Master Thesis Econometrics & Management Science

FEM21030-16

Multivariate Pairs Trading Using Temporal Dependence Structures

Author:
Jelmer Keuzenkamp
(380267)

Supervisors:
Dr. Erik Kole
Arnout Tilgenkamp

Co-reader:
Dr. Chen Zhou

January 25, 2018

Abstract

Multivariate pairs trading is a strategy that tries to exploit inefficiencies in the relative value pricing between stocks and baskets of related assets. In this research we set up such strategies, where the baskets are identified by elastic net regularization. The elastic net rigorously combines Lasso and ridge regression resulting in compact and robust baskets. We model the spread dynamics of pairs using copula-based (semi-)parametric time series models. This type of model allows for a range of dependency structures and marginals that can be modelled separately. This affords great flexibility and opens up a new way on modelling the spread dynamics. We investigate the strategies' performance for the Japanese universe using daily stock prices of the Nikkei 225 index constituents ranging from August 3rd 2001 to April 1st 2010. We find that the generalization to account for nonlinear associations in the spread dynamics does not necessarily lead to more trading opportunities. However, this generalization leads to 'better' trading signals, i.e. a higher rate of trades that converges. After optimizing the input parameters we find annualized Sharpe ratios up to 1.35 for the copula model and 0.69 for its linear counterpart. Hedging the positions by the identified baskets rather than only stocks' sector members lowers the portfolios' standard deviations, but does not lead to significant higher Sharpe ratios.

Keywords Pairs Trading, Copulas, Elastic Net regularization

Contents

1 Introduction

Pairs trading is a form of statistical arbitrage that exploits short-term deviations from a long-run equilibrium between two securities. Statistical arbitrage encompasses a variety of strategies and is characterized by market neutrality and rule-based signal generations. The idea behind pairs trading is fairly easy. The first step is finding two securities that have moved together historically. Then we monitor the spread between them in a subsequent period. If this spread widens, short the winner and buy the loser. Here it is important that the securities follow a long-run equilibrium, then they mean-revert and money can be made.

In practice one often uses a time series framework that imposes a process on the spread dynamics. Linear autoregressive (AR) models are a standard way of doing this. There is a large number of generalizations towards nonlinear models. Most of these approaches rely on the assumption of Gaussian residual dependence, for example Neumann et al. (1998), Elliott et al. (2005) and Hamilton (1989). However, by allowing for more different dependence structures we can describe this dependence in a much finer way.

The idea of trading individual pairs can be generalized by trading groups of stocks together: one can trade on the relative value pricing between stocks and baskets of comoving assets. Chen et al. (2012) and Perlin (2007) show that this multivariate type of pairs trading results in higher and more robust annual excess returns compared to univariate pairs trading. For this type of pairs trading it is important to recover the baskets properly since this handles an important trade-off. On one hand including a large set of stocks forces the residuals to have small variance. This diminishes the opportunity of making money, especially in the presence of transaction costs. On the other hand, omitting too many important stocks might mean that not all diversification opportunities are exploited. In practice it is hard to identify the baskets. This is caused by changing market conditions and highly correlated stock returns that cause multicollinearity issues.

The question that guides this research is whether we can set up profitable multivariate pairs trading strategies using linear autocorrelation or nonlinear autocopula models for the residuals. This research question leads to different sub-questions. First of all, we are interested in the question whether the generalization to account for both linear and nonlinear associations in residuals lead to more or better trading opportunities and higher profits. Secondly, we are interested in the question whether the pairs trading performance becomes better if we trade stocks pairwise with a basket of comoving stocks rather than their sector members.

We perform our analysis on the Japanese universe using daily stock prices of the Nikkei 225 index constituents ranging from August 3rd 2001 to April 1st 2010. The Japanese universe is interesting to investigate because it has a complex structure: many Japanese firms are part of so called keiretsu organizations. These organizations consist of several firms that have seemingly unrelated businesses across different industries. These type of business structures cause complexity to the ownership structure and the interconnection between firms. Because of these phenomena it is sometimes hard to identify the baskets, or pricing the assets, which might give rise to more arbitrage opportunities.

The methodology in this paper is twofold. For the first part of the methodology, the identification of the baskets, we use the elastic net estimation of Zou and Hastie (2005) that bridges between Lasso and ridge regression. The elastic net estimator both benefits from the sparsity of the Lasso and the stability of the ridge regression. This stability ensures that the elastic net is quite robust to multicollinearity. This is an important benefit for the identification of the baskets from an universe with highly correlated stock returns. Next to that, using the elastic net overcomes caveats of other approaches that have been used in the field of multivariate pairs trading. As an example, the PCA approach as in Avellaneda and Lee (2010) relies on subjective criteria that determine the number of factors or explained variance *a priori*. The elastic net is more rigorous and incorporates this choice into its optimization process, moreover it delivers a structure that is easier to interpret.

The second part of the methodology focuses on modelling the residuals. By doing so we can determine whether current residuals, or a accumulation of residuals, deviate from their historic relation. When this deviation is significant we either long the stock and short its basket or *vice versa*, depending on the direction of this deviation. The proposed trading strategies use a 250-day estimation period and a 125-day out-of-sample trading period for which the estimated parameters, dependency in the residuals and trading bands are held constant. To model the residuals we both need to specify the marginal behavior and the temporal dependence structure. Here we use two different approaches. In the first approach we specify the residuals under a fully Gaussian specification. This boils down on an AR(1) model. In the second approach, as opposed to the former approach, we allow for both linear and nonlinear associations in residuals. We do this by using a copula-based (semi-)parametric time series model. This type of model allows for a range of dependence structures and marginals, which can be modelled separately. We estimate the model parameters in multiple steps using so-called multi-stage maximum likelihood (MSML) estimation. This multi-step estimation eases the estimation procedure, although the estimates lose efficiency (Patton et al., 2012).

We find that both the copula and the benchmark approach are unable to generate profitable trades with Sharpe ratios that vary between 0.07 and -0.99. The copula approach captures a wider variety of specifications for which the benchmark does not suffice: 50% of the stocks has the best fit to a Student-t copula and 30% of the stocks has the best fit to a Frank copula. We can only improve upon the benchmark with the parametric copula model but not enough to be profitable. This is due to a high fraction of trades that does not converge. We can improve the performance by optimizing the input parameters, which results in annualized Sharpe ratios up to 0.69 for the benchmark approach and up to 1.35 for the copula approach.

The generalization to account for nonlinear associations between residuals does not necessarily lead to more trading opportunities. However, for the copula approaches we do observe 'better' trading signals, i.e. a higher rate of convergence. Especially the parametric copula approach benefits from this.

We find that the elastic net selects on average 49% of the stocks that belong to the same industry group. Although this industry connection is relatively strong, the baskets on average contain more stocks than the number of firms in most industry groups. We observe that hedging the positions by the baskets, rather than the sector members, lowers the portfolios' standard deviations significantly. However, Sharpe ratios show that neither the industry approach nor the basket approach performs better consistently.

This research contributes to existing methods: both the way of identifying multivariate pairs with elastic net regularization and the copula-based time series frameworks on modelling the residual processes are novelties. This contribution is large within this field of pairs trading that is heavily exploited by many researchers. Further, this research contributes to the literature by applying the methods to an empirical data set. Previous research lack empirical evidence of pairs trading strategies: most researchers only consider a few pairs for which they test the strategies. From a practical point of view combining the Japanese universe and the newly proposed methods make that the results are potentially useful for practitioners, hedge funds and other trading companies. Our results reveal empirical violation of equilibrium asset pricing. In line with Avellaneda and Lee (2010) we show that we can profit from this upon a linear dimension. We find that we can increase profits by also taking into account nonlinear associations in the residuals. The simplicity of linear associations may induce increased arbitrage activity, resulting in lower profits.

The remainder of this report is structured as follows. Section 2 focuses on literature from previous researches conducted on this subject which provides the starting point of our research. Section 3 describes the data needed for our research. Section 4 describes the methodology and techniques, whereas section 5 provides the results. Finally section 6 concludes and section 7 provides directions to extend this work for future research.

2 Literature

In the scientific arena there are different approaches to set up a pairs trading strategy. Gatev et al. (2006) form the basis for most pairs trading approaches. Their approach has n stocks under consideration for which the sum of Euclidean squared distances of all $n(n-1)/2$ possible combination of pairs is calculated. The top 20 pairs with minimum historic distance are considered in a subsequent trading period. Trades are opened when the spread diverges by more than two historical standard deviations and closed upon mean reversion or at the end of the trading period. The strategy of Gatev et al. (2006) yields average annualized excess returns up to 11%. Besides, their 'pairs' effect differs from previously documented reversal profits.

An other frequently used approach is the cointegration approach. Economically speaking this approach is more reliable than the distance approach since it relies on a long-term equilibrium. The idea is finding cointegrated pairs and use the cointegration relation to construct a spread. Vidyamurthy (2004) is the most cited author in this domain. He proposes a theoretical framework that consists of three key steps. The first step entails a preselection of potentially cointegrated pairs. In the second step one tests for tradability/cointegration. Last, entry signals are triggered when the spread deviates more than two standard deviations from its mean. Caldeira and Moura (2013) investigate the empirical performance applied to 50 liquid stocks of the Brazilian stock index IBovespa. They show statistically significant excess returns after consideration of transaction costs.

An approach that has been discussed more recently and less extensively is the copula approach. The use of copulas is motivated from the fact that former approaches do not fully describe the dependence structure between stocks. Under the consensus that stock returns are rarely joint normal, and thus nonlinear, the conventional approaches no longer appear optimal: trades are based on non-optimal positions due to a loss of dependency information. Xie et al. (2014) make use of this copula framework. They fit bivariate copulas between pairs of stocks which are identified by minimum distance. The conditionals of the copulas indicate the relative value pricing of these pairs. Their framework triggers trading signals once the relative value exceeds the quantiles of the conditional copulas multiple days in a row.

Krauss and Stübinger (2017) extend on the research of Xie et al. (2014) by selecting pairs based on their profitability and by introducing individualized exit rules based on the cumulative returns extracted from a pseudo-trading period. Both Xie et al. (2014) and Krauss and Stübinger (2017) show that the copula approach is a powerful alternative compared to traditional pairs trading techniques: their empirical work verifies the ability to generate higher profits compared to the distance approach as in Gatev et al. (2006).

Perlin (2007) introduces a multivariate version of pairs trading. Here one security is traded against a weighted portfolio of comoving securities instead of a single stock. Perlin (2007) finds that for the Brazilian market the multivariate approaches beat several benchmarks of randomized portfolios. More importantly, he finds that the multivariate approach delivers more robust excess returns compared to univariate pairs trading. Avellaneda and Lee (2010) improved on the work of Perlin (2007). They use a multivariate approach where they trade securities paired with their PCA factors. Their PCA-based strategies beat the alternative of only using stocks' corresponding sector ETF as a factor. As an alternative to their PCA-approach we identify the portfolio of comoving securities using the elastic net estimator of Zou and Hastie (2005). This elastic net delivers a sparse solution and does not rely on subjective criteria that determine the number of factors *a priori*.

Different from our copula-based Markov process, other time series approaches have been used to model the spread dynamics. Elliott et al. (2005) describe the spread with a mean-reverting Gaussian Markov chain which is observed in Gaussian noise. They model this with a state space model and estimate its parameters using the EM algorithm. Various papers extend on the state space representation and the EM estimation, for example Do et al. (2006) and Triantafyllopoulos and Montana (2011).

An other time series approach is modelling the spread dynamics with an Ornstein Uhlenbeck (OU) process. Bertram (2010) form the theoretical foundation for this. Cummins and Bucca (2012) use their approach to 861 energy futures and find daily returns of their top strategies ranging from 0.07 to 0.55 per cent between 2003 and 2010. Zeng and Lee (2014) discuss different trading rules for this framework. They show that it is optimal to set trading thresholds symmetric around the mean of the spread. Similar as for the state space model this OU approach relies on the assumption of normality. As argued in Bertram (2010) this is in conflict with the stylized facts of financial returns.

Our research has both similarities and extensions compared to previous research. Similar as in Perlin (2007) and Avellaneda and Lee (2010) we apply a multivariate pairs trading framework. We extend on their approaches by introducing elastic net regularization to identify the multivariate pairs, rather than PCA or correlation-based measures. Avellaneda and Lee (2010) model the spread dynamics using an AR(1) model, where the spread is defined as the accumulation of residuals. In our approach we first estimate the residual process and subsequently construct the spread as the accumulation of transformed residuals. This approach makes it easier to generalize towards nonlinear dynamics, and does not require multivariate copulas as in Stübinger et al. (2016). We apply the theoretical framework of Xie et al. (2014), but redefine it in a time-series setting. Further, we extend on their approach by comparing the nonlinear methods to a fully linear specification and by considering both a parametric and semi-parametric copula approach.

3 Data

In this research we use daily price observations (close-to-close) of the Nikkei 225 index constituents. The sample runs from August 3rd 2001 to April 1st 2010. This amounts to a total of 2126 daily observations after excluding weekends, holidays and observances. We transform the daily stock prices to log returns by taking the log difference of the stock prices, i.e. $r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1})$, where $P_{i,t}$ is the stock price of firm i at time t. We split this data set up into 15 periods of estimation and evaluation, using a moving window. The estimation periods consist of 250 observations, the evaluation periods consist of 125 observations.

Due to the changing composition of the Nikkei 225 constituents we have to cope with missing values. Ignoring the stocks with missing values over the full sample will cause survivorship bias. That is why we solve this problem by removing stocks that have one or more missing values within a period that we observe them. In total we have 45 out of 225 constituents that enter or exit the index during the full sample period. So it is worth taking these firms into account.

Table 1 summarizes the data set on industry group level. Here we make use of firms' market capitalizations and firms' global industry classification standard (GICS) code, where the latter helps us to find the sector and industry in which each company is assigned to. We see that the number of firms corresponding to an industry group heavily differs. In particular MTRLS, CAPGD, TRANS and TECHH constitute a lot of firms. On the other hand we see that other industry groups as ENERGY, CONSV, MEDIA, HPROD and HCEQS only constitute a few firms. The minimum number of firms per industry shows us the number of firms that exit or enter the Nikkei index and their distribution over the industry groups. Further we see in table 1 that, apart from SMCEQ, a lot of industry groups show similarities regarding their market capitalization. Further table 1 summarizes the average annualized return and its standard deviation, calculated per industry group. These yearly mean log returns vary between -11.5% and 4.0% and the standard deviations between 20.4% and 45.0%. The overall average mean return equals -1.9%, implying that there is on average a downward trend in the Japanese market over the period 2001-2010. This period of economic stagnation is often referred to as 'The Lost Decade' (Harari, 2013).

Table 1: Summary statistics of the Nikkei 225 constituents

This table summarizes the daily log-returns from August 3rd 2001 until April 1st 2010 of the 24 industry groups of the Nikkei 225 constituents. The table shows the industry groups' abbreviations and the (minimum) number of firms, assigned to each industry group. Further the table shows the percentage of average total market capitalization, the percentage of average yearly mean and their corresponding standard deviations.

	Industry group	Firms (min)	Market Cap.(%)	Mean (%)	Std. (%)
ENRGY	Energy	3 (1)	1.63	4.0	37.3
MTRLS	Materials	35 (30)	2.18	0.6	29.5
CAPGD	Capital Goods	47 (43)	2.97	1.5	31.8
COMPR	Commercial & Professional Serv.	3 (3)	2.60	−3.6	27.0
TRANS	Transportation	15 (14)	9.20	−2.3	20.4
AUTOC	Automobiles & Components	13 (13)	3.38	0.1	30.5
CONDA	Consumer Durables & Apparel	9 (8)	0.25	−5.4	31.2
CONSV	Consumer Services	1 (1)	1.51	−3.5	37.3
MEDIA	Media	3 (1)	2.30	−1.9	29.5
RTLNG	Retailing	6 (4)	4.95	−2.4	32.2
FSRET	Food & Staples Retailing	3 (2)	2.40	−1.1	32.8
FBVTA	Food, Beverage & Tobacco	13 (10)	3.89	2.6	22.6
HPROD	Household & Personal Products	2 (2)	2.69	1.0	23.5
HCEQS	Health Care Eq. & Serv.	2 (2)	4.90	6.5	31.1
PHBLS	Pharmaceuticals, Biotechnology & Life Sciences	9 (7)	6.97	−1.4	23.1
BANKS	Banks	11 (3)	3.56	−6.9	39.4
DIVFN	Diversified Financials	4 (4)	4.51	−11.5	40.1
INSUR	Insurance	6 (0)	3.22	2.5	37.7
SFTSV	Software & Services	6 (3)	3.41	−10.5	33.2
TECHH	Technology Hardware & Eq.	15 (15)	1.80	−4.9	30.5
SMCEQ	Semiconductors & Semiconductor Eq.	4 (3)	18.68	−5.0	45.0
TELSV	Telecommunication Serv.	4 (3)	4.91	−7.5	29.4
UTLTS	Utilities	5 (5)	3.75	0.9	18.7
RLEST	Real Estate	6 (5)	4.33	3.1	38.4
	Average	9.38 (7.50)	4.17	-1.9	31.3

4 Methodology

4.1 General framework

We consider a multivariate pairs trading strategy that is based on exploiting relative mispricing between a stock and a portfolio of assets that should price the corresponding stock properly. We write this general framework as

$$r_{i,t} = \beta_{i,0} + \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j} r_{j,t} + \varepsilon_{i,t}, \tag{4.1}$$

where $r_{i,t}$ is stock return i measured at time t and $r_{j,t}$ for $j = 1, \dots, i-1, i+1, \dots, N$ represent the other $(N-1)$ stock returns that act as regressors. Including all stocks forces the residuals to have small variance which diminishes the opportunity of making money, especially in the presence of transaction costs. Therefore we want to impose restrictions on the possible number of stocks that affects a stock return i. We identify those stocks using a penalized regression technique: the elastic net regularization. This technique extracts the relevant stocks from the full set of regressors by forcing coefficients of unimportant stocks to zero. So in this step we form our multivariate pairs.

The identification of the baskets is the first component of this research and is explained in section 4.2. The baskets may consist of firms within the same industry group or, more generally, firms that are closed-connected in a different way. For example, firms connected with their keiretsu organization by shared ownership and other types of collaboration. These keiretsu organizations consist of several businesses that are seemingly unrelated and might operate in different industry groups. That is why we want to allow for structures of interactions beyond industry level. It is more difficult to identify such structures, however this might give rise to a better pairs trading performance.

Different from classical portfolio theory is that we are not only interested in finding a proper set of stocks that explains a certain stock return, but also in the residual that remains. This motivates the second component of our analysis which involves model building for the residual process $\varepsilon_{i,t}$. We consider two different approaches for this. The first approach models the residual process using an AR(1) structure. This means that we have linear dependence and normally distributed error terms. We refer to this approach as the benchmark approach which we describe in section 4.3.1. In the second approach we use a more general framework where we relax the linearity and normality assumptions of the benchmark approach. We refer to this approach as the copula approach which we describe in section 4.3.2.

The pairs trading strategies that we propose consists of two periods: a T-day estimation period and a P-day out-of-sample trading period. In the first period we identify the baskets for every stock, calculate the dependence structure of the residuals and set up bounds that trigger trading signals when we observe significant excursions in the error terms multiple days in a row.

During the estimation period we have all stocks under consideration. But, during the trading period we do not want to have all stocks under consideration. Otherwise we would end up with too many trades that are less likely to be profitable, especially in the presence of transaction costs. Therefore we filter the stocks that we follow in the subsequent trading period. For this trading period we hold the output of the estimation period (estimated parameters, dependency in the residuals and trading bands) constant. During the trading period we start trading the selected pairs when we observe a trading signal. After such signal we hold the pairs until the signal indicates to close the position. During the trading period we reconsider this signal on a daily basis. At the end of a trading period we start a new trading period where we again, from that moment on, use the past T days for estimation.

4.2 Identifying baskets

We want to extract the relevant stocks $r_{j,t}$ from a larger set of potential regressors. Including too many regressors leads to overfitting, resulting in trading decisions that are primarily based on noise. That is why we want to reduce the dimensionality to some extent. The sparsity that is inherent to this approach is an added benefit: trading fewer stocks is easier and better to understand. For this reason we propose an estimator that does both parameter estimation and variable selection simultaneously: the elastic net estimator (Zou and Hastie, 2005). This estimator is related to the seminal work of Tibshirani (1996) and Hoerl and Kennard (1970) that respectively introduces forms of penalized regression with L_1-penalty (Lasso) and a squared L_2-penalty (ridge-regression) on the regression coefficients:

$$\hat{\beta}_i^{Lasso} = \underset{(\boldsymbol{\beta}_i)}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} (r_{i,t} - \beta_{i,0} - \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j} r_{j,t})^2 + \lambda_i \sum_{\substack{j=1 \\ j \neq i}}^{N} |\beta_{i,j}| \right\} \tag{4.2}$$

$$\hat{\beta}_i^{ridge} = \underset{(\boldsymbol{\beta}_i)}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} (r_{i,t} - \beta_{i,0} - \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j} r_{j,t})^2 + \lambda_i \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j}^2 \right\}. \tag{4.3}$$

Figure 1: Graphical illustration of penalized regression

These graphs show the penalty functions of both ridge regression (a) and Lasso (b). The dotted lines represent the penalty functions and the ovals correspond to confidence ellipses.

(a) ridge regression **(b)** Lasso

For penalized regression the confidence ellipse illustrated in figures 1a and 1b has to find a point where it minimizes the sum of squared residuals and at the same time satisfies the imposed restriction. Graphically speaking this implies that the ellipse is tangent to the penalty function. From the figures it becomes clear that for ridge regression this tangent point can only be close to the vertical axis, whereas Lasso provides a tangent point that exactly lies on one of the axis. That is why Lasso provide parsimonious models and act as variable selection, but also shows that ridge regression is often more stable since Lasso can be quite abrupt in determining the optimal position of the confidence ellipse. The elastic net tries to combine both this parsimony of Lasso and the stability of ridge regression: it introduces a penalty term that is a convex combination of the sum of absolute values of the parameters and the sum of squared values of the parameters. We can now define this naive elastic net (NEN) as follows

$$\hat{\boldsymbol{\beta}}_i^{nen} = \underset{(\boldsymbol{\beta}_i)}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} (r_{i,t} - \beta_{i,0} - \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j} r_{j,t})^2 + \lambda_{1,i} \sum_{\substack{j=1 \\ j \neq i}}^{N} |\beta_{i,j}| + \lambda_{2,i} \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j}^2 \right\}. \tag{4.4}$$

If we now define $\alpha_i = \lambda_{2,i}/(\lambda_{1,i} + \lambda_{2,i}) \in [0, 1]$, then solving for $\boldsymbol{\beta}_i$ in (4.4) is equivalent to the optimization

$$\hat{\boldsymbol{\beta}}_i^{nen} = \underset{(\boldsymbol{\beta}_i)}{\operatorname{argmin}} \left\{ \sum_{t=1}^{T} (r_{i,t} - \beta_{i,0} - \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j} r_{j,t})^2 + \lambda_i \left((1 - \alpha_i) \sum_{\substack{j=1 \\ j \neq i}}^{N} |\beta_{i,j}| + \alpha_i \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j}^2 \right) \right\}. \tag{4.5}$$

The function $(1 - \alpha_i) \sum_{\substack{j=1 \\ j \neq i}}^{N} |\beta_{i,j}| + \alpha_i \sum_{\substack{j=1 \\ j \neq i}}^{N} \beta_{i,j}^2$ is called the elastic net penalty. The strength of each penalty is defined by $\alpha_i \in [0, 1)$ which makes that NEN bridges between Lasso ($\alpha_i \to 1$) and ridge regression ($\alpha_i = 0$).

