UNIVERSITY OF AMSTERDAM

MSc. Computational Science

Track: Computational Finance

Master Thesis

High-Frequency Statistical Arbitrage Trading using Dynamic Copulas

by

Haro de Jong

11040890

December 8, 2017

Examiner/Supervisor:	*Assessor:*	*Assessor:*
Prof. Dr. B.D. Kandhai	Dr. S. Sourabh	I. Anagnostou MSc

Abstract

Pairs trading is a popular statistical arbitrage trading strategy employed by many financial markets practitioners. The strategy relies on exploiting temporary mispricings between financial instruments that will eventually revert back to their fair value. In this thesis we propose a dynamic copula pairs trading framework suitable for high-frequency trading and test its performance on a large set of stock pairs. High-frequency financial data is known for exhibiting stylized facts such as volatility clustering and seasonality. As such, the conventional copula method will not accurately reflect the characteristics of stock pairs. Therefore, we model the stock returns using a MC-GARCH type model and use a rolling estimation window for the copula parameters. For the modeling of the joint stock returns we consider both Archimedean and Elliptical copulas. We select suitable pairs based on their correlation during the formation period, optimize the trading parameters during a pseudo-trading period and evaluate our strategy during an out-of-sample trading period. We apply our trading strategy on highly liquid stocks that are traded on the NASDAQ or NYSE. We form four sector portfolios: Financials, Energy, Industrials and Technology. Each portfolio is constructed by selecting the top-5 pairs with the highest correlation during the formation period within their sector. Our empirical analysis finds that the student's t-copula is the best suitable copula for the modeling of the joint distribution of the filtered returns, and by utilizing this copula while ignoring the cost of trading, each sector portfolio is able to produce a mean daily return of at least 3% with annualized Sharpe ratios between 10 and 20 and Sortino ratios between 16 and 28. Furthermore, none of the portfolios are greatly exposed to drawdown risk, tail-risk or systematic sources of risk. When including transaction costs of 0.75, 1.5 or 3 basis points per round-trip we observe that the mean daily return drops significantly to 1%, -1% and -5% respectively. In addition, when imposing a wait-one-bar trading rule to account for delayed execution, returns declined with approximately 50%. Hence, the empirical results show that for participants that are able to achieve little or no transaction costs and that are able to execute their trades fast and efficiently, our proposed dynamic copula trading method can produce significant economical returns while incorporating little risk.

1 Introduction

Pairs trading is generally defined as a statistical arbitrage strategy that capitalizes on the temporary relative mispricing between two instruments whose prices are expected to converge due to strong historical co-movements. By going long on the relatively undervalued stock and short on the relatively overvalued stock, a profit can be made by unwinding the position upon convergence of the stock prices to their fair values. Pairs trading is one of the most common statistical arbitrage strategies that is used by professional traders, institutional investors, and hedge funds.

Throughout the years different pairs trading strategy have been proposed. Firstly, the distance method of Gatev et al. (2006) is one of the first pairs trading strategies which dates back to the mid-1980's. It uses the distance between normalized prices to form a spread that is used as a criteria to judge the degree of mispricing between two instruments. Once the spread exceeds a certain threshold, a long/short position is opened. When the spread diverges back to zero, the position is closed and the profit/loss is taken. Secondly, the cointegration approach outlined in Vidyamurthy (2004) is an attempt to parameterize pairs trading by exploring the possibility of cointegration (Engle and Granger, 1987). Under the assumption that there exists a linear relationship between two non-stationary stock price series, one can identify a stationary spread series using a linear combination of these two non-stationary series. Then, any divergence in the spread from zero should eventually diverge back to its mean and similar to the distance method we open a long/short position when the spread exceeds a certain threshold. In general, the distance and cointegration methods are referred to as the conventional pairs trading methods.

The main assumption of these methods is that there exists a linear relationship between the stocks and thus one can use correlation and cointegration to measure the dependency. This assumptions would be valid if financial data where normally distributed, however they are hardly normally distributed in practice (Ling, 2006). Hence, as an alternative to overcome this assumption Ferreira (2008) was the first to introduce a copula pairs trading method. Liew and Wu (2013) and Xie et al. (2014) extended on this work and further outlined the copula pairs trading frameworks. The copula pairs trading approach aims to generalize the conventional pairs trading methods. The foundation for copulas was laid by Sklar's theorem (Sklar, 1959). It provided the link between marginal distributions and their corresponding joint distribution. With copulas, the estimation of marginal and joint distributions is separated, this allows for the relaxation of the normality assumption. In addition, there exists explicit functions for the copulas, allowing for a better evaluation of the dependency between stocks.

More recently Zhi et al. (2017) proposed the dynamic copula trading method which further generalizes the work of Xie et al. (2014) by incorporating a dynamic method to account for stylized facts of stock returns such as volatility clustering. Zhi et al. (2017) test the performance of their frameworks on daily data and finds that the dynamic copula trading method is able to outclass both the static copula method as well as the conventional methods. The aim of this thesis is to investigate the performance of the dynamic copula pairs trading method on high-frequency data. By trading on a higher frequency the number of trading opportunities increases but also new statistical and computational challenges arise such as seasonality and slippage. We aim to overcome these challenges by utilizing the MC-GARCH model of Engle & Sokalska (2012) which is able to capture the daily, intraday and diurnal volatility that is present in high-frequency data. Our dataset consists of over 80 US equity stocks that are listed on the NYSE or NASDAQ. For our analysis we construct portfolios with pairs that are part of the following sectors: Financials, Technology, Energy, and Industrials. We construct pairs based on their correlation during a formation period of 12 months. Then we optimize the trading parameters using a pseudo-trading period of 6 months. And finally, we evaluate the performance of each portfolio using an out-of-sample period of 12 months.

The remainder of this thesis is organized as follows: In Chapter 2 we provide an extensive literature review of the most important results regarding copula-based pairs trading and high frequency statistical arbitrage and we further elaborate on our research goal and our approach. In Chapter 3 we outline the most important characteristics of our dataset. Then, in Chapter 4 we present the dynamic copula pairs trading method. Thereafter, in Chapter 5 we further elaborate on the trading method employed and the evaluation criteria. In Chapter 6 we present our empirical findings, we investigate a single stock example and perform an industry wide portfolio analysis. Finally, in Chapter 7 we give our conclusion and provide directions for further research.

2 Background

2.1 Literature Review

In this section we cover the most important results regarding high-frequency trading, statistical arbitrage, and the copulas pairs trading method.

As of today, there exists multiple approaches to statistical arbitrage pairs trading using copulas. Firstly, Ferreira (2008) introduced the return-based copulas pairs trading method. With this method, pairs are created with commonly applied co-movement metrics, i.e., correlation or cointegration criteria. Then the distribution of the log returns of each stock is estimated using either parametric or non-parametric distributions. Using the probability integral transform method the data is transformed to uniform variables and the copula parameters are estimated. With the estimated copula parameters the conditional marginal distribution is determined and the trading strategy can be executed. The strategy is tested out-of-sample on a daily dataset of 12 months for a single stock pair and was found to be highly profitable. The return-based copula pairs trading strategy is further outlined by Stander et al. (2013). If the conditional marginal probability of a stock is higher (lower) than 0.5, the stock is considered overvalued (undervalued) relative to its peer. The authors suggest to trade when the conditional probabilities are in the tail regions of their conditional distribution function, i.e., below their 5% and above their 95% confidence level. Furthermore, they investigate a set of 22 different Archimedean copulas and suggest exiting the position as soon as it is profitable. Liew and Wu (2013) compares the method to conventional pairs trading methods such as the distance and cointegration method using daily stock data. Data from the first 24 months was used to find the optimal parameters and the subsequent 12 months are used as the trading period. Liew and Wu (2013) deviate from Stander et al. (2013) in the sense that they consider five copulas (Gumbel, Student-t, Normal, Frank and Clayton) that are most commonly used in financial applications and advocate to reverse the position once the conditional probabilities cross the boundary of 0.5 again. Their empirical results demonstrate that the dependence structure of the Gumbel copula fitted the data best and that the copula approach for pairs trading is superior to the conventional methods. However, the drawbacks of this approach is that the entry and exit signals do not take the possible convergence or divergence of the pairs into account.

Next, the level-based copula method is similar to the return-based method but deviates in terms of the method used to enter or exit a position. By subtracting 0.5 from the conditional marginal probability, and accumulating the mispricing over several periods to a misprice index, the strategy takes the time structure of the mispricing into account. Xie et al. (2014) used the strategy for a large-sample analysis of 10 years of utility industry data with a one day holding period. Their results demonstrate the superiority of the method over the conventional distance method. Over the long data period, the top 5 pairs identified indicate that the distance method produces insignificant excess returns, in contrast to the proposed copula method, which produced a 3.6% annualized excess return. They conclude that the proposed copula method better captures the dependency structure and provides more trading opportunities with higher excess returns and profits than the traditional approach. Rad et al. (2015) performed a study on the performance of three pairs trading strategies - the distance, cointegration and copula method - on the entire US equity market from 1962 to 2014. In terms of economic outcomes, the distance, cointegration and copula methods show a mean monthly excess return of 38, 33 and 5 basis points after transaction costs, respectively. Even though they find that the performance of the copula method is weaker than the distance and cointegration methods, their results still provide some important insights. First, in recent years the distance and cointegration strategies suffer from a decline in trading opportunities, whereas the copula method remains stable in presenting such opportunities. Second, the copula method shows returns comparable to those of other methods in its converged (long/short positions that mean-revert) trades, even though its relatively high proportion of unconverged (long/short positions that do not mean-revert) trades countervails a considerable portion of such profits. Therefore, any attempt to increase the ratio of converged trades or limit their losses would result in enhanced performance outcomes. Third, the copula methods unconverged trades exhibit higher risk-adjusted performance than those of any other strategy which further motivates the use of such strategies. Finally, they find that the Student-t is selected as the copula that provides the best fit for the dependence structure across stock pairs in pairs trading on the US equity market. As with the return-based method, the level-based method also has its drawbacks. For example, it does not differentiate between pairs that reach the critical levels using many small mispricings or pairs that have a few large mispricing steps.

The aforementioned copula methods all utilize a static model to estimate the joint distribution of stock pair's returns. However, it is show by Ang & Chen (2002), that financial assets have different correlations of stock returns between market upturns and downturns. Therefore, Zhi et al. (2017) proposed a dynamic copula framework for pairs trading that uses the copula-GARCH model and a rolling window formation period to account for the dynamic dependency structure. The framework is tested on 10 years of daily stock data of three Asia-Pacific indices using a rolling estimation window of 6 months. Their results show that the dynamic copula method is generally able to produce higher excess returns, Sharpe ratios and Sortino ratios across all three markets compared to the distance and non-dynamic copula pairs trading method.

Then, turning our attention to high frequency pairs trading, Bowen et al. (2010) examined the characteristics of high frequency pairs trading using a sample of FTSE100 constituent stocks for the period January to December 2007. They showed that the excess returns of the strategy are extremely sensitive to both transaction costs and speed of execution. When introducing transaction costs the excess returns of the strategy were reduced by more than 50%. Likewise, when implementing a wait one period restriction on execution the positive return was completely eliminated. Miao (2014) investigated a high frequency and dynamic pairs trading system using a two-stage correlation and cointegration approach. The strategy was applied to equity trading in U.S. equity markets. The proposed pairs trading system was tested for out-of-sample testing periods with 15-minute stock data from 2012 and 2013. The strategy yields cumulative returns up to 56.58%, which exceeded the S&P 500 index performance by 34.35% over a 12-month trading period. The proposed trading strategy achieved a monthly 2.67 Sharpe ratio and an annual 9.25 Sharpe ratio. Furthermore, the proposed pairs trading system performed well during the two months in which the S&P 500 index had negative returns.

2.2 Research Statement

As of today, there are no official publications on the performance of high-frequency copula pairs trading strategies. This thesis makes a first attempt to fill this gap by adapting the dynamic copula pairs trading method of Zhi et. al (2017) to high-frequency data and testing it on a large set of highly liquid stock pairs. Our dataset consists of 2 years of 1-minute last-traded prices ranging from 2015 until 2016. We use a dependence criteria based pairs selection, taking into account a 12 month formation period, consisting of overlapping 6 month estimation periods and a subsequent 6 month pseudo-trading period. We group our stocks into four sectors: Energy, Technology, Financials and Industrials. The top-5 most attractive pairs within each sector that have the strongest dependence in the formation period are transferred to a 12 month out-of-sample trading period. We evaluate the performance of the trading strategy based by its risk-return characteristics and compare the results with an S&P500 Buy-and-Hold strategy.

3 Data

Our dataset consists of 1-minute and daily last-traded OHLC (Open-High-Low-Close) prices of over 80 stocks that are traded on the NYSE or NASDAQ, and is obtained from Algoseek.com. Our universe of stocks can be divided into four sector groups: Energy, Technology, Financials and Industrials. Furthermore, each stock is part of the most liquid traded stocks on the NYSE or NASDAQ. A complete list of stocks considered can be found in Appendix A. The advantages of using such dataset are three-fold. First, the highly liquid stocks serve as a robust test for the trading strategy, as investor and analyst have excessively studied these large capitalization stocks. Secondly, according to the efficient market hypothesis the prices of traded assets reflect all known information at any given time, however, due to liquidity demands this hypothesis does not always hold allowing arbitrage strategies to be profitable. Third, due to the high liquidity, market friction will have less impact on the performance of our strategy. The sample period of the 1-minute data ranges from 15:31 GMT, January 5, 2015 through 22:00 GMT, December 31, 2016. It comprises 502 trading days with a total of approximately 194500 observations, where a normal trading day comprises 390 observations. The daily data sample runs from January 2, 2012 until December 31, 2016 and thus comprises a total of 1322 trading days.

In Table 1 we display the average last traded price, maximum last traded price and the minimum last traded price of all the stocks considered for the complete sample period. The table shows that there is a wide variety between the stock prices of our stock universe with the average last traded price ranging from \$13.70 to \$348.00. Furthermore, the maximum and minimum last traded price show that the stocks also exhibit strong fluctuations. These fluctuations are necessary for any kind of trading strategy to be profitable.

