff-Nielsen (1977). It can

be represented as a *normal mean-variance mixture*. More precisely, a random variable X is said to have a generalised hyperbolic distribution if

$$X \stackrel{\mathrm{d}}{=} \mu + W\gamma + \sqrt{W}\sigma Z \tag{5.4.1}$$

where $\mu \in \mathbb{R}$, $\gamma \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$, $Z \sim N(0,1)$ and $W \sim N^-(\lambda, \chi, \psi)$. This is, the mixing variable W has a *generalised inverse Gaussian* (GIG) distribution with density

$$f(w) = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{w^{\lambda-1}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{w} + \psi w\right)\right\} \ , \ w > 0 \ , \tag{5.4.2}$$

and the parameters satisfying $\chi > 0$, $\psi \geq 0$ if $\lambda < 0$; $\chi > 0$, $\psi > 0$ if $\lambda = 0$; and $\chi \geq 0$, $\psi > 0$ if $\lambda > 0$. K_λ is a modified *Bessel* function of the third kind with integral representation

$$K_\lambda(x) = \int_0^\infty v^{\lambda-1} \exp\left\{-\frac{1}{2}x\left(v + \frac{1}{v}\right)\right\} dv \ , \ x > 0 \ .$$

The density function of a univariate GH distribution can, thus, be written as

$$f(x) = c\frac{K_{\lambda-\frac{1}{2}}\left(\sqrt{\left(\chi + (x-\mu)^2\sigma^{-2}\right)\left(\psi + \gamma^2\sigma^{-2}\right)}\right) \exp\left\{(x-\mu)\gamma\sigma^{-2}\right\}}{\left(\sqrt{\left(\chi + (x-\mu)^2\sigma^{-2}\right)\left(\psi + \gamma^2\sigma^{-2}\right)}\right)^{\frac{1}{2}-\lambda}} \tag{5.4.3}$$

with normalising constant

$$c = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{\left(\psi + \gamma^2\sigma^{-2}\right)^{\frac{1}{2}-\lambda}}{\sqrt{2\pi}\sigma K_\lambda\left(\sqrt{\chi\psi}\right)} \ .$$

From the *normal mean-variance mixture* representation (5.4.1) one can easily see that μ is a location parameter and σ a dispersion parameter. With the parameter γ we control the skewness of the distribution. If $\gamma = 0$ then the distribution is symmetric. The role of the remaining parameters λ, χ and ψ is less obvious. They control the general shape of the distribution, and thus, also the tails. For example, if $\lambda = -\frac{\nu}{2}$, $\chi = \nu$, $\psi = 0$ and $\gamma = 0$ we obtain a t distribution with ν degrees of freedom. If $\lambda = -\frac{1}{2}$, then we get a distribution that is known as *normal inverse Gaussian* (NIG). Finally, if $\lambda = 1$, then we just have a *hyperbolic* distribution. Some illustrative density plots for different parametrisations of the univariate GH as well as some more analytical information on the GH and the GIG can be found in appendix A.

One last remark concerns the parametrisation of the GH. The parametrisation as chosen here is very natural from a *normal mean-variance mixture* point of view. Unfortunately, we have an identifiability problem if we choose the parametrisation in this way as $\mathrm{GH}(\lambda, \chi, \psi, \mu, \sigma^2, \gamma)$ and $\mathrm{GH}(\lambda, \chi k^{-1}, \psi k, \mu, \sigma^2 k, \gamma k)$ are identical for $k > 0$. This problem can, however, be solved by introducing a suitable constraint. For instance, setting the dispersion parameter σ equal to one or requiring the expectation of the mixing variable W to be one. The parameters can then be estimated by the usual maximum likelihood method. Finally, the chosen parametrisation is not the only possible one. In the literature on GH distributions there are at least two other very popular parametrisations.

Table 5.3 summarises the estimated parameters of the fitted GH distributions for the returns of the four funds and the S&P 500 index. None of them show one of the previously

	LPB 20	LPB 50	PLB 20	PLB 50	S&P 500
$\widehat{\lambda}$	0.4678	0.1782	1.0467	-2.0845	0.1207
$\widehat{\chi}$	0.0868	0.2039	0.0000	2.2963	0.1805
$\widehat{\psi}$	1.2870	1.0458	2.0935	0.0428	0.9202
$\widehat{\mu}$	0.0000	0.0001	0.0001	0.0000	0.0009
$\widehat{\sigma}$	0.0042	0.0031	0.0048	0.0034	0.0128
$\widehat{\gamma}$	0.0002	0.0002	0.0003	0.0003	-0.0006

Table 5.3: Estimated parameters of the generalised hyperbolic distribution.

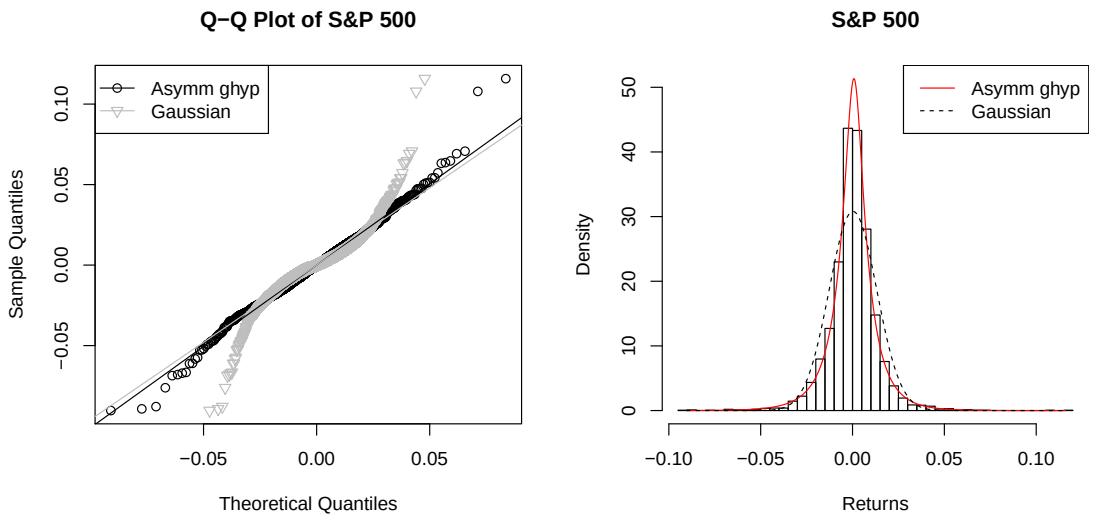

Figure 5.8: The Q-Q plot and histogram of the S&P 500 index.

