Cointegration Based Algorithmic Pairs Trading

DISSERTATION
of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor Oeconomiae

submitted by

Markus Harlacher

from

Schöfflisdorf (Zürich)

Approved on the application of

Prof. Dr. Heinz Müller

and

Prof. Francesco Audrino, PhD

Dissertation no. 4600

D-Druck Spescha, St. Gallen 2016

The University of St. Gallen, School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, October 24, 2016

The President:

Prof. Dr. Thomas Bieger

To my family

Contents

Abstract

This work analyses an algorithmic trading strategy based on cointegrated pairs of assets. The centrepiece of such a strategy is the discovery of tradable linear combinations consisting of two or more financial assets. An intuitive attempt to identify such linear asset combinations is based on statistical tests. An unrestricted testing of all possible linear combinations leads, however, to a multiple testing problem. Carrying out a sufficiently large number of tests always gives significant results, even if not a single combination is truly cointegrated. This is in the nature of statistical tests. Well established correction methods like the popular Bonferroni correction turn out to be too conservative in such cases and make it impossible to discover even truly cointegrated combinations that would be highly profitable if applied in the proposed investment strategy.

A possible way to mitigate this problem can lie in the effective pre-partitioning of the considered asset universe with the purpose of reducing the number of feasible combinations and, therefore, the number of statistical tests, to combinations with an increased potential of being profitable. This is the main contribution of this dissertation. Besides analysing the robustness of established cointegration tests with respect to particular strategy-relevant features, the main focus lies on possible ways to pre-partition the overall set of admissible assets. A carefully carried out back-testing finally inspects the effectiveness of the proposed methods.

The back-testing results, which are based on an asset universe consisting of S&P 500 stocks and a time period stretching from January 1995 up to December 2011, show a very favourable picture. Apart from an attractive rate of return and a significantly smaller volatility as compared to the S&P 500 index, the returns of the applied pairs-trading strategies showed only a marginal correlation with the overall market returns.

Zusammenfassung

In dieser Arbeit wird eine algorithmische Anlagestrategie untersucht, die auf paarweise cointegrierten Assets basiert. Das Kernstück solcher Strategien bildet das Auffinden von handelbaren Linearkombinationen aus mindestens zwei Assets. Ein intuitiver Ansatz, um solche Linearkombinationen aufzufinden, basiert auf statistischen Tests. Ein uneingeschränktes Testen aller möglichen Linearkombinationen führt jedoch zum Problem des multiplen Testens. So werden bei einer genügend grossen Anzahl Tests immer signifikante Resultate gefunden, auch wenn in Wahrheit keine einzige Kombination cointegriert ist. Dies liegt in der Natur statistischer Tests. Bekannte Korrekturmethoden wie beispielsweise die häufig angewendete Bonferroni-Korrektur sind in solchen Fällen zu konservativ und verunmöglichen auch das Auffinden von real cointegrierten Linearkombinationen, die für die Anlagestrategie äusserst profitabel wären. Ein möglicher Ansatz zur Linderung dieses Problems kann in einer wirkungsvollen Vorgruppierung des Anlageuniversums liegen, um damit die Anzahl Kombinationsmöglichkeiten, und damit die Anzahl statistischer Tests, auf Kombinationen mit grossem Profitabilitätspotential zu reduzieren. Dies ist denn auch der Kern dieser Dissertation, die sich neben der Frage der Robustheit von bekannten Cointegrationstests in Bezug auf die für eine solche Anlagestrategie relevanten Eigenschaften vor allem mit möglichen Varianten der Vorselektion beschäftigt. Die Wirksamkeit der vorgeschlagenen Methoden wird anschliessend mit einem sorgfältig durchgeführten Back-Testing überprüft.

Die Back-Testing-Resultate basierend auf einer Periode von Januar 1995 bis Dezember 2011 und einem Anlageuniversum, das aus Aktien der im S&P 500 Index geführten Unternehmen besteht, zeigen ein äusserst vielversprechendes Bild. Neben einer ansprechenden Rendite und einer gegenüber dem S&P 500 Index deutlich tieferen Volatilität besitzen solche Anlagestrategien auch nur eine marginale Korrelation mit dem Gesamtmarkt.

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
$\stackrel{\mathrm{d}}{\to}$	Converges in distribution
$N\left(\mu, \sigma^2\right)$	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
$SE(\cdot)$	Standard error of an estimated parameter
$A^\top$	Transposed of matrix A
$tr\,(A)$	Trace of matrix A
$\boldsymbol{\nabla}$	Function gradient: $\boldsymbol{\nabla} f = \left(\frac{\partial f}{\partial x_1} \dots \frac{\partial f}{\partial x_d}\right)^\top$
∇^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 $\mathcal{M}$

Chapter 1

Introduction

A pairs trading strategy that is based on statistical procedures is a special form of statistical arbitrage investment strategy. More generally, it belongs to the set of relative value arbitrage investment strategies, which encompasses a large variety of different investment strategies that share some particular features. The most essential feature of all those strategies is probably the fact that not only does the asset side follow a stochastic process, but so does the liability side. While stochastic behaviour of the values on the asset side is rather common for any financial fund, non-deterministic values on the liability side are clearly less usual. Thus, the stochastics with regard to a fund's surplus come in this case not only from the asset side, but also from the liability side. This is a particular feature that all relative value arbitrage funds have in common. However, how the two sides are created and what kind of relationship exists between them depends on the particular type of strategy. Statistical arbitrage funds, in particular, construct their long- and short-positions based on statistical models and typically have an almost market-neutral trading book. A special form of this strategy, sometimes referred to as pairs trading, tries to determine pairs of assets, where one of them is purchased while the other is sold short - that is to say, the asset is borrowed and then sold without actually owning it. In this particular form, two assets with an assumed price relationship are traded. If the implied price relationship really holds true and the two prices move largely together, the idea is to exploit the mean reverting behaviour of the corresponding price spread between the two securities. Thus, when the price of one security increases relative to the price of the pair-security, which means that there is a short-term deviation from the long-term equilibrium relationship, the strategy suggests selling that security short and using the proceeds to purchase (i.e. go long in) the pair-security. If the price spread again approaches its equilibrium level, the strategy leads to a gain by reverting the corresponding positions. As one asset is financed by short-selling another asset, the value of the financing is obviously subject to random fluctuations too.

1.1 The Contribution of Relative Value Investments

At first sight, it might seem odd to establish an investment strategy with an additional random part, as it is already challenging to handle investments with a stochastic asset side only. However, a stochastic liability side can potentially offer some distinct advantages. In particular, a relative value arbitrage fund needs only to make relative statements with respect to security prices. This means that the long position is established by assets that the portfolio manager considers to be undervalued relative to some other assets that, in this case, must establish the corresponding short position. So, there is no statement about whether an asset by itself is correctly priced in absolute terms. There are just relative value statements. This also explains the name of this kind of investment style. Thus, from such an investment portfolio, we would presumably expect returns that show very distinct behaviour with respect to movements in the overall financial market, at least as opposed to "conventional" investments consisting of only (stochastic) long positions. We would probably expect that our long-short portfolio may show a profit even when the overall financial market suffers heavy losses. Such long-short investments, however, do not necessarily have to be market-neutral[1] in their general form, although there exists a particular form aiming for an explicit market neutral-investment.

The aim of relative value investments is, thus, not primarily to achieve returns that are as high as possible, but rather to obtain returns that show a negligible dependence structure with a chosen benchmark or market portfolio while still generating returns above a rate that is considered risk-free. That the term "risk-free" can be a bit tricky should have become clear after the

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

sovereign debt crisis of 2008. So, many allegedly "risk-free" assets may turn out to be riskier than actually desired. In addition, it is not completely tortuous to assume that the development of their risk profiles are significantly correlated with greater turbulences on the stock markets. The aim of a pairs trading strategy is, thus, to contribute to an investor's overall portfolio diversification. That this is an important aspect of an investment is probably most highlighted in the popular investment portfolio concept proposed by Markowitz (1953) and Tobin (1958). In their mean-variance framework, they consider, in addition to the mean and the variance of the individual asset returns, their pairwise Pearson correlations. Clearly, whether the Pearson correlation is a suitable dependence measure in a particular case is a valid question. However, as long as we are willing to assume an elliptical distribution framework, employing linear correlation as a proper dependence measure may be absolutely reasonable.

Anyway, whatever dependence measure is used, there is little doubt about the importance of a favourable dependence structure of the assets in a portfolio. It is exactly there that a relative value arbitrage strategy, like a pairs trading investment, has its biggest potential. This is presumably also the reason why such investments have become so attractive in the past few years, not only for institutional investors but also for private ones.

1.2 The Asset Selection

According to the description so far, almost any investment style with both a stochastic asset and a stochastic liability side can be said to pursue a relative value arbitrage strategy. Even a fund that has its asset and liability side constructed by the pure feelings of its portfolio manager belongs to this category. The very particular way that we construct our asset and liability side determines more closely the type of relative value arbitrage fund. One such way is to employ statistical methods in order to tackle the asset selection problem. This type of strategy is often called statistical arbitrage investment. This means that, in a statistical arbitrage investment, there exist statistical relationships between the fund's asset and liability sides. If such relationships consist of only two securities each (i.e. one security on the asset side and one on the liability side), we call this special form of statistical arbitrage pairs trading. There are other forms of pairs trading, however, which may not be based on statistical relationships. Thus, the term "pairs trading" is not unique. In this dissertation we will look at an algorithmic pairs trading investment strategy that is based on the statistical concept of cointegration.

1.3 The Aim of this Analysis

As a cointegration based pairs trading strategy uses statistical methods in order to establish the fund's asset and liability sides, the strategy also faces the typical problems that are entailed by the employed methods. When using statistical tests as the criteria to find tradeable linear combinations, we obviously run into the problem of multiple testing. To make this point clear, if we want to engage in a pairs trading strategy and have an asset universe of 100 securities, we will obtain 4,950 combinations on which we could potentially trade. So, there are 4,950 decisions to take. If we have a statistical test procedure at hand that allows us to draw a decision about the marketability of a particular pair, we would have to perform 4,950 tests. The tests could be constructed with the null hypothesis H_0 stating that "the tested pair is not tradeable". If H_0 is then rejected on a chosen significance level for a particular pair, we would consider the pair for trading. Now, what if the null hypothesis is actually true for all tested pairs? This would mean that in fact none of the possible combinations is eligible for trading. By carrying out 4,950 independent tests, we would, nevertheless, expect to obtain around 50 alleged pairs for trading at the one per cent significance level. The probability of committing at least one type I error would be $1 - (1 - 0.01)^{4950} \approx 1$.

Hence, the procedure on which we base our decision with respect to an appropriate selection of the securities is certainly at the centre of a statistical pairs trading investment style. Therefore, it is worth investing some time and effort here to make a considerable contribution.

In addition, long-short investment strategies are usually considered absolute return strategies, meaning that they have only negligible exposure with respect to a representative market index. Even though it is not the aim to come up with an investment strategy that is able to achieve uncorrelated or even independent returns regarding such a market index, it will nevertheless be very interesting to analyse the correlation structure of the proposed strategy with a broad market index.

1.4 The Structure of this Work

We start with a short description of the general idea of this type of strategy in Chapter 2, where we also briefly discuss the pioneering work of Gatev, Goetzmann, and Rouwenhorst (1999). The discussion in Chapter 2 also provides the necessary background knowledge for a sound understanding of what is presented as a potential solution to multiple testing. Chapter 3 gives some basic knowledge about price generating processes that will be used in the subsequent chapters. Readers who are already familiar with the commonly used price generating approaches may want to skip this chapter. Chapter 4 continues then with a short review of four popular cointegration tests: the Augmented Dickey-Fuller Test, the Phillips-Perron Test, the Kwiatkowski-Phillips-Schmidt-Shin Test and the Phillips-Ouliaris Test. Experts on these tests may also want to skip this chapter, the purpose of which is to provide the knowledge necessary in order to be able to understand the analysis in Chapter 5. There we conduct a brief simulation study that is intended to check the sensitivity and specificity of the proposed selection procedure under different settings. This will give us the answer to the question of whether, or in what particular cases, the employed methods really work in the expected way. This also gives us the opportunity to examine some particular situations where it is not clear *ex ante* how well the chosen tests are able to handle them. With the results of the simulation study in mind, we will be able to implement a simple cointegration based pairs trading strategy. In order to tackle the still unsolved multiple testing problem, we will discuss three different ideas based on a pre-partitioning of the asset universe in Chapter 6. Having identified potentially profitable pairs for trading, we will have to decide on a trading rule - that is, we have to define the kinds of event that trigger a selected pair to open and to close again. This is addressed in Chapter 7. The corresponding back-testing is finally carried out and discussed in Chapter 8. There we do a proper back-testing on the US stock market for the time period from 1996 up to 2011. The actual performance of our prototype funds in comparison to the

S&P 500 index is also comprehensively analysed there. We finally conclude the study with a brief summary and give a short outlook with respect to further research in this area.

Chapter 2

A First Overview

Before we highlight the idea of cointegration and pre-selecting securities in terms of pairs trading, we first start our overview by having a look at the relevant literature, which is then followed by a short summary of the presumably first published pairs trading approach. The chapter is finally completed by a brief discussion about the pros and cons of employing transformed versus untransformed price time series to match the pairs.

2.1 Literature Review

Despite of the fact that pairs trading strategies have already been applied in financial markets for several decades, the corresponding academic literature is not as rich as for other finance topics. As already mentioned in Chapter 1, there are many ways of trading pairs. When Do, Faff, and Hamza (2006) introduced their pairs trading approach, which they called *the stochastic residual spread model*, they summarised in their literature review three existing approaches which they considered as the main methods in the context of pairs trading. The first one they labelled *the distance method*. This method encompasses among others the scheme employed by Gatev, Goetzmann, and Rouwenhorst (2006). The second one they called *the stochastic spread method*. This method refers to the approach described in detail by Elliott, Van Der Hoek, and Malcolm (2005). Last but not least, the third main method they considered worth mentioning was *the cointegration method*. Huck (2010) proposed an extension to those approaches by combining forecasting techniques and multi-criteria decision making methods and found promising results in a back-testing study carried out on the S&P 100. Papadakis and Wysocki (2007) extended the approach proposed by Gatev et al. (2006) by considering the impact of accounting information events, such as earnings announcements and analysts' earnings forecasts. In their back-testing study, which was carried out on the US stock market for the time period from 1981 up to 2006, they found that many pairs trades were triggered around accounting events and that positions that were opened just before such events were significantly more profitable. Do and Faff (2008) examined again the strategy established by Gatev et al. (2006) on an extended data set stretching from 1962 up to 2009 and reported in their study a tendency of a generally diminishing profitability of that particular strategy, but also mentioned that the strategy performed quite strongly during periods of prolonged turbulences.

In his book about pairs trading, Vidyamurthy (2004) suggested different techniques to successfully detect and trade on pairs. He was probably the first one who proposed the application of the cointegration concept in the context of pairs trading. As we will see later, cointegration is indeed a very perspicuous approach to find suitable assets for the purpose of pairs trading. So, it is no surprise that it has been recommended also by other influential authors like, for instance, Wilmott (2006). Unfortunately, neither Vidyamurthy (2004) nor Wilmott (2006) provide any back-testing results. They just proposed the concept but no empirical analysis. Other studies, such as the one by Do et al. (2006), provide results for a few preselected pairs but not for a broad market. Bogomolov (2010) conducted a cointegration based back-testing on the Australian stock market for the time period stretching from 1996 up to 2010 and reported that the method showed a good performance and led to almost market-neutral returns. A back-testing of the cointegration approach based on the American stock market for the period stretching from 1996 up to 2011 carried out by Harlacher (2012) showed a similar result with a very appealing performance and hardly any dependence with the overall market.

At this point it is worth mentioning that both Bogomolov (2010) as well as Harlacher (2012) executed an unconditional cointegration approach, running the strategy on a broad asset universe without any further restrictions, following basically the idea of the approach by Gatev et al. (2006). The application of the cointegration method on a bigger set of assets without imposing any further restrictions is, however, not unproblematic as we will discuss later in more detail.

An interesting pairs trading back-testing based on the same approach and time interval as employed by Gatev et al. (2006) was also carried out on the Brazilian market by Perlin (2009). The study examined the approach with daily, weekly as well as monthly prices and confirmed the good results that are already reported by other studies for other markets also for the Brazilian market.

Bowen, Hutchinson, and O'Sullivan (2010) analysed a pairs trading rule on the FTSE 100 constituent stocks in the context of high frequency trading for the time interval stretching from January to December 2007. They reported that the strategy's excess returns have little exposure to traditional risk factors and are only weakly related to the overall market, which is in line with most other pairs trading back-testing studies based on daily data. As an important issue they mentioned the many transactions of the strategy in this context. Also Nath (2003) implemented a high frequency pairs trading scheme, however, not in a stock market environment but in the highly liquid secondary market for U.S. government debt.

The studies mentioned so far focused on the matching of pairs and the subsequent trading. The study carried out by Andrade, Di Pietro, and Seasholes (2005), however, focussed on a possible explanation for the excess returns obtained by a typical pairs trading strategy and attached it mainly to uninformed demand shocks. In particular, it is argued that pairs traders try to keep the relative prices in line and the excess returns are, thus, a compensation for providing liquidity during times of differential market stress. This view is also confirmed by Engelberg, Gao, and Jagannathan (2009).

2.2 A Short Summary of the Presumably First Published Pairs Trading Approach

As already mentioned before, the presumably first academic pairs trading paper was the work of Gatev et al. (1999). As this work basically laid the ground for many subsequent publications in the area of pairs trading, it makes sense to briefly discuss this first approach in more detail. This should also enable us to see the crucial differences between the distance method and the cointegration method, i.e. the approach on which we are going to focus later, and it should, hence, make us understand why the cointegration method should not be applied an a broad asset universe in the same way as we can apply the distance method.

The strategy proposed by Gatev et al. (1999) can be considered as a non-parametric approach that consists of two stages, i.e. the pairs-formation period and the trading period. For the first stage, i.e. the so-called pairs-formation stage, they rather arbitrarily choose a period of 12 months to find suitable pairs. Candidates are found by constructing 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 select for the second stage, i.e. the so-called trading period, the top pairs with the smallest price distance and trade them during a period of at most 6 months. As the core of this procedure relies on the squared Euclidean norm, it is usually referred to in the literature as the *minimum distance method* or just *distance method*.

With a set of promising pairs available, it is necessary to determine a suitable trading rule, i.e. a rule according to which positions are opened and closed again. The rule proposed by Gatev et al. (2006) is a rather simple one. A position in a pair is opened by going one dollar long in the lower priced stock and one dollar short in the higher priced stock when the two normalised prices diverge by more than two historical standard deviations[2]. Positions are unwound when the two prices converge again - that is to say, when the price difference between them becomes zero. A particular trading period is limited to 6 months. This means that any open position is closed at the end of a trading period, independent of any price convergence occurrence.

Employing this strategy on the US equity market for the period from 1962 to 2002, Gatev et al. (2006) reported average annualised excess[3] returns of up

[1] Reinvested dividends are included in the index.
[2] The standard deviation is estimated during the formation stage.
[3] *Excess* means here just the return component that exceeds the return of a risk-free investment.

to 11%. This is quite an impressive result. In addition, they mentioned that portfolios consisting of more than one pair showed significant diversification benefits. This means that an increasing number of pairs in a particular portfolio results in a lower standard deviation. An interesting observation in their study was the fact that the maximum realised return of a portfolio remained quite stable when the number of pairs was increased. The minimum realised return, however, increased when additional pairs were added to the portfolio.

A key characteristic of this method is the absence of any model assumptions. In order to match pairs, it is not necessary to estimate any parameter values. There is just an empirical distance measure telling us that two normalised prices moved largely together in the past, leaving us with the hope that this will also continue in the future.

2.3 The Idea of Cointegration in Terms of Pairs Trading

In contrast to the distance method, as discussed in Section 2.2, the cointegration method is a model based, parametric approach that assumes the existence of a common stochastic trend in the price series of two financial assets. In order to understand the concept, it is important to be familiar with the idea behind *stationary* and *non-stationary* time series. These two terms describe an important characteristic of any time series, such as prices of financial assets, for example. A short summary of the main points regarding stationarity as well as an exact definition of cointegration is given in Chapter 4.

For the moment we can be less precise and just reason what we need to have in order to implement an effective pairs trading strategy. The distance method, as explained above, just takes two securities that moved closely together in the period prior to the trading period. Moving together means, thus, that the prices of the two securities do not drift apart too much, which basically means that the price spread between them shows some mean reverting behaviour. In the ideal case, the spread time series has a constant mean over time. This is now where the cointegration concept comes into play. An important feature of two cointegrated time series is that there exists a linear combination of the two non-stationary time series that has a constant mean. From this point of view, it is clear that in case we have two assets with cointegrated price time series, they would be ideal for pairs trading. The only problem that remains, is the issue that we do not know whether two time series are actually cointegrated or not. Fortunately, there have been statistical procedures developed to test for that feature. One of the most commonly used test procedure is the one proposed by Granger (1981). It is a two-step procedure that fits in the first step a linear regression of one series on the other by using *ordinary least squares* (OLS), i.e.

$$y_t = \alpha + \beta x_t + \varepsilon_t \ , \ \text{for } t = 1, \ldots, T \tag{2.3.1}$$

with ε_t being the error term, and then checks in the second step whether ε_t can be assumed to be *weakly stationary*. Frequently used tests for this purpose are the augmented Dickey-Fuller test (ADF), the Phillips-Perron test and the KPSS test. An important remark in this regard concerns the critical values to be used in those tests. As the spread process is not a directly observed time series but results from an estimated relationship according to (2.3.1) with correspondingly estimated coefficients, it is necessary to adjust the critical values in the tests as proposed by Phillips and Ouliaris (1990). In addition, there is another problem with this approach. In particular, it is unknown which time series to take on the left hand side of (2.3.1) and which one on the right hand side. This is, the estimated values of the cointegration vector are not independent of the chosen normalisation. Therefore, we actually have to test in both directions with that approach.

The studies of Bogomolov (2010) and Harlacher (2012) both applied the cointegration approach on two different stock markets in a very similar style like Gatev et al. (2006) tested the distance method on real data. They used a 12 months matching period prior to each trading period, which they also limited to a maximum of 6 months. They estimated spread time series for each pair combination and applied a stationarity test on each of them. Bogomolov (2010) used the *Dickey-Fuller* test[4] and took as trading candidates all pairs that showed a test statistic greater than the 5% critical value according to the test distribution. The candidates with the lowest spread process standard deviation were then taken for trading. For the back-testing Bogomolov (2010)

[4]For more details about that test, see Section 4.2.

considered two portfolios, one consisting of 5 pairs and one of 20 pairs. By contrast, Harlacher (2012) did not use any critical test values but just took the pairs with the highest test statistic according to the Phillips-Ouliaris cointegration test[5]. In a single test the Phillips-Ouliaris cointegration test rejects the null hypothesis of two processes being not cointegrated at a lower significance level for higher values of the corresponding test statistic. For the back-testing, Harlacher (2012) used two portfolios, one consisting of 20 pairs and one of 50 pairs. There were no further restrictions in the formation of the trading portfolios.

The results of all pairs-portfolios looked quite favourable over the tested periods. Even though the back-testings of the two studies were carried out on two completely different markets, the return processes of their portfolios looked remarkably similar. In both studies, the investments in the pairs-portfolios showed hardly any dependence with the stock index that served as the asset universe from which the pairs were taken. Another striking feature was the strong increase in value of the pairs-portfolios during the recent financial crisis. This is, however, also a fact that Do and Faff (2008) reported in their review of the distance method based on more recent data.

As already mentioned by Harlacher (2012), using critical values of any statistical test when testing several hundred pairs, is basically meaningless. In this context, it is, however, also questionable to which extent we can actually rely on a pairs formation method that uses any kind of statistics stemming form a statistical test. In any case, we should somehow tackle the multiple testing problem to get more confidence in the cointegration method. In this context it makes clearly sense to discuss in more detail different test procedures that can be used for this purpose. This is what we are going to do in Chapter 4. As we will see, the different tests all have their very particularities. They make different assumptions about the data generating process, and so it is not too surprising that they can come to different conclusions when the same data is tested. Before we apply the tests on real data, we will briefly check in Chapter 5 how the different tests perform under different assumptions. This will help us select the most appropriate test for our purpose based on our believes about the true data generating process.

[5]See section 4.5 for more details about this test procedure.

2.4 Pre-Selecting Securities

With no doubt, in order to have two securities that move largely together, we would expect the underlying companies to be "similar" to some extent. What we exactly mean by the term "similar" still has to be clarified. However, one of the first ideas we might have, leads us to the conclusion that two companies operating in two completely different economic sectors are less likely to have cointegrated security prices. So, we could, for example, say that two similar companies must be operating in the same economic sector. A partition that is based on the companies' economic main activities is, hence, not the worst idea to start with.

When attempting to allocate different companies to predefined economic sectors, we will, unfortunately, quickly realise that this task can be tricky in some cases. For example, what if some companies operate in several economic sectors simultaneously? The decision to be taken is, however, just a binary one. This is, a company that is active in several economic sectors would necessarily have to be assigned to one single predefined sector. From this point of view, the mentioned partitioning method is probably too crude and we should aim for a method that is able to account for multi-sector activities. Therefore, a method that gives us a kind of exposure of a company to a particular economic sector might serve our purpose much better.

In addition to the sector membership, there are probably some other aspects that influence a company's share price. Therefore, Vidyamurthy (2004) suggested to employ the *Arbitrage Pricing Theory* (APT) initially proposed by Ross (1976). It is a very general asset pricing model with a linear return-generating index model at its centrepiece. Assuming the *law of one price* to hold and market participants to have homogeneous expectations, it claims that the return on any stock is linearly related to a set of so-called risk factors, i.e.

$$E\left(r_i\right) = r_f + \sum_{j=1}^{k} \beta_{i,j} r_j^* \tag{2.4.1}$$

where r_j^* can be seen as a risk premium of risk factor j and $\beta_{i,j}$ is the risk exposure of asset i to risk factor j. The r_f denotes in this case the risk-free return.

Without repeating the whole APT, which is a very interesting theory on its own, the relevant statement in the context of a pairs trading strategy is that assets with virtually identical risk exposures should yield the same expected returns.

Unlike the purely statistical methods of Section 2.2 and Section 2.3, the APT is 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 pairs trading 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 that has very similar risk factor exposures, reducing the systematic risk of the pair position to a minimum.

Unfortunately, the part of the APT we are actually interested in, i.e. the return-generating index model, is usually not provided in the academic literature. Neither becomes Vidyamurthy (2004) more specific about the appropriate risk factors in his proposal. In addition, the nice property with regard to the returns of two securities does not guarantee the existence of a weakly stationary spread process for the two securities. Even if the signal to noise ratio in the model is such that the two securities show highly correlated returns, the spread process of the two prices is as such, without any further restrictions, not necessarily weakly stationary.

A more detailed discussion on the asset selection procedure and the handling of the multiple testing problem is given in Chapter 6.

2.5 The Question of Transforming Prices

The cointegration approach as suggested by Vidyamurthy (2004) uses transformed security prices by taking the logarithm of them, which implies for the spread process the following relationship:

$$m_t = \log\left(P_t^A\right) - \beta \log\left(P_t^B\right) \ . \tag{2.5.1}$$

A clear motivation for that is, however, not provided. Neither is there an obvious reason why we should do that. Why not just taking price levels directly? The subsequent considerations are intended to obtain a deeper understanding about this question. For instance, using (2.5.1) with price levels directly, where the prices at time $t+1$ are given by

$$\begin{aligned}
P_{t+1}^A &= P_t^A e^{r_{t+1}^A} \\
P_{t+1}^B &= P_t^B e^{r_{t+1}^B}
\end{aligned}$$

with r^i, $i \in \{A, B\}$, denoting the continuously compounded return[6] with respect to the time period stretching from t to $t+1$, the spread value at time $t+1$ can be expressed by

$$\begin{aligned}
m_{t+1} &= P_t^A e^{r_{t+1}^A} - \beta P_t^B e^{r_{t+1}^B} \\
&= m_t + P_t^A \left(e^{r_{t+1}^A} - 1 \right) - \beta P_t^B \left(e^{r_{t+1}^B} - 1 \right)
\end{aligned} \tag{2.5.2}$$

which actually implies

$$m_{t+1} = m_t \quad \Leftrightarrow \quad P_t^A \left(e^{r_{t+1}^A} - 1 \right) = \beta P_t^B \left(e^{r_{t+1}^B} - 1 \right) \ .$$

If we assume for the moment that the two prices are in equilibrium at time t, i.e. m_t corresponds to the value around which the spread process fluctuates in a steady way over time, it is interesting to see what happens to the spread value m if both assets generate the same return in the period from t to $t+1$. In particular, if we have $r_{t+1}^A = r_{t+1}^B$, we obtain

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

which means that as long as the equilibrium spread value is not equal to zero, the spread value changes in spite of identical returns. This is somehow counter intuitive, especially in light of our reasoning in Section 2.4. Establishing a price relationship on price levels directly means that we can express the price of asset A at time t according to

$$P_t^A = \alpha + \beta P_t^B + \varepsilon_t \tag{2.5.3}$$

with ε_t having an equilibrium value of zero. If we now require for our spread process an equilibrium value of zero, we correspondingly need to have $\alpha = 0$.

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

By contrast, if we take log-transformed prices, the question of a non-zero spread process equilibrium does not pop up. To see that, let us consider again

$$
\begin{aligned}
P_{t+1}^{A} &= P_{t}^{A} e^{r_{t+1}^{A}} \\
P_{t+1}^{B} &= P_{t}^{B} e^{r_{t+1}^{B}}
\end{aligned}
$$

as before, but this time writing our linear combination in terms of log-prices. In this case, the spread at time $t+1$ is given by

$$
\begin{aligned}
m_{t+1} &= \log\left(P_{t}^{A} e^{r_{t+1}^{A}}\right) - \tilde{\beta}\log\left(P_{t}^{B} e^{r_{t+1}^{B}}\right) \\
&= \log\left(P_{t}^{A}\right) + r_{t+1}^{A} - \tilde{\beta}\log\left(P_{t}^{B}\right) - \tilde{\beta} r_{t+1}^{B} \\
&= m_t + r_{t+1}^{A} - \tilde{\beta} r_{t+1}^{B}
\end{aligned}
\tag{2.5.4}
$$

which implies that

$$
m_{t+1} = m_t \quad \Leftrightarrow \quad r_{t+1}^{A} = \tilde{\beta} r_{t+1}^{B} \ .
$$

It is important to keep in mind that the estimated slope parameter and the constant now have a different interpretation as compared to the relation with price levels directly. When working with price levels, we have the advantage of a straight forward interpretation of the linear relationship. A conintegration vector $(1\ \beta)^{\top}$ just means to sell (buy) one share of A and buy (sell) β shares of asset B in order to get a tradable pair. The equilibrium relationship between the prices of the two assets A and B is then assumed to be $P_t^A = \alpha + \beta P_t^B$. This simple interpretation is, unfortunately, not true in case we transform the price time series by using the logarithm. Here, a conintegration vector $\left(1\ \tilde{\beta}\right)^{\top}$ implies an equilibrium relationship of the form

$$
P_t^A = c\left(P_t^B\right)^{\tilde{\beta}}
$$

with $c = e^{\tilde{\alpha}}$. This means in fact, that the c acts as the β of the untransformed price equation. With $\tilde{\beta} = 1$ we have a linear relationship in the two prices, too, and we need to buy (sell) one share of asset A and to sell (buy) c shares of asset B in order to obtain a tradable pair. The equilibrium spread value is here implicitly assumed to be zero. Working with log-transformed prices gives us, however, the advantage of considering non-linear relationships too, by allowing $\tilde{\beta}$ to be different from 1.

An important question is, thus, whether it is reasonable to assume a zero equilibrium value of the spread process. The case of $\alpha \neq 0$ actually means that holding a portfolio according to (2.5.3) must return in expectation a non-zero cash flow for the portfolio holder. In particular, one has to pay a premium for holding asset A over an equivalent position in asset B if $\alpha > 0$ and obtains a premium for such positions if $\alpha < 0$. Now, why should some securities trade at a premium or discount, respectively? Vidyamurthy (2004) argues that such differences could occur, for example, in cases where one asset is more liquid than the pair asset, or when one of the pair assets is a takeover target. He also sees the chance that pure charisma on the part of some assets can justify a non-zero α. For some readers such explanations may look plausible, for others not. Probably, there are for any explanation attempt also some stronger or weaker counter arguments. So, one could argue with respect to the liquidity premium argument that pairs trading works anyway only in highly liquid markets, so that there are hardly any liquidity differences that would justify such a premium. With respect to the takeover premium argument, one could say that a company that is a takeover target is in an exceptional state, which makes any past pricing information of that company unreliable in terms of future developments. As a pairs trader one would usually exclude such companies as soon as there is any rumour in the market about the takeover.

For the remaining part of this work, the spread process is always assumed to have a zero mean value. In addition, to capture potential non-linear price relationships, it is reasonable to base the analysis on log-transformed price series according to (2.5.1).

Chapter 3

Price and Spread Generating Processes

Before we start analysing different aspects of cointegration based pairs trading strategies, it will be beneficial to get a sound understanding of the assumed underlying data generating processes. This is very important as our model based approach relies on certain assumptions. Understanding the data generating process helps assessing whether the assumptions made are plausible or not. In addition, it makes us aware of potential problems and general weaknesses of the selected model.

This chapter discusses, thus, the necessary basics and lays the ground for the analyses in subsequent chapters. We start with a brief discussion about popular continuous time processes and will then lead over to more specific discrete time processes that are based on the discussed processes but are better suited for our needs. The introduction to the basic models is basically taken over from Harlacher (2012) Chapter 3.

3.1 A Popular Price Process for Common Stocks

The price behaviour of a publicly traded financial asset is usually considered as a stochastic process. For a particular asset it is often reasonable to assume that the generated returns are independent of the asset's price level. Hull (2009), for example, suggests in this respect a constant expected return. In particular, the return r of a riskless financial asset with price S_t at time t can be expressed as

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

with μ being a constant. By rearranging (3.1.1) to

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

and letting $\Delta t \longrightarrow 0$, we can write the expression above as

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

which is an ordinary differential equation. This equation can easily be solved, so that we get the solution

$$S(t) = S_0 e^{\mu t} \tag{3.1.4}$$

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

Expression (3.1.4) is obviously a smooth deterministic function in t that depends only on the starting value S_0 and the chosen parameter μ. In order to get a roughly realistic asset price process, we clearly need to add some randomness to that expression. A very simple way to do that is just employing an additive noise term which is dependent on the time interval Δt, i.e.

$$\frac{\Delta S_{t+\Delta t}}{S_t} = \mu \Delta t + \epsilon_{t,\Delta t} \; . \tag{3.1.5}$$

By introducing a random part, we need to define a proper probability model with a corresponding probability space $(\Omega, \mathcal{F}, P)$ so that we can state something about the probability distribution of our random term $\epsilon_{t,\Delta t}$. If we are willing to assume that $\epsilon_{t,\Delta t}$ is normally distributed with $E[\epsilon_{t,\Delta t}] = 0$ and $Var(\epsilon_{t,\Delta t}) = \sigma^2 \Delta t$, for any positive values of t and Δt, we will finally be able to express our process in closed form, which is clearly quite handy for many applications. So, we can write

$$\epsilon_{t,\Delta t} \stackrel{\mathrm{d}}{=} \sigma \sqrt{\Delta t} \; \varepsilon_t$$

with $0 < \sigma < \infty$ and

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

Translating our discrete time model now into a continuous time process by letting $\Delta t \longrightarrow 0$, we obtain

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

where we set

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

which is known as Wiener-process. The ω indicates that $S(t,\omega)$ is a random process defined on the filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P\right)$. It is, however, more convenient to omit the argument ω in subsequent discussions as it just makes the notation unnecessarily complicated.

The dynamics as described by the stochastic differential equation (SDE) (3.1.6) is known as *geometric Brownian motion*. This is a special case of a so-called Itô-process[1]. By taking $f : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}\ ,\ (X_t, t) \mapsto \log(X_t)$, we can derive a closed-form solution for (3.1.6), i.e.

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

with $S_0 = S(0) > 0$ and $W(t) \sim N(0,t)$. As the price of a common stock can only take on positive values, our price process should reflect that feature. As we can see, (3.1.7) can indeed only take on positive values. Figure 3.1 illustrates an example path of the process as given by (3.1.7) with $S_0 = 30$, $\mu = 0.03$ and $\sigma = 0.2$. At first sight, the depicted price path looks indeed like the trajectory of a non-dividend-paying stock. A small flaw in (3.1.7) is, regrettably, the fact that we assumed the disturbance term ϵ in (3.1.5) to be normally distributed, which implies at the same time that our process generates normally distributed continuously compounded returns. This is, to be honest, not very realistic, as the distributions of daily stock returns have typically more mass in the tails then the normal distribution. To account for

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

that fact we could, of course, adjust the disturbance distribution accordingly. Unfortunately, abandoning the normal distribution would then also mean losing many of the nice analytical properties of the Itô-process. If we stay, however, with the normal distribution, we can also derive handy expressions for the (conditional) expected value and the variance of the process.

The dotted green line in Figure 3.1 shows the conditional expectation of S_t, given S_0, which can easily be derived from (3.1.7). In particular, we get

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

which is the same expression as in (3.1.4). In addition, we can also derive an expression for the conditional variance, which 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

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

Now, plugging this together with (3.1.8) in (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 two expressions (3.1.8) and (3.1.10) it is obvious that such a process is non-stationary, i.e. neither the expected value nor the variance converges to a finite value when t becomes large. In particular, we have

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

for any $S_0 > 0$.

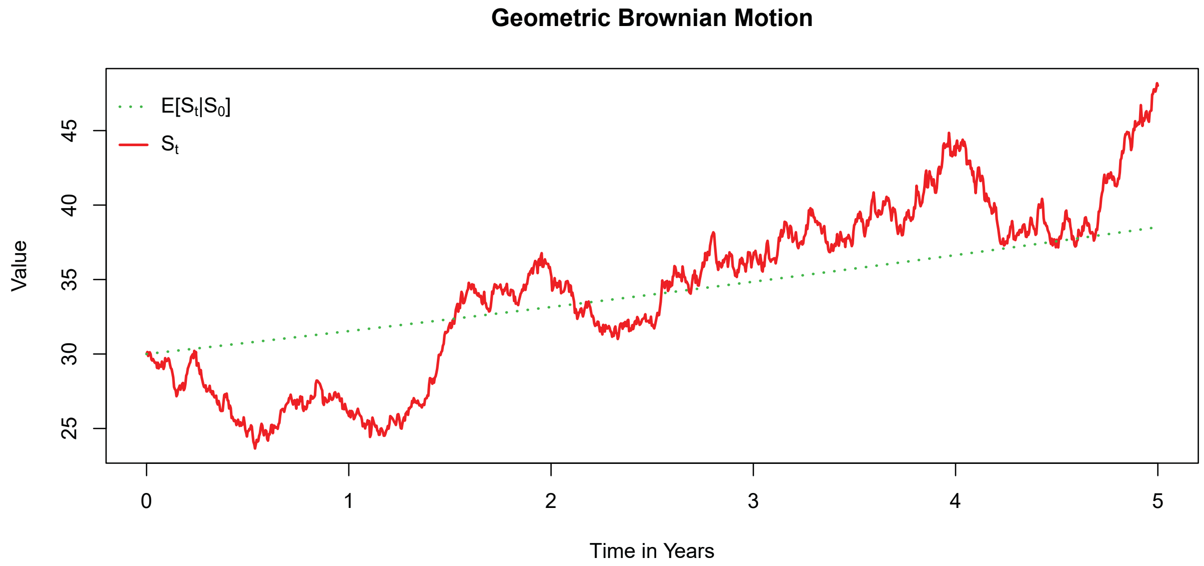

Figure 3.1: A typical sample path of a Geometric Brownian Motion.

3.2 A Spread Process

In terms of pairs trading, we are certainly not only interested in stochastic processes that are appropriate to describe the price process of stocks and similar assets, but also in processes that are suitable to describe the spread between two pair-assets. On the one hand, such a process should probably share many of the stochastic features of the process as described in Section 3.1. On the other hand, there must clearly be some crucial differences. The spread process of two pair-assets is usually assumed to fluctuate around a constant mean. We should, therefore, look as well at mean-reverting processes with a finite mean.

In order to obtain an suitable spread-process, it is probably not the worst idea to take (3.1.6) as a base and look for appropriate adjustments. We have seen in Section 3.1 that the conditional expectation of the process was determined by the first part of (3.1.6). Modifying that first part to some extent, allows us to rewrite it as

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

with $\kappa \geq 0$, $\theta \in \mathbb{R}$ and $dW(t)$ denoting again a Wiener-process. This expression already allows us to see the mean-reverting behaviour as the process is pulled downwards if the value of the process lies above the parameter θ, i.e.

the drift term becomes negative. On the other hand, the process is pushed upwards if $X_t < \theta$. Instead of applying $\sigma X(t)dW(t)$ as diffusion component, it is for many financial applications more appropriate to use $\sigma\sqrt{X(t)}dW(t)$ and restrict θ to positive values. 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 a so-called *square-root diffusion process*, which is a very popular process in short term interest rate modelling[2]. A major drawback of (3.2.2) is the fact that it cannot be solved in closed-form.

Nevertheless, the conditional expectation of X_t, given a starting value X_0, can be derived via the integral form of the corresponding stochastic differential equation

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

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, which is 0.

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 be quickly solved by using the superposition principle. It follows that

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

with

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

[2]See for example the work of Cox, Ingersoll, and Ross (1985).

for all $X_0 > 0$. For the unconditional expectation we can state accordingly

$$E_{t_0}\left[X(t)\right] = E_{t_0}\left[E_{t_1}\left[X(t) \mid X(t_1)\right]\right] \xrightarrow{t\to\infty} \theta \tag{3.2.8}$$

for any point in time t with $t_0 < t_1 < t$. So, for any given starting value X_0, (3.2.2) drags any subsequent value towards its mean level θ. The mean-reversion speed is thereby controlled by the parameter κ. The parameter σ is still related to the disturbance of the process, as in (3.2.1).