Apart from the fact that the elastic net can incorporate Lasso as a special case, there is also a more general connection between them. We can show this link by introducing an augmented data set $(\boldsymbol{r}_i^*, \mathbf{R}^*)$ obtained from our original data set $(\boldsymbol{r}_i, \mathbf{R})$. Here $\boldsymbol{r}_i$ is a $T \times 1$ vector of returns corresponding to stock i and $\mathbf{R}$ is a $T \times (N-1)$ matrix consisting of stock returns $r_{j,t}$. We define the augmented data set $(\boldsymbol{r}_i^*, \mathbf{R}^*)$ as

$$\mathbf{R}^*_{(T+M)\times(N-1)} = \frac{1}{\sqrt{1+\lambda_{2,i}}} \begin{pmatrix} \mathbf{R} \\ \sqrt{\lambda_{2,i}}\mathbf{I}_M \end{pmatrix}, \qquad \boldsymbol{r}^*_{i\,(T+M)} = \begin{pmatrix} \boldsymbol{r}_i \\ \mathbf{0}_M \end{pmatrix}. \tag{4.6}$$

Here our original $\mathbf{R}$ and $\boldsymbol{r}_i$ of length T are extended to length $T+M$ through matrices $\mathbf{I}_M$ and $\mathbf{0}_M$, where $M = N-1$. Further we define $\gamma_i = \lambda_{1,i}/\sqrt{1+\lambda_{2,i}}$ and $\boldsymbol{\beta}_i^* = \sqrt{1+\lambda_{2,i}}\boldsymbol{\beta}_i$. Now we can define the criterion in (4.4) as:

$$L(\gamma_i, \boldsymbol{\beta}_i^*) = \sum_{t=1}^{T} (r^*_{i,t} - \beta^*_{i,0} - \sum_{\substack{j=1 \\ j\neq i}}^{N} \beta^*_{i,j} r^*_{j,t})^2 + \gamma_i \sum_{\substack{j=1 \\ j\neq i}}^{N} |\beta^*_{i,j}| \tag{4.7}$$

If we now solve for $\boldsymbol{\beta}_i^*$ and obtain the corresponding estimate $\hat{\boldsymbol{\beta}}_i^*$ it turns out that the NEN estimator is just a scalar multiple of the Lasso problem (4.7), i.e.

$$\hat{\boldsymbol{\beta}}_i^{nen} = \frac{1}{\sqrt{1+\lambda_{2,i}}}\hat{\boldsymbol{\beta}}_i^*, \quad \text{where} \quad \hat{\boldsymbol{\beta}}_i^* = \underset{(\boldsymbol{\beta}_i^*)}{\operatorname{argmin}} \left\{ L(\gamma_i, \boldsymbol{\beta}_i^*) \right\}. \tag{4.8}$$

In (4.4) we optimize over different combination of $\lambda_{1,i}$ and $\lambda_{2,i}$, by holding $\lambda_{2,i}$ constant. Therefore we shrink the parameters twice. In an orthogonal design it is straightforward to show this double shrinkage. Kim and Swanson (2013) show that in this setting we can rewrite (4.2) and (4.3) as

$$\hat{\boldsymbol{\beta}}_i^{Lasso} = \left(|\hat{\boldsymbol{\beta}}_i^{ols}| - \frac{\lambda_{1,i}}{2} \right)^+ \operatorname{sgn}(\hat{\boldsymbol{\beta}}_i^{ols}), \tag{4.9}$$

$$\hat{\boldsymbol{\beta}}_i^{ridge} = \frac{\hat{\boldsymbol{\beta}}_i^{ols}}{1+\lambda_{2,i}}, \tag{4.10}$$

where z^+ denotes the positive part, which is z if $z > 0$ and 0 otherwise. Next to that, Zou and Hastie (2005) show that the expression of $\hat{\boldsymbol{\beta}}_i^{nen}$ in this setting equals

$$\hat{\boldsymbol{\beta}}_i^{nen} = \frac{\left(|\hat{\boldsymbol{\beta}}_i^{ols}| - \frac{\lambda_{1,i}}{2} \right)^+}{1+\lambda_{2,i}} \operatorname{sgn}(\hat{\boldsymbol{\beta}}_i^{ols}). \tag{4.11}$$

So the solution of the naive elastic net is a combination of (4.9) and (4.10) and is subject to the shrinkage of both $\lambda_{1,i}$ and $\lambda_{2,i}$. Zou and Hastie (2005) show that empirically the NEN does not perform well because of this double shrinkage: the double shrinkage does not reduce the variance much and may lead to extra bias. We can solve this problem by correcting the double shrinkage. The corrected elastic net estimate of (4.7) is defined as

$$\hat{\boldsymbol{\beta}}_i^{en} = \sqrt{1+\lambda_{2,i}}\hat{\boldsymbol{\beta}}_i^* \tag{4.12}$$

Further, we already found that

$$\hat{\boldsymbol{\beta}}_i^{nen} = \frac{1}{\sqrt{1+\lambda_{2,i}}}\hat{\boldsymbol{\beta}}_i^*, \tag{4.13}$$

therefore,

$$\hat{\boldsymbol{\beta}}_i^{en} = (1 + \lambda_{2,i})\hat{\boldsymbol{\beta}}_i^{nen}. \tag{4.14}$$

So we scale the NEN such that we revert back the ridge shrinkage. Scaling is a simple way to do this, besides the scaled estimator retains the desirable properties of the NEN. From the orthogonal design we have another justification for the scaling factor $(1 + \lambda_{2,i})$. As argued in Donoho et al. (1995) the Lasso estimator is minimax optimal in this setting. From (4.11) we see that this implies that the NEN is not optimal. By scaling the estimator with $(1 + \lambda_{2,i})$ we do achieve this optimality.

The simple augmentation of the elastic net solves the problem of the 'normal' Lasso that can select at most T regressors. The sample size of $\mathbf{R}^*$ is now $T + M$. This implies that the NEN can incorporate all M regressors, also in cases where $M >> T$. This problem might be relevant in this research as the number of regressors might exceed the number of observations. At least this augmentation does not impose restrictions on the choice of the length of the estimation period.

More relevant for our research is the second problem of the Lasso. As argued by Zou and Hastie (2005) Lasso might have problems in selecting variables when regressors do have high pairwise correlation. In those cases Lasso selects randomly among them, not necessarily the most relevant one. One also refers to this as the failure of the 'grouping effect'. This has a negative effect on the performance of Lasso. The elastic net solves this problem because the quadratic part of the penalty encourages grouping effect and stabilizes the regularization path, but the L_1 part still ensures a sparse solution. As the variables that we use are stock returns, we suspect high correlation among variables over the full sample, or over a certain period of time. That is why the use of the elastic net is relevant in our application.

The choice of the tuning parameters $\lambda_{1,i}$ and $\lambda_{2,i}$ is an important one. Those parameters determine the trade-off between variance and bias, and must prevent us from 'noise trading'. There are different approaches to set these tuning parameters. Similar as in Zou and Hastie (2005) we choose to do cross-validation (CV) on a two dimensional surface. In the first step we pick a grid of values for $\lambda_{2,i}$. We opt for a limited range of grid values for $\lambda_{2,i}$. For practical purposes we consider a range of grid values that are more tilted toward Lasso. This might be at the cost of the optimal solution, but gives me a solution that is easier to interpret. We base this choice on the trade-off between limiting the computational burden on one hand and providing a sufficient number of possibilities to optimize over on the other hand. In the next step, for given $\lambda_{2,i}$, we select $\lambda_{1,i}$ by K-fold CV. This means that the $\lambda_{1,i}$ with smallest prediction error (PE) is chosen, which is defined as:

$$\text{PE}_i^{cv} = \frac{1}{T}\sum_{t=1}^{T}(r_{i,t} - \hat{r}_{i,t}^{-k(i,t)})^2, \tag{4.15}$$

where $r_{i,t}^{-k(i,t)}$ is the fitted value for observation t of stock i computed with the $k(t)$th part removed and $k \in \{1, 2, \ldots, 10\}$. So for the k-th part we fit the model to the other $K-1$ parts of the data and calculate the PE of the fitted model when predicting the k-th part of the data. In the last step we consider all $\lambda_{2,i}$ and their corresponding $\lambda_{1,i}$. We choose the $\lambda_{2,i}$ that gives the smallest error PE_i^{cv}.

4.3 Error modeling

Our modelling is separated into two steps. In the first step we identify the baskets. This is done by elastic net. Now we focus on the residuals that remain. To model the residuals we use two different approaches that are discussed in subsections 4.3.1 and 4.3.2. In subsection 4.3.3 we explain the differences and similarities between both approaches.

4.3.1 Benchmark approach

In the first approach we consider our linear model (4.1) where the error term $\varepsilon_{i,t}$ follows an AR(1) process

$$\varepsilon_{i,t} = \varphi_i \varepsilon_{i,t-1} + \eta_{i,t}, \qquad t = 1, \ldots, T, \tag{4.16}$$

where $\eta_{i,t} \sim \mathcal{N}(0, \sigma_i^2)$. So in this benchmark approach we assume linear dependence and normality of $\varepsilon_{i,t}$. The AR(1) process may be written as $(1 - \varphi L)\varepsilon_{i,t} = \eta_{i,t}$, where L is the lag operator. The necessary condition for stationarity is that the roots of $(1 - \varphi B)$ lie outside the unit circle, i.e. $|\varphi_i| < 1$. Under stationary conditions holds that $\varepsilon_{i,t}$ has unconditional distribution $\varepsilon_{i,t} \sim \mathcal{N}(0, \frac{\sigma_i^2}{(1-\varphi_i^2)})$

4.3.2 Copula approach

The probabilistic properties of $\varepsilon_{i,t}$ are determined by the joint distribution function of $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$, say $H_i(\varepsilon_{i,t}, \varepsilon_{i,t-1})$. Sklar (1959) describes that we can express this joint distribution in terms of a marginal distribution for $\varepsilon_{i,t}$ and a copula function of $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$. The marginal distribution characterizes the marginal behavior such as the fat-tailedness of the time series, while the copula function characterizes the temporal dependence structure such as non-linear and asymmetric dependence of the time series. This separation of marginal behavior and dependency structure affords great flexibility and opens up a new way on modelling our residuals. We can write this relation of Sklar (1959) as

$$H_i(\varepsilon_{i,t}, \varepsilon_{i,t-1}) = C_i(G_i(\varepsilon_{i,t}), G_i(\varepsilon_{i,t-1}); \alpha_i). \tag{4.17}$$

Here we consider $C_i(\cdot, \cdot; \alpha_i)$ as the true parametric copula up to unknown α_i, where α_i is a vector that contains the copula dependence parameters of a stock i. Further we denote $G_i(\cdot)$ as the marginal distribution of $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$. Here we assume that we only have one marginal distribution since the time series $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$ are quite similar. Using (4.17) we also obtain the conditional density of $\varepsilon_{i,t}$ given $\varepsilon_{i,t-1}$ which is given by

$$h_i(\varepsilon_{i,t}|\varepsilon_{i,t-1}) = g_i(\varepsilon_{i,t})c_i(G_i(\varepsilon_{i,t}), G_i(\varepsilon_{i,t-1}); \alpha_i). \tag{4.18}$$

Here $c_i(\cdot,\cdot;\alpha_i)$ is the copula density of $C_i(\cdot,\cdot;\alpha_i)$, $h_i(\cdot|\epsilon_{i,t-1})$ is the conditional density function and $g_i(\cdot)$ is the density of the marginal distribution $G_i(\cdot)$. So to describe the conditional distribution $h_i(\varepsilon_{i,t}|\varepsilon_{i,t-1})$ we both need to specify the copula and the marginal distribution. We estimate the copula parametrically and use for the estimation of the marginal $G_i(\cdot)$ both a parametric and a non-parametric approach. Further, to ease the notation, we define the transformed processes $\{v_{i,t} : v_{i,t} \equiv G_i(\varepsilon_{i,t})\}$ and $\{u_{i,t} : u_{i,t} \equiv G_i(\varepsilon_{i,t-1})\}$ and denote $C_i(u_{i,t}, v_{i,t}; \alpha_i)$ as the copula that gives the joint distribution of $u_{i,t}$ and $v_{i,t}$.

4.3.3 Benchmark versus copula approach

Chen and Fan (2006) show that all copula-based first order Markov models $(G_i(\cdot), C_i(\cdot,\cdot;\alpha_i))$ can be transformed in terms of an autoregression transformation model. Now, for ease of notation, we drop the subscript i. This means that $\{\varepsilon_t : t = 1, \ldots, T\}$ is generated from

$$\Lambda_{1,\theta_1}(G(\varepsilon_t)) = \Lambda_{2,\theta_2}(G(\varepsilon_{t-1})) + \sigma_{\theta_3}(G(\varepsilon_{t-1}))\zeta_t, \tag{4.19}$$

where $\Lambda_{1,\theta_1}(\cdot)$ is an increasing parametric function, $\Lambda_{2,\theta_2}(\cdot)$ and $\sigma_{\theta_3}(\cdot) > 0$ are also parametric functions and $\{\zeta_t\}$ is iid with probability density $f_\zeta(\cdot;\theta_4)$. Further we assume that ζ_t is independent of ε_{t-1}. The copula density function of process $\{\varepsilon_t\}$ is defined as

$$c(u_t, v_t; \alpha) = f_\zeta\left(\frac{\Lambda_{1,\theta_1}(v_t) - \Lambda_{2,\theta_2}(u_t)}{\sigma_{\theta_3}(u_t)}; \theta_4\right) \times \frac{d\Lambda_{1,\theta_1}(v_t)}{dv}, \tag{4.20}$$

where α consists of elements θ_1, θ_2, θ_3 and θ_4. In appendix A we elaborate on this type of autoregression transformation models and show the relation when we drop the assumption of ζ_t being independent of ε_{t-1}.

To show the relation between the benchmark approach and the copula approach, we suppose that ε_t and ε_{t-1} follow a Gaussian copula. The Gaussian copula with dependence parameter α is defined as

$$C(u_t, v_t; \alpha) = \Phi_2\left(\Phi^{-1}(u_t), \Phi^{-1}(v_t); \alpha\right), \tag{4.21}$$

where $\Phi_2(\cdot)$ and $\Phi(\cdot)$ are respectively the distribution functions of a standard normal and standard bivariate normal distribution. This means that in equation (4.19) we set $\Lambda_{1,\theta_1}(v_t) = \Phi^{-1}(v_t)$, $\Lambda_{2,\theta_2}(u_t) = \alpha\Phi^{-1}(u_t)$, $\sigma_{\theta_3}(u_t) = \sqrt{1-\alpha^2}$ and $f_\zeta(\cdot;\theta_4)$ is the standard normal density and $\alpha = \theta_2 = \theta_3$. Substituting gives that $\{\Phi^{-1}G(\varepsilon_t)\}$ is a Gaussian process represented by

$$\Phi^{-1}(G(\varepsilon_t)) = \alpha\Phi^{-1}(G(\varepsilon_{t-1})) + \zeta_t, \tag{4.22}$$

where $\zeta_t \sim \mathcal{N}(0, 1-\alpha^2)$. If we also specify $G(\cdot)$ and take as marginal the standard normal distribution, i.e. $G(\cdot) = \Phi(\cdot)$, we see that ε_t again follows an AR(1) process. By varying the specification of $G(\cdot)$ we allow (4.22) to generate processes that are described by a Gaussian copula but non-normal marginals. However we know that a Gaussian copula can not capture nonlinear and asymmetric dependence. The representation above shows us that in this copula approach we extend on the benchmark approach by allowing for more complicated dependence structures among ε_t and ε_{t-1} and by relaxing the assumption of normal marginals $G(\cdot)$.

4.4 Parameter estimation

The type of models we consider in this research consist of different sets of parameters. The first set of parameters corresponds to the β parameters of (4.1) which we estimate by elastic net. The procedure that remains consists of estimating the other sets of parameters that belong to the residual models.

4.4.1 Benchmark approach

The benchmark approach has one set of parameters belonging to its error model. Let $\theta_i = (\varphi_i, \sigma_i^2)'$ denote the vector of parameters from our AR(1) model. To estimate θ_i we use maximum likelihood estimation. The joint density function of $\varepsilon_{i,0}, \dots, \varepsilon_{i,T}$ can be written as:

$$h_i(\varepsilon_{i,0}, \dots, \varepsilon_{i,T}) = \prod_{t=1}^{T} h_i(\varepsilon_{i,t}|\varepsilon_{i,t-1}, \dots, \varepsilon_{i,0}) f_i(\varepsilon_{i,0}) \tag{4.23}$$

Since the process is AR(1) past values only influence a current value via the latest value, therefore we have the Markov property. This means that the conditional distribution of $\varepsilon_{i,t}|\varepsilon_{i,t-1}, \dots, \varepsilon_{i,0}$ is the same as the distribution of $\varepsilon_{i,t}|\varepsilon_{i,t-1}$. The latter is a distribution we can easily derive either directly by conditioning on the information set at time t or indirectly by using the closed-form expression for conditional normals. For both cases we find that

$$\varepsilon_{i,t}|\varepsilon_{i,t-1} \sim \mathcal{N}(\mathrm{E}(\varepsilon_{i,t}|\varepsilon_{i,t-1}), \mathrm{V}(\varepsilon_{i,t}|\varepsilon_{i,t-1})) \tag{4.24}$$

where,

$$\mathrm{E}(\varepsilon_{i,t}|\varepsilon_{i,t-1}) = \varphi_i \varepsilon_{i,t-1} \qquad \mathrm{V}(\varepsilon_{i,t}|\varepsilon_{i,t-1}) = \sigma_i^2 \tag{4.25}$$

The conditional likelihood function can now be written as

$$\mathcal{L}_i(\theta_i) = h_i(\varepsilon_{i,0}, \dots, \varepsilon_{i,T}|\varepsilon_{i,0}) = \prod_{t=1}^{T} h_i(\varepsilon_{i,t}|\varepsilon_{i,t-1}) f_i(\varepsilon_{i,0}) \tag{4.26}$$

So we can find the parameters $(\varphi_i, \sigma_i)'$ by optimizing the log-likelihood for every stock i:

$$\check{\theta}_i = \underset{\theta_i}{\operatorname{argmax}}\, \ell_i(\theta_i) = \underset{\theta_i}{\operatorname{argmax}} \sum_{t=1}^{T} h_i(\varepsilon_{i,t}|\varepsilon_{i,t-1}). \tag{4.27}$$

The function $h_i(\varepsilon_{i,t}|\varepsilon_{i,t-1})$ is known and thus we can find $\check{\theta}_i$ by optimizing

$$\underset{\theta_i}{\operatorname{argmax}} \left\{ -\frac{1}{2} \log\left(\frac{\sigma_i^2}{1-\varphi_i^2}\right) - \frac{\varepsilon_{i,1}^2}{2\sigma_i^2/(1-\varphi_i^2)} - (T-1)\frac{(\log(2\pi) + \log(\sigma_i^2))}{2} - \sum_{t=2}^{T} \left(\frac{(\varepsilon_{i,t} - \varphi_i \varepsilon_{i,t-1})^2}{2\sigma_i^2}\right) \right\}.$$

The estimator $\check{\theta}_i$ is asymptotically normal distributed:

$$\sqrt{T}(\check{\theta} - \theta) \to \mathcal{N}(0, V_{MML}) \text{ as } T \to \infty. \tag{4.28}$$

We can also obtain a consistent estimator of the asymptotic variance, say $\check{V}_{MML}$ which is calculated as $\check{B}_i^{-1}\check{\Sigma}\check{B}_i^{-1}$. First we estimate $\check{B}_i$

$$\check{B}_i = \frac{1}{T-1}\sum_{t=2}^{T} \check{H}_i, \tag{4.29}$$

where $\check{H}_i$ is calculated as

$$\check{H}_i = \begin{pmatrix} \nabla_{11,i} & 0 \\ \nabla_{21,i} & \nabla_{22,i} \end{pmatrix} \qquad \begin{aligned} \nabla_{11,i} &= \frac{\partial^2 \ell_i}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_i'}, \\ \nabla_{21,i} &= \frac{\partial^2 \ell_i}{\partial \theta_i \partial \boldsymbol{\beta}_i}, \\ \nabla_{22,i} &= \frac{\partial^2 \ell_i}{\partial \theta_i \partial \theta_i'}, \end{aligned} \tag{4.30}$$

where $\ell_i(\theta_i)$ the log-likelihood obtained from (4.26). Further, $\check{\Sigma}$ is calculated as the outer product of the gradients:

$$\check{\Sigma}_i = \frac{1}{T-1}\sum_{t=2}^{T} \mathbf{s}_{i,t}\mathbf{s}_{i,t}', \tag{4.31}$$

where

$$\mathbf{s}_{i,t} = \left(\frac{\partial \ell_i}{\partial \boldsymbol{\beta}_i}, \frac{\partial \ell_i}{\partial \theta_i}\right) \tag{4.32}$$

4.4.2 Copula approach general

The type of model we have for the copula approach has two sets of unknown parameters belonging to its error model. Correspondingly, as described by Patton et al. (2012), we can estimate these sets of parameters using multiple steps of estimation here. This multi-stage estimation to estimate the copula dependence parameter extends on the procedure of Genest et al. (1995) for bivariate copula models with iid observations now applied to the class of univariate copula-based (semi-)parametric time series models. The main advantage of this multi-stage estimator is that the estimation problem is greatly simplified.

In the first step of estimation we fit the marginal distribution. Here we make use of different approaches: a fully parametric approach where we estimate the marginals' parameters by maximum likelihood and a semi-parametric approach where we make use of the fact that $G_i(\cdot)$ can be estimated by the empirical distribution function (EDF) $\hat{G}_i(\cdot)$ (Genest et al., 1995). The second step is the estimation of the copula dependence parameter(s) α_i. The estimator for α_i differs from a normal maximum likelihood estimator because we separately estimate the marginals' parameters from the copula parameters. Such models require that the parameters are separable. Clearly, this type of estimation is asymptotically less efficient than one-stage maximum likelihood estimation. However, studies as Joe (2005) show that this loss in efficiency is not great in many cases.