discussed "special" cases. The returns of the funds are slightly skewed to the right (i.e. positively skewed), while the S&P 500 returns are slightly negatively skewed. The goodness of fit with respect to the distribution tails is probably best inspected by a quantile-quantile plot as depicted in the Figures 5.8, 5.9 and 5.10. Besides the GH quantiles we also plot the quantiles of a normal distribution for comparison. The plot on the right of each q-q plot depicts the scaled histogram of the returns and the corresponding GH and normal densities. Concerning the S&P 500 returns the fitted GH distribution suits quite well apart from two observations on the upper tail. The returns of the LPB funds are a bit less well captured by the estimated GH distribution. Here, the upper as well as the lower tail of the distribution are not quite appropriate. But as it is well stressed by comparing the empirical quantiles of these returns with the quantiles of a Gaussian distribution, we see that these returns here are especially heavy tailed. The returns of the PLB funds correspond again slightly better to the fitted distributions. Especially the upper tail is captured quite well. The PLB 50 fund shows one rather extreme observation on the lower tail that is quite far off.

If we now employ these univariate distributions in order to transform the returns into "pseudo" observations in the $[0, 1]$ space, the scatter plots of the transformed data may be used to make a decision with regard to the dependence structure, and thus, the appropriate copula to be employed. Figure 5.11 shows the scatter plots of the transformed data. At

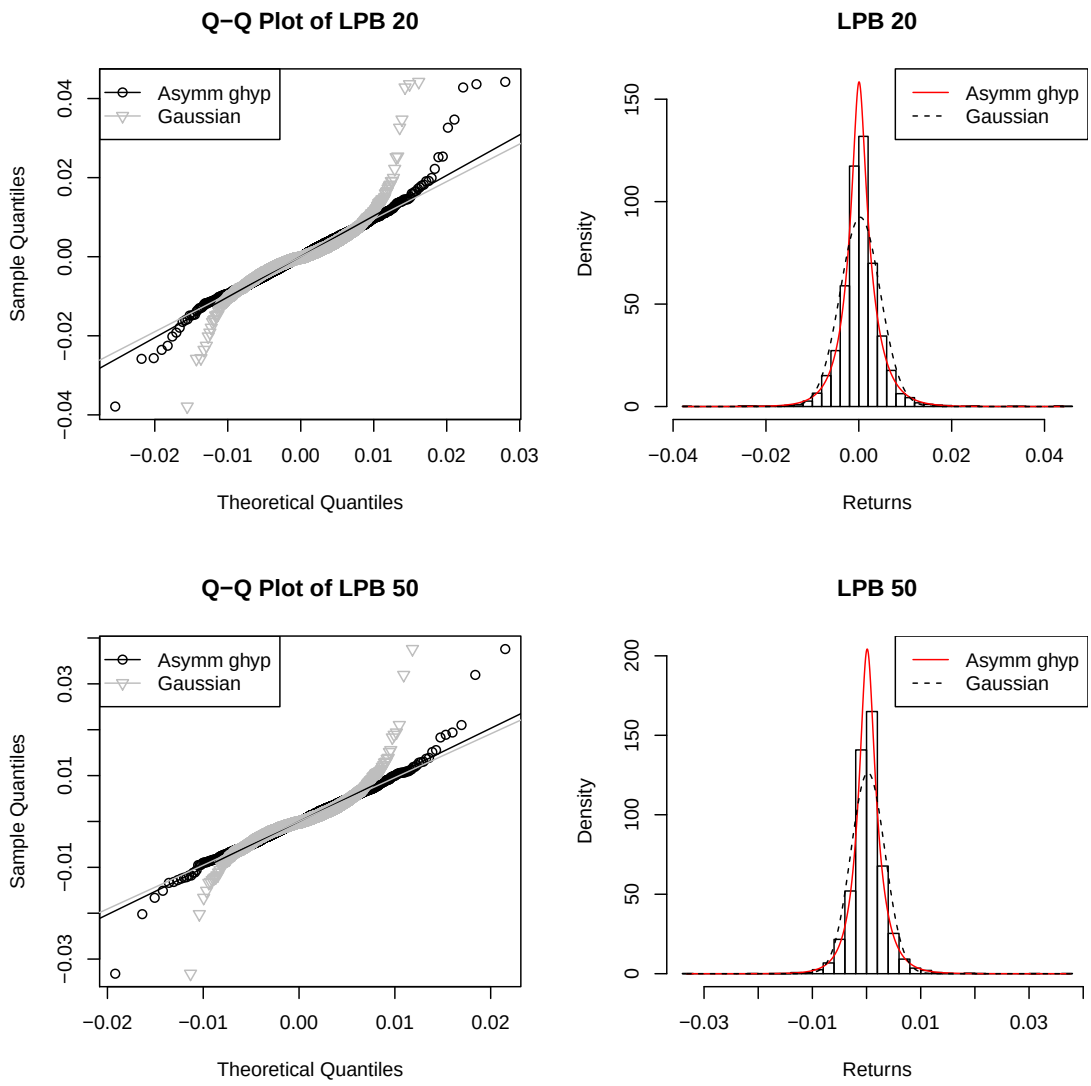

Figure 5.9: The Q-Q plots and histograms of the LPB funds.