Focussing a bit closer on the disturbance part of (3.2.2) reveals, however, an important feature of the process that makes it inappropriate to model the kind of spread process we have in mind. As long as the disturbance term depends on the current level of the process, the process cannot cross the zero line as the diffusion term gets zero as soon as X_t goes towards zero. According to our discussion in Chapter 2, we need however a process that is able to take on positive as well as negative values, fluctuating around a long-term mean that is likely to be zero. So, an easy way to adjust process (3.2.2) to our needs is just by omitting X_t in the disturbance term and letting all the rest untouched. By this, we get

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

with the same interpretations of the parameters, but with a diffusion term that is now independent of the current value $X(t)$. The solution of the dynamics described by (3.2.9) (often with $\theta = 0$) is called *Ornstein-Uhlenbeck* process. As we have just made a small change in the disturbance part, without changing anything in the mean-reversion part, the conditional and unconditional expectation of (3.2.9) is still the same as for the square-root diffusion process. By using Itô's lemma together with $f : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$, $(X_t, t) \mapsto X_t e^{\kappa t}$ we obtain

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

So, we can write

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

This expression allows us to 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 easily derived, given (3.2.10). In particular, we get

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

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

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

For any particular value of X_0 and $t > 0$, the distribution of X_t is given by

$$X_t \sim N\left(\theta - \left(\theta - X_0\right) e^{-\kappa t}, \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa t}\right)\right) \, . \qquad (3.2.13)$$

Figure 3.2 depicts an example path with $X_0 = 0.7$, $\theta = 0$, $\kappa = 15$ and $\sigma = 0.2$. According to (3.2.13), we can also depict the path of the expected value (green dotted line) together with the bounds of a 95% confidence interval (blue dashed lines).

3.3 More General Processes

The processes described in the previous sections are well known processes that lie at the centrepiece of many pricing theories. It is, of course, also

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

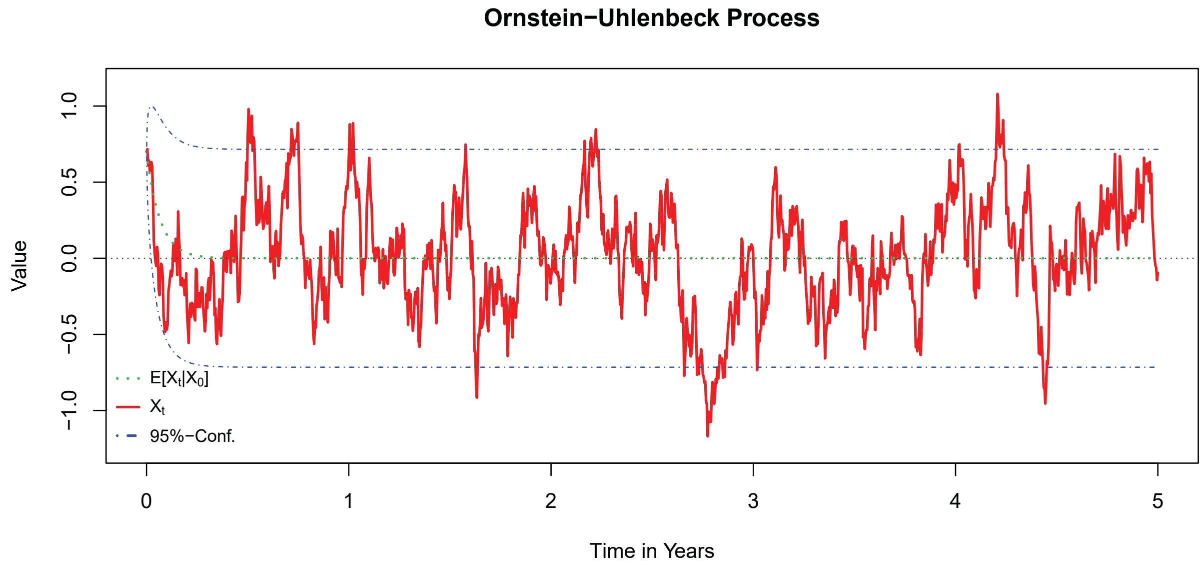

Figure 3.2: A typical sample path of an Ornstein-Uhlenbeck-process.

possible to combine them in order to obtain more general processes. For example, to account for heteroskedasticity in stock prices over time, we could use a slightly modified version of (3.1.6) in the sense that we replace the fix disturbance parameter σ by a time varying process $\sqrt{\nu_t}$. This is

$$dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dW_t \ . \tag{3.3.1}$$

The process ν_t could then be modelled, for instance, by a square-root diffusion process like

$$d\nu_t = \kappa\left(\theta - \nu_t\right) dt + \sigma\sqrt{\nu_t} dW_t \ . \tag{3.3.2}$$

Such a stochastic volatility model, i.e. combining (3.3.1) with (3.3.2), is used, for example, in the *Heston model* proposed by Heston (1993).

Obviously, a more complex process dynamics may better reflect the features we actually observe in real financial markets. However, more complex processes have hardly ever simple closed form solutions. This actually means that we have to evaluate approximations of such processes by numerical methods. At the same time we can also say that asset prices are anyway only observable in discrete time. A transaction takes place only when the ask price meets the bid price. It is, therefore, worth taking the next step to move from continuous time processes to discrete time processes.

An easy but powerful way to obtain discrete time processes from the previously discussed continuous time processes, is by applying the so-called *Euler* scheme. It is a scheme that works quite generally. For example, the discrete time process $\left\{\hat{X}_h, \hat{X}_{2h}, \ldots, \hat{X}_{mh}\right\}$, $h \in (0, +\infty)$, consisting of m time steps and generated by

$$\hat{X}_{th} = \kappa\theta h + (1 - \kappa h)\,\hat{X}_{(t-1)h} + \sigma\sqrt{h}\zeta_t \tag{3.3.3}$$

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

$$\hat{X}_t = \kappa\theta + (1 - \kappa)\,\hat{X}_{t-1} + \sigma\zeta_t \tag{3.3.4}$$

which is a simple AR(1) process. By considering the characteristic equation of (3.3.4), i.e. $z - (1 - \kappa) = 0$, we see that we have to restrict the parameter κ to the interval $0 < \kappa < 2$ in order to have the root of the characteristic equation to lie inside the unit circle. This is a necessary and sufficient condition for an AR(1) process to be weakly stationary. In our case, it is actually sufficient to require that $-1 < z < 1$, as we can only obtain one real valued root here. This implies, at the same time, that $(1 - \kappa)$ must lie between -1 and 1. As $(1 - \kappa)$ corresponds to the AR coefficient, the restriction $0 < \kappa < 2$ does, however, not guarantee a realistic behaviour for the spread process we have in mind. A negative AR coefficient seems less realistic as this suggest a very jumpy process. It is therefore reasonable to restrict the AR coefficient to only positive values, which actually means to restrict the admissible values of κ to the interval $0 < \kappa < 1$.

Chapter 4

Testing for Cointegration

In section 2.3 we briefly introduced the concept of cointegration. Even tough there are mean-reverting long-short trading schemes that do not explicitly test for cointegration, but instead have their own way to determine mean-reverting asset combinations, many statistical arbitrage strategies are implicitly based on the cointegration concept with relationships that are assumed to hold at least temporarily. The reason why such schemes do not explicitly test for cointegration might be explained by the considerable weaknesses of many popular tests.

Before we examine in chapter 5 some general features of these tests and analyse what some deviations from the standard assumptions imply for the test results, we briefly introduce the most frequently used residual based cointegration test procedures and state the required time series definition for clarity in this chapter.

The reason why we limit ourselves to residual based tests is actually because they are more intuitive. Many experienced statisticians would probably even tell us that it is better to just have a look at a times series plot of the residuals instead of performing a cointegration test in order to take the decision of whether some variables can be assumed to be cointegrated. Indeed, this is probably not the worst idea if we are experienced enough and if we just have to check a few time series. In an algorithmic trading scheme, as the one proposed in this work, where we usually have to consider several hundreds or even thousands of linear combinations, this will not be feasible without inserting some powerful restrictions. We will discuss later in Chapter 6 some restrictions that could possibly be applied to limit the amount of combinations to be tested to a set of most promising matches. However, as long as we want to let a computer take the decision, we need a well defined criterion based on which the computer is able to decide. For this purpose, cointegration test procedures may be well appropriate.

4.1 Preliminaries

Similarly to our discussion in Chapter 3 about stochastic processes, we define a discrete time series as a set of random variables. This is

$$\{X(\omega, t), \omega \in \Omega, t \in \mathbb{N}\} \tag{4.1.1}$$

with sample space Ω. The general time series definition is broader than that of (4.1.1) and not limited to just equidistant points in time at which the process can be observed. The time series definition according to (4.1.1) will, however, be sufficient for our purpose. In the subsequent notation we will, in addition, just write X_t in order to avoid a too messy notation.

An observed time series $\{x_1, x_2, \ldots, x_T\}$ is, thus, just one single realisation of the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_T)^\top$ with multivariate distribution function $G_T(\boldsymbol{x})$. With this in mind, we can now define the term "stationarity", which we have already previously used in this work. We are becoming now a bit more precise with this term. It is one of the key elements we will refer to throughout the whole further discussion. We distinguish in general between *strict stationarity* and *weak stationarity*.

Definition 4.1.1. *A time series is said to be strictly stationary, if the joint distribution of $X_t, X_{t+1}, \ldots, X_{t+k-1}, X_{t+k}$ corresponds to the joint distribution of $X_s, X_{s+1}, \ldots, X_{s+k-1}, X_{s+k}$ for every combination of t, s and k.*

Unfortunately, regarding practical applications, we will never be able to check this kind of stationarity based on a data sample. Therefore, one uses in general a weaker form of stationarity definition.

Definition 4.1.2. *A time series is said to be weakly stationary, if*

i.) $E[X_t] = \mu < \infty$, *for all* $t \in \mathbb{N}$

ii.) $E[(X_t - \mu)(X_{t+h} - \mu)] = \gamma(h)$, *for all* $t \in \mathbb{N}$, $h \in \mathbb{N}_0$

Weak stationarity is sometimes also referred to as *second-order stationarity* or *covariance stationarity.* It is this kind of stationarity we mean when we talk about stationarity throughout this analysis.

Another often used term is the "order of integration". It is a summary statistic for a time series, denoted $I(d)$, that reports the minimum number of differences required in order to obtain a weakly stationary series.

Having established all the prerequisites, we can now formulate the definition of cointegration.

Definition 4.1.3. *The components of a vector* $\boldsymbol{x}_t$ *are said to be cointegrated of order d, b, denoted* $CI(d, b)$*, with* $b > 0$*, if*

i.) all components of $\boldsymbol{x}_t$ *are integrated of order d, denoted* $I(d)$*, and*

ii.) a vector $\boldsymbol{\beta} \neq \mathbf{0}$ *exists such that* $\boldsymbol{\beta}^\top \boldsymbol{x}_t$ *is integrated of order d-b, i.e. I(d-b).*

The vector $\boldsymbol{\beta}$ *is called cointegrating vector.*

Price series of many financial assets are usually assumed to be $I(1)$. The relevant cases for us are, thus, cointegrated series of the type $CI(1, 1)$. Hence, Engle and Granger (1987) proposed a simple two step procedure to check whether two price series can be assumed to be cointegrated. In a first step, they suggest to estimate a regression by OLS and then to test the residual series for weak stationarity by employing the augmented Dickey-Fuller (ADF) unit root test. Hence, as long as we deal with $I(1)$ price series, cointegration is related to the concept of stationarity. The kind of stationarity test we may use in such a two step procedure is obviously not limited to the ADF test. So, we can use any kind of stationarity test suitable for our purpose. As we will see in the subsequent sections of this chapter, different tests have different features and are, thus, more or less appropriate for our purpose, depending on the assumptions we reasonably have to make in our case.

The tests we are going to discuss in the next few sections are all residual based tests. There are, of course, many other more or less established tests available. To consider all of them would very likely cover enough material to write a whole dissertation. We, therefore, have to limit ourselves to some of them and choose the probably most common ones that differ in some for our purpose relevant features. The pioneering augmented Dickey-Fuller test may be a good initial point to start with. We will then briefly discuss the Phillips-Perron, the KPSS and the Phillips-Ouliaris test.

4.2 The Augmented Dickey-Fuller Unit Root Test

The ADF test according to Said and Dickey (1984) is basically an extension of the *Dickey-Fuller* (DF) unit root test initially proposed by Dickey and Fuller (1979). While the DF test is only valid for time series that are characterised by an AR(1) process with a white noise error term, the ADF test can accommodate general ARMA(p,q) processes of unknown order. The null hypothesis assumes that the series to be tested is $I(1)$. This is also the reason why it is called *unit root test*. The corresponding test statistic is obtained by first estimating the following regression[1]

$$\Delta x_t = \beta_0 + \beta_1 t + (\phi - 1)\, x_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta x_{t-j} + u_t \tag{4.2.1}$$

where u_t is assumed to be white noise. It is important to mention that the distribution of u_t is not limited to a normal distribution. It can have any type of distribution. β_0 and β_1 should be set to 0 if there is no reason to assume the existence of deterministic parts like a constant and / or a time trend in (4.2.1). Of course, it raises the question whether it makes sense to consider deterministic terms as long as we are just interested in employing the test in terms of cointegration testing where we usually hand over the residuals of the regression of the first step. The residuals of an OLS regression are by construction a demeaned series. As we will see later in this chapter, the distribution of the test statistic depends on the inclusion of deterministic terms, like a constant and / or a time trend. When we use the ADF test in the context of cointegration testing, we will, however, include a constant and / or time trend in the first step of the Engle-Granger procedure, i.e. in the OLS regression step. For example, if we include a constant term α in the regression

$$y_t = \alpha + \beta x_t + \epsilon_t \tag{4.2.2}$$

[1] The regression is usually estimated by OLS.

we will then take the residuals of (4.2.2) and use the ADF test without any deterministic terms, i.e.

$$\Delta\widehat{\epsilon}_t = (\phi - 1)\,\widehat{\epsilon}_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta\widehat{\epsilon}_{t-j} + u_t \ . \tag{4.2.3}$$

This is because the constant term β_0 of (4.2.1) is already considered by the α in (4.2.2). Even though we do not include a constant term in the actual test equation (4.2.3), from which we finally obtain the test statistic, we nonetheless have to use the critical values according to the test statistic distribution that includes a constant term. The same logic holds also true for any other deterministic term, like a linear (or if required) quadratic time trend. Therefore, it makes sense to mention the deterministic parts in the test formulation, albeit they will never be included in the test equation but rather in the OLS regression expression. As already mentioned, the handling of the deterministic parts is very crucial as these parts influence the asymptotic distribution of the test statistic.

It is worth mentioning at this point that many unit root or stationarity tests allow to calculate more than just one test statistic. It is, however, not so clear to which extend, if at all, the different test statistics of a particular test really differ from each other in terms of rejecting or not rejecting the null hypothesis. Hence, for the tests we discuss here, we stick to the test statistic that is related to the t-statistic known from ordinary linear regression. The test statistic t_{ADF} is then obtained by

$$t_{ADF} = \frac{\widehat{\phi} - 1}{SE\left(\widehat{\phi}\right)} \tag{4.2.4}$$

which is actually calculated in the same way as the usual t-statistic in an ordinary regression. The hypotheses are:

$$H_0 : \phi = 1 \tag{4.2.5}$$

$$H_1 : -1 < \phi < 1 \ . \tag{4.2.6}$$

Under H_0, the distribution of the test statistic t_{ADF} is, however, not t-distributed and does neither converge to a standard normal distribution. Asymptotically, t_{ADF} has a so-called *Dickey-Fuller* distribution, which depends on the specifications of (4.2.1) in terms of the chosen deterministic parts, and has no closed form representation. For example, if we assume a constant but no time trend in (4.2.1), the limiting distribution of t_{ADF} is given by

$$t_{ADF} \xrightarrow{\text{d}} \frac{\frac{1}{2}\left(W(1)\right)^2 - \frac{1}{2} - W(1)\int_0^1 W(r)dr}{\left(\int_0^1 \left(W(r)\right)^2 dr - \left(\int_0^1 W(r)dr\right)^2\right)^{\frac{1}{2}}} \tag{4.2.7}$$

where $W(r)$ is a *Wiener-process* as discussed in Chapter 3. Even though the corresponding critical values according to (4.2.7) can easily be obtained by simulations, there is a less time consuming way provided by MacKinnon (1996), which is called *response surface method*. An extensive discussion about the DF distribution including the derivation of (4.2.7) can be found in Hamilton (1994) (Chapter 17).

The ADF test is a one-sided left-tailed test. Figure 4.1 shows the lower tail of a DF distribution according to (4.2.7) compared to the lower tail of a standard normal distribution. It becomes clear that if we employed the standard normal distribution for t_{ADF}, we would reject H_0 far too often. The standard normal distribution is not even close to the DF distribution.

Figure 4.1: Distribution of the ADF/PP Teststatistic.

An additional fact we have to consider in our particular case, i.e. when we want to use this test to check for cointegration, is that we do not just take an observable time series and test whether it has a unit root or not. We use the ADF test in a two step procedure on the residuals of a regression. In particular, the problem we face is that the time series we test for a unit root is an estimated series based on the residuals of the OLS regression of the first step. This requires an adjustment of the critical values of the test statistic. Phillips and Ouliaris (1990) provide an asymptotic theory for this case with the corresponding limiting distributions. The distributions are similar to (4.2.7) in the sense that they can also be characterised by Wiener-processes.

Figure 4.1 also shows the distribution of the ADF test statistic under the null hypothesis when the tests are used on the residuals of a cointegration regression including a constant term but no deterministic time trend. We can, clearly, spot a considerable difference to the distribution that applies in cases of directly observable time series.

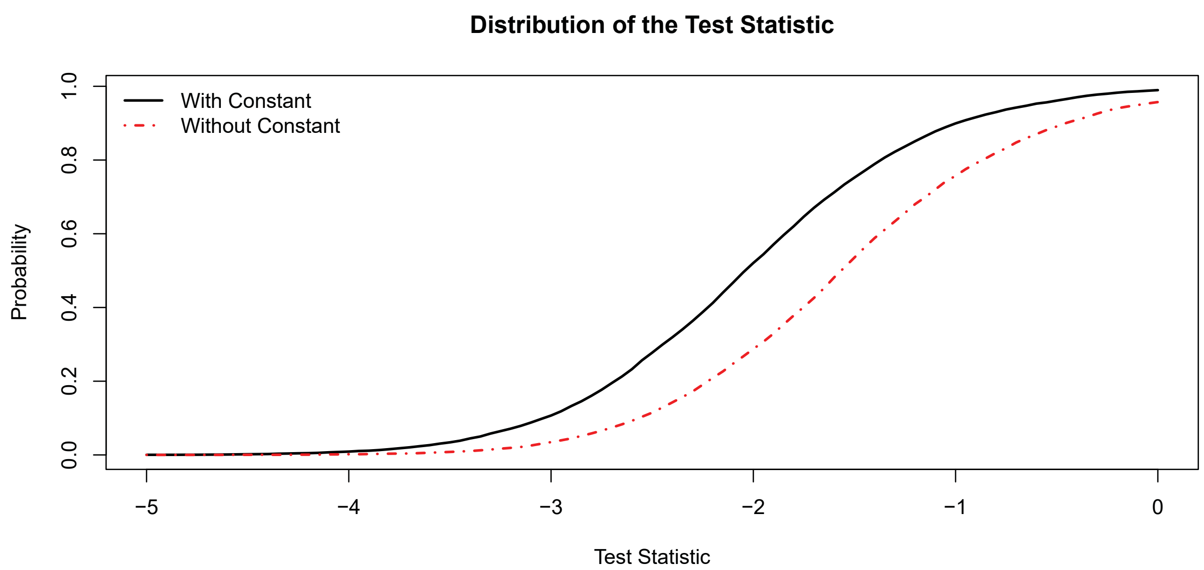

Figure 4.2: Distribution of the ADF/PP Test Statistic with and without Constant.

Figure 4.2 shows the distribution of the ADF test statistic applied in the context of cointegration testing and highlights the difference in the distributions when a constant term is included or excluded[2].

By choosing an appropriate value of the lag terms k in (4.2.1), the ADF test

[2]As already mentioned, in the cointegration testing the constant is included in the regression part.

can deal with any kind of ARMA(p,q) structure in the residual series. The question is, thus, how we best choose a value for k. On the one hand, if we do not include enough lags in (4.2.1), the remaining serial correlation in the residual series will bias the ADF test. On the other hand, if we choose to include too many lags, the power of the test will suffer. Ng and Perron (1995) propose to estimate (4.2.1) first with a particular maximum number k_{max} of lag terms and then having a look at the t-statistic of the last lag term. If it has an absolute value of more than 1.6, then k_{max} should be the number of lag terms to be used. If the absolute value of the corresponding t-statistic is smaller (or equal to) 1.6, then the lag length should be reduced by one and the whole procedure should be repeated again. To determine the upper bound of lag terms, Schwert (1989) suggests as a rule of thumb to take

$$k_{max} = \left[12 \left(\frac{T}{100} \right)^{\frac{1}{4}} \right] \tag{4.2.8}$$

where $[\cdot]$ indicates in this case the integer operator returning the integer part of the corresponding argument.

As an alternative, one can also think of taking the AIC or the BIC to determine the lag length k.

4.3 The Phillips-Perron Unit Root Test

Another unit root test, with the same hypotheses as in the ADF test, is the one proposed by Phillips and Perron (1988), which is known under the name Phillips-Perron (PP) test. This kind of test is particularly popular in the analysis of financial time series. This is mainly because of the way the test deals with serial correlation and heteroskedasticity in the errors. In particular, the PP test regression is given as follows

$$x_t = \beta_0 + \beta_1 t + \rho x_{t-1} + v_t \tag{4.3.1}$$

where v_t is the error term. The error term is assumed to have a constant mean of zero and a constant correlation structure. It is, however, allowed to be heteroskedastic. Again, the two parameters β_0 and β_1 are set to zero if there are good reasons to assume that there is no constant term and / or deterministic time trend. The serial correlation and the potential heteroskedasticity in v_t is then corrected by modifying the test statistic t_{PP}. The modified PP test statistic z_t can finally be obtained as follows

$$z_{PP} = \sqrt{\frac{\widehat{\gamma}_0^2}{\widehat{\lambda}^2}} \frac{\widehat{\rho} - 1}{SE\left(\widehat{\rho}\right)} - \frac{1}{2} \frac{\widehat{\lambda}^2 - \widehat{\gamma}_0^2}{\widehat{\lambda}^2} \frac{SE\left(\widehat{\rho}\right) T}{\widehat{\gamma}_0^2} \tag{4.3.2}$$

with

$$\widehat{\gamma}_j^2 = \frac{1}{T} \sum_{i=j+1}^{T} \widehat{v}_i \widehat{v}_{i-j} \tag{4.3.3}$$

and

$$\widehat{\lambda} = \widehat{\gamma}_0^2 + 2 \sum_{j=1}^{q} \left(1 - \frac{j}{q+1}\right) \widehat{\gamma}_j^2 \ . \tag{4.3.4}$$

Expression (4.3.4) is just the heteroskedasticity and autocorrelation consistent estimator of the error variance with a chosen number of lags q according to Newey and West (1987). It is, in principle, not the only possible choice for $\widehat{\lambda}$, but the most commonly used. In the case where there is no serial correlation in the error term, (4.3.2) boils down to the usual t-statistic.

Under the null hypothesis, i.e. $\rho = 1$, the Phillips-Perron test statistic z_{PP} has the same asymptotic distribution as the ADF t_{ADF}-statistic. Like with the ADF test, the PP test is a one-sided left-tailed test too. When the test is applied on the residuals of an OLS regression in the sense of a two step cointegration test, Phillips and Ouliaris (1990) show that the limiting distribution is the same as the one we have to consider when we use the ADF test to test for cointegration. Regarding financial time series, the PP test has the advantage of being robust to general forms of heteroskedasticity in the error term of (4.3.1). In addition, there is no need to determine the number of lagged differences k, which can have a significant effect in the ADF test. There is, however, still the need to choose an appropriate number q when estimating the Newey-West heteroskedasticity and autocorrelation consistent error variance. This choice should, however, not have a major effect. As with the ADF test, there are no restrictions on the distribution of the error term v_t.

4.4 The KPSS Stationarity Test

The KPSS test according to Kwiatkowski, Phillips, Schmidt, and Shin (1992) can similarly like the ADF and the PP test be used to test for cointegration. Unlike the ADF and the PP test, the KPSS test is actually a stationarity test, meaning that the null hypothesis assumes a stationary series. The corresponding test statistic is derived by the following model

$$\begin{aligned} x_t &= \beta t + r_t + \epsilon_t \\ r_t &= r_{t-1} + u_t \end{aligned} \tag{4.4.1}$$

where u_t is assumed to be white noise, satisfying $E[u_t] = 0$ and $Var(u_t) = \sigma_u^2$. As with the other tests, previously discussed, the deterministic time trend with β is optional, i.e. it should be put to zero if there are no reasons to assume such a trend. The initial value r_0 takes here the role of a deterministic constant, if necessary. The error term ϵ_t is assumed to fluctuate around zero with a constant correlation structure. Similar to the error term of the PP test regression, ϵ_t is allowed to be heteroskedastic. Obviously, the second equation of (4.4.1) is a pure random walk. As already mentioned, the null hypothesis assumes the process $\{x_t\}$ to be stationary, this is

$$H_0 : \sigma_u^2 = 0 \tag{4.4.2}$$

$$H_1 : \sigma_u^2 > 0 \ . \tag{4.4.3}$$

So, under the null hypothesis r_t is just a constant.

The test statistic t_{KPSS}, sometimes also referred to as *score statistic* or *Lagrange Multiplier* (LM), is given as follows:

$$t_{KPSS} = \frac{\sum\limits_{t=1}^{T} \widehat{S}_t^2}{T^2 \widehat{\lambda}_\epsilon^2} \tag{4.4.4}$$

where

$$\widehat{S}_t^2 = \sum_{j=1}^{t} \widehat{\epsilon}_j \tag{4.4.5}$$

and $\widehat{\lambda}^2$ is again a autocorrelation and heteroskedasticity consistent estimator of the variance of ϵ_t like, for instance, the Newey-West estimator according to (4.3.4). Under H_0 the distribution of the test statistic depends on the inclusion of deterministic parts like a constant and / or a time trend. For $\beta = 0$, i.e. without deterministic time trend, and $r_0 = 0$, we have

$$t_{KPSS}^{(0)} \xrightarrow{\mathrm{d}} \int_0^1 W^2(r)\, dr \tag{4.4.6}$$

where $W(r)$ denotes a Wiener process according to our discussion in Chapter 3.

In the case where we assume $r_0 \neq 0$, but still have $\beta = 0$, i.e. assuming no time trend but a non-zero constant, the limiting distribution of the test statistic (4.4.4) is

$$t_{KPSS}^{(r_0)} \xrightarrow{\mathrm{d}} \int_0^1 V^2(r)\, dr \tag{4.4.7}$$

with $V(r) = W(r) - rW(1)$, which is a so-called *Brownian Bridge.*

Finally, if we include a constant and a time trend, i.e. $r_0 \neq 0$ and $\beta \neq 0$, the distribution to consider in this case is

$$t_{KPSS}^{(r_0,\beta)} \xrightarrow{\mathrm{d}} \int_0^1 V_2^2(r)\, dr \tag{4.4.8}$$

with

$$V_2(r) = W(r) + \left(2r - 3r^2\right) W(1) + 6r\,(r-1) \int_0^1 W(s)\, ds\ . \tag{4.4.9}$$

The distributions according to (4.4.6) - (4.4.8) can, again, be obtained by simulation. The KPSS test is a one-sided right-tailed test, which means that we reject the null hypothesis at the α % error probability if the corresponding test statistic t_{KPSS} exceeds the $(1-\alpha)$ % quantile of the appropriate distribution.

Similar to the case of the ADF and the PP test, if we apply the KPSS test on the residuals of an OLS regression in terms of a two step cointegration test, the corresponding critical values have to be adjusted. This is, the limiting distributions are given in terms of mixtures of Brownian bridges and vector Brownian motions. Shin (1994) provides a table with corresponding critical values.

4.5 The Phillips-Ouliaris Test

The tests discussed so far are all well established. With respect to applications in an algorithmic trading scheme they have, however, the disadvantage of not being invariant to the formulation of the regression equation in step one. This means, the outcome of the test depends on the chosen normalisation of the regression equation. In the case of a pairs trading scheme with two assets x and y, we would have to run $y_t = \alpha + \beta x_t + \epsilon_t$ as well as $x_t = \alpha + \beta y_t + \varepsilon_t$. With an asset universe of 100 securities this means that we have to run $\frac{100!}{98!} = 9,900$ tests. As we will discuss in more detail in Chapter 6, the number of tests to be run should be limited. Phillips and Ouliaris (1990) proposed a test procedure in their work that is independent of the chosen normalisation. It is based on a first order vector autoregressive model (VAR) of the form

$$\boldsymbol{z}_t = \hat{\Pi}\boldsymbol{z}_{t-1} + \hat{\boldsymbol{\xi}}_t \ . \tag{4.5.1}$$

The test statistic is then calculated as

$$\hat{P}_z = T \cdot tr\left(\hat{\Omega} M_{zz}^{-1}\right) \tag{4.5.2}$$

with

$$\hat{\Omega} = \frac{1}{n}\sum_{t=1}^{n} \hat{\boldsymbol{\xi}}_t \hat{\boldsymbol{\xi}}_t^\top + \frac{1}{n}\sum_{s=1}^{l} \omega\left(s,l\right) \sum_{t=s+1}^{n} \left(\hat{\boldsymbol{\xi}}_t \hat{\boldsymbol{\xi}}_{t-s}^\top + \hat{\boldsymbol{\xi}}_{t-s} \hat{\boldsymbol{\xi}}_t^\top\right) \tag{4.5.3}$$

and

$$M_{zz} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{z}_t \boldsymbol{z}_t^\top \ . \tag{4.5.4}$$

The weighting function $\omega\left(s,l\right)$ is suggested to be $1 - \frac{s}{l+1}$, which depends on a user chosen lag window l.

Again, the limiting distributions of the test statistic under H_0, which can again be expressed as mixtures of a Brownian bridges and vector Brownian motions, are only obtained by simulation. A table with critical values can be found in Phillips and Ouliaris (1990).

Chapter 5

Test Analytics

In Chapter 4 we briefly discussed some well known residual based tests which can be employed to test for cointegration. An interesting question is now, how suitable these tests are for our purpose. Even tough there is a whole asymptotic theory for the different test statistics, we are probably not only interested in the type I error (or false negative rate) but also in the type II error (also known as false positive rate) and, thus, the power of the considered tests. This is a well known weakness of basically all existing cointegration tests, including the ones discussed in Chapter 4. Certainly, the specificity of a statistical test depends on the true data generating process. Intuitively, we would assume that the closer the true process is to a unit root process, the more difficult it becomes for a statistical test to come up with the correct decision. It is also clear that there are other influencing factors, such as a finite length of the analysed time series or the existence of heteroskedasticity, for example, that make it even more difficult for such tests to live up to the user's expectations.

In this chapter we will, thus, briefly check how the different tests, we have discussed so far, behave under different assumptions with respect to the data generating process. We should always keep in mind that we just need a mean-reverting process in order to successfully trade on pairs. If the remaining requirements of a stationary process are given as well, everything gets clearly easier, especially in terms of finding an effective trading rule, but we have to keep in mind that they are not essential for our purpose.

5.1 The Sensitivity and Specificity of Tests

As already mentioned, in the proposed algorithmic pairs trading scheme of this work, we use the conintegration test as a classifier in order to take a decision about the tradability of a particular pair of assets. So, it is important for us to understand how well the employed classifier actually works under different settings.

		Test Decision	
		H_0	H_1
True	H_0	Sensitivity $1 - \alpha$	Type I Error α
	H_1	Type II Error β	Specificity $1 - \beta$

Table 5.1: The Sensitivity and Specificity of Statistical Tests

Table 5.1 shows the 4 possible cases we face when we analyse the performance of a statistical test, where α denotes the true probability of committing a type I error. The probability of a type II error is denoted as β. Usually, the critical value of the test statistics, which we take as a decision criterion, is set according to a predefined value of the type I error probability α. Typically, we chose here values of 1% or 5%. By setting the critical value, the type II error probability is actually also set. It depends, however, on the distribution of the test statistic under H_1. In general, it is, of course, hard to say anything about the true data generating process. As we have already discussed in Chapter 3, there are some well established price generating processes in finance we may use in our case to generate the time series of our assets. A geometric Brownian motion or a process according to the Heston model is at least able to generate time series that look quite similar to what we actually observe in real markets. With respect to a potentially existing relationship between assets, it is clearly more challenging to come up with a suitable formulation. As we have also seen in Chapter 3, the Ornstein-Uhlenbeck process could be a valid candidate to model the spread series of two assets. Transformed in a daily observable discrete time series, it is, however, just a simple AR(1) process. On the one hand, one could clearly argue that an AR(1) process is not flexible enough to capture the true underlying dynamics. On the other hand, it is a sparse process with many desirable features and it is for sure a good point to start with.

Having determined the data generating process, we can examine the performance of a particular test by plotting the sensitivity against the specificity. A powerful test would show a curve that decreases only very slowly up to values close to 1 and drops to zero just close before 1. A weaker test starts to decrease faster. The minimum boundary for admissible tests would show a linearly decreasing curve with a slop of -1. This would mean that the test decides on pure randomness. So, the area under the curve should be as large as possible, i.e. it should be close to 1. Accordingly, we could define a quantity based on the area under the sensitivity/specificity curve as a criterion for the goodness of a test. One possibility is, for example,

$$Q := 2A - 1 \tag{5.1.1}$$

where A is the area under the sensitivity/specificity curve of a particular test with $Q \in [0, 1]$. This is the case because the area under the minimum boundary is 0.5.

Criterion (5.1.1) can, however, only make precise distinctions if the curves to compare are symmetric. Unfortunately, it does not take the shape of the curve into account. This is clearly an issue because we are primarily interested in small α-values and, therefore, focused on the right part of the plot. So, if the curves turn out to be very asymmetric, we could restrict the admissible area under the curve to the right part of the plot.

5.2 Test Performance under Different Price Dynamics

In order to analyse the performance of the different tests we introduced in Chapter 4, it makes sense to run simulations in three different settings. In all three settings we generate time series of 250 data points and employ 10,000 iterations. Determining the length of the time series appropriately is important as it clearly has an influence on the test performance. As already mentioned, Gatev et al. (2006) used a calibration period on market data of one year. The choice of using 250 data points is actually based on the same idea, which means that the 250 data points should roughly represent the number of trading days during one year. In order to obtain the desired sensitivity/specificity curve, we need to run our tests under the alternative hypothesis by using the critical values according to the test statistic distribution under the null hypothesis. In particular, for each chosen type I error probability α under H_0 we get a corresponding critical value, with which we can then classify each run into "H_0 rejected" or "H_0 not rejected" according to our cointegration test. The portion of the non-rejected tests is then taken as an estimate of the type II error probability β. As mentioned in Chapter 4, for the considered tests there are no closed form solutions with respect to the test statistic distributions under H_0. The corresponding distributions have to be simulated. So, in the case with cointegrated pairs[1], we imply the following relationship for two price series S_t^A and S_t^B:

$$\log\left(S_t^B\right) = \beta_0 + \beta_1 \log\left(S_t^A\right) + m_t \tag{5.2.1}$$

with m_t being an AR(1)-process. The m_t can be seen as the price series of a portfolio of two assets, with one asset *long* and the other asset *short*. Realistic values for β_0 depend on the relativity of the share denomination of the two assets under scrutiny. In particular, for two companies with the same share denomination, β_0 should reasonably be 0. For β_1 we probably best assume a value of 1, implying a linear relationship. The price series S_t^A and S_t^B can then be generated by a geometric Brownian motion according to (3.1.6). For the drift term we could take $\mu = 0.02$ and for the diffusion $\sigma = 0.25$, which corresponds to an annual expected return of 2% with an annual volatility of 25%. These values are just picked from a range of realistic values as they have been observed so far in real markets. Taking slightly different values, for example, a drift of 0.01 and a diffusion of 0.3, should not influence the results in a crucial way. For the spread series m_t, it makes sense to use

$$m_t = \gamma m_{t-1} + \kappa_t \tag{5.2.2}$$

with κ *i.i.d.* $N(0, 0.015)$. Here again, one can certainly argue about the "correct" standard deviation of κ_t. Time series of spreads the like we are looking at, are not directly observable. The variance of m_t depends in our setting also on γ. So, whether $Var(\kappa) = 0.015^2$ is a good choice, is a question of personal believes with regard to the true spread process. Our choice here leads at least to realistic looking time series and is, at the end of the day,

[1] This corresponds to H_0 in the KPSS test.

just a best guess assumption. With respect to the autoregressive behaviour, it makes sense to assume γ-values that lie in the interval $[0.9, 1]$, where every value of $\gamma < 1$ generates a cointegrated pair of assets suitable for pairs trading. Therefore, we run our simulation with 4 different γ-values, i.e. 0.9, 0.925, 0.95 and 0.975. In order to generate the setting of non-cointegrated pairs, we can either set $\gamma = 1$ or just take two independently generated processes according to (3.1.6).

Another issue worth mentioning concerns the way we generate the time series of cointegrated pairs according to (5.2.1). If $\log\left(S_t^A\right)$ and m_t are generated independently and we set $\beta_0 = 0$ and $\beta_1 = 1$ then $\log\left(S_t^B\right)$ will fluctuate systematically stronger than $\log\left(S_t^A\right)$ for every sample we create. There is, however, no plausible reason why this should be the case in real markets. In fact, both sample series should fluctuate roughly by the same amount, at least as long as the two assets have more or less the same share denomination. The crucial point here is that m_t is actually a term that encompasses a lot of different components that influence the value of S_t^B as well as the value S_t^A simultaneously. This is unlike the situation we usually desire in a linear regression. It seems odd to assume that $\log\left(S_t^A\right)$ and m_t are completely uncorrelated in our setting. The correlation structure is, however, not constant over the whole sample. We should keep in mind that $\log\left(S_t^A\right)$ is a non-stationary process while m_t is stationary as long as $\gamma < 1$. This makes it hard to employ a linear transformation in the usual way to generate the desired correlation structure without losing control over the set parameter values. If $\log\left(S_t^A\right)$ and m_t were both stationary autoregressive processes, it would be easier, though still challenging, to derive and implement an appropriate simulation framework. However, the asset prices we have in mind are hardly ever stationary. Therefore, carrying out a simulation study under such assumptions would most likely miss the point. The question is, thus, how we can implement a realistic simulation framework under non-stationary price time series. A heuristic solution unfolds by considering the procedure we apply when we test for cointegration with residual based tests as discussed in Chapter 4. In order to obtain the time series which we test for stationarity, we regress one price time series on another one. The relationship we employ for the regression is the same as the one given by (5.2.1). If we estimate the parameters $\widehat{\beta}_0$ and $\widehat{\beta}_1$ by OLS in a setting where m_t is not uncorrelated with

$\log\left(S_t^A\right)$, we will obtain biased parameter estimates. So, the idea is basically to use biased parameter values to generate the time series in the simulation. In particular, if we want the variance of $\log\left(S_t^A\right)$ be roughly the same as the variance of $\log\left(S_t^B\right)$, we will have to set $\widehat{\beta}_1$ to a value slightly smaller then 1. This implies a negative correlation structure between $\log\left(S_t^A\right)$ and m_t over the observed sample. The size of the bias is, as already mentioned, influenced by the correlation structure between $\log\left(S_t^A\right)$ and m_t, but also by the sample standard deviation of $\log\left(S_t^A\right)$ and m_t, and for $\widehat{\beta}_0$, the bias even depends on the sample means of $\log\left(S_t^A\right)$ and $\log\left(S_t^B\right)$. Following a heuristic approach, we obtain together with the already chosen parameters realistic looking processes by setting $\widehat{\beta}_0 = 0.02$ and $\widehat{\beta}_1 = 0.97$.

In order to generate the random numbers, a *combined multiple recursive generator* (CMRG) according to L'Ecuyer (1999) is probably the most appropriate one. It has a period of around 2^{191} and is well suited also for parallel implementations of the computing framework. For the generation of the autoregressive process (5.2.2), it is advisable to employ a *burn in period*. Hence, it is good practice to generate spread series of 350 data points and then to strip off the first 100 observations, leaving us with a random spread series of 250 data points. While an AR(1) process with $\gamma = 0.9$ looks typically very much like a stationary process when it is plotted, an AR(1) with $\gamma = 0.975$ looks very often already like a unit root process, i.e. a process with $\gamma = 1$. Referring again to Chapter 4, any value of $\gamma < 1$ corresponds to H_1 for the Dickey-Fuller, the Phillips-Perron as well as the Phillips-Ouliaris test, but to H_0 for the KPSS test. A value of $\gamma = 1$ corresponds, thus, to H_0 for the Dickey-Fuller, the Phillips-Perron and the Phillips-Ouliaris test, but to H_1 for the KPSS test. Hence, we set $\gamma = 1$ to generate the cases with the non-cointegrated pairs.

Figure 5.1 depicts the corresponding sensitivity/specificity plots. In Table 5.2 we list the respective Q-values according to (5.1.1) for our four tests.