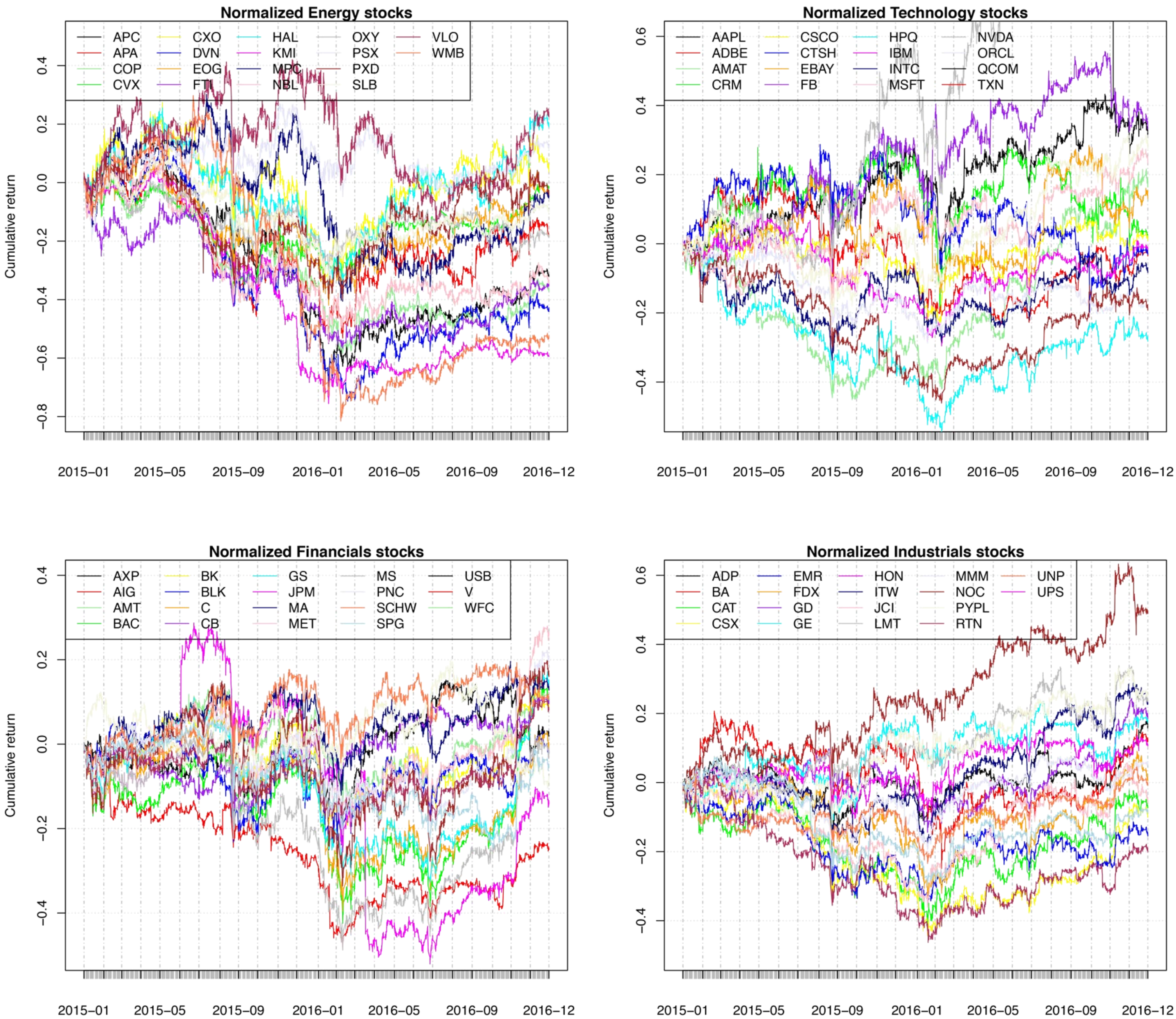
Table 1: Portfolios statistics of last traded stock pirces of complete sample period

Financials	Avg. Price $	Max Price $	Min Price $	**Industrials**	Avg. Price $	Min Price $	Max Price $
AIG	58.00	67.00	48.43	ADP	86.13	125.69	44.71
AMT	100.80	118.20	83.10	BA	86.67	103.85	65.94
AXP	70.86	93.91	50.28	CAT	138.24	160.05	102.15
BAC	16.00	23.39	11.00	CSX	78.69	97.35	56.38
BK	40.78	49.51	32.30	EMR	29.78	37.66	21.34
BLK	348.48	399.46	280.55	FDX	52.75	62.71	41.29
C	49.81	61.30	34.57	GD	163.81	201.54	119.79
CB	116.29	133.76	96.11	GE	143.54	180.01	121.64
GS	180.98	245.39	138.24	HON	28.56	33.00	20.30
JPM	37.11	51.87	23.11	ITW	106.55	119.32	90.63
MA	94.25	108.92	77.67	JCI	100.50	127.99	79.03
MET	64.69	87.32	50.07	LMT	37.26	46.12	27.12
MS	32.84	44.01	21.17	MMM	220.27	269.41	185.46
PNC	91.57	118.52	77.43	NOC	162.41	182.27	134.40
SCHW	30.51	40.57	21.53	PYPL	189.97	253.75	141.69
SPG	194.17	229.00	171.02	RTN	121.45	151.10	95.51
USB	43.13	52.68	37.07	UNP	94.00	124.46	67.47
V	74.30	83.94	60.55	UPS	103.26	120.58	87.33
Technology	Avg. Price $	Max Price $	Min Price $	**Energy**	Avg. Price $	Max Price $	Min Price $
AAPL	112.36	133.87	89.50	APA	53.91	71.85	32.20
ADBE	89.07	111.06	69.03	APC	64.05	95.80	28.18
AMAT	22.14	33.63	14.30	COP	50.30	70.10	31.06
CRM	72.56	84.44	52.65	CVX	98.07	119.00	69.78
CSCO	28.32	31.95	22.47	CXO	114.10	147.54	70.00
CTSH	59.74	69.77	45.52	DVN	44.42	70.44	18.08
EBAY	26.45	33.16	21.53	EOG	84.48	109.25	57.21
FB	102.85	133.37	73.60	HAL	41.53	56.05	27.65
HPQ	13.70	18.66	8.92	KMI	27.05	44.70	11.23
IBM	152.93	176.23	116.93	MPC	45.39	60.19	29.28
INTC	32.73	38.36	25.02	NBL	37.61	53.64	23.79
MSFT	50.95	64.10	39.84	OXY	73.13	83.68	58.24
NVDA	38.67	119.81	18.96	PSX	80.27	94.09	57.38
ORCL	40.01	45.23	33.16	PXD	151.04	194.82	103.58
QCOM	60.11	75.27	42.28	SLB	79.46	95.11	59.61
				VLO	60.12	73.87	43.50
				WMB	34.57	61.11	10.32

Note: Portfolios average last traded price, minimum last traded price, maximum last traded price of 1-minute close bar

To get a better sense of these fluctuations and the co-movements between the stocks. We compute for each industry the cumulative returns of each stock for the complete sample period and plot them in Figure 1. These graphs show how profitable each individual stock was since the beginning of our sample period, the fluctuations over time, and display the co-movements between the stocks. For each industry, we observe a strong co-movement between the stocks with cumulative returns generally moving in the same direction.

Figure 1: Cumulative returns of sector portfolios stocks



4 Copula

In this chapter we outline the dynamic copula estimation framework employed for the trading strategy. First we give a general introduction to copulas. Then we outline our high-frequency dynamic copula method, thereafter we define the copula functions that we consider and finally we elaborate on our estimation method.

Copulas are described as 'functions that join or couple multivariate distribution functions to their one dimensional marginal distribution function' by Nelson (2006). An n-dimensional copula is a function $C : [0,1]^n \to [0,1]$ that satisfies the following properties:

- $\forall u = (u_1, ..., u_n) \in [0,1]^n : \min\{u_1, ..., u_n\} = 0 \implies C(u) = 0.$
- $C(1, ..., u_i, 1, ..., 1) = u_i \quad \forall u_i \in [0,1]$
- $V_C([a,b]) \geq 0$, where $V_c([a,b])$ denotes the C-volume of the hyperrectangle $[a,b] = \prod_{i=1}^{n} [a_i, b_i], a_i \leq b_i \forall i \in \{1, ..., n\}$

Sklar's theorem (Sklar, 1959) is central to the theory of copulas, and is used to establish the relationship between the multivariate distribution function and the univariate margins. Let $H_{X_1,...,X_n}$ be an n-dimensional joint distribution function with marginals F_{X_i} $(i = 1, 2..., n)$. Then, there exists an n-copula C which satisfies the following equation for all $(x_1, ..., x_n) \in R^n$:

$$H_{X_1,...,X_n}(x_1, ..., x_n) = P(X_1 \leq x_1, ..., X_n \leq x_n) = C(F_{X_1}(x_1), ..., F_{X_n}(x_n)) \tag{1}$$

If the marginals are continuous then C is unique; otherwise C is uniquely determined on $Ran(F_1) \times ... \times Ran(F_d)$. Consequently, if F_{X_i} are distribution functions and C is a copula, the function $H(.)$ is a joint distribution function with marginal distribution functions $F_{X_1}, ..., F_{X_n}$. Implying that copula are not only joint distribution functions, but joint distribution functions can also be written in terms of a copula and a marginal distribution.

In the remainder of this paper we focus on the 2-dimensional copula specification. An extension to Sklar's theorem is provided by Patton (2006) that allows for conditional marginal distributions. Consider the random variables X and Y and conditional variable W. Denote $F_{X|W}(.|w)$ as the conditional distribution of $X|W$, and $F_{Y|W}(.|w)$ as the conditional distribution of $Y|W$. In addition, let $F_{XY|W}(.|w)$ be the joint conditional distribution of $(X,Y)|W$, and W be the support of w. Assume that $F_{X|W}(.|w)$ and $F_{X|W}(.|w)$ are continuous in x and y for all $w \in W$. Then there exists a unique conditional copula $C(.|w)$ such that:

$$F_{X,Y|W}(x,y|w) = C(F_{X|W}(.|w), F_{Y|W}(.|w)|w) \\ \forall(x,y) \in \Re \times \Re \text{ and each } w \in W \tag{2}$$

Since the converse of this result also holds, we can link together any two univariate distribution, of any type, with any copula and as a result have a valid bivariate distribution.

4.1 High Frequency Dynamic Copula

For our high frequency dynamic copulas pairs trading method we make use of the framework proposed by Zhi et al. (2017), which in turn is an extension to the works of Xie & Wu (2013) and Xie et al.(2014). The conventional copula pairs trading methods make use of a static copula to estimate the joint distribution of stock returns. However, stock returns exhibit characteristics such as volatility clustering and correlation between stock returns often differs substantially over time. Hence, the dynamic copula method accounts for these phenomena by filtering the returns and using a rolling estimation window.

Assume that stock X and Y are candidates for pairs trading. Their minute closing prices, i.e., the last price quoted within a minute, are defined as P_t^X and P_t^Y. Hence, we can compute the log-returns as,

$$r_t^k = \log(P_t^k) - \log(P_{t-1}^k), \text{ with } k = X, Y \tag{3}$$

High frequency financial returns are known for exhibiting stylized features such as volatility clustering, seasonality and heavy tails in the distribution. To account for these characteristics in the copula estimation we filter the returns of these features using generalized autoregressive conditional heteroskedaticity (GARCH) type models.

In the past, conventional GARCH models (Engle 1982, Bollerslev 1986) have been used to model intraday stock prices for different frequencies but where found unsuccessful. As a result Andersen and Bollerslev (1997, 1998) proposed a decomposition of intraday returns by means of the Fourier Flexible Function (FFF) method from Gallant (1981, 1982). However, the FFF method relies on incorporating announcement effects in the model, these announcement effect are often not easily observed by market participants and therefore, a different deseasonalization of high-frequency returns, which builds on the work of Anderson and Bollerslev, was introduced by Engle and Sokalska (2012). Their Multiplicative Component GARCH (MC-GARCH) model decomposes the intraday returns into multiplicative components. The conditional variance can then be expressed as a product of daily, diurnal and stochastic intraday volatility components. We can summarize the model mathematically as follows:

$$
\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\epsilon_t &= \sigma_t z_t, \quad z_t \sim D(0,1) \\
\sigma_t &= \sqrt{h_d s_i q_t}
\end{aligned} \tag{4}
$$

where μ_t is the mean (assumed 0), h_d is the daily variance, s_i is the seasonal variance, q_t is the intraday variance, z_t is the error term from a standardized distribution. For our approach we select the exponential GARCH (EGARCH) model of Nelson (1991) to model the daily and intraday volatility components. The EGARCH model extends on the GARCH model by allowing for the capture of leverage effects in the returns. The model is defined as follows,

$$
\ln(\sigma_t^2) = \omega + \sum_{i=1}^{q} \{\alpha_i z_{t-i} + \gamma_i(|z_{t-i}| - E|z_{t-i}|)\} + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}^2)
$$

where the coefficient α_i captures the sign effect and γ_i the size effect. For example, if $\alpha_i < 0$ future conditional variances will increase relatively more when a negative shock occurs compared to when a positive shock occurs. In addition, α_i, β_j and ω are not restricted in this case for σ_t^2 to be non-negative. Note that the daily variance component must be determined exogenous in order to avoid look-ahead bias. Hence, we base the daily variance on 1-step-ahead forecasts of the EGARCH model. In addition, we select the Student's t-distribution as the standardized distribution for both the daily and intraday model. Its probability density function is given by:

$$
t(z;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{(\nu-2)\nu\pi}\Gamma(\nu/2))} \left(1 + \frac{z^2}{(\nu-2)}\right)^{-\left(\frac{\nu+1}{2}\right)} \tag{5}
$$

with ν as the degrees of freedom, zero skewness and excess kurtosis equal to $6/(\nu-4)$ for $\nu > 4$. The student's t-distribution is selected for both models since daily stock prices are known to exhibit non-Gaussian dynamics as shown by Fama and Blume (1966). The distribution is fat tailed, which means that extreme price movements occur much more often than predicted given a Gaussian model. Gerig et al. (2009) investigated intraday price fluctuations for several stocks and found that these returns exhibit similar fat-tailness in the distribution, furthermore they provide evidence that, for all stock, the returns can be best explained using a Student's t-distribution.

Then, we estimate the seasonal component by taking the average of the intraday returns deflated by the daily variance over each minute bin i. i.e.,

$$
s_i = \frac{1}{D} \sum_{d=1}^{D} \frac{r_t}{h_d} \tag{6}
$$

Finally, we estimate the resulting intraday volatility after filtering the returns by the daily volatility and seasonality also using an EGARCH model. Throughout this paper we use an EGARCH(1,1) with student's t-distribution for the daily volatility forecasts and the intraday volatility use an EGARCH(1,2) with student's t-distribution for the intraday volatility. Thus, we do not perform an extensive model selection procedure. For the remainder of this thesis we refere to this model specification as MC-EGARCH, for more details on the estimation of the MC-GARCH model see Engle & Sokalska (2012).

Then, let $W_X = \{\sigma_t^X\}$ and $W_Y = \{\sigma_t^Y\}$ under the assumption that the volatility of stock X does not influence the volatility of stock Y. By Patton (2006), there exists a copula linking the conditional marginal distributions of $F_{X|W_X}$ and $F_{Y|W_Y}$. Hence, we can estimate a copula, C, based on the values of $u_t = F_{X|W_X}(e_t^X) = F_{X|W_X}(x|\sigma_t^X) = t_\nu(r_t^X/\sigma_t^X)$ and $v_t = F_{Y|W_Y}(e_t^Y) = F_{Y|W_Y}(y|\sigma_t^Y) = t_\nu(r_t^Y/\sigma_t^Y)$, with e_t^X and e_t^Y as the realized residuals of stocks X and Y at time t respectively. Thus,

$$C(u_t, v_t|W_X, W_Y) = C(F_{X|W_X}, F_{Y|W_Y}) \tag{7}$$

After obtaining the joint distribution of minute returns, we can denote the degree of mispricing using the conditional probability. Let $MI_t^{X|Y}$ and $MI_t^{Y|X}$ be the mispricings of the two stocks at time t be defined as follows:

$$\begin{aligned} MI_t^{X|Y} &= P(r_t^X < e_t^X | r_t^Y = e_t^Y, W_X, W_Y) \\ MI_t^{Y|X} &= P(r_t^Y < e_t^Y | r_t^X = e_t^X, W_X, W_Y) \end{aligned} \tag{8}$$

The misprice indices indicate whether the return of stock X (Y) is high or low at minute t, given the information of the return of stock Y (X) on the same minute and the relation between the two stock returns. A value of 0.5 for $MI_t^{X|Y}$ is interpreted as 50% chance for the price of stock X, to be below its current realization, given the current price of stock Y. Accordingly, conditional probability values above 0.5 show that chances for the stock price to fall below its current realization is higher than they are for it to rise, while values below 0.5 predict an increase in the stock price compared to its current value is more probable than a decrease. The conditional probabilities can be calculated by taking partial derivatives of the copula function w.r.t. to u and v.

$$\begin{aligned} MI_t^{X|Y} &= \frac{\partial C(u_t, v_t|W_X, W_Y)}{\partial v_t} \\ MI_t^{Y|X} &= \frac{\partial C(u_t, v_t|W_X, W_Y)}{\partial u_t} \end{aligned} \tag{9}$$

In the next chapter we will further outline how these mispricing indices can be used to create a trading strategy.