4.4.3 Parametric copula approach

To begin with, we discuss the fully parametric case. In the fully parametric case we specify $G_i(\cdot)$ as the CDF of a Student-t distributions characterized by the degrees of freedom ν_i. We opt for the degrees of freedom that maximizes the log-likelihood function given by

$$\frac{1}{T}\sum_{t=2}^{T}\log\left(\frac{1}{\pi(\nu_i-2)}\frac{\Gamma(\frac{\nu_i+1}{2})}{\Gamma(\frac{\nu_i}{2})}\left[1+\frac{(\frac{\varepsilon_{i,t}-\hat{\mu}_i}{\hat{\sigma}_i})^2}{\nu_i-2}\right]^{-\frac{\nu_i+1}{2}}\right), \tag{4.33}$$

where $\hat{\mu}_i$ and $\hat{\sigma}_i$ are sample estimates. In the second step we estimate the copula dependence parameters. For this purpose we need the conditional distribution $h_i(\varepsilon_{i,t}|\varepsilon_{i,t-1})$. In this parametric case we can identify the conditional distribution of $\varepsilon_{i,t}$ given $\varepsilon_{i,t-1}$ as $h_i(\varepsilon_{i,t}|\varepsilon_{i,t-1}) = g_i(\varepsilon_{i,t})c(G_i(\varepsilon_{i,t}), G_i(\varepsilon_{i,t-1}); \alpha_i)$ and its corresponding log-likelihood function as

$$\ell_i(\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}\log g_i(\varepsilon_{i,t}) + \frac{1}{T}\sum_{t=2}^{T}\log c_i(G_i(\varepsilon_{i,t}), G_i(\varepsilon_{i,t-1}); \alpha_i), \tag{4.34}$$

where the marginal $G_i(\cdot)$ is the CDF of a Student-t distribution characterized by the degrees of freedom ν_i obtained from optimizing (4.33). The first term in (4.34) does not depend on α_i so can be ignored in the optimization. Now optimizing the log-likelihood function gives the parametric estimate $\hat{\alpha}_i$ which is defined as

$$\hat{\alpha}_i = \underset{\alpha_i}{\text{argmax}}\frac{1}{T}\sum_{t=2}^{T}\log c_i(G_i(\varepsilon_{i,t}), G_i(\varepsilon_{i,t-1}); \alpha_i). \tag{4.35}$$

Under some regularity conditions provided by White (1996) this multi-stage estimator $\hat{\alpha}_i$ is asymptotically normal distributed:

$$\sqrt{T}(\hat{\alpha}_i - \alpha_i) \to \mathcal{N}(0, V_{MSML}) \text{ as } T \to \infty. \tag{4.36}$$

Now we obtain the estimator of the asymptotic variance, say $\hat{V}_{MSML}$, by calculating $\hat{B}_i^{-1}\hat{\Sigma}_i\hat{B}_i^{-1}$. Compared to one-stage MLE the outer-matrix $\hat{B}_i$ takes a different form for multi-stage MLE. The outer-matrix reflects the presence of estimated parameters in the log-likelihood and is estimated as

$$\hat{B}_i = \frac{1}{T-1}\sum_{t=2}^{T}\hat{H}_i, \tag{4.37}$$

where $\hat{H}_i$ is calculated as

$$\hat{H}_i = \begin{pmatrix}\nabla_{11,i} & 0\\ \nabla_{21,i} & \nabla_{22,i}\end{pmatrix} \qquad \begin{aligned}\nabla_{11,i} &= \frac{\partial^2}{\partial\gamma_i\partial\gamma_i'}\log g_i(\varepsilon_{i,t}),\\ \nabla_{21,i} &= \frac{\partial^2}{\partial\alpha_i\partial\gamma_i}\log c_i(u_{i,t}, v_{i,t}; \hat{\alpha}_i),\\ \nabla_{22,i} &= \frac{\partial^2}{\partial\alpha_i\partial\alpha_i'}\log c_i(u_{i,t}, v_{i,t}; \hat{\alpha}_i),\end{aligned} \tag{4.38}$$

where γ_i consist of the parameters that correspond to the marginal distribution $G_i(\cdot)$. Further, the inner-matrix $\hat{\Sigma}_i$ is estimated in a similar way as for one-stage MLE and is calculated as

$$\hat{\Sigma}_i = \frac{1}{T-1}\sum_{t=2}^{T} \mathbf{s}_{i,t}\mathbf{s}'_{i,t}, \tag{4.39}$$

where

$$\mathbf{s}_{i,t} = \left(\frac{\partial}{\partial \gamma_i}\log g_i(\varepsilon_{i,t}), \frac{\partial}{\partial \alpha_i}\log c_i(u_{i,t}, v_{i,t}; \hat{\alpha}_i)\right) \tag{4.40}$$

4.4.4 Semi-parametric copula approach

Now we discuss the case where we estimate the copula semi-parametrically. For the parametric approach holds that $G_i(\cdot)$ is subject to estimation error. Next to that it is arguable which distribution fits the data best. To overcome these problems we now propose a semi-parametric case where we replace $G_i(\cdot)$ in equation (4.34) by the estimated distribution $\hat{G}_i(\cdot)$, which is defined as

$$\hat{G}_i(\varepsilon) = \frac{1}{T+1}\sum_{t=1}^{T} \mathbb{I}\{\varepsilon_{i,t} \leq \varepsilon\} \rightarrow G_i(\varepsilon) \text{ as } T \rightarrow \infty. \tag{4.41}$$

Here we use the rescaled EDF to ensure that the criterion function is well defined for all finite T. The use of this EDF not only changes the estimation procedure but also its asymptotic distribution. The semiparametric estimator $\tilde{\alpha}_i$ of α_i is defined as

$$\tilde{\alpha}_i = \underset{\alpha_i}{\text{argmax}} \frac{1}{T}\sum_{t=2}^{T} \log c_i(\hat{G}_i(\varepsilon_{i,t}), \hat{G}_i(\varepsilon_{i,t-1}); \alpha_i). \tag{4.42}$$

Chen and Fan (2006) provided regularity conditions for this estimator $\tilde{\alpha}$ under which the following asymptotic distribution is derived:

$$\sqrt{T}(\tilde{\alpha}_i - \alpha_i) \rightarrow \mathcal{N}(0, V_{SPML}) \text{ as } T \rightarrow \infty, \tag{4.43}$$

We estimate V_{SPML} by its consistent estimator, say $\tilde{V}_{SPML}$, which is calculated as $\tilde{B}_i^{-1}\tilde{\Sigma}_i\tilde{B}_i^{-1}$. The outer matrix B_i^{-1} is an inverse Hessian and estimated by,

$$\tilde{B}_i = -\text{E}\left[\frac{\partial^2}{\partial \alpha_i \partial \alpha'_i}\log c_i(u_{i,t}, v_{i,t}; \tilde{\alpha}_i)\right] = -\frac{1}{T-1}\sum_{t=2}^{T}\frac{\partial^2}{\partial \alpha_i \partial \alpha'_i}\log c_i(u_{i,t}, v_{i,t}; \tilde{\alpha}_i). \tag{4.44}$$

The inner matrix, Σ_i, for this semiparametric estimator is not simply the gradient of the log-likelihood with respect to α_i. Due to the presence of the empirical distribution functions we have to estimate our dependence parameters in multiple stages. That is why the estimator loses efficiency and two additional terms appear in the objective function, namely $\tilde{Q}_{1,i,t}$ and $\tilde{Q}_{2,i,t}$:

$$\tilde{\Sigma}_i = \frac{1}{T-1}\sum_{t=2}^{T} \mathbf{s}_{i,t}\mathbf{s}'_{i,t} \tag{4.45}$$

$$\mathbf{s}_{i,t} = \frac{\partial}{\partial \alpha_i} \log c_i(u_{i,t}, v_{i,t}; \tilde{\alpha}_i) + \tilde{Q}_{1,i,t} + \tilde{Q}_{2,i,t} \tag{4.46}$$

$$\tilde{Q}_{1,i,t} = \frac{1}{T-1} \sum_{s=2, s\neq t}^{T} \frac{\partial^2}{\partial \alpha_i \partial u} \log c_i(u_{i,t}, v_{i,t}; \tilde{\alpha}_i)(\mathbb{I}\{u_{i,t} \leq u_{i,s}\} - u_{i,s}) \tag{4.47}$$

$$\tilde{Q}_{2,i,t} = \frac{1}{T-1} \sum_{s=2, s\neq t}^{T} \frac{\partial^2}{\partial \alpha_i \partial v} \log c_i(u_{i,t}, v_{i,t}; \tilde{\alpha}_i)(\mathbb{I}\{v_{i,t} \leq v_{i,s}\} - v_{i,s}) \tag{4.48}$$

For the case where $G_i(\cdot)$ is completely known both terms $\tilde{Q}_{1,i,t}$ and $\tilde{Q}_{2,i,t}$ disappear from the expression for $\mathbf{s}_{i,t}$. However, in this semi-parametric case $G_i(\cdot)$ is unknown and the asymptotic variance of $\tilde{\alpha}_i$ thus depends on the estimation error in the empirical distribution function via $\tilde{Q}_{1,i,t}$ and $\tilde{Q}_{2,i,t}$. However, we note that the asymptotic variance of the parametric estimator $\tilde{\alpha}_i$ does not depend upon the estimation error from our model (4.1), whereas $\hat{\alpha}_i$ does. So for the semi-parametric case we can ignore the estimation error from equation (4.1) for the purpose of inference and estimation of the copula.

4.5 Copulas applied in this thesis

So far we did not specify the copula $C_i(\cdot, \cdot; \alpha_i)$ for a specific stock i. The literature comprises a large number of different copulas that represent a certain dependence structure. In this research we consider three copulas: the Gaussian copula, the Student-t copula and the Frank copula. Besides that those copulas are commonly used in finance, they also can account for the (possible) negative dependence in our application. In what follows we give a short overview of respectively the Gaussian copula, the Student-t copula and the Frank copula.

Gaussian copula

The bivariate Gaussian copula with dependence parameter $\rho_i \in (-1, 1)$ is given by

$$C(u_{i,t}, v_{i,t}; \rho_i) = \Phi_2\left(\Phi^{-1}(u_{i,t}), \Phi^{-1}(v_{i,t}); \rho_i\right), \tag{4.49}$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal, $\Phi_2(\cdot)$ is the joint cumulative distribution function of a bivariate normal distribution with zero means and covariance matrix Σ. When ρ_i approaches 1 and -1, the Gaussian copula attains the Fréchet–Hoeffding upper and lower bound respectively. Further, when $\rho_i = 0$ we attain the case of independence. The Gaussian copula is flexible as is allows an equal degree of positive and negative dependence and includes both Fréchet bounds in its permissible range.

Student-t copula

The bivariate Student-t copula distribution is given by

$$C(u_{i,t}, v_{i,t}; \rho_i, \nu_i) = t_{2,\nu_i}(t_{\nu_i}^{-1}(u_{i,t}), t_{\nu_i}^{-1}(v_{i,t})), \tag{4.50}$$

where $\nu_i \in (0, \infty)$ represents the degrees of freedom, t_{2,ν_i} the joint cumulative Student-t distribution function, t_{ν_i} the univariate Student-t distribution function and $\rho_i \in (-1, 1)$ the copula parameter. Similar as the Gaussian copula, the Student-t copula allows an equal degree of positive and negative dependence and includes both Fréchet bounds in its permissible range. As opposed to the Gaussian copula we do not attain the case of independence when $\rho_i = 0$. In this case the Student-t copula still shows dependence since uncorrelated multivariate Student-t random variables are not independent. Further we note that when $\nu \to \infty$, then $C(u_{i,t}, v_{i,t}; \rho_i, \nu_i) \to \Phi_2\left(\Phi^{-1}(u_{i,t})\Phi^{-1}(v_{i,t}); \rho_i\right)$.

Frank copula

The Frank copula is a so called Archimedean copula. This family of copulas are built on a generator function $\psi(\cdot)$ and can be written as:

$$C(u_{i,t}, v_{i,t}; \theta_i) = \psi^{-1}(\psi(u_{i,t}) + \psi(v_{i,t})). \tag{4.51}$$

For the Frank copula this generator function is defined as:

$$\psi(u_{i,t}) = -\ln\left(\frac{\exp(-\theta_i u_{i,t}) - 1}{\exp(-\theta_i) - 1}\right). \tag{4.52}$$

So the copula is defined as follows:

$$C(u_{i,t}, v_{i,t}; \theta_i) = -\frac{1}{\theta_i}\ln\left[1 + \frac{(\exp(-\theta_i u_{i,t}) - 1)(\exp(-\theta_i v_{i,t}) - 1)}{(\exp(-\theta_i) - 1)}\right], \tag{4.53}$$

where $\theta_i \in (-\infty, \infty) \setminus \{0\}$ is the copula parameter. The Frank copula allows for the maximum range of dependence. This means that the dependence parameter permits the approximation of the upper and lower Fréchet–Hoeffding bounds when θ_i goes to ∞ or $-\infty$. So the Frank copula permits both positive and negative dependence in the data, which is symmetric in both tails. When θ_i approaches 0 the case of independence will be attained.

To determine the best fitting copula (Gaussian, Student-t or Frank) for a specific stock we use the Akaike information criterion (AIC). We calculate this criterion for all copulas and choose the copula that has minimum AIC. Here we opt for a dynamic approach and thus refit the dependence structure every estimation window. Although this is computationally more intensive, it is more realistic to allow for a dependence structure that may change over time. The dynamic AIC criterion that we obtain is defined as:

$$AIC_{i,s} = -2\ell_{i,s}(\alpha_{i,s}) + 2k_{i,s}, \tag{4.54}$$

where $\ell_{i,s}$ is the optimized log-likelihood function for stock i in estimation period $s \in \{1, 2, \ldots, S\}$. Further k denotes the number of parameters in α_i.

4.6 Signal generation

After the estimation of the parameters $\beta_{i,j}$ and the parameters corresponding to the residual processes $\{\varepsilon_{i,t}\}$, we need to decide which stocks to trade and when to trade on the mispricing that we observe. In subsection 4.6.2 we explain how we specify these important trading rules. But first we explain in the following subsection 4.6.1 how we extract this mispricing from our models.

4.6.1 Detecting mispricing

The dependence structure captured in our models for the residuals gives an indication of the relative behavior of error terms $\varepsilon_{i,t}$ and its lagged version $\varepsilon_{i,t-1}$. We can use this dependency to detect excursions of residuals and argue whether a current observation deviates from the historic dependence structure between $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$. To do this we define a mispricing index for both the benchmark approach and the copula approach. This index is defined as:

$$MI_{i,t}(\varepsilon_{i,t}, \varepsilon_{i,t-1}) = P(\varepsilon_{i,t} < \varepsilon | \varepsilon_{i,t-1}). \tag{4.55}$$

For the benchmark approach we make distributional assumptions and thus we know the conditional distribution of $\varepsilon_{i,t}$ given $\varepsilon_{i,t-1}$. This implies that we can easily derive the mispricing index for this case as:

$$\begin{aligned} MI_{i,t}(\varepsilon_{i,t}, \varepsilon_{i,t-1}) &= P(\varepsilon_{i,t} < \varepsilon | \varepsilon_{i,t-1}) \\ &= P(G_i(\varepsilon_{i,t}) < G_i(\varepsilon) | G_i(\varepsilon_{i,t-1})) \\ &= P(v_{i,t} < v_i | u_{i,t}) = v_i \sim \mathcal{U}(0,1). \end{aligned} \tag{4.56}$$

For the copula approach the calculation of the mispricing index is more complicated. In this case we make use of the fact that the distribution of $v_{i,t}$ conditional to $u_{i,t}$, $C(v_{i,t}|u_{i,t}; \alpha)$, is equivalent to $\frac{\partial}{\partial u} C(u_{i,t}, v_{i,t}; \alpha_i)$. That is why for the copula approach we define the mispricing index as:

$$\begin{aligned} MI_{i,t}(\varepsilon_{i,t}, \varepsilon_{i,t-1}) &= P(\varepsilon_{i,t} < \varepsilon | \varepsilon_{i,t-1}) \\ &= \frac{\partial}{\partial u} C(u_{i,t}, v_{i,t}; \alpha_i) \\ &= C(v_{i,t} | u_{i,t}; \alpha_i) = v_i \sim \mathcal{U}(0,1). \end{aligned} \tag{4.57}$$

Similar as for the benchmark case this index $MI_{i,t}$, or conditional distribution, also follows an uniform distribution on [0,1].

To detect mispricing from uniforms (4.56) and (4.57) we make use of the following reasoning. When we observe a value of 0 for $(MI_{i,t} - 0.5)$ we can say that this reflects no mispricing: a value of $\varepsilon_{i,t}$ is neither too high nor too low given $\varepsilon_{i,t-1}$ and their historical relation. The historical data simply indicates that on average there is equal number of observations for $\varepsilon_{i,t}$ being smaller or larger than the realization of $\varepsilon_{i,t}$ given $\varepsilon_{i,t-1}$. A similar reasoning can be used to verify that we have an indication for mispricing when $(MI_{i,t} - 0.5)$ sufficiently deviates from 0.

The measure $MI_{i,t}$ has several advantages. First of all, a same degree of mispricing is reflected by the same numerical value. This makes that we can easily compare different mispricing indexes of different stocks. This characteristic overcomes, for instance, the shortcomings of the distance measure which can have identical numerical values although the degree of mispricing differs. Secondly, the measure is easy to calculate and thus easy to implement. Last, we can use the same measure for both our benchmark approach and our copula approach. This enables a fair comparison between the performance of both measures. The major difference between the measures $MI_{i,t}$ is that for the copula case the measure contains information on both linear and nonlinear associations between residuals.

Similar as in Xie et al. (2014) we accumulate individual mispricing values of a series of subsequent observations. This sum of conditional probabilities can be seen as a degree of cumulative mispricing. This approach is appealing as it reflects the mispricings over multiple days and thus retains the time structure, i.e. trading signals are not only based on the last return and assess how pairs trade subsequently (convergence/further divergence). Compared to the mispricing indices the cumulative mispricing will result in a more stable trading strategy since we do not trade on every single excursive mispricing.

We capture the dynamics of our cumulative mispricing measure with a time dependent spread, say $S_{i,t}$, which is defined as

$$S_{i,t} = S_{i,t-1} + (MI_{i,t} - 0.5) \text{ for } t = 1, \ldots, T + P, \tag{4.58}$$

with $S_{i,0} = 0$ and $S_{i,T+1} = 0$. With other words, $S_{i,t}$ is the sum of a series of uniform random variables provided that we ignore the correlation between them and view the situation as akin to a pure random walk. By construction $S_{i,t}$ is not a stationary time series: it can either move up, move down or fluctuate around zero. However, pairs trading is based on the assumption of mean-reversion indicating that $S_{i,t}$ has a tendency to revert back to zero when it has moved away significantly.

The properties of $S_{i,t}$ depend on the time-series correlation between $e_{i,t} = (MI_{i,t} - 0.5)$ and $S_{i,t-1}$, where $e_{i,t}$ is uniform on (-0.5, 0.5). In general we can classify three mechanisms that are relevant here for the $S_{i,t}$ series.

- There is no time-series correlation between $e_{i,t}$ and $S_{i,t-1}$. In this case $S_{i,t}$ should be a purely random walk. There is no arbitrage opportunity in this case.

- There is positive time-series correlation between $e_{i,t}$ and $S_{i,t-1}$. In this case $S_{i,t}$ has a tendency to diverge, which generally results in a loss.

- There is negative time-series correlation between $e_{i,t}$ and $S_{i,t-1}$. This means that $S_{i,t}$ has a tendency to converge when it is far away from zero, which generally results in profits.

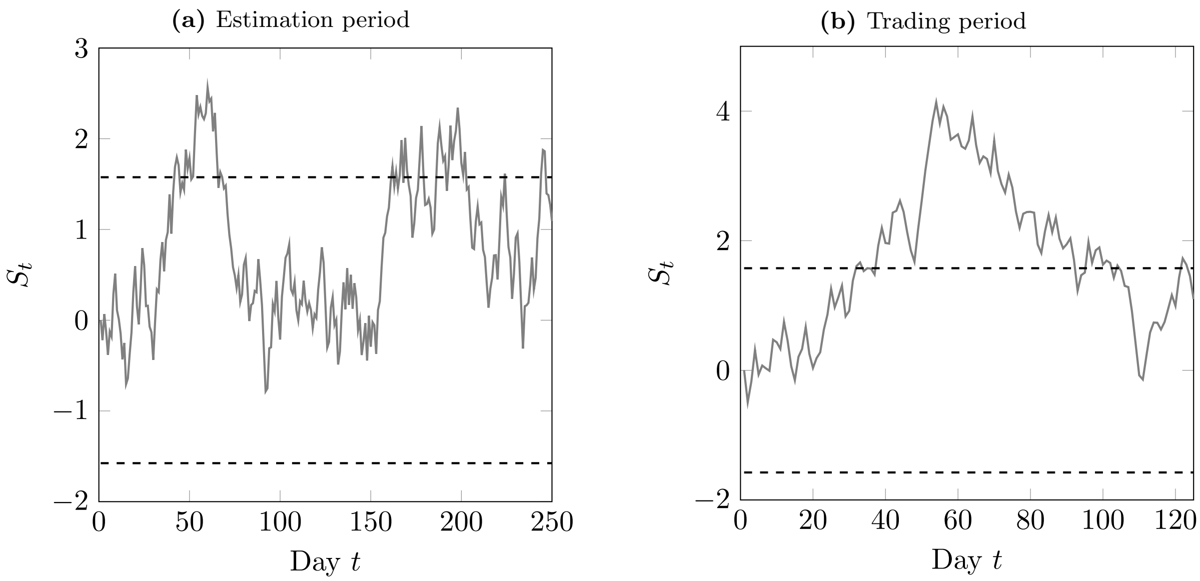
Empirically $S_{i,t}$ alternates across the three mechanisms. This results in both profits and losses. As long as the convergent mechanism dominates here we can profit from our pairs trading strategies.

4.6.2 Trading decisions

To extract trading signals from the spread series we use a similar approach as in Landgraf (2016). We define trading bounds $D_i = 2\sigma_{s,i}$ and $-D_i = -2\sigma_{s,i}$, where $\sigma_{s,i}$ denotes the standard deviation of the spread corresponding to stock i, only using the in-sample observations $S_{i,0}, \ldots, S_{i,T}$. So rather than prefixing all trigger points we use the data to specify them and allow for stock specific heterogeneity by calculating D_i and $-D_i$ for the stock specific spread series. These trading bounds indicate when the accumulation of mispricings is worth trading on. So when the spread $S_{i,t}$ crosses D_i or $-D_i$ we open a position. In the figures below we show an example of the in-sample and out-of-sample spread dynamics and its corresponding trading bounds.

Figure 2: Illustration of spread dynamics and bounds

Illustration of spread dynamics S_t and its bounds $(D_i, -D_i)$ for a 250-day in-sample estimation period and a 125-day out-of-sample trading period for the stock of Toyota Tsusho Corp.