As expected, the higher the γ-value, the worse is the performance of the tests. The KPSS test shows the worst performance for all chosen γ values. The other three tests lie very close together. Even though the Phillips-Perron test shows the best Q-values for all γ-values, it is hard to say whether it significantly outperforms the ADF or the PO test. While the results for $\gamma = 0.9$ and

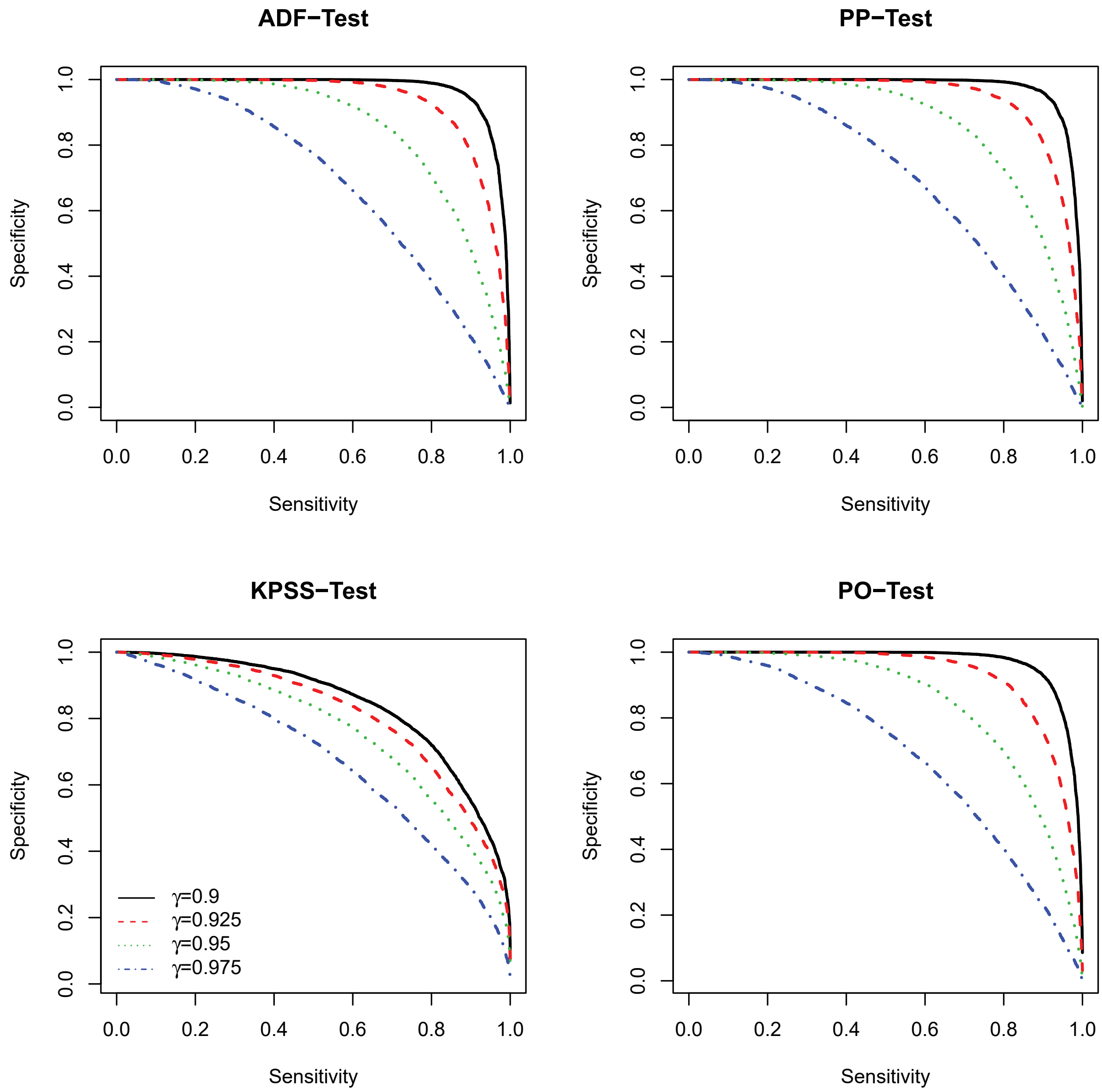

Figure 5.1: Sensitivity / Specificity Plots according to a GBM Price Dynamics.

	$\gamma = 0.9$	$\gamma = 0.925$	$\gamma = 0.95$	$\gamma = 0.975$
Dickey-Fuller	0.9445	0.8657	0.6914	0.3674
Phillips-Perron	0.9527	0.8777	0.7051	0.3796
KPSS	0.6882	0.6262	0.5243	0.3431
Phillips-Ouliaris	0.9398	0.8547	0.6767	0.3635

Table 5.2: Sensitivity / Specificity Quantifier with GBM Price Dynamics

$\gamma = 0.925$ look still very promising, the power of the tests decreases rapidly the further we go. A non-linear decrease in the power of a test is, however, quite natural and nothing we should be surprised of.

Although the price process (3.1.6) already exhibits heteroskedasticity in the sense that the volatility level rises as the price level moves up, we can still go one step further and augment the heteroskedastic behaviour by allowing the volatility to change also on the same price level. This is a particular feature of the Heston model. So, in the second setting we use the Heston model to generate the price series. The equilibrium volatility level θ according to the dynamics of (3.3.2) is set equal to the volatility level we used in setting 1, i.e. 0.25. For the diffusion parameter σ of (3.3.2) we choose 0.05 in order to obtain reasonable values for the process. Finally, we have to decide on κ, the parameter that controls the mean-reversion speed. Here we consider a value of 0.3 as a good choice that leads to realistic price movements. Figure 5.2 depicts an example of how the diffusion level may evolve over time.

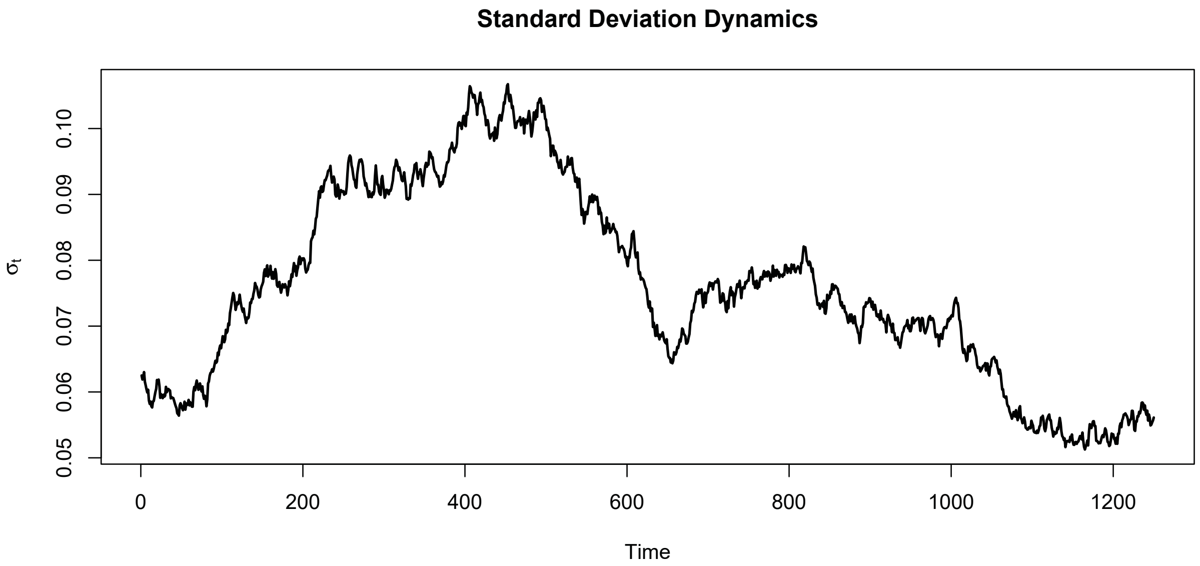

Figure 5.2: Standard Deviation Dynamics under the Heston Model.

An example price path under the Heston price dynamics with the chosen parameter values is shown in Figure 5.3.

In addition to the price series modification, we also introduce heteroskedasticity in the spread series m_t (i.e. the portfolio price). Figure 5.4 shows an example of spread returns with the introduced heteroskedasticity. There we can clearly see some volatility clusters.

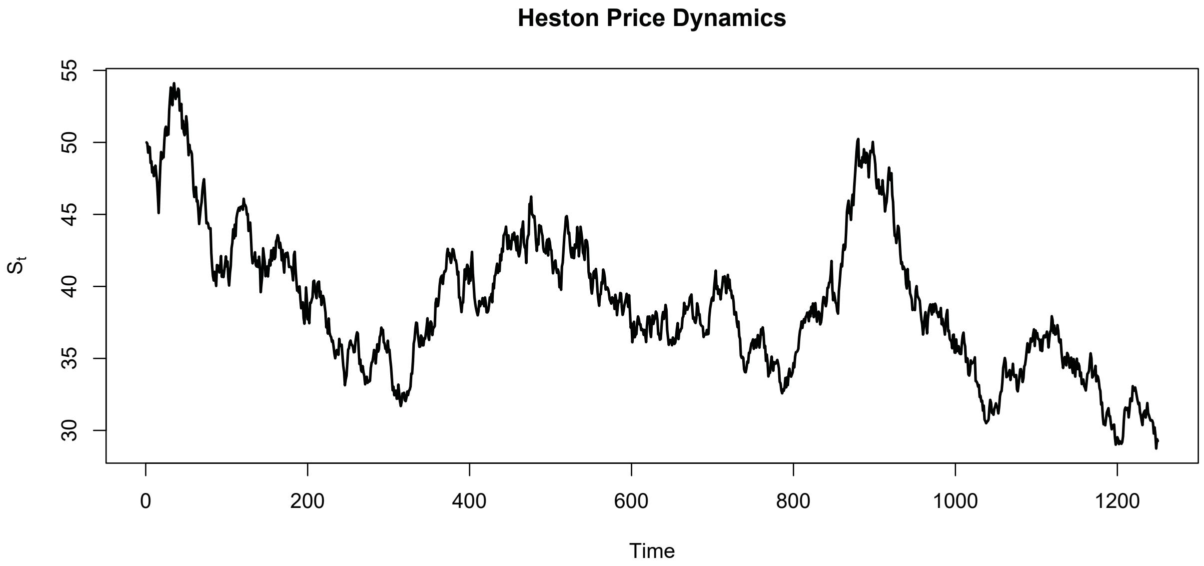

Figure 5.3: Price Process Dynamics under the Heston Model.

Repeating the simulation of setting 1 but now with the heteroskedastic price and spread dynamics, leads, however, to a very similar result as in setting 1. The sensitivity/specificity curves look very similar to the ones in Figure 5.1. Table 5.3 reports the corresponding Q-values. From there it becomes clear that the introduction of heteroskedasticity has basically no effect on the goodness of the tests.

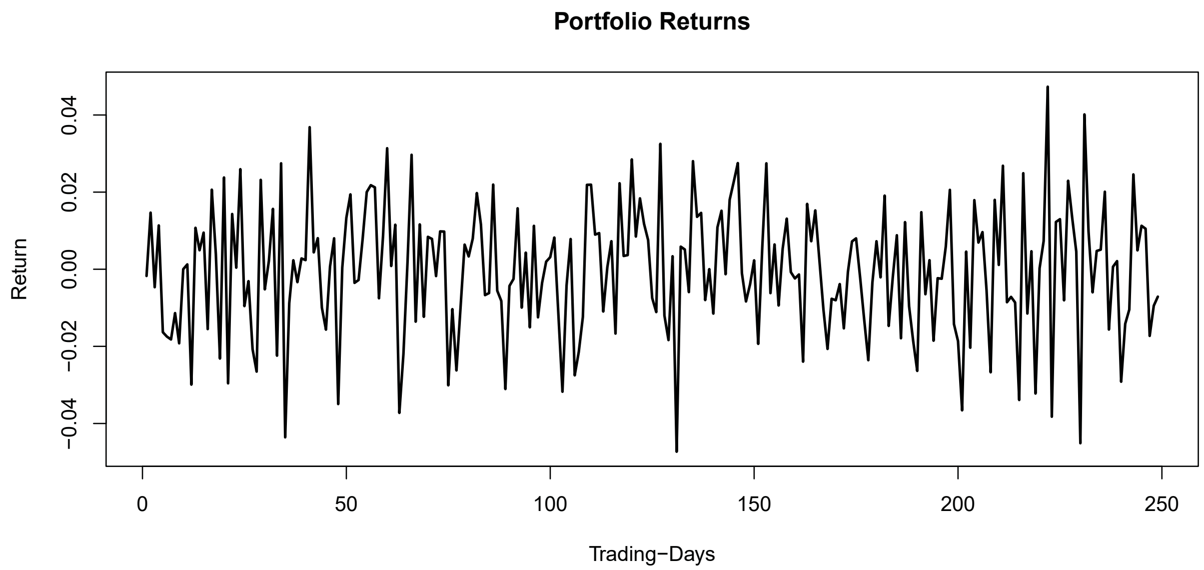

Figure 5.4: An Example of Spread-Returns with Heteroskedasticity.

	$\gamma = 0.9$	$\gamma = 0.925$	$\gamma = 0.95$	$\gamma = 0.975$
Dickey-Fuller	0.9453	0.8667	0.6936	0.3724
Phillips-Perron	0.9547	0.8788	0.7066	0.3839
KPSS	0.6888	0.6270	0.5244	0.3411
Phillips-Ouliaris	0.9432	0.8603	0.6837	0.3694

Table 5.3: Sensitivity / Specificity Quantifier with Heston Price Dynamics

The main change in setting 2, as compared to setting 1, is the introduction of a time-varying diffusion parameter. We could go even one step further and use the idea of time-variation not only with respect to the volatility but also in terms of the correlation structure. In particular, we assumed so far a constant correlation structure in the spread series m_t. Obviously, as with heteroskedasticity, a changing correlation structure clearly violates the definition of a stationary process. A constant correlation structure is, however, not really necessary in order to apply an effective pairs trading strategy.

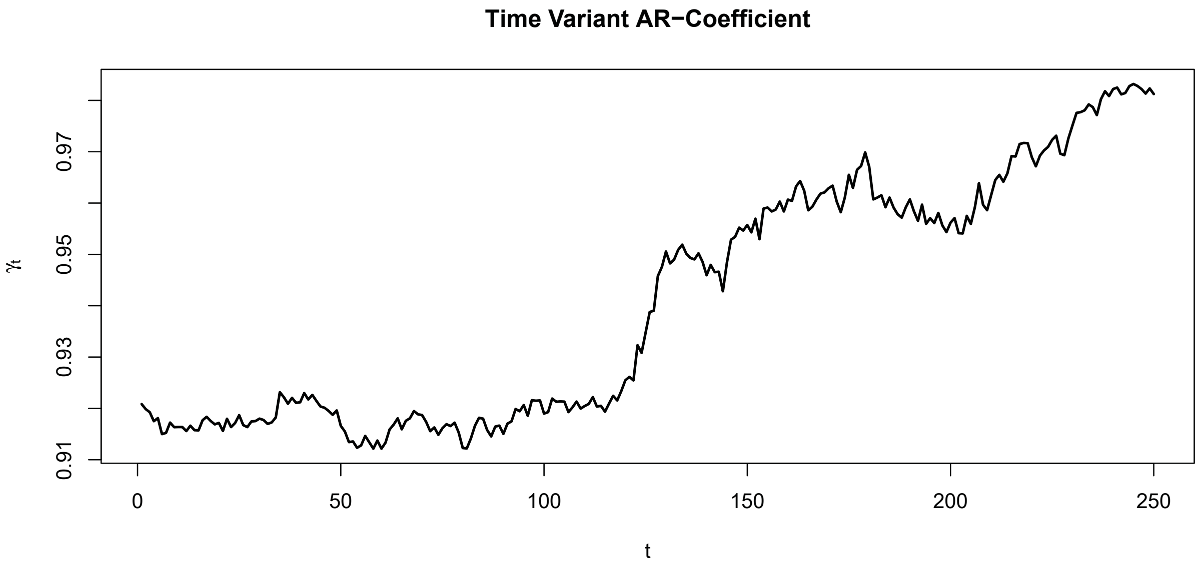

Figure 5.5: An Example Path with a Time-Varying AR-Coefficient.

If we want to stay in general within our AR(1) framework, a time-varying correlation structure can easily be implemented by making the γ parameter in (5.2.2) varying over time. As already mentioned before, it is reasonable to assume that γ lies in the interval $[0.9, 1[$. A way to achieve such values randomly, while still getting some rational transition from one point in time to the next, is the following one

$$\gamma_t = \frac{9}{10} + \frac{e^{\eta_t}}{10\,(1+e^{\eta_t})} \tag{5.2.3}$$

with

$$\eta_t = \eta_{t-1} + \zeta \tag{5.2.4}$$

and ζ being independent and standard normally distributed. An example of such a parameter path is depicted in Figure 5.5.

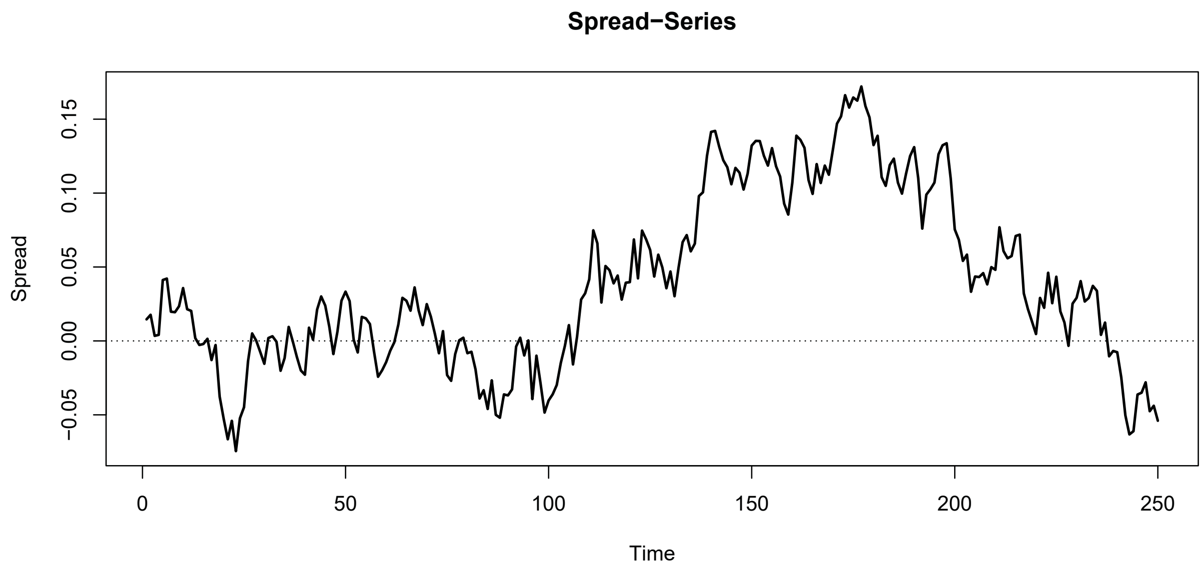

Figure 5.6: An Example of Spread-Returns with Time-Varying AR-Coefficient.

An example of a resulting spread time series is given in Figure 5.6. Without any doubt, the depicted process is not stationary but it is still mean-reverting. Such a process would, thus, still be appropriate for a pairs trading scheme. It is, however, also clear that it is a lot more challenging to find an effective trading rule in such a case.

So, if we repeat the simulations of setting 1 and 2 but this time with a time-varying γ_t according to (5.2.3) and (5.2.4), we can again plot the sensitivity/specificity curves of the four different tests. These are depicted in Figure 5.7. The corresponding Q-values are listed in Table 5.4.

The null hypothesis is in this case a bit over-stretched to “the assets are not tradable” for the ADF, the PP and the OU tests and to “the assets are

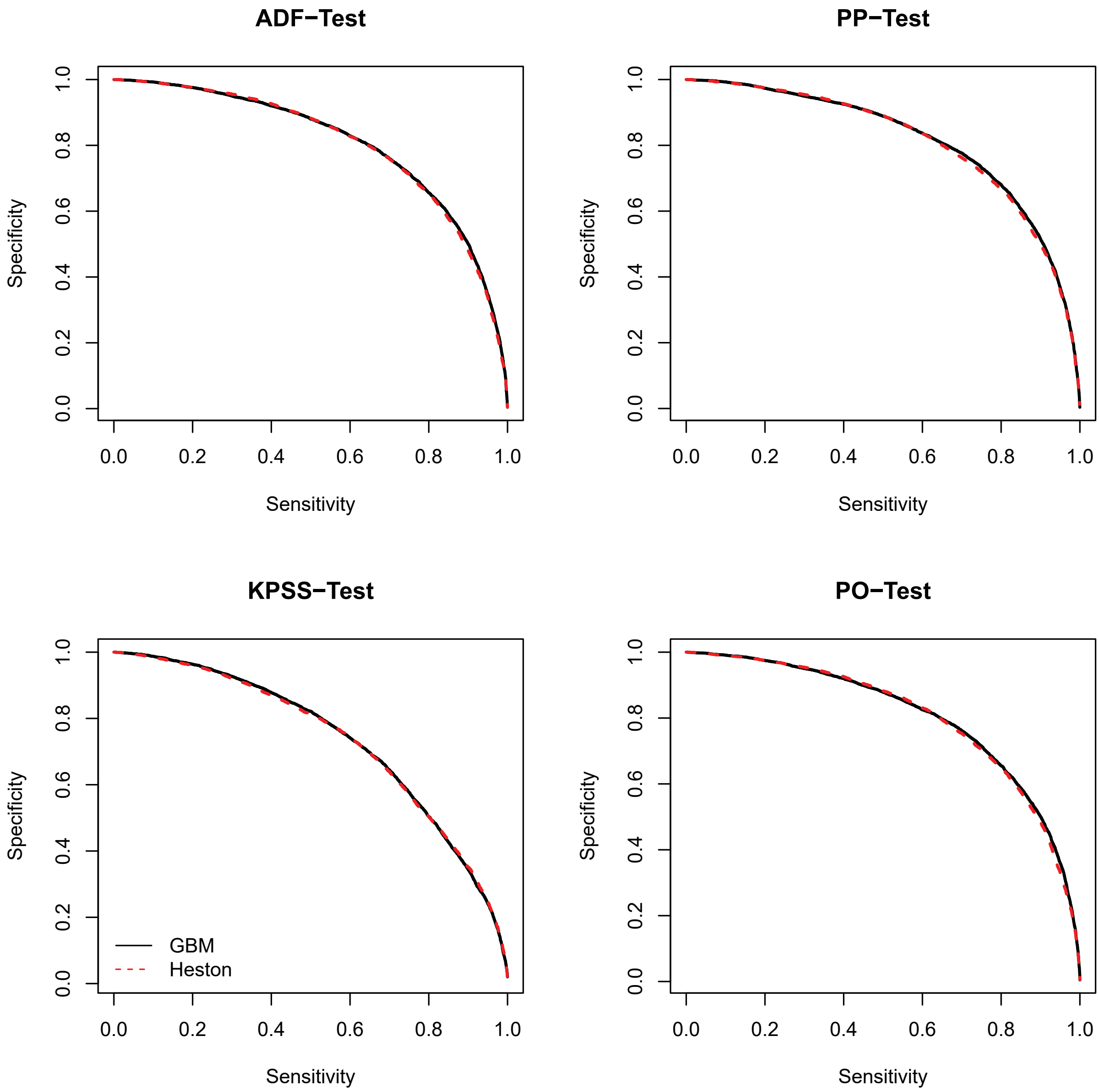

Figure 5.7: Sensitivity / Specificity when AR Coefficient is Varying over Time.

	GBM	Heston
Dickey-Fuller	0.6123	0.6078
Phillips-Perron	0.6274	0.6203
KPSS	0.4749	0.4704
Phillips-Ouliaris	0.6118	0.6059

Table 5.4: Sensitivity / Specificity Quantifier for Processes with Time-Varying AR-Coefficient

tradable" for the KPSS test. So, we should keep in mind that we stretch the purpose of our four tests quite a bit in this manner.

When looking at the results, we can say that here again, heteroskedasticity does not really have an impact. The augmented Dickey-Fuller test, the Phillips-Perron test and the Phillips-Ouliaris test show quite similar results and still perform quite decently under these circumstances. So, they may be very useful classifiers in quite a general setting and we can definitely employ them to select assets for a pairs trading scheme. If we do so, we will inevitably have to determine a critical test value according to which we finally allocate a particular pair to the *tradabe* or *non tradable* group. As we have seen in this chapter, under the more realistic settings, i.e. with a γ value close to 1 or a time-varying γ_t, the power of our cointegration tests is rather weak. For example, as we can directly see from Figure 5.7, if we chose a critical value that corresponds to a type I error of 5% under H_0, we have a specificity in setting 3 with the ADF test of merely 36%, meaning that we identify truly tradable pairs in only 36% of all cases. Hence, we should not generally limit ourselves to critical values that correspond to the typical 1%, 5% or 10% type I error under H_0. A critical value that corresponds to a 1% type I error under H_0 is very likely not suitable to identify tradable pairs. Values of 5% or 10% may be more adequate. It may be appropriate to go even further. This increases, however, the risk of implementing a pairs trading scheme on pairs that are in fact not cointegrated and, thus, of just running a random gamble.

Hence, the results of this chapter show that our cointegration tests may be too week in order to live up to our aspirations. Therefore, it may be wise to look for some helper procedure that can support or supplement a pure cointegration test based decision making.

Chapter 6

Approaching the Multiple Testing Problem

The key to success in terms of a statistical arbitrage trading strategy in general, and a pairs trading rule in particular, is a favourable security selection. Trading on spreads that are not mean-reverting can cause bankruptcy quite quickly. As we have just seen in Chapter 5, a pairs selection based on cointegration tests only may be problematic due to the limited power of the statistical tests. However, the test performance is not the only problem in our context. Running tests on a large number of pairs also leads us to the well known *multiple testing problem*. It is, therefore, advisable to expand the pairs selection procedure in a way that makes it more reliable.

6.1 Highlighting the Problem

As we have seen in Chapter 5, the usual test procedures to detect stationary, or at least mean-reverting spreads, reveal some considerable weaknesses. In addition, due to the nature of typical statistical tests, we will always find cases that reject the null hypothesis as long as we use enough samples, even if there is no actual cointegration relationship between the considered variables. This is not a problem with cointegration tests in particular. It is true for basically all statistical tests.

So far, almost all published empirical work on cointegration based statistical arbitrage rules take quite a broad asset universe and apply one of the established cointegration test procedures on it without any further restrictions.

This is problematic from a statistical point of view. The reason for that is quickly explained.

Let's assume that we have an asset universe of 100 assets that fulfil the basic trading requirements for our purpose. Let's further assume that not a single cointegration relationship truly exists between those assets. If we were to pursue a simple pairs trading strategy with a pairs selection procedure that uses a cointegration test like the Phillips-Ouliaris test, for instance, which, as we have seen in Chapter 4, has the nice property that it is not sensitive to the chosen normalisation of the cointegration vector, we would nevertheless find about 50 "tradable" pairs on the 1% significance level. The problem here comes not from the chosen cointegration test. It is the nature of our test application that leads to the problem, because with 100 assets, we would have to perform 4,950 tests. The critical value of the test statistic is chosen in a way that under the null hypothesis, i.e. if the null hypothesis is in fact true, we would observe values of the test statistic exceeding this critical value only in 1% of all cases. This is nothing else then our well known type I error. It says that if we apply a particular test and use the corresponding critical value on the chosen significance level α as a decision criterion to decide on the null hypothesis, we will make α errors in 100 tests if the null hypothesis holds true. Testing for tradable pairs with a cointegration test that is invariant to the normalisation of the cointegration vector means in this case that we check all the possible combinations we can build out of 100 elements, i.e. $\frac{100!}{2!\,98!} = 4,950$. Of course, if we apply a cointegration test that is sensitive to the normalisation of the cointegration vector, we will have to test in both ways, which means that we will have to test on a variation of 2 out of 100, resulting in $\frac{100!}{98!} = 9,900$ tests.

So, an unrestricted testing of asset combinations for tradability by using cointegration tests is very prone to produce spurious results. Trading on such results would be a high risk gamble. Nevertheless, the idea of employing cointegration tests to test for tradability of particular asset combinations still remains valid. However, we have to restrict this kind of testing to combinations for which we already have a reasonable belief in terms of their suitability for a successful pairs trading.

In the following, we will discuss some ways that could help us mitigate the discussed weaknesses of a purely cointegration test based pairs selection procedure.

6.2 A Simple Pre-Selection

As already mentioned, performing cointegration tests on linear combinations based on the whole set of generally suitable securities has considerable deficits. The number of admissible combinations must somehow be reduced by excluding combinations that have only a very slim change of leading to pairs trading gains. By this, the number of tests could be reduced. In this sense, it is very natural to abstain from asset combinations that have very little in common. The question is, why the security prices of two companies should largely move together when they operate in completely different fields. For example, what does a tire manufacturer and an insurance company have in common? So, it seems natural to focus on groups of securities that are somehow similar. The next question to ask is then, of course, what the term “similar” actually means in this case and how it could be measured. For common stocks it could mean that the underlying companies must be engaged in similar economic activities. So, a first simple approach could be to partition the whole set of admissible securities into industry groups. For instance, if we take the stocks of the S&P 500 index as our overall asset universe, we could group the companies into 10 industries according to the *FTSE Industry Classification Benchmark* (ICB).

The ICB, which is owned by FTSE International, sub-divides stock exchange listed companies in different groups according to their main business activities in terms of turn over. There are 4 levels of segregation, with the *industry* level being the highest one and consisting of 10 industry categories. On the second level we have the allocation to *supersectors*, consisting of 19 categories. On the third and fourth level, the ICB even provides a mapping for the companies on 41 *sectors* and 114 *subsectors*. Table 6.1 gives a short overview of the industry and supersector categories together with the underlying number of sectors.

A more detailed description of the segregation and the corresponding grouping criteria can be found on the ICB website.

Industry	Supersector	Nr. of Sectors
Oil & Gas	Oil & Gas	3
Basic Materials	Chemicals	1
	Basic Materials	3
Industrials	Construction & Materials	1
	Industrial Goods & Services	6
Consumer Goods	Automobiles & Parts	1
	Food & Beverage	2
	Personal & Household Goods	4
Health Care	Health Care	2
Consumer Services	Retail	2
	Media	1
	Travel & Leisure	1
Telecommunications	Telecommunications	2
Utilities	Utilities	2
Financials	Banks	1
	Insurance	2
	Real Estate	2
	Financial Services	3
Technology	Technology	2

Table 6.1: FTSE Industry Classification Benchmark

By looking at the ICB industry classification already reveals that such a classification could be very useful for our purpose. The next question is now, based on which level one should do the grouping. On the one hand, the first level, i.e. the 10 industry categories, might still be too crude. On the other hand, partitioning instead according to the second level, i.e. the 19 supersegments, might be too restrictive already.

A general problem of a classification scheme like the one provided by the ICB, is its binary nature. This is, a company either belongs to a particular group or not. There is no partial allocation possible. However, in the overall market we observe many companies that are active in several categories. Especially bigger companies that are of particular interest for a pairs trader due to their very liquid stocks, are often active in several categories. Bigger companies often have the means that allow them to reduce their industry exposure by diversifying their business activities. From this point of view it seems beneficial to have a more flexible partitioning scheme then just a binary one with 100% in one industry and 0% in all the others.

6.3 A Heuristic Pre-Selection Based on Indexes

Considering a particular ICB industry of a desired market as a portfolio of companies that generate their main turnovers in that industry, we can, in principle, calculate the prices and, therefore, also the returns of such a portfolio over time. Such price series are indeed provided by *FTSE International* for many markets. With such price series available, we could actually also obtain a rough estimate to which extent a particular company generates its gains in a specific industry. With I industries to be considered, a simple way to do that is by employing the following regression:

$$r_j = \beta_{j,0} + \sum_{i=1}^{I} \beta_{j,i} r_i + \varepsilon_j \tag{6.3.1}$$

with r_j denoting the return of the company we want to analyse and r_i being the return generated by industry i. The ε_j captures together with $\beta_{j,0}$ the company specific return that cannot be allocated to any of the I industries. Finally, the $\beta_{j,i}$ measures the extent to which company j generates its returns in industry i. So, for each company j we get an exposure vector $\boldsymbol{\beta}_j$ of dimension $I \times 1$, consisting of all the $\beta_{j,i}$.

In the ideal case, the exposure vector of a company j would contain only non-zero entries for industries in which the company is indeed operating and the non-zero entries would add up to 1. In addition, the company specific part $\beta_{j,0}$ would be zero. In this case, the exposure vectors represent a point in the I-dimensional space. Two points lying very close to each other would, thus, be considered as representing two similar securities. Consequently, we would consider two securities with exposure points laying far away from each other as dissimilar.

In order to partition the whole asset universe into homogeneous groups, we have to define an appropriate similarity measure $s(A,B)$ or dissimilarity measure $d(A,B)$ for two assets A and B. In partitioning exercises it is usually more convenient to work with dissimilarities. However, having one, we can always obtain the other too. For example, if we work with a dissimilarity measure, we can easily obtain a similarity measure out of it by simply using $s(A,B) = \frac{1}{1+d(A,B)}$, with $s(A,B) \in [0,1]$.

An easy and straight forward dissimilarity measure we could apply in our case is the Euclidean distance between the exposure points of two assets A and B. This is

$$d(A,B) = ||\boldsymbol{\beta}_A - \boldsymbol{\beta}_B||_2 = \sqrt{\sum_{i=1}^{I} (\beta_{A,i} - \beta_{B,i})^2} \tag{6.3.2}$$

where $\beta_{A,i}$ is the exposure value of asset A on industry i according to (6.3.1) and $\beta_{B,i}$ is the corresponding exposure value of asset B on industry i according to (6.3.1).

While (6.3.2) is probably a good point to start with, it might be beneficial to consider a small adjustment to it. Notably, when we look again at (6.3.1), we see that the industry returns are linearly translated to the return of a particular company and companies with different exposure vectors generate different returns. However, a difference in a specific industry exposure component has more influence on the return difference between two considered companies if that industry return has a higher variance. This is, for any $i = 1, 2, 3, \ldots, I$, a difference in β_i is worse for industry returns r_i with larger fluctuations. Hence, we should not consider the exposure differences in any dimension in exactly the same way but employ instead a deviation weighting based on the fluctuations of the industry returns. A simple extension of (6.3.2) is, thus

$$d(A,B) = \sqrt{(\boldsymbol{\beta}_A - \boldsymbol{\beta}_B)^\top V_I (\boldsymbol{\beta}_A - \boldsymbol{\beta}_B)} \tag{6.3.3}$$

with V_I denoting the $I \times I$ covariance matrix of the industry returns.

6.3.1 The Partitioning

Having defined a suitable dissimilarity measure, we can now focus on the partitioning of our asset universe into m homogeneous groups $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_m$.

With homogeneous groups we mean groups of companies with very similar economic activities. Accordingly, a heterogeneous group would, thus, consist of companies with rather different economic activities and would therefore not qualify as a group where we expect to find profitable security combinations for pairs trading. With a dissimilarity measure such as (6.3.2) or (6.3.3), we can achieve an adequate grouping in many ways. Employing the popular *K-means* or the *K-medoids* algorithm is as valid as utilising an *agglomerative* or *divisive clustering*. The later two belong to the so-called *hierarchical clustering* methods and offer the advantage over the K-means or K-medoids algorithm that they do not depend on the chosen starting configuration and the selected number of groups. After having run the clustering algorithm once, we have the optimal partitions for any desired number of groups. In addition, hierarchical clustering methods allow us to graphically illustrate the groupings in a dendrogram. When operating on high-dimensional spaces, agglomerative (buttom-up) methods are usually better suited and, in particular, computationally less expensive then divisive (top-down) methods. According to such considerations, an agglomerative clustering seems at least to be a good choice, although it is clearly not a compulsive choice for our purpose.

So, with an agglomerative clustering we start by calculating the dissimilarity value between every security in the overall sample and then pool the two companies with the smallest value together. For the next step we have to determine the dissimilarity values between the newly developed groups as well as between the groups and the remaining single observations. This is done by utilising a so-called *linkage*. The three most commonly used ones are the *single linkage*, the *complete linkage* and the *average linkage*. Single linkage agglomerative clustering defines the dissimilarity value between a group $\mathcal{G}_i$ and any remaining single observation j as follows:

$$d(\mathcal{G}_i, j) = \min\left(d(i,j) | i \in \mathcal{G}_i\right) \; . \tag{6.3.4}$$

Between two groups $\mathcal{G}_j$ and $\mathcal{G}_i$, the dissimilarity value is correspondingly defined as:

$$d(\mathcal{G}_i, \mathcal{G}_j) = \min\left(d(i,j) | i \in \mathcal{G}_i, j \in \mathcal{G}_j\right) \; . \tag{6.3.5}$$

This is a more general formulation as $\mathcal{G}$ can also be seen as a group that consists of just one single observation. Measure (6.3.5) actually means that we just take the dissimilarity value between the two most similar observations that belong to two different groups. Even though this is a very natural way to pool two groups together, (6.3.5) has the disadvantage that some observations of two particular groups that are pooled together can be very dissimilar. In order to avoid that, we could employ the complete linkage instead, which is defined as:

$$d(\mathcal{G}_i, \mathcal{G}_j) = \max\left(d(i,j) | i \in \mathcal{G}_i, j \in \mathcal{G}_j\right) \ . \tag{6.3.6}$$

Complete linkage agglomerative clustering ensures, thus, that all group members of a newly pooled group are not too dissimilar. Some kind of middle ground between the single and complete linkage represents the average linkage agglomerative clustering, which uses the following definition:

$$d(\mathcal{G}_i, \mathcal{G}_j) = \frac{\sum_{i \in \mathcal{G}_i} \sum_{j \in \mathcal{G}_j} d(i,j)}{N_{\mathcal{G}_i} N_{\mathcal{G}_j}} \tag{6.3.7}$$

where $N_{\mathcal{G}_i}$ and $N_{\mathcal{G}_j}$ denote the number of observations in group $\mathcal{G}_i$ and group $\mathcal{G}_j$, respectively.

Of the three presented definitions, the complete linkage agglomerative clustering seems the most appropriate for our purpose as it makes sure that the companies with the largest dissimilarity values in a particular group are not too unalike.

So, starting with single companies, we pool in each step single observations or groups together and finally end up with one big group that consists of all companies. By this, we get a hierarchical tree structure, which can be illustrated by a dendrogram. Figure 6.1 illustrates an example of such a dendrogram. Obviously, there is still the question of where we have to cut the tree in order to obtain a useful partition. There is no uniquely correct answer to that question. A natural suggestion is to just look at the dendrogram and cut the tree at a point where we see longer distances in the nodes. For example, in Figure 6.1 we would probably take a partition of just two groups, with the 16 elements on the left hand side in one group and all the rest in the other group. Another possibility is to cut the tree according to a maximum number of securities we would like to have in a group. This is clearly an issue for us as the goal of the whole grouping exercise is finally a reduction in the amount of statistical cointegration tests that have to be carried out. This

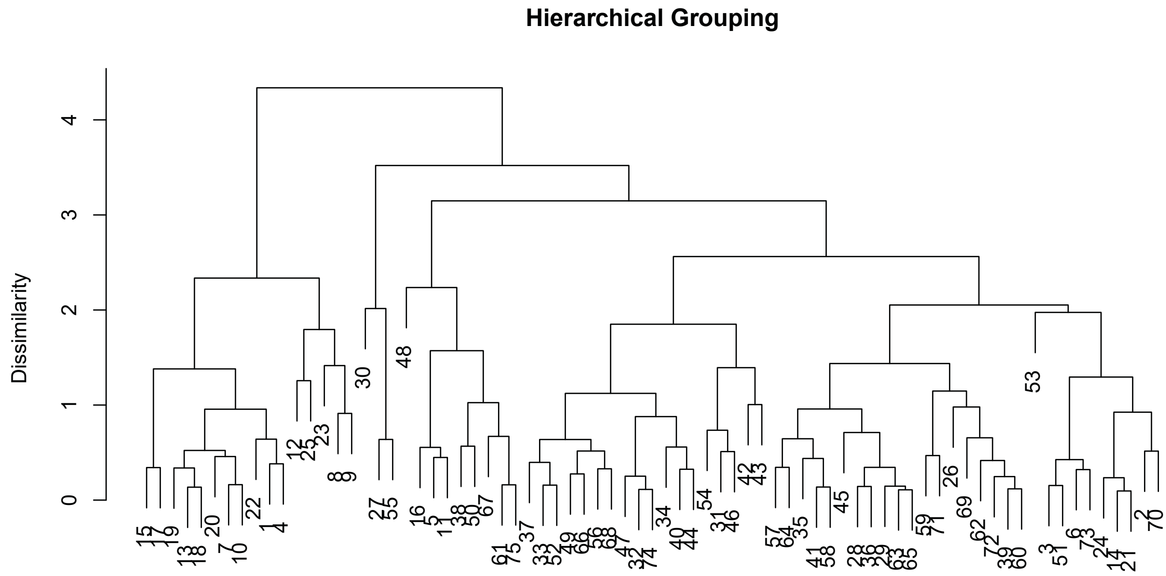

Figure 6.1: Example of a Dendrogram.

criterion indicates, however, only how many groups we should have at least but it does not tell us how many groups we should take at most. So, we need an additional goodness criterion. Such a key figure could be derived from the so-called *silhouettes*, which we will discuss in more detail in Section 6.3.2.

6.3.2 Silhouettes for a Good Partition

So far, we have not yet mentioned, into how many groups we should optimally partition our asset universe. That is a question that is not directly answered by any partitioning algorithm. As we have 10 indexes, it might be reasonable to start with 10 groups too. The question of how many groups should then finally be used, is probably best answered by checking how well the different groups are separated from each other. Our partitioning algorithm will always provide us with a result, no matter how badly the groups are delimited after all. To tackle this problem, Rousseeuw (1987) provides a useful quantity to determine for each observation how distinctly it is allocated to its current group. This quantity is called *silhouette width*.

To be more precise, let us denote $\mathcal{G}$ as the group to which observation i is currently allocated. So, we can define $a(i)$ as the average dissimilarity of object i to all other objects of its group. Furthermore, we specify $\bar{d}\,(i, \mathcal{H})$ as the average dissimilarity of observation i to all objects of group $\mathcal{H}$. We then determine the smallest $\bar{d}\,(i, \mathcal{H})$ for all groups $\mathcal{H} \neq \mathcal{G}$ and denote it as $b\,(i)$.

Hence, the silhouette width of a particular observation i is defined as

$$\mathcal{S}(i) := \frac{b\,(i) - a\,(i)}{\max\left\{a\,(i)\,, b\,(i)\right\}}\ . \tag{6.3.8}$$

As can be seen from the definition above, the silhouette width of any single observation lies always between -1 and 1. In addition, the bigger the silhouette width of a particular point, the better it fits to the group to which it is currently allocated. So, when deciding on the goodness of an overall partition, we should consider the silhouette widths of all points, i.e. of all single observations. Many small or even negative silhouette widths indicate a suboptimal partition. Hence, the average silhouette width of the whole partition may serve as a key figure in determining the optimal number of groups. Figure 6.2 illustrates a *silhouette* example with 150 observations and 2 groups. The silhouette values of every observation is depicted with a grey bar and the values of the group members are taken together and ordered from highest to lowest. High values and a sharp drop of the grey area towards the end of the group with no negative values are indications for a good partition.