4.2 Copula Functions

For our research we consider two copula families to model the dependence structure between stock pairs, the Elliptical copulas and the Archimedean copulas. Archimedean copulas are copulas that have an explicit closed form solution and are therefore easy to estimate and define. For our research we only consider the three most popular ones, the Clayton, Gumbel and Frank copula. Elliptical copulas differ from the Archimedean classes of copulas in the sense that only implicit analytical expressions are available. These copulas are derived from the related elliptical distribution (e.g. normal, Student-t distribution). Since we are interested in modeling the dependence structure between two stocks we present the most important properties of the bivariate Elliptical and Archimedean copulas in the next section. Note, for simplicity we refer to the conditional marginal distributions as $F_{i|W_i} = F$ and to the uniform data as $u_t = u$ and $v_t = v$.

4.2.1 Elliptical Copulas

As stated before, Elliptical copulas are copulas that are generated by an elliptical distribution. The general analytical form of an Elliptical copula is given by:

$$C_\rho(u,v) = F_\rho(F^{-1}(u), F^{-1}(v)) \tag{10}$$

where $F_\rho(.)$ is the bivariate elliptical distribution with ρ as the correlation coefficient and F^{-1} inverse univariate distribution function. The most common elliptical distributions are the Gaussian and Student's t-distribution.

Gaussian Copula The bivariate Gaussian copula is parameterized by a linear correlation coefficient ρ and defined as:

$$C_\rho^{Gauss}(u,v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{11}$$

Where $\Phi(.)$ is the standard bivariate Gaussian distribution function given by:

$$\Phi_\rho(a,b) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{a} \int_{-\infty}^{b} \exp(-(x^2 - 2\rho xy + y^2)/2(1-\rho^2) dx\, dy \tag{12}$$

and $\Phi^{-1}(.)$ is the inverse of the univariate Gaussian distribution function. Then, the conditional probabilities of the Gaussian copula are,

$$\frac{\partial C_R^{Gauss}(u,v)}{\partial v} = \Phi\left(\frac{\Phi^{-1}(u) - \theta\Phi^{-1}(v)}{\sqrt{(1-\theta^2)}}\right) \tag{13}$$

$$\frac{\partial C_R^{Gauss}(u,v)}{\partial u} = \Phi\left(\frac{\Phi^{-1}(v) - \theta\Phi^{-1}(u)}{\sqrt{(1-\theta^2)}}\right) \tag{14}$$

Students t-Copula Contrary to the Gaussian copula, the student's t-copula is parametrized by two parameters, ν and ρ. Including the degrees of freedom parameter ν provides additional flexibility in controlling the structure of the copula. We can express the bivariate Student t-copula as:

$$C_{\nu,\rho}^t(u,v) = T_{\nu,\rho}(T_\nu^{-1}(u), T_\nu^{-1}(v)) \tag{15}$$

With $T_{\nu,\rho}(.)$ as the bivariate Student t-distribution defined as:

$$T_{\nu,\rho}(a,b) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{a} \int_{-\infty}^{b} \left(1 + \frac{x^2 + y^2 - 2\rho xy}{(1-\rho^2)\nu}\right)^{-\frac{\nu+2}{2}} dx\ dy \tag{16}$$

and T_ν^{-1} denotes the inverse of the univariate Student's t-distribution with ν degrees of freedom. Then, the conditional Students t-copula probabilities are given by,

$$\frac{\partial C_{\nu,\rho}^t(u,v)}{\partial v} = T_{\nu+1}\left(\sqrt{\frac{\nu+1}{\nu + T_\nu^{-1}(v)^2}} \times \frac{T_v^{-1}(u) - \rho T_\nu^{-1}(v)}{\sqrt{1-\rho^2}}\right) \tag{17}$$

$$\frac{\partial C_{\nu,\rho}^t(u,v)}{\partial u} = T_{\nu+1}\left(\sqrt{\frac{\nu+1}{\nu + T_\nu^{-1}(u)^2}} \times \frac{T_v^{-1}(v) - \rho T_\nu^{-1}(u)}{\sqrt{1-\rho^2}}\right) \tag{18}$$

Furthermore, note that the Students t-copula is such that as ν increases, it approaches the Gaussian copula.

4.2.2 Archimedean Copulas

A copula C is called Archimedean if it can be written in the form:

$$C(u,v;\theta) = \psi^{-1}(\psi(u;\theta) + \psi(v;\theta)) \tag{19}$$

with some generator function $\psi : [0,1] \times \Theta \to [0,\infty]$ that is continuous, strictly decreasing and convex such that $\psi(1;\theta) = 0$. θ is a parameter within some space Θ. In addition $\psi^{[-1]}$ is the pseudo-inverse of ψ defined as:

$$\psi^{[-1]}(t;\theta) = \begin{cases} \psi^{-1}(t;\theta) & \text{if } 0 \le t \le \psi(0;\theta) \\ 0 & \text{if } \psi(0;\theta) \le t \le \infty \end{cases} \tag{20}$$

Moreover ψ^{-1} generates an Archimedean copula in dimension 2 if and only if it is 2-monotone, i.e. $\psi \in C^2(0,\infty)$ and $(-1)^k\psi^{-1,(k)} \geq 0$ for any $k = 1, 2$. For our research we consider the most common Archimedean Copulas: the Clayton, Frank and Gumbel copula.

Clayton The Clayton copula captures lower-tail dependence of the data, the generator function and its corresponding inverse are given by:

$$\psi_\theta(t) = \frac{1}{\theta}(t^{-\theta} - 1) \qquad \psi_\theta^{-1}(t) = (1 + \theta t)^{-1/\theta} \tag{21}$$

with $\theta = 2\tau(1-\tau)^{-1}$ where τ is the Kendall's correlation coefficient. Then we can define the bivariate distribution of the Clayton copula by:

$$C_\theta^{Clayton}(u, v) = [\max\{u^{-\theta} + v^{-\theta} - 1, 0\}]^{-1/\theta} \quad \theta \in [-1, \infty]\backslash\{0\} \tag{22}$$

and, taking partial derivatives of the copula yields the conditional copula functions:

$$\begin{aligned}
\frac{\partial C_\theta^{Clayton}(u, v)}{\partial v} &= v^{-(\theta+1)}(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-1} \\
\frac{\partial C_\theta^{Clayton}(u, v)}{\partial u} &= u^{-(\theta+1)}(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}-1}
\end{aligned} \tag{23}$$

Frank The Frank copula does not capture specific tail dependence, its generator function and corresponding inverse are given by:

$$\begin{aligned}
\psi_\theta(t) &= -\log\left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}\right) \\
\psi_\theta^{-1}(t) &= -\frac{1}{\theta}\log(1 + \exp(-t)(\exp(-\theta) - 1))
\end{aligned}$$

Where θ is determined using MLE similar to the elliptical copulas. Then, the Frank copula function is given by:

$$C_\theta^{Frank}(u, v) = -\frac{1}{\theta}\log\left[1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1}\right] \quad \theta \in \mathbb{R}\backslash\{0\} \tag{24}$$

and, taking partial derivatives of the copula yields the conditional copula functions:

$$\begin{aligned}
\frac{\partial C_\theta^{Frank}(u, v)}{\partial v} &= \frac{\exp(-\theta v)(\exp(-\theta u) - 1)}{(\exp(-\theta) - 1) + (\exp(-\theta u) - 1)(\exp(-\theta v) - 1)} \\
\frac{\partial C_\theta^{Frank}(u, v)}{\partial u} &= \frac{\exp(-\theta u)(\exp(-\theta v) - 1)}{(\exp(-\theta) - 1) + (\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}
\end{aligned} \tag{25}$$

Gumbel The Gumbel copula only captures upper-tail dependence that is present in the data, the generator function and its corresponding inverse are given by:

$$\psi_\theta(t) = (-\log(t))^\theta \qquad \psi_\theta^{-1}(t) = \exp(-t^{1/\theta}) \tag{26}$$

with $\theta = (1-\tau)^{-1}$ where τ is the Kendall's correlation coefficient. Then we can define the bivariate distribution of the Gumbel copula by:

$$C_\theta^{Gumbel}(u,v) = exp[-[(-\log(u))^\theta + (-\log(v))^\theta]^{1/\theta}] \quad \theta \in [1,\infty) \tag{27}$$

Finally, the conditional copulas are determined by taking the partial derivatives:

$$\begin{aligned} C(v|u) &= \frac{\partial C_\theta^{Gumbel}(u,v)}{\partial v} = \frac{C_\theta^{Gumbel}(u,v)}{v} \times [(-\log(u))^\theta + (-\log(v))^\theta]^{\frac{1-\theta}{\theta}} \times (-\log(v))^{\theta-1} \\ C(u|v) &= \frac{\partial C_\theta^{Gumbel}(u,v)}{\partial u} = \frac{C_\theta^{Gumbel}(u,v)}{u} \times [(-\log(u))^\theta + (-\log(v))^\theta]^{\frac{1-\theta}{\theta}} \times (-\log(u))^{\theta-1} \end{aligned} \tag{28}$$

4.3 Estimation

The properties of the copulas allow for flexible multivariate distribution that can be constructed with pre-specified, discrete and/or continuous marginal distributions and copula functions that represent the desired dependence structure. To estimate the joint distribution we make use of a two step procedure where (1) the parameters of the EGARCH models for the marginal distributions of intraday returns are estimated using maximum likelihood estimation (MLE), and (2) conditional on the these estimated marginals the copula parameters are estimated also using MLE.

To estimate the copula we first describe the filtered returns by their empirical cumulative distribution functions F_X and F_Y. Then, we plug them into the copula density yielding the log-likelihood function with parameter set θ:

$$\ell(\theta) = -\sum_{i=1}^{N} \log[C_\theta(F_X(e_t^X), F_Y(e_t^Y))] \tag{29}$$

Then, with the maximum value of $\ell(\theta)$, we can define the Akaike Information Criterion (AIC) of Akaike (1973) and Bayesian Information Criterion (BIC), from Schwarz (1978) as follows:

$$AIC = -2\ell(\theta) + 2k \tag{30}$$

$$BIC = -2\ell(\theta) + log(N)k \tag{31}$$

where k are the number of parameters of the copula function and N the number of observations. The copula with the best overall fit minimizes the AIC and BIC values.

5 Copula Pairs Trading

In this chapter we outline our approach for the dynamic copula pairs trading strategy. First we elaborate on the construction of the formation and trading period. Then, we further examine the properties of the mispricing indices. Next, we shed light on the cost of trading for pairs trading and finally we introduce the performance evaluation procedure used to evaluate the trading results.

5.1 Formation Period

Our formation period T_f consists of 12 months of 1-minute observations, i.e. approximately 98000 observations, and runs from January 5, 2015 until December 31, 2015. Initially we split the formation period into two moving sub periods, a rolling estimation period and a pseudo-trading period. Thus, the estimation period initially consists of the first 6 months of data and the pseudo-trading period of the other 6 months. The size of the estimation window is in accordance with the work of Zhi et al. (2017) and provides a good trad-off between accuracy and computation time. However, further research could investigate the optimal estimation window.

5.1.1 Pairs selection

The first step is identifying suitable stock pairs for our trading algorithm, for this we need stocks that show strong co-movement. Past literature has proposed several methods for pairs selection, such as the Euclidean Distance Metric, ADF test and correlation measures. In our case we base our pairs-selection on the non-linear correlation measure Spearman's ρ correlation. A high correlation coefficient suggests a close co-movement of a stock pair. For selecting the pairs we compute the overall correlation coefficients of the complete formation period for each stock pair that belongs to the same sector. We select our pairs for the industry portfolios based on the top-5 pairs with the highest correlation, which are not below 0.8 and thus highly correlated. The Spearman's ρ correlation is calculated by first ranking the return series of stock X and Y, then we define the difference between the ranks of returns as $d_t = rk(r_t^X) - rk(r_t^Y)$ with $rk(.)$ as the ranked series, and the squared sum of the distance as $D = \sum_{i=1}^{N} d_i^2$. Then the Spearman's ρ is given by:

$$\rho = 1 - \frac{6D}{N(N^2 - 1)} \tag{32}$$

5.1.2 Estimation period

During the initial estimation period we use 6 months of minute data and 4 years of daily data to estimate the optimal MC-EGARCH models for the uniform transformation of the returns. Then, we estimate and select the optimal Archimedean (Gumbel, Frank, Clayton) or Elliptical (Normal and Student's t) copula using our method outlined in Section 4.3. The optimal MC-EGARCH model and copula are then transferred to the pseudo-trading period.

5.1.3 Pseudo-trading period

During the pseudo-trading period we re-estimate the optimal MC-EGARCH and copula model on a rolling basis with a look-back period equal to 6 months, thus, equal to the estimation period. Every 390 data points we re-estimate the parameters of our models ensuring that at the start of each trading day we are using the most recent parameters. With the estimated models we can construct $MI_t^{X|Y}$ and $MI_t^{Y|X}$, the mispricing indices. We define two trading indicators m_t^x and m_t^y that are set to zero at the start of the trading period. During the trading period, m_t^x and m_t^y are updated every minute as follows,

$$\begin{aligned} m_t^x &= m_{t-1}^x + MI_t^{X|Y} - 0.5 \\ m_t^y &= m_{t-1}^y + MI_t^{Y|X} - 0.5 \end{aligned} \tag{33}$$

where 0.5 comes from the relative mispricing outlined in the previous chapter. In addition, D is defined as the trade-entery point and S as the stop-loss (exit) trigger. Hence, there are four possible cases for opening a position (given that no trade is open), they are summarized in Table 2.

Table 2: Trade conditions and positions for copula pairs trading strategy

Open Trade	Long	Short	Close	Stop Loss
$m_t^x > D$	Y	X	$m_t^x < 0$	$m_t^x > S$
$m_t^x < -D$	X	Y	$m_t^x > 0$	$m_t^x < -S$
$m_t^y > D$	X	Y	$m_t^y < 0$	$m_t^y > S$
$m_t^y < -D$	Y	X	$m_t^y > 0$	$m_t^y < S$

All opening trades are closed at the end of the trading period regardless of the value of m_t^x and m_t^y. Xie et al. (2016) and Zhi et al. (2017) set $D = 0.6$ and $S = 2$ for daily data, we however, are investigating minute data and thus perform our own backtest procedure to find the optimal S and D for each pair individually.

5.2 Trading Period

The trading period consist of the second half of our data and runs from January 4, 2016 until December 31, 2016. The trading period is used to test the pairs trading strategy out-of-sample. Again we use a rolling-estimation method to update the copula models every 390 minutes with a 6-month look back. Implying that we do use data from the formation period for our trading period to estimate the parameters. Then, we execute the dynamic copula pairs trading strategy using the optimal trade entry point and stop-loss levels found during the pseudo-trading period. Thus, during the trading period our trade entry and stop-loss levels are fixed for the complete period. More sophisticated methods for selecting the entry point and stop-loss levels could be fruitful but this method has the smallest risk of overfitting on the data.

5.3 Mispricing Index Behaviour

Xie et. al (2014) extensively elaborates on the properties of the mispricing indices that are used for the copula trading strategy. In general, the copula trading strategy is build on the concept of combing a discrete step with a continuous state stochastic process. The cumulative distribution $F(.)$ of any continuous random variable X is seen as a uniform random variable from zero to one. Since $MI_t^{X|Y}$ and $MI_t^{Y|X}$ are conditional cumulative distributions, they are also uniform (0,1). Hence, $(MI_t^{X|Y} - 0.5)$ and $(MI_t^{Y|X} - 0.5)$ follow a uniform (-0.5, 0.5) distribution. m_t^X and m_t^Y, which are accumulations of the 1-minute $(MI_t^{X|Y} - 0.5)$ and $(MI_t^{Y|X} - 0.5)$, are the sums of a series of uniform random variables between -0.5 and 0.5 assuming that there is no correlation between them.