The positions that we take during the out-of-sample trading period differ for each trigger point. Once we find that $S_{i,t}$ crosses $D_i = 2\sigma_{s,i}$ there is strong positive comovement among the errors and we thus short 1 yen in stock i and long $\beta_{i,j}$ yen in the jth stock of the basket of stock i. Reversely, if we find that $S_{i,t}$ crosses $-D_i = -2\sigma_{s,i}$ there is strong negative comovement among the errors and we thus long 1 yen in stock i and short $\beta_{i,j}$ in the jth stock of the basket of stock i. We close the position if the spread series returns to zero or at the end of a trading period. Further, to limit the losses when $S_{i,t}$ diverges we introduce stop-loss bounds: $SL_i = 8\sigma_{s,i}$. When $S_{i,t}$ crosses either SL_i or $-SL_i$ we close the opened position and reset $S_{i,t}$ to zero. This stop-loss will be triggered only a few times but might improve the performance significantly.

Given that the measure $MI_{i,t}$ is comparable across stocks we can use it to set up a selection criterion that determines which pairs potentially generate profitable trading opportunities. This selection process filters the N stocks under consideration during the estimation period to k stocks under consideration during the trading period. Optimally we want to select stocks that have a strong in-sample mean reversion in their spread series. Under the assumption that this mean reversion is likely to hold out-of-sample, it is also likely that we can make profitable trades after we observe a mispricing. As a heuristic for mean reversion we count the number of zero-crossings, i.e. the number of times the spread series $S_{i,t}$ crosses zero. Then, at the end of a estimation period, we transfer the pairs with the top k number of zero-crossings to the subsequent trading period.

Because the spreads start at zero it is much more likely that zero-crossings appear in the first few steps of the motion: a small up-move followed up by a larger down-move directly causes a zero-crossing. Moreover zero-crossing on itself does not generate profit. To overcome these problems we count the number of zero-crossings which started initially outside the bounds D_i or $-D_i$. If the number of zero-crossings is equal between stocks we distinguish between them by selecting stocks that have maximum total distance. This total distance is calculated as the sum of distances between the initial starts outside the bounds D_i or $-D_i$ and zero.

After we have set up all trades it might be possible that the total net position of the created portfolio has exposure to the market. Optimally we want the positions such that our portfolio is not affected by the market's direction, i.e. the portfolio is market-neutral. To assure market-neutrality of the strategies we invest in the Nikkei index itself. We do this by calculating

$$MN_t = \sum_{i=1}^{N} \beta_i w_{i,t} \text{ for } t = T+1, \ldots, T+P, \tag{4.59}$$

where $w_{i,t}$ is the total position in stock i at time t and β_i the exposure of stock i to the market. Here $w_{i,t}$ consists of either a single position in a specific stock or the sum of multiple positions when stock i both act as a selected stock that we trade versus its basket and simultaneously act as a part of the basket for an other selected stock at time t. More specifically we define

$$w_{i,t} = \sum_{j \in N^*} w_{i,t,j}, \tag{4.60}$$

where N^* is the number of selected. Our strategies are market neutral if $MN_t = 0$. When we observe that $MN_t > 0$ or $MN_t < 0$ we invest in the Nikkei index such that $MN_t = 0$ to ensure market neutrality.

4.7 Excess return computation

The calculation of excess returns is a simple approach to evaluate our trading strategy. The full set of trades forms a portfolio of pairs. Because pairs open and close at various points over time the calculation of the excess return is not trivial. However this is important to asses our strategy: pairs can have both a positive cash flow when the spread series reverts back to zero or a negative cash flow when the spread series has not reverted at the end of the trading period. As noted by Gatev et al. (2006) there are several ways to measure this excess return computation: either by calculating the return on committed capital or by calculating the return on the fully invested return. The first measurement takes into account all the money that is prepared for potential trades, so also money that is committed for a potential pairs trade when the position is not open during the trading period. This measure marks the portfolio return as zero on days that no trades occurred. This is the difference with the second measurement that only considers money currently being traded and thus disregards trading days in which we do not observe trading signals. So the first measurement is more conservative. We compute the daily portfolio returns on the long and short positions as value weighted returns:

$$r_{p,t} = \frac{\sum_{i=1}^{N} w_{i,t} r_{i,t}}{\sum_{i=1}^{N} w_{i,t}} \tag{4.61}$$

$$w_{i,t} = w_{i,t-1}(1 + r_{i,t-1}), \tag{4.62}$$

where $r_{p,t}$ is the portfolio return at time t and $w_{i,t}$ defines weights as in (4.60). Similar as for the case where we make the portfolio market neutral we take into account that the weights $w_{i,t}$ can consist of both a long or short position in a specific stock i and a position in a stock part of a basket for an other stock j at the same time t.

4.8 Performance measures

Next to calculating the excess returns and considering its distributive characteristics we want to use some other measures to assess the performance of the pairs trading strategies. Here it is important to capture different characteristics such as the risk-return profile and the portfolios' tail risk behavior. Besides, we want to do some analysis how the portfolio returns relate to previously documented reversal strategies. In this section we give an overview of all performance measures we use.

4.8.1 Risk-return profile

The most widely used performance measure to assess the risk-return profile is the Sharpe ratio as in Sharpe (1994). This measure is defined as

$$SR = \frac{\mu}{\sigma}, \tag{4.63}$$

where $\mu = \frac{N}{P}\sum_{t=1}^{P} r_{p,t}$ is the annualized average excess return and $\sigma = \sqrt{\frac{N}{(P-1)}\sum_{t=1}^{P}(r_{p,t}-\bar{r})^2}$ is the annualized standard deviation. Further P is the number of trading days in a trading period, N is defined as the number of trading days per year and $\bar{r} = \mu/N$ is the average portfolio return. We set N equal to 245 which is the average number of trading days in Japan.

4.8.2 Tail-risk behavior

In order to assess the tail risk behavior of the strategies we simply look at the characteristics of the distribution of excess returns such as the skewness and kurtosis. But, differently, we also want to look at the maximum drawdown (MDD) (Magdon-Ismail and Atiya, 2004) of the strategy. The MDD can be interpreted as an indicator of downside risk over a certain time period. The MDD is defined as

$$MDD = \max_{\tau \in (0,P)} \left| \frac{\max_{t\in(0,\tau)}(p(t)-p(\tau))}{\max_{t\in(0,\tau)} p(t)} \right| \tag{4.64}$$

where $p(t) = \prod_{t=1}^{\tau}(1+r_{p,t})$ is the portfolio value at time t. Using this MDD we can define several tail-risk measures. First of all we use the Calmar ratio (CR) as in Young (1991) which is defined as

$$CR = \frac{\left[\sum_{t=1}^{P} r_{p,t}\right]\frac{N}{P}}{MDD}, \tag{4.65}$$

and similarly we use the Burk ratio (BR) as in Burke (1994) which is defined as

$$BR = \frac{\left[\sum_{t=1}^{P} r_{p,t}\right]\frac{N}{P}}{\sqrt{\sum_{k=1}^{10}\max\{MDD_k\}^2}}. \tag{4.66}$$

The first measure is a ratio that compares the annualized excess returns to the maximum drawdown over the period $(0, P)$. The second measure is a ratio that compares the annualized excess returns to the sum of the 10 largest MDDs over the period $(0, P)$.

4.8.3 Systematic risk exposure

Because pairs trading strategies are based on the relative value of stocks it might be possible that the strategies simply exploit patterns that are known to be profitable. To investigate this relation to previously reversal strategies we regress the daily committed capital returns of our strategies on the four-factor model of Carhart (Carhart, 1997). This model consist of the market excess return, the size factor, the value factor and the momentum factor. The first 3 factors are constructed as in Fama and French (1992). The excess market factor is calculated by using the Nikkei index in excess to the one-month Japanese T-bill rate. For the size factor, small minus big (SMB), we sort the stocks on market capitalization and create two value weighted portfolios. The small portfolio contains stocks with smaller than median market capitalization and the big portfolio contains stocks with larger than median market capitalization. Now we create SMB by subtracting the big portfolio returns from the small portfolio returns. In a similar way we create the value factor. Now we sort stocks on price to book ratio. The value factor, high minus low (HML), is created by subtracting the value-weighted high portfolio return minus the value-weighted low portfolio return. Here we use the 30th and 70th percentile as cutoff points to indicate whether stocks belong to the high or low portfolio. Our last factor, the momentum factor, is created as in Carhart (1997). For this factor we again sort stocks, but now based on their cumulative return for month $t-12$ to month $t-2$. The factor returns are calculated by subtracting loser portfolio returns from the winner portfolio returns. Again we use the 30th and 70th percentile as cutoff points to indicate stocks as loser, winner or neither of them. After we have constructed all factors we perform the regression that reads as

$$r_{t,p} = \alpha + \boldsymbol{\beta} \boldsymbol{f}_t + \omega_t \tag{4.67}$$

where $\boldsymbol{f}_t$ is a vector that consist of the factor returns at time t, $r_{t,p}$ the strategies' excess portfolio returns and ω_t denotes an error term. To determine the significance of the coefficients of (4.67) we make use of t-statistics with Newey-West (NW) standard errors (Newey and West, 1986). These standard errors are more robust for autocorrelation and heteroskedasticity. The NW standard errors of the estimator $\hat{\boldsymbol{\beta}}$ are given by the square roots of the diagonal elements of of the matrix

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \frac{1}{T^*}(\mathbf{F}'\mathbf{F})^{-1}\hat{\mathbf{V}}(\mathbf{F}'\mathbf{F})^{-1}, \tag{4.68}$$

where

$$\hat{\mathbf{V}} = \frac{1}{T^*}\sum_{t=1}^{T^*} e_t^2 \boldsymbol{f}_t \boldsymbol{f}_t' + \frac{1}{T^*}\sum_{t=1}^{T^*-1}\sum_{j=t+1}^{T^*} w_{j-i} e_t e_j (\boldsymbol{f}_t \boldsymbol{f}_j' + \boldsymbol{f}_j \boldsymbol{f}_t'). \tag{4.69}$$

Here $\mathbf{F}$ is a $(T^* \times 4)$ is matrix with the factor returns, T^* is the total number of observations, e_t are the residuals from regression (4.67) and w corresponds to the so called Barlett kernel and has weights $w_b = \frac{b}{(T^*)^{1/3}}$.

5 Results

5.1 Pairs formation

We start our analysis with the elastic net estimation. The elastic net estimators of all stocks form a sparse matrix with the structure of interactions between a single stock i and a basket of stocks j. The baskets consist of stocks that are close connected to its corresponding linked stock. This connection can have different sources, such as stocks belonging to the same industry, stocks from related industries, or stocks that belong to the same keiretsu organization.

For ease of interpretation we aggregate the results on industry group level in this subsection. We define an industry connection between industry group a and industry group b as a connection between stock $i \in a$ and stock $j \in b$ imposed by the elastic net estimator, i.e. $\beta_{i,j} \neq 0$. The fractions in figure 3 indicate the selected number of stocks relative to the number of stocks assigned to the same industry group. The fractions are calculated as an average over all estimations periods and take into account the varying number of firms that are assigned to an industry group in a specific estimation period. This number varies because we only take into account stocks that have no missing observations over an estimation period.

Figure 3 summarizes the compositions of the baskets on industry group level. It is clear that we see a diagonal structure: stocks are on average linked to 49% of the stocks that belong to the same industry group. This effect is stronger for some industries compared to other industries. For example, stocks that belong to HPROD or HCEQS are almost in all cases linked to stocks of the same industry group. This makes sense since these industry groups only consist of two firms and thus the strength of the industry connectedness does not 'suffer' from the sparsity imposed by the elastic net regularization. On the other hand, stocks within MRTLS and CAPGD seem to have weaker connections. Again one has to take into account the size of the industry groups: MRTLS and CAPGD respectively consist of 35 and 47 firms. The elastic net forces a small number of connections which implies a small fraction of selected stocks.

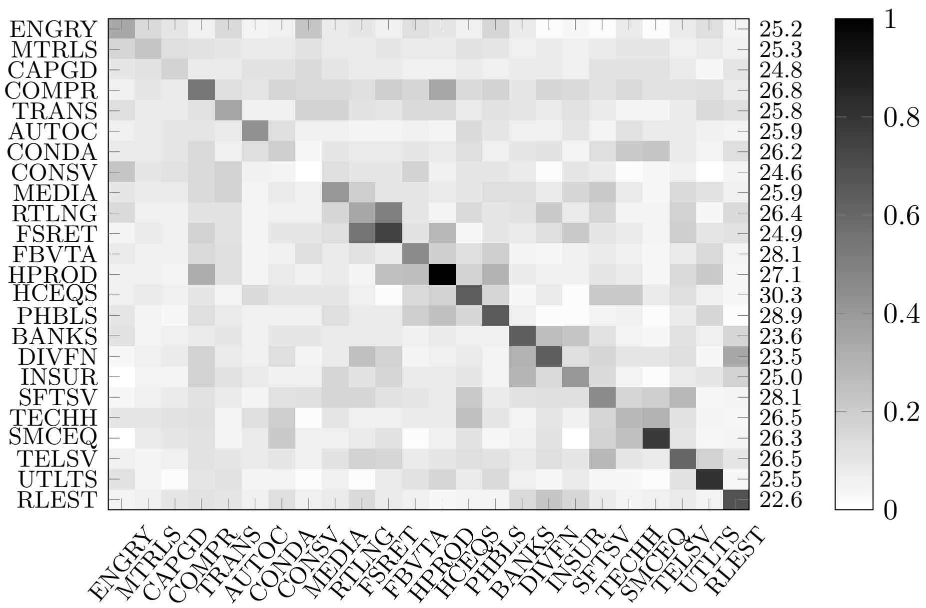
Next to these industry connections we see connections on cross-industry level. On average the baskets consist of 22.8 stocks with an average standard deviation of 13.8 stocks. Although the industry connection is quite strong, the baskets on average contain more stocks than the number of firms in most industry groups. This difference is caused by stocks that belong to other industries. This motivates why it might be better to trade stocks versus their basket of stocks rather than, for example, only versus their own industry group members.

The column on the right hand side of figure 3 shows the average number of times that firms of a specific industry are selected across all industries. Those numbers indicate the average importance of a sector. We observe that those averages do not differ a lot and range from 30.3 times until 22.6 times. Remarkable is that HCEQS has the highest average, although HCEQS only consists of two firms. Apparently a lot of stocks are related to those firms. More generally we see that the industries with the largest number of firms are not necessarily the industries that most frequently appear in a basket.

For verification purposes, figures 13 and 12 respectively show the fractions that belong to figure 3 and the average number of connections between industry groups. They can be found in appendix B.

Figure 3: Elastic net output

This table shows the number of selected stocks in an industry group as a fraction of the number of total stocks in an industry group. The fractions are calculated as an average over estimation windows $s = 1, \dots, S$. A stock j is considered as selected when $\beta_{i,j}^{(s)} \neq 0$ and correspond to a connection between industry a and b when firm $i \in a$ and firm $j \in b$, evaluated at estimation window s. Further, the column on the right hand side shows the average number of times a firm of a specific industry is selected across all industries.



5.2 One pair, one cycle: Shiseido Co Ltd

In order to illustrate the functionality of the methods we show an example where we analyze a specific stock: Shiseido Co Ltd. This firm produces cosmetics and is assigned to the industry group HPROD. We do this small analysis for the first period of estimation and trading in the sample. This block consists of a 250-day estimation period that runs from 03/08/2001 until 07/08/2002 and a 125-day trading period that runs from 08/08/2002 until 13/02/2003.

Basket composition

As a first step of the analysis we identify the basket of Shiseido. Table 2 shows the elastic net output for this stock under consideration.

Table 2: Elastic net results Shiseido

This table shows the elastic net coefficient estimates for Shiseido. The model is described as $r_{i,t} = \beta_{i,0} + \sum_{\substack{j=1 \\ j\neq i}}^{N} \beta_{i,j} r_{j,t} + \varepsilon_{i,t}$, where $r_{i,t}$ describes the stock returns of Shiseido and $r_{j,t}$ describes the j-th stock return in the Nikkei universe. The elastic net selects the stocks as indicated in the table with corresponding weights and stocks' industry group.

Weight	**Stock**	**Industry (firms)**
0.0989	Kao Corp	HPROD (2)
0.0061	Astellas Pharma Inc	PHBLS (7)
0.0007	Takota Pharmaceutical	PHBLS (7)
0.1015	Sumitomo Pharma Co Ltd	PHBLS (7)
0.0476	Kirin Holdings Co Ltd	FBVTA (10)
0.0623	Sapporo Holdings Ltd	FBVTA (10)
0.0406	Kubota Corp	CAPGD (43)
0.0046	Furukawa Co Ltd*	CAPGD (43)
0.0012	Hino Motors Ltd	CAPGD (43)
0.0276	OKUMA Corp	CAPGD (43)
0.0205	Sumitomo Heavy Industries Ltd	CAPGD (43)
0.0470	Keio Corp	COMPR (3)
0.0592	Dai Nippon Printing Co Ltd*	COMPR (3)
0.0009	Shizuoka Bank Ltd	BANKS (3)
0.0109	Citizen Watch Co Ltd	TECHH (15)
0.0030	Nippon Kayaku Co Ltd*	MTRLS (30)
0.0215	Toray Industries Inc	MTRLS (30)
0.0260	Tokai Carbon Co Ltd	MTRLS (30)
0.0025	Sumitomo Chemicals Co Ltd	MTRLS (30)
0.0026	Tokuyama Corp	MTRLS (30)
0.0724	ANA Holdings Inc	TRANS (15)
0.0052	Central Japan Railway Co	TRANS (15)

* Keiretsu members of Shiseido.

Table 2 shows that the basket of stocks contains 22 stocks. This means that the other 162[1] coefficients are forced to zero by the elastic net regularization. The structure that we obtain has several patterns. First of all, the basket consists of large firms such as Kubota Corp, Shizuoka

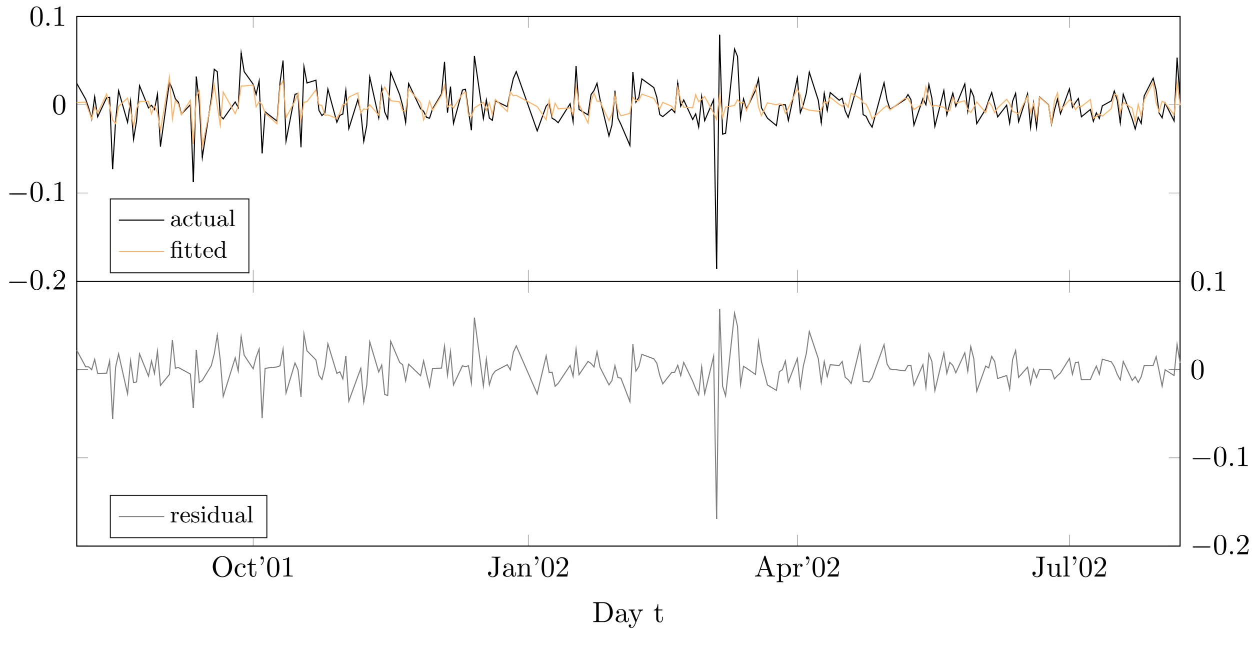
[1]The first estimation window has 184 stocks that have available observations over this estimation sample.

Bank Ltd, Sumitomo Heavy Industries Ltd and Sumitomo Chemicals Co Ltd. Secondly, the basket contains firms that are member of the same keiretsu organization (Daiichi Kangyo group): Nippon Kayaku Co Ltd, Dai Nippon Printing Co Ltd and Furukawa Co Ltd. Further we see that Kao Corp is selected with relatively large weight. This makes sense since this is the only other company in the industry group HPROD and also a large producer of cosmetics (as well as Shiseido). Last, we see that firms of related industry groups are selected such as the pharmaceutical firms Astellas Pharma Inc, Takota Pharmaceutical and Summitomo Pharma Co Ltd.

Figure 4 shows the actual, fitted and residual time series that we obtain from the elastic net estimates. We observe that the residuals do not look completely identical to a white noise series. It looks like the residuals have some heteroskedastic and autocorrelated behavior. We can test whether the autocorrelation is different from zero by performing a Ljung-Box test at the first lag. This delivers a p-value smaller than 0.01, indicating that there is significant autocorrelation in the residuals up to the 1%-level.

Figure 4: Actual, fitted and residuals

This figure shows the daily return series for the stock Shiseido Co Ltd, its fitted values and the corresponding residuals over the estimation period that runs from 03/08/2001 until 07/08/2002.



Parameter estimation

Table 3a shows the parameters estimates of our AR(1) benchmark model. This yields $\varphi_i = -0.21$ and $\sigma_i = 0.02$. Further we report the estimates' standard errors, both naive standard errors and multi-stage standard errors. The latter also takes into account the estimation uncertainty obtained from estimating (4.1). As expected the standard errors are higher in the multi-stage case.

In the copula approach we need to estimate both the marginal distribution of the error terms and the dependence structure. In the parametric case this marginal, $G_i(\cdot; n_i)$, is the CDF of a Student-t distribution with $\hat{n}_i = 9.30$ degrees of freedom. In the semi-parametric case this marginal is the EDF.

Tables 3b and 3c summarize the estimation results that belong to the dependence structure. According to the AIC values in table 3b we see that the Student-t copula has the lowest AIC value. This means that the Student-t copula is the best fitting copula among the copulas that we consider. This is also the case for the semi-parametric approach in table 3c, although the AIC values are quite different.

Further we observe the copula parameters in tables 3b and 3c. It is clear that there is quite some difference between the parametric and semi-parametric approach. This difference is caused by the use of the EDF. Because we have a relatively short estimation horizon the EDF lacks data in the tails. This causes estimation error in the EDF. That is why in the parametric case the copula dependence parameters better take into account possible observations in the tails, resulting here in slightly more negative dependence. The negative dependence between $\varepsilon_{i,t}$ and $\varepsilon_{i,t-1}$ will result in mean-reversion in the spread series.