6.3.3 Qualifications

The technique discussed so far in this section aims at partitioning a broad universe of securities with very different economic profiles into homogeneous groups. The goal of this exercise is to obtain groups of securities with very similar economic profiles. There are, however, many uncertainties in the approach stated so far. Firstly, employing indexes like the ICB price series to capture the economic risk profile of a security may be problematic. Such series are created by an initial classification of the companies according to their main industry exposure. So, a company that is part of a particular price index may have quite some exposure in other industries too. The revenues of a company are not split and allocated to the corresponding industries where the company earned the money. Therefore, an ICB industry index contains also components that are extraneous to the corresponding industry. This is less problematic if the extraneous contributions are random. Unfortunately, this

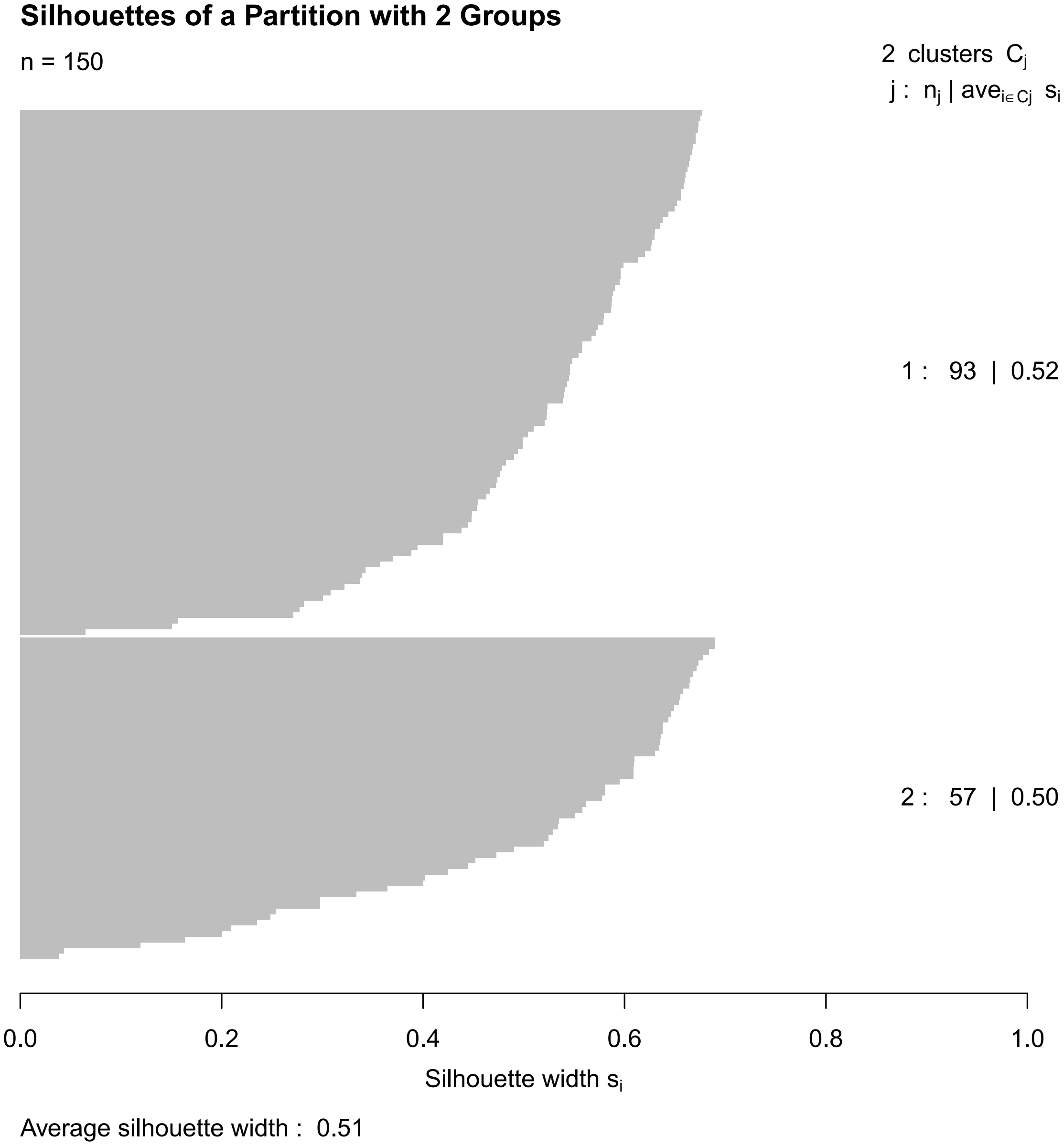

Figure 6.2: A Silhouette Plot Example.

is unlikely to be the case. The diversification strategy of many companies is based on their key strengths and this implies that their secondary businesses are not independent of their main business. If we examine the correlation structure of the US industry indexes, we will observe a picture like the one in Figure 6.3. There we depict on the upper triangular panels the bivariate scatter plots of the index returns of the year 1995 together with a locally-weighted polynomial regression according to Cleveland (1981). On the lower triangular panels we print an estimate of the Pearson linear correlation. As can be seen in Figure 6.3, some of the industry indexes show quite heavily

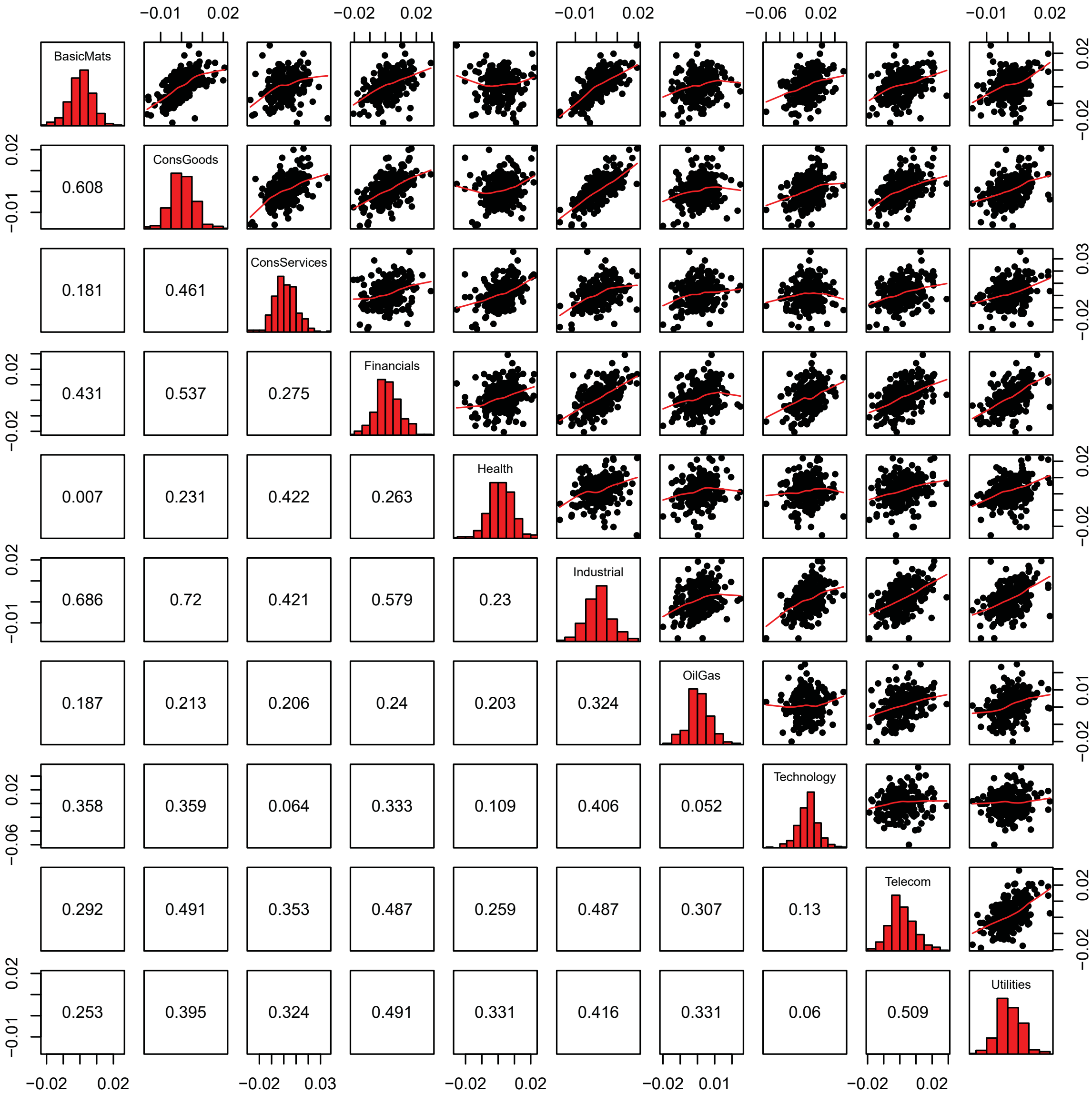

Figure 6.3: Correlation Structure between Industry Indexes.

correlated returns with other industry indexes. Using the proposed regression model (6.3.1) with considerably correlated indexes is not unproblematic as the parameter estimates, i.e. the $\widehat{\beta}_i$, are in this case badly determined. In other words, the design matrix of an OLS estimation of (6.3.1) could be ill-conditioned.

Clearly, there are ways to mitigate the correlation problem of explanatory variables in a linear regression. One could, for instance, orthogonalise the explanatory variables. However, the problem with the orthogonalisation is that we would have to do this on the base of sample data and by this, we would include additional uncertainty into our model. Nevertheless, it might be beneficial to strip off at least the effects of the overall market from the index returns. Figure 6.4 shows the bivariate industry scatter plots together

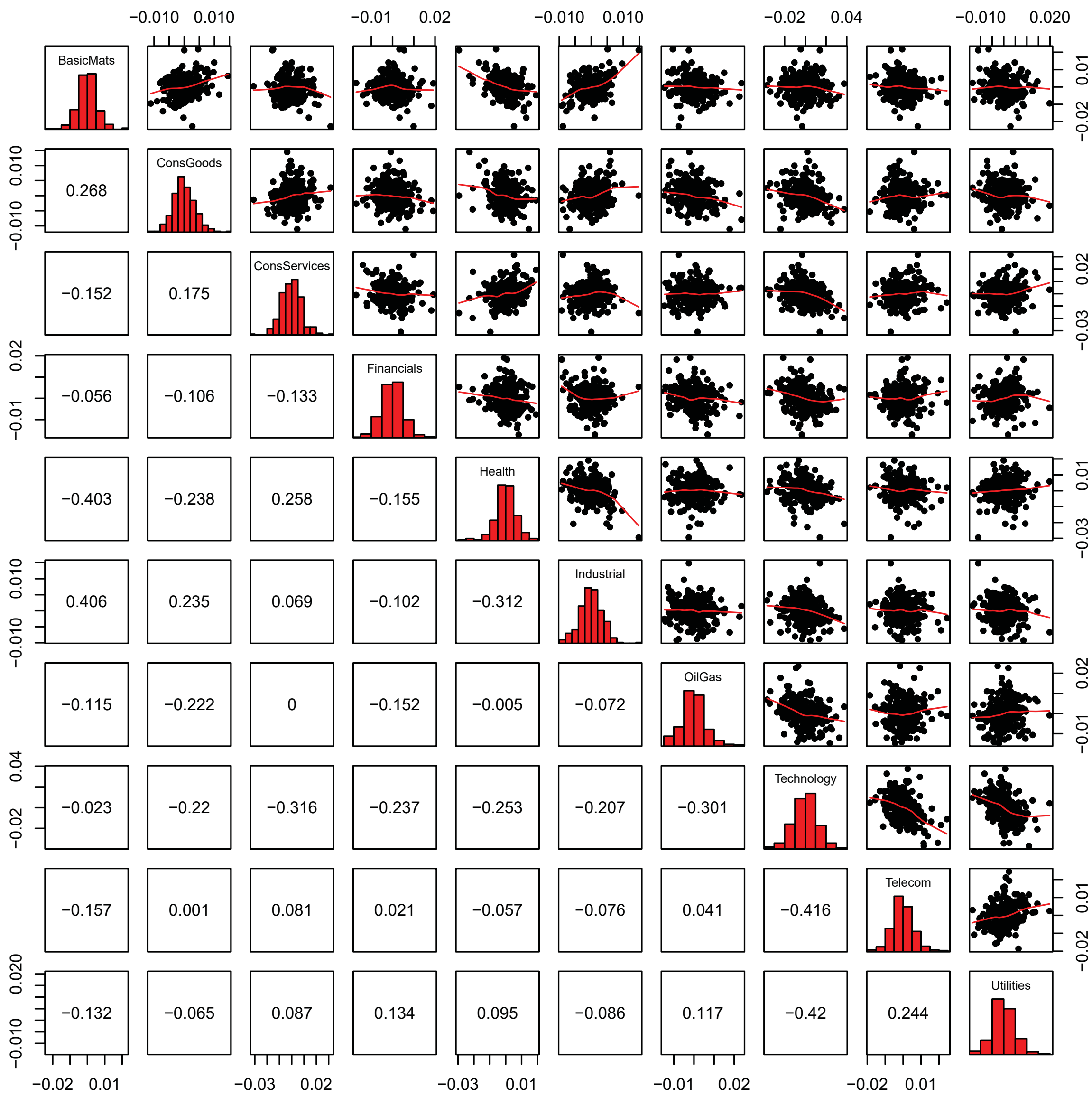

Figure 6.4: Market Adjusted Correlation Structure between Industry Indexes.

with the corresponding correlation estimates for the market adjusted index returns. The difference as compared to Figure 6.3 is quite striking. The strong correlations between Basic Materials, Consumer Goods and Industrials are clearly lessened, even though there is still some positive correlation left after the market correction. While the correlation structures between all the indexes are positive in the unadjusted setting, we now obtain also negative linear dependence structures in the adjusted setting.

Obviously, the proposed partitioning method as outlined in this section is only a heuristic approach with some clear deficits. It is also wrong to believe that groups of assets with similar risk profiles contain cointegrated securities for sure. So far, we have just assumed that it is more likely to find cointegrated or finally tradable securities among similar assets, which seems to be reasonable from a theoretical point of view. After all, the proposed method of this section may still be powerful enough to contribute considerably to the mitigation of the multiple testing problem.

6.4 The Common Trend Model

So far, we have considered a rather "quick and dirty" first approach in Section 6.2 and a more sophisticated but still heuristic approach in Section 6.3. The underlying assumption has been that securities with a similar risk profile are more likely to be suitable for pairs trading. So, it is time to shed light on the relationship between similar, in terms of economic risk factors, and statistically cointegrated assets. The link is best established on the base of the *common trend model* initially proposed by Stock and Watson (1988).

The common trend model makes use of the usual practice in time series analysis to decompose a series into a trend component, a cycle component, a seasonal component and a remainder. Often one assumes an additive structure as, for example,

$$x_t = \eta_t + \psi_t + \varsigma_t + \xi_t \tag{6.4.1}$$

where η represents the trend component, ψ the cyclical component and ς the seasonal component. The remaining part ξ is assumed to be weakly stationary.

Regarding financial time series, it is usually rather difficult to justify a cyclical and / or a seasonal component. Therefore, it makes sense in our case to assume a reduced version of (6.4.1), i.e.

$$x_t = \eta_t + \xi_t \ . \tag{6.4.2}$$

The trend component η_t in (6.4.2) is of stochastic nature, usually integrated of order one, that is $I(1)$. In addition, according to our discussion about security price processes in Chapter 3, it makes sense to assume that such an additive structure is more appropriate for log-transformed prices. So, if we have two assets A and B with price processes $S_{A,t}$ and $S_{B,t}$, we can write according to (6.4.2)

$$\begin{aligned} \log\left(S_{A,t}\right) &= p_{A,t} = \eta_{A,t} + \xi_{A,t} \\ \log\left(S_{B,t}\right) &= p_{B,t} = \eta_{B,t} + \xi_{B,t} \ . \end{aligned}$$

If it holds that

$$\eta_{A,t} = \gamma \eta_{B,t}$$

with $\gamma \in \mathbb{R}$, then there exists a linear combination of the two log-prices $p_{A,t}$ and $p_{B,t}$ that is weakly stationary. In particular, by taking the difference $p_{A,t} - \gamma p_{B,t}$ we obtain

$$\begin{aligned} p_{A,t} - \gamma p_{B,t} &= \eta_{A,t} - \gamma \eta_{B,t} + \xi_{A,t} - \gamma \xi_{B,t} \\ &= \xi_{A,t} - \gamma \xi_{B,t} \\ &= \epsilon_t \end{aligned} \tag{6.4.3}$$

which is a weakly stationary series as the stochastic trend cancels out. Equation (6.4.3) can now be rearranged, so that we obtain

$$\begin{aligned} p_{A,t} &= \gamma p_{B,t} + \epsilon_t \\ &= \alpha + \gamma p_{B,t} + \varepsilon_t \end{aligned} \tag{6.4.4}$$

with $\epsilon_t = \alpha + \varepsilon_t$. This is exactly the cointegration case, i.e. the relationship we obtain if the two price series are $CI(1,1)$. Hence, cointegration means that two time series have a common stochastic trend.

6.5 Risk Factors, Returns and Cointegration

With the common trend model in mind, we can now establish a link between risk factors and cointegration.

6.5.1 Risk Factors and Returns

By elaborating a bit more on the idea we briefly discussed in Section 2.4, we can establish a relationship between economic risk factors and statistical cointegration. As already mentioned, the *Arbitrage Pricing Theory* (APT) is a well established pricing theory. The theory consists, however, of much more than we actually need in our case. The building block that is relevant to us, is the idea of a return generating index or factor model[1]. Figure 6.5

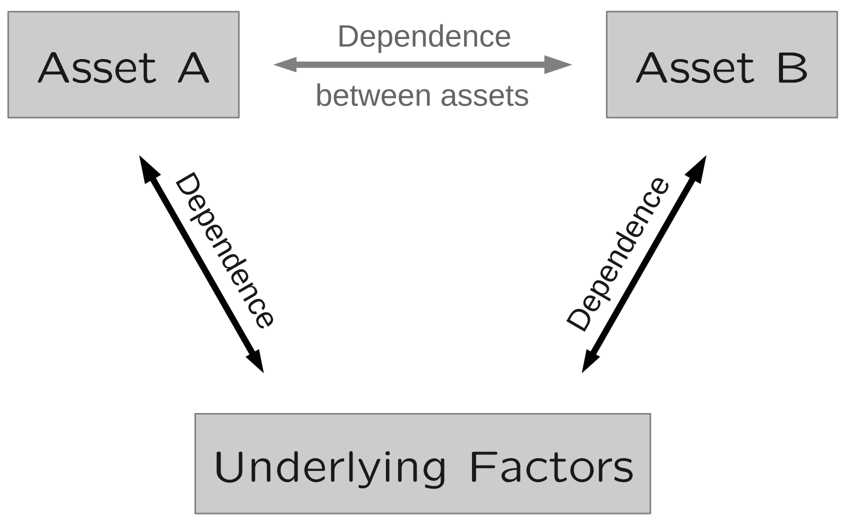

Figure 6.5: Dependence Structure between Assets and underlaying Factors.

illustrates the idea graphically. In particular, it is assumed that two assets depend on each other only by some underlying factors. This is, there exists a relationship between some underlying factors and the assets under scrutiny, so that the dependence between the assets exists only due to their common relation with the factors. The general structure of the return generating process for an asset i is assumed to depend on the k factors in an additive way

$$r_{t,i} = \beta_{0,i} + \sum_{j=1}^{k} \beta_{i,j} r^{e}_{t,j} + u_{t,i} \tag{6.5.1}$$

[1]Depending on the literature, different authors use the two terms *factor* and *index* in this context synonymously. The reader should be aware of the fact that the term *factor* means something different here as compared to the usual statistical understanding.

where $r^e_{t,j}$, with $j \in \{1, 2, \dots, k\}$, are in the standard model either the relative or absolute changes of the factors at time t. The $\beta_{0,i}$ is together with $u_{t,i}$ interpreted as the idiosyncratic part of the observed return. In addition to the linear dependence structure, there are some other assumptions usually made in order to obtain a reasonable working model. These are:

Assumption 1: $E\left[u_{t,i}\right] = 0$, for all assets.

Assumption 2: $E\left[u_{t,i}\left(r^e_{t,j} - E\left[r^e_{t,j}\right]\right)\right] = 0$, for all assets and risk factors.

Assumption 3: $E\left[u_{t,i}u_{t,h}\right] = 0$, for all i and h with $i \neq h$.

From Assumption 1 it follows that the expected idiosyncratic return component of an asset i is just $\beta_{0,i}$. If the factors represent the so-called systematic risk, an asset with zero exposure to the risk factors, which actually means that all the β's are zero, should according to the established financial theory generate in expectation the same return as the risk-free rate. In this case, it holds true that $\beta_{0,i}$ is equal to the risk-free rate for all assets. The other two assumptions follow directly from the model formulation. Clearly, if the dependence of two assets is only established through their exposure to the risk factors, the remaining return parts that are not related to the risk factors (according to Assumption 2) must be independent and, therefore, uncorrelated (Assumption 3).

There is a version of this model with orthogonal factor returns. Even though the original factor returns might not be orthogonal, one can easily make them orthogonal. This has the advantage of obtaining better determined parameter estimates. In addition, it leads to some simplifications regarding the analysis and the handling of the model. The price to pay is, however, an additional data manipulation that actually just ensures orthogonality of the current sample at hand. If we do not orthogonalise our factor returns, we should formulate 6.5.1 in vector notation in order to avoid too messy expressions in the further analysis. With $\boldsymbol{\beta}_i = (\beta_{i,0}, \beta_{i,1}, \dots, \beta_{i,k})^\top$ and $\boldsymbol{r^e} = (1, r^e_1, \dots, r^e_k)^\top$ we obtain

$$r_i = \boldsymbol{\beta}_i^\top \boldsymbol{r^e} + u_i \ . \tag{6.5.2}$$

If this relationship together with the mentioned model assumptions holds true, we can express the variance of any asset i as

$$\mathrm{Var}\left(r_i\right) = \boldsymbol{\beta}_i^\top V \boldsymbol{\beta}_i + \sigma_i^2 \tag{6.5.3}$$

and the covariance of any two assets i and h as

$$\mathrm{Cov}\left(r_i, r_h\right) = \boldsymbol{\beta}_i^\top V \boldsymbol{\beta}_h \tag{6.5.4}$$

with $\sigma_i^2 = E\left[u_i^2\right]$, and V being a $(k+1)\times(k+1)$ matrix containing the $k \times k$ covariance matrix of the factor changes. The first column and the first row of matrix V consist of only zeros.

6.5.2 Risk Factors and Cointegration

As already mentioned in Section 6.4, cointegration between two time series means that the two series follow a common stochastic trend. The link between cointegration and the risk factors according to (6.5.1) can, therefore, be made by employing the stochastic trend model. In particular, we assume that the log-prices of our securities are generated according to (6.4.2). This is

$$\log\left(S_t\right) = \eta_t + \xi_t \ .$$

The continuously compounded returns of a security A for a time span of m periods, with $m \in \mathbb{N}$, can then be written as

$$\begin{aligned} r_{t+m,A} &= \log\left(S_{t+m,A}\right) - \log\left(S_{t,A}\right) \\ &= \eta_{t+m} - \eta_t + \xi_{t+m,A} - \xi_{t,A} \\ &= \Delta(\eta_{t+m}) + \Delta(\xi_{t+m,A}) \ . \end{aligned} \tag{6.5.5}$$

The return part stemming from the specific part $\Delta(\xi_{t+m,A})$ should in expectation just be equal to the m-period risk-free rate $r_f(m)$ as there is usually no risk premium paid in the market for holding idiosyncratic risks. The systematic returns of a security should then be generated by the stochastic trend part $\Delta(\eta_{t+m})$ only. Therefore, if it holds true that the risk factors of (6.5.1) describe the changes in the stochastic trend component completely, i.e.

$$\Delta(\eta_{t+m}) = \sum_{j=1}^{k} \beta_{A,j} r^e_{t+m,j} \tag{6.5.6}$$

and

$$\Delta(\xi_{t+m,A}) = m\beta_{0,A} + u_{t+m,A} \\ = r_f(m) + u_{t+m,A} \tag{6.5.7}$$

then the corresponding β coefficients are equal up to a multiplicative constant γ for cointegrated securities. In vector notation this means for any two cointegrated securities A and B that we have

$$\boldsymbol{\beta}_A = \gamma \boldsymbol{\beta}_B \ , \tag{6.5.8}$$

meaning that the two exposure vectors lie on the same line passing through the origin.

Even though the proposed model looks promising at first glance, there are some considerable problems with it, especially with respect to practical applications. First of all, the requirement of a complete description of the trend component by the factors according to (6.5.6), is quite a strong one. We will never be able to check whether we have the "right" factors, let alone having all of them. In addition, it is not even guaranteed that such factors are measurable at all. The biggest issue in terms of practicality is, however, the fact that even if we have all the necessary factor and exposure information, (6.5.8) does not imply the existence of cointegration between two time series. So, having two securities with exposure vectors lying on each other, does not mean that they make a good pair for trading. However, the fact that (6.5.6) must hold for two cointegrated time series as long as all model assumptions are fulfilled, implies that the model is at least suitable to reduce the set of all possible pairs to a smaller subset of promising candidates.

An issue we have to consider if we want to apply our model on real data, is the estimation of the exposure vectors. This is actually not very difficult. However, we have to be aware of the fact that even if our model holds true and we have indeed a cointegrated security combination at hand, we will hardly ever observe the corresponding exposure vectors laying on the same line passing through the origin. In addition, we might be able to describe the changes in the stochastic trend only by approximations of the true factors. Hence, if we want to use the risk factor model as a grouping criterion similar to the approach discussed in Section 6.3, we have to come up with a decent decision criteria like the dissimilarity (or similarity) measures as discussed previously. If two assets have a common stochastic trend, we know that in the ideal case the exposure vectors of the two assets would lie on the same line passing through the origin. This means that the angle δ between such exposure vectors would be either 0 or 180 degrees. As already mentioned in Section 6.3, it is usually more convenient to define a dissimilarity measure. In this case, a useful criteria would be

$$d(A, B) = |\sin(\delta)| = \sqrt{1 - \left(\frac{\boldsymbol{\beta}_A^\top \boldsymbol{\beta}_B}{\|\boldsymbol{\beta}_A\| \cdot \|\boldsymbol{\beta}_B\|}\right)^2}. \tag{6.5.9}$$

The values of this measure lie between 0 and 1. So, in the ideal case of two cointegrated securities A and B, $d(A, B)$ would be 0. A special feature of this measure is its non-linear behaviour. Small deviations from the corresponding line passing through the origin translate almost linearly into the dissimilarity measure. Bigger deviations, however, translate into the dissimilarity measure less than linear. This can distort the picture of a partition. So, the question is, in which way we intend to use the dissimilarity measure.

Similar to the extension of the dissimilarity measure in Section 6.3, we can also extend (6.5.9) by including an exposure weighting according to the covariance matrix of the factor changes. To be more precise, if we denote the $k \times k$ covariance matrix of the factor changes as M, we can write:

$$d(A, B) = \sqrt{1 - \left(\frac{\boldsymbol{\beta}_A^\top M \boldsymbol{\beta}_B}{\left(\boldsymbol{\beta}_A^\top M \boldsymbol{\beta}_A\right)^{0.5} \cdot \left(\boldsymbol{\beta}_B^\top M \boldsymbol{\beta}_B\right)^{0.5}}\right)^2}. \tag{6.5.10}$$

The part in the brackets has now a very neat interpretation. In our model framework it is according to (6.5.3) and (6.5.4) of Section 6.5.1 just the linear correlation, $\rho^s_{A,B}$, of the systematic risk components of two assets A and B. This is

$$d(A, B) = \sqrt{1 - \left(\rho^s_{A,B}\right)^2}. \tag{6.5.11}$$

With such a distance measure, we could now allocate each observation to the object with the smallest dissimilarity value. Such an allocation would, however, not be optimal. We can illustrate this on a fictive example with four companies A, B, C and D together with their corresponding k-dimensional exposure vectors and the following dissimilarity matrix that contains all the pairwise dissimilarity values

$$\mathcal{D} = \begin{bmatrix} 0.00 & 0.03 & 0.05 & 0.02 \\ 0.03 & 0.00 & 0.09 & 0.07 \\ 0.05 & 0.09 & 0.00 & 0.11 \\ 0.02 & 0.07 & 0.11 & 0.00 \end{bmatrix} .$$

The element in the third row and second column is the dissimilarity value between company B and C. The diagonal elements are all zero as an object referring to itself is per definition perfectly similar. Hence, according to $\mathcal{D}$ we would allocate company A to company D and company B to company C. The dissimilarity value between the companies B and C is, however, with a value of 0.09 larger than the value between company B and company A, with a dissimilarity value of only 0.03. The allocation of company B to company C is even the worst for company B, as all the other dissimilarity values are smaller. It makes, therefore, sense not to restrict the allocation of the elements to just one matching partner but to allow an element to have several matches instead. Considering the issue that the estimated exposure vector is not exact, it is probably better to just pool promising securities together, rather than directly allocating them to each other. So, we can utilise again the same approach as discussed in Section 6.3.1 and Section 6.3.2 in order to achieve homogeneous groups of securities.

Chapter 7

Trading Rules and the Calculation of Returns

So far, we have just been considering the way of finding profitable pairs. We have, though, not yet mentioned the exact implementation of the strategy we have in mind. This is the topic of this chapter. So, the back-testing of Chapter 8 will finally be based on the terms described here.

7.1 Pairs Formation and Trading Rule

Before we can do the back-testing, we need to define the trading rule we want to employ. In order to be comparable with the strategy pursued by Gatev et al. (2006), it makes sense to use the same step by step approach as described there. It consists of a pairs formation period and a subsequent trading period. So, for the pairs formation period we choose a length of one year. Given a particular starting point, we take the observed data of the period reaching back one year from there, which gives us about 260 trading days, to match our pairs. Using a pairs formation period of roughly 260 trading days seems adequate as the period length should be long enough to obtain a reasonable amount of data points to reliably estimate the model parameters but at the same time, it should not reach to far into the past in order to avoid any structural breaks in the data. After having determined the pairs, we trade them during the subsequent 130 trading days. To be more precise, we look at the spread $m_t = \log(S_{A,t}) - \alpha - \beta \log(S_{B,t})$, which should be zero in equilibrium. A value of m_t that is smaller than zero implies that security A

is undervalued in terms of security B, in which case we short-sell security B and purchase security A. If m_t is larger than zero, we short-sell security A and purchase security B. Thus, having determined the pairs, the trading rule is based on the spread value m_t. So, when we say that one particular asset is overvalued or undervalued, we only make a relative statement, meaning that our valuation is based on the price of the corresponding pair-asset. The long-short mechanism with a pair works, however, only if the corresponding β is positive. If we have a relationship with a negative β, we would either have to buy both assets or to short-sell both assets. Such cases are, however, rather unlikely, especially in stock markets.

Having determined the relationship between the prices of two pair-assets, the spread value m_t is immediately obtained as the prices $S_{A,t}$ and $S_{B,t}$ are directly observable at time t. With this in mind we can now continue and discuss adequate trading rules. Thereby, it is advisable to consider the whole pairs trading investment as a portfolio of spreads. This is very convenient and gives us the ability to choose appropriate optimisation rules with respect to the overall investment and to control its risks. In the subsequent sections we discuss two very different cases in this regard. In Section 7.1.1 we will introduce a simple static trading approach, while a much more complex dynamic trading strategy will be explained in Section 7.1.2.

7.1.1 A Simple Trading Rule

At the very beginning, it is important to stress that a pairs trading strategy, as discussed so far, needs proper capital. It is incorrect to believe that we could just take the proceeds of the assets we have sold short to finance our long position without investing any proper capital. In particular, any short position has to be backed by an appropriate collateral. With respect to the capital that is used for this purpose, we may assume an investment return equal to the risk-free rate as the capital is usually placed in the money market during the investment period.

In order to formulate a suitable trading strategy, we have to focus on the spread series m_t. To give a short illustration, we can take the two companies *Verizon Communications Inc.* and *Bellsouth Corp*, for example. These two companies get a significant cointegration test statistic for the period stretching from January 4th, 1995 up to December 31st, 1995. If we regress the log-prices of *Verizon* on the log-prices of *Bellsouth*, we get the following estimated coefficients: $\widehat{\alpha} = 1.45$ and $\widehat{\beta} = 0.64$. The resulting spread time series m_t is depicted together with the corresponding correlogram in Figure 7.1. We can see there that the spread time series in the fitting period looks

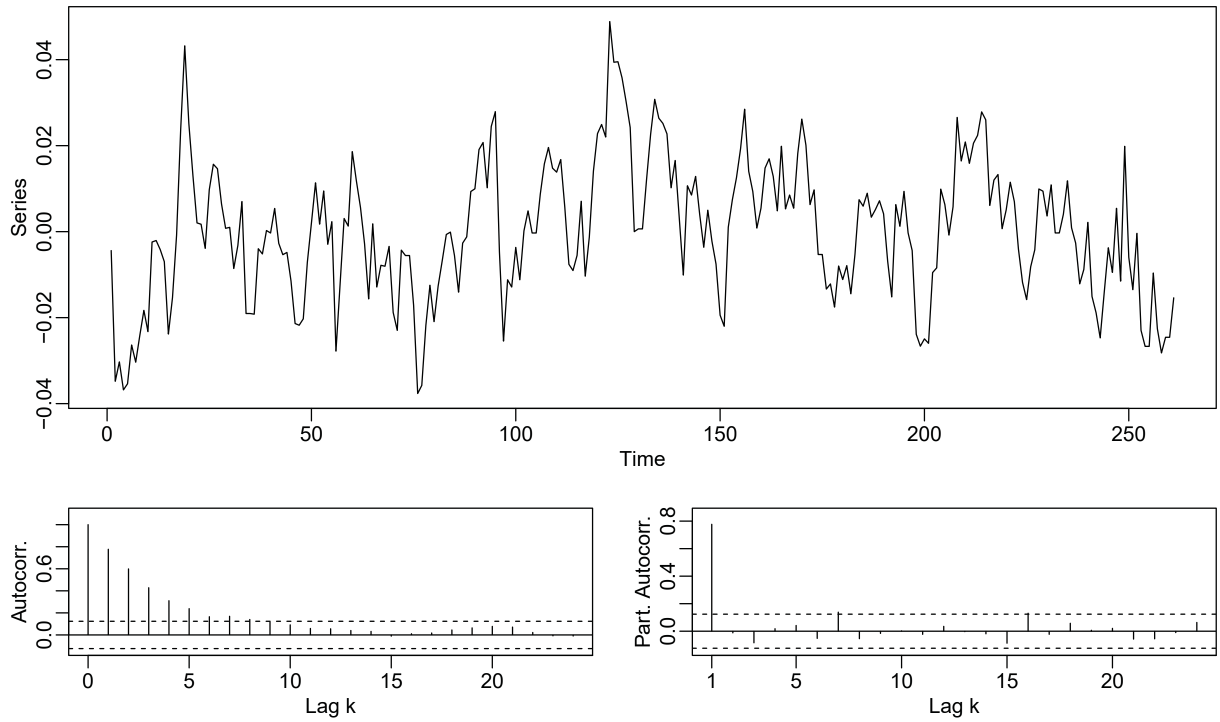

Figure 7.1: Spread time series of a potential pair.

very much like an AR(1) process suitable for our purpose. So, an obvious rule is to open a pair-position whenever the spread series shows a peak, i.e. a relative maximum or minimum, and to close it again, whenever it crosses the zero-line. Obviously, it is far from clear in advance, when the process reaches such a peak. However, we could try to model the spread series in this case as an AR(1) process. With that, we could make a forecast for the next value or the next few values of the spread time series and base the decision of whether to open or close a pair position now or later on the forecast values. Unfortunately, the assumption about the kind of process at hand may be wrong. Furthermore, there is always a general parameter risk involved in every statistical model. This fact is well illustrated by having a closer look at Figure 7.2. There, we see how the spread series as determined during the matching period, would have developed in the subsequent trading period, i.e. in the period stretching from January 2nd, 1996 up to July 3rd,

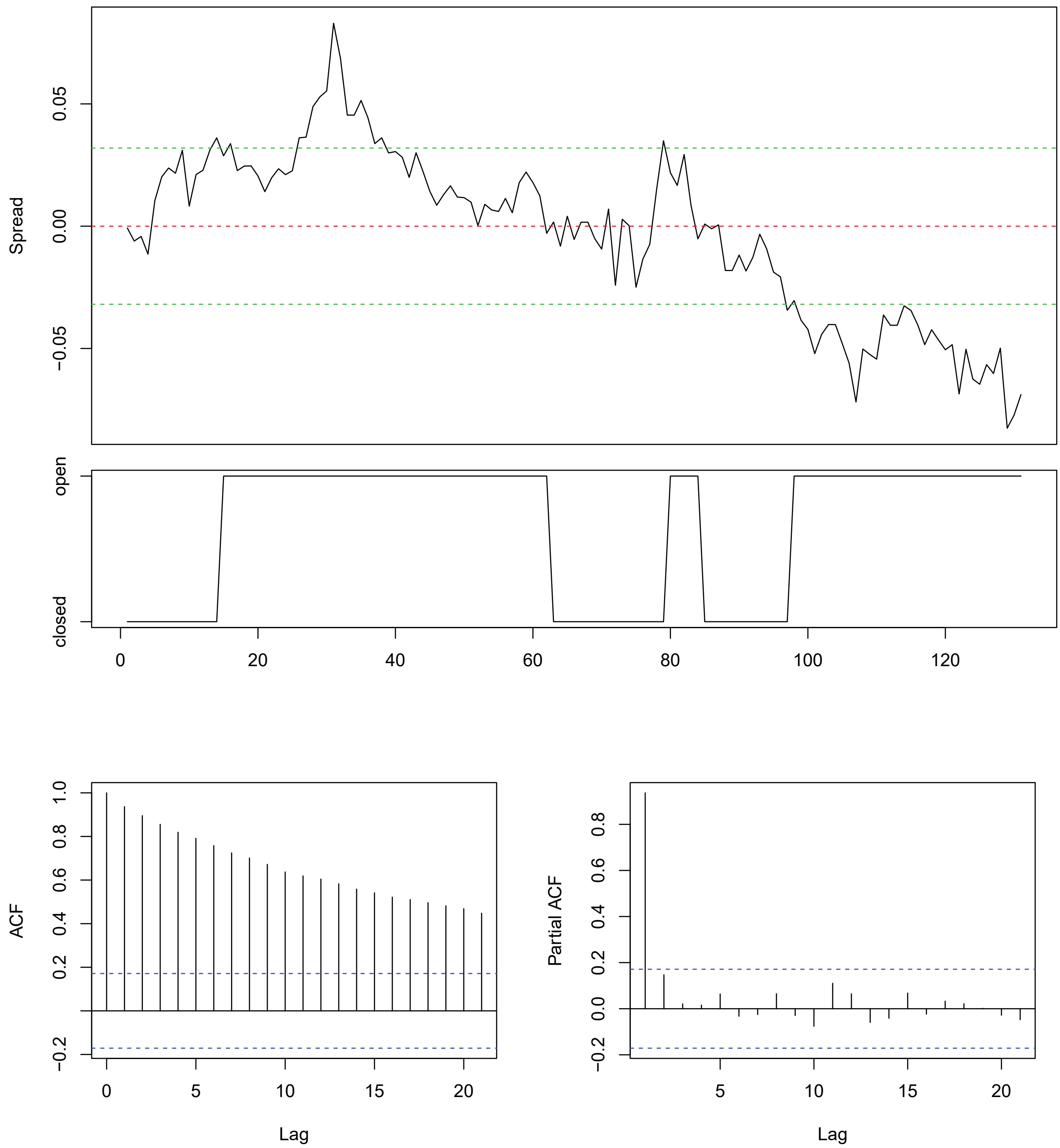

Figure 7.2: Spread time series of a fitted pair in the trading period.

1996. Even though the spread time series crosses the zero-line several times, it is unlikely to be weakly stationary. Most notably towards the end of the trading period, the spread departs considerably from the zero-line. This can also be noticed in Figure 7.3, where we see that the distance between the log-prices of the two stocks becomes smaller towards the end of the trading period. Nevertheless, a simple trading rule that has still potential to work under such circumstances, is a simple deviation based rule. For instance, the

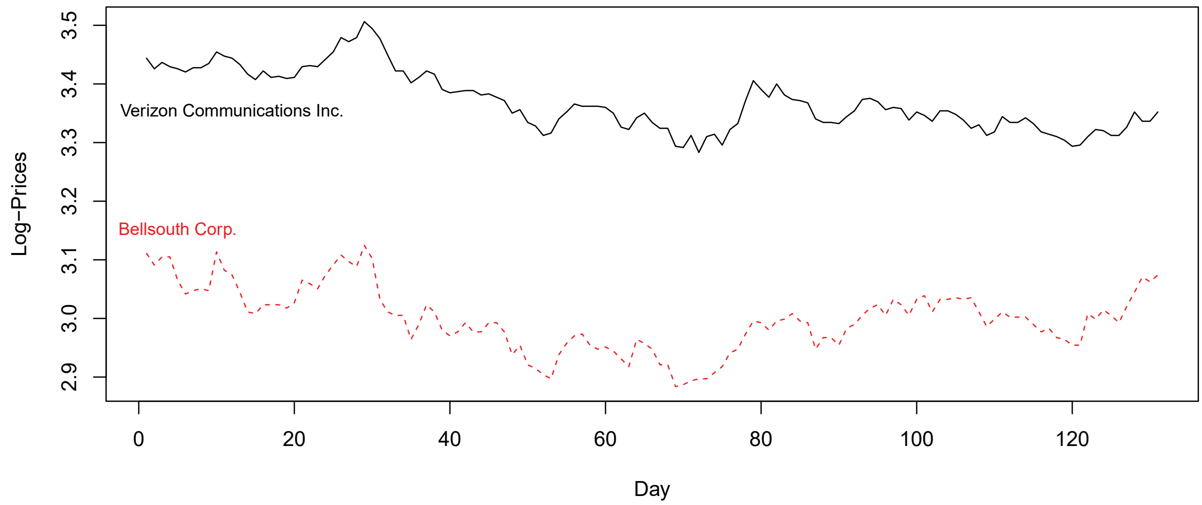

Figure 7.3: Log-price series of a potential pair in the trading period.

green dashed lines in Figure 7.2 mark the positive and negative deviation from the equilibrium relationship by two standard deviations of the spread time series as calculated during the matching period. So, the pair is opened when the spread series deviates by more than two historical standard deviations from its assumed equilibrium and is closed again when the spread is in equilibrium, i.e. when it crosses the zero-line or, in case we limit the trading period of each run, at the end of the trading period, regardless of the actual spread value.

This example shows with the spread value behaviour during the last few days of the trading period that a deviation from the assumed equilibrium value may become quite large, translating in a massive trading loss. The loss could even be so high that the portfolio becomes worthless. We should, thus, consider a stop-loss rule in order to limit the loss potential and to avoid bankruptcy.