Mathematically if we define a series, where $F_t = F_{t-1} + e_t$, then e_t follows an i.i.d. uniform distribution from -0.5 to 0.5, and $F_0 = 0$. Then this series has the same characteristics as the m_t^X and m_t^Y series within a trading period (from opening to closing a position). This holds for both m_t^x and m_t^y because of the assumption that R_i^X is independent of R_j^X and R_i^Y is independent of R_j^Y for $i \neq j$. This sum of conditional probabilities minus its mean (0.5) values for every trading minute can be seen as a similar measurement to the degree of cumulative mispricing. It adds up the minor effect from each minute and provides a cumulative indicator.

If we assume that e_t follows a uniform distribution from -0.5 to 0.5 without time-series correlation, the F_t becomes equivalent to a pure random walk. In this case there is no arbitrage opportunity because there is an equal chance of further divergence or convergence. Even though the expected value of F_t is equal to zero, it is not a stationary time series. It can move strictly up or down, or fluctuate within a certain range. However, the pairs trading strategy relies on the assumption that F_t converges to zero when it is far from zero, i.e. mean-reverting. Furthermore, the strategy is at risk (Arbitrage Risk) when F_t has a tendency towards further divergence when the series is far from zero. In both cases e_t is not uncorrelated and thus we can identify the following three cases for F_t.

- Random Walk: No time-series correlation on the residual term e_t with lagged F_{t-1}, implying F_t is a purely random walk. There is no arbitrage opportunity in this case and thus no pairs trading strategy can be used.

- Mean-Reverting: There is negative correlation between the residuals term e_t and F_{t-1}, this causes F_t to have a tendency to converge when it is away from zero. Pairs trading strategies can be employed in this case.

- Arbitrage Risk: There is positive correlation between residuals e_t and F_{t-1}, this causes F_t to have a tendency to diverge further from zero when it is already far from zero. Pairs trading strategies make losses on average in this case.

In practice, F_t alternates between the three cases, resulting in both profits and losses in pairs trading. By introducing a stop-loss mechanism the losses can be minimized such that the strategy remains profitable.

To test the behavior of our mispricing indices we make use of the Hurst Exponent of Hurst (1951). The Hurst Exponent provides a scalar value that identifies whether a series is mean reverting, random walking or trending. The generalized Hurst exponents, $H_q = H(q)$, for a time series $g(t)(t = 1, 2, ..)$ are defined by the scaling properties of its structure functions $S_q(\tau)$

$$S_q(\tau) = \langle |g(t+\tau) - g(t)|^q \rangle_t \sim \tau^{qH_q(\tau)} \tag{34}$$

where $q > 0$, τ is the time lag and averaging is over the time window $t >> \tau$. In the case of we set $q = 2$ and $H_2(\tau) = H$, giving us the following relationship:

$$\langle |g(t+\tau) - g(t)|^2 \rangle_t \sim \tau^{2H} \tag{35}$$

Then, our mispricing indices can be characterized as follows:

- $H < 0.5$, the index is mean reverting

- $H = 0.5$, the index is a random walk

- $H > 0.5$, the index is trending

In addition to the characterization of the mispricing indices, the Hurst Exponent also indicates the extent to which a series behaves. For example, a value close to 0 implies that the index is highly mean reverting, or a value close to 1 implies that the index is strongly trending.

5.4 Trading Costs

The explicit trading costs for a pairs trading strategy comprise two round trip commissions per pair trade, short selling fees, and the implicit cost of the market impact. Determining the correct cost of trading is problematic because it varies with the sample period, with the size of the trade, and due to technical aspects such as the type of brokers assigned to execute the trade and the investment style underlying the trade.

Do and Faff (2012) have extensively discussed trading costs that arise within pairs trading frameworks. They estimate commissions per trade for institutional traders to decline from 10 basis points (bps) in 1998 to 7-9 bps in 2007-2009. Retail traders trade at around 10 bps according to Bogomolov (2013). In current markets, the commission fees of major online brokers vary from 1 bps to 2 bps per trade depending on the stock and size that is traded. Furthermore, transaction costs decline with the traders volume, reducing it even further to 0.1 bps per trade.

In the case of pairs trading, trades are measured in round-trips (RT) trades. A RT consists of 4 transactions, i.e. entering a position in stock X and Y and subsequently exiting the position. Hence, to evaluate our trading strategy we investigate three RT commission profiles: Low, Mid and High, with costs equal to 0.75 bps, 1.5 bps and 3 bps per RT respectively.

In addition to commissions, another important cost of trading, despite our highly-liquid stock universe, is market impact. Market impact, also known as slippage or market friction, is the difference between the expected price and the executed price of a trade. To investigate the effect of market impact we consider the one-bar-waiting rule. This rule assumes that the executed price is not the currents bar price but the price of the next bar, i.e. delayed execution. Even though this method will most likely over estimate the actual market impact encountered during a real life trading situation, it is still a good test for assessing the robustness of the strategy in light of this trading cost.

5.5 Performance Evaluation

To evaluate the performance of the trading strategy we investigate its risk and return characteristics and compare it to a Buy-and-Hold strategy of the S&P500 index. We compare each strategy in terms of return distribution, Value at Risk, drawdown, risk-return ratios and exposure to common risk factors. Below we define the most important evaluation statistics.

5.5.1 Return Calculation

For the calculation of the return of our strategy we make use of the equal weighted return. This return calculation method assumes that during every trade we buy/sell an equal amount of stocks. The cumulative return is then calculated by summing the simple returns generated by each trade.

5.5.2 Risk-Return ratios

Sharpe ratio The Sharpe ratio (Sharpe, 1975) is a measure of risk-adjusted portfolio performance, and measures the excesses return per unit of derivation. The Sharpe ratio formula is given by:

$$S = \sqrt{K}\left(\frac{r_p - r_f}{\sigma_p}\right) \tag{36}$$

where r_p is the expected portfolio return, r_f is the risk free rate, σ_p is the portfolio standard deviation, and K is the total number of returns. Since our returns are based on minute data we set $r_f = 0$, and $K = 252 \times 6.5 \times 60$ to compute the annualized Sharpe.

Sortino ratio The Sortino ratio (Sortino, 1994) is a modification of the Sharpe ratio, using downside deviation instead of the overall standard deviation as the measure of risk. By using the downside deviation, the Sortino ratio differentiates between harmful volatility and the total volatility. The ratio is defined as:

$$S = \sqrt{K}\left(\frac{r_p - r_t}{TDD}\right) \tag{37}$$

where r_p is the expected portfolio return, r_t is the risk free rate, K is the total number of returns, and TDD is the target downside deviation. The TDD can be obtained as follows:

$$TDD = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(\min(0, r_i - r_t))^2} \tag{38}$$

where r_i is the i^{th} portfolio return.

Upside Potential Ratio The Upside Potential Ratio is introduced by Sortino, Van der Meer and Plantinga (1999). It is an alternative to the Sortino and Sharpe Ratio by extending the measurement of only upside on the numerator, and only downside of the denominator of the ratio equation. It is defined as follows:

$$UPR_{mar} = \frac{\sum_{i=1}^{K} \imath^{+}(r_i - r_{mar})p_i}{\sqrt{\sum_{i=1}^{K} \imath^{-}(r_i - r_{mar})^2 p_i}} \tag{39}$$

with $\imath^- = 1$ if $r_i \leq r_{mar}$, $\imath^- = 0$, if $r_t > r_{mar}$, $\imath^+ = 1$ if $r_t > r_{mar}$ and $\imath^+ = 0$ if $r_t \leq r_{mar}$. Furthermore, p_i is the probability of an observation, i.e. $p_i = 1/T$, and r_{mar} is the minimal acceptable rate of return.

Omega Ratio The Omega ratio was introduced by Keating & Shadwick (2002) and is defined as the probability weighted ratio of gains versus losses for some threshold return target. Omega is calculated by creating a partition in the cumulative return distribution in order to create an area of losses and an area for gains relative to this threshold. The ratio is calculated as:

$$\Omega(r) = \frac{\int_r^{\infty}(1 - F(x)dx}{\int_{-\infty}^{r} F(x)dx} \tag{40}$$

where $F(.)$ is the cumulative distribution function of the returns and r is the target return threshold defining what is considered a gain versus a loss. For our analysis we set $r = 0$.

5.5.3 Drawdown Measures

Maximum Drawdown The Maximum Drawdown (MDD) can be defined as the largest percentage loss of an investment over a given period of time. The formal definition is given by:

$$MDD(T) = max_{\tau \in (0,T)}[max_{t \in (0,\tau)} X(t) - X(\tau)] \tag{41}$$

Where X(.) is the cumulative return of the portfolio. The MDD is a useful way to assess the relative riskiness of one strategy versus another, as it focuses on capital preservation, which is a key concern for most investors. However, this measure only tells us something about the maximum possible loss and not the frequency of the losses.

Calmar Ratio The Calmar ratio measures return versus drawdown risk and is calculated by dividing the annual return by the observed Maximum Drawdown.

$$Calmar\ Ratio = \frac{AnnualReturn}{MDD} \tag{42}$$

5.5.4 Risk measures

Value at Risk (VaR) Value-at-Risk (VaR) is a measure of the potential loss in value of a portfolio over a given time period and was first introduced by J.P. Morgan (1996). Currently it is widely adopted by financial risk managers. The usual time horizon used for calculating VaR is 1 to 10-days. VaR can be defined as a single estimate, and is calculated such that the probability of a N-period return smaller than the VaR is equal to α.

$$\alpha = P(R_N < VaR) \tag{43}$$

where R_N is the N-period return and the negative sign is due to the convention of reporting VaR as a positive number, i.e., loss. The VaR itself can be calculated by rearranging this equation to,

$$VaR(\alpha, N) = F_N^{-1}(\alpha) \tag{44}$$

Where F_N^{-1} is the inverse distribution of the N-period return and the VaR is equal to the α% quantile of the distribution. There is no universal method to calculate VaR thus we consider two methods: the Historical VaR and Modified VaR.

The Historical method assumes that the portfolio returns in the future will follow the same pattern as they did in the past. The rate of returns are calculated from the data that we have and then we organize these returns from worst to best in a histogram. For the given confidence level, 95% or 99% is the most common confidence levels, we look for the worst 5% or 1% respectively of the outcomes. Then we can say that the loss for a given period of time will not exceed this worst outcome with probability 95% or 99%. The weakness of the historical approach is that it relies on the assumption that history will repeat itself which is far from the truth.

The Modified VaR, also known as Cornish-Fisher VaR, is an alternative approach to calculate VaR. If the return of a portfolio is not Gaussian distributed then the classical VaR method is no longer an efficient measure of risk. This method takes into account the higher moments, skewness and kurtosis by utilizing the Cornish-Fisher Expansion of Cornish and Fisher (1937). When the returns have negative skewness or fat-tails the Cornish-Fisher VaR will give a larger estimation for the loss than the usual VaR. On the other hand, when returns possess positive skewness or are leptokurtic, the loss estimation will be smaller than traditional VaR.

Expected Shortfall (ES) Expected Shortfall (ES), also known as Conditional Value-at-Risk, is an alternative measure to Value-at-Risk and has been proposed by Basel Committee banking regulation. VaR tells us the loss at a particular quantile q. It therefore tells us nothing about what the distribution looks like below q. ES on the other hand gives the average loss in the tail below q and answers the question What is the expected loss if things do get bad?. The common definition for Expected Shortfall is:

$$ES_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_\beta(X)d\beta \tag{45}$$

The main difference between VaR and ES is that while ES is always a coherent measure of risk, VaR sometimes fails the property of subaddivity which means that the risk in a diversified portfolio is higher than in an undiversified portfolio. Similar to the VaR we consider the historical ES and Cornish-Fisher ES.

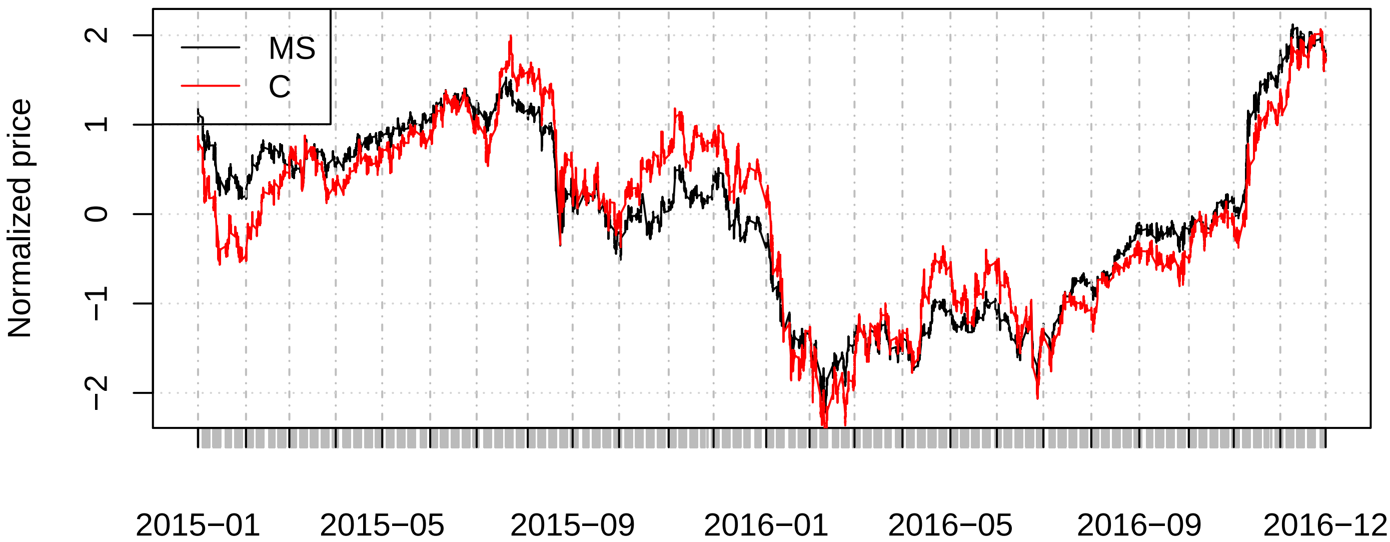
6 Empirical Analysis

This chapter provides empirical evidence of the dynamic copula pairs trading strategy on high frequency data. Our analysis is divided into two parts. In part one we analyze the performance of the dynamic copula trading strategy on a single stock pair in order to extensively outline our estimation and parameter selection approach. Then, in the second part we perform a sector wide analysis by constructing four portfolios consisting of stock pairs with the highest dependence criteria. In this section we mainly focus on the risks and returns of the strategy.

6.1 Single Pair Analysis

In this section we analyze the performance of the copula pairs trading strategy for a single stock pair. For our analysis we consider the stocks of Morgan Stanley (MS) and Citigroup (C) which are both part of the financial sector, traded on the NYSE, and very similar in terms of their business operations. The time periods considered are January 2, 2015 until December 30, 2015 (formation period) and January 2, 2016 to December 31, 2016 (trading period). To show the correlation between the price movements of the stocks, a plot of the normalized prices can be found in Figure 3.