Tables 3b and 3c also summarize different types of standard errors corresponding to the estimates α_i. First we consider naive standard errors, where the estimation error of earlier stages is ignored. Besides, we consider MSML standard errors using the theory of Patton et al. (2012) and Chen and Fan (2006). These MSML standard errors take into account the extra uncertainty from estimating the marginal distribution parameters, including $\beta_{i,j}$ from (4.1). We observe that the naive standard errors are smaller than the MSML standard errors, which is a predictable outcome. In the parametric case this difference is larger than in the semi-parametric case. This makes sense because in the semi-parametric case, as showed in subsection 4.4.4, the MSML standard errors only require adjustments for the estimation error coming from the EDF.

Table 3: Estimation output

This table shows the estimation output of the benchmark approach and the copula approaches. Table 3a shows the AR(1) estimates and standard errors for the benchmark approach. Tables 3b and 3c show the copula parameters, standard errors and corresponding AIC values for respectively the parametric and semi-parametric case. Here the AIC values are based on the copula part of likelihood (4.34). We both report naive standard errors, ignoring the estimation error in previous steps, and correct MSML standard errors using the asymptotic theory of Patton et al. (2012) and Chen and Fan (2006).

(a) Benchmark fit

$\theta_i = (\varphi_i, \sigma_i)'$	SE (naive)	SE (MSML)
(-0.21,0.02)'	(0.06,1.35e-9)'	(0.08,1.90e-9)'

(b) Copula fit (parametric)

Copula	α_i	SE (naive)	SE (MSML)	AIC
Gaussian	-0.19	0.10	0.15	-7.59
Frank	-1.91	0.58	0.71	-7.25
Student-t	(-0.26,5.09)'	(0.11,0.10)'	(0.13,0.15)'	-14.10

(c) Copula fit (semi-parametric)

Copula	α_i	SE (naive)	SE (MSML)	AIC
Gaussian	-0.17	0.06	0.07	-5.50
Frank	-1.03	0.39	0.43	-4.74
Student-t	(-0.17,4.82)'	(0.07,0.08)'	(0.07,0.09)'	-10.40

Relative pricing

After we have selected the baskets and have estimated the parameters of the residual models, we can extract the relative pricing using the spread series $S_{i,t}$. Figures 5a and 5b show the in-sample movements of $S_{i,t}$, out-of-sample movements of $S_{i,t}$ and the corresponding bounds D_i and $-D_i$.

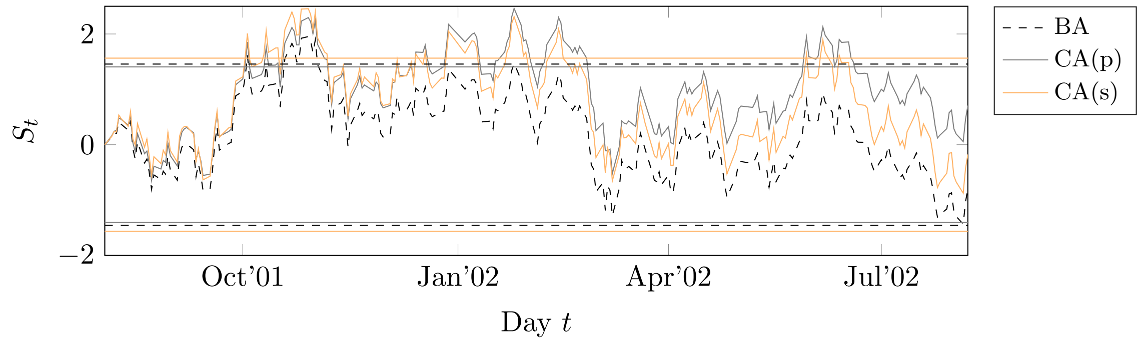
In this example the Student-t copula with around five degrees of freedom is the best fitting copula. So the benchmark and the copula approaches do not only differ by their marginals. Since the estimated degrees of freedom for the marginal t-distribution is quite high ($n_i = 9.30$) the t-distribution does not have fat tails and the EDF forms a good approximation. This is reflected by the roughly similar behavior of the spread series of the copula approaches.

Eventually we base our investment strategy on the out-of-sample spread movements. We form this spread by holding the dependence structure, estimated parameters and bounds constant. We transform new incoming return observations into our mispricing indices and argue whether current cumulative mispricing values are worth trading on. Figure 5b shows that for all approaches we observe a trading signal around the 15th of September 2002. The spreads cross the upper bound D_i so we short 1 unit of Shiseido and long the stocks in the basket with the weights as in (4.60), (4.61) and (4.62), where $N^* = 1$ and $w_{i,t}$ as in table 2. Now we hold this position until we observe a signal to close it. However, we observe for all approaches that the spread does not revert back. This means that we hold the positions until the end of the trading period. Then we reverse the positions with the initial weights corrected for the returns that we have observed over the holding period. Trading this pair will cause disastrous losses. This shows us the importance of not trading every stock. From the in-sample dynamics we already see that this stock has relatively weak mean-reversion in its spread dynamics and thus is unattractive for a potential trade.

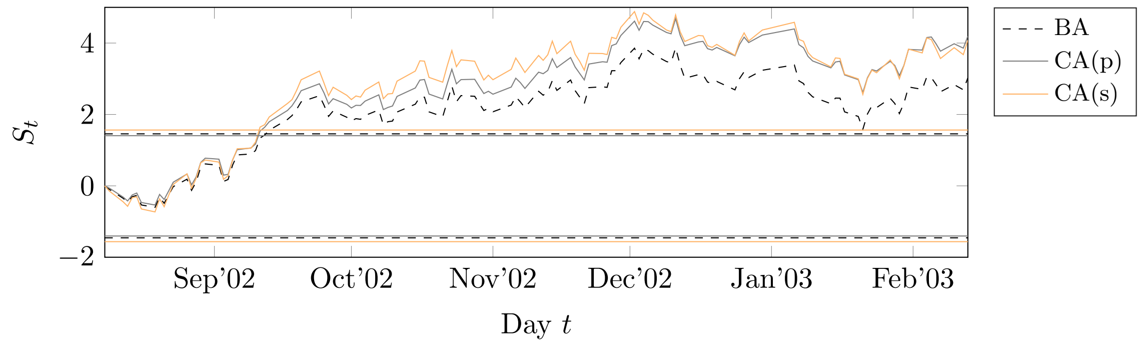
Figure 5: Illustration of Shiseido: $S_{i,t}$ and bounds

Spread dynamics $S_{i,t}$ and its bounds $(D_i, -D_i)$ for the 250-day estimation period and the 125-day trading period for the stock of Shiseido Co Ltd. The spreads are constructed using the benchmark approach (BA), the parametric copula approach (CA(p)) and the semiparametric copula approach (CA(s)).

(a) Estimation period



(b) Trading period



5.3 Copula distribution

Theoretically the benchmark and the copula models are encompassing models, so under some conditions the copula models do not have a lot of added benefit. Therefore we now zoom in on the copula specification. Here we consider all stocks and the whole sample period that consists of 15 periods of estimation and trading.

The copula distribution is shown in figures 6a and 6b for respectively the parametric and semi-parametric model. In almost all estimation periods the Student-t copula has the best fit to most of the stocks. It captures on average the dependence structure of 50% of the stocks. The Frank copula and the Gaussian copula follow up capturing around 30% and 20% of the stocks, indicating that the majority of stocks has a dependency structure in their residual process that can not be captured by our benchmark model.

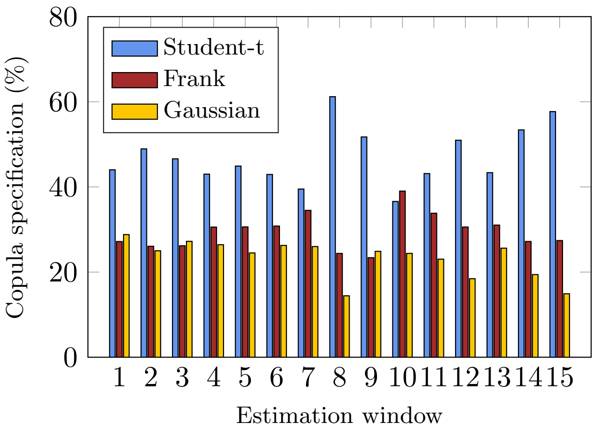
During the trading periods we do not follow all stocks. Therefore we also discuss the copula specifications of only the selected stocks. We do this for both the parametric and semi-parametric approach, because the selected stocks may differ when differences arise in the in-sample spread dynamics. The parametric case in figure 6c shows that the order of importance among copulas remains similar after selecting. However, different from the in-sample case the Student-t copula does not consistently dominate the other copulas. The Gaussian copula dominates in the first few estimation windows, whereas the Student-t and the Frank copula are more important in later estimation windows. For the semi-parametric case in figure 6d the copula distributions does not remain similar to the parametric case. In later estimation windows the Student-t copula is less important, whereas the Gaussian copula is relatively more important.

For large degrees of freedom the parametric copula model with a Student-t copula converges to our benchmark model. That is why we further investigate the degrees of freedom that correspond to figures 6c and 6d. Tables 4a and 4b show that the degrees of freedom of the copula are on average 4.87. It makes sense that those degrees of freedom are not too large, since the AIC criterion would have preferred the Gaussian copula otherwise. For the parametric copula model we impose a Student-t distribution for the marginals. Table 4a shows that the corresponding degrees of freedom are on average 17.05.

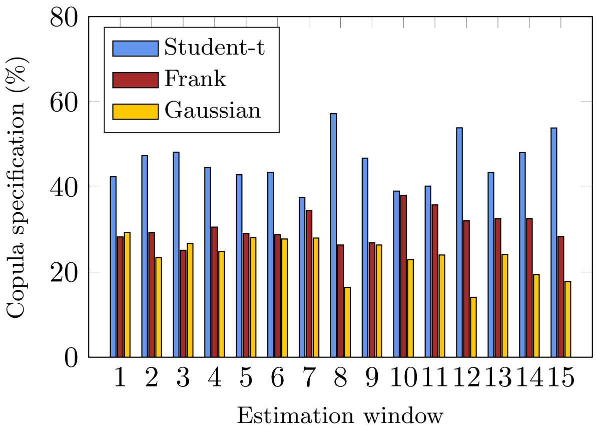
Figure 6: Copula specification over time

This figure shows the copula specifications over all estimation periods $s \in 1, \ldots, 15$. The bars indicate the fraction of stocks with a certain dependence structure (Student-t, Frank or Gaussian). The copulas are selected based on their AIC criterion and estimated either parametrically or semi-parametrically. Figures 6a and 6b show the copula specifications considering all stocks. Figures 6c and 6d show the copula specifications after selecting the top k mean-reverting stocks.

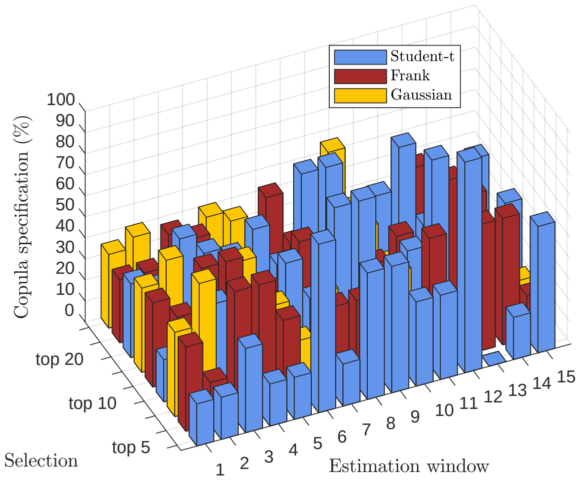
(a) Parametric estimation



(b) Semi-parametric estimation



(c) Selection parametric



(d) Selection semi-parametric

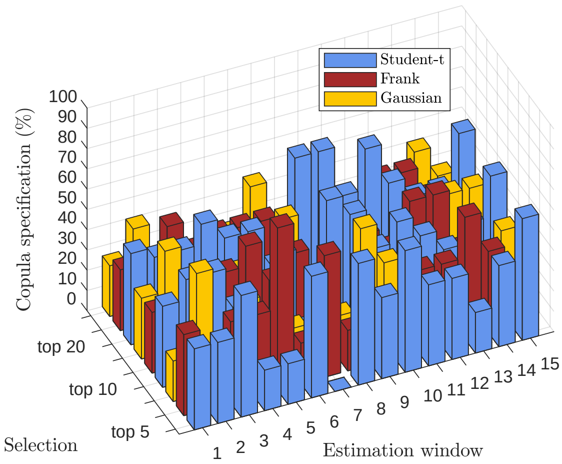


Table 4: Degrees of freedom

This figure shows the harmonic average degrees of freedom (dof) for stocks that have the best fit to the Student-t copula, i.e. $1/m_i$, where $m_i = \mathbb{E}(1/n_i)$. These averages only include the selected stocks. Figure 4a shows the parametric case with the dof for both the copula and the marginal distribution. Figure 4b shows the semi-parametric case with the dof for the copula.

(a) Copula fit (parametric)

Selection	Copula	Marginal
Top 5	5.28	16.45
Top 10	5.26	17.11
Top 20	5.49	17.58

(b) Copula fit (semi-parametric)

Selection	Copula
Top 5	4.67
Top 10	4.91
Top 20	5.02

5.4 Strategy performance

Table 5 summarizes the descriptive statistics of the annualized excess returns for all approaches. Both the benchmark approach and the copula approaches do not perform well: the portfolios have negative annualized average excess returns, irrespective of how we estimate the marginal distributions or how many stocks we select. We observe huge differences between the three approaches. A few of them are noteworthy.

First of all, we observe differences between the copula approaches. The performance of the semi-parametric approach depends on two counteracting effects. It either benefits from the fact that the copula parameters do not depend on the marginals' parameters or it suffers from the estimation error in the EDF. These effects cause differences in the spread movements of both approaches. Since we select stocks based on their in-sample spread movements different stocks are selected for both approaches, causing differences in their portfolio returns. In all cases the semi-parametric approach performs worse.

Notably is the medians of zero in table 5 that we obtain for the committed capital return streams. Our strategies need a few days before observing the first trading signals at the beginning of a trading period. This motivates why it is worth reporting the returns on employed capital streams that disregards these trading days. Here it means that for the employed capital streams the average excess returns are more negative for most approaches.

Further table 5 shows that some portfolio distributions have strong leptokurtic behavior. During most trading days we hold a portfolio with a large number of pairs, especially for the top 20 portfolios. This causes that the portfolio returns are smaller, but also more robust. However, during some rare days we have a relatively small number of pairs in our portfolio. If such trading days go together with large variances in the pairs' returns, a large part of the variance in the portfolio returns is caused by these rare events.

These rare events, or minimum and maximum values, are still modest when taking into account that this data set contains more than 10 years of data covering highly volatile states. Although the strategy is market-neutral we still have exposure to idiosyncratic risk for the individual pair constituents. The modest values for the maximum and minimum are explained by the fact that the idiosyncratic variability in individual stock prices is to some extent diversified by holding the basket of stocks. A similar reasoning explains why the standard deviations are quite modest. Most portfolio returns obtain their minimum and maximum value in September 2008, which was a turbulent month after the collapse of Lehman Brothers on September 15.

Table 6 shows the annualized risk-characteristics by means of the Sharpe ratio and the portfolios' tail-risk behavior by means of several drawdown measures (MDD,CR,BR). We base these measures on the committed capital return streams which is more conservative and therefore adopted here. From the Sharpe ratios we conclude that the parametric copula approach where we trade the top 5 pairs based on their mean reversion is the 'best' performing strategy. Similarly the Calmar ratio and the Burk ratio are highest for this portfolio.

All in all we observe that the Japanese stock market is quite efficient in our dimensions. Taking into account nonlinear associations between residuals does not improve a lot on finding better arbitrage opportunities. In the next section we further zoom in on the strategies by looking at the trading characteristics and the portfolios' composition.

Table 5: Excess return distribution

Summary statistics of the annual excess returns on portfolios of pairs between August 8th 2002 and April 1st 2010. The portfolio of pairs consists of stocks paired with a basket of comoving stocks and is constructed by means of the benchmark approach, the parametric copula approach or the semi-parametric copula approach. The table depicts both committed and employed capital return streams. The top k portfolios include the top ranked pairs with maximum number of zero-crossings in their in-sample spread dynamics. The t-statistics in this table are calculated with Newey-West (NW) standard errors.

COMMMITTED CAPITAL	Benchmark			Copula (parametric)			Copula (semi-parametric)			
	top 5	top 10	top 20	top 5	top 10	top 20	top 5	top 10	top 20	Nikkei
Average excess return (%)	-0.67	-6	-4.49	0.57	-2.5	-3.26	-7.5	-7	-5.47	1.80
t-statistic (NW)	-0.25	-2.98	-2.25	0.21	-1.29	-1.59	-2.61	-2.88	-2.87	0.21
Distribution										
Median (%)	0	0	0	0	0	0	0	0	0	12.74
Standard deviation (%)	7.64	6.09	6.16	7.67	5.94	5.97	8.8	7.1	5.6	25.36
Skewness	1.42	-0.36	2.84	0.61	0.34	3.24	0.01	-0.14	-0.25	-0.44
Kurtosis	19.62	8.83	60.35	11.19	6.87	68.23	7.83	9.99	8.83	11.04
Minimum (%)	-2.47	-3.01	-3.17	-2.7	-1.65	-3.17	-3.29	-3.69	-2.17	-12.11
Maximum (%)	5.71	2.56	6.76	4.46	2.1	6.76	3.52	2.6	2.13	13.23
Observations < 0 (%)	44.05	47.63	48.53	44.43	46.45	48.11	47.25	49.12	48.53	48.11

EMPLOYED CAPITAL	Benchmark			Copula (parametric)			Copula (semi-parametric)		
	top 5	top 10	top 20	top 5	top 10	top 20	top 5	top 10	top 20
Average excess return (%)	-0.76	-6.59	-4.89	0.65	-2.77	-3.54	-8.49	-7.72	-5.94
t-statistic (NW)	-0.25	-2.98	-2.27	0.21	-1.29	-1.59	-2.59	-2.86	-2.83
Distribution									
Median (%)	0.03	-4.63	-4.44	-0.11	-4.34	-4.22	-8.23	-7.84	-4.53
Standard deviation (%)	8.13	6.38	6.43	8.15	6.26	6.21	9.37	7.46	5.83
Skewness	1.33	-0.33	2.73	0.57	0.33	3.12	0.03	-0.11	-0.22
Kurtosis	17.29	8.05	55.49	9.92	6.2	62.93	6.92	9.06	8.13
Minimum (%)	-2.47	-3.01	-3.17	-2.7	-1.65	-3.17	-3.29	-3.69	-2.17
Maximum (%)	5.71	2.56	6.76	4.46	2.1	6.76	3.52	2.6	2.13
Observations < 0 (%)	50	52.25	52.85	50.12	51.54	52.2	53.53	54.18	52.72

Table 6: Performance measures

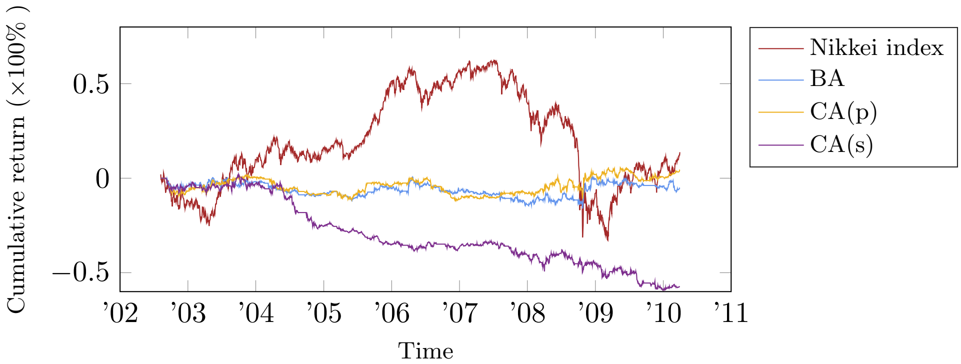
This table shows four performance measures per strategy: annualized Sharpe ratio (SR), maximum drawdown (MDD), Calmar ratio (CR) and the Burk ratio (BR). The measures are based on committed capital return streams evaluated over a trading window.

	Benchmark			Copula (parametric)			Copula (semi-parametric)		
	top 5	top 10	top 20	top 5	top 10	top 20	top 5	top 10	top 20
μ (%)	-0.67	-6	-4.49	0.57	-2.5	-2.5	-7.5	-7	-5.47
SR	-0.09	-0.99	-0.73	0.07	-0.42	-0.55	-0.85	-0.99	-0.98
MDD (%)	3.12	2.40	2.49	2.97	2.14	2.33	3.09	2.66	2.08
CR	-0.47	-2.33	-2.50	0.02	-1.35	-2.44	-2.44	-2.83	-3.05
BR	-0.27	-1.25	-1.24	0.02	-0.61	-1.08	-1.18	-1.44	-1.52

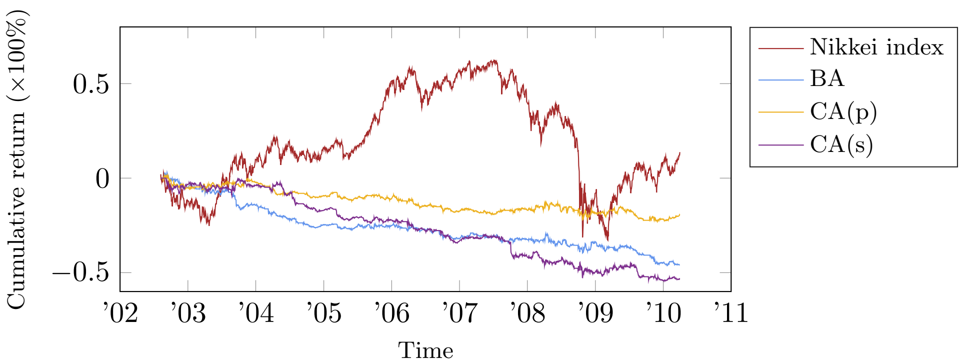
Figure 7: Cumulative returns

The graphs below show the cumulative returns (×100%) over time for different strategies over the period August 8th 2002 until April 1st 2010. The portfolio returns are based on the benchmark appraoch (BA), the parametric copula approach (CA(p)) and the semi-parametric copula approach (CA(s)).

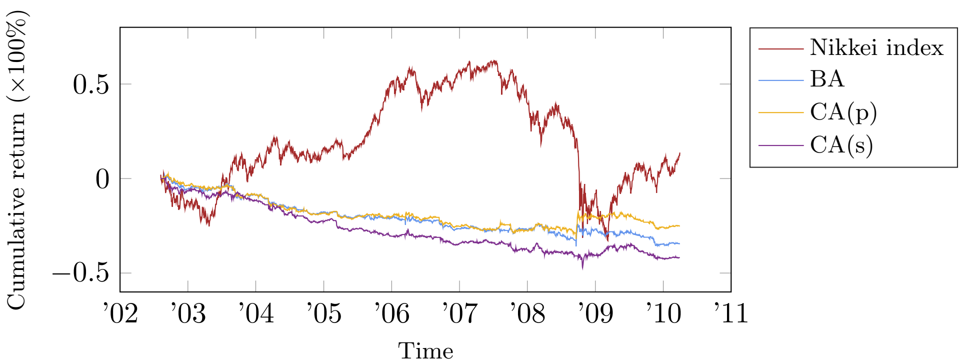
(a) Top 5



(b) Top 10



(c) Top 20



5.5 Trading statistics

Table 7 summarizes the trading statistics and the composition of the pairs trading portfolios. This shows us more about how often we trade a typical pair, its average holding period and its characteristics.