7.1.2 A Dynamic Trading Rule

The simple trading rule, as illustrated in the previous section, is for sure a good point to start with and it will certainly qualify as a reasonable trading rule for us. Needless to say, there is a vast plenitude of other, more sophisticated rules we could apply instead and it is probably worth briefly highlighting at least one of them. A very interesting and technically advanced dynamic trading rule has, for instance, been proposed by Kim, Primbs, and Boyd (2008) and Mudchanatongsuk, Primbs, and Wong (2008). So, we take the opportunity to examine their proposal in the remaining part of this section to become a flavour of what kinds of other procedures are currently provided in the literature. At the end of this section we shall, however, discuss why the presented model is not appropriate for practical implementations. As opposed to the simple trading rule, the considered dynamic trading rule does not only decide on whether to open or close a specific pair at a particular point in time, but also decides by how much a pair position should be opened. It, therefore, decides on how much weight we should allocate to an individual position at a any point in time by considering the development of the whole portfolio of pairs as opposed to just focussing on one single pair position at once. As already previously mentioned, we need some proper capital as collateral which is in general assumed to be invested in the money market and to earn the corresponding interest rate. Kim et al. (2008), however, do not take that assumption and propose instead an approach where only the part of the proper capital that is not explicitly used for financing a corresponding trade position is invested in the money market. According to their assumptions, this part of the capital evolves according to

$$dB_t = rB_t dt \tag{7.1.1}$$

with r denoting the continuously compounded risk-free rate. To capture the mean-reverting behaviour of the spread time series m_t, the presumably simplest continuous time process suitable for this purpose, i.e. an Ornstein-Uhlenbeck process as discussed in Section 3.2, is proposed. This is, for each of the n pairs in the portfolio, it is assumed that their spread time series $m_{i,t}$, with $i = 1, 2, \ldots, n$, follow a mean-reverting Ornstein-Uhlenbeck process, where the corresponding parameters are allowed to have individual values for each process. Storing the values of the spread processes for each point in time t into a vector $\boldsymbol{m}_t = (m_{1,t}, m_{2,t}, \ldots, m_{n,t})^\top$, we can write all processes together in a more compact form, as

$$\boldsymbol{dm}_t = \Theta\left(\boldsymbol{\mu} - \boldsymbol{m}_t\right) dt + S\boldsymbol{dW}_t \tag{7.1.2}$$

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 $\boldsymbol{dW}_t = \left(dW_{1,t}, dW_{2,t}, \ldots, dW_{n,t}\right)^\top$, which is the vector that contains the standard Brownian motions of the n spread time series. Furthermore, we use $\boldsymbol{h}_t = \left(h_{1,t}, h_{2,t}, \ldots, h_{n,t}\right)^\top$ as the vector that contains the amount of spread units held at time t for each spread m_i, with $i = 1, 2, \ldots, n$. It indicates how much weight or capital we should allocate to the different spreads at a particular point in time. Denoting I_t as the total value of the whole pairs investment at time t, $\boldsymbol{h}_t^\top \boldsymbol{m}_t$ is the total value of the positions in the underlying securities of the spreads and $I_t - \boldsymbol{h}_t^\top \boldsymbol{m}_t$ is the amount of money invested in the money market. Thus, the dynamics of the total investment can be written as

$$dI_t = \boldsymbol{h}_t^\top \boldsymbol{dm}_t + \left(I_t - \boldsymbol{h}_t^\top \boldsymbol{m}_t\right) r dt \; . \tag{7.1.3}$$

Plugging the spread dynamics of (7.1.2) into (7.1.3) leads then to

$$dI_t = \left(r\left(I_t - \boldsymbol{h}_t^\top \boldsymbol{m}_t\right) + \boldsymbol{h}_t^\top \Theta \left(\boldsymbol{\mu} - \boldsymbol{m}_t\right)\right) dt + \boldsymbol{h}_t^\top S \boldsymbol{dW}_t \; . \tag{7.1.4}$$

Having determined the investment value dynamics according to (7.1.4), we can now, in conjunction with an appropriately chosen utility function $U\left(I_t\right)$, formulate a stochastic optimal control problem to maximize the expected utility for a given investment expiration time T. The utility function plays an important role as we need to quantify our preferences somehow. Neglecting for the moment all kinds of transaction costs, the control problem is then given by

$$\begin{gathered} \max_{\boldsymbol{h}} E\left[U\left(I_T\right)\right] \\ \text{subject to} \end{gathered} \tag{7.1.5}$$

$$\begin{aligned} \boldsymbol{dm}_t &= \Theta\left(\boldsymbol{\mu} - \boldsymbol{m}_t\right) dt + S\boldsymbol{dW}_t \\ dI_t &= \left(r\left(I_t - \boldsymbol{h}_t^\top \boldsymbol{m}_t\right) + \boldsymbol{h}_t^\top \Theta\left(\boldsymbol{\mu} - \boldsymbol{m}_t\right)\right) dt + \boldsymbol{h}_t^\top S\boldsymbol{dW}_t \end{aligned}$$

with given starting values $\boldsymbol{m}_0 = \left(m_{1,0}, m_{2,0}, \ldots, m_{n,0}\right)^\top$ 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 necessary value function is correspondingly defined as

$$J\left(t, I, \boldsymbol{m}\right) = \sup_{\boldsymbol{h}} \; E\left[U\left(I_T\right)\right] \tag{7.1.6}$$

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

$$\begin{aligned}-\frac{\partial J}{\partial t} = \sup_{\boldsymbol{h}} \; & \left[\frac{\partial J}{\partial I}\left(r\left(I - \boldsymbol{h}^{\top}\boldsymbol{m}\right) + \boldsymbol{h}^{\top}\Theta\left(\boldsymbol{\mu} - \boldsymbol{m}\right)\right) + \boldsymbol{\nabla}_{\boldsymbol{m}}^{\top}\Theta\left(\boldsymbol{\mu} - \boldsymbol{m}\right)\right. \\ & \left. + \frac{1}{2}\frac{\partial^2 J}{\partial I^2}\boldsymbol{h}^{\top}\Sigma\boldsymbol{h} + \frac{1}{2}tr\left(S^{\top}\nabla_{\boldsymbol{m}}^2 S\right) + \boldsymbol{h}^{\top}\Sigma\boldsymbol{\nabla}_{\boldsymbol{m}I}\right]\end{aligned} \tag{7.1.7}$$

with $\Sigma = SS^{\top}$, $\boldsymbol{\nabla}_{\boldsymbol{m}}$ 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 spread processes, it should be positive definite. Furthermore, admissible utility functions feature a diminishing marginal utility, which implies that $\frac{\partial^2 J}{\partial I^2}$ is negative. Hence, we can state that the expression in the square brackets of (7.1.7) is a concave function in $\boldsymbol{h}$. As a result, we just have to consider the first order conditions

$$\frac{\partial J}{\partial I}\left(-r\boldsymbol{m} + \Theta\left(\boldsymbol{\mu} - \boldsymbol{m}\right)\right) + \frac{\partial^2 J}{\partial I^2}\Sigma\boldsymbol{h} + \Sigma\boldsymbol{\nabla}_{\boldsymbol{m}I} = \boldsymbol{0} \tag{7.1.8}$$

in order to express the control vector as

$$\boldsymbol{h} = -\frac{1}{J_{II}}\Sigma^{-1}\left[J_I\left(\Theta\left(\boldsymbol{\mu} - \boldsymbol{m}\right) - r\boldsymbol{m}\right) + \Sigma\boldsymbol{\nabla}_{\boldsymbol{m}I}\right] \tag{7.1.9}$$

with $\frac{\partial J}{\partial I} = J_I$ and $\frac{\partial^2 J}{\partial I^2} = J_{II}$. If we substitute this expression back into (7.1.7), we obtain the following partial differential equation for the value function

$$\begin{aligned}0 =& J_t + J_I r I + \boldsymbol{\nabla}_{\boldsymbol{m}}^{\top}\Theta\left(\boldsymbol{\mu} - \boldsymbol{m}\right) + \frac{1}{2}tr\left(S^{\top}\nabla_{\boldsymbol{m}}^2 S\right) \\ & - \frac{1}{2J_{II}}\left[J_I\left(\Theta\left(\boldsymbol{\mu} - \boldsymbol{m}\right) - r\boldsymbol{m}\right) + \Sigma\boldsymbol{\nabla}_{\boldsymbol{m}I}\right]^{\top} \\ & \Sigma^{-1}\left[J_I\left(\Theta\left(\boldsymbol{\mu} - \boldsymbol{m}\right) - r\boldsymbol{m}\right) + \Sigma\boldsymbol{\nabla}_{\boldsymbol{m}I}\right]\end{aligned} \tag{7.1.10}$$

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 in cases where the utility functions are of the following two types, i.e.

$$U(I) = \begin{cases} \frac{I^{1-\gamma}}{1-\gamma} & , \gamma \neq 1 \ , \ \gamma > 0 \\ \log I & , \gamma = 1 \end{cases} \tag{7.1.11}$$

or

$$U(I) = -\frac{1}{\gamma} e^{-\gamma I} \ , \ \gamma > 0 \ . \tag{7.1.12}$$

If we cannot solve the problem analytically, we will, nevertheless, be able to evaluate (7.1.10) numerically and to employ the resulting values to calculate the control vector in this way.

As becomes clear from the previous explanations, this approach is computationally more expensive than the simple trading rule of Section 7.1.1. On the one hand, it may lead to a better result in terms of a higher profit and / or a less risky price path of the whole portfolio of pairs. On the other hand, there are besides of the more challenging implementation, some other major problems too. For example, it relies heavily on the assumption that each spread time series follows an Ornstein-Uhlenbeck process and it is an open question, how optimal the results would be if the true processes deviate slightly form that assumption. Another issue concerns the amount of transactions. Adjusting the portfolio weights continuously over time also means higher transaction cost. However, one of the biggest drawbacks of the approach is the parameter estimation problem, especially with respect to the drift term. This is already a major problem in terms of normal price time series. With respect to spread time series, which are not directly observable but have to be estimated instead, we have to admit that there is just no convincing parameter estimation procedure available yet that would make it worth implementing the approach. The prospect of success of that dynamic trading rule seems in this respect rather limited. So, we will not pursue that trading rule any further in the sequel of this work.

7.2 Return Calculations

Regarding investments consisting of only long-positions, the return calculation is usually not a big issue. Returns are just relative values. More precisely, they measure one value relative to another value. The return of an asset in a long position with respect to a particular period of time 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 is in this case just the money earned on one monetary unit of the whole investment, i.e. the amount an investor gets for investing one single dollar.

With respect to a pairs trading investment, the return calculation might not be so obvious at first sight. However, this is only the case if we believe in a self-financing portfolio where the investor does not have to invest any single dollar of his own money. This means, the long position is financed by the proceeds received from the short position. However, as we have already stated above, a pairs trading strategy does indeed require some proper capital. This is, in order to short-sell assets, we need to borrow them first. This can usually only be done by depositing other assets or just money of the same value, or at least a fraction of it. This is the so-called *collateral*. The borrower of any assets usually has to pay an *initial margin*. This is a particular amount of money paid at the beginning of the borrowing period, which serves as a buffer with regard to 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*, obliging him to provide additional funds. The capital put aside to serve as a collateral typically yields the short-term interest rate. Under these circumstances, we can interpret the change in value of the pair-positions relative to the amount of the collateral as excess return, i.e. the return earned above the risk-free rate.

With regard to return calculations, we should distinguish between the value of a pair-position and the cash flows it generates. While the cash flows are important regarding liquidity issues of the fund, it is actually only the value of the fund, which also comprises the unrealised gains and losses of unclosed pair-positions, that matters for return calculations. Hence, in order to obtain proper returns for our pairs trading strategy, we should first calculate the value of the fund for each point in time and then calculate the simple returns from those values. The value of the fund, denoted as V_t, can be expressed as:

$$V_t = C_t + \sum_{i=1}^{n} h_{i,t} \left(\upsilon_{\tau_i,t}^{(i,long)} - \upsilon_{\tau_i,t}^{(i,short)} \right) \tag{7.2.1}$$

where $\upsilon_{\tau_i,t+1}^{(i,long)}$ stands for the value of a one dollar long position with transaction time τ_i, where $0 \leq \tau_i \leq t$, and $h_{i,t}$ is the weight that is put on the pair-position i at time t. The term $\upsilon_{\tau_i,t+1}^{(i,short)}$ denotes the corresponding short position of pair i, i.e. the position that is sold short in order to receive one dollar at time τ_i. Thus, the term $\upsilon_{\tau_i,t}^{(i,long)} - \upsilon_{\tau_i,t}^{(i,short)}$ just represents the net value of pair i at time t. To be more specific, if we open a pair position at time τ, we borrow as many shares of the relatively overpriced asset in order to obtain one dollar when we sell this position. With the one dollar received from that selling, we buy as many shares of the relatively underpriced asset as we obtain for that amount of money. So, at time τ, the net value of any pair-position is zero. For $t > \tau$ these net positions represent the unrealised gains or losses of the pair-positions that were opened and have not yet been closed. The term C_t stands, by contrast, for money we actually hold at time t, i.e. it consists of the collateral and the cumulative realised gains and losses of closed pair-positions. It is usually assumed to be interest bearing. If we assume an initial margin of 100% at the beginning of each trading period t_0, then the weights h_{i,t_0} must incipiently sum up to C_{t_0}. If we employ a simple trading rule with equal weights, as introduced in Section 7.1.1, we just allocate $\frac{C_{t_0}}{n}$ to each pair-position at the beginning of every trading period.

Chapter 8

Back-testing on the S&P 500

After the theoretical considerations of the previous sections, we now want to see how the proposed strategies would have performed in the past if we had applied them on 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 stretching from the beginning of 1996 up to the end of 2011.

With respect to the pairs formation and the trading periods, we follow the same approach as employed by Gatev et al. (2006). In particular, the pairs formation period is always one year followed by a trading period of 6 months. After these 6 months, we take again the past 12 months for a new matching and use the new pairs in the subsequent 6 months lasting trading period. As trading rule we take the simple approach as introduced in Section 7.1.1 with an implemented stop-loss rule at -20%. This means that a pair is immediately closed as soon as it reaches a cumulative return of -20%.

8.1 The S&P 500

Before we start, it might be advisable to briefly reason about the choice of performing the back-testing on 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 a pairs trading 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 a particular asset.

Regarding the implementation of our back-testing, it is absolutely key to avoid any *survivorship bias*. Ex-post we are always smarter. So, we have to make sure that we only use the information available at the point in time when we do the calibration and the trading. A major problem in not carefully enough designed back-testing studies is the neglect of failed companies because they no longer exist at the time when the study is carried out. If we only take those companies into account that have been successful enough to survive until the end of a specific period, we will end up with a performance that is higher then it would actually have been at the time. Any company that goes bankrupt during the analysed period can cause a considerable loss in an investment strategy. There is no exception in our case. In particular, if we have spotted during the matching period a pair of two assets, A and B, on which we then trade in the subsequent period, it may be the case that company A faces some company specific problems which results in a decreased share price and an increased spread. Consequently, we would buy the shares of company A and short-sell the shares of company B. If the situation of company A becomes even worse so that it finally goes bankrupt, its share price would fall even further. Holding a long position in asset A and a short position in asset B, would then also translate in a decrease of the value of our portfolio, resulting in a potentially heavy loss. Thus, any serious back-testing must take this into account, even though the back-testing becomes much more expensive that way. Often it is already very difficult to acquire the relevant information about the historical asset universe of a particular market, and if the required information is available, the data collection and preparation process for the testing is usually much more time consuming.

In the back-testing we pursue in the following sections, we consider only stocks listed on the S&P 500 index. In particular, for every selection period we choose only pairs out of the asset universe that make up the S&P 500 index. Hence, the composition of the asset universe is not constant over time. Some companies vanish from the index at some point in time and new companies enter instead. 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 roughly 1,000 entries.

8.2 Strategy Implementation

The chosen back-testing time period from the beginning of 1996 up to the end of 2011, is with no doubt a very interesting one as it encompasses *bullish* as well as *bearish* markets, including the financial crisis of 2008. As a suitable benchmark we consider the S&P 500 index as it is a valid choice for a passive investment in a well diversified portfolio. In addition, the S&P 500 is also the market from which we take the assets involved in our strategy. So, the difference in performance can in this way directly be allocated to the active fund management.

8.2.1 Fixed Index Grouping

By subdividing the S&P 500 assets into the 10 industry groups according to the FTSE Industry Classification, we perform a cointegration test for the assets in each group by employing the Phillips-Perron procedure as discussed in Chapter 4. We restrict the pair formation to assets of the same industry group, i.e. only assets belonging to the same industry group are allowed to make a pair. For each industry group and period we then take the combinations with the highest cointegration test statistics, which must at least exceed the 1% significance level according to the corresponding test distribution. As transaction cost are always a big issue for active trading strategies, we have to be concerned about the number of transactions and should try to keep it low. So, the number of pairs in our portfolio should be limited somehow. Furthermore, in the back-testing studies by Gatev et al. (2006) and Harlacher (2012) it is reported that an increasing number of pairs in the portfolio has rather a diminishing effect on the returns even before accounting for transaction cost. To be more precise, in the back-testing study by Harlacher (2012), where portfolios consisting of 20 and 50 pairs are analysed, the smaller portfolios always led to higher average returns before transaction cost. So, restricting the size of the overall portfolio to a maximum of 50 pairs seems reasonable. Clearly, when carrying out a proper back-testing, we always have to be very careful with such decisions in order to not distort the results too much by acting according to information we would not have had at the time. Already knowing the results of other back-testing studies performed on similar time spans is, thus, tricky.

If we have naturally defined groups that are stable over time, like in this case here, where we use the 10 industry groups according to the FTSE Industry Classification, we may want to consider a further restriction with respect to the number of pairs we are willing to consider for our strategy. In particular, the cap we set for the entire portfolio can be broken down to the industry group level. A fact that we should always keep in mind, is that the statistics we obtain from the cointegration tests, are not really designed for the purpose we actually use them. So, they are very likely appropriate to give us valuable indications, but we should always be accessible for additional means to optimise our strategy. Setting an upper limit for each industry group, too, may lead to a better diversified portfolio overall and, finally, to a better performance. Furthermore, we get the opportunity to analyse the performance contribution of each group individually. This may lead to further conclusions and appropriate actions with respect to the trading rule implementation in the future. If we decide to apply a sub-limit to our groups, we will have to answer the question of how we should derive it from the cap we set for the entire portfolio. A sophisticated answer to that question may be complex. However, a simple and intuitive way to set a sub-limit for each group is by just dividing the overall portfolio cap by the number of considered groups, leading to the same sub-limit for each group. In our case, we would obtain by this a sub-limit of 5 pairs for each industry group. As this may be beneficial and causes not too much effort, we will implement this 5 pairs per industry sub-limit for the Fixed Index Grouping strategy.

Figure 8.1 depicts the number of chosen pairs per industry and period. As can be seen there, the imposed maximum number of pairs is not always binding. Especially for the *Telecommunications* industry (*Industry 9*) it is rather exceptional to have five pairs in a particular period. A possible explanation for this could clearly be the smaller number of assets in this industry group over all 32 periods. The dashed black line indicates the average number of

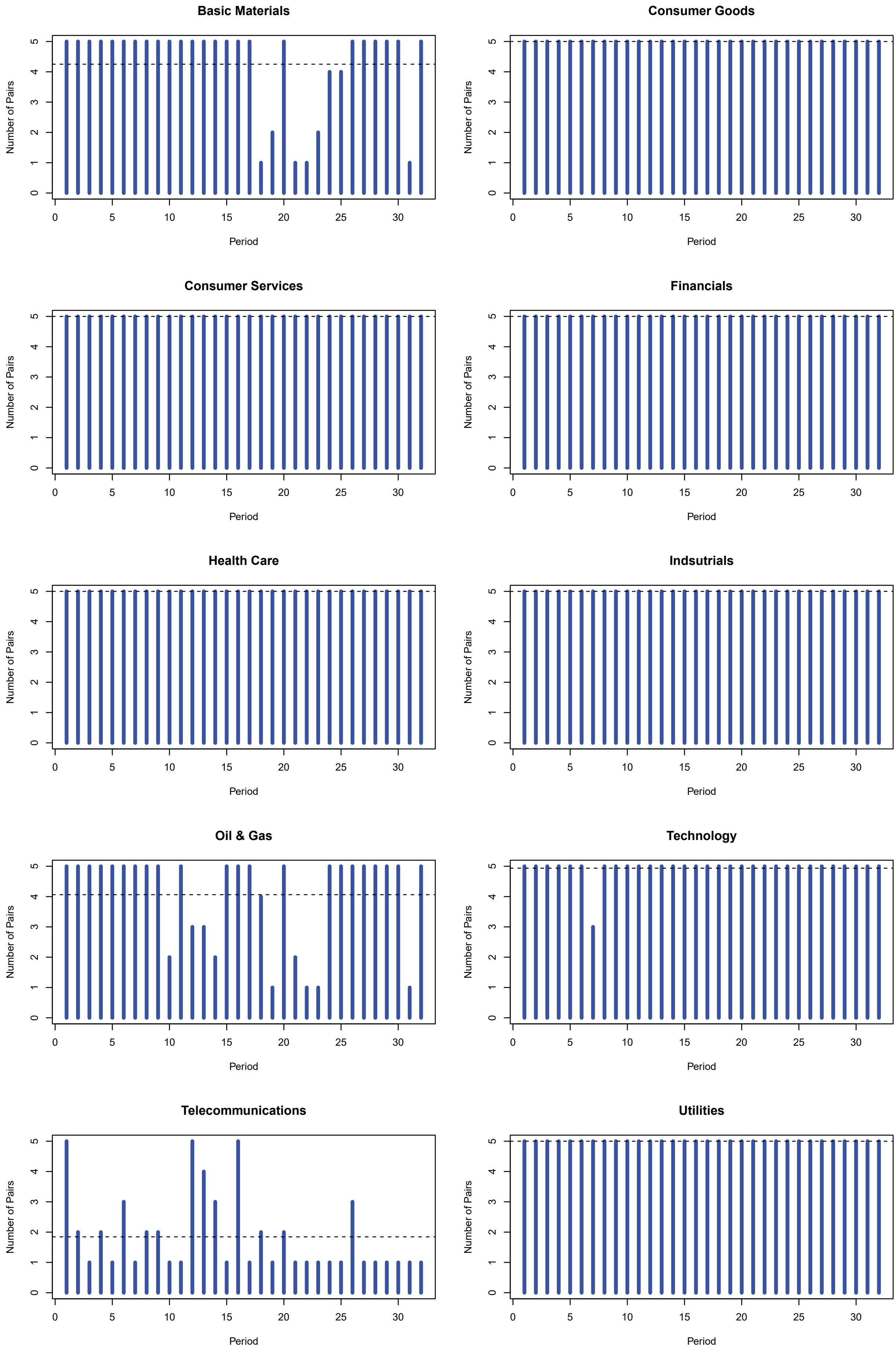

Figure 8.1: Fixed Index Grouping: Number of Pairs per Period and Industry.

pairs in the corresponding sub-industry over all periods.

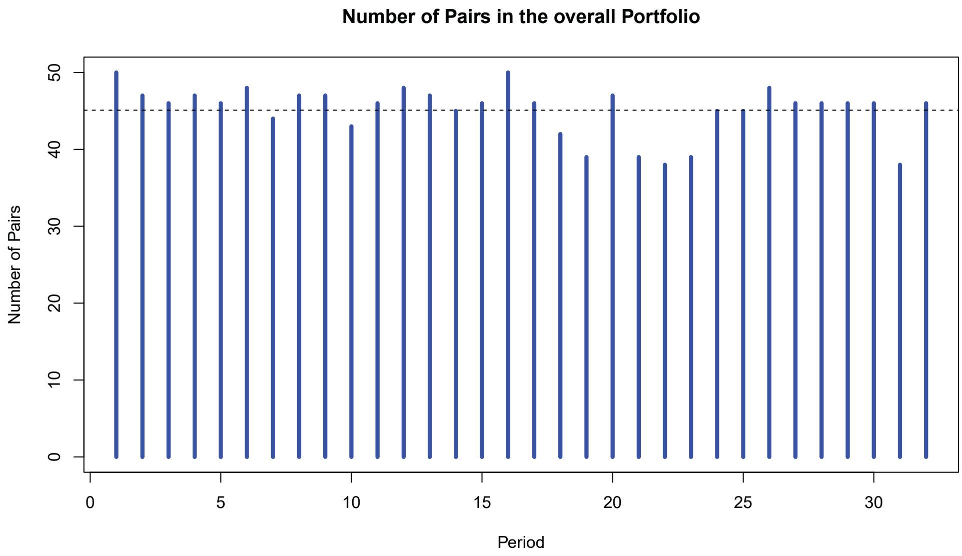

Figure 8.2: Fixed Index Grouping: Number of Pairs per Period.

Figure 8.2 depicts the total number of pairs per period for the whole portfolio. It is important to note that the depicted numbers are just the numbers of pairs that are allowed to be traded according to our trading rule. It is, however, possible that some pairs will never trade during the whole trading period, i.e. they remain always closed. To show this, we plot the value developments of the different industry-pairs (without considering interest rates for the moment) for the first trading period in Figure 8.3. In the graph for the *Industrials*, for instance, we can see such a case, where some pairs do not open during the whole trading period.

In the graphs of Figure 8.3 we also see curves that move closely together or lie even perfectly on each other. These are in fact pairs consisting of the same assets. The trading trigger may, however, be different. This is due to the fact that we test for cointegration in both directions, i.e. we use for each pair two different test equations. The test equations serve then as the base for the trading trigger. An obvious question to be asked is, of course, whether we should allow the same pair for trading twice. One could argue that in this case we focus too much on a particular pair. We could, however, also say that a pair that has a significant test result in either way of the test, and both test statistics rank among the highest, has a higher credibility of being a profitable match and we should therefore put a bit more weight on

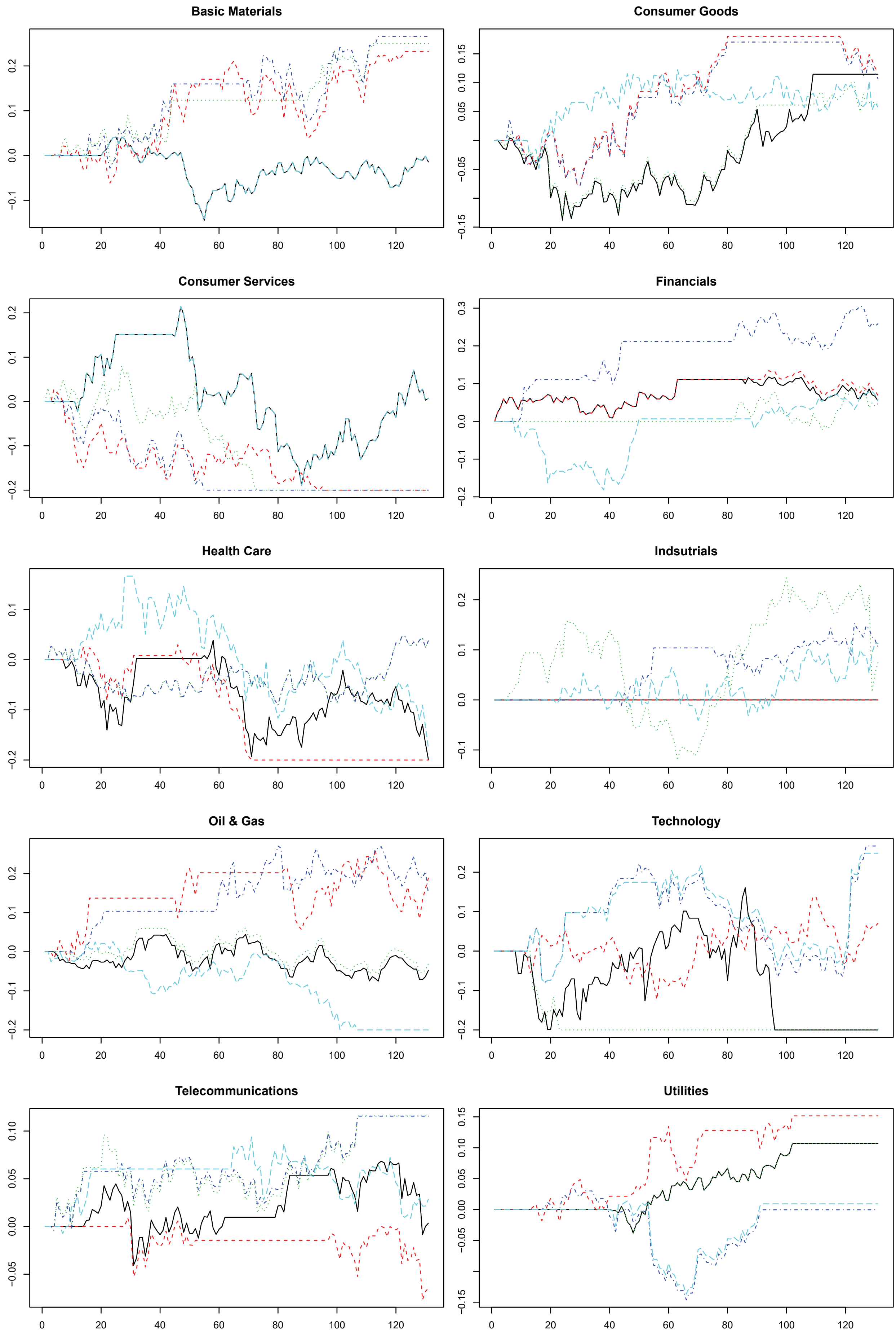

Figure 8.3: Pairs Development per Industry in the 1st Period.

it. In addition, having the same pair twice in the trading set, can also have an additional advantage. If the trading triggers are similar enough, we could potentially reduce the number of transactions by investing just the double amount in that pair. As already mentioned, the number of transactions can be critical for the whole strategy. Clearly, there is no final right or wrong. At the end of the day it is a question of personal believes whether one should put such a restriction in place or not. For this analysis we do not further restrict our set of admissible pairs and allow pairs to appear twice in any trading period.

Figure 8.4 depicts the number of transactions for each industry and period. The dashed line indicates the average number of transactions over all periods for that industry. Figure 8.5 visualises for each period the total number of transactions over all industries, where the dashed line again indicates the average number of transactions over all periods. As we can see there, in the first six months trading period there are in total 344 transactions. This number considers all transactions, i.e. also the final transaction at the end of the period when we ultimately close every remaining open pair. So, when having just one pair that opens only once during the six month trading period, we will have four transactions. This is, two transactions when opening the pair, by buying one asset and selling the corresponding pair-asset, and again two transactions when the trade is finally closed. The reported numbers are, thus, always multiples of four. Over all, we have on average 251 transactions per period, which means that we roughly have two transactions per day. The number of transactions varies, though, quite a bit over time, from 172 up to 408.

Regarding the performance of this trading strategy, we see in Figure 8.6 the value development of one invested dollar in this fund over the considered period from the beginning of 1996 until the end of 2011. In comparison to the S&P 500 index the value develops almost linearly. The fund's value at the end of the considered period is 2.82, which corresponds to an average yearly return of 6.70%. Transaction costs are not yet considered here as these costs can vary substantially, depending on the circumstances of a particular fund operator. It is, however, clear that transactions are not for free and will, thus, have an effect on the fund's performance. Another interesting feature is that the development of the fund seems to be almost independent of the

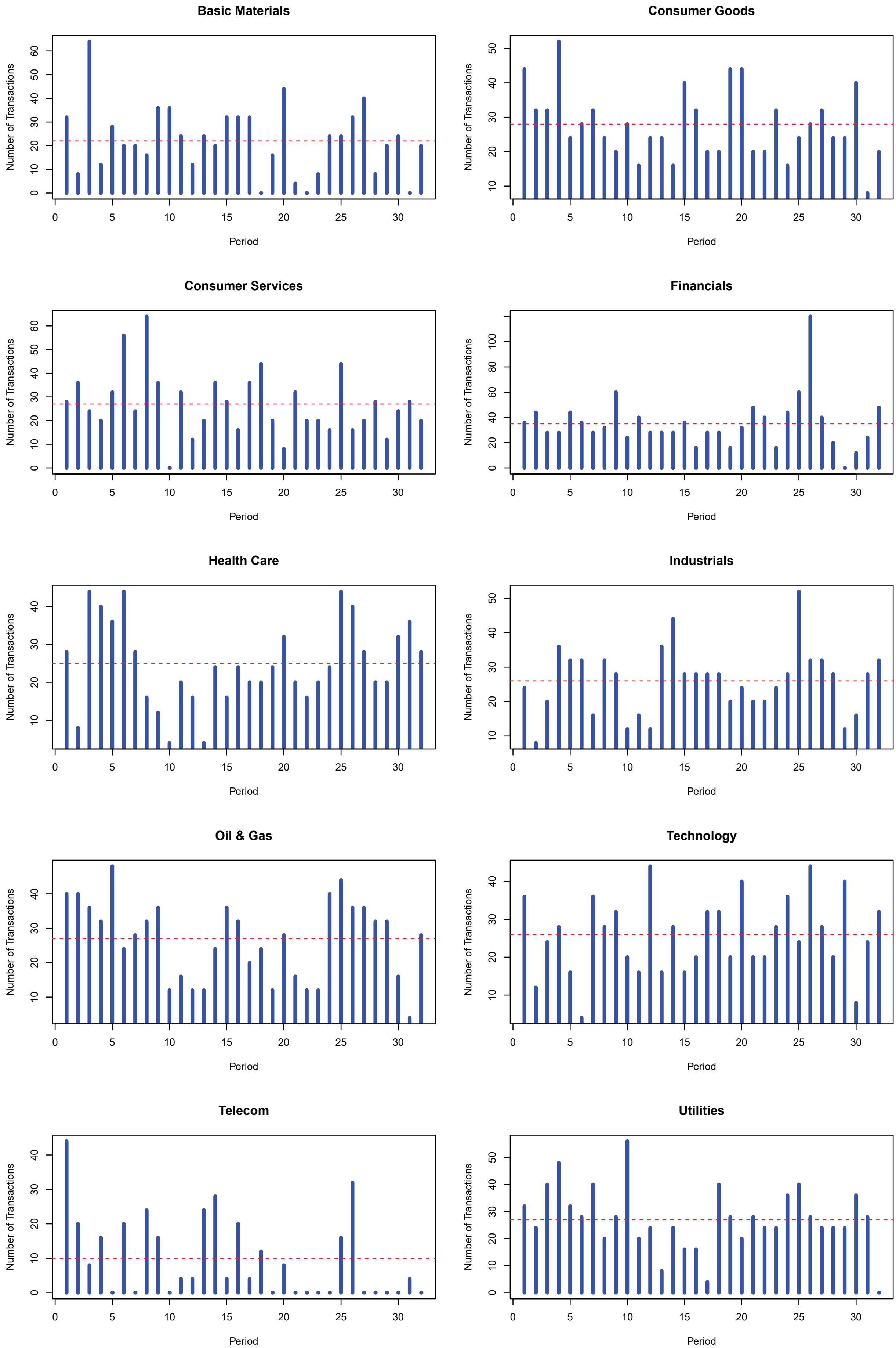

Figure 8.4: Number of Transactions per Industry and Period.

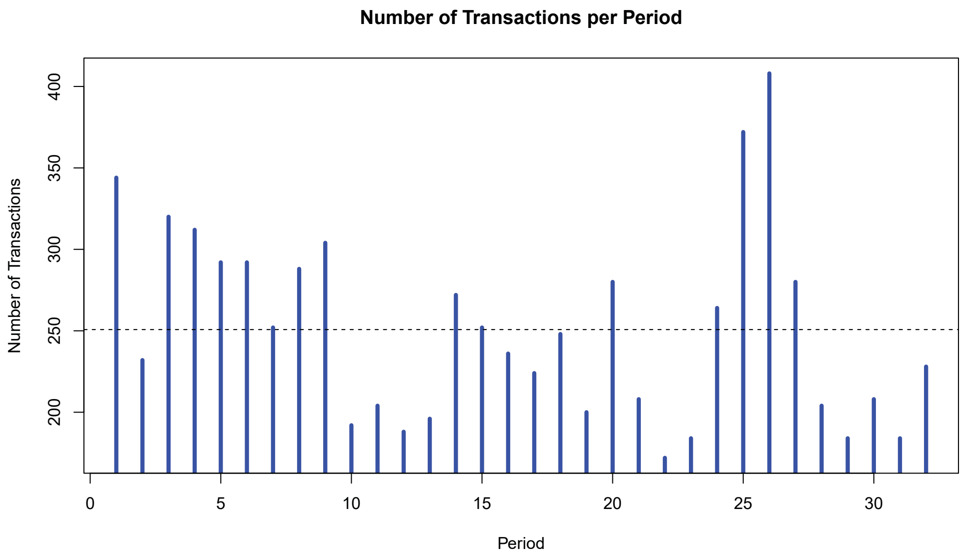

Figure 8.5: Number of Transactions per Period.

development of the S&P 500 index. In fact, the linear correlation of the fund returns and the returns of the S&P 500 index is only 0.0739. By examining the two trajectories in Figure 8.6 more closely, it stands out that the price path of the fund fluctuates considerably less than the price path of the S&P 500 index. Indeed, the standard deviation of the fund returns scaled to one year is a mere 5.01% as compared to the S&P 500 index returns with a four times higher number, i.e. 20.84%. Figure 8.18 illustrates this comparison graphically.

A similar comparison can be done on a lower level too, i.e. on the value development of the fund's industry components. So, we can also compare the industry contributions to the overall fund performance with the corresponding *FTSE* industry index. Figure 8.7 shows indeed a very interesting picture as it reveals the contributions of the different industry components to the overall fund performance. It shows us that the positive contributions come from *Consumer Goods*, *Consumer Services*, *Financials*, *Oil & Gas*, *Telecommunications* as well as *Technology*. Even though the two sectors *Industrials* and *Basic Materials* have positive values at the end of the back-testing period, the values cannot beat an investment in the risk-free rate. Even worse is the performance of the *Health Care* and the *Utilities* sector. There we even lose money. Interestingly, both industries started quite promising at the beginning of the back-testing period. Towards the end of the year 1999,

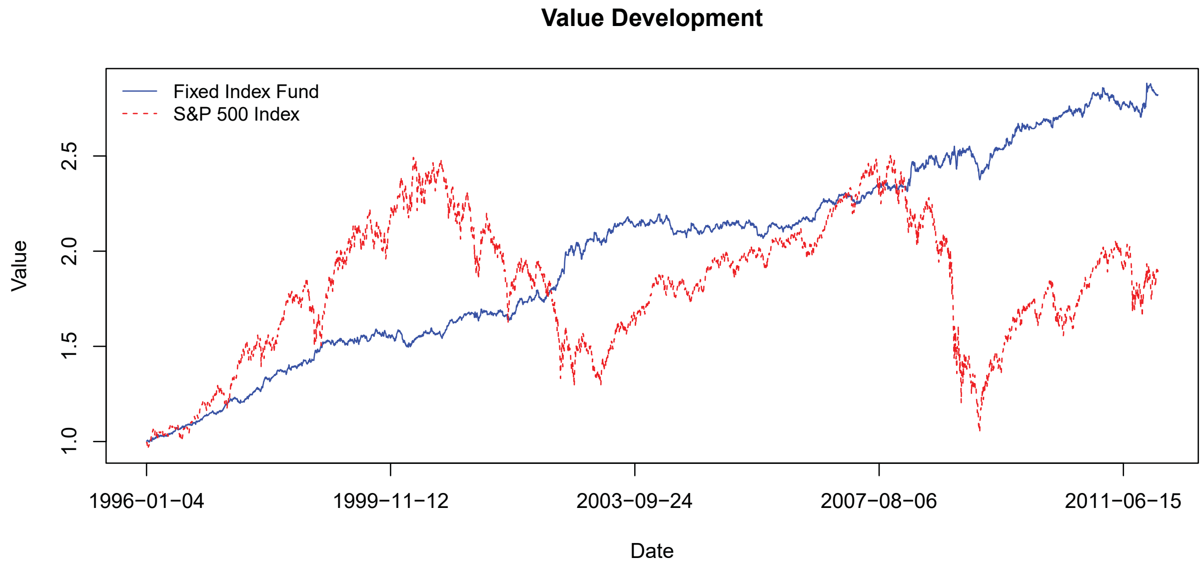

Figure 8.6: Fixed Index Grouping: Value Development.

however, both sub-portfolios started to become unprofitable.

Period 13	**Company 1**	**Company 2**	**Coint. Stat.**
Pair 1	Bellsouth Corp	Qwest Communication Inc	-4.3179
Pair 2	Sprint PCS	Century Link Inc	-4.1394
Pair 3	Bellsouth Corp	Nextel Communications Inc	-4.1069
Pair 4	Sprint	AT&T Inc	-4.0953

Table 8.1: Telecommunication Pairs of Period 13.

Period 14	**Company 1**	**Company 2**	**Coint. Stat.**
Pair 1	Qwest Communication Inc	AT&T Inc	-4.1701
Pair 2	AT&T Inc	Qwest Communication Inc	-4.1683
Pair 3	Verizon Communications Inc	Qwest Communication Inc	-4.0641

Table 8.2: Telecommunication Pairs of Period 14.

Period 25	**Company 1**	**Company 2**	**Coint. Stat.**
Pair 1	Century Link Inc	Sprint Nextel Corp	-3.9673

Table 8.3: Telecommunication Pairs of Period 25.