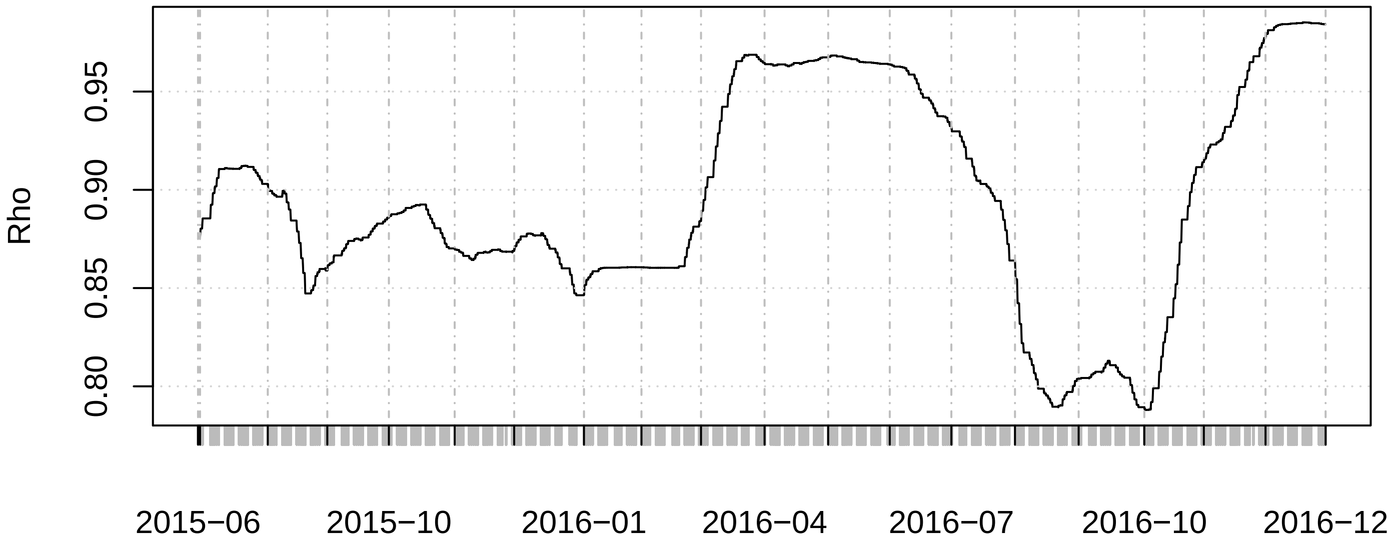
Figure 3: Morgan Stanley (MS) - Citigroup (C) 1-minute normalized stock prices



Note: Stock price normalization performed by subtracting the mean of the prices and dividing it by the standard deviation of the stock prices.

The figure shows that the two stock prices move very similar during both time periods. In addition, the Spearman's ρ correlation coefficient between the two stocks during the formation period is 0.794, and 0.958 during the trading period. In Figure 4 we represent the rolling 48000 minutes Spearman's ρ correlation between the two stocks. The y-axis indicates the last measured point for calculating the correlation.

Figure 4: Morgan Stanley (MS) - Citigroup (C) rolling Spearman's ρ correlation



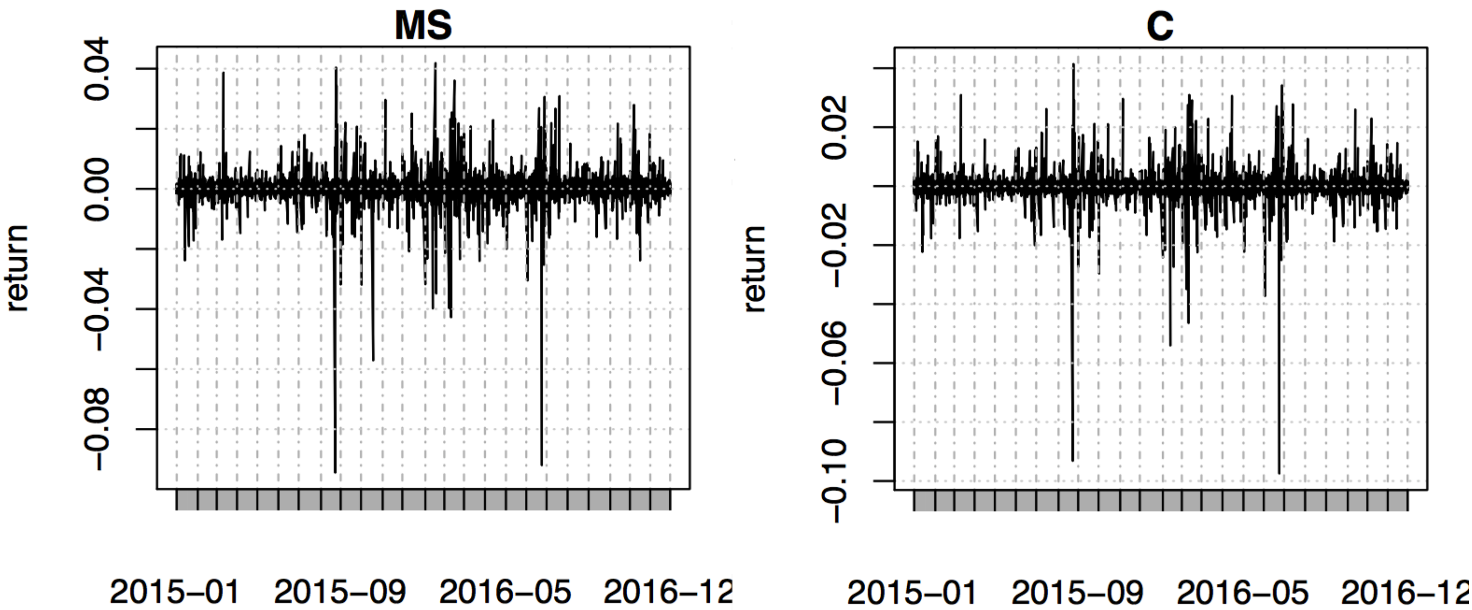
Note: Estimation window is 48000 minutes, y-axis indicates the last measured Spearman's ρ correlation.

The graph shows that the dynamic dependence is relatively stable and fluctuates between 0.75 and 0.99. Hence, given the overall correlation during the formation and trading period and the high rolling correlation we can conclude that this pair exhibits the dependence needed for our dynamic trading strategy.

6.1.1 Uniform Transformation

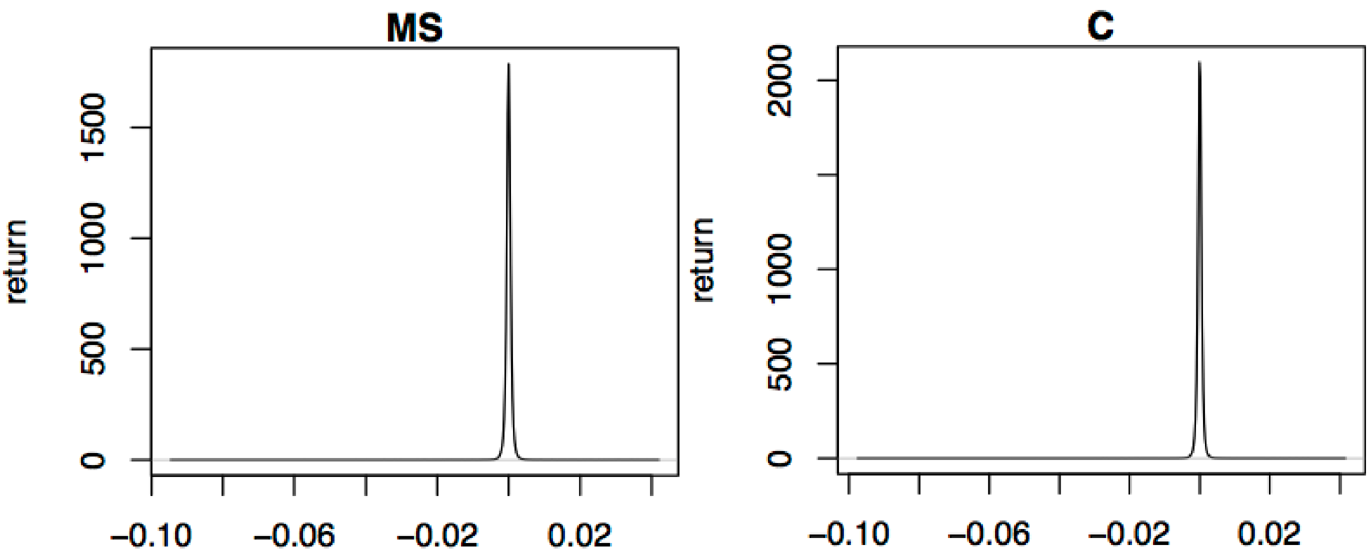
Our first objective is to find a proper model specification for the individual stock returns that can subsequently be used for the uniform transformation. In Figure 5, 6 and 7 we present the log-returns, density of the log-returns and the autocorrelation function of the log-returns for both stocks respectively.

Figure 5: Morgan Stanley (MS) and Citigroup (C) log-returns



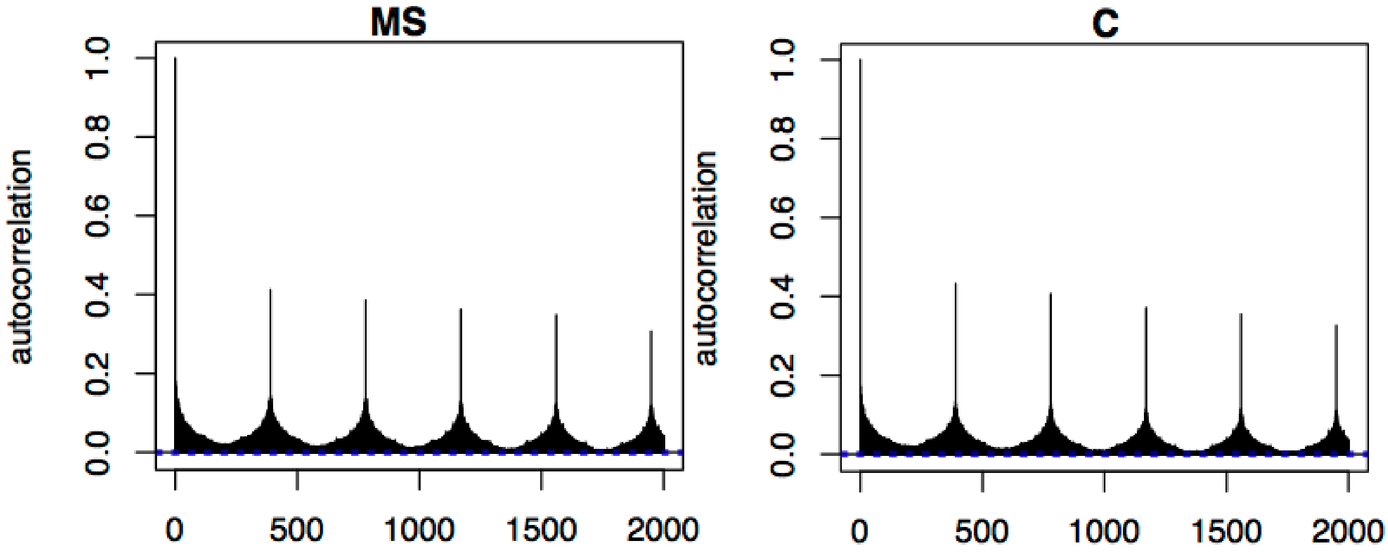
Note: Log-returns of the complete sample period for both stocks.

Figure 6: Morgan Stanley (MS) and Citigroup (C) density of log-returns



Note: Density estimated using kernel density estimation

Figure 7: MS & C autocorrelation functions of absolute log-returns

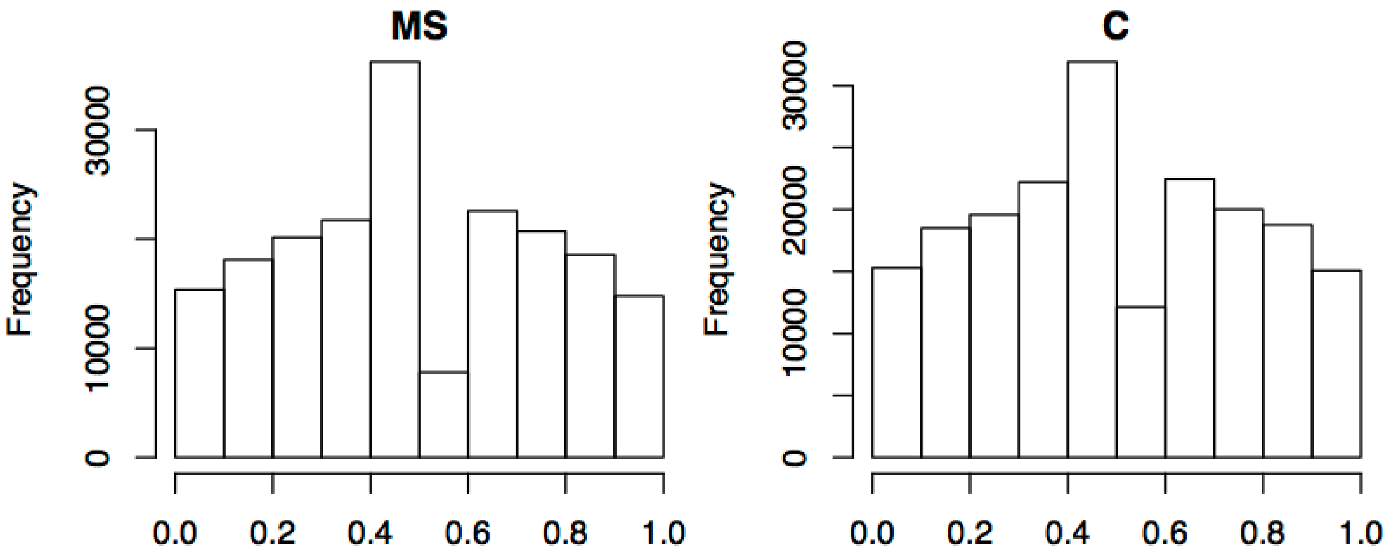


Note: First 2000 lags of the autocorrelation function of the absolute log-returns.

By inspection of these figures we observe that both stocks possess similar characteristics. Firstly, the data exhibits volatility clustering where high returns are followed by low returns. Secondly, the return distributions are fat tailed and highly peaked, i.e. leptokurtic. And finally, the data exhibits a strong presence of seasonality due to the repeating pattern of approximately 390 minutes (1 trading day) in the autocorrelation function. Thus, a proper model specification for the individual stock returns must incorporate these characteristics. Hence, we adopt the modeling approach by Engle & Sokalska (2012).

For our uniform transformation we estimate a MC-EGARCH model on the first 48000 minutes and then forecast the daily volatility 1 day ahead and the intraday volatility 390 minutes ahead (1 trading day), with the forecasted daily and intraday volatility we transform the returns to uniform scale using the inverse student's t-distribution. This process is performed rolling until the complete sample is transformed to uniform. In 8 we present the resulting histogram of both stocks after applying the transformation. Both histograms show that the stock returns are between 0 and 1, as required.

Figure 8: Morgan Stanley (MS) - Citigroup (C) Histogram of returns after uniform transformation



Note: Resulting histogram of the MC-EGARCH uniform transformation method using 1-day-ahead forecasting for the daily volatility component, 390-minutes-ahead forecasting for the intraday volatility component and basing the diurnal component on the first 48000 minutes of the sample. The innovations are modeled using a student's t-distribution.

6.1.2 Copula Selection

In order to select the best copula function we make use of the procedure as outlined in Section 4.3. We estimate the copula functions using a look-back period that is in accordance with the uniform transformation, i.e. 48000 observations. However, we fix the copula during the estimation period based on the copula that was found to be the overall best fitting copula during this period. In Table 3 we present the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values of the copulas during the estimation period.

Table 3: Copula estimation goodness-of-fit results

	AIC	BIC
Archimedean Copulas		
Clayton	-22785.79	-22777.01
Frank	-23843.57	-23834.8
Gumbel	-27088.87	-27080.1
Elliptical Copulas		
Gaussian	-26428.71	-26419.93
Student's t	-30117.52	-30099.97

Note: Copula estimation goodness-of-fit results using the first 48000 observations of our sample.

As mentioned in Section 4.3 the copula with the lowest AIC and BIC is the preferred model for modeling the joint stock distribution. The results show that the Student's t-copula is the best fitting Elliptical copula and also the overall best fitting copula with an AIC of -30117.52 and BIC of -30099.97. Furthermore, we find that the Gumbel copula is the best fitting Archimedean copula, and ranks second best as the overall best fitting copula with an AIC and BIC of -27088.87 and -27080.1 respectively. However, it is clear that the Student's t-copula is the superior copula and thus we select this copula for modeling the joint distribution of both stocks.