To begin with, table 7 shows the average number of pairs traded. It becomes clear that on average 89-95% of the selected pairs open during the 125-day trading periods, and on average more than once. We categorize these trades into four categories: 1. A fully converged trade; 2. A partly converged trade; 3. A partly diverged trade; 4. A fully diverged trade. The first category consists of trades where the spread returns to zero. The second category consists of trades for which the spread only reverts toward zero, without crossing it. The third category consists of trades for which the spread series diverge, but does not cross the stop-loss bound. The last category consists of trades for which we do trigger this stop-loss. Trades in the first categories can be considered as profitable trades, whereas trades in the bottom categories can be considered as unprofitable trades. Over all strategies the first two categories capture on average around 45%. So the majority of trades that we open does not converge, causing huge losses.

The last lines of the upper panel of table 7 show the average time that pairs are open in days and its standard deviation. We observe that we hold a position on average between 55 and 71 days. These averages contain all types of trades. Trades that fully converge generally have a shorter holding period than other type of trades. This is because it mostly takes more time to reach either the stop-loss bounds or the end of a trading period.

All in all, we see that the generalization to also account for nonlinear associations does not necessarily lead to more trading opportunities. We observe that neither the benchmark nor the copula approach generates more trading signals. However, we observe that the copula approaches have a higher rate of convergence indicating that we observe 'better' trading signals.

The lower panel of table 7 summarizes the pairs portfolio composition. First we consider the market capitalization of a stock in our portfolios. The market capitalization tells us the average size of a stock that is mispriced relative to its basket. We observe that the average stock in the top 5, top 10 and top 20 pairs belongs to the fifth and sixth deciles.

Now we consider the average sector weights of our portfolios. There are large differences in the number of stocks assigned to an industry group. Therefore we include the sectors' contribution to the Nikkei index in parenthesis. We observe that, on average, the sector weights in the top 5, top 10 and top 20 portfolios correspond to the sectors' share in the Nikkei index. We observe the largest differences for TECHH and COMPR for which the portfolio weights are consistently higher than their sectors' share and, *vice versa*, for MTRLS , CONDA, MEDIA, INSUR for which the portfolio weights are consistently lower then their sectors' share.

Table 7: Trading statistics and portfolio composition

This table shows the trading statistics and portfolio composition during the 125-day trading periods. The statistics are calculated as an average over 15 blocks of trading periods that range between August 8th 2002 and April 1st 2010. I categorize the total number of trades per strategy into four categories: 1) converged trades; 2) partly converged trades; 3) partly diverged trades; 4) diverged trades. Further I categorize sectors based on their 4-digit GICS classification and indicate them by their abbreviations.

	Benchmark			Copula (parametric)			Copula (semi-parametric)		
	top 5	top 10	top 20	top 5	top 10	top 20	top 5	top 10	top 20
Trading statistics									
Avg. number of pairs traded	4.73	9.47	16.6	4.47	9.07	16.67	4.8	9.27	18.73
Avg. number of trades per pair	1.39	1.35	1.28	1.26	1.25	1.29	1.53	1.38	1.33
Type of trade (%)									
converged	34.34	29.17	27.5	34.12	30.59	32.72	35.78	32.63	31.45
partly converged	14.14	14.58	19.06	16.47	14.71	16.05	11.01	12.11	15.59
partly diverged	9.09	9.90	8.13	9.41	7.65	6.79	14.68	11.05	7.80
diverged	42.42	46.35	45.31	40.00	47.06	44.44	38.53	44.21	45.16
Avg. time pairs are open in days	62.94	61.69	61.3	70.96	65.07	62.83	55.92	59.43	61.62
Std. of time open in days	14.06	11.97	8.29	14.39	10.32	7.30	17.37	10.45	8.77
Pairs portfolio composition									
Avg. size decile of stock	5.69	5.93	5.67	6.01	6.11	5.84	5.82	5.57	5.56
Avg. sector weights (%)									
ENRGY (0.67)	0	2.00	1.33	1.33	2.00	1.00	0	0.67	0.67
MTRLS (15.68)	12.00	13.33	13.67	13.33	12.00	13.00	9.33	13.33	13.00
CAPGD (22.89)	21.33	18.67	20.00	18.67	19.33	20	20.00	21.33	23.33
COMPR (1.51)	5.33	3.33	2.33	4.00	2.67	2.33	2.67	3.33	2.67
TRANS (7.44)	8.00	7.33	7.67	12.00	8.67	9.00	4.00	3.33	5.00
AUTOC (6.53)	5.33	7.33	7.67	5.33	5.33	6.67	6.67	6.00	8.33
CONDA (4.46)	4.00	3.33	4.00	2.67	3.33	4.00	4.00	5.33	3.67
CONSV (0.50)	1.33	1.33	1.00	1.33	1.33	1.00	2.67	2.00	1.00
MEDIA (0.97)	0	0	0.33	0	0	0.67	0	0	0
RTLNG (2.14)	1.33	1.33	2.00	2.67	2.00	2.00	1.33	0.67	2.33
FSRET (1.20)	1.33	1.33	1.67	0	0.67	1.67	5.33	4.00	2.33
FBVTA (5.46)	6.67	5.33	6.33	8.00	6.67	5.67	4.00	7.33	5.00
HPROD (1.01)	1.33	0.67	0.33	1.33	0.67	0.67	0	0.67	0.67
HCEQS (1.01)	0	0.67	1.33	1.33	0.67	2.00	0	0	0.67
PHBLS (3.72)	5.33	3.33	4.00	2.67	2.00	4.00	8.00	6.00	5.00
BANKS (3.28)	1.33	2.00	3.00	4.00	4.67	3.33	4.00	3.33	4.00
DIVFN (2.01)	0	4.00	2.00	0	2.67	2.33	4.00	2.00	2.00
INSUR (0.73)	0	0	0.33	0	0	0.67	0	0	0
SFTSV (2.58)	4.00	2.00	2.33	2.67	2.00	1.33	2.67	4.00	3.00
TECHH (7.54)	9.33	10.67	9.33	9.33	13.33	11.00	10.67	9.33	9.00
SMCEQ (1.71)	2.67	2.67	1.67	2.67	2.00	1.33	2.67	1.33	2.00
TELSV (1.91)	0	2.00	1.00	2.67	1.33	1.33	5.33	2.67	2.00
UTLTS (2.51)	4.00	3.33	2.67	4.00	2.67	2.00	0	0	1.33
RLEST (2.51)	5.33	4.00	4.00	0	4.00	3.00	2.67	3.33	3.00

5.6 Pairs trading by industry group

In this subsection we relate the performance of the strategies to the case where we hedge stocks with their industry group indices. As shown in subsection 5.1 the share of industry peers in the basket is relatively large. However the baskets are, on average, larger in size than most industry groups. This makes that there are differences between the baskets and the industry group indices. The main question here is whether the sector indices form a 'better' pair, which possibly improves the performance.

To form the pairs in this industry approach we do not use elastic net estimation. Here a simple regression model where we regress stock returns on their industry group returns suffice, where the industry returns are calculated as capitalization-weighted returns. From the regression coefficient that we obtain for every stock i, say β_i, we long or short stock i and hedge the positions by β_i units in its corresponding industry group. The steps for the error modelling, parameter estimation and signal generation remain the same.

Table 8 summarizes the excess returns of the portfolio of pairs for the industry approach. Similar to the basket approach we still have negative average annualized returns. The portfolios' standard deviations for the industry approach are significantly higher compared to the basket approach. Besides, the minimum and maximum values are more extreme. However, based on the Sharpe ratios in table 6 and 9 neither the basket approach nor the industry approach consistently performs better. This also holds in terms of the drawdown measures. Compared to the basket approach we observe larger maximum drawdowns for the industry approach. However, the Burke ratios and the Calmar ratios indicate that neither the basket approach nor the industry approach performs consistently better.

Table 8: Industry group pairs trading: excess return distribution

Summary statistics of the annual excess returns on portfolios of pairs between August 8th 2002 and April 1st 2010. The portfolio of pairs consists of stocks paired with its industry group index and is constructed by means of the benchmark approach or the parametric or semiparametric copula approach. The table depicts both committed and employed capital return streams. The top k portfolios include the top ranked pairs with maximum number of zero-crossings. The t-statistics in this table are calculated with Newey-West (NW) standard errors.

COMMMITTED CAPITAL

	Benchmark			Copula (parametric)			Copula (semi-parametric)			
	top 5	top 10	top 20	top 5	top 10	top 20	top 5	top 10	top 20	Nikkei
Average excess return (%)	-2.38	-1.37	-2.95	-3.67	-4.75	-6.26	-4.39	-5.78	-7.36	1.80
t-statistic (NW)	-0.72	-0.5	-0.99	-0.95	-1.55	-2.01	-1.2	-1.87	-2.74	0.21
Distribution										
Median (%)	0	0	0	0	0	0	0	0	0	12.74
Standard deviation (%)	10.3	8.25	8.94	11.42	9.07	9.34	11.37	9.72	8.31	25.36
Skewness	-0.46	-0.12	2.96	0.11	-0.32	2.56	1.01	1.86	3.35	-0.44
Kurtosis	15.56	6.25	74.03	10.3	7.63	62.98	36.62	52.16	91.2	11.04
Minimum (%)	-7.26	-2.8	-4.62	-4.41	-4.08	-4.62	-6.23	-4.88	-4.62	-12.11
Maximum (%)	5.19	2.23	10.51	6.06	3.01	10.51	10.51	10.51	10.51	13.23
Observations < 0 (%)	43.79	44.85	46.08	45.97	45.71	46.72	45.6	47.47	49.33	48.11

EMPLOYED CAPITAL

	Benchmark			Copula (parametric)			Copula (semi-parametric)		
	top 5	top 10	top 20	top 5	top 10	top 20	top 5	top 10	top 20
Average excess return (%)	-2.79	-1.53	-3.26	-4.12	-5.29	-6.86	-4.9	-6.31	-7.96
t-statistic (NW)	-0.72	-0.5	-0.99	-0.94	-1.54	-2.01	-1.21	-1.88	-2.73
Distribution									
Median (%)	-5.28	-0.87	-2.33	-5.9	-2.03	-3.21	-3.36	-5.74	-5.62
Standard deviation (%)	11.16	8.73	9.4	12.11	9.57	9.78	12.01	10.16	8.64
Skewness	-0.42	-0.11	2.82	0.12	-0.29	2.45	0.97	1.79	3.23
Kurtosis	13.24	5.58	67	9.15	6.86	57.5	32.81	47.78	84.49
Minimum (%)	-7.26	-2.8	-4.62	-4.41	-4.08	-4.62	-6.23	-4.88	-4.62
Maximum (%)	5.19	2.23	10.51	6.06	3.01	10.51	10.51	10.51	10.51
Observations < 0 (%)	51.44	50.3	50.94	51.74	50.86	51.23	50.92	51.86	53.31

Table 9: Industry group pairs trading: performance measures

This table shows four performance measures per strategy: annualized Sharpe ratio (SR), maximum drawdown (MDD), Calmar ratio (CR) and the Burk ratio (BR). The measures are based on committed capital return streams evaluated over a trading window.

	Benchmark			Copula (parametric)			Copula (semi-parametric)		
	top 5	top 10	top 20	top 5	top 10	top 20	top 5	top 10	top 20
μ (%)	-2.38	-1.37	-2.95	-3.67	-4.75	-4.75	-4.39	-5.78	-7.36
SR	-0.23	-0.17	-0.33	-0.32	-0.52	-0.67	-0.39	-0.59	-0.89
MDD (%)	4.08	3.03	3.21	4.40	3.16	3.31	4.29	3.62	3.11
CR	-0.88	-0.63	-1.46	-0.91	-1.60	-2.86	-1.18	-1.69	-3.04
BR	-0.44	-0.34	-0.72	-0.53	-0.76	-1.20	-0.60	-0.91	-1.45

5.7 Robustness check: back-testing input parameters

Our strategies depend on various input parameters, such as the trading bounds and the observation period of a pair. Here a s-day observation period means that we only take into account the cumulative mispricing of a pair of the last s days (4.58). In this subsection we verify the sensitivity of the results to those input parameters. Currently the observation period for pairs is 125 days. It is doubtful whether we include too much noise here: as shown in subsection 5.5 a large fraction of trades that we open diverges. So even though our strategies can show positive returns in its converged trades, the high proportion of diverged trades countervails a portion of these profits. Therefore we verify the results using a shorter observation periods ranging around 60 days, which is used in Avellaneda and Lee (2010). Further we consider a range of trading bounds such that we open and close positions earlier, which can be considered as more conservative trading bounds.

Table 10 tabulates annualized Sharpe ratios for our strategies with different input parameters. We observe that the results are quite sensitive to the choice of the input parameters. This is partially due to differences in the set of stocks selected for trading. This set contains at maximum 10% of the stocks in the universe, so even small changes in the input parameters heavily changes the stocks selected for trading.

Table 10 shows that the average performance is improved by considering shorter observation periods. This makes sense because recent mispricing values are more relevant than mispricing values from a longer time ago. Especially the parametric copula approach benefits from the use of shorter observation periods. This effect is less clear for the benchmark approach, but the Sharpe ratios for the 40-day and 50-day observation period are on average higher than for the 125-day specification. The semi-parametric copula approach only results in higher average Sharpe ratios for the 50- and 60-day observation periods. Further we note that the Sharpe ratios that correspond to our initial input parameters (open = D, close = 0, observation period = 125d.), are slightly higher in all cases compared to the Sharpe ratios in table 6. This is because in this case also the in-sample spread dynamics only depend on the last 125 days, which may result in a different set of stocks selected for trading.

We observe that the strategies' performance is on average slightly improved by opening trades earlier than the initial case of D and deteriorated by closing trades earlier than the case of 0. This especially holds for the 40-, 50- and 60-day observation periods. For the 125-day observation period we can, instead, on average do better by closing positions earlier. On average the parametric copula approach is the best performing approach, followed up by respectively the benchmark approach and the semi-parametric copula approach. These differences mainly arise for the case of the 40-day observation periods.

Table 10: Robustness check: varying input parameters

This table shows annualized Sharpe ratios for all strategies with various input parameters.

Top 5		Benchmark				Copula (parametric)				Copula (semi-parametric)			
		125 d.	60 d.	50 d.	40 d.	125 d.	60 d.	50 d.	40 d.	125 d.	60 d.	50 d.	40 d.
Trading bounds													
open = $1D$	close = $0.5D$	-0.04	0.21	0.39	-0.51	-0.55	0.56	-0.40	0.59	0.11	0.39	-0.19	0.25
	close = $0.4D$	0.05	0.57	0.42	-0.45	-0.70	0.25	0.40	0.47	0.15	0.54	-0.16	-0.04
	close = $0.2D$	0.10	0.23	0.41	-0.70	-0.13	0.15	0.65	0.78	-0.35	0.03	0.55	0.06
	close = $0D$	-0.17	-0.13	0.33	-0.54	0.10	0.12	0.62	0.46	-0.45	0.41	0.85	-0.17
open = $0.9D$	close = $0.5D$	-0.15	-0.40	-0.08	0.08	-1.02	0.08	-0.09	0.84	-0.40	-0.01	0.08	0.07
	close = $0.4D$	0.18	0.04	0.24	0.41	-0.84	-0.05	0.27	0.78	-0.35	-0.26	-0.34	0.20
	close = $0.2D$	0.01	-0.28	0.30	0.48	-0.19	0.18	0.19	1.02	-0.16	-0.06	0.30	-0.32
	close = $0D$	0.24	0.19	0.58	-0.05	0.01	0.08	0.03	0.92	0.42	0.31	0.50	-0.22
open = $0.8D$	close = $0.5D$	-0.04	0.10	0.16	-0.01	-0.20	0.31	-0.51	-0.15	0.24	0.10	0.29	-0.41
	close = $0.4D$	0.03	0.45	0.03	0.24	0.08	-0.26	0.12	0.25	0.12	0.31	0.04	-0.36
	close = $0.2D$	0.35	0.24	0.20	0.13	0.31	0.04	-0.08	0.55	0.36	0.43	0.03	-0.20
	close = $0D$	0.49	0.19	0.68	-0.05	-0.04	0.17	0.15	0.84	0.52	-0.05	0.00	-0.21
open = $0.7D$	close = $0.5D$	-0.07	-0.17	0.30	0.19	0.34	0.13	-0.23	0.13	0.41	-0.10	-0.08	-0.51
	close = $0.4D$	-0.08	-0.07	0.21	0.51	0.56	0.26	0.20	0.12	0.11	0.31	0.06	-0.55
	close = $0.2D$	0.67	-0.02	0.16	0.62	0.48	0.35	0.13	1.11	0.28	-0.02	0.20	0.10
	close = $0D$	0.25	0.18	0.60	0.43	-0.01	0.29	0.15	1.06	-0.21	0.31	0.27	0.23

Top 10		Benchmark				Copula (parametric)				Copula (semi-parametric)			
		125 d.	60 d.	50 d.	40 d.	125 d.	60 d.	50 d.	40 d.	125 d.	60 d.	50 d.	40 d.
Trading bounds													
open = $1D$	close = $0.5D$	0.13	0.37	0.23	-0.23	-0.39	0.30	0.05	0.58	0.08	0.21	0.06	0.12
	close = $0.4D$	0.04	0.23	0.22	-0.18	-0.15	-0.02	0.76	0.76	0.09	0.16	0.08	0.20
	close = $0.2D$	0.39	0.13	0.38	-0.22	0.47	0.07	0.78	0.98	-0.21	0.02	0.35	-0.14
	close = $0D$	-0.49	-0.15	0.48	0.08	0.12	0.28	0.26	0.79	-0.39	0.29	0.68	0.32
open = $0.9D$	close = $0.5D$	0.10	-0.13	-0.40	0.35	-0.90	-0.08	-0.11	0.75	-0.53	0.18	-0.01	-0.11
	close = $0.4D$	0.09	-0.01	0.01	0.38	-0.88	0.26	0.49	0.89	-0.45	0.08	-0.16	-0.22
	close = $0.2D$	-0.03	-0.07	0.39	0.40	0.08	0.35	0.51	1.02	-0.58	-0.15	0.20	-0.29
	close = $0D$	0.09	-0.14	0.49	0.09	-0.05	0.31	0.51	1.10	0.13	0.22	0.20	-0.12
open = $0.8D$	close = $0.5D$	0.03	-0.10	-0.11	0.16	-0.05	0.00	-0.29	0.53	0.00	-0.17	0.30	-0.02
	close = $0.4D$	-0.05	0.09	-0.19	0.14	0.14	0.01	0.24	0.53	0.03	0.15	-0.30	-0.43
	close = $0.2D$	0.35	0.06	0.11	0.39	-0.01	-0.10	0.22	0.92	0.08	0.40	-0.06	0.19
	close = $0D$	0.35	-0.09	0.41	0.18	-0.25	0.32	0.29	1.17	0.30	0.13	-0.25	0.10
open = $0.7D$	close = $0.5D$	0.00	-0.07	-0.01	0.27	0.47	0.23	-0.28	0.17	0.71	-0.13	-0.14	-0.38
	close = $0.4D$	-0.21	0.21	0.23	0.69	0.47	0.25	0.45	0.45	0.33	0.19	0.06	-0.53
	close = $0.2D$	-0.09	-0.12	0.17	0.57	0.10	0.34	-0.02	1.02	0.06	0.35	-0.03	0.38
	close = $0D$	-0.22	0.09	0.38	0.34	-0.14	0.30	-0.12	1.14	-0.33	0.65	0.37	0.12

Top 20		Benchmark				Copula (parametric)				Copula (semi-parametric)			
		125 d.	60 d.	50 d.	40 d.	125 d.	60 d.	50 d.	40 d.	125 d.	60 d.	50 d.	40 d.
Trading bounds													
open = $1D$	close = $0.5D$	0.30	0.06	0.22	0.39	-0.51	-0.07	0.21	0.53	0.16	0.13	0.34	-0.07
	close = $0.4D$	0.24	0.23	0.27	0.12	-0.27	0.16	0.65	0.87	0.20	0.15	0.03	-0.12
	close = $0.2D$	0.25	0.04	0.32	0.04	0.11	0.56	0.91	1.35	-0.20	0.06	0.52	-0.06
	close = $0D$	0.13	-0.07	0.27	0.14	-0.05	0.86	0.49	1.12	-0.08	-0.12	0.33	0.11
open = $0.9D$	close = $0.5D$	-0.16	0.09	-0.23	0.39	-0.86	-0.03	0.02	0.77	-0.31	0.16	0.04	0.14
	close = $0.4D$	-0.16	-0.13	-0.23	0.18	-0.41	0.50	0.42	0.88	-0.38	0.20	0.16	-0.08
	close = $0.2D$	0.29	-0.04	0.05	0.44	-0.09	0.48	0.45	1.24	-0.23	0.31	-0.06	-0.20
	close = $0D$	-0.29	0.17	0.63	0.54	0.07	0.75	0.69	1.19	0.16	0.12	-0.10	-0.18
open = $0.8D$	close = $0.5D$	0.01	-0.14	0.10	0.09	-0.05	-0.01	0.19	0.57	-0.04	-0.12	0.04	-0.04
	close = $0.4D$	-0.14	-0.07	0.16	0.07	-0.05	0.45	0.57	0.95	-0.31	-0.07	0.02	-0.20
	close = $0.2D$	-0.13	0.14	0.13	0.37	-0.15	0.27	0.23	1.16	-0.16	-0.06	0.09	0.05
	close = $0D$	-0.14	-0.15	0.27	0.30	-0.24	0.16	0.61	1.12	0.38	0.12	0.42	0.31
open = $0.7D$	close = $0.5D$	-0.10	-0.22	-0.25	0.27	0.05	0.11	-0.07	0.54	0.03	-0.09	-0.11	-0.24
	close = $0.4D$	-0.38	-0.04	-0.01	0.14	0.30	0.59	-0.01	0.81	-0.24	-0.34	-0.21	-0.44
	close = $0.2D$	-0.47	-0.04	-0.21	0.68	0.18	0.34	-0.58	1.07	-0.14	0.40	-0.07	0.34
	close = $0D$	-0.67	-0.58	0.09	0.69	0.14	0.30	-0.06	0.99	-0.03	0.30	0.44	0.29

5.8 Risk characteristics

In the previous subsection we have been able to generate a profitable pairs trading strategy under certain input parameters. Especially for the input parameters (open = $0.7D$, close = 0, observation period = 40d.) we consistently have positive Sharpe ratios over all approaches. In this subsection we investigate how these portfolio returns relate to other well-known factors. We investigate this relation by regressing the portfolio returns of our strategies on the factor returns. This provides perspective of the systemic risk exposure of the pairs portfolios. In this analysis we use the factors of Fama and French (1992) and Carhart (1997).