Quite an interesting value development can be observed in the *Telecommunications* sector. As can be seen in the corresponding plot of Figure 8.4,

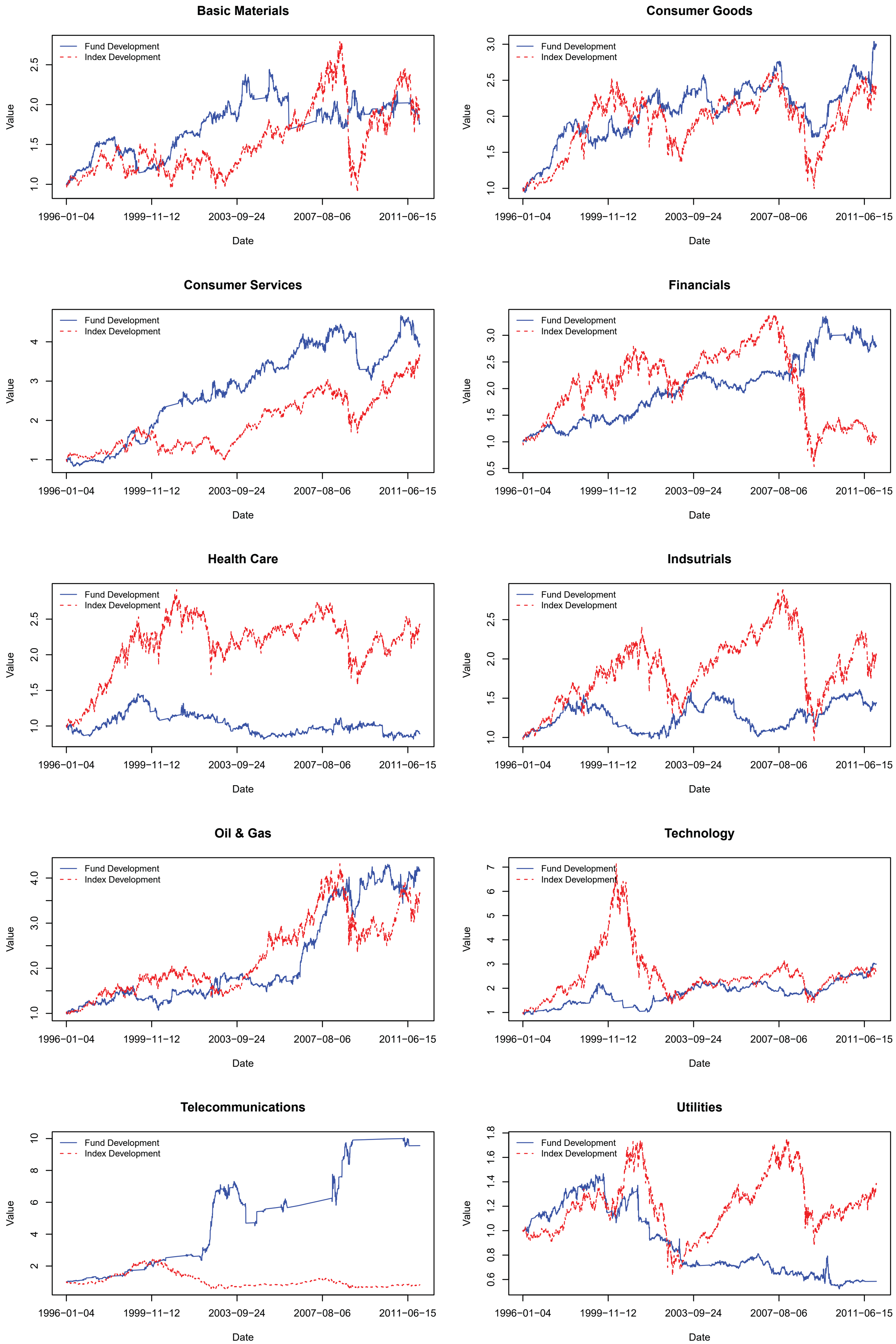

Figure 8.7: Fixed Index Grouping: Sub-Industry Value Development.

there are several periods in the whole back-testing time span without any transactions. This is also visible in the value development curve. Although the strategy is highly profitable in this industry, it is striking that the main contributions to the top performance come from two sharp increases in the portfolio value in period 13, 14 (i.e. in the year 2002) and period 25 (i.e. in the first half of the year 2008). In fact, the industry portfolio return in period 13 is +38.56% and in period 14 it is a whopping +83.38%, which means that we have an overall performance in the year 2002 of 154.09% on the Telecommunication portfolio. In the first half of the year 2008 (i.e. period 25) we observe a return of +36.12%. An obvious question is, thus, whether there is something wrong here or whether this is just a rare but, nevertheless, plausible strategy inherent feature. Table 8.1, 8.2 and 8.3 report the corresponding pairs and companies involved in the trading of the periods 13, 14 and 25 together with the cointegration test statistic obtained in the precedent fitting period for each pair.

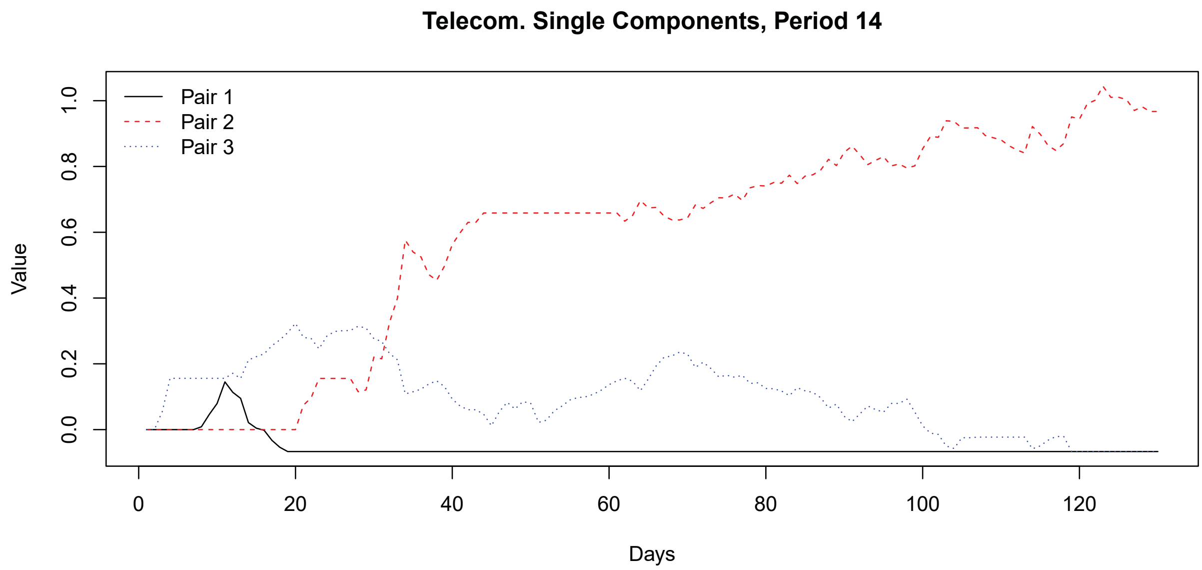

Figure 8.8: Development of Telecom. Single Components in Period 14.

Inspecting the three tables does, however, not reveal any considerable irregularities. The pairs are made of independent companies and the corresponding cointegration statistics show appropriate values. However, an important point to note here is the number of pairs available for trading. In period 13 there are four pairs that we could possibly trade. In period 14 we have only 3 pairs to trade, where two of them even consist of the same two companies.

In period 25 there is merely one pair available for trading. The number of admissible pairs is crucial with respect to the capital we put in a single pair. If we have one dollar available for a particular industry group and five admissible pairs in that group, we reserve 0.2 dollars (20%) for each pair. If only one pair opens during the whole trading period, we will have used only 20% of the available trading capital. 80% of our funds performs in this case just according to the risk-free rate. If we have, however, only one tradable pair available, we will allocate 100% of our capital to that trade if the pair opens. This can make the value development of a group quite volatile or jumpy. While this may explain the sharp increase in value of the Telecommunication portfolio for period 25 satisfactorily, it may not be fully convincing in terms of explaining the value development in period 13 and 14. By examining the three pairs in period 14 (where we observe an increase in value of 83.38%) more closely, we observe that the big increase stems from Pair 2 (i.e. AT&T Inc and Qwest Communication Inc) only. Figure 8.8 shows the development of one dollar invested to equal parts in the three pairs. While pair 1 and 3 are knocked-out by the imposed stop-loss rule, the value of pair 2 quadruples. At the end of the period we obtain minus \$ 0.07 from pair 1, plus \$ 0.97 from pair 2 and minus \$ 0.07 from pair 3, leaving us with a total of plus \$ 0.83 for an investment of \$ 1. This illustrates that we may have such extreme developments, even if they are rather exceptional.

8.2.2 Flexible Index Grouping

In Chapter 6 we argued that it might be beneficial to have a less rigid grouping. In Section 6.3 we, therefore, introduced a procedure to classify the assets at the beginning of each trading period according to their industry exposure. As the returns of all *FTSE* industry indexes are considerably influenced by the overall market, it is suggested to remove the market component from the industry index returns before running (6.3.1). So, for every trading period we orthogonalise the observed returns of each industry with respect to the S&P 500 index returns first, before we regress the returns of our assets on the 10 industry returns to obtain the exposure vectors $\boldsymbol{\beta}$.

Having estimated an exposure vector for each asset of our asset universe, we have to focus on the classification mechanism in order to come up with an appropriate grouping. Following our argumentation of Section 6.3, we employ (6.3.3) as dissimilarity criterion.

An important question we have already addressed briefly in Section 6.3.1, is the optimal number of groups we should consider. As discussed in Section 6.3.2, the silhouette plots may serve as valuable indicators for an adequate partitioning. Typically we would have a look at these plots before every new trading period and decide according to them. However, in order to make our results reproducible, we should employ a quantitative decision criteria such as the average silhouette width as briefly mentioned in Section 6.3.2. A good partition would typically show a high average silhouette width. So, we can check this criteria at the beginning of each trading period for different numbers of groups and then take the number with the highest average silhouette width. This means, though, that the number of groups may change form one period to the next period as we update our information every period. There is, in principle, nothing wrong with a dynamic number of groups. We should, however, put some restrictions in place to ensure some basic requirements. In the Fixed Index Grouping approach we had 10 groups that were given to us by definition. If we assume for the moment that the 500 assets of our asset universe were equally distributed among the 10 groups, we would have 50 assets per group, which means that we would still have 1,225 possible asset combinations in each group. This is still a large number and we would rather prefer a smaller one. So, it makes sense to set a minimum number of groups we are willing to accept. Similarly, it is also advisable to set an upper limit in order to avoid any ill partitioning with groups containing only one asset. So, a good suggestion may be to set the lower boundary to 10 and the upper boundary to 100.

When having a look at some random samples with regard to the 260 days testing-period before the first trading period, it is not surprising to see that the largest component in the exposure vector is usually the one for the industry where the asset is allocated according to the FTSE Industry Classification Benchmark.

Figure 8.9 depicts the number of groups we choose for each period on the base of the mentioned criteria. The black dashed line indicates the average number of groups over all periods, i.e. roughly 13 groups. More details can be

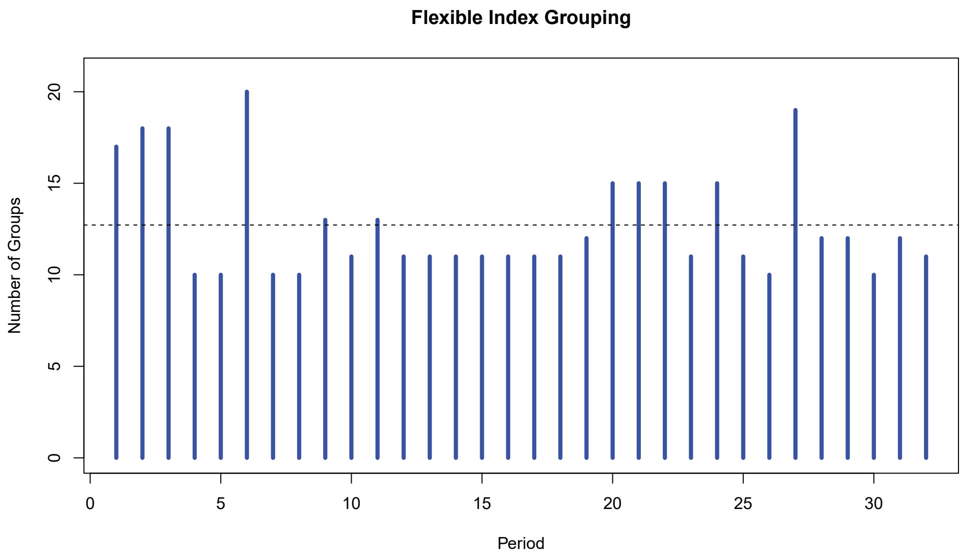

Figure 8.9: Flexible Index Grouping: Number of Groups per Period.

found in Appendix A by inspecting Figure A.1, A.2, A.3 and A.4, where the average silhouette widths are plotted for each period for different numbers of groups, i.e. for numbers between 8 and 50. It is worth mentioning that the upper boundary of 100 groups is never binding. However, the limit we set on the lower end has an influence on the chosen number of groups.

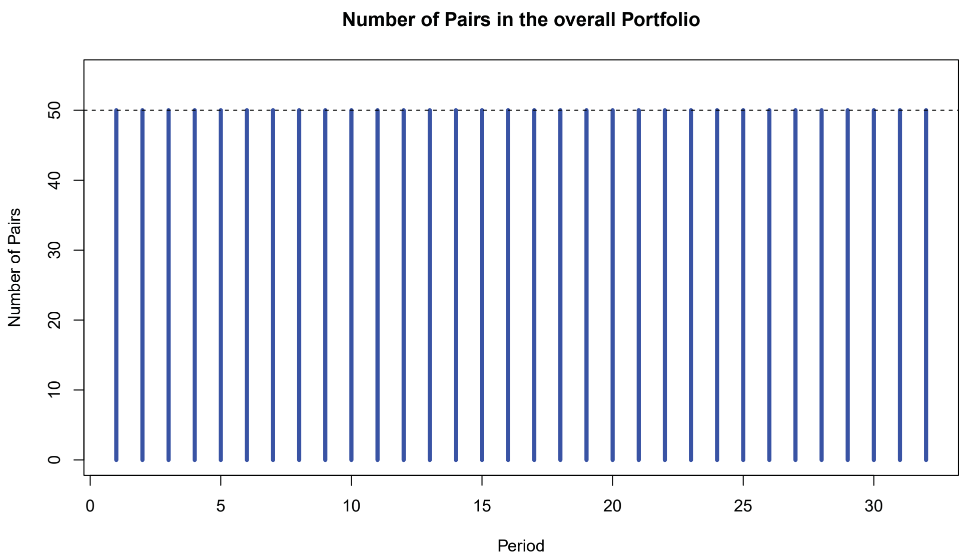

Figure 8.10: Flexible Index Grouping: Number of Pairs per Period.

Similar to the Fixed Index Grouping strategy we impose for the strategy here also a restriction with respect to the number of pairs available for trading. It makes sense to use the same overall restriction of 50 pairs. In the Fixed

Index Grouping strategy we went even one step further and set a sub-limit on each industry. In this setting, however, this does not make too much sense anymore as we have no naturally defined groups anymore. To be more precise, a particular company can now be allocated to group 5 in period 1 and to group 10 in period 2. Being in a setting with a changing number of groups also means that two companies can be in the same group in one period but in different groups in the subsequent period. So we leave it at the restriction of 50 pairs on the overall portfolio. Figure 8.10 shows again the number of pairs per period in the overall portfolio. In the case here as opposed to the Fixed Index Grouping strategy, the restriction of 50 pairs is always binding.

Figure 8.10 shows again the number of transactions for the 32 trading periods. On average we have 304 transactions per period, which is about 20% more than what we observed in the Fixed Index Grouping strategy. Striking is the large jump in the number of transactions in period 26, i.e. during the second half of the year 2008). There we observe 800 transactions, which is more than 6 transactions per day. Such a peak is also visible in the Fixed Index Grouping strategy. There as well, the number of transactions increase sharply in the periods 25 and 26 (i.e. in the year 2008), although not as high as in this strategy here.

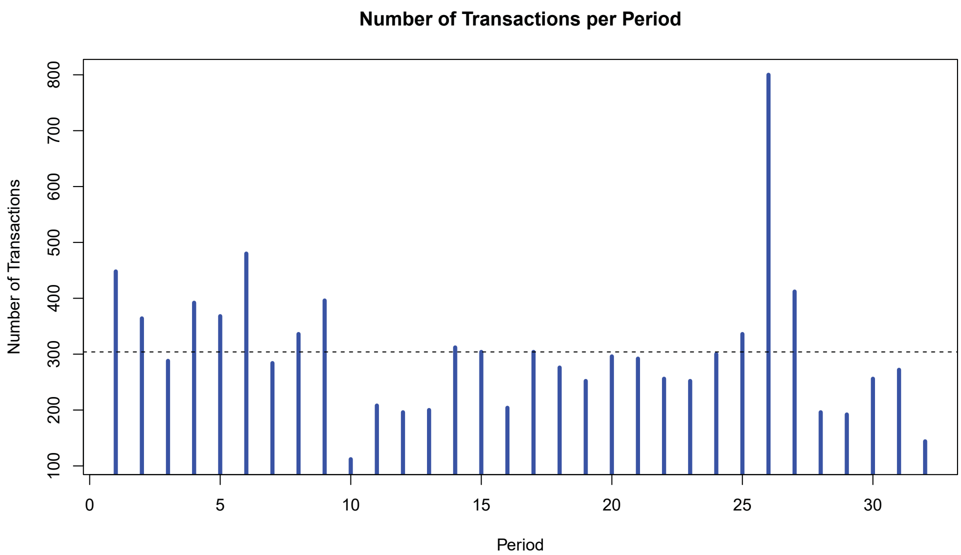

Figure 8.11: Flexible Index Grouping: Number of Transactions per Period.

In order to see how one invested dollar would have performed in comparison to the corresponding development of the S&P 500 index, we can have a look at Figure 8.12. The fund's end value is 2.94, which corresponds to an average yearly return of 6.97%, before considering transaction cost. This is a bit more than what we get with the Fixed Index Grouping strategy of the previous section. However, the better result of this strategy results mainly from a sharp increase in value during the financial crisis in the year 2008. The yearly standard deviation of the returns is with 6.87% a bit higher than the one of the Fixed Index Grouping strategy but still far smaller than the 20.84% of the S&P 500 index. The linear correlation of the Flexible Index Grouping strategy returns with the S&P 500 index returns is with 0.0736 also quite low.

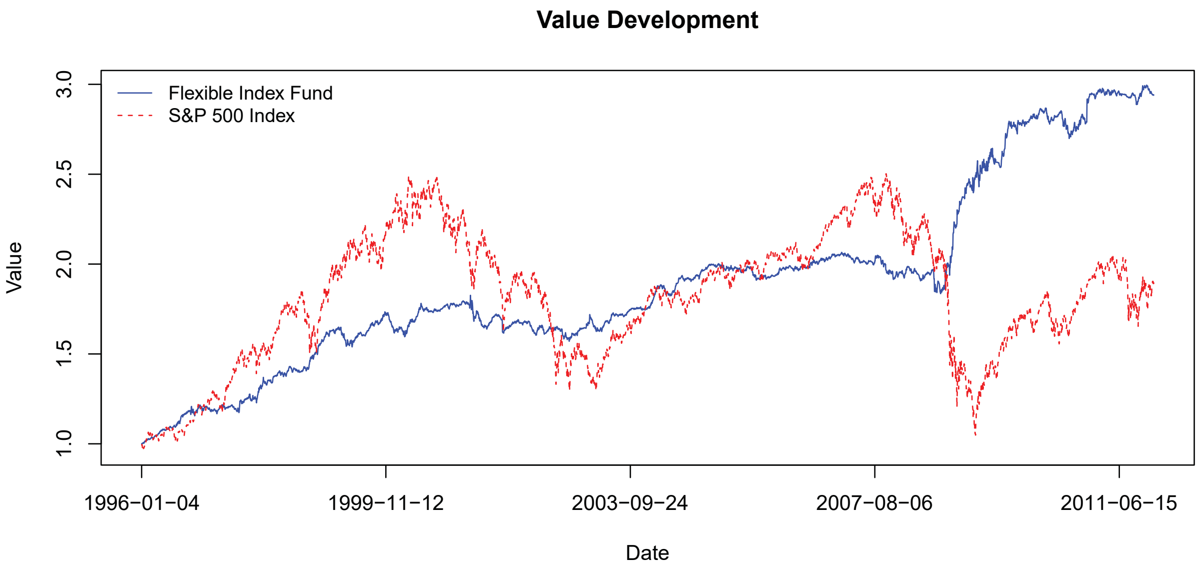

Figure 8.12: Flexible Index Grouping: Value Development.

8.2.3 Flexible Grouping Based on Fundamentals

Another way of grouping based on fundamental values and its underlying idea of the common trend model was proposed in Section 6.5. Applying such a grouping technique requires to answer the important question of which factors to be used in (6.5.1), so that the model requirements are fulfilled as good as possible.

Finding "good" factors in an empirical way is obviously inappropriate in a back-testing study. We, therefore, have to come up with a good guess based on theoretical considerations and evaluate our choice then ex-post.

A possible set of factors could be the following ones:

i.) Market Index

ii.) Company Size Index (Small Minus Big)

iii.) Company Value / Growth Index (High Minus Low book to market ratio)

iv.) Company Quality Index (Quality Minus Junk)

v.) Credit-Spread

vi.) Term-Spread

vii.) USD FX (trade-weighted)

viii.) Real Estate Index

ix.) Oil Price

x.) Volatility Index (VIX)

The first factor is motivated by the well known Capital Asset Pricing Model (CAPM). The second two factors are the Fama-French factors according to Fama and French (1993). The SMB factor is said to measure the excess returns of small companies over big companies. The HML factor is thought to measure the excess returns of *value stocks* over *growth stocks*. Fama and French (1993) consider higher returns as a reward for bearing higher risks. This is, if we see higher returns for higher book to market ratios, then shares with a higher book to market ratio must be riskier. Smaller companies can be seen as being riskier than larger companies because they are typically less diversified and have less market power. So, they are more prone to become bankrupt in an economic downturn. Companies with a low book to market ratio are typically companies, where investors assume a strong growth. Therefore, they are often called *growth stocks*. This is usually characteristic for a whole business sector. On the other side, there are the so-called *value stocks*. These are companies usually operating in a well established and more or less saturated business sector. An exceptionally high book to market ratio can also mean that a company is in distress or that investors see some problems with respect to future earnings.

The company quality factor is based on the work of Asness, Frazzini, and Pedersen (2014). They establish a high quality security for which it is assumed that investors would typically pay a higher price as the underlying company is considered safe, profitable, growing and well managed. Their Quality Minus Junk (QMJ) factor goes long in high-quality shares and short in low quality shares. Andrea Frazzini (Frazzini) provides a data library with time series of the QMJ factor according to Asness et al. (2014) together with time series of the Fama-Franch factors.[1] The time series used in the empirical analysis of this work are taken from their data library.

The credit-spread is thought to measure the propensity of default of a particular company. If the spread between AAA-corporate bonds and US government bonds widens, it is expected that some companies react more sensitive to this than others. The ones that do not react too much, i.e. the ones with a small exposure, are probably thought to have a characteristic that is economically more sound.

The term-spread usually gives us some information about the current economic environment and the market participant's expectations with respect to future economic conditions. As with the other proxies to measure the economic conditions, every company can be affected in a different way, depending on its risk profile.

The trade-weighted USD foreign exchange is assumed to measure a company's exposure towards exports. The idea is to distinguish between companies with a more import / export oriented profile and companies that buy and sell their products mainly in the US. A company that mainly provides local services in the US has clearly less exposure to that factor than a global goods trader. Two such companies will, thus, have a different risk profile and this is exactly what we try to distinguish.

The real estate index should measure the exposure with respect to the US housing market. There might be some criticism with this index. One could argue that we only consider the housing market exposure because of the recently experienced housing bust, which is certainly very problematic in a back-testing study. This is clearly a valid point. However, in order to distinguish service companies, in particular banks, with respect to their economic activities, the housing exposure is an essential factor. We could also argue that housing crises have already occurred in many other countries before and

[1] Data Source: `http://www.econ.yale.edu/~af227/data_library.htm`

that they occur in general quite frequently. It is, thus, a factor that should always be considered.

Including the oil price as a factor can also be justified. Oil is a commodity that is relevant for many industries. As we observe, and have always observed quite frequently, bigger shocks in the oil market, the oil price dependency of a company is clearly a relevant feature of that company's profile.

Finally, there is an index with respect to the implied volatility in the market. A well known index for this purpose is the VIX. The implied volatility can be seen as a measure of financial market turbulences. It usually rises whenever there is uncertainty in the market. The exposure to the VIX is, thus, assumed to capture the robustness of a company with respect to difficult financial market environments.

Figure 8.13 shows the bivariate dependence structure of the discussed factors for the first fitting period. As one would expect, there are some factors that are correlated with the market. Especially the VIX shows quite some negative correlation with the market. A less pronounced, but still noticeable correlation structure is observed for the HML, the SMB, the term-spread and the property index. So, we could think again of an orthogonalisation of these factors with respect to the market for every period. However, an orthogonalisation is an interference into the data structure and should only be done with reservation. As roughly half of the factors are less or even completely unproblematic regarding their correlation structure with the market and the factors besides the market factor show a correlation structure among each other that can be neglected, we will abstain from the mentioned orthogonalisation and leave all factors as they are.

The mentioned factors may not be perfect. It is finally a matter of personal believes which factors one wants to include. The proposed model of this section is admittedly simple and there is with no doubt considerable room for improvements. However, the proposed approach seems appropriate enough for a first run.

So, with the chosen factors and the correspondingly fitted parameters according to (6.5.1), we can now group the assets in compliance with (6.3.5) by using the distance measure as defined by (6.5.10). By doing so, we have again to take a decision with regard to the number of groups we thinks is

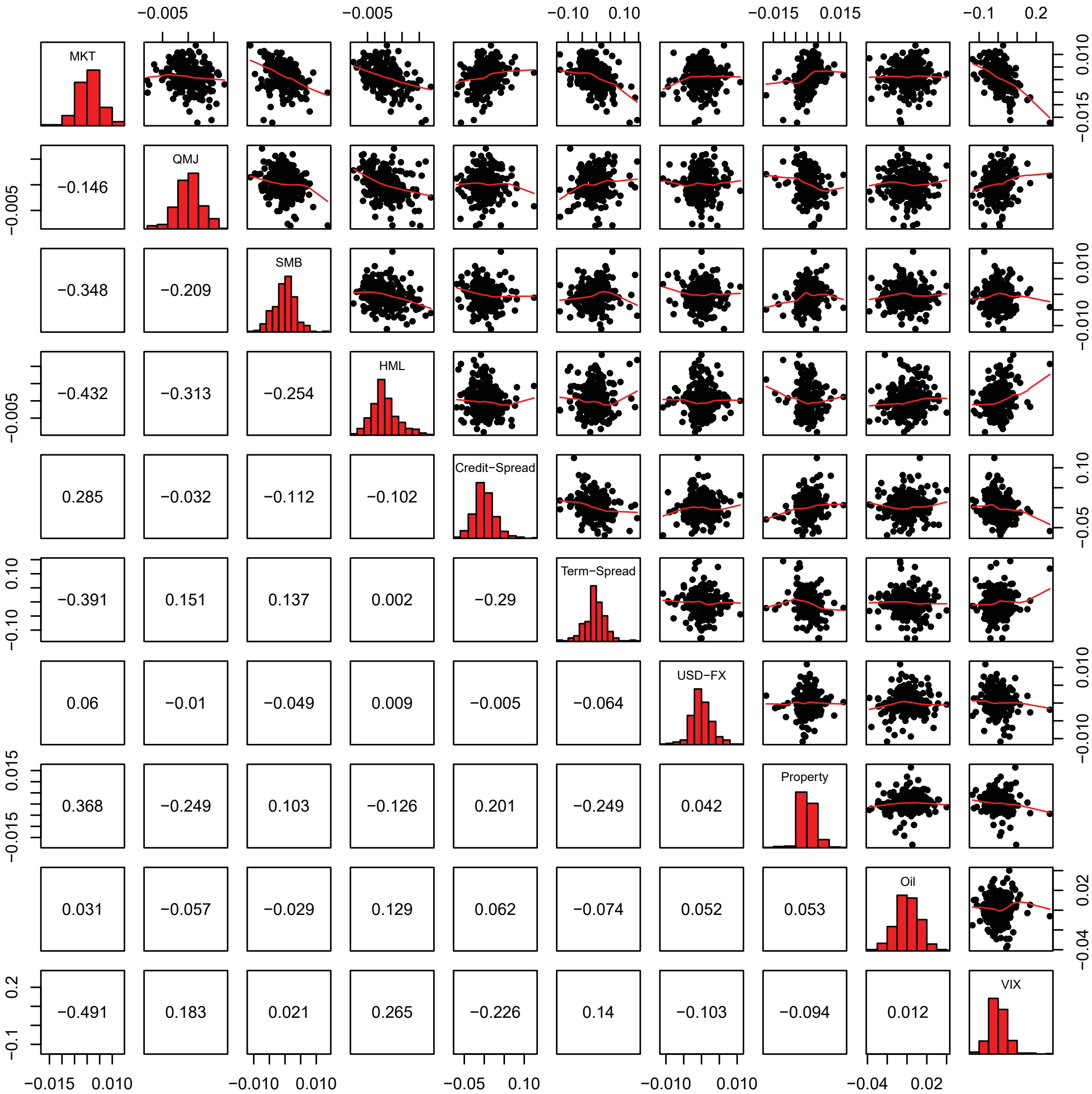

Figure 8.13: Correlation Structure between Fundamental Factors.

appropriate. To stay in line with the grouping of Section 8.2.2, we employ the same procedure based on the average silhouette width as discussed in the previous section.

Figure 8.14 shows again the number of groups we determined for each period. As can be seen there, we start with more than 50 groups in the first period. The number of groups falls, though, quite quickly towards 10 groups only. The black dashed line indicates the average number of groups over all periods, which is 13 in this case. For more details, the interested reader is again referred to Appendix A and Figure A.5, A.6, A.7 and A.8, where the average silhouette widths are plotted for different numbers of groups, i.e. for numbers of groups between 8 and 50. Similar to what we have already seen in the case of the Flexible Index Grouping strategy, the upper boundary of 100 groups is never binding. However, the lower boundary of 10 groups is binding for many periods.

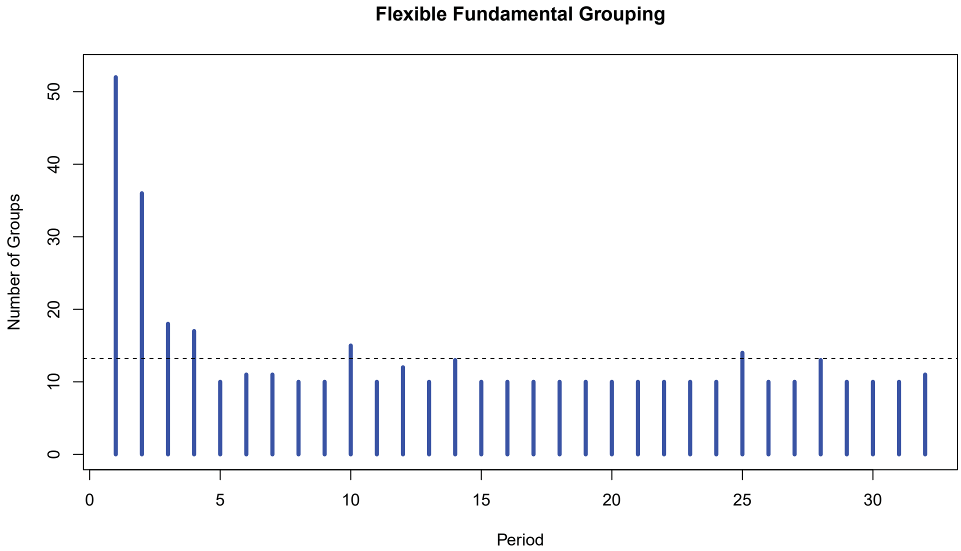

Figure 8.14: Flexible Fundamental Grouping: Number of Groups per Period.

Figure 8.15 depicts the number of pairs we consider for each of the 32 trading periods. Similarly to section 8.2.2, we choose for each period 50 pairs.

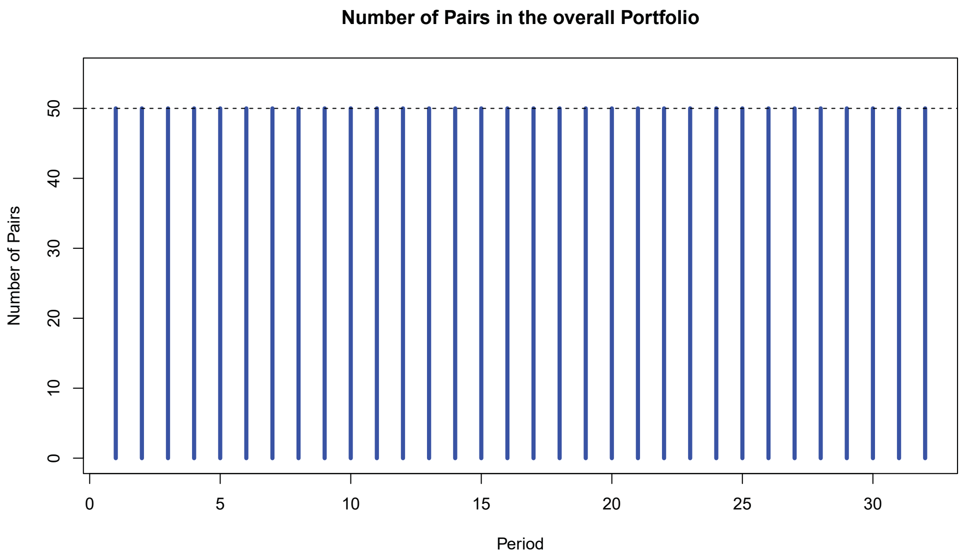

Figure 8.15: Flexible Fundamental Grouping: Number of Pairs per Period.

As in the case with the Flexible Index Grouping strategy of section 8.2.2, the groups are determined in a dynamic way and are allowed to change their characteristic from period to period. It is, therefore, not reasonable to impose any sub-limits with respect to the number of admissible pairs on a group level.

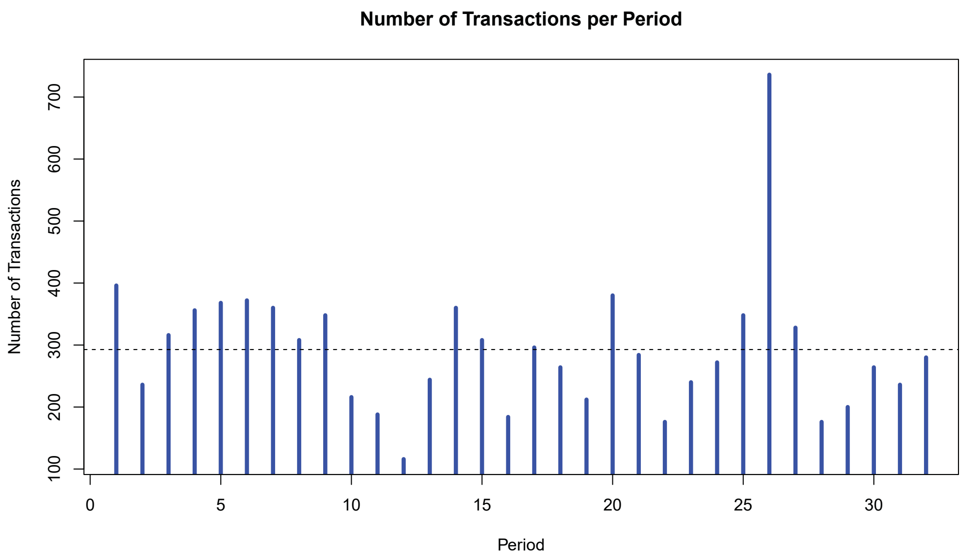

Figure 8.16: Flexible Fundamental Grouping: Number of Transactions per Period.

Figure 8.16 depicts the number of transactions for each period. On average we have 299 transactions per period, which is very similar to the average number of transactions observed for the Flexible Index Grouping strategy. Also the transaction pattern looks very similar to the one of the previous section, with a pronounced peak in period 26 (i.e. in the second half of the year 2008).

The development of one invested dollar in comparison to the corresponding development of the S&P 500 index is depicted in Figure 8.17. The fund's end value (before accounting for transaction cost) is 2.95, which corresponds to an average yearly return of 7.00%. This is the best result of all the proposed strategies, including the S&P 500 index, although it is only slightly higher than the end value of the Flexible Index Grouping strategy. The yearly standard deviation of the returns is with 6.52% a bit smaller than the one of the Flexible Index Grouping strategy but clearly larger than standard deviation of the Fixed Index Grouping strategy of Section 8.2.1.

Similar to what we have already observed in the previous section, the value of the fund increases sharply in the year 2008. The linear correlation with the S&P 500 index returns is with a value of 0.0456 again very low.

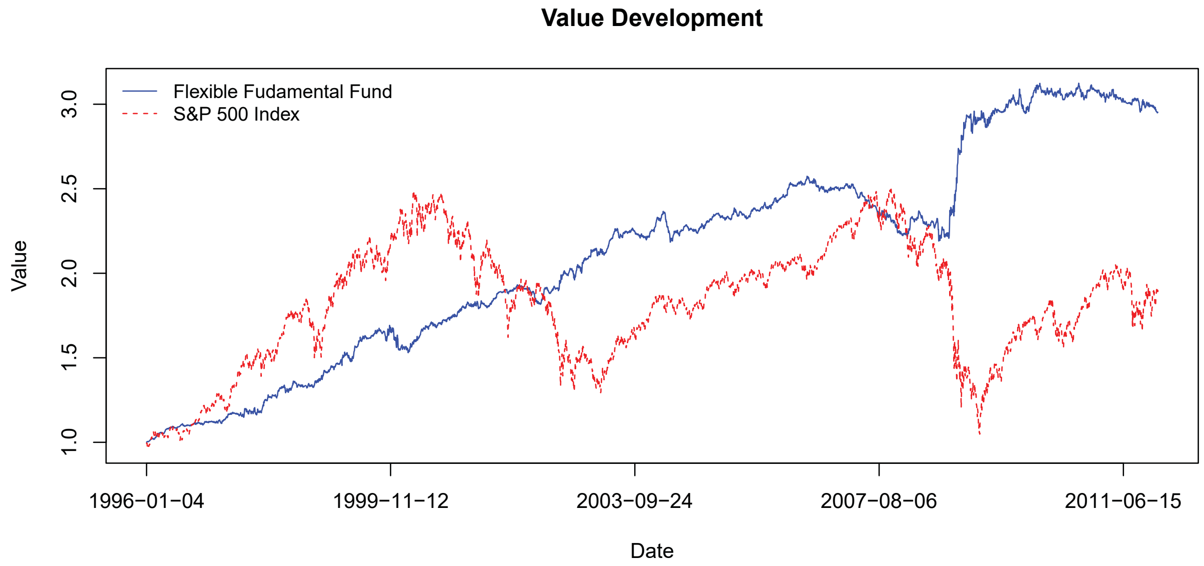

Figure 8.17: Flexible Fundamental Grouping: Value Development.

8.2.4 Summary of the Results

Table 8.4 summarises the results of an investment in the three proposed fund types as well as in the S&P 500 index.

	Fix. Index	Flex. Index	Flex. Fundam.	S&P 500
Daily Mean	0.00025	0.00026	0.00026	0.00017
Yearly Mean	0.06698	0.06974	0.07000	0.04506
Daily Mean Excess	0.00012	0.00013	0.00013	0.00006
Yearly Mean Excess	0.03114	0.03381	0.03407	0.01571
Daily Stand. Dev.	0.00311	0.00427	0.00405	0.01296
Yearly Stand. Dev.	0.05010	0.06868	0.06519	0.20839
Daily Min.	-0.01516	-0.03995	-0.03296	-0.09035
Daily Max.	0.01799	0.04440	0.04295	0.11580
Skewness (daily)	0.47993	0.64149	0.87393	-0.03345
Kurtosis (daily)	5.62431	13.12635	12.98142	10.51187
Value at the end	2.82163	2.94058	2.95255	2.02408

Table 8.4: Summary statistics of the pairs trading funds

In the first two lines we report the geometric means of the daily and the yearly returns, respectively. The pairs trading funds achieve annual average returns between 6.7% and 7.0%. By comparison, an investment in the S&P 500 index yielded over the same period only 4.5% p.a. on average. An investment in the S&P 500 index performed clearly worst if we consider the whole investment period. By looking at the figures of the corresponding value developments, i.e. Figure 8.6, 8.12 and 8.17, we realise, however, that during particular sub-periods the index investment developed better than any of the proposed pairs trading funds. The mean excess return is just the return in excess of the risk-free rate. The corresponding excess returns are reported in line 3 and 4 of Table 8.4.