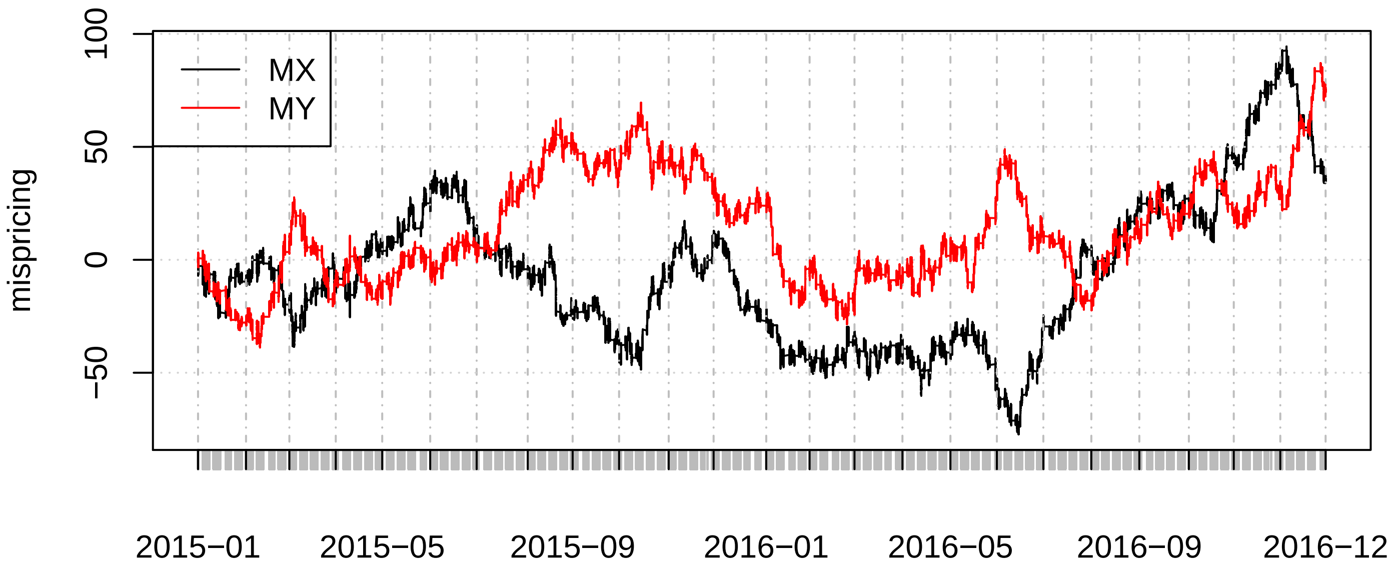
Similar to the uniform transformation we re-estimate the copula parameters every 390 minutes with a look-back window of 48000 minutes, implying that we update the copula parameters at the start of every trading day. Updating our parameters every minute is unfeasible in our case since this would drastically increase the computation time of the strategy during the trading days.

6.1.3 Mispricing Index

With the optimal copula selected we can construct the mispricing indices as shown in Section 5.1.3. In Section 5.3 we outlined how the trading strategy depends on the behavior of the mispricing index, i.e. mean-reversion and/or -diversion. In this section we investigate the behavior of our mispricing indices.

In Figure 9 we present a plot of the two mispricing indices for the pseudo-trading and actual trading period. We refere to the mispricing index of MS as MX or index X and to the mispricing index of C as MY or index Y. The figure shows that both indices exhibit mean-reversion from January 2, 2015 until July 1, 2016. From July 1, 2016, until December 30, 2016, they start to trend upwards. In addition, the overall Hurst Exponent of index X is 0.477 and the Hurst Exponent of index Y is 0.478. Hence, even though the indices are mean-reverting they are only slightly mean-reverting given that the lower the Hurst Exponent the more mean-reverting our series is. Hence, we impose a stop-loss criteria in order to force the mispricing indices to become more mean-reverting and subsequently limit the losses of the trading strategy.

Figure 9: Morgan Stanley (MS) - Citigroup (C) mispricing indices constructed using the conditional student's t-copula



Note: MX refers to the mispricing index of MS and MY refers to the mispricing index of C

6.1.4 Pseudo-trading Period

The pseudo-trading period is used to determine the optimal trading parameters S and D, i.e. our stop-loss criteria and trade-enter trigger point. Selecting the parameters can be done using one of the preferred performance evaluation criteria of Section 6.5, each having their own pros and cons. In our case we optimize our strategy based on the total yield produced versus the number of round-trips needed. To find the optimal S and D we make use of a brute-force optimization method that evaluates our strategy for all $S \in (0.1, 3.0)$ and $D \in (-0.5, min\{5, S\})$ with increments of 0.1. In Table 4 and 5 we present a part of the cumulative returns and number of round-trip optimization results. For the computation of the cumulative returns we assume an equal weighted position in both stocks, for example 1 short and 1 long position.

We observe that the maximum cumulative return of 126 % is attained with $S = 0.9$ and $D = -0.2$. For this return, a total of 11543 round-trips have been made. The second highest cumulative return of 124% is generated with $S = 1.3$ and $D = -0.2$ with a total of 9609. Thus, approximately 2000 round-trips less, but only a 2% decrease in return. Then, the third highest return is 123% with $S = 1.7$ and $D = -0.2$ and with 7720 round-trips. Thus, another 2000 round-trips less but only a 1% decrease in return. Hence, by increasing the stop-loss point the number of trades is decreased but the profitability remains stable as long as the trade entry point is kept constant. Given that the number of trades will have a significant impact on the profitability of our strategy we opt for the trading parameters that needs the least number of round-trips while still producing a high yield. Hence, we select $S = 1.7$ and $D = -0.3$ as our out-of-sample trading parameters.

Table 4: Pseudo-trading 6-months cumulative returns table

D \ S	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
-0.5	65%	71%	81%	81%	82%	77%	76%	70%	69%	69%	78%	77%	75%
-0.4	96%	102%	112%	105%	96%	94%	90%	90%	99%	93%	107%	98%	98%
-0.3	103%	98%	113%	97%	96%	104%	114%	115%	113%	109%	113%	104%	99%
-0.2	102%	95%	126%	119%	113%	119%	124%	120%	122%	114%	123%	117%	103%
-0.1	102%	96%	116%	111%	111%	117%	120%	107%	119%	110%	103%	90%	94%
0	80%	75%	102%	111%	97%	99%	97%	84%	99%	87%	100%	96%	96%
0.1	74%	82%	75%	73%	52%	77%	78%	66%	64%	57%	62%	76%	69%

Note: Returns are computed assuming an equal position in both stocks

Table 5: Pseudo-trading 6-months number of round-trips table

D \ S	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
-0.5	10523	9818	9208	8752	8372	7993	7648	7328	7030	6729	6541	6346	6155
-0.4	12046	11055	10226	9546	9062	8602	8170	7809	7497	7206	7044	6763	6513
-0.3	13190	11991	10997	10158	9654	9110	8652	8291	7886	7524	7323	7002	6789
-0.2	13791	12596	11543	10715	10171	9609	9078	8764	8323	7969	7720	7421	7145
-0.1	13660	12666	11634	10805	10157	9625	9153	8764	8292	7917	7660	7361	7142
0	13183	12101	10989	10364	9685	9379	8819	8362	8145	7762	7461	7212	7016
0.1	12770	11458	10249	9331	8505	8044	7593	6993	6662	6247	5889	5677	5467

Note: A single round-trip consists of 4 trades, one long/short to enter a postion and one short/long to exit a position

6.1.5 Trading results

In the previous section we selected the optimal trading parameters by backtesting our strategy on 6 months of data. Now, we will investigate the trading performance out-of-sample using the last 12 months our data. Similar to the pseudo-trading results we assume an equal weighted position in both stocks for the computation of the returns. The daily return characteristics of Table 6 show that on average our strategy produces a return of approximately 1% per day with a standard deviation of 0.011. Furthermore, the returns are positively skewed and have excess kurtosis implying that the returns are peaked and fat-tailed, i.e. leptokurtic.

To produce these results the strategy performed a total of 14970 round-trips as shown in Table 7. The average holding period was 6 minutes, the maximum holding period was 93 minutes, and the minimum holding period was 1 minute. Implying that the strategy is trading on a high to mid frequency. Approximately 77% of the trades produced a positive return and 23% produced a negative return.

Table 6: Return characteristics

	Minute	Daily
Mean	0.000027	0.0107
Minimum	-0.023	-0.250
Maximum	0.029	0.0625
Standard Deviation	0.000619	0.0119
Skewness	1.853	0.512
Kurtosis	134.672	2.175

Table 7: Quantity of trades characteristics

Total Number of RT	14970
Number of Winning RT	11525
Number of Lossing RT	3445
Percentage of Winning RT	0.769873
Percentage of Loss RT	0.230127
Mean Hold (min)	6
Min Hold (min)	1
Max Hold (min)	93

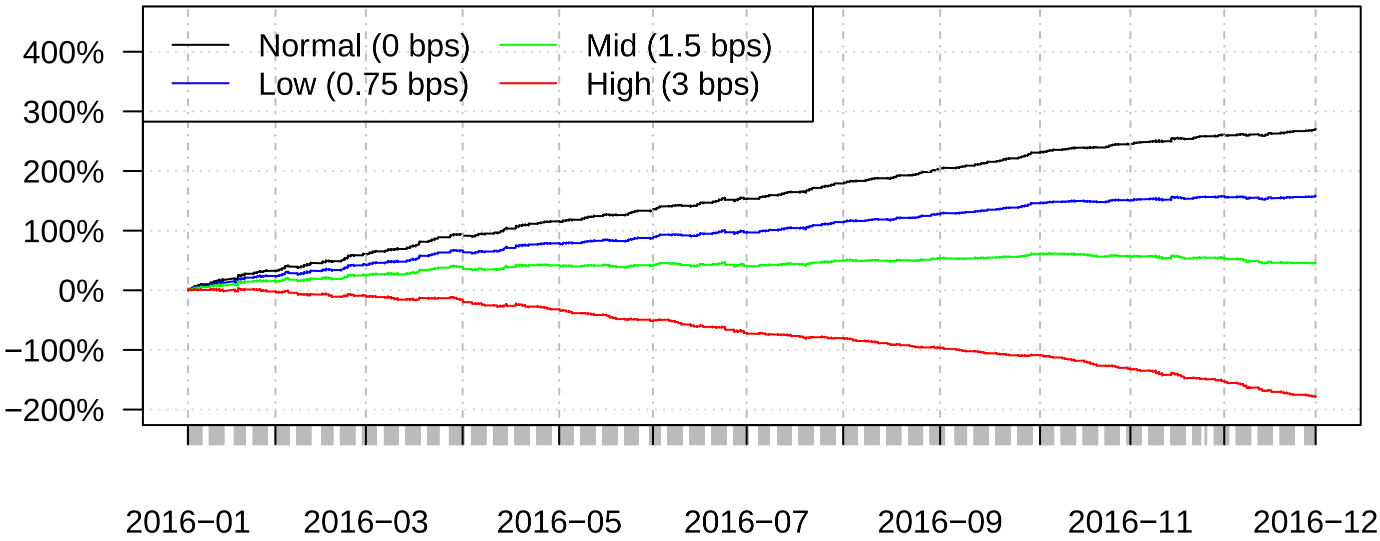
In Table 8 we present the annualized risk-return statistics of our strategy with and without transaction costs. Without transaction costs, our strategy produces a high cumulative annual return of 270% with a Sharpe Ratio of 13.9 and Sortino Ratio of 21.1. Given our trading frequency and return characteristics these high Sharpe and Sortino ratios are not surprising. However, when including transaction cost of just 0.75 bps per round-trip the annual return drops to 158%, the Sharpe Ratio to 8.2 and Sortino Ratio to 12.3. Still a relatively strong performance, but it makes clear that transaction costs significantly impact the profitability of the strategy. If we further increase the cost of trading to 1.5 bps or even 3 bps per round-trip the annual return, Sharpe and Sortino ratios decrease further and even become highly negative. Hence, the dynamic copula method is able to identify profitable trading opportunities but transaction costs have a significant impact on the actual profitability of the strategy. In Figure 15 we present the evolution of the equity curves for each transaction cost profile. This figure also visualizes the observed maximum drawdown for each transaction cost profile from Table 8.

Table 8: Annualized return statistics

	Normal	Low Adjusted	Mid Adjusted	High Adjusted
Cumulative Return	270%	158 %	46%	-178%
Sharpe Ratio	13.9	8.2	2.4	-9.4
Sortino Ratio	21.1	12.3	3.6	-13.6
Maximum Drawdown	5.9%	6.5%	16.0 %	84.0%

Note: Low, Mid, and High include 0.75, 1.5 or 3 basis points per round-trip respectively, returns are computed assuming an equal position in both stocks

Figure 10: Morgan Stanley (MS) - Citigroup (C) dynamic copula pairs trading method results with transaction cost profiles included



To investigate the effect of market impact we present the return characteristics when the wait-one-bar rule is imposed in Table 9. In this case the executed price is the price of the next bar and not the current bar. We observe that with the rule imposed, the mean daily returns drops with more than 50% to 0.49%, thus the rule has a significant effect on the profitability of our strategy. This implies that besides transaction costs, the profitability of the strategy is also heavily influenced by the speed of execution and consequently the price for which it is executed.

Table 9: Return characteristics with wait-one-bar rule imposed

	Minute	Daily
Mean Return	0.00001	0.0049
Minimum Return	-0.0236	-0.0286
Maximum Return	0.029	0.0389
Standard Deviation Return	0.0006	0.0112
Skewness Return	0.6937	0.113
Kurtosis Return	107.98	0.508

Hence, we can conclude for this particular stock pair, the dynamic copula trading method is able to identify profitable arbitrage opportunities during the intraday market period on the 1-minute frequency. However, the profitability of the strategy is heavily influenced by transaction cost and market impact.

6.2 Industry Analysis

In this section we investigate the performance of the dynamic copula trading strategy on high frequency data for a large set of pairs across sectors. We examine the performance of our trading strategy for stocks that are matched only within four large sector groupings: Financials, Industrials, Technology and Energy. Within each sector we select the most liquid traded stocks and calculate their correlation during the speudo-trading period. From this we select the top-5 pairs with the highest correlation of each sector, which gives us a total of 20 pairs. We evaluate the performance of each sector portfolio based on their return distribution, risk-return characteristics, Value at Risk, drawdown and exposure to market risk. As a benchmark we compare the strategy to a S&P 500 Buy-and-Hold strategy.

6.2.1 Portfolio construction

To construct each portfolio we select five pairs from each sector with the highest Spearman's ρ correlation coefficient. In Appendix B we present the correlation matrices of each sector. In Table 10 we present the correlation coefficients and trade-entry (D) and stop-loss (S) parameters used for the trading period for each selected pair. Note that an individual stock can be part of different pairs and thus is traded multiple times in the portfolio. We observe that each pair had a correlation of at least 0.81 during the formation period, i.e. in the year 2015. Thus, each pair was at least highly correlated during the formation period. Furthermore, our backtest results shows that each pair has a trade-entry point and stop-loss trigger in the neighborhood of -0.2 and 1.5 respectively.