Table 11 summarizes the regression coefficients, its corresponding Newey-West (NW) t-statistics as well as the R-squared of each regression. At first we observe that the R^2 values are generally very low and over all cases at maximum 0.0076. So the risk factors are only able to explain a small fraction of total variability in the daily portfolio returns. Furthermore we find that the parametric copula approach generates statistically significant alphas for the top 5 and top 10 stocks selection. For the other approaches the daily risk-adjusted returns are still positive, but not significantly.

The factor loadings show that the market exposures are small and mostly insignificant implied by the market neutrality of the strategies. Further, the exposures to the other two Fama-French factors are mostly insignificant. For the momentum factor we do find significant exposure in two of the cases, but overall the exposures are not large enough to explain the pairs trading returns.

Table 11: Exposure to common risk factors

This table summarizes the risk exposure of the daily portfolio returns to the excess market portfolio (MKT-RF), the small minus big (SMB) portfolio, the high minus low (HML) portfolio and the momentum (MOM) portfolio. The exposures are calculated for the portfolios constructed with the benchmark approach, the parametric copula approach and the semi-parametric copula approach. The table shows the regression coefficients, t-statistics (NW) and the R^2 resulting from the regression.

	INTERCEPT		MKT-RF		SMB		HML		MOM		
	α ($\times 10^{-2}$)	t-stat	β_m ($\times 10^{-2}$)	t-stat	β_s ($\times 10^{-2}$)	t-stat	β_h ($\times 10^{-2}$)	t-stat	β_{mm} ($\times 10^{-2}$)	t-stat	R^2 ($\times 10^{-2}$)
Basket											
Benchmark											
top 5	0.01	1.15	0.41	1.02	1.33	1.27	1.03	1.11	-1.75	-2.00	0.33
top 10	<0.01	0.73	0.14	0.34	0.24	0.23	-0.93	-1.01	-0.54	-0.62	0.28
top 20	0.01	1.08	-0.08	-0.20	0.05	0.05	-0.59	-0.63	-0.43	-0.49	0.30
Copula (parametric)											
top 5	0.02	3.73	-1.07	-2.69	-0.49	-0.47	1.78	1.93	-0.81	-0.93	0.42
top 10	0.02	2.40	-0.27	-0.68	-0.35	-0.33	1.36	1.48	0.01	0.01	0.21
top 20	0.01	1.42	-0.14	-0.35	-0.17	-0.16	0.52	0.56	0.01	0.02	0.08
Copula (semi-parametric)											
top 5	<0.01	0.60	-0.01	-0.03	-1.13	-1.08	1.32	1.43	1.94	2.22	0.55
top 10	<0.01	0.24	-0.50	-1.26	-0.77	-0.74	0.43	0.46	1.86	2.12	0.76
top 20	<0.01	0.44	-0.23	-0.57	-0.23	-0.22	-0.13	-0.14	0.67	0.77	0.19

6 Conclusion

In this research we focus on the relative value pricing between an individual stock and a corresponding basket of stocks that should price the stocks properly. We identify these baskets with the elastic net estimator. The elastic net estimator 'selects' on average 49% of stocks that belong to the same industry group. Further we observe that the baskets are, on average, larger in size than most industry groups. This makes that the cross-industry members of the baskets are not negligible. The number of times that a specific industry group is selected across all baskets does not greatly differ. This number ranges between 22.6 times and 30.3 times. Here we find that the industry groups with the largest share in the Nikkei index are not necessarily the industries that most frequently appear in a basket.

After we perform the elastic net estimation we model the residual dynamics. We base trading decisions on the accumulation of conditional probabilities of new incoming errors exceeding their historical relation. This probability depends on how the error terms behaved marginally and the temporal dependence structure of stocks' error terms. In the benchmark approach we assume a fully Gaussian specification. As a extension we propose a second approach where we allow for a variety of dependence structures and marginals. We do this by assuming a copula relation between current error terms and its lagged version.

It is possible that the benchmark and the copula models are encompassing models. This happens when we mainly prefer the Gaussian copula or have large degrees of freedom. In such cases there is not much benefit of considering the copula models. That is why we summarize the copula distribution and the corresponding degrees of freedom. We find that the Student-t copula and the Frank copula capture the dependence of around 50% and 30% of the stocks. After selection, these copulas still capture the majority of stocks. Further we find that the degrees of freedom that belong to the Student-t copula are around 4.87, whereas for the marginal these are around 17.05.

The above-mentioned findings suggest that the copula models capture a broader range of specifications. However, we find that both the benchmark and the copula approach are unable to generate profitable trades. The parametric copula approach can improve upon the benchmark approach, but not enough to be profitable. The semi-parametric copula is the worst performing strategy. The estimation error in the EDF dominates the higher estimation quality in this semi-parametric approach. The Sharpe ratios of all different specifications range from 0.07 until -0.99.

We further zoom in on the performance by considering the trading statistics. We find that the generalization to also account for nonlinear associations does not necessarily lead to more trading opportunities. However, the copula approaches do have a higher rate of convergence indicating that we observe 'better' trading signals. Trading opportunities mostly appear for companies assigned to TECHH and COMPR, whereas for MTRLS, CONDA, MEDIA and INSUR we do not observe a lot of signals. Over all strategies we find a large fraction of trades does not converge, which mainly results in unprofitable trades.

In hope of improving the performance we consider trading stocks versus their sector member, rather than its basket. We find that hedging the positions by the baskets lowers the portfolios' standard deviations significantly, but does not improve the Sharpe ratios consistently.

As a second way of improving the performance we consider different trading bounds and observation periods. We observe that observing pairs for a shorter period of time and opening trades earlier generally improves the performance. We find Sharpe ratios up to 1.35 for the parametric copula approach and 0.69 for the benchmark approach. Further we find that the parametric copula approach dominates both the semi-parametric and the benchmark approach.

At last we investigate how the pairs trading returns using our optimized input parameters relate to the factor returns of Fama and French (1992) and Carhart (1997). We find that size, value and momentum are unable to explain our portfolio returns: the obtained R^2 are very low and at most 0.0076.

7 Discussion

A first point of discussion here is the time frame that we use in this research. For later research it would be interesting to see whether pairs trading profits can be found using a more recent data set and in markets other than Japan. On top of that, the proposed strategies are applied based on a daily trading frequency. One can also apply the proposed strategies to a high-frequency framework. The methodology that we use is not dependent on the time frame used. Moreover, the shorter time-interval might give rise to finding more arbitrage opportunities.

A next potential shortcoming is the use of the multi-stage maximum likelihood (MSML) that loses efficiency by separately estimating the parameters that correspond to the marginal distribution of the residuals and the parameters that correspond to the dependence structure. This estimator eases the estimation procedure, however this also means that we have additional uncertainty in the trading signals. Chen et al. (2006) propose an one-stage variant of this estimator that attains full efficiency, but also is more difficult to implement.

To deal with multicollinearity in our application we introduced the elastic net estimator. Despite its robustness for multicollinearity, the estimates are still point estimates and we do not know much about its uncertainty. To better deal with this problem one can consider the Bayesian Lasso introduced by Park and Casella (2008). The Bayesian Lasso allows for this model uncertainty by averaging over different specifications which further robustifies our approach since we do not rely on a single model. In future research such estimates can be considered to identify the baskets. However, the Bayesian Lasso does not deliver a sparse solution so the structure is more complex and thus harder to interpret.

For future research one can take into account the results of Papanicolaou and Yeo (2016). They show that one can improve the performance by only allowing for trading activity when the goodness-of-fit of the estimation for trading signals is sufficiently high. In our framework this can be implemented by using the R^2 obtained from the residual process estimation as a extra selection criterion. In appendix A we show a small implementation of this. One can further investigate different selection methods, varying not only the selection criteria but also the number of stocks selected for trading. Increasing the latter will further robistify the strategies' performance.

As explained in subsection 4.6.1, $S_{i,t}$ (temporarily) behaves as a purely random walk when there is no time-series correlation between $e_{i,t} = (MI_{i,t} - 0.5)$ and $S_{i,t-1}$. We have seen that this happens for the majority of trades when $S_{i,t}$ contains the cumulative mispricing of the last 125 days. Apparently we include too much noise here, which causes that $S_{i,t-1}$ is not strongly correlated with current mispricings. As shown in subsection 5.7 we can improve on this by considering shorter observation periods. In future research one can further discuss the sensitivity for a wider range of observation periods. Alternatively, one can generate trading signals from the mispricing indices $MI_{i,t}$ directly and open/exit trades when the conditional probabilities are well in the tail regions of their distribution function. In this approach we do not rely on the observation periods which also solves our issue. However the disadvantage of this approach is the complete loss of time-structure and we cannot asses how pairs trade subsequent to a signal.

Further, in future research one might consider different trading bounds. The spread series $S_{i,t}$ is by construction not stationary. That is why the choice of $D = 2\sigma_s$ is misleading since the variance of $S_{i,t}$ changes over time. In future research one can improve on this by determining the trading thresholds using the standard deviation across $MI_{i,t}$, which is a stationary series.

Last, to put the results into perspective one can investigate whether the pairs trading profits found in subsection 5.7 do survive transaction costs. This plays a major role in determining whether the strategies are trade-able in practice.

References

Avellaneda, M. and Lee, J.-H. (2010). Statistical arbitrage in the us equities market. *Quantitative Finance*, 10(7):761–782.

Bertram, W. K. (2010). Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical Mechanics and its Applications*, 389(11):2234–2243.

Burke, G. (1994). A sharper sharpe ratio. *Futures*, 23(3):56.

Caldeira, J. and Moura, G. V. (2013). Selection of a portfolio of pairs based on cointegration: a statistical arbitrage strategy.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Chen, H. J., Chen, S. J., and Li, F. (2012). Empirical investigation of an equity pairs trading strategy.

Chen, X. and Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, 130(2):307–335.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101(475):1228–1240.

Cummins, M. and Bucca, A. (2012). Quantitative spread trading on crude oil and refined products markets. *Quantitative Finance*, 12(12):1857–1875.

Do, B., Faff, R., and Hamza, K. (2006). A new approach to modeling and estimation for pairs trading. In *Proceedings of 2006 Financial Management Association European Conference*, pages 87–99.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 301–369.

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

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

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

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, pages 543–552.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.

Harari, D. (2013). Japan's economy: from the "lost decade" to abenomics. *House of Commons Library, Standard Note SN06629. London: Oct*, 24.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.

Kendall, M. G. (1948). Rank correlation methods.

Kim, H. H. and Swanson, N. R. (2013). Mining big data using parsimonious factor and shrinkage methods.

Krauss, C. and Stübinger, J. (2017). Non-linear dependence modelling with bivariate copulas: statistical arbitrage pairs trading on the s&p 100. *Applied Economics*, pages 1–18.

Landgraf, N. (2016). High-frequency copula-based pairs trading on u.s. goldmine stocks. *Erasmus University Thesis Repository.*

Magdon-Ismail, M. and Atiya, A. F. (2004). Maximum drawdown.

Neumann, M. H., Kreiss, J.-P., et al. (1998). Regression-type inference in nonparametric autoregression. *The Annals of Statistics*, 26(4):1570–1613.

Newey, W. K. and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Papanicolaou, G. and Yeo, J. (2016). Risk control of mean-reversion time in statistical arbitrage.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Patton, A. J. et al. (2012). Copula methods for forecasting multivariate time series. *Handbook of economic forecasting*, 2:899–960.

Perlin, M. (2007). M of a kind: A multivariate approach at pairs trading.

Sharpe, W. F. (1994). The sharpe ratio. *The journal of portfolio management*, 21(1):49–58.

Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges.* Université Paris 8.

Stübinger, J., Mangold, B., and Krauß, C. (2016). Statistical arbitrage with vine copulas. Technical report, FAU Discussion Papers in Economics.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Triantafyllopoulos, K. and Montana, G. (2011). Dynamic modeling of mean-reverting spreads for statistical arbitrage. *Computational Management Science*, 8(1):23–49.

Vidyamurthy, G. (2004). *Pairs Trading: quantitative methods and analysis*, volume 217. John Wiley & Sons.

White, H. (1996). *Estimation, inference and specification analysis.* Number 22. Cambridge university press.

Xie, W., Liew, R. Q., Wu, Y., and Zou, X. (2014). Pairs trading with copulas. *The Journal of Trading*, 11(3):41–52.

Young, T. W. (1991). Calmar ratio: A smoother tool. *Futures*, 20(1):40.

Zeng, Z. and Lee, C.-G. (2014). Pairs trading: optimal thresholds and profitability. *Quantitative Finance*, 14(11):1881–1893.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

Appendix

A. Analytical results

In this section we give some derivations of results that we need during the analysis.

Analytical results copulas

At first we need the copula density to estimate $\tilde{\alpha}$ in (4.42). Secondly we derive $C(v|u)$ which we use to calculate the spread series. We derive those results for the copulas we consider in this research: Gaussian copula, Student-t copula and Frank copula.

Gaussian copula

The bivariate Gaussian copula with dependence parameter ρ_i is given by

$$C(u_{i,t}, v_{i,t}; \rho_i) = \Phi_2\left(\Phi^{-1}(u_{i,t}), \Phi^{-1}(v_{i,t}); \rho_i\right), \tag{7.1}$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal, $\Phi_2(\cdot)$ is the joint cumulative distribution function of a bivariate normal distribution with zero means and covariance matrix Σ. The copula density is given by

$$\begin{aligned} c(u_{i,t}, v_{i,t}; \rho_i) &= \frac{\partial^2}{\partial u \partial v} C(u_{i,t}, v_{i,t}; \rho_i) = \frac{\phi_2(\Phi^{-1}(u_{i,t}), \Phi^{-1}(v_{i,t}); \rho_i)}{\phi(\Phi^{-1}(u_{i,t}))\phi(\Phi^{-1}(v_{i,t}))} \\ &= \frac{1}{\sqrt{1-\rho_i^2}} \exp\left(\frac{2\rho_i \Phi^{-1}(u_{i,t})\Phi^{-1}(v_{i,t}) - \rho_i^2(\Phi^{-1}(u_{i,t})^2 \Phi^{-1}(v_{i,t})^2)}{2(1-\rho_i^2)}\right). \end{aligned} \tag{7.2}$$

The conditional bivariate copulas can be derived by taking the partial derivative of equation (7.1):

$$C(v_{i,t}|u_{i,t}) = \frac{\partial}{\partial u} C(u_{i,t}, v_{i,t}; \rho_i) = \Phi\left(\frac{\Phi^{-1}(v_{i,t}) - \rho_i \Phi^{-1}(u_{i,t})}{\sqrt{1-\rho_i^2}}\right) \tag{7.3}$$

The copula parameter $\rho_i \in [-1, 1]$ can be calibrated by $\hat{\rho}_i = \sin(\frac{\pi}{2}\hat{\tau}_i)$. Here $\hat{\tau}_i$ is the correlation rank coefficient as in Kendall (1948) which is defined as

$$\tau_i = \frac{2(N_c^i - N_d^i)}{T(T-1)}, \tag{7.4}$$

where N_c^i and N_d^i represent the number of concordant and discordant pairs $(\varepsilon_{i,t}, \varepsilon_{i,t-1})$.

Student-t copula

The bivariate Student-t copula distribution is

$$C(u_{i,t}, v_{i,t}; \rho_i, n_i) = t_{2,n_i}(t_{n_i}^{-1}(u_{i,t}), t_{n_i}^{-1}(v_{i,t})), \tag{7.5}$$

where $n_i \in (0, \infty)$ represents the degrees of freedom, t_{2,n_i} the joint cumulative Student-t distribution function, t_{n_i} the univariate Student-t distribution function and ρ_i the copula parameter. The density of the copula is defined as

$$c(u_{i,t}, v_{i,t}; \rho_i, n_i) = \frac{\partial^2}{\partial u \partial v} C(u_{i,t}, v_{i,t}; \rho_i, n_i) = \frac{f_{2,n_i}(t_{n_i}^{-1}(u_{i,t}), t_{n_i}^{-1}(v_{i,t}))}{f_{n_i}(t_{n_i}^{-1}(u_{i,t})) f_{n_i}(t_{n_i}^{-1}(v_{i,t}))}$$
$$= \frac{1}{\sqrt{1-\rho_i^2}} \frac{\Gamma(\frac{n_i+2}{2})\Gamma(\frac{n_i}{2})}{(\Gamma(\frac{n_i+1}{2}))^2} \frac{\left[\left(1+\frac{\psi_{1,i}^2}{n_i}\right)\left(1+\frac{\psi_{2,i}^2}{n_i}\right)\right]^{\frac{n_i+1}{2}}}{\left[1+\frac{1}{n_i(1-\rho_i^2)}(\psi_{1,i}^2 - 2\rho_i\psi_{1,i}\psi_{2,i} + \psi_{2,i}^2)\right]^{\frac{n_i+2}{2}}}, \tag{7.6}$$

where f_{2,n_i} is the joint density of a two-dimensional random vector from a multivariate Student-t distribution. Further f_{n_i} is the density of a univariate Student-t distribution. Further I have defined $\psi_i = (t_{n_i}(u_{i,t})^{-1}, t_{n_i}(v_{i,t})^{-1})'$ to ease the notation. The conditional bivariate Student-t copula is defined as:

$$C(v_{i,t}|u_{i,t}) = \frac{\partial}{\partial u} C(u_{i,t}, v_{i,t}; \rho_i, n_i) = t_{(n_i+1)}\left(\sqrt{\frac{n_i+1}{n_i + (t_{n_i}^{-1}(u_{i,t}))^2}}\left(\frac{t_{n_i}^{-1}(v_{i,t}) - \rho_i t_{n_i}^{-1}(u_{i,t})}{\sqrt{1-\rho_i^2}}\right)\right) \tag{7.7}$$

Similar to the case of the Gaussian copula we can calibrate $\rho_i \in [-1, 1]$ by $\hat{\rho}_i = \sin(\frac{\pi}{2}\hat{\tau}_i)$. However different in this case is the extra estimation of the degrees of freedom n_i. For given $\hat{\rho}_i$ I estimate this parameter numerically, opting for n_i that maximizes the log-likelihood in (4.34).

Frank copula

The Frank copula is a so called Archimedean copula. This family of copulas are built on a generator function $\psi(\cdot)$ and can be written as:

$$C(u_{i,t}, v_{i,t}; \theta_i) = \psi^{-1}(\psi(u_{i,t}) + \psi(v_{i,t})). \tag{7.8}$$

For the Frank copula this generator function is defined as:

$$\psi(u_{i,t}) = -\ln\left(\frac{\exp(-\theta_i u_{i,t}) - 1}{\exp(-\theta_i) - 1}\right), \tag{7.9}$$

where $\theta_i \in (-\infty, \infty) \setminus \{0\}$ is the copula parameter. So the copula and its density are defined as follows:

$$C(u_{i,t}, v_{i,t}; \theta_i) = -\frac{1}{\theta_i}\ln\left[1 + \frac{(\exp(-\theta_i u_{i,t}) - 1)(\exp(-\theta_i v_{i,t}) - 1)}{(\exp(-\theta_i) - 1)}\right] \tag{7.10}$$

$$c(u_{i,t}, v_{i,t}; \theta_i) = \frac{\partial^2}{\partial u \partial v} C(u_{i,t}, v_{i,t}; \theta_i) = \frac{-\theta_i(\exp(-\theta_i) - 1)(\exp(-\theta_i(u_{i,t} + v_{i,t})))}{((\exp(-\theta_i u_{i,t}) - 1)(\exp(-\theta_i v_{i,t}) - 1) + (\exp(-\theta_i) - 1))^2} \tag{7.11}$$

The bivariate conditional distributions are now defined as follows:

$$C(v_{i,t}|u_{i,t}) = \frac{\partial}{\partial u} C(u_{i,t}, v_{i,t}; \theta_i) = \frac{(\exp(-\theta_i u_{i,t}) - 1)(\exp(-\theta_i v_{i,t}) - 1) + (\exp(-\theta_i v_{i,t}) - 1)}{(\exp(-\theta_i u_{i,t}) - 1)(\exp(-\theta_i v_{i,t}) - 1) + (\exp(-\theta_i) - 1)} \tag{7.12}$$

For the copula parameter θ_i we do not have a closed form solution. That is why I calculate θ_i numerically, opting for θ_i that maximizes the log-likelihood function (4.34).