first sight it seems difficult to see any dependence structure at all. From these pictures it seems, however, reasonable to conclude that the choice has to be made between the Gaussian and the t copula. As we have already seen, the correlations between the fund returns and the returns of the S&P 500 index are close to zero. This influences, of course, the parameters of the copulas. In order to contrast this picture a bit, we can have a look at a stylised example. Figure 5.12 depicts some simulated data of a Gaussian and a t copula, both with the same dispersion parameter $\varrho = 0.05$. Now, as the t copula goes to a Gaussian copula with increasing parameter ν, we can stress the difference between the two copulas by choosing this parameter to be a low positive number, such as $\nu = 2$, for example. As can be seen, there is clearly some difference in the pattern of the two copulas. In the case of the t copula we see that the distribution of the points is a bit more sparse on the margins between the four corners. Such a pattern is, at least in tendency, also visible in the plots in Figure 5.11. This obviously argues for the t copula. In addition, as the t copula goes to a Gaussian copula with increasing ν, and we have well enough data

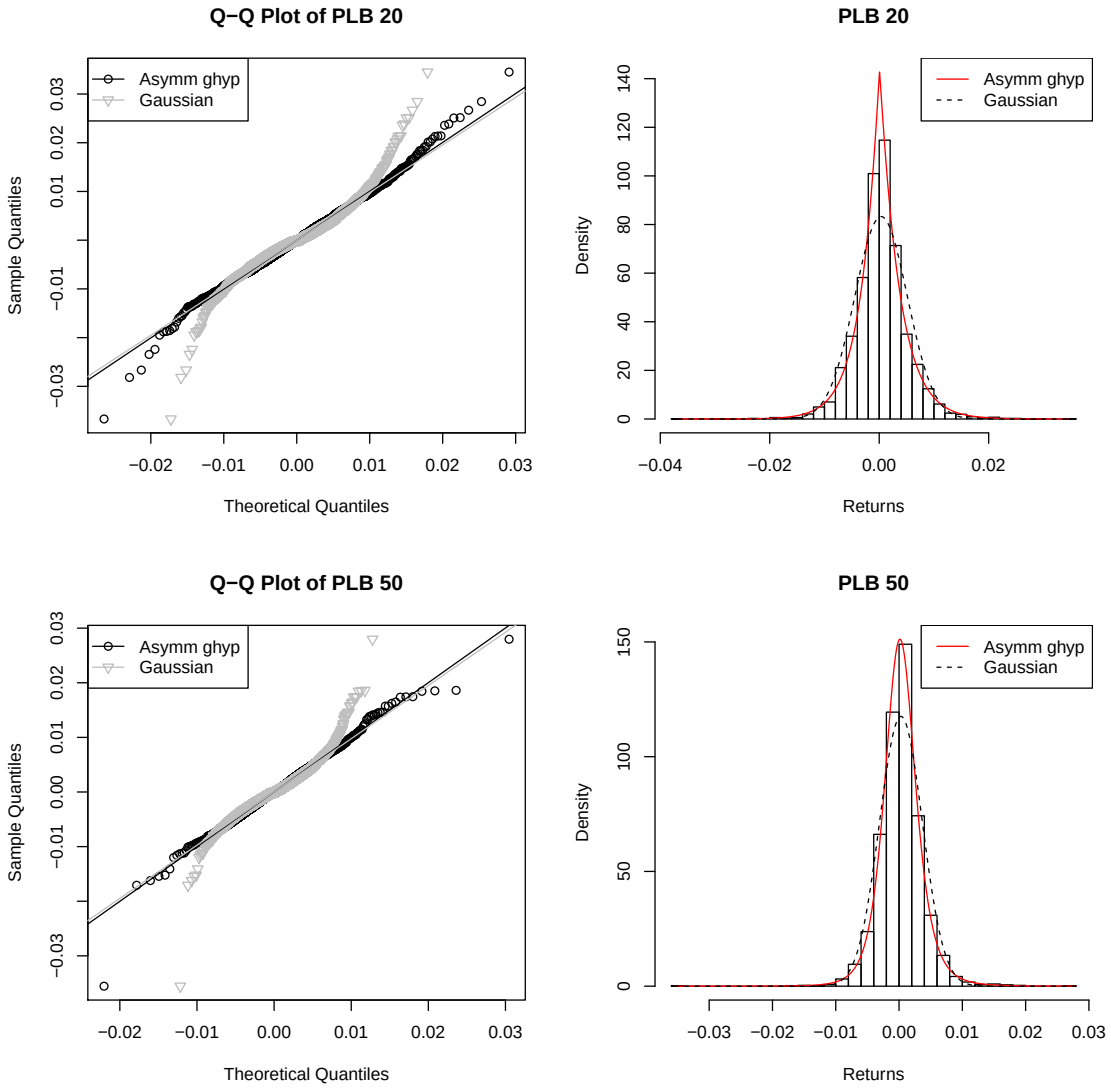

Figure 5.10: The Q-Q plots and histograms of the PLB funds.

points concerning the parameter estimation, choosing the t copula is clearly a reasonable decision.

When we now fit the t copula to our data by employing the maximum likelihood procedure according to (5.3.17), we obtain the parameter estimates as reported in Table 5.4.

Indeed, the estimates of the parameter ν are roughly between 5 and 6.5, which clearly supports our initial impression we got from inspecting the scatter plots of the transformed data. The t copula seems, thus, to be a good choice in this respect. The fact that all the $\widehat{\varrho}$ are rather small numbers in absolute terms is not surprising either, as there is a one-to-one relationship between this copula parameter and Kendall's rank correlation, which again, can be understood as a measure of concordance in the bivariate case. Looking again at Figure 5.11, a concordant behaviour of the fund and the S&P 500 index returns is hardly visible there.