Table 10: Sector Portfolios pairs Spearman's ρ correlation coefficients and trade entry (D) and stop-loss (S) parameters

Pair	ρ	S	D	Pair	ρ	S	D
Panel A: Financials				Panel C: Technology			
JPM-AIG	0.974	1.7	-0.2	FB-ADBE	0.889	1.7	-0.3
JPM-GS	0.956	1.4	-0.3	CRM-ADBE	0.876	1.7	-0.3
JPM-C	0.950	1.5	-0.3	QCOM-ORCLE	0.853	1.5	-0.1
BK-AIG	0.949	1.2	-0.3	ORCL-AMAT	0.818	1,3	-0.2
GS-AIG	0.948	1.2	-0.4	EBAY-ADBE	0.816	1.3	-0.3
Panel B: Industrials				Panel D: Energy			
EMR-CAT	0.916	1.6	-0.2	NBL-APA	0.962	1.4	-0.1
UTX-MMM	0.912	1.9	-0.2	SPY-HAL	0.962	1.5	-0.2
MMM-ITW	0.907	1.9	-0.1	DVN-APC	0.958	1.5	-0.2
CSX-EMR	0.890	2.3	-0.2	COP-APA	0.957	1.5	-0.3
CSX-CAT	0.889	1.4	-0.3	EOG-COP	0.954	1.3	-0.1

Note: Portfolios are constructed by selecting the top-5 pairs with the highest Spearman's ρ correlation within their respective sector during the formation period

For the computation of the portfolio returns we assume that an equal weight of each pair in the portfolio where an equal number of stocks is traded for each stock in the pair. The overall return is then computed by aggregating the returns of each pair over the trading period. Thus, each pair is considered independently and has an equal weight in the portfolio.

6.2.2 Copula selection

In this section we select a single copula function from our universe of copulas for the modeling of the joint stock returns. Specifically, we compare the Archimedean copulas (Clayton, Frank, Gumbel) and the two Elliptical copulas (Gaussian and Student's t) regarding the goodness of fit to our uniform transformed data using a MC-EGARCH(1,2)-std model (see Section 5.1). In Table 11 we rank each copula based on their average AIC and BIC values of all the considered pairs. The test statistics show that the student's t-copula is ranked first in terms of the average AIC and BIC, the Gaussian copula is ranked second and the Gumbel, Frank and Clayton copula third, fourth and fifth respectively. Hence, we select the student's t-copula as the preferred copula to model the joint distribution of each stock pair.

Table 11: Copula goodness-of-fit estimation results for all portfolio pairs

	Avg. AIC	Avg. BIC
Archimedean Copulas		
Clayton	-54658.72	-54648.03
Frank	-64533.32	-64522.65
Gumbel	-65222.65	-65211.95
Elliptical Copulas		
Gaussian	-68157.31	-68146.61
Student's t	-70394.92	-70374.55

Note: The average AIC and BIC are computed by summing all AIC and BIC values of each pair and then dividing them by 20

6.2.3 Return characteristics and trading statistics

This section is devoted to the return characteristics and trading statistics produced by each portfolio during the trading period, i.e. the year 2016. Table 12 depicts the daily return characteristics of each industry portfolio. We observe that each portfolio produces a significant economical mean returns of at least 3% per day. Furthermore, we also find that these returns statistically significant with Newey-West (NW) t-statistics of above 6. Then, each portfolio has a return distribution with positive skew and excess kurtosis compared to a normal distribution, implying that the returns are leptokurtic. Notably the returns of the Technology portfolio are highly peaked and fat-tailed with an excess kurtosis of 14.4 and skewness of 1.1. Comparing the portfolios with the S&P 500 Buy-and-

Hold strategy we find that the mean-returns are a factor 10 times larger for the trading strategy compared to the benchmark. However, the benchmarks returns are negatively skewed with excess kurtosis implying that this strategy has a smaller probability of losing a lot and a high probability of earning a little. Finally, we find that between 29 and 58 days of the 225 trading days the portfolios produced a negative return versus 115 for the benchmark.

Table 12: Daily return characteristics of top-5 pairs portfolio of each sector versus S&P500 Buy-and-Hold

	Financials	Industrials	Technology	Energy	S&P500
Mean return	0.030	0.038	0.038	0.033	0.0003
Standard error (NW)	0.003	0.002	0.004	0.005	0.0005
t-Statistic (NW)	17.33	12.78	8.631	6.251	0.725
Maximum	0.154	0.217	0.381	0.214	0.024
Minimum	-0.054	-0.057	-0.162	-0.181	-0.037
Median	0.031	0.0372	0.0281	0.0312	0.0005
Standard Deviation	0.025	0.033	0.045	0.052	0.008
Skewness	0.183	0.612	1.197	0.200	-0.432
Kurtosis	2.328	2.853	14.393	1.783	2.305
Obs. with return < 0	30	29	31	58	115

In Table 13 we report the quantity of trade statistics for each portfolio. For each portfolio an average of approximately 75000 round trips need to be made in order to produce the results of Table 12. Of all the trades that are executed in each portfolio, approximately 72% where profitable and 28% where not.

Hence, from these tables we can conclude that our strategy can produce economically and statistically significant returns across sectors. However, this comes at the cost of a large number of trades that have to be executed on a short frequency and thus exposes the strategy to significant trading costs.

Table 13: Portfolio quantity of trade statistics

	Financials	Industrials	Technology	Energy
Total Number of RT	82419	72730	80255	70642
Number of Winning RT	59605	54407	58587	50523
Number of Loss RT	22814	18323	21668	20119
Percentage of Winning RT	72.3	74.8	73.0	71.5
Percentage of Losing RT	27.7	25.2	27.0	28.5

6.2.4 Value at Risk

To further investigate the tail risk of the returns Table 14 reports the daily Value at Risk (VaR) measures of each portfolio. We investigate two methods of VaR computation, the historical VaR and the Cornish-Fisher (CF) VaR. In addition, we also present the historical Expected Shortfall (ES) and Cornish-Fisher ES, which is the expected loss beyond the VaR. For more details on each method, see Section 6.5.4.

Based on the Historical 5% VaR we can conclude with 95% confidence that the daily loss will not exceed 0.79%, 1.23%, 2.29% or 3.97% for the Financials, Industrials, Technology and Energy portfolio respectively. Comparing this to our benchmark we can conclude that only the Technology and Energy portfolio exhibit more risk. However, when a loss does occur the largest loss that can occur are 2.3%, 2.7%, 7.8% and 7.8% for the Financials, Industrials, Technology and Energy portfolio respectively which is larger than the benchmark's Historical 5% ES of 2.2%. The Historical 1% VaR and ES display similar characteristics. But in this case the Financials and Industrials portfolio's Historical 1% VaR also exceeds the benchmark's Historical 1% VaR. Furthermore, the Technology and Energy have a Historical 1% ES of 12.6% and 13.0% respectively which is almost 4 times larger than the benchmarks Historical 1% ES of 2.8%.

The CF VaR and CF ES show that when accounting for the excess kurtosis and skewness in the return distribution, the 5% VaR and is smaller for the Financials, Industrials and Technology portfolio but larger for the Energy portfolio compared to the benchmark. However, the CF 5% ES is larger for the Financials and Energy portfolio, but smaller for the Industrials and Technology portfolio. In the case of the 1% CF VaR and CF ES we observe that all portfolios have a higher estimated compared to the benchmark.

Hence, based on these results we can conclude that the portfolios exhibit more tail risk compared to the S&P500 Buy-and-Hold benchmark.

Table 14: Daily Value at Risk measures of top-5 sector pairs portfolio versus S&P500 Buy-and-Hold

	Financials	Industrials	Technology	Energy	S&P 500
Historical VaR 5%	0.0079	0.0123	0.0229	0.0397	0.0130
CF VaR 5%	0.0086	0.0081	0.0064	0.0474	0.0138
Historical ES 5%	0.0230	0.0272	0.0782	0.0782	0.0198
CF ES 5%	0.0229	0.0164	0.0064	0.0748	0.0220
Historical VaR 1%	0.0352	0.0352	0.0755	0.0916	0.0243
CF VaR 1%	0.0383	0.0413	0.1545	0.1013	0.0254
Historical ES 1%	0.0492	0.0449	0.1265	0.1305	0.0289
CF ES 1%	0.0726	0.0932	0.1545	0.1630	0.0273

6.2.5 Annualized risk-return characteristics

In Table 15 we discuss the annualized risk-return ratios. These metrics provides us with insight of the return that is obtained per unit of risk taken. We observe that the Financials, Industrials, Technology and Energy portfolios produce an annual cumulative return of 850.5%, 972.9%, 960% and 848.1% respectively. Furthermore, the annualized Sharpe and Sortino ratios of each portfolio exceed 11 and 17 respectively. Comparing these results with the benchmark we can conclude that the portfolios significantly out-perform it. The S&P500 benchmark produces an annual cumulative return of just 8.7% with a Sharpe of 0.7 and Sortino of 0.9. The Omega and Upside Potential ratios display similar results, the Financials, Industrials, Technology and Energy portfolios have a Omega ratio of 27, 22.7, 10.9 and 5.9 respectively and Upside Potential ratio of 1.64, 2.2, 0.98, 1.19 respectively. The benchmark produces a Omega Ratio and Upside Potential Ratio of 1.13 and 0.65 respectively. Hence, the portfolios also out-perform the benchmark in terms of these two risk-return measures. These result are not surprising given our trading frequency, the high mean returns and low risk of the portfolios that where observed in the previous sections.

Table 15: Annualized Risk-Return of portfolios versus S&P500 Buy-and-Hold

	Financials	Industrials	Technology	Energy	S&P 500
Cumulative Return	850.5 %	972.9	960.0%	848.1%	8.69%
Sharpe Ratio	20.6	18.7	14.4	11.5	0.6
Sortino Ratio	29.5	29.1	21.5	17.6	0.9
Omega Ratio (L = 0%)	27.0	22.7	10.9	5.9	1.1
Upside Potential Ratio (MAR=0%)	1.64	2.20	0.98	1.19	0.65

6.2.6 Drawdown measures

In this section we investigate the downside risk characteristics of each portfolio. In Table 16 we display the maximum drawdown (MDD) measures for each portfolio. Besides historical MDD we also consider the Monte-Carlo maximum drawdown (MC-MDD).

The MC-MDD is determined as follows: First, we randomly sample the order of each trade without replacement, with the newly ordered trades we construct a new equity curve and from the new equity curve we compute the maximum drawdown. We repeat this procedure 10000 times and take the 95% quantile of the resulting distribution to find the 95% MDD. The resulting value is our MC-MDD at 95% level.

From the historical MDD we can conclude that during the trading period we have observed a maximum drawdown of 3.9%, 6.5%, 9% and 11.4% for the Financials, Industrials, Technology and Energy portfolio respectively. The benchmark historical MDD was 10.5% and thus only the Energy portfolio produced a larger drawdown.

The Monte Carlo MDD lets us scramble the order of the trades in the backtest to provide a better understanding of possible future performance, based on the assumption that future trades will have similar characteristics to historical trades but in an unknown order. Similar to VaR, the 95% MC-MDD tells us that out of our 10000 simulations, 95% had a better outcome than the MC-MDD value. For the portfolios, the MC-MDD range between 7.6% and 14.4%, thus larger than the hisorical MDD but smaller than the benchmark MC-MDD of 15.1%.

Then, the Calmar ratios show us that the Financials, Industrials, Technology and Energy portfolio are able to cover the MDD 212, 148, 106 and 74 times respectively. Implying that the portfolios were able to recover from a drawdown quickly. The results are similar for the MC-MDD with 97, 128, 67 and 58 times respectively. Thus, we expect that the strategy will also be able to recover quickly from future drawdowns. The benchmark is not even able to cover the historical MDD or MC-MDD once, further underlining the superior performance of the dynamic copula trading strategy.

Table 16: Drawdown measures and ratios of portfolios and S&P500 Buy-and-Hold

	Financials	Industrials	Technology	Energy	S&P 500
Historical MDD	0.039	0.065	0.090	0.114	0.105
MC-MDD 95%	0.087	0.076	0.143	0.144	0.151
Calmar Ratio Historical MDD	212.9	148.7	106.6	74.0	0.87
Calmar MC-MDD 95%	97.0	128.4	67.3	58.8	0.60

6.2.7 Transaction costs

In Section 7.2.2 we showed that each portfolio relies on approximately 75,000 trades to produce its return characteristics. In Section 6.4 we discussed that each trade comes at the cost in the form of commission and short-selling fees, i.e. transaction costs. Hence, in this section we evaluate the robustness of the trading strategy in light of these trading costs. We consider three profiles for the transaction costs, namely Low, Mid and High as outlined in Section 6.4 and in Table 17 we present the transaction cost adjusted return characteristics for each portfolio. The transaction cost adjusted return characteristics show that once transaction costs are factored into the equation, the profitability of the strategy drops significantly. When only adjusting the returns with the Low transaction cost profile, the mean returns of the Financials, Industrials, Technology and Energy portfolio drop from 3%, 3.8%, 3.8% and 3.3% to 0.9%, 1,6%, 1.5%, 1,2%. Thus, a decrease of 50% to 66%. When adjusting the returns with the Mid and High transaction costs we see that our each portfolio produces significant negative mean returns, Sharpe ratios, Sortino ratios and consequently cumulative returns. Hence, our trading strategy is not robust against the Mid to High transaction cost profile. Implying that using the dynamic coplua trading method on the 1-minute frequency is only suitable for investors that are able to achieve at least the Low transaction cost profile. However, in the previous section concerning the single pair analysis we found that this particular pair was able to produce a positive return despite the Mid transaction costs. This gives evidence that the construction of portfolios could be improved by taking the trading statistics obtained during the formation period into account when selecting suitable pairs for trading.

Table 17: Trading cost adjusted return characteristics

TC profile	Low	Mid	High
Panel A: Financials			
Mean return	0.009	-0.012	-0.064
Standard error (NW)	0.002	0.002	0.002
t-Statistic (NW)	5.37	-7.97	-32.93
Cumulative return	232.0%	-385.0%	-1621.8%
Sharpe Ratio	5.7	-9.4	-39.9
Sortino Ratio	7.8	-12.7	-50.7
Panel B: Industrials			
Mean return	0.016	-0.0046	-0.047
Standard error (NW)	0.002	0.002	0.002
t-Statistic (NW)	5.76	-1.62	-17.49
Cumulative return	427.6%	-117.3%	-1208.5%
Sharpe Ratio	8.3	-2.3	-23.6
Sortino Ratio	12.6	-3.4	-33.6
Panel C: Technology			
Mean return	0.015	-0.007	-0.053
Standard error (NW)	0.004	0.004	0.004
t-Statistic (NW)	3.47	-1.82	-12.84
Cumulative return	381.2 %	-197.1%	-1354.4%
Sharpe Ratio	5.7	-2.9	-20.5
Sortino Ratio	8.5	-4.3	-28.9
Panel D: Energy			
Mean return	0.012	-0.0083	-0.050
Standard error (NW)	0.004	0.004	0.003
t-Statistic (NW)	2.54	-1.83	-12.88
Cumulative return	318.9%	-211.2%	-1270.0%
Sharpe Ratio	4.3	-2.8	-17.3
Sortino Ratio	6.5	-4.3	-25.0

6.2.8 Market friction

In this section we investigate the effect of market friction on each portfolio. In Table 18 we present the trading results of each portfolio when the wait-one-bar rule is imposed. Comparing the wait-one-bar trading results with the previous results we observe that the mean return of each portfolio drops significantly. The Financials portfolio drops from 0.03 to 0.006, the Industrial portfolio from 0.038 to 0.019, the Technology portfolio from 0.038 to 0.012 and the Energy portfolio from 0.033 to 0.006. However, the standard deviation of the returns remains fairly stable. Finally, the number of days with a loss has increases with a factor of 2 to 3. Hence, we find evidence that delayed execution can have a significant impact on the profitability of the strategy in a real-life market situation.