Transformation model

As demonstrated in equation (4.19) all the copula-based first order Markov models can be transformed in terms of an autoregression transformation model. Define $\{v_{i,t} : v_{i,t} \equiv G_i(\varepsilon_{i,t})\}$ and $\{u_{i,t} : u_{i,t} \equiv G_i(\varepsilon_{i,t-1})\}$ and denote $C_i(u_{i,t}, v_{i,t}; \alpha_i)$ as the copula in which the joint distribution of $u_{i,t}$ and $v_{i,t}$ is given. Further we define its corresponding density as $c_i(u_{i,t}, v_{i,t}; \alpha_i)$. Let $\Lambda_{1,\theta_{1(i)}}(\cdot)$ be any increasing parametric function and $\Lambda_{1,\theta_{1(i)}}(\cdot) = E\{\Lambda_{1,\theta_{1(i)}}(G_i(\varepsilon_{i,t}))|G_i(\varepsilon_{i,t-1}) = u_{i,t})\}$, then

$$\Lambda_{1,\theta_{1(i)}}(G_i(\varepsilon_{i,t})) = \Lambda_{2,\theta_{2(i)}}(G_i(\varepsilon_{i,t-1})) + \sigma_{i,\theta_{3(i)}}(G_i(\varepsilon_{i,t-1}))\zeta_{i,t}, \tag{7.13}$$

or, equivalently

$$v_{i,t} = G_i(\varepsilon_{i,t} = \Lambda^{-1}_{1,\theta_{1(i)}}(\Lambda_{2,\theta_{1(i)}}(u_{i,t}) + \sigma_{i,\theta_{3(i)}}(u_{i,t})\zeta_{i,t})), \tag{7.14}$$

where the conditional density of $\zeta_{i,t}$ given $u_{i,t}$ is

$$\begin{aligned} f_{\zeta_i|u_{i,t})}(\zeta_i) &= c_i(u_{i,t}, \Lambda^{-1}_{1,\theta_{1(i)}}(\Lambda_{2,\theta_{1(i)}}(u_{i,t}) + \sigma_{i,\theta_{3(i)}}(u_{i,t})\zeta_i))/D_i(u_{i,t}) \\ &= c_i(G_i(\varepsilon_{i,t-1}, \Lambda^{-1}_{1,\theta_{1(i)}}(\Lambda_{2,\theta_{1(i)}}(G_i(\varepsilon_{i,t-1}) + \sigma_{i,\theta_{3(i)}}(G_i(\varepsilon_{i,t-1})\zeta_i))/D_i(G_i(\varepsilon_{i,t-1}) \end{aligned} \tag{7.15}$$

where,

$$D_i(u_i) = \frac{d\Lambda^{-1}_{1,\theta_{1(i)}}(\Lambda_{2,\theta_{1(i)}}(u_i) + \sigma_{i,\theta_{3(i)}}(u_i)\zeta_i))}{d\zeta_i} \tag{7.16}$$

and satisfies the condition that

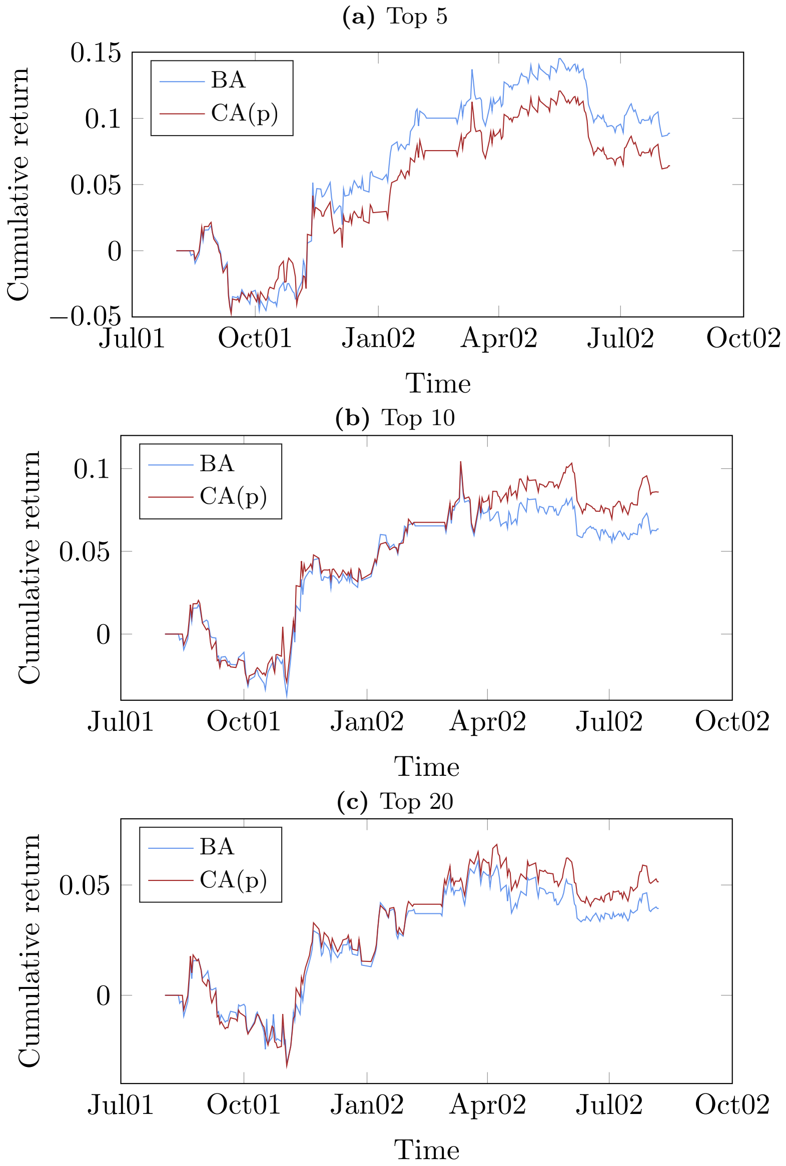
$$\Lambda_{2,\theta_{1(i)}}(u_i) = E[\Lambda_{1,\theta_{1(i)}}(v_{i,t})|u_{i,t} = u_i] = \int_0^1 \Lambda_{1,\theta_{1(i)}}(v_i) \times c_i(u_i, u)du \tag{7.17}$$

Risk control

As shown by Papanicolaou and Yeo (2016) we can improve the performance by allowing the trading activity only when the goodness-of-fit of the estimation for trading signals is sufficiently high. In this subsection we show that by doing so indeed improves the performance for the first two trading periods.

Figure 8: Cumulative returns after risk control

The graphs below show the cumulative returns (×100%) over time for different strategies over the period August 8th 2002 until April 1st 2003. The portfolio returns are based on the benchmark approach (BA) and the parametric copula approach (CA(p)) after applying risk control and by selecting the top k mean-reverting pairs.



B. Tables and figures

Table 12: Number of connections on industry group level

This table shows the average number of connections on industry group level. This average is calculated over all 15 estimation windows and rounded down to the nearest integer value. A connection between industry group a and b is made when $\beta^{(s)}_{i,j} \neq 0$ where firm i in industry group a and firm j in industry group b, evaluated at estimation window s. The numbers in the axis indicate the industry group and are defined as before: 1: ENRGY; 2: MTRLS; 3: CAPGD; 4: COMPR; 5: TRANS; 6: AUTOC; 7: CONDA; 8: CONSV; 9: MEDIA; 10: RTLNG; 11: FSRET; 12: FBVTA; 13: HPROD; 14: HCEQS; 15: PHBLS; 16: BANKS; 17: DIVFN; 18: INSUR; 19: SFTSV; 20: TECHH; 21: SMCEQ; 22: TELSV; 23: UTLTS; 24: RLEST.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	2	12	8	0	4	1	0	0	0	2	1	3	0	1	4	1	0	0	1	7	1	0	2	0
2	10	239	185	10	36	46	29	3	10	25	11	33	7	8	31	15	9	10	22	57	16	8	17	12
3	10	170	330	9	35	79	47	4	6	19	11	32	7	5	15	18	20	11	30	95	23	15	7	21
4	0	4	9	4	8	5	3	0	2	2	0	8	4	1	5	1	2	0	1	5	0	1	1	3
5	1	29	44	6	58	10	10	4	4	7	4	22	6	1	13	7	2	5	9	11	0	6	12	6
6	2	32	68	4	11	76	18	1	4	2	4	6	2	2	2	7	2	3	4	23	6	2	3	5
7	1	28	49	3	7	22	22	1	1	5	2	6	2	2	3	7	8	4	4	27	10	0	4	6
8	0	4	5	1	4	0	1	0	1	0	0	3	0	0	0	0	0	0	1	0	0	0	0	0
9	0	6	6	1	7	0	1	0	2	1	0	2	0	0	3	3	0	0	3	1	0	5	3	0
10	0	13	21	1	7	5	4	0	1	16	6	10	2	1	5	4	2	4	6	10	1	6	2	1
11	0	8	10	0	9	3	4	0	1	7	4	8	3	0	8	2	2	1	0	5	0	1	4	2
12	1	22	26	6	18	3	5	2	1	13	6	56	4	4	18	6	3	1	12	8	1	2	7	2
13	0	1	3	4	5	1	2	0	0	0	3	5	2	1	4	1	1	0	1	2	0	3	4	2
14	2	5	6	2	0	4	1	0	0	0	0	3	0	2	1	0	2	2	3	9	1	0	0	0
15	1	13	11	3	13	2	6	0	3	6	3	27	5	2	38	8	2	2	7	4	1	3	8	1
16	1	7	21	1	9	6	5	0	2	3	1	8	2	0	4	23	5	8	5	9	1	2	4	10
17	0	5	19	3	1	2	6	0	1	3	2	4	0	2	1	6	10	3	3	5	0	2	0	9
18	0	3	7	0	2	2	3	0	0	5	0	2	1	1	2	8	3	12	2	2	0	2	3	1
19	0	19	33	2	9	4	9	1	1	8	3	15	1	5	8	5	2	2	17	7	4	6	2	4
20	7	47	93	6	9	27	32	1	3	6	3	16	2	9	9	13	2	7	12	60	11	3	8	4
21	0	17	27	0	1	4	8	0	1	3	0	2	0	1	1	0	1	1	5	13	6	0	1	2
22	1	2	14	1	10	1	1	0	3	7	2	2	1	0	5	2	2	4	6	4	0	5	4	3
23	0	9	7	0	7	1	1	0	0	3	3	5	2	0	5	1	1	0	2	2	0	3	17	0
24	0	12	20	1	6	5	8	0	0	0	1	3	1	2	0	5	3	2	5	5	2	2	2	12

Table 13: Fraction of industry group members selected in baskets

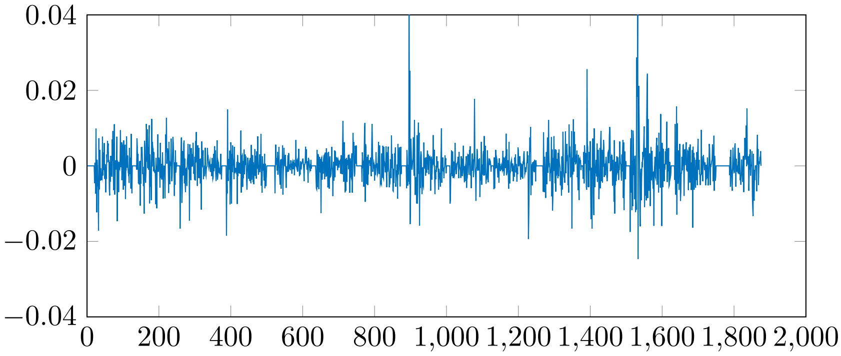
This table shows the average number of connections on industry group level. This average is calculated over all 15 estimation windows and rounded down to the nearest integer value. A connection between industry group a and b is made when $\beta^{(s)}_{i,j} \neq 0$ where firm i in industry group a and firm j in industry group b, evaluated at estimation window s. The numbers in the axis indicate the industry group and are defined as before: 1: ENRGY; 2: MTRLS; 3: CAPGD; 4: COMPR; 5: TRANS; 6: AUTOC; 7: CONDA; 8: CONSV; 9: MEDIA; 10: RTLNG; 11: FSRET; 12: FBVTA; 13: HPROD; 14: HCEQS; 15: PHBLS; 16: BANKS; 17: DIVFN; 18: INSUR; 19: SFTSV; 20: TECHH; 21: SMCEQ; 22: TELSV; 23: UTLTS; 24: RLEST.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0.33	0.15	0.1	0.08	0.16	0.06	0.08	0.23	0.08	0.1	0.08	0.14	0.1	0.07	0.17	0.09	0.03	0.03	0.03	0.09	0.02	0.08	0.14	0.06
2	0.17	0.22	0.13	0.11	0.1	0.09	0.09	0.12	0.09	0.09	0.1	0.09	0.09	0.11	0.1	0.07	0.08	0.07	0.09	0.1	0.11	0.07	0.08	0.07
3	0.11	0.13	0.17	0.08	0.08	0.11	0.12	0.15	0.11	0.08	0.08	0.08	0.07	0.09	0.07	0.08	0.09	0.07	0.11	0.12	0.11	0.09	0.04	0.1
4	0.07	0.1	0.08	0.53	0.14	0.1	0.17	0.16	0.16	0.14	0.2	0.16	0.34	0.16	0.18	0.1	0.17	0.14	0.12	0.15	0.11	0.11	0.14	0.08
5	0.14	0.08	0.08	0.13	0.36	0.08	0.08	0.19	0.18	0.13	0.1	0.14	0.15	0.08	0.12	0.1	0.08	0.11	0.09	0.06	0.06	0.09	0.15	0.13
6	0.07	0.08	0.1	0.11	0.08	0.44	0.13	0.08	0.06	0.06	0.05	0.07	0.06	0.15	0.1	0.07	0.06	0.1	0.06	0.12	0.09	0.09	0.06	0.06
7	0.09	0.08	0.1	0.15	0.07	0.14	0.19	0.04	0.1	0.08	0.09	0.11	0.09	0.14	0.07	0.1	0.12	0.05	0.13	0.21	0.23	0.09	0.05	0.14
8	0.23	0.1	0.12	0.16	0.19	0.07	0.05	0	0.13	0.1	0.1	0.19	0.07	0.1	0.11	0.08	0.02	0.1	0.08	0.03	0.04	0.07	0.01	0.05
9	0.1	0.08	0.09	0.14	0.18	0.05	0.09	0.07	0.4	0.2	0.11	0.11	0.08	0.1	0.14	0.13	0.09	0.17	0.22	0.09	0.04	0.16	0.11	0.07
10	0.15	0.07	0.07	0.11	0.12	0.05	0.07	0.07	0.17	0.35	0.51	0.11	0.06	0.14	0.1	0.13	0.22	0.09	0.17	0.05	0.06	0.18	0.04	0.15
11	0.06	0.09	0.08	0.19	0.11	0.06	0.1	0.1	0.14	0.55	0.73	0.11	0.27	0.04	0.1	0.08	0.14	0.21	0.11	0.08	0.06	0.2	0.11	0.13
12	0.09	0.08	0.06	0.15	0.13	0.06	0.08	0.13	0.09	0.12	0.09	0.45	0.21	0.14	0.2	0.06	0.04	0.07	0.09	0.07	0.04	0.08	0.15	0.06
13	0.07	0.06	0.06	0.32	0.14	0.05	0.08	0.07	0.08	0.06	0.24	0.26	1	0.18	0.32	0.09	0.08	0.07	0.11	0.09	0.04	0.15	0.22	0.05
14	0.07	0.09	0.07	0.1	0.06	0.15	0.1	0.1	0.1	0.07	0.03	0.14	0.18	0.63	0.17	0.04	0.08	0.03	0.21	0.22	0.09	0.13	0.07	0.04
15	0.1	0.06	0.04	0.13	0.09	0.08	0.06	0.07	0.13	0.09	0.08	0.2	0.24	0.16	0.65	0.07	0.03	0.03	0.06	0.07	0.03	0.09	0.17	0.03
16	0.12	0.06	0.07	0.08	0.1	0.07	0.1	0.1	0.08	0.09	0.09	0.08	0.09	0.07	0.07	0.63	0.26	0.23	0.12	0.05	0.04	0.12	0.07	0.17
17	0.04	0.08	0.09	0.18	0.09	0.07	0.13	0.05	0.09	0.25	0.19	0.06	0.07	0.09	0.06	0.31	0.63	0.13	0.17	0.11	0.1	0.13	0.04	0.33
18	0	0.06	0.05	0.18	0.12	0.09	0.08	0.07	0.17	0.13	0.17	0.08	0.09	0.1	0.06	0.3	0.15	0.4	0.15	0.06	0.02	0.08	0.1	0.19
19	0.04	0.07	0.09	0.12	0.09	0.05	0.13	0.13	0.17	0.17	0.12	0.1	0.08	0.22	0.08	0.12	0.14	0.13	0.46	0.16	0.19	0.29	0.04	0.06
20	0.1	0.1	0.12	0.13	0.05	0.13	0.2	0.03	0.1	0.07	0.07	0.08	0.09	0.24	0.1	0.06	0.1	0.07	0.17	0.28	0.31	0.12	0.06	0.05
21	0	0.09	0.1	0.12	0.05	0.09	0.21	0.08	0.06	0.09	0.11	0.03	0.06	0.11	0.05	0.08	0.11	0.01	0.18	0.25	0.77	0.1	0.04	0.05
22	0.08	0.06	0.08	0.11	0.1	0.09	0.1	0.07	0.11	0.19	0.16	0.08	0.1	0.14	0.12	0.1	0.14	0.1	0.28	0.1	0.09	0.59	0.18	0.1
23	0.11	0.05	0.03	0.1	0.12	0.05	0.05	0.03	0.07	0.03	0.08	0.12	0.16	0.08	0.15	0.06	0.03	0.08	0.03	0.04	0.02	0.12	0.81	0.03
24	0.05	0.07	0.1	0.11	0.11	0.07	0.13	0.07	0.08	0.14	0.09	0.07	0.05	0.06	0.06	0.16	0.23	0.17	0.08	0.05	0.06	0.09	0.05	0.68

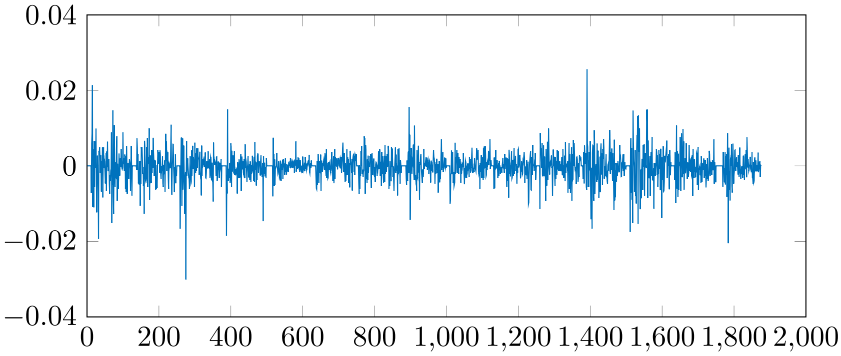
Figure 9: Returns benchmark approach

The graphs below show the returns over time for the benchmark strategies over the period August 8th 2002 until April 1st 2010.

(a) Top 5



(b) Top 10



(c) Top 20

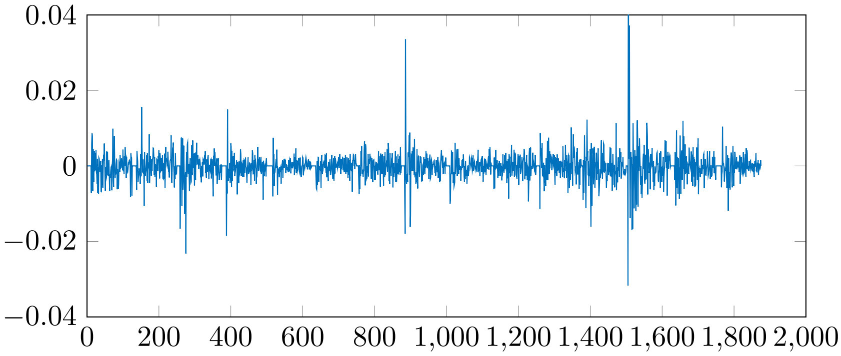


Figure 10: Returns copula approach (parametric)

The graphs below show the returns over time for the parametric copula strategies over the period August 8th 2002 until April 1st 2010.

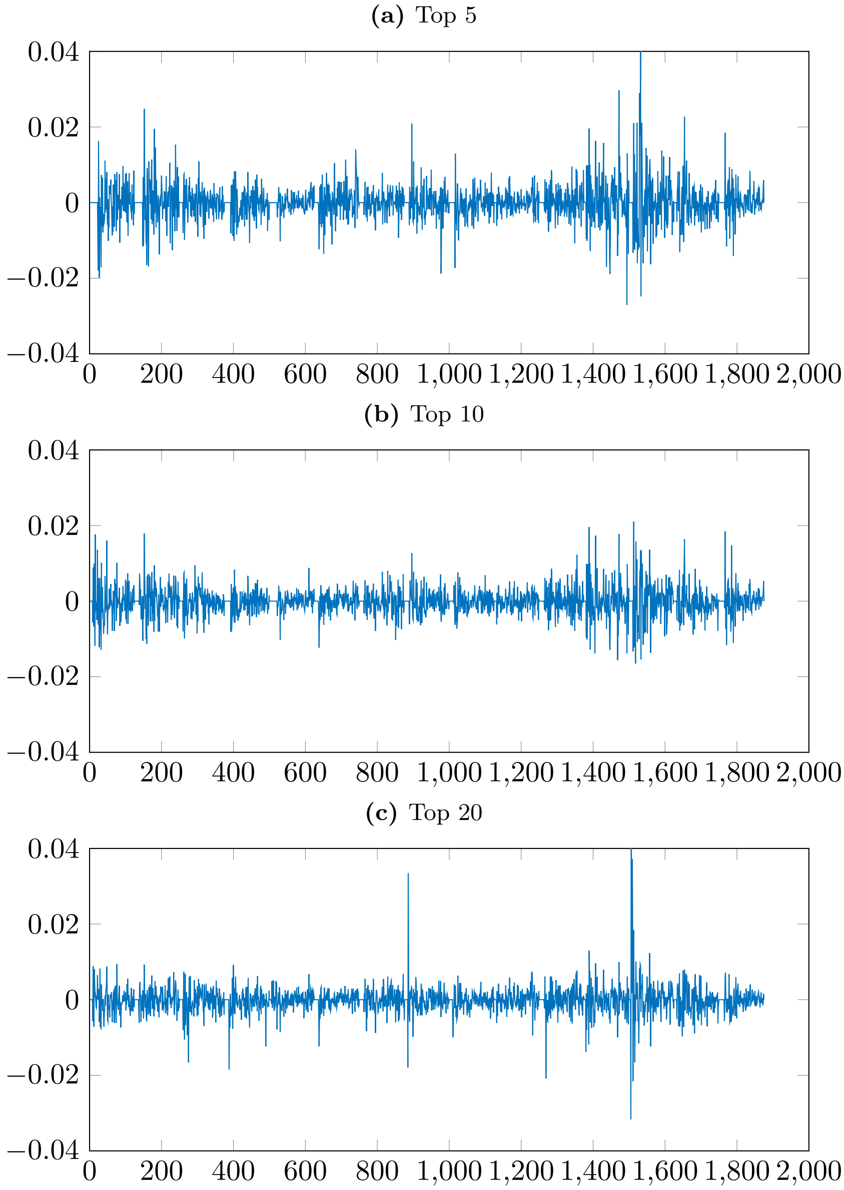
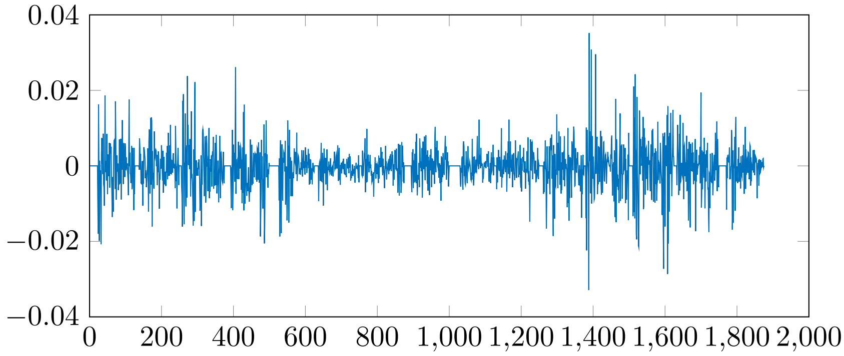


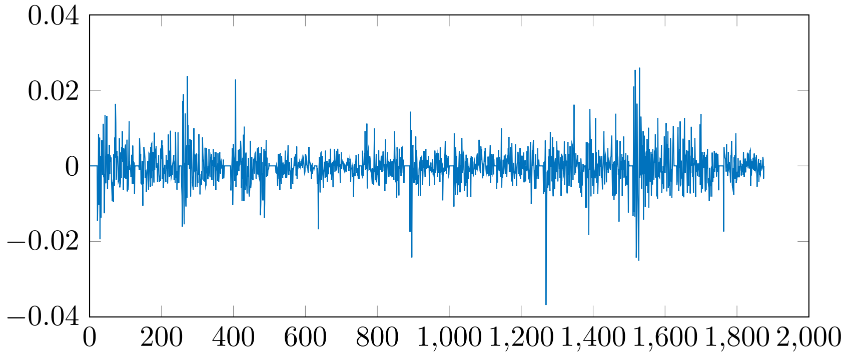
Figure 11: Returns copula approach (semi-parametric)

The graphs below show the returns over time for the semi-parametric copula strategies over the period August 8th 2002 until April 1st 2010.

(a) Top 5



(b) Top 10



(c) Top 20