Apart from the point estimates of the copula parameters, there are also estimates of the

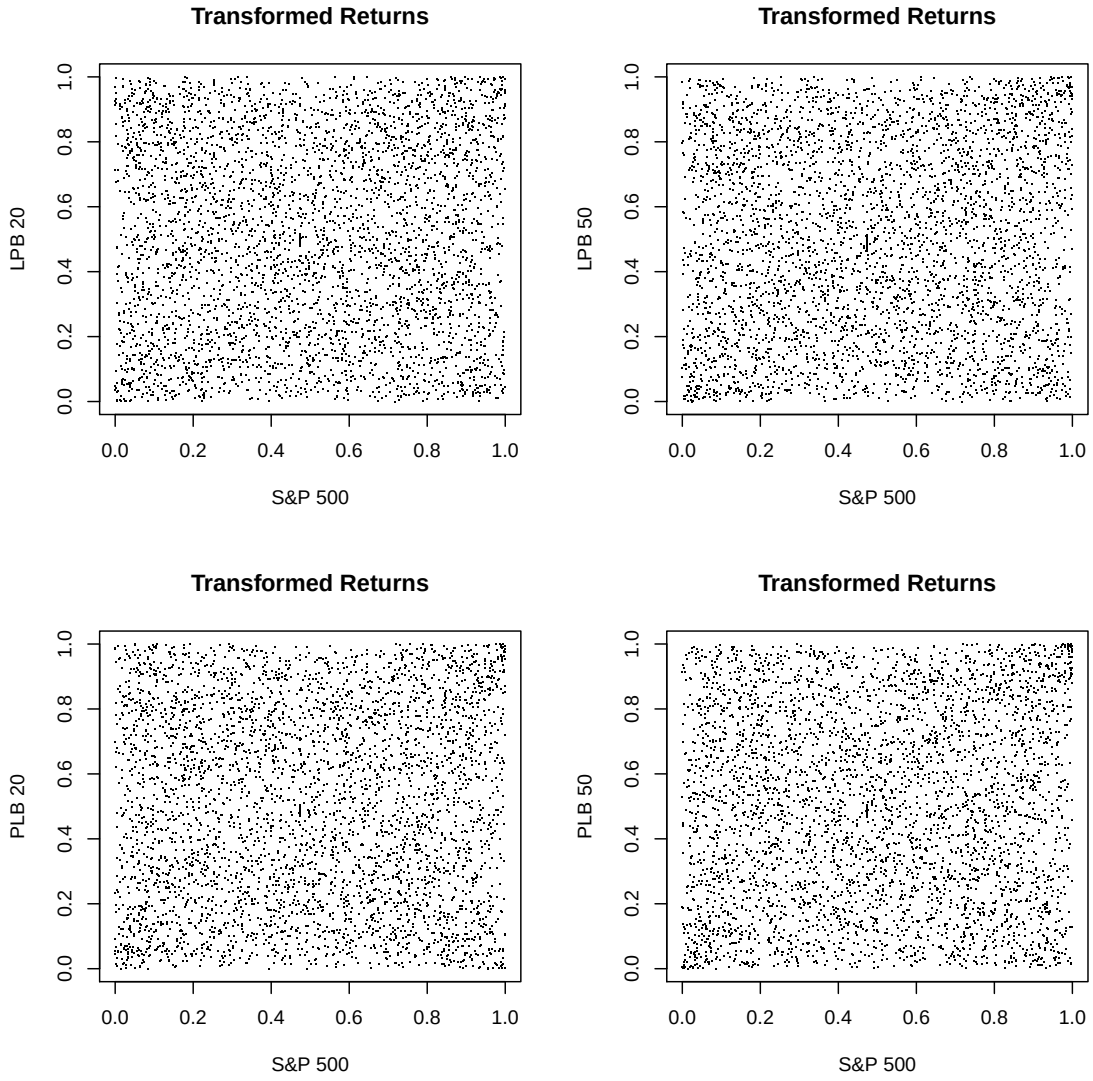

Figure 5.11: Scatter plots of the transformed data.

corresponding standard errors reported. However, these calculations are based on the assumption that we use actually observed and not (artificially) transformed data. What this implies for the estimated values, is difficult to say. However, the results may at least serve as a rough guideline.

Having estimated all the necessary parameters, we now have complete descriptions of the bivariate return distributions. Figure 5.13 shows the contour plots of the fitted return distributions. There we see very well the different dispersion in the univariate distributions on the margins. Even more illustrative are the contour lines of the corresponding densities. These are depicted in Figure 5.14 together with the scatter plots. The contour lines look indeed quite fancy.

Clearly, in this particular case here, we could probably say that an elliptical distribution and the linear correlation would also do the job. The linear correlation, as we have estimated it, seems to describe the data structure appropriately in terms of performance