Apart from achieving a high investment return, we are also interested, by how much our returns fluctuate over time. This is reported in line 5 and 6 of Table 8.4. There we show the standard deviations of the daily and yearly returns. As can be stated again, the S&P 500 index performs worst with a yearly standard deviation of 20.84%. The best result is achieved by the Fixed Index Fund strategy with a yearly standard deviation of only 5.01%. A very interesting picture in this regard is Figure 8.18. It shows the returns of all the funds and the S&P 500 index over time. So, we can see quite well, when and how strongly the returns fluctuated over the whole analysed period. Thus, an investment in the S&P 500 index is from a pure risk avoiding perspective rather less desirable. In this respect, we may also want to know what the worst return was on any trading day during the whole investment period. The corresponding number is stated in line 7 of the mentioned table. The worst outcome of the Fixed Index Fund strategy on any trading day during the observed time span is a mere -1.52%. This is very remarkable. However, also the other two pairs trading strategies perform quite well in this respect. So, non of the pairs trading funds ever had a daily return smaller than -4%. By contrast, the S&P 500 index shows a worst return that is smaller than -9%. At this point we have, however, also to mention that the highest daily return is achieved by the S&P 500 index too. It shows a maximum return of 11.58%, while the best daily return achieved by any pairs trading fund is just 4.44%. In line 9 and 10 we give some additional information about the univariate distribution of the returns, i.e. we report there the empirical skewness and the kurtosis of the observed returns. Interestingly, all the pairs trading fund returns are positively skewed, while the returns of the S&P 500 index show a negative skewness. With a kurtosis that is significantly larger

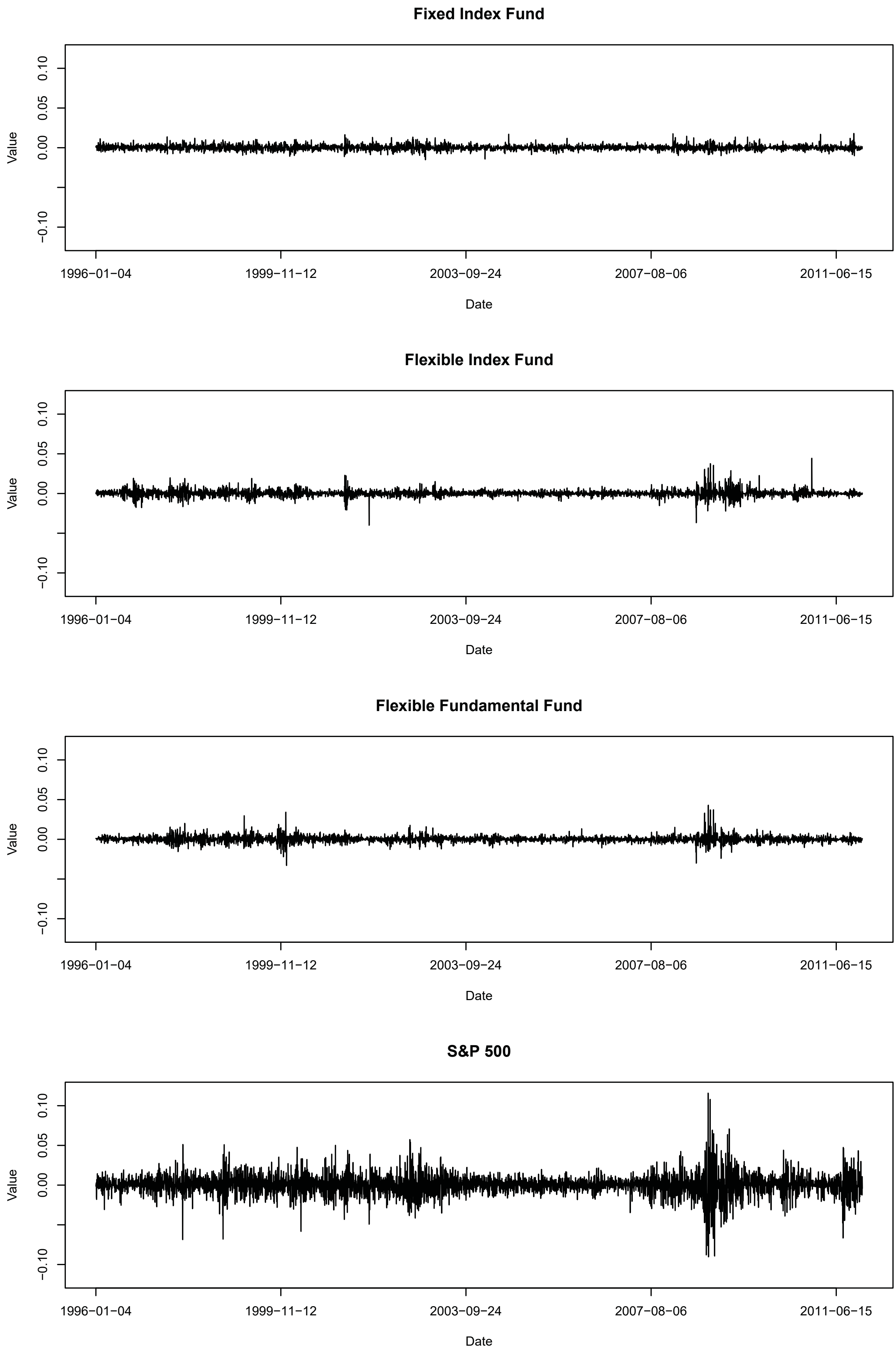

Figure 8.18: Returns Volatility Plots.

than 3, all pairs trading funds, as well as the S&P 500 index, show so-called excess kurtosis, which means that their distributions have considerably more mass in the tails than normal distribution has. This is, however, nothing atypical for daily returns. Finally, we report in line 11 of Table 8.4 for each fund the 2011 value of a $ 1.00 investment that was executed at the beginning of 1996.

Obviously, an investment yielding high returns with limited fluctuations, is attractive. Investigating the value developments, as depicted in Figure 8.6, 8.12 and 8.17, more closely, reveals another remarkable aspect. The value development of the S&P 500 index shows two major incidences with respect to financial markets. In particular, in mid 2000 we observe the so-called *dot-com bubble burst* and towards the end of 2007 we spot the so-called *sub-prime crisis*. During both crises the S&P 500 index lost pretty much in value, while the pairs trading funds were hardly affected by the dot-com bubble burst and even positively affected by the sub-prime crisis. Only the Flexible Index Fund was negatively affected by the dot-com bubble burst but to a far lesser extent than the S&P 500 index. This implies that the dependence structure between our pairs trading funds and the S&P 500 index must be quite attractive.

Figure 8.19 summarises the value developments of the three pairs trading funds again in one plot, making them better comparable. During the dot-com bubble burst in the year 2000 and slightly after that, the investments in the Flexible Index Fund and the Flexible Fundamental Fund developed differently. However, after that, i.e. from the year 2003 onwards, it is remarkable how similar the two investments evolved. Most interestingly, both investments decreased in value months before the sub-prime crisis, while the S&P 500 index was still rising. During the crisis, however, both investments jumped up sharply, while the S&P 500 index had a deep fall. Only the Fixed Index Fund followed its own rules and developed quite independently form the S&P 500 index and the other two pairs trading investments.

In the whole analysis so far, we have not consider any transaction cost yet. Incorporating transaction cost is, however, important if we want to compare different types of investments. On the one hand, we considered an investment in the S&P 500 index, which is often seen as a “passive” investment. Passive means in this sense that there is no further rule behind that needs to be

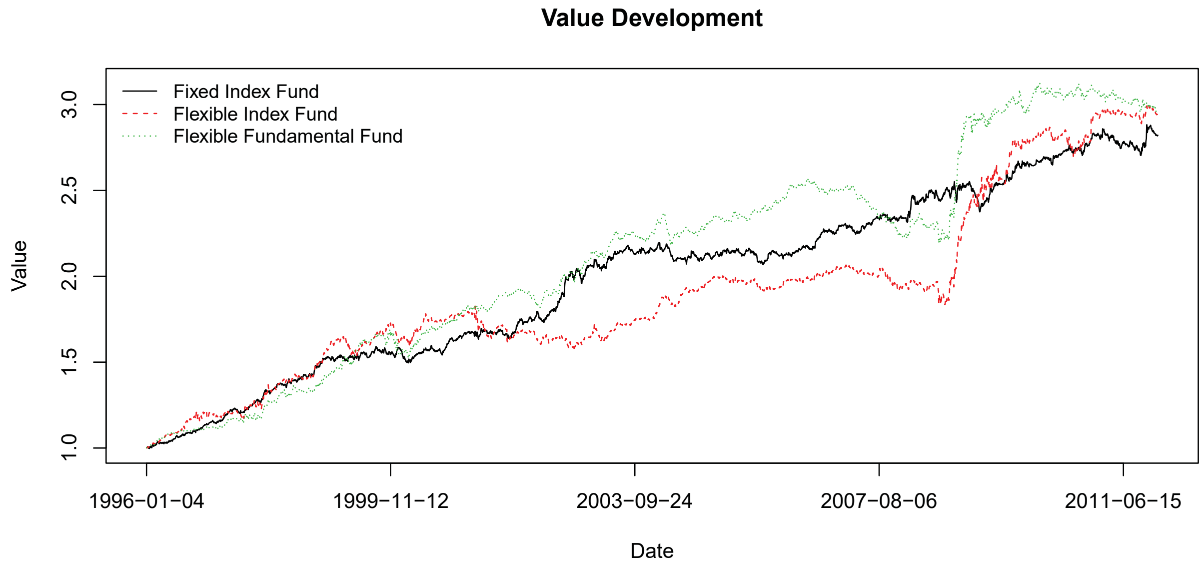

Figure 8.19: Comparison of the Value Developments.

considered. Basically, we “buy” the index at the beginning of our investment period and sell it again at the end of the period. In between there are no transactions intended. On the other hand, we have our pairs trading funds, which require a lot of transactions. As transactions are not for free, it is clear that the number of transactions have a significant effect on the performance of such investment strategies. When comparing our pairs trading strategies with an investment in the S&P 500 index, we could plead that investing money in a portfolio that replicates the S&P 500 index requires quite some transactions too. Just buying one share of each company that is part of the index would already require 500 transactions. Selling them all again at the end of the investment period would require another 500 transactions and as the composition of the index changes over time, there would be several additional transactions in between. Even though there are other options than physically replicating an index portfolio, non of them is free of charge. How much money finally has to be paid in terms of fees and commissions, is very individual and strongly depends on the concrete circumstances of an investor. Therefore, it is in general very hard to incorporate transaction costs in a study like this one here. The comparison between our pairs trading funds and the S&P 500 index has, for sure, its limitations but it is definitely not spurious.

Nevertheless, in order to judge on the actual profitability of our pairs trading

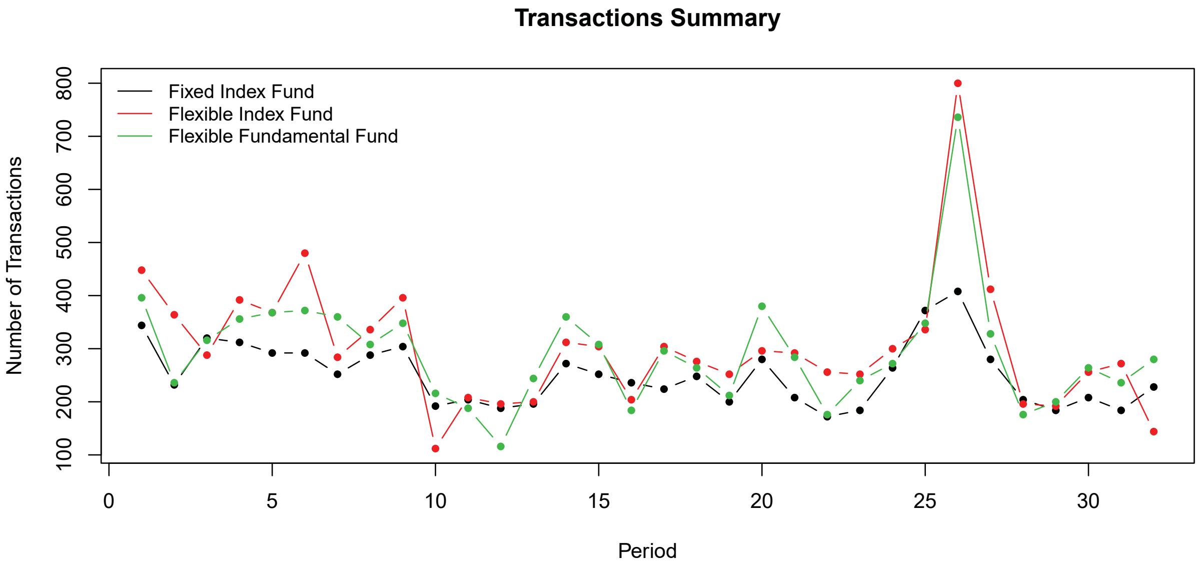

Figure 8.20: Summary: Number of Transactions.

investments, we should give some guidelines with respect to transaction cost. Figure 8.20 shows again the number of transactions of each of the three funds over the whole 32 periods lasting investment period. Table 8.5 reports the geometric means of the daily and the yearly fund returns together with the fund's values at the end of the year 2011 with incorporated transaction fees of 5, 10, 15 and 20 basis points.

It is important to keep in mind that one transaction concerns only one pair. So, if we have 50 pairs for a trading period, one transaction is only based on 2% of the fund's total value. In this respect, the number of transactions we reported for our three strategies, do not cause dramatic cost. In fact, the three proposed pairs trading strategies look still promising even after accounting for transaction cost. As can be seen in Table 8.5, if we assume transaction cost of 15 basis points, all three funds show still a better value development during the considered 16 years investment period than the S&P 500 index. Only if we have transaction cost of 20 basis points, we see that two of our funds lie below the value development of the S&P 500 index, for which we have, however, not yet considered the index replication cost as discussed before.

	Cost	Fix. Index	Flex. Index	Flex. Fundam.
Value at the end	5 bp	2.58225	2.66799	2.68851
Daily mean return	5 bp	0.00023	0.00024	0.00024
Yearly mean return	5 bp	0.06108	0.06325	0.06376
Value at the end	10 bp	2.36317	2.42067	2.44809
Daily mean return	10 bp	0.00021	0.00021	0.00022
Yearly mean return	10 bp	0.05522	0.05681	0.05755
Value at the end	15 bp	2.16268	2.19627	2.22916
Daily mean return	15 bp	0.00017	0.00019	0.00019
Yearly mean return	15 bp	0.04939	0.05040	0.05138
Value at the end	20 bp	1.97920	1.99267	2.02981
Daily mean return	20 bp	0.00016	0.00017	0.00017
Yearly mean return	20 bp	0.04359	0.04403	0.04524

Table 8.5: Summary Statistics with Incorporated Transaction Cost

8.3 Performance Measurement

In the previous section we have looked so far at the average return, the corresponding standard deviation, skewness and kurtosis. However, when talking about performance, we should analyse some additional aspects.

Even though we are clearly interested in high returns, it is not the only decisive criteria. There is usually a tradeoff between the rate of an asset's return and its risk. An investment is usually considered risky if we can loose a substantial amount of money. From this point of view, it is the downward potential of an investment that matters. So, the left tail of the return distribution is certainly important when we talk about risk. There are some well known risk management measures like the *Value at Risk* (VaR) or the *Expected Shortfall* (ES) that are especially designed to capture this kind of risk. When looking at the finance literature, we find, however, quite often that one just employs a statistical dispersion measure, like the standard deviation for example, as a risk measure. There are good reasons for that. First of all, it is a measure that we can estimate quite easily from our data. Clearly, when looking again at the finance literature, we see that there are many different, more or less demanding ways to estimate such a value. However, often it is already useful when just the empirical standard deviation of the returns is employed. Obtaining this number is clearly a lot easier than getting an estimate of the VaR or the ES, where we first have to determine the corresponding distribution tail in a decent way. Taking the standard deviation of the returns[2] as a measurement of risk is, though, not only justified by its simplicity but also by its ability to serve as a measurement to capture the risk of obtaining unpleasant investment results. If the returns fluctuate strongly, we can loose quite a substantial amount of money in a very short time. So, we can say that the higher the standard deviation of the returns, the riskier is the investment. With this risk perspective, we can apply a very intuitive performance measure, this is the so-called *Sharpe Ratio*, as defined by

$$SR_i = \frac{\overline{r}_i - \overline{r}_f}{\sigma_i} \tag{8.3.1}$$

with $\overline{r}_i$ denoting the average return of asset i and $\overline{r}_f$ standing for the average risk-free return. The standard deviation of the returns of asset i is given by σ_i. The Sharpe ratio can, thus, be interpreted as measuring the risk premium of an investment per unit of risk.

In Table 8.6 we report in the first line an estimate of the Sharpe ratio for all three fund types as well as for the S&P 500 index. The reported figures of the pairs trading funds are quite impressive. This comes at no surprise. We have already seen that the fund returns show rather limited fluctuations as compared to their average rate of return.

Another often reported performance figure in empirical back-testing studies with respect to asset management strategies, is the so-called α. It is related to the famous *Capital Asset Pricing Model* (CAPM) as proposed by Sharpe (1964), Lintner (1965) and Mossin (1966). As it is usually taught in finance lectures, investors may only expect to earn a return over the risk-free rate if they are willing to take on higher risks. As already mentioned, risk can be defined in many ways. If we base our risk definition on the return fluctuations of an investment, as discussed above, we may follow the popular portfolio theory approach according to Markowitz (1953). This theory states that there is only a reward for bearing so-called *systematic risk*, i.e. risk that cannot be diversified away. The so-called *unsystematic risk* can be avoided

[2]This is often termed "volatility" in this context.

by holding a large, well diversified portfolio. It actually means that there are investments with different fluctuation levels but the same average rate of return. So, for a given average rate of return, a rational investor would always choose the investment with the lowest risk, which is in this case the investment with the smallest standard deviation of the returns. On this basis, Tobin (1958) defined a *market portfolio*, which is the result of a mean-variance optimisation under the consideration of all possibly available assets. If there is such a market portfolio, which contains only systematic risk, a rational investor should, depending on his risk appetite, only decide about the capital distribution between the market portfolio and a riskless asset that yields the corresponding risk-free rate.

If we stay in this framework and are able to determine such a market portfolio, we could estimate for each single asset its systematic risk component. In order to do that in a practical exercise, one often refers to the CAPM. From a theoretical point of view, the CAPM reflects the relationship between the risk (in terms of the standard deviation of the asset returns) and reward (in terms of the expected returns). In particular, the CAPM should provide us with a benchmark return that enables us to evaluate any investment project. The proposed relationship regarding the return of an asset A is thereby given as follows:

$$E\left[r_A\right] - r_f = \beta_A \left(E\left[r_M\right] - r_f\right) \tag{8.3.2}$$

with $\beta_A = \frac{\mathrm{Cov}(r_A, r_M)}{\mathrm{Var}(r_M)}$, and $E\left[r_M\right]$ denoting the expected return of the market portfolio.

For any given β_i, (8.3.2) gives us a relationship between the expected excess return of the market portfolio and the risk premium of asset i. β_i captures thereby the systematic risk part of asset i. To be more precise, we have $\beta_i = \frac{\rho_{i,M}\sigma_i}{\sigma_M}$, where σ denotes the standard deviation of the corresponding returns and $\rho_{i,M}$ stands for the linear correlation[3] between the market returns and the returns of asset i. So, according to the CAPM, the risk premium is higher for more volatile assets that are at the same time stronger correlated with the market portfolio. The α we have mentioned earlier, becomes relevant if the observed excess returns deviate systematically from the ones predicted by the CAPM. Such an α is then interpreted as the part of the risk premium

[3]This is the linear correlation according to Bravais-Pearson.

that does not depend on the market portfolio. Finance people often speak in this context about *absolute* returns as the return part expressed by α should persist independently of the dynamics of the market portfolio. This is related to the CAPM expression as given by (8.3.2) and the way we usually try to estimate the β parameter, where the α is then just a constant added to (8.3.2) in order to explain the difference in expectations.

Finally, it is worth mentioning that the market portfolio as discussed so far, is only a theoretical construct. In practical applications we just have to take a proxy, i.e. a really existing financial portfolio that is as broad and well diversified as possible. In many applications the S&P 500 index is already considered as a valid choice for a market portfolio proxy. What we finally choose as a decent market portfolio proxy, depends on the feasible investment universe of a particular investor. If this investment universe is reduced to very liquid US stocks, then the S&P 500 index may qualify indeed as a market portfolio proxy. So, if we take the S&P 500 index as our relevant market portfolio, we have already all the values we need in order to calculate the corresponding βs. Figure 8.21 depicts the bivariate dependence structure between the returns of our three pairs trading funds and the S&P 500 index returns. In Table 8.6 we report the estimated linear correlations $\widehat{\rho}$ between the returns of our pairs trading funds and the S&P 500 index. There we also report the estimated β parameters. As expected, the $\widehat{\beta}$ values of the pairs trading strategies are quite small. Nevertheless, we have seen in the previous section that the pairs trading strategies yielded significant excess returns. Hence, the three strategies must show some positive α values. In fact, by adding a constant parameter α to (8.3.2), we can estimate the corresponding values. These results are also reported in Table 8.6. As expected, the estimated β-coefficients are pretty small for all pairs trading funds, though still significantly different from zero. With $\widehat{\alpha}^y$ we report the estimated yearly excess return part that is linearly independent of our chosen market portfolio. The corresponding t-values $t_{\widehat{\alpha}^y}$ indicate that the estimated α values are highly significant.

Hence, if we consider the discussed α in the CAPM context and the Sharpe ratio as relevant performance measures, we can say that our three pairs trading strategies performed indeed well over the analysed 16 years time interval.

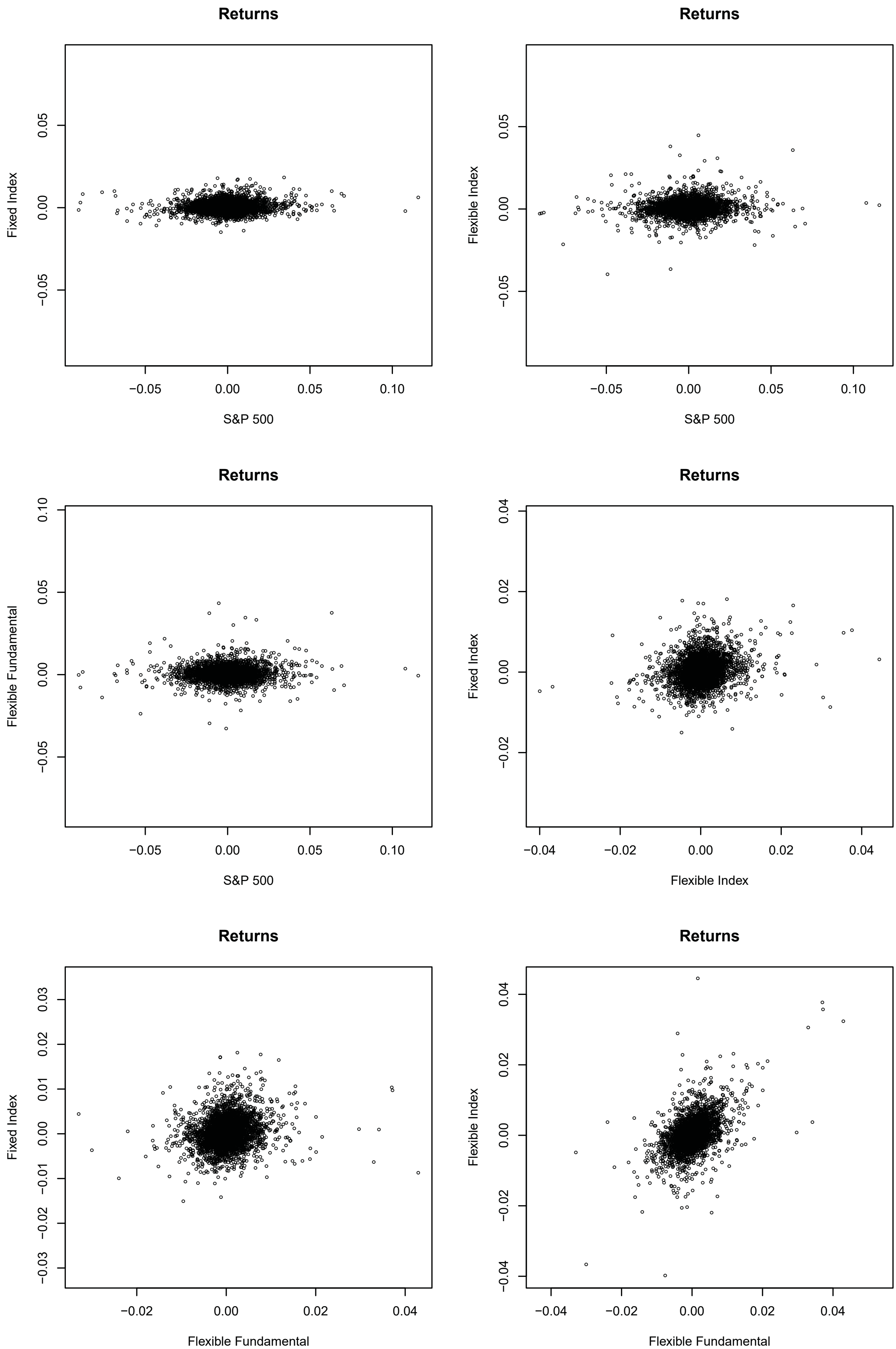

Figure 8.21: Returns Dependence Plots.

	Fixed Index	Flexible Index	Flexible Fundamental	S&P 500
$\widehat{SR}$	0.6388	0.5195	0.5478	0.1337
$\widehat{\alpha}^y$	0.0309	0.0334	0.0338	0.0000
$t_{\widehat{\alpha}^y}$	2.5165	2.0393	2.1639	
$\widehat{\beta}$	0.0177	0.0243	0.0142	1.0000
$t_{\widehat{\beta}}$	4.7585	4.7461	2.9298	
$\widehat{\rho}$	0.0739	0.0736	0.0456	1.0000

Table 8.6: Performance Figures Summary.

8.4 Risk Measures

As already mentioned in the previous section, a comprehensive analysis should also consider some appropriate risk measures. So far we have just used the standard deviation of the returns to measure our investment risk. Often we are, however, more concerned about the negative deviations then the positive ones. This is especially important when we need to know how much capital we need to put aside in order to avoid bankruptcy in a bad moment. We are, thus, especially interested in the left tail of our return distribution or, to put it differently, in the potential losses.

As a loss is per definition a negative monetary value, we deal with positive loss values when we analyse them. So, we transform our negative returns into positive relative loss numbers. By this, we have to focus on the right tail of the loss distribution, which is also handier from an analytical point of view. In addition, it is usually sufficient to focus only on the excess distribution over a particular threshold u. This is:

$$F_u(x) = P(X - u \leq x \mid X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \tag{8.4.1}$$

for $0 \leq x < x_F - u$, with x_F defining the finite or infinite right endpoint of distribution F.

Without going into too much details of extreme value theory, we can make use of the theorem proposed by Pickands (1975), Balkema and De Haan (1974), which basically says that for a large class of underlying distribution functions F, the conditional excess distribution F_u is well approximated by a *Generalised Pareto Distribution* (GPD). More precisely, they say that

$$\lim_{u \to x_F} \sup_{0 \le x < x_F - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0 \tag{8.4.2}$$

if and only if F lies in the *maximum domain of attraction* of the *Generalised Extreme Value* distribution (GEV).[4]

The Generalised Pareto distribution is given by

$$G_{\xi,\beta} = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-\frac{1}{\xi}} & \text{, for } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{, for } \xi = 0 \end{cases} \tag{8.4.3}$$

where $\beta > 0$, $x \ge 0$ when $\xi \ge 0$, and $0 \le x \le -\frac{\beta}{\xi}$ when $\xi < 0$. ξ is the shape parameter and β the scale parameter. A random variable X that is distributed according to such a GPD has only a finite expected value for $\xi < 1$. In this case we have

$$E[X] = \frac{\beta}{1 - \xi} \ . \tag{8.4.4}$$

If we want to use a GPD to describe the right tail of our loss distribution, we need to determine a decent threshold above which we fit the GPD to our data. A possible way to get an idea of where to set the threshold, is by using the mean excess function

$$e(u) = E[X - u \mid X > u] \ . \tag{8.4.5}$$

Provided that $\xi < 1$, the mean excess function of the GPD is given by

$$e(\nu) = \frac{\beta + \xi(\nu - u)}{1 - \xi} \tag{8.4.6}$$

with $u \le \nu < \infty$ if $0 \le \xi < 1$ and $u \le \nu \le u - \frac{\beta}{\xi}$ if $\xi < 0$. Expression (8.4.6) is obviously a linear function in ν. This fact can now be used as a diagnostic means. In particular, we plot the sample mean excess function, which is given by

$$e_n(\nu) = \frac{\sum_{i=1}^{n} (X_i - \nu) \mathbb{I}_{\{X_i > \nu\}}}{\sum_{i=1}^{n} \mathbb{I}_{\{X_i > \nu\}}} \tag{8.4.7}$$

[4] More details about extreme value theory can be found in (McNeil, Frey, and Embrechts, 2005), Chapter 7.

with $\mathbb{I}_{\{X_i>\nu\}}$ denoting the characteristic function.

The sample mean excess function plots of the relative losses of our three funds and the S&P 500 are depicted in Figure 8.22. With respect to the sample

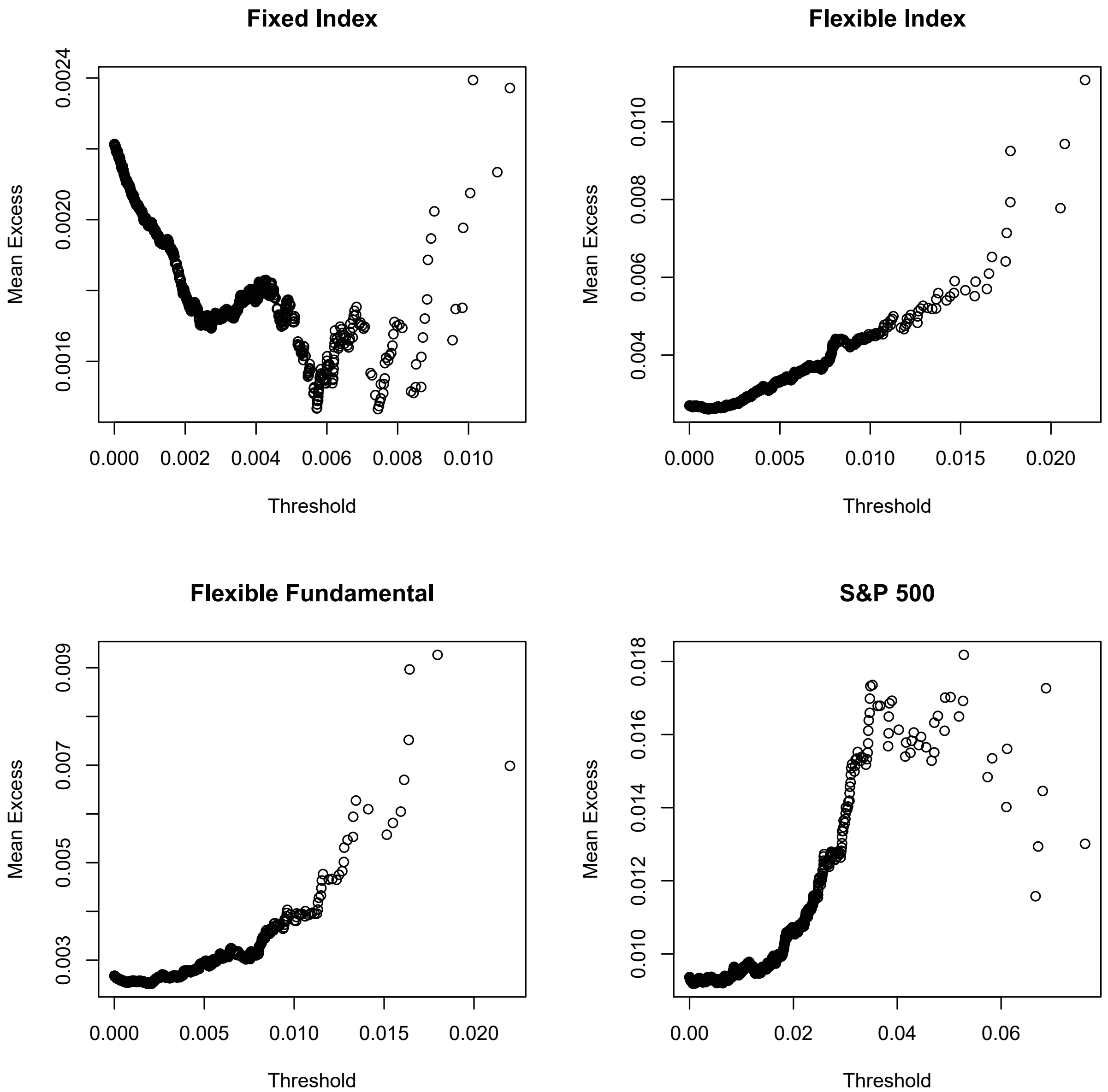

Figure 8.22: Sample Mean Excess Plots.

mean excess function it is clear that points for higher thresholds become less reliable and fluctuate more randomly as the number of observations gets smaller and smaller. The idea of such a mean excess plot is, thus, to find a threshold above which we are willing to assume a linear behaviour of the points. If the points show an upward trend, this indicates a GPD with a positive shape parameter ξ. A negative trend would suggest a GPD with a negative shape parameter. Similarly, a horizontal line would suggest a shape parameter of zero and, therefore, an exponential distribution.

Examining our four plots, we see that the previously described task of finding a suitable analysis threshold is not always so trivial. The easiest one is probably the plot of the Flexible Index fund on the upper right. There we could quite easily justify to chose a threshold of approximately 0.005. A similar choice can probably be made for the Flexible Fundamental fund. For the Fixed Index fund we could potentially assume a horizontal line at some point slightly after 0.0025. The plot for the S&P 500 looks a bit special. There it makes, however, sense to put less weight on the points on the very right and assume an increasing line. Hence, a threshold at around 0.0125 seems to be reasonable.

	Fixed Index	Flexible Index	Flexible Fundam.	S&P 500
u	0.0025	0.0050	0.0050	0.0125
$P(X \leq u)$	0.6450	0.8604	0.8613	0.7570
N_u	667	267	260	489
$\widehat{\xi}$	0.0018	0.1826	0.1333	0.1379
$SE\left(\widehat{\xi}\right)$	0.0403	0.0711	0.0653	0.0493
$\widehat{\beta}$	0.0017	0.0027	0.0025	0.0082
$SE\left(\widehat{\beta}\right)$	0.0001	0.0003	0.0002	0.0005

Table 8.7: Fitted Generalised Pareto Distribution

Table 8.7 reports the estimated parameters with the associated standard deviations. The estimates are obtained by the corresponding maximum likelihood estimator. Obviously, the number of observations used in the parameter estimation depends on the chosen threshold. So, with a relatively low threshold, we get quite some more data points for the Fixed Index fund and the S&P 500 index.

While the estimated shape parameters of the Flexible Index fund, the Flexible Fundamental fund and the S&P 500 index are significantly positive, we could probably use the tail of an exponential distribution with respect to the Fixed Index fund. This is basically what we already suspected when we looked at the mean excess plots of Figure 8.22.

Figure 8.23 shows how the results with respect to $\widehat{\xi}$ change if we use a different

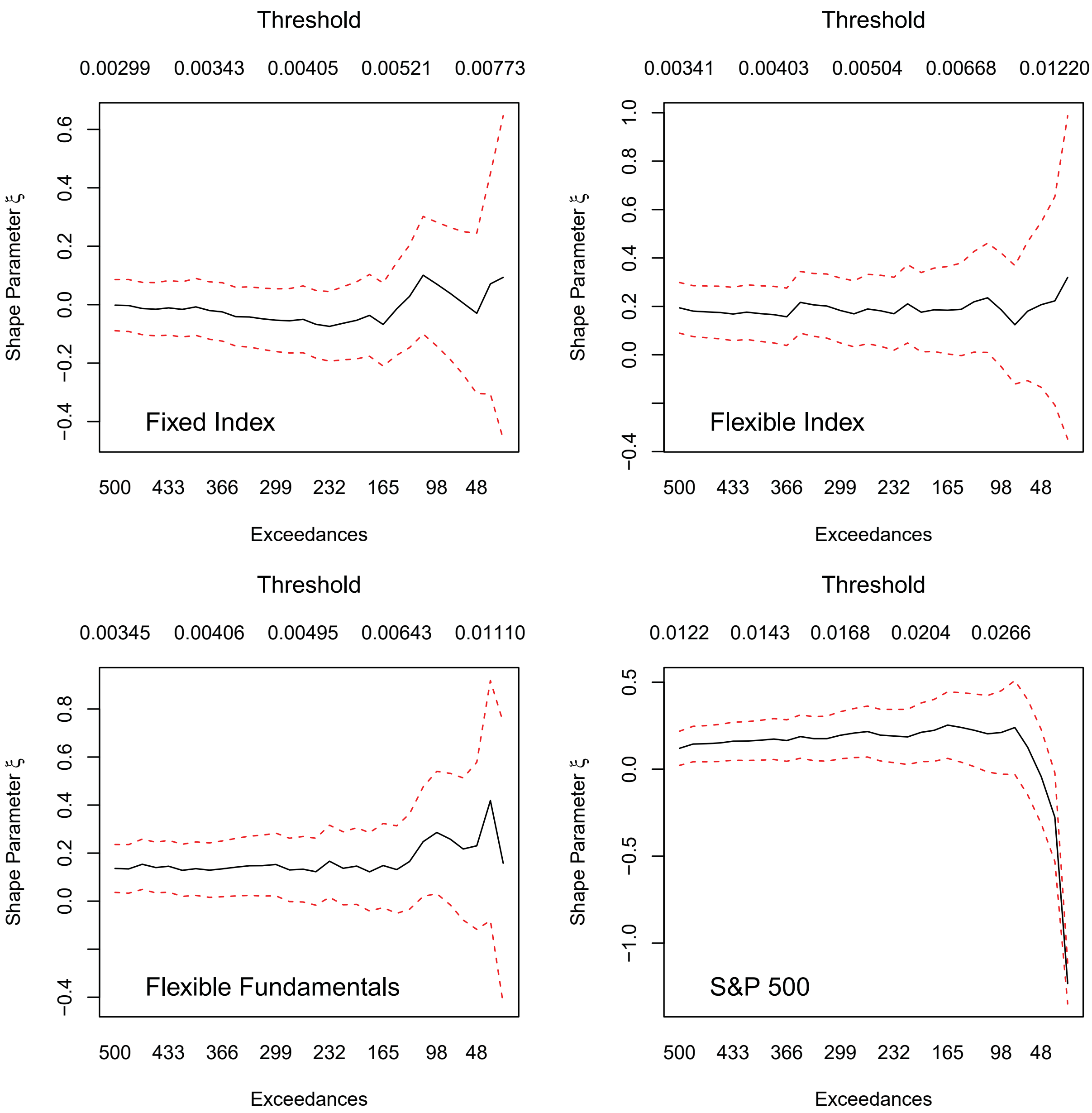

Figure 8.23: Shape Parameter ξ Robustness.

threshold u. The results look, however, quite robust.

In order to use the obtained excess distributions as right tail of our corresponding loss distributions, we can use the following relation:

$$\begin{aligned} P(X > x) &= P(X > u)\, P(X > x \mid X > u) \\ &= (1 - F(u))\, P(X - u > x - u \mid X > u) \\ &= (1 - F(u))\,(1 - F_u(x - u)) \ . \end{aligned} \tag{8.4.8}$$

For $F(x)$ we can simply use the empirical distribution or, if necessary, we may try to fit a suitable type of distribution to the data below the threshold u. If we use the empirical distribution, we can write the right tail of our loss distribution according to (8.4.8) as:

$$P(X > x) = \frac{N_u}{n}\left(1 + \widehat{\xi}\frac{x - u}{\widehat{\beta}}\right)^{-\frac{1}{\widehat{\xi}}} \tag{8.4.9}$$

where $\frac{N_u}{n}$ is the estimator of $(1 - F(u))$, with $N_u = \sum_{i=1}^{n} \mathbb{I}_{x_i > u}$.

The fitted tails are displayed in Figure 8.24 (with a black solid line). As can be seen there, the GPD model fits the observed data of all the funds pretty well.

Having a good estimate for the relevant tail, we can calculate the most common capital related risk measures, i.e. the Value at Risk (VaR) and the Expected Shortfall (ES). For a random loss variable X and a chosen level $\alpha \in (0, 1)$ the VaR is defined as:

$$VaR_\alpha(X) = \inf\{x \in \mathbb{R} \mid P(X > x) \leq 1 - \alpha\} \ . \tag{8.4.10}$$

It is basically nothing else than the α-quantile of the distribution function F_X. Having VaR_α, we can define the ES_α of a random loss variable X as:

$$ES_\alpha(X) = \frac{1}{1 - \alpha}\int_\alpha^1 VaR_v(X)\, dv \ . \tag{8.4.11}$$

In our case, where we have a continuous distribution function, we can also write the expected shortfall as:

$$ES_\alpha(X) = \frac{1}{1 - \alpha}\int_{VaR_\alpha(X)}^{\infty} x\, dF(x) = E\left[X \mid X \geq VaR_\alpha(X)\right] \tag{8.4.12}$$

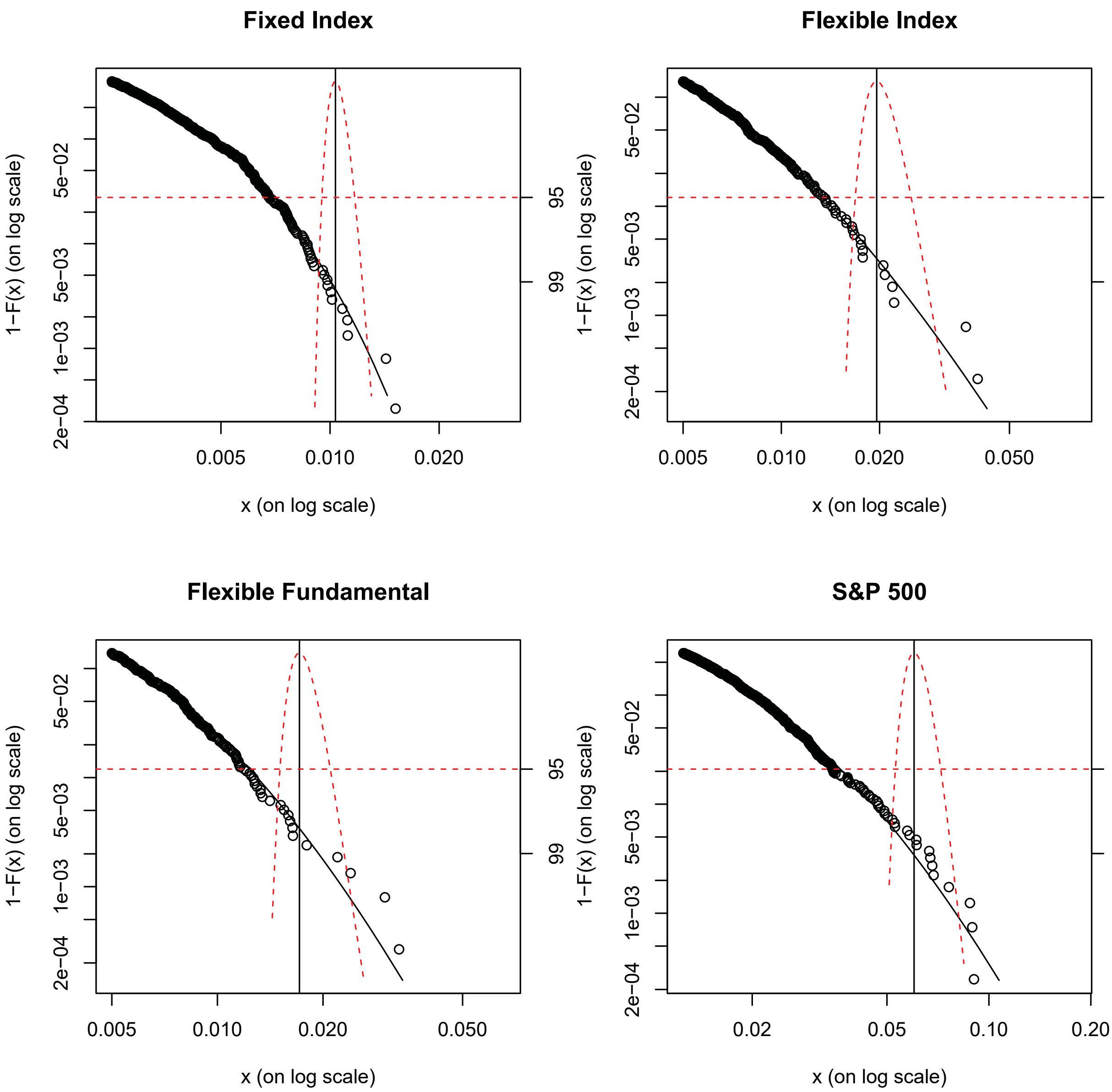

Figure 8.24: Fitted Tails and Shortfall Estimates.

which has the nice interpretation of a conditional expected loss, given that the loss exceeds the $VaR_\alpha(X)$. By using (8.4.9), we can get an estimate of the VaR by

$$\widehat{VaR}_\alpha = u + \frac{\widehat{\beta}}{\widehat{\xi}}\left(\left(\frac{n(1-\alpha)}{N_u}\right)^{-\widehat{\xi}} - 1\right) . \qquad (8.4.13)$$

In order to obtain an estimate of the ES, we can use (8.4.12) together with the mean excess function (8.4.6), and we get

$$\widehat{ES}_\alpha = \frac{\widehat{VaR}_\alpha}{1-\widehat{\xi}} + \frac{\widehat{\beta} - \widehat{\xi}u}{1-\widehat{\xi}} . \qquad (8.4.14)$$

With this, we can now get the corresponding risk measures. Table 8.8 reports the estimated risk measure values together with the associated 95% upper and lower confidence bands that are obtained by the profile likelihood curve.