Table 18: Daily return characteristics of top-5 pairs portfolio of each sector with wait-one-bar rule imposed

	Financials	Industrial	Technology	Energy
Mean return	0.006	0.019	0.012	0.006
Standard error (NW)	0.001	0.001	0.003	0.003
t-Statistic (NW)	5.682	10.108	3.209	2.198
Standard Deviation	0.022	0.033	0.038	0.047
Obs. with return < 0	91	60	78	116

6.2.9 Common risk factors

Finally, in this section we analyze the exposure of the strategies to common systematic sources of risk. For this we employ the Fama-French five-factor model (FF5), following Fama (2015). The model is defined as follows,

$$R_t - R_{Ft} = a + b(R_{Mt} - R_{Ft}) + s(SMB_t) + h(HML_t) + r(RMWt) + c(CMA_t) + \epsilon_t \quad (46)$$

In this model R_t is the daily portfolio return, R_{ft} is the risk-free return, R_{Mt} is the market return, SMB_t is the return on a diversified portfolio of small stocks minus the return of a diversified portfolio of big stocks, HML_t is the difference between the returns on diversified portfolios of high and low B/M stock, RMW_t is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and CMA_t is the difference between the returns on diversified portfolios of the stocks of low and high investment firms. The regression results for each portfolio can be found in Table 19. All data related to these models is downloaded from Kenneth R. Frenchs website.

Table 19: Portfolio exposures to systematic sources of risk

	Financial	Industrial	Technology	Energy
(Intercept)	0.029	0.037	0.037	0.032
	(0.001)***	(0.002)***	(0.002)***	(0.003)***
Market	-0.002	0.002	-0.006	-0.004
	(0.002)	(0.002)	(0.003)	(0.004)
SMB	0.001	0.001	0.001	-0.0007
	(0.001)	(0.003)	(0.005)	(0.006)
HML	-0.005	-0.001	0.010	-0.009
	(0.003)	(0.004)	(0.006)	(0.007)
RMW	-0.001	-0.004	-0.003	-0.002
	(0.004)	(0.006)	(0.008)	(0.009)
CMA	0.005	0.005	-0.023	-0.002
	(0.005)	(0.007)	(0.010)	(0.011)
R^2	0.018	0.011	0.026	0.019
Adj. R^2	-0.0017	-0.0088	0.0063	-0.0002
Num. obs	252	252	252	252

Note: $^{***}p < 0.01$, $^{**}p < 0.05$

Generally, the R^2 of each portfolio regression is small, implying that the factors are unable to explain much of the variation in the returns of our strategy. The regression intercept coefficient shows that each portfolio produces economically and statistically significant alphas of approximately 3% per day, which is similar to the raw returns we found of Table 12. Since our strategy is budget-neutral, as expected, we observe that our strategy has no statistically significant loading on the market implying that our strategy does not depend on the performance of the overall market. Furthermore, none of the portfolios have statistical significant loading different from zero on the SMB, HML, RMW or CMA factors. Implying that our portfolios are not exposed to stocks with small capitalization, stocks with low Book-to-Market ratios, firms with weak profitability or firms with high investments.

7 Conclusion

Pairs trading is generally defined as a statistical arbitrage strategy that capitalizes on the temporary relative mispricing between two instruments whose prices are expected to converge due to strong historical co-movements. Pairs trading is one of the most common statistical arbitrage strategies that is used by professional traders, institutional investors, and hedge funds. In this paper we investigate the performance of the dynamic copula pairs trading method on high-frequency intraday data. The conventional copula method assumes a static relationship for both the marginal and joint distribution of stock pairs. However, (high-frequency) financial data are known for exhibiting stylized facts such as volatility clustering and seasonality. As such, the conventional copula method will not accurately reflect the characteristics of the stock pairs.

We model the stock returns using a MC-GARCH model with an EGARCH model with students-t innovations for both the daily and intraday volatility components, and use a rolling window for the estimation of the copula parameters to account for dynamic dependency structures between stocks. We select suitable pairs based on their correlation during the formation period, optimize the trading parameters during the pseudo-trading period and evaluate our strategy during an out-of-sample trading period.

We apply the dynamic copula method on highly liquid stocks that are traded on the NYSE or NASDAQ and compare it to an S&P500 Buy-and-Hold strategy. For our main analysis we form four sector portfolios consisting of five stock pairs with the highest correlation coefficient during the formation period. Ignoring the cost of trading we find that each portfolio is able to produce an average daily return of at least 3% with Sharpe ratios between 11 and 20 and Sortino ratios between 17 and 29. In addition, the cumulative returns are able to cover the observed maximum drawdown between 74 and 148 times and the Monte Carlo maximum drawdown between 58 and 128 times. Implying that the portfolios exhibit little drawdown risk. In terms of tail-risk we found that the portfolios exhibit more tail risk compared to the benchmark strategy but given the superior mean returns of the strategy the tail-risk is relatively small. Finally, the Fama-French five-factor model regression results showed that non of the factors had a statistical significant loading on the returns. Thus, the portfolios also exhibit little systematic sources of risk.

However, to produce these results the strategy had to perform approximately 75,000 transactions for each portfolio. Implying that the method is able to identify many profitable arbitrage opportunities but is also exposed to a significant cost of trading. When imposing a transaction costs ranging from 0.75 bps, 1.5 bps or 3 bps per round-trip we observe that the average daily return drops significantly to approximately 1%, -1%, and -5% respectively. In addition, when imposing a wait-one-bar trading rule to account for possible delayed execution in a real market situation, returns declined with approximately 50%. Hence, only for market participants that are able to achieve little to no transaction costs and that are able to execute their trades fast and efficiently the dynamic copula trading method can produce significant economical returns while incorporating little risk.

Despite these results, there still exist a vast amount of areas for further studies. First, for the modeling of the marginal stock returns we assumed an EGARCH model with student's t-innovations for the MC-GARCH specification. Thus, we did not perform an extensive model selection procedure for the modeling of the stock returns. Finding a better MC-GARCH specification that for example incorporates realized mesures could improve the detection of arbitrage opportunities. Second, we fix the rolling estimation window of the MC-GARCH model and copulas at 48000 datapoints. Future studies could investigate the effect of changing the width of the look-back window. Third, we optimize our trade-entry and stop-loss parameters during the 6 months before the trading period and select the optimal values based on the ratio between cumulative returns and number of round-trips. Thus, more sophisticated parameter selection methods that incorporates risk, the cost of trading, and possible changes in the behavior of the mispricing indices (mean-reversion and/or -diversion) could potentially improve the profitability of the strategy, especially after adjusting for transaction costs. Fourth, our single stock and portfolio analysis showed that an individual stock is able to cover the Mid transaction cost profile of 1.5 bps per round-trip but the portfolios were not. Hence, a copula-based pair selection method as discussed by Krauss and Stubinger (2015) could improve the selection of suitable stock-pairs for the dynamic copula trading strategy.

References

1. Ang, A. and Chen, J. (2002). *Asymmetric correlations of equity portfolios.* Journal of Financial Economics, 63, 443-494

2. Andersen, T.G. and Bollerslev, T. (1997). *Intraday Periodicity and Volatility persistence in Financial Markets.* Journal of Empirical Finance, 4, 115-158.

3. Andersen, T.G. and Bollerslev, T. (1998). *Answering the skeptics: Yes, standard volatility models do provide accurate forecasts.* International Economic Review, 39, 885-905.

4. Bollerslev, T. (1986). *Generalized Autoregressive Conditional Heteroskedasticty.* Journal of Econometrics, 30, 307-327.

5. Bowen, D., Hutchinson, M.C. & O'Sullivan, N., (2010). *High Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution and Patterns in Returns.* Journal of Trading, 5, 31-38.

6. Bogomolov, T. (2013). *Pairs trading based on statistical variability of the spread process.* Quantitative Finance, 13, 14111430.

7. Cornish, E., and Fisher, R., (1937). *Moments and Cumulants in the Specification of the Distribution.* Revue de lInstitut International de Statistique, 5, 307-320.

8. Do, B. and Faff, R. (2012). *Are pairs trading profits robust to trading costs?* Journal of Financial Research, 35, 261287.

9. Engle R.F., Granger C. W. J., (1987). *Co-Integration and Error Correction: Representation, Estimation, and Testing.* Econometrica, 55, 251-276.

10. Engle, R.F. and Sokalska, M.E. (2012). *Forecasting Intraday Volatility in the US Equity Market. Multiplicative component GARCH.* Journal of Financial Econometrics, 10, 54-83.

11. Engle, R.F. (1982). *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.* Econometrica, 50, 987.

12. Ferreira, L. (2008). *New tools for spread trading. Futures: News, Analysis& Strategies for Futures, Options.* Derivatives Traders, 37, 3841.

13. Fama, E.F. and French, K.R. (2015). *A five-factor asset pricing model.* Journal of Financial Economics, 116, 1, 1-22

14. Gallant, A.R. (1981). *On The Bias in Flexible Functional Forms and an Essentially Unbiased Form*, Journal of Econometrics, 15, 211-45.

15. Gallant, A.R. (1982). *Unbiased Determination of Production Technologies*, Journal of Econometrics, 20, 285-323.

16. Gateve, E., Goetzmann, W.N., and Rouwenhorst, K.G., (2006). *Pairs Trading: Performance of a Relative Value Arbitrage Rule.*The Review of Financial Studies, 19, 797-827.

17. Gerig, A., Bicente, J., and Fuentes, M.A. (2009) *Model for Non-Gaussian Intraday Stock Returns.* Physical review. E, Statistical, nonlinear, and soft matter physics.

18. Hurst, H.E. (1951). *Long-term storage of reservoirs: an experimental study.* Transactions of the American Society of Civil Engineers, 116, 770-799.

19. J.P.Morgan. (1996). *RiskMetrics Technical Document(4th edition)*, Morgan Guaranty Trust Company, NewYork, 1-58.

20. Krauss, C. and Stübinger, J. (2015). *Nonlinear dependence modeling with bivariate copulas: Statistical arbitrage pairs trading on the S&P 100.* IWQW Discussion Paper Series, 15/2015.

21. Keating, C. and Shadwick, W. (2002). *A universal Performance Measure.* Journal of Performance Measurement, Spring, 59-84.

22. Miao, G.J., (2014). *High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach.* International Journal of Economics and Finance, 6.

23. Nelsen, R. B. (2006). An Introduction to Copulas. New York: Springer.

24. Nelson, D.B. (1991). *Conditional Heteroskedasticity in Asset Returns: A New Approach.* Econometrica, 59, 347-370.

25. Patton A.J. (2006). *Modeling Asymmetric Exchange Rate Dependence.* International Economic Review, 47, 527-556.

26. Rad, H., Rand Kwong Yew Low, R.K.Y., & Faff, R., (2016) *The profitability of pairs trading strategies: distance, cointegration and copula methods.* Quantitative Finance, 16:10, 1541-1558.

27. Sklar A. (1959). *Fonctions de Repartition a n Dimensions et Leurs Marges.* Publications de lInstitut de Statistique de LUniversite de Paris. 229-231.

28. Stander, Y., Marais, D., and Botha, I. (2013). *Trading strategies with copulas.* Journal of Economic and Financial Sciences, 6, 83107.

29. Sharpe, W.F. (1975). *Adjusting for Risk in Portfolio Performance Measurement.* Journal of Portfolio Management, Winter, 29-34.

30. Sortino, F. A., and Lee N.P. (1994). *Performance Measurement in a Downside Risk Framework.* Journal of Investing, 3, 3, 59-64.

31. Sortino, F., van der Meer, R. A. H. and Plantinga, A. (1999). *The Dutch triangle - A framework to measure upside potential relative to downside risk.* Journal of Portfolio Management. 26, 1, 50 - 58.

32. Vidyamurthy, G. (2004). *Pairs Trading, Quantitative Methods and Analysis*, John Wiley & Sons, Canada.

33. Xie, W. and Wu, Y., (2013). *Copula-based pairs trading strategy.*

34. Xie, W., Liew, R.Q., Wu, Y. and Zou, X., (2014). *Pairs Trading with Copulas.* The Journal of Trading, 11, 41-52.

35. Zhi, Z. Wenjun, X, Yuan, W., and Liming, X. (2017). *Dynamic Copula Framework for Pairs Trading.* Working paper.

Appendix A

List of stocks considered for the construction of pairs, orderd by industry.

Symbol	Name	Exchange
	Panel A: Financials	
AIG	AMERICAN INTERNATIONAL GROUP INC	New York Stock Exchange Inc.
AMT	AMERICAN TOWER REIT CORP	New York Stock Exchange Inc.
AXP	AMERICAN EXPRESS	New York Stock Exchange Inc.
BAC	BANK OF AMERICA CORP	New York Stock Exchange Inc.
BK	BANK OF NEW YORK MELLON CORP	New York Stock Exchange Inc.
BLK	BLACKROCK INC	New York Stock Exchange Inc.
C	CITIGROUP INC	New York Stock Exchange Inc.
CB	CHUBB LTD	New York Stock Exchange Inc.
GS	GOLDMAN SACHS GROUP INC	New York Stock Exchange Inc.
JPM	JPMORGAN CHASE & CO	New York Stock Exchange Inc.
MA	MASTERCARD INC CLASS A	New York Stock Exchange Inc.
MET	METLIFE INC	New York Stock Exchange Inc.
MS	MORGAN STANLEY	New York Stock Exchange Inc.
PNC	PNC FINANCIAL SERVICES GROUP INC	New York Stock Exchange Inc.
SCHW	CHARLES SCHWAB CORP	New York Stock Exchange Inc.
SPG	SIMON PROPERTY GROUP REIT INC	New York Stock Exchange Inc.
USB	US BANCORP	New York Stock Exchange Inc.
V	VISA INC CLASS A	New York Stock Exchange Inc.
	Panel B: Industrials	
ADP	AUTOMATIC DATA PROCESSING INC	NASDAQ
BA	BOEING	New York Stock Exchange Inc.
CAT	CATERPILLAR INC	New York Stock Exchange Inc.
CSX	CSX CORP	NASDAQ
EMR	EMERSON ELECTRIC	New York Stock Exchange Inc.
FDX	FEDEX CORP	New York Stock Exchange Inc.
GD	GENERAL DYNAMICS CORP	New York Stock Exchange Inc.
GE	GENERAL ELECTRIC	New York Stock Exchange Inc.
HON	HONEYWELL INTERNATIONAL INC	New York Stock Exchange Inc.
ITW	ILLINOIS TOOL INC	New York Stock Exchange Inc.
JCI	JOHNSON CONTROLS INTERNATIONAL PLC	New York Stock Exchange Inc.
LMT	LOCKHEED MARTIN CORP	New York Stock Exchange Inc.
MMM	3M	New York Stock Exchange Inc.
NOC	NORTHROP GRUMMAN CORP	New York Stock Exchange Inc.
PYPL	PAYPAL HOLDINGS INC	NASDAQ
RTN	RAYTHEON	New York Stock Exchange Inc.
UNP	UNION PACIFIC CORP	New York Stock Exchange Inc.
UPS	UNITED PARCEL SERVICE INC CLASS B	New York Stock Exchange Inc.

Panel C: Technology		
AAPL	APPLE INC	NASDAQ
ADBE	ADOBE SYSTEM INC	NASDAQ
AMAT	APPLIED MATERIAL INC	NASDAQ
CRM	SALESFORCE.COM INC	New York Stock Exchange Inc.
CSCO	CISCO SYSTEMS INC	NASDAQ
CTSH	COGNIZANT TECHNOLOGY SOLUTIONS COR	NASDAQ
EBAY	EBAY INC	NASDAQ
FB	FACEBOOK CLASS A INC	NASDAQ
HPQ	HP INC	New York Stock Exchange Inc.
IBM	INTERNATIONAL BUSINESS MACHINES CO	New York Stock Exchange Inc.
INTC	INTEL CORPORATION CORP	NASDAQ
MSFT	MICROSOFT CORP	NASDAQ
NVDA	NVIDIA CORP	NASDAQ
ORCL	ORACLE CORP	New York Stock Exchange Inc.
QCOM	QUALCOMM INC	NASDAQ
Panel D: Energy		
APA	APACHE CORP	New York Stock Exchange Inc.
APC	ANADARKO PETROLEUM CORP	New York Stock Exchange Inc.
COP	CONOCOPHILLIPS	New York Stock Exchange Inc.