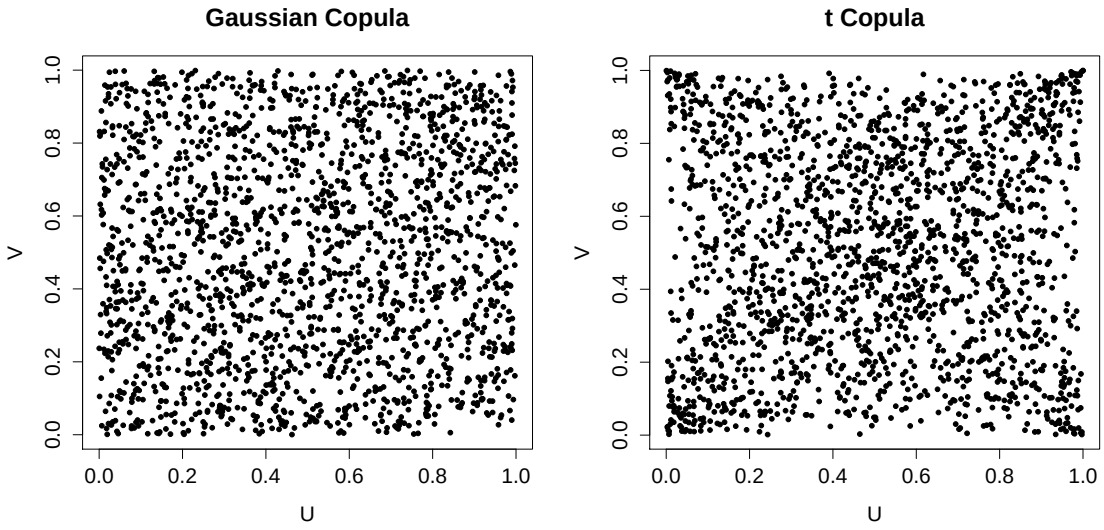

Figure 5.12: Gaussian vs. t copula.

	LPB20	LPB50	PLB20	PLB50
$\widehat{\varrho}$	0.0301	0.0667	0.0351	0.0729
$\widehat{se}\,(\varrho)$	0.0175	0.0176	0.0176	0.0177
$\widehat{\nu}$	6.3616	5.9678	5.4313	5.1669
$\widehat{se}\,(\nu)$	0.7683	0.7199	0.5707	0.5317

Table 5.4: Estimated parameters of the t copulas.

measurement in a CAPM framework. However, having a suitable multivariate description of the data in terms of a multivariate distribution at hand, offers, of course, much more opportunities than a simple performance measurement by means of the CAPM. It is especially interesting with regard to the risk monitoring and management of a whole investment position. Here one can think of an investor who desires to hold a part of his wealth in a passive well diversified portfolio, such as a fund that replicates the S&P 500 index, and another part in an actively managed portfolio, such as a statistical arbitrage hedge fund. With the discussed methods one can obtain an appropriate multivariate distribution that is flexible enough to describe almost any kind of dependence structure. Having the distribution of the return vector then enables the risk manager to draw conclusions about the distribution of the profit and loss of the overall investment position, which can then again be used in the calculation of any risk measures such as the expected shortfall (ES) or the value at risk (VaR), for example.

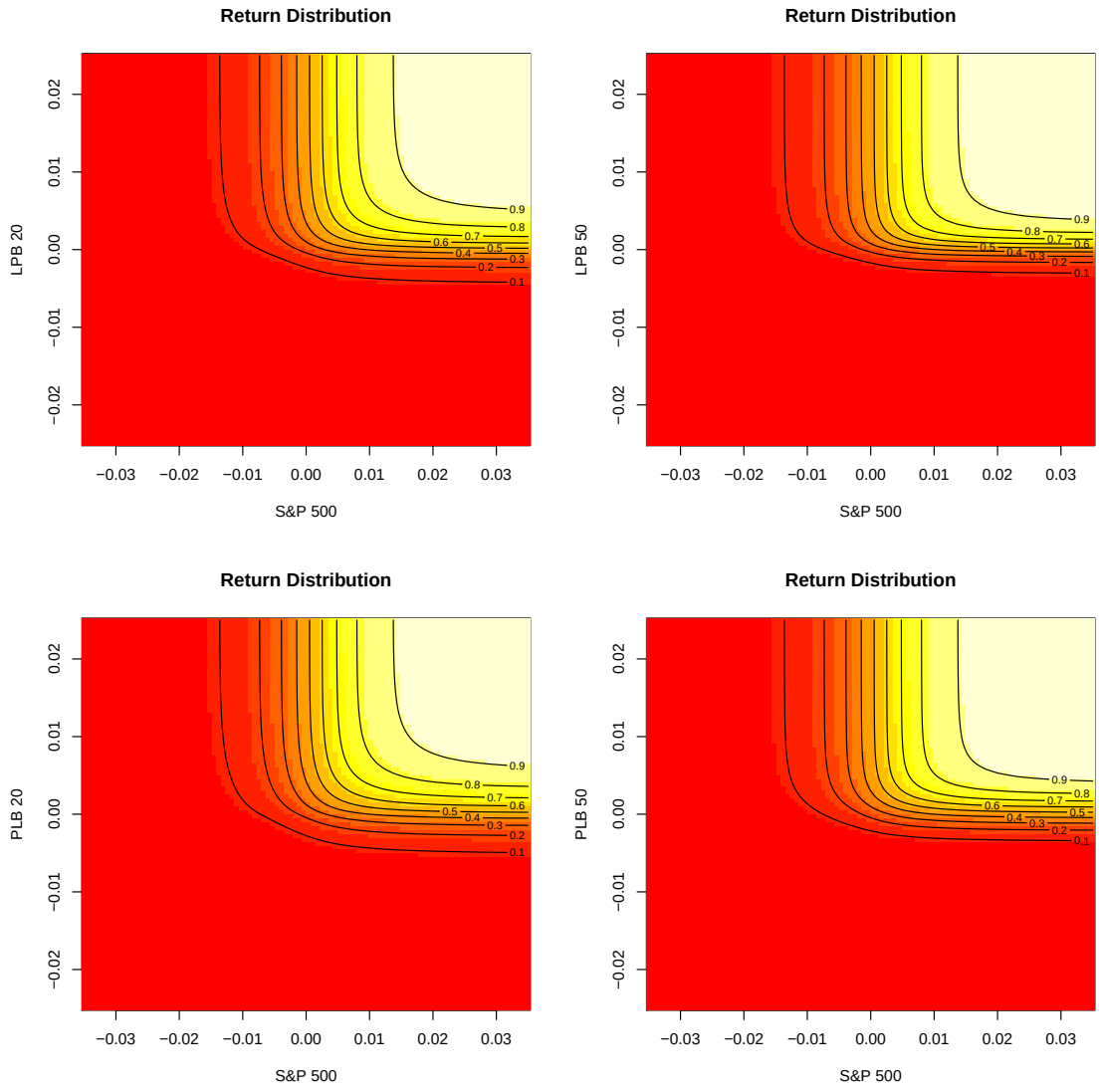

Figure 5.13: Contour plots of the fitted bivariate distributions.