	Fixed Index	Flexible Index	Flexible Fundam.	S&P 500
$\widehat{ES}_{0.99}$ Point Estimate	0.01035	0.01960	0.01713	0.06012
$\widehat{ES}_{0.99}$ Upper Bound	0.01168	0.02498	0.02093	0.07204
$\widehat{ES}_{0.99}$ Lower Bound	0.00951	0.01693	0.01511	0.05297
$\widehat{VaR}_{0.99}$ Point Estimate	0.00863	0.01421	0.01298	0.04536
$\widehat{VaR}_{0.99}$ Upper Bound	0.00934	0.01602	0.01444	0.05042
$\widehat{VaR}_{0.99}$ Lower Bound	0.00811	0.01294	0.01193	0.04173
$\widehat{ES}_{0.95}$ Point Estimate	0.00758	0.01209	0.01112	0.03881
$\widehat{ES}_{0.95}$ Upper Bound	0.00813	0.01362	0.01230	0.04285
$\widehat{ES}_{0.95}$ Lower Bound	0.00717	0.01108	0.01031	0.03596
$\widehat{VaR}_{0.95}$ Point Estimate	0.00586	0.00807	0.00777	0.02698
$\widehat{VaR}_{0.95}$ Upper Bound	0.00613	0.00855	0.00819	0.02832
$\widehat{VaR}_{0.95}$ Lower Bound	0.00565	0.00766	0.00740	0.02564

Table 8.8: Risk Measures Summary

The corresponding $\widehat{ES}_{0.99}$ values together with the associated profile likelihood curves are also plotted in Figure 8.24 (by the red dashed lines).

The values in Table 8.8 confirm the picture we have already seen by looking at the volatilities and the value development plots of the different funds.

For both risk measures and both chosen security levels α, the values of all the pairs trading funds are significantly smaller than the ones of the S&P 500 index. Concerning only the pairs trading funds, the Fixed Index fund shows the smallest values, while the Flexible Index fund and the Flexible Fundamental fund show results that are quite closely together.

Chapter 9

Conclusion

We have briefly discussed the pioneering work of Gatev et al. (2006) in the context of pairs trading. Pairs trading belongs to the broader category of relative value arbitrage investments. We 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.

9.1 Summary

After a short overview of the currently most popular methods in the academic literature in the field of pairs trading, we focused on an algorithmic cointegration based pairs trading approach. Even though cointegration is not a necessary precondition for two assets making up a tradable pair, we used an approach that takes the decision with respect to the tradability of a particular pair according to a cointegration test. It was therefore necessary to obtain a good understanding of the tests we considered. We decided to go with four well established residual based test procedures and analysed some important properties of them by an illustrative simulation study. There we saw that the differences between the ADF, PP and OU tests are not really crucial. The KPSS test, however, showed a considerably weaker performance. While heteroskedasticity seemed to be no real issue for neither of the four tests, a high or time-varying persistence level in the spread series can be a serious problem for them.

By illustrating that an unconditional testing approach is problematic in terms of a multiple statistical testing, we introduced three pre-partitioning methods in order to mitigate the problem. Starting with a very simple partitioning based on industry definitions, we went to more flexible methods with the Flexible Index and the Flexible Fundamental partitioning. While the Flexible Index method can basically be seen as a more flexible extension of the simple industry classification, the Flexible Fundamental approach basically derives from the idea of the common trend model.

The subsequent back-testing of the models on an asset universe consisting of the S&P 500 stocks, showed really satisfying 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 in terms of a considerably lower volatility, VaR and ES. In addition, the back-tested strategies showed only a very minor correlation with the S&P 500 index, which is clearly a very desirable feature with respect to portfolio diversification.

9.2 Outlook

During the analysis we have come across several aspects that might be worth being considered again in more detail in a subsequent study.

Obviously, a possible extension of the discussed models could be the expansion of the *pairs-term* of only 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. It is clear that in such a case a proper pre-partitioning would become even more crucial as the number of possible combinations in an unrestricted setting would just explode.

We have also noted that the trading rule based on the two standard deviations might not be optimal. One should probably spend some more effort in researching a simple but more effective trading rule.

Finally, the proposed pre-partitioning methods, especially with respect to the Flexible Fundamental approach, may still be improved. So, one could still look for ways to improve the grouping. It is certainly an open question whether the proposed factors in the Flexible Fundamental model are appropriate enough or whether they should be extended or substituted by other factors.

Bibliography

Andrade, S., V. Di Pietro, and M. Seasholes (2005). Understanding the profitability of pairs trading. *Unpublished working paper, UC Berkeley, Northwestern University*.

Asness, C., A. Frazzini, and L. H. Pedersen (2014). Quality minus junk. *Available at SSRN 2312432*.

Balkema, A. and L. De Haan (1974). Residual life time at great age. *The Annals of Probability*, 792–804.

Bogomolov, T. (2010). Pairs trading in the land down under. In *Finance and Corporate Governance Conference, Bundoora, Australia*.

Bowen, D., M. C. Hutchinson, and N. O'Sullivan (2010). High frequency equity pairs trading: transaction costs, speed of execution and patterns in returns. *Journal of Trading 5*(3), 31–38.

Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *American Statistician*, 54–54.

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.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*, 427–431.

Do, B. and R. Faff (2008). Does naïve pairs trading still work. Technical report, Working Paper.

Do, B., R. Faff, and K. Hamza (2006). A new approach to modeling and estimation for pairs trading. In *Proceedings of 2006 Financial Management Association European Conference*.

Elliott, R., J. Van Der Hoek, and W. Malcolm (2005). Pairs trading. *Quantitative Finance 5*(3), 271–276.

Engelberg, J., P. Gao, and R. Jagannathan (2009). An anatomy of pairs trading: the role of idiosyncratic news, common information and liquidity. In *Third Singapore International Conference on Finance*.

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.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*(1), 3–56.

Frazzini, A. Data library. `http://www.econ.yale.edu/~af227/data_library.htm`. Accessed: 2016-01-30.

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.

Granger, C. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics 16*(1), 121–130.

Hamilton, J. D. (1994). *Time series analysis*. Princeton: University Press.

Harlacher, M. (2012). Cointegration based statistical arbitrage. *Master's thesis, ETH Zürich*.

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.

Huck, N. (2010). Pairs trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research 207*(3), 1702–1716.

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

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

Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics 54*(1), 159–178.

L'Ecuyer, P. (1999). Good parameters and implementations for combined multiple recursive random number generators. *Operations Research 47*(1), 159–164.

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.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics 11*, 601–618.

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.

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.

Nath, P. (2003). High frequency pairs trading with us treasury securities: Risks and rewards for hedge funds. *Available at SSRN 565441*.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Ng, S. and P. Perron (1995). Unit root tests in arma models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association 90*(429), 268–281.

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

Papadakis, G. and P. Wysocki (2007). Pairs trading and accounting information. Technical report, Citeseer.

Perlin, M. S. (2009). Evaluation of pairs-trading strategy at the brazilian financial market. *Journal of Derivatives & Hedge Funds 15*(2), 122–136.

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

Phillips, P. C. B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika 75*, 335–346.

Pickands, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 119–131.

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

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics 20*, 53–65.

Said, S. E. and D. A. Dickey (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika 71*(3), 599–607.

Schwert, G. W. (1989). Tests for unit roots: A Monte Carlo investigation. *Journal of Business & Economic Statistics 7*(2), 147–159.

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

Shin, Y. (1994). A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric theory 10*(01), 91–115.

Stock, J. H. and M. W. Watson (1988). Testing for common trends. *Journal of the American statistical Association 83*(404), 1097–1107.

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 Silhoutte Widths

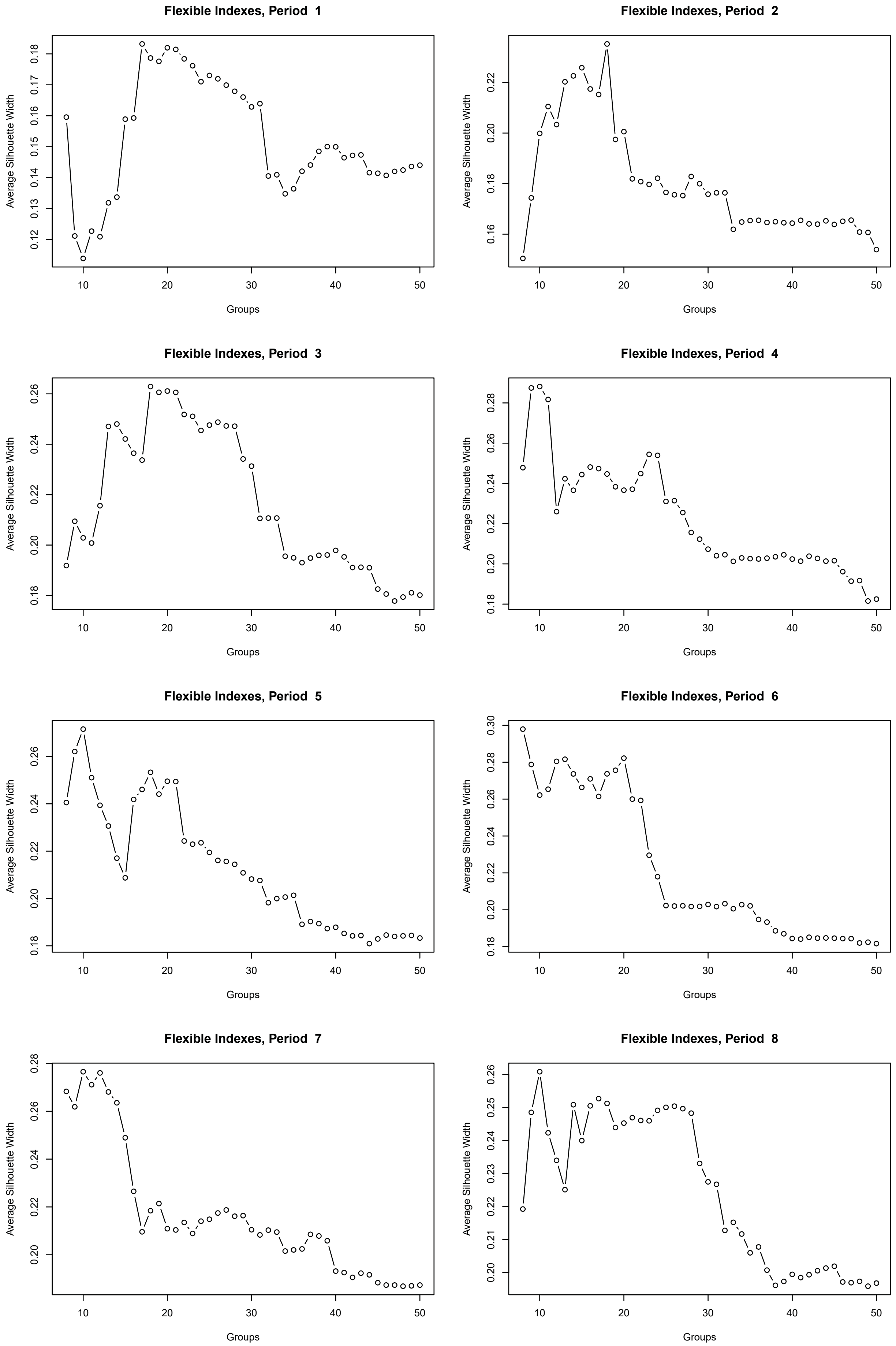

Figure A.1: Flexible Indexes, Periods 1 to 8.

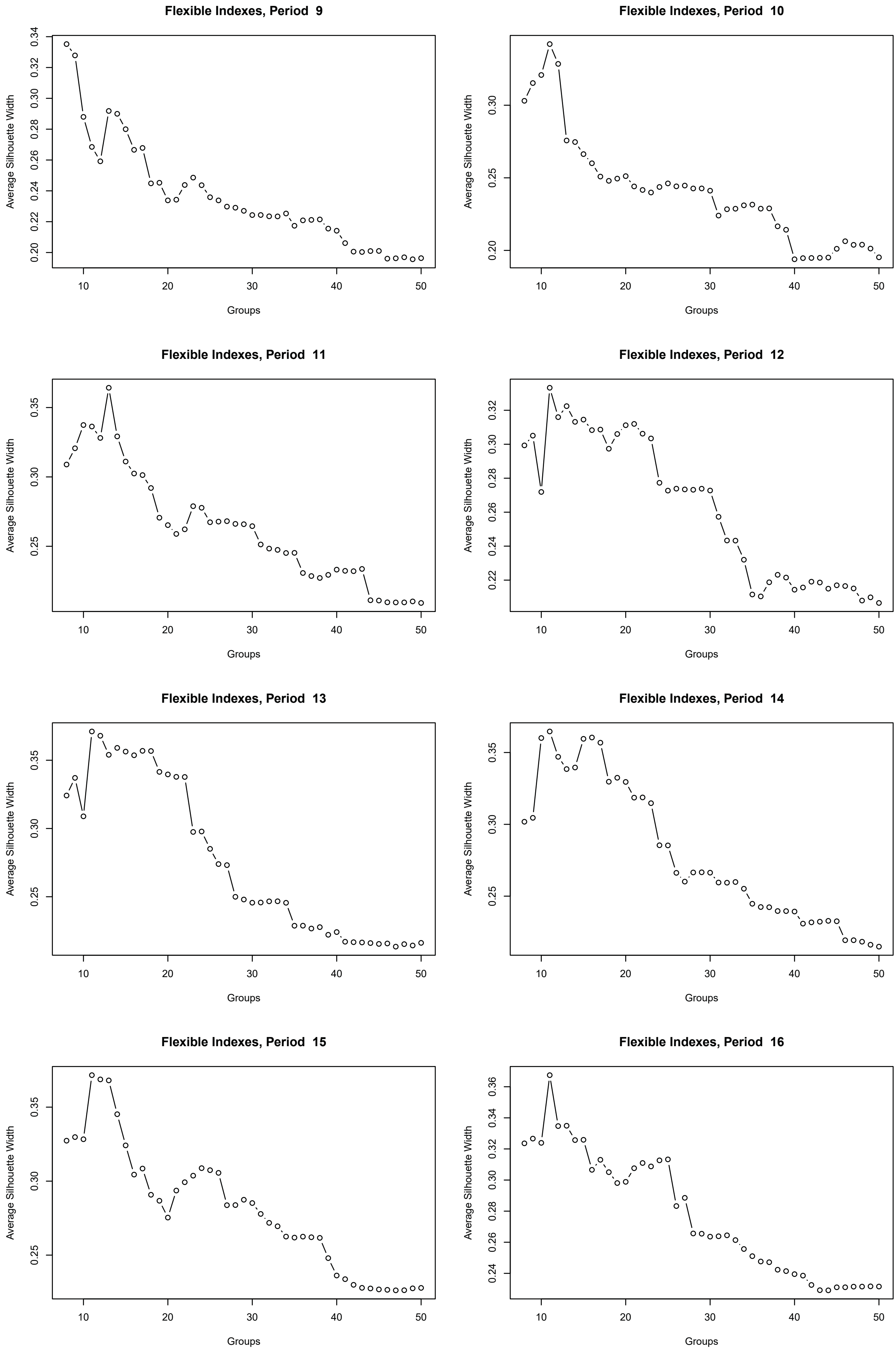

Figure A.2: Flexible Indexes, Periods 9 to 16.

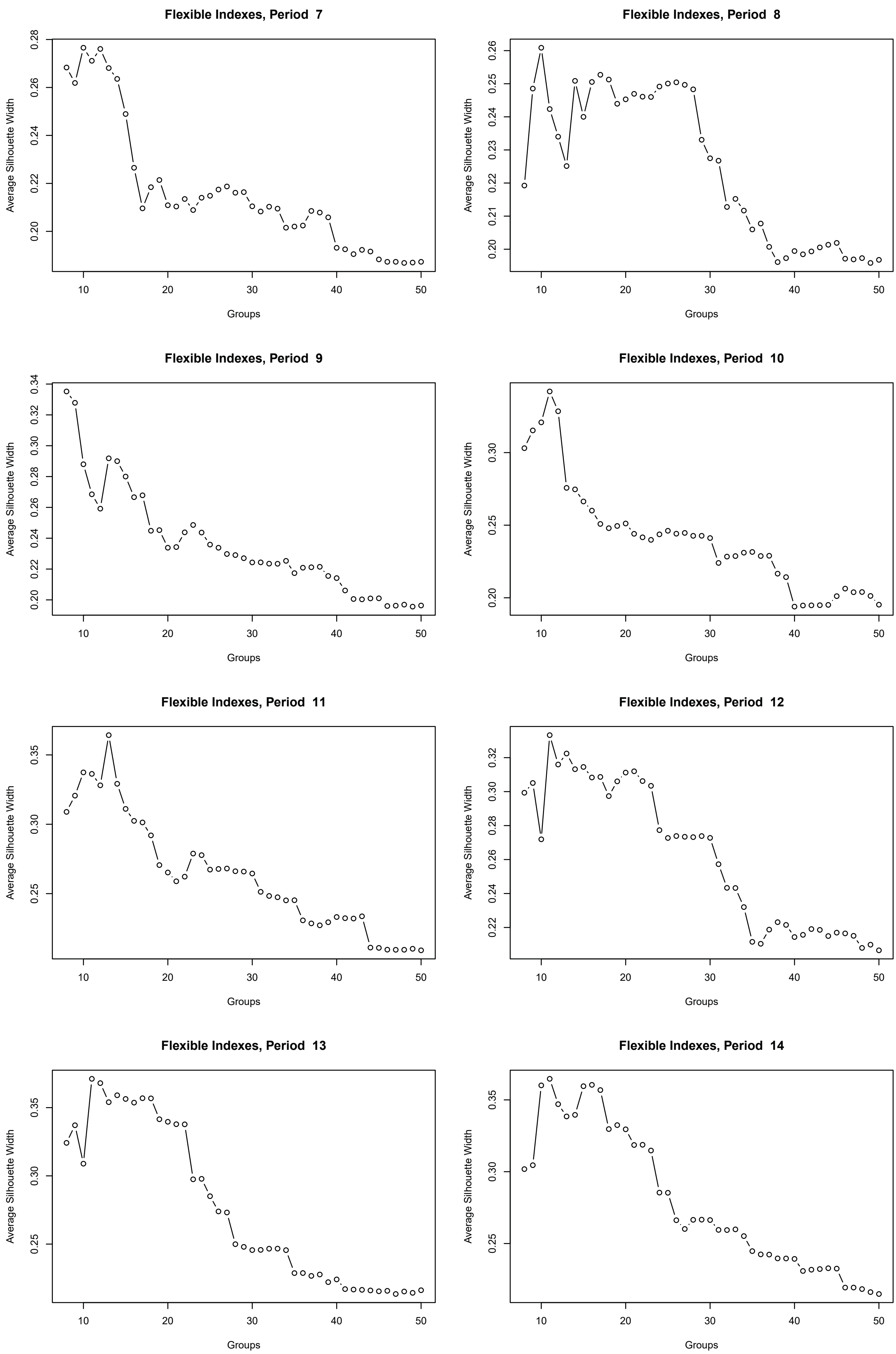

Figure A.3: Flexible Indexes, Periods 17 to 24.

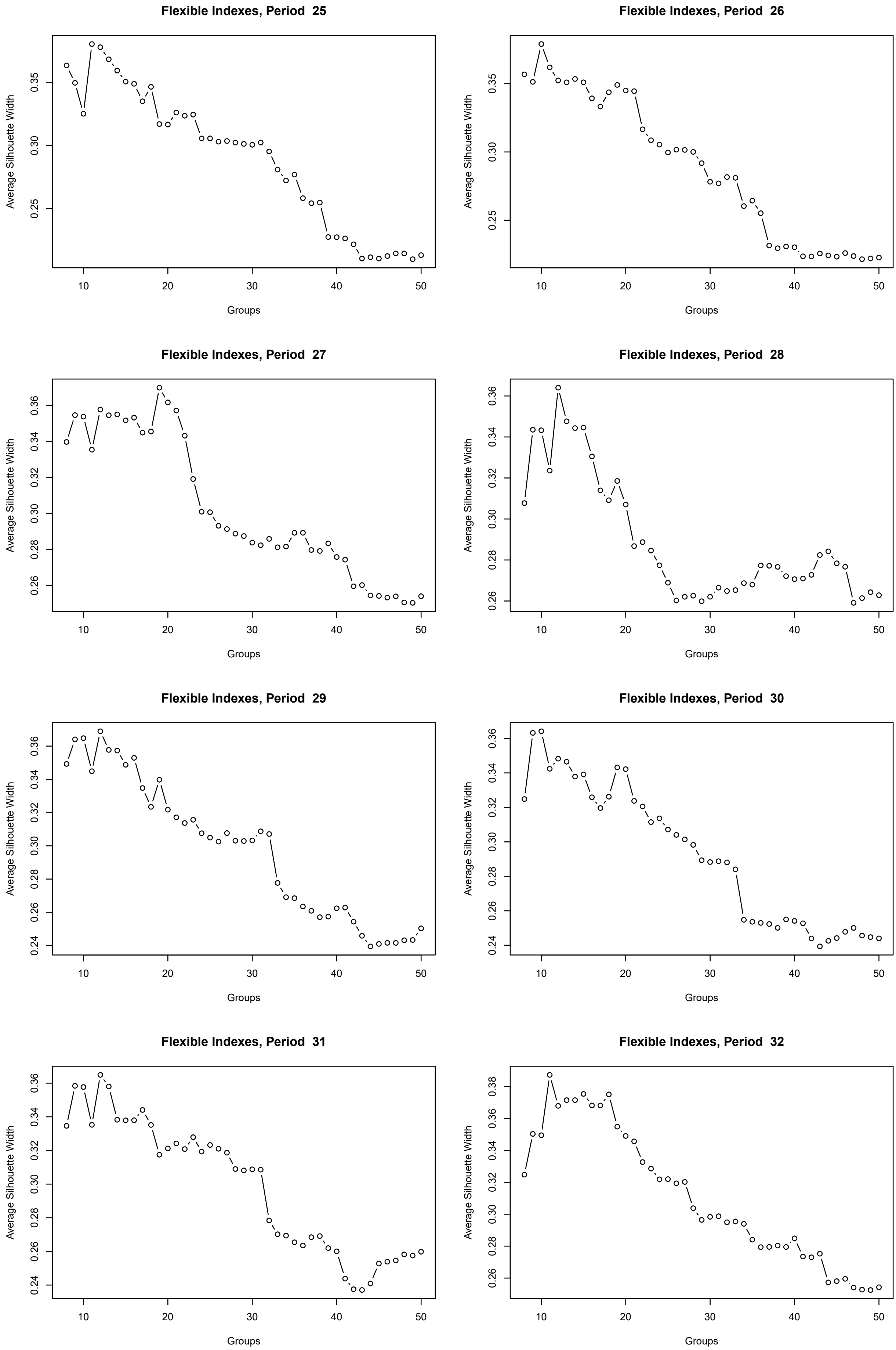

Figure A.4: Flexible Indexes, Periods 25 to 32.

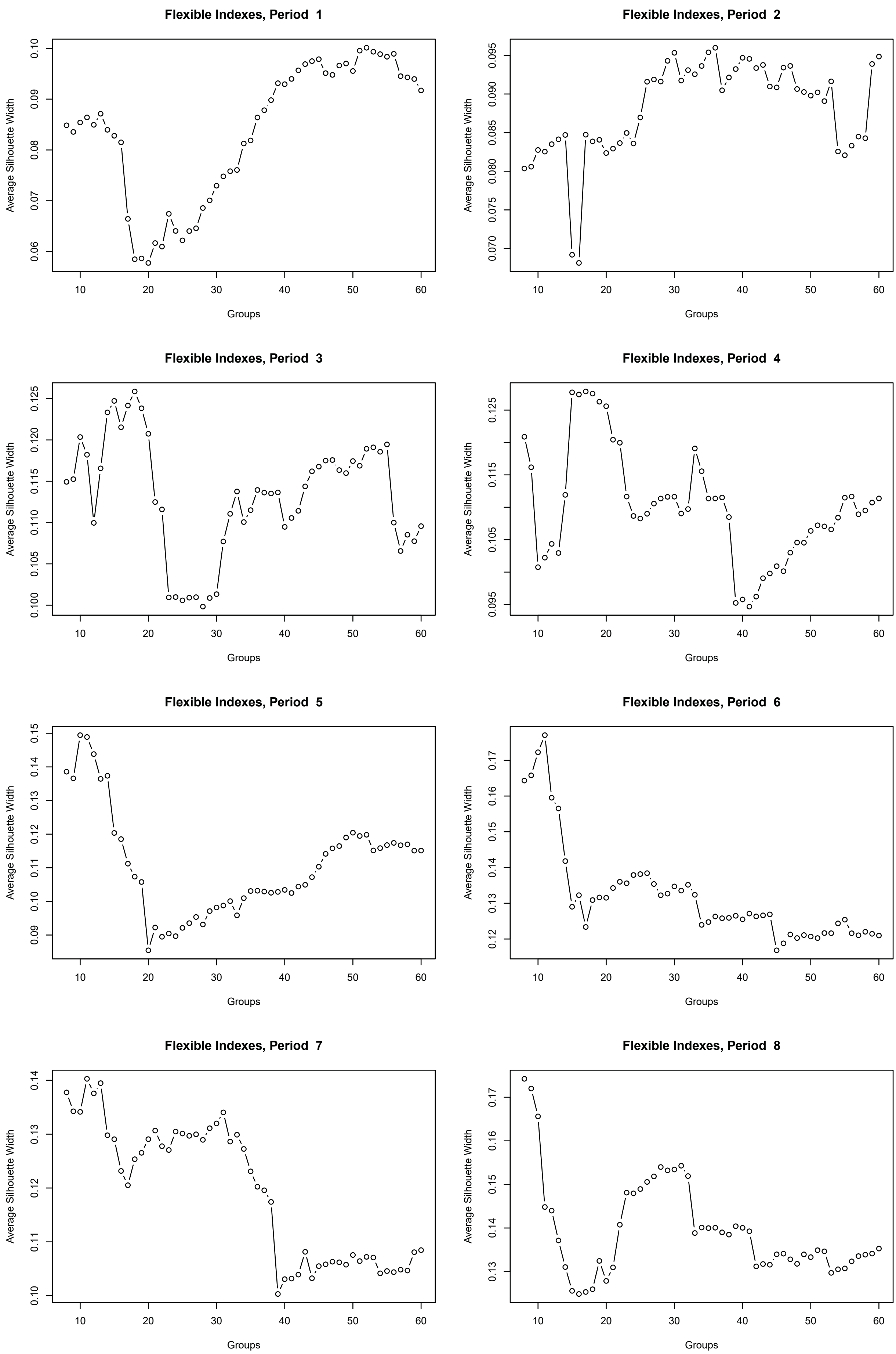

Figure A.5: Flexible Fundamentals, Periods 1 to 8.

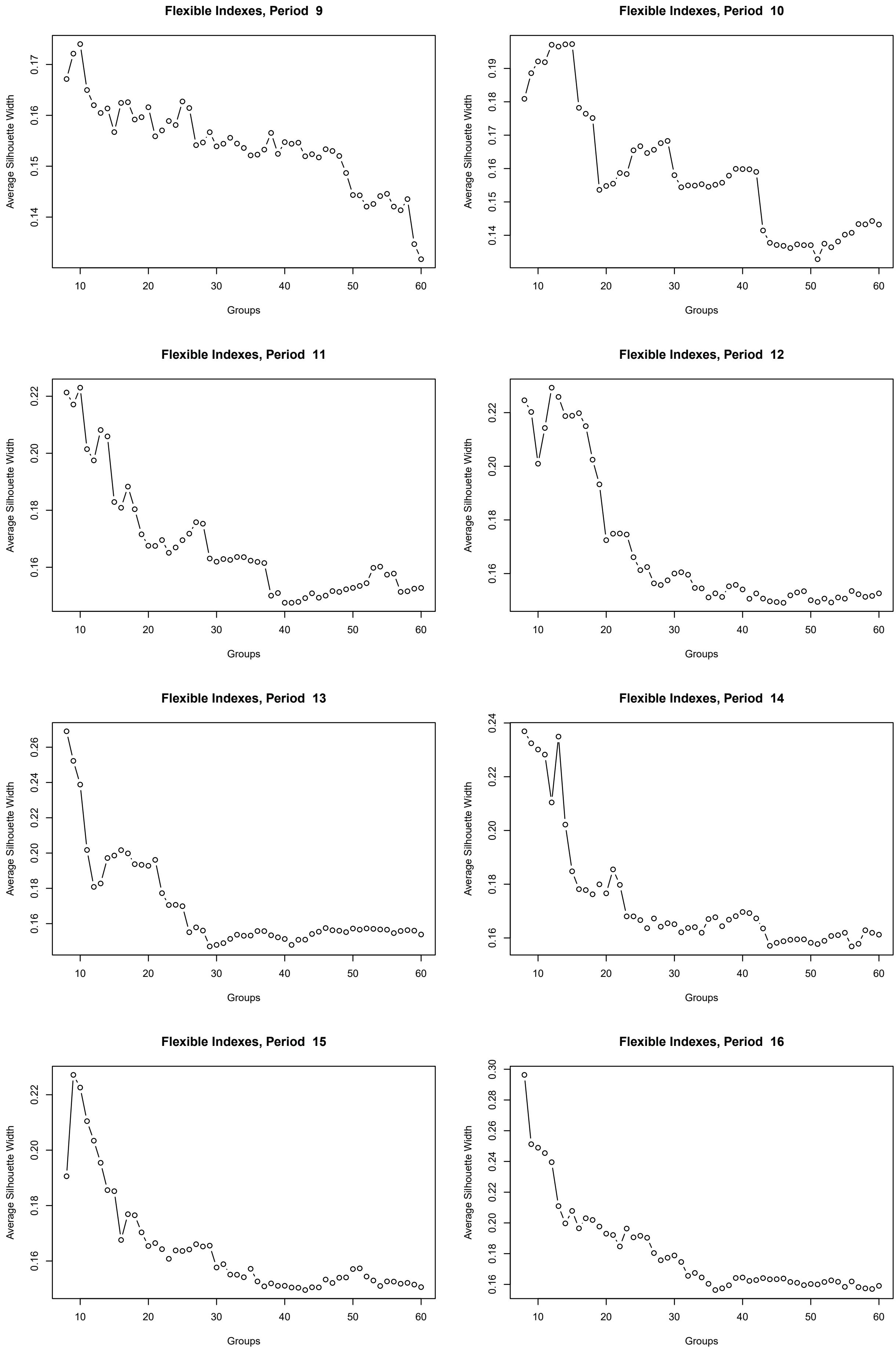

Figure A.6: Flexible Fundamentals, Periods 9 to 16.

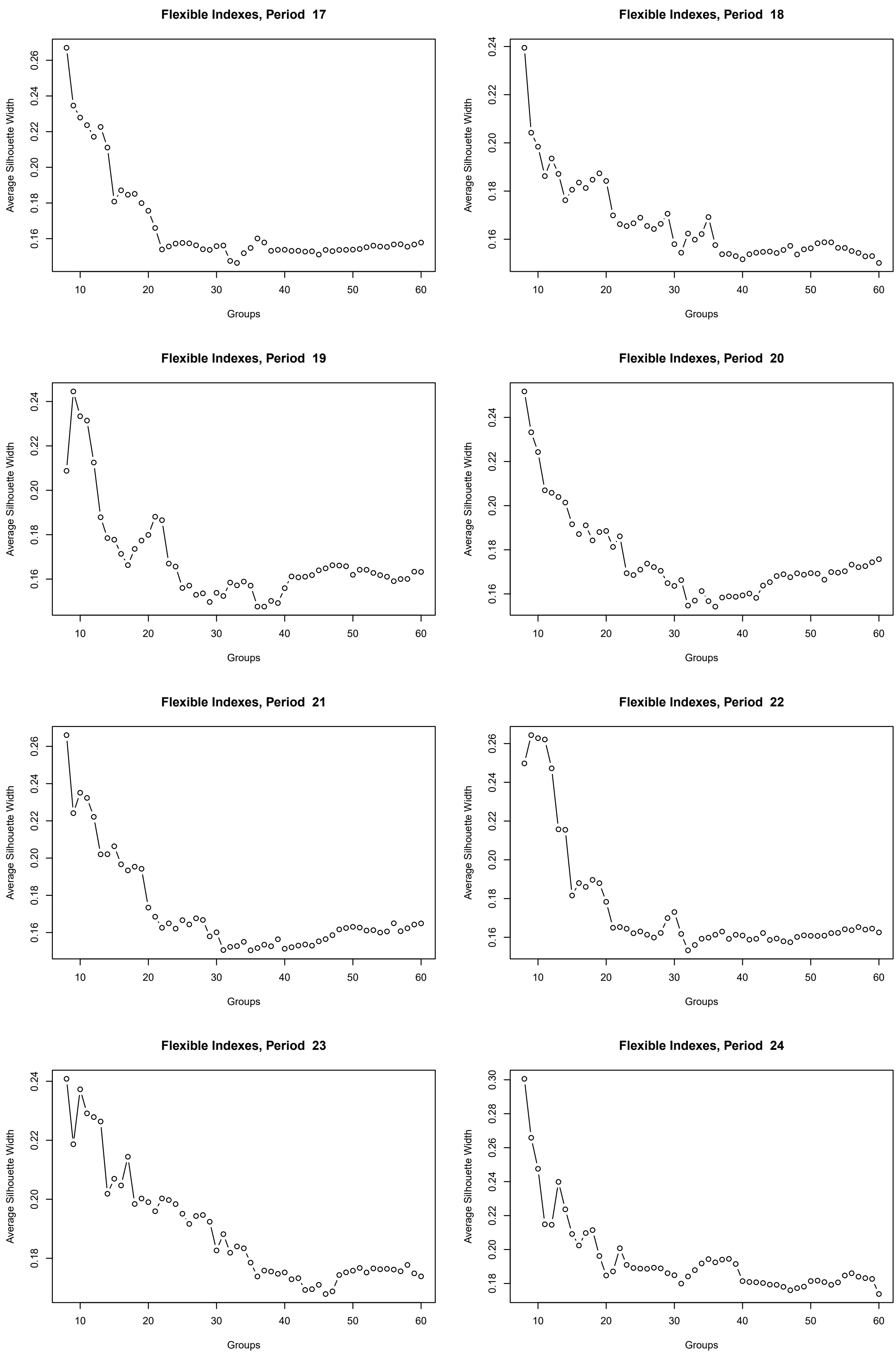

Figure A.7: Flexible Fundamentals, Periods 17 to 24.

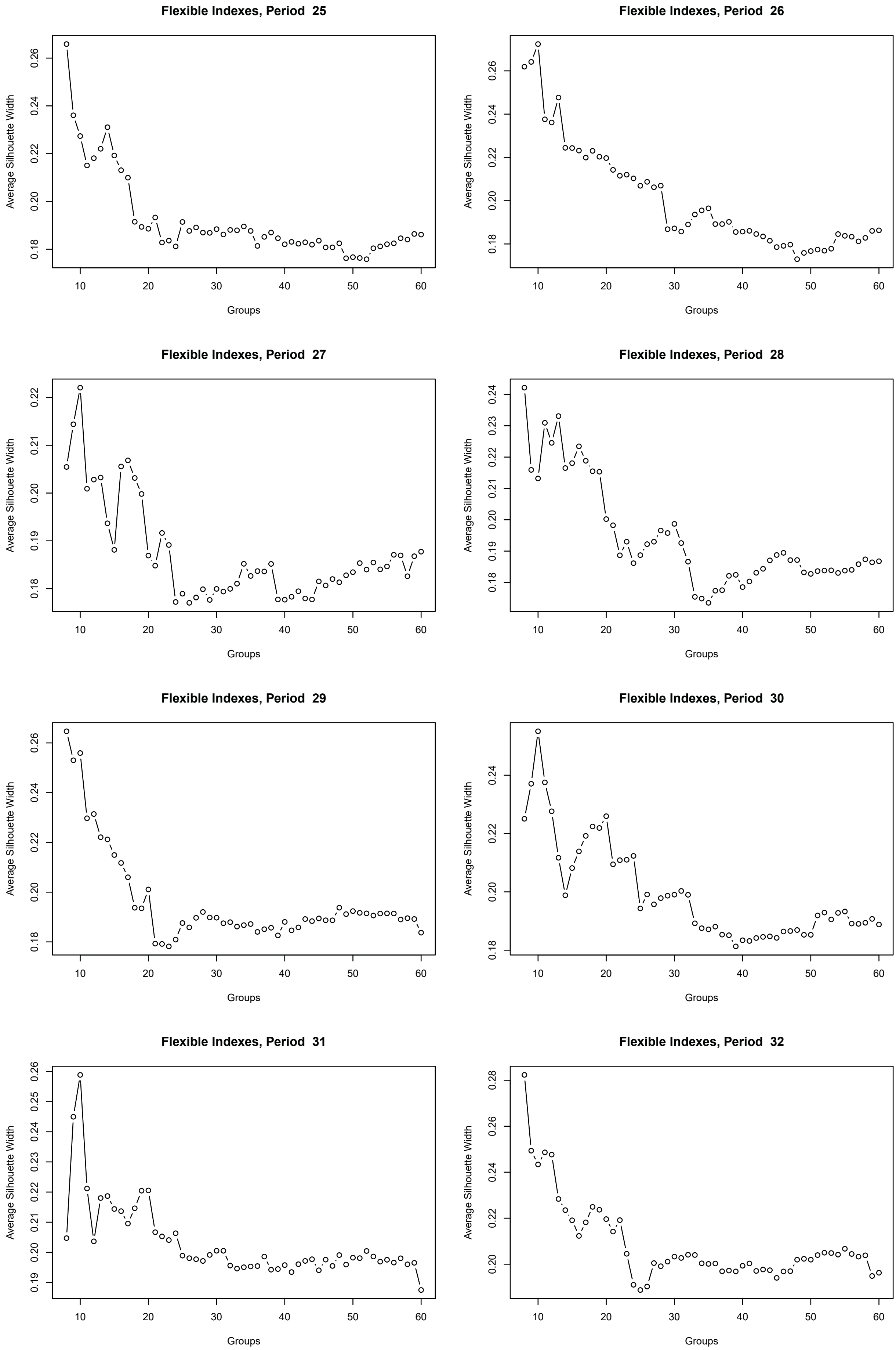

Figure A.8: Flexible Fundamentals, Periods 25 to 32.

Appendix B

Trading Periods

	Start Date	End Date
Period 1	1996-01-03	1996-07-03
Period 2	1996-07-04	1997-01-02
Period 3	1997-01-03	1997-07-03
Period 4	1997-07-04	1998-01-02
Period 5	1998-01-05	1998-07-03
Period 6	1998-07-06	1999-01-01
Period 7	1999-01-04	1999-07-02
Period 8	1999-07-05	2000-01-03
Period 9	2000-01-04	2000-07-03
Period 10	2000-07-04	2001-01-01
Period 11	2001-01-02	2001-07-03
Period 12	2001-07-04	2002-01-01
Period 13	2002-01-02	2002-07-03
Period 14	2002-07-04	2003-01-01
Period 15	2003-01-02	2003-07-03
Period 16	2003-07-04	2004-01-01
Period 17	2004-01-02	2004-07-02
Period 18	2004-07-05	2005-01-03
Period 19	2005-01-04	2005-07-04
Period 20	2005-07-05	2006-01-02
Period 21	2006-01-03	2006-07-03
Period 22	2006-07-04	2007-01-01
Period 23	2007-01-02	2007-07-03
Period 24	2007-07-04	2008-01-02
Period 25	2008-01-03	2008-07-03
Period 26	2008-07-04	2009-01-01
Period 27	2009-01-02	2009-07-03
Period 28	2009-07-06	2010-01-01
Period 29	2010-01-04	2010-07-02
Period 30	2010-07-05	2010-12-31
Period 31	2011-01-03	2011-07-01
Period 32	2011-07-04	2011-12-30

Table B.1: Start and End Dates of the 32 Trading Periods

Curriculum Vitae

Markus Harlacher, born on the 2nd of March 1976 in Winterthur as son of Hans-Heinrich and Elisabeth Harlacher-Matzinger

Education

2015 -	Actuarial Studies SAV, Swiss Federal Institute of Technology (ETH), Zurich
2008 - 2016	Doctoral Studies in Finance, University of St. Gallen (HSG), St. Gallen
2009 - 2012	Studies in Mathematics and Statistics (M. Sc. ETH), Swiss Federal Institute of Technology (ETH), Zurich
2003 - 2008	Studies in Economics and Finance (M. A. HSG), University of St. Gallen (HSG), St. Gallen

Work Experience

2014 -	Pricing Actuary P&C, AXA Winterthur, Winterthur
2013 - 2014	Pricing Actuary P&C, Swiss Re, Zurich
2007 - 2013	Research and Teaching Assistent, Department of Mathematics and Statistics, University of St. Gallen (HSG), St. Gallen
2006 - 2007	Research and Teaching Assistent, Swiss Institute of Banking and Finance, Univeristy of St. Gallen (HSG), St. Gallen