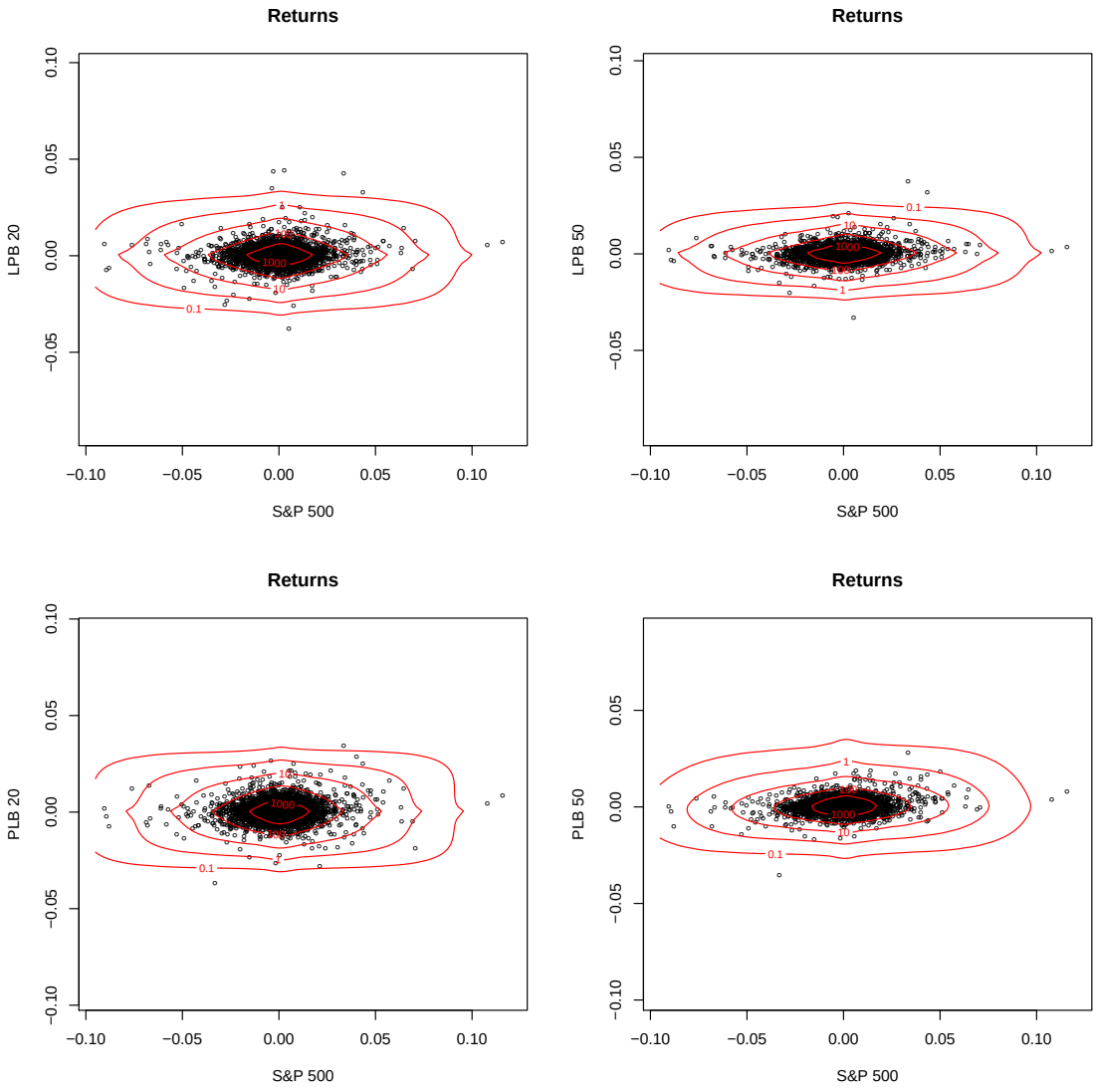

Figure 5.14: Return scatter plots with contour lines of the funds vs. the S&P 500 index.

Chapter 6

Conclusion

We have briefly discussed what we could term the main publicly available ideas in the field of a statistical arbitrage investment strategy focussing on pairs, and analysed in some detail a cointegration based approach. The thesis is now concluded with a short summary and an outlook with respect to a potentially promising extension.

6.1 Summary

After a short overview with respect to three main methods in the field of statistical arbitrage aimed at finding tradeable pairs, we focused on the cointegration based model. There we showed that this model can be consistent with the APT-approach. In a simulation study we found that a high persistence level in the spread time series may cause problems regarding the testing for cointegration. We also saw that the Phillips-Ouliaris trace statistic test has problems rejecting the null hypothesis when the starting values of the two series lie far apart. The subsequent back-testing of the model on an asset universe consisting of the S&P 500 stocks showed astonishingly good results. All the different strategy versions we tested exhibited not only a higher average return over the whole analysed time period as compared to the S&P 500 index, but also offered a much more favourable risk profile as determined by classical performance measurements. Also the advanced dependence analysis in the last part of the thesis attested our four prototype funds very convenient results in the sense that their returns depend only very weekly on the returns of the S&P 500 index, which is clearly a very desirable feature with respect to portfolio diversification.

6.2 Outlook

A very promising extension of the discussed models could be the expansion of the *pairs-term* of two assets to a *portfolio-term* of several assets. This is, one would use more than two assets and just look for a linear combination of them that is mean-reverting. In this case one might use the APT-method to obtain a smaller subset of portfolio-candidates. This would reduce the total number of possible combinations to a feasible amount that could be calculated through by the cointegration based method. In this case, with a given set of relevant risk factors, an obvious way to proceed would be to define a distance

measure with respect to the β-values of relation (2.2.1) and then determine a subset of similar assets accordingly. In order to do that, it would be advisable to standardise the risk factors before estimating the β-values so that they are on the same scale.

Bibliography

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences 353*(1674), 401–419.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Cox, J., J. Ingersoll, and S. Ross (1985). A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, 385–407.

Drouet-Mari, D. and S. Kotz (2001). Correlation and dependence. *London: Imperial College*.

Embrechts, P., A. McNeil, and D. Straumann (1999). Correlation: Pitfalls and alternatives. *Risk 5*, 69–71.

Engle, R. and C. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: Journal of the Econometric Society 55*(2), 251–276.

Frees, E. and E. Valdez (1998). Understanding relationships using copulas. *North American Actuarial Journal 2*, 1–25.

Gatev, E., W. Goetzmann, and K. Rouwenhorst (1999). Pairs trading: Performance of a relative-value arbitrage rule. *Unpublished Working Paper*.

Gatev, E., W. Goetzmann, and K. Rouwenhorst (2006). Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies 19*(3), 797–827.

Genest, C. and A. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering 12*(4), 347–368.

Genest, C., M. Gendron, and M. Bourdeau-Brien (2009). The advent of copulas in finance. *The European Journal of Finance 15*(7-8), 609–618.

Genest, C. and J. MacKay (1986). The joy of copulas: Bivariate distributions with uniform marginals. *American Statistician 40*, 280–283.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies 6*(2), 327–343.

Hull, J. (2009). *Options, futures and other derivatives*. Pearson Prentice Hall.

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall/CRC.

Joe, H. and J. Xu (1996). The estimation method of inference functions for margins for multivariate models. *Department of Statistics, University of British Colombia, Technical Report 166*.

Kim, S., J. Primbs, and S. Boyd (2008). Dynamic Spread Trading. *Unpublished Working Paper*.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The review of economics and statistics 47*(1), 13–37.

Markowitz, H. (1953). Portfolio Selection. *Journal of Finance 7*, 77–91.

McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative risk management: Concepts, techniques, and tools*. Princeton University Press Princeton, NJ.

Mikosch, T. (2006). Copulas: Tales and facts. *Extremes 9*(1), 3–20.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica 34*, 768–783.

Mudchanatongsuk, S., J. Primbs, and W. Wong (2008). *Optimal pairs trading: A stochastic control approach*. American Control Conference.

Nelsen, R. (1999). *An introduction to copulas*. Springer.

Øksendal, B. (2007). *Stochastic differential equations: An introduction with applications*. Springer Verlag.

Phillips, P. and S. Ouliaris (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica: Journal of the Econometric Society 58*(1), 165–193.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3).

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of finance 19*(3), 425–442.

Tobin, J. (1958). Liquidity preference as behavior towards risk. *The Review of Economic Studies 25*(2), 65–86.

Vidyamurthy, G. (2004). *Pairs trading: quantitative methods and analysis*. Wiley.

Wilmott, P. (2006). *Paul Wilmott on quantitative finance*. Wiley Chichester.

Appendix A

The GIG and the GH Distribution

This part provides some additional information with respect to less commonly used univariate probability distribution functions like the *Generalised Inverse Gaussian* and the *Generalised Hyperbolic* distribution.

A.1 Generalised Inverse Gaussian

The random variable X is said to have a generalised inverse Gaussian distribution (GIG), denoted as $X \sim N^{-}(\lambda, \chi, \psi)$, if its probability density function is

$$f_X(x) = \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}\right)^{\lambda}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{x} + \psi x\right)\right\} \ , \ x > 0$$

with K_λ denoting a modified Bessel function of the third kind with index λ and parameters satisfying $\chi > 0, \psi \geq 0$ if $\lambda < 0$; $\chi > 0, \psi > 0$ if $\lambda = 0$; and $\chi \geq 0, \psi \geq 0$ if $\lambda > 0$.

The moment generating function is given by

$$M_X(t) = \left(\frac{\psi}{\psi - 2t}\right)^{\frac{\lambda}{2}} \frac{K_\lambda\left(\sqrt{\psi(\chi - 2t)}\right)}{K_\lambda(\chi\psi)} \ .$$

From the moment generating function the expectation and the variance are easily determined as

$$E[X] = \left(\frac{\chi}{\psi}\right)^{\frac{1}{2}} \frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}$$

and

$$Var(X) = \left(\frac{\chi}{\psi}\right)\left[\frac{K_{\lambda+2}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)} - \left(\frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\right)^2\right] \ .$$

A.2 Generalised Hyperbolic

A random variable X is said to have a generalised hyperbolic distribution, written as $X \sim GH(\lambda, \chi, \psi, \mu, \sigma^2, \gamma)$ if

$$X \overset{\mathrm{d}}{=} \mu + W\gamma + \sqrt{W}\sigma Z$$

where $\mu \in \mathbb{R}$, $\gamma \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$, $Z \sim N(0,1)$ and $W \sim N^-(\lambda, \chi, \psi)$. It is a special case of a so-called *normal mean-variance mixture* distribution.

Obviously, the conditional distribution of $X \mid W = w$ is normal, i.e.

$$X \mid W = w \sim N\left(\mu + w\gamma, w\sigma^2\right)$$

Its probability density function is given by

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X|W}(x \mid w) f_W(w)\, dw \\ &= c \frac{K_{\lambda-\frac{1}{2}}\left(\sqrt{\left(\chi + (x-\mu)^2\sigma^{-2}\right)(\psi + \gamma^2\sigma^{-2})}\right) \exp\left\{(x-\mu)\gamma\sigma^{-2}\right\}}{\left(\sqrt{\left(\chi + (x-\mu)^2\sigma^{-2}\right)(\psi + \gamma^2\sigma^{-2})}\right)^{\frac{1}{2}-\lambda}} \end{aligned}$$

with normalising constant

$$c = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{\left(\psi + \gamma^2\sigma^{-2}\right)^{\frac{1}{2}-\lambda}}{\sqrt{2\pi}\sigma K_\lambda\left(\sqrt{\chi\psi}\right)} .$$

The moment generating function is best derived form its stochastic representation as a normal mean-variance mixture. This is

$$\begin{aligned} M_X(t) &= E\left[E\left[e^{tX} \mid W\right]\right] \\ &= e^{t\mu}\left(\frac{\psi}{\psi - 2t\gamma - t^2\sigma^2}\right)^{\frac{\lambda}{2}} \frac{K_\lambda\left(\sqrt{\psi(\chi - 2t\gamma - t^2\sigma^2)}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)} \ , \ \chi \geq 2t\gamma + t^2\sigma^2 . \end{aligned}$$

Also its unconditional expectation and variance can be easily derived from the normal mean-variance mixture representation. This is

$$\begin{aligned} E[X] &= \mu + E[W]\gamma \\ &= \mu + \left(\frac{\chi}{\psi}\right)^{\frac{1}{2}} \frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\gamma \end{aligned}$$

and

$$\begin{aligned} Var(X) &= E[Var(X \mid W)] + Var(E[X \mid W]) \\ &= \left(\frac{\chi}{\psi}\right)^{\frac{1}{2}} \frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\sigma^2 + \left(\frac{\chi}{\psi}\right)\left[\frac{K_{\lambda+2}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)} - \left(\frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}\right)^2\right]\gamma^2 . \end{aligned}$$

Figure A.1 illustrates some GH densities for different values of the shape parameters λ,χ and ψ.

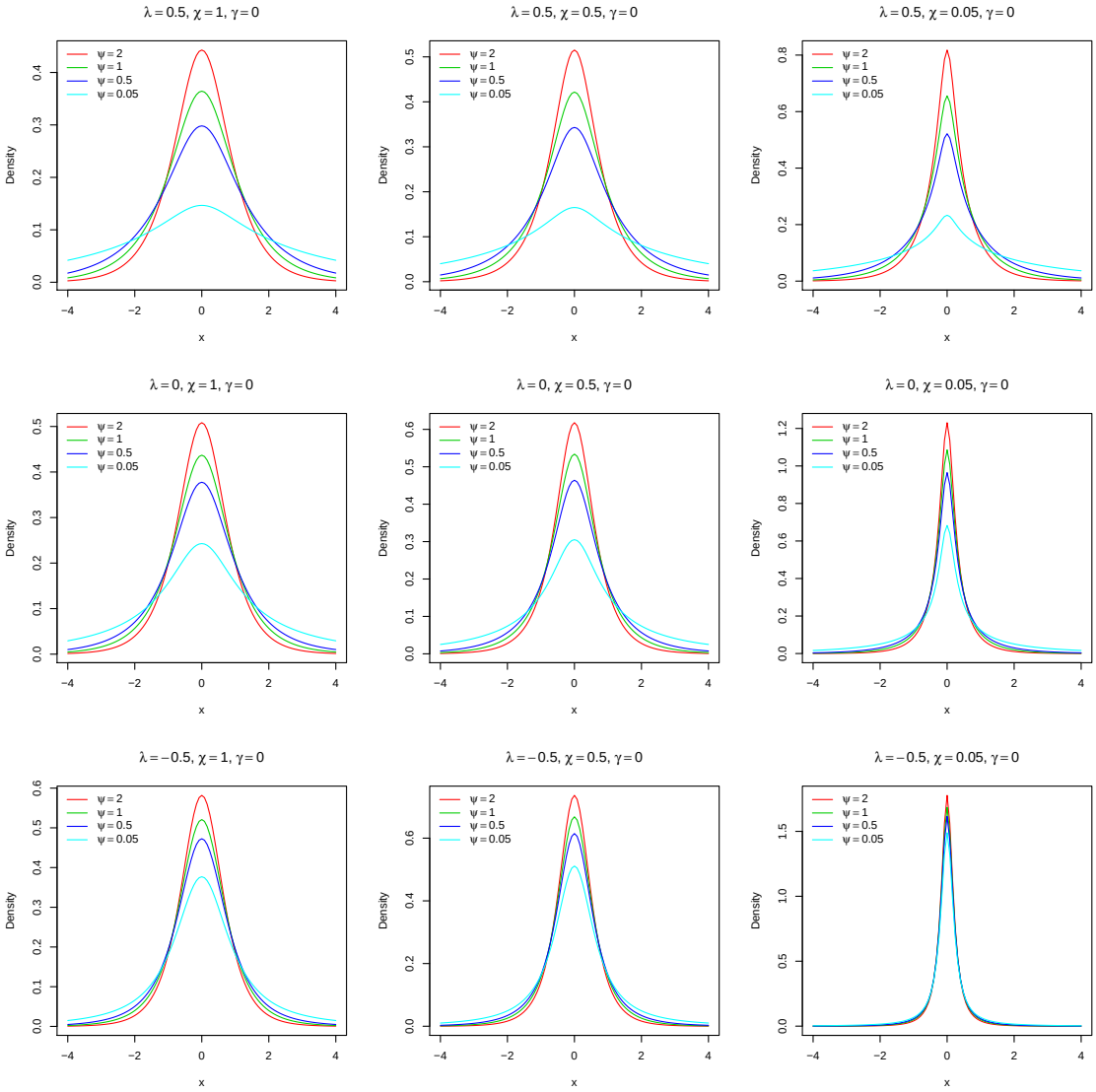

Figure A.1: GH densities for different parameter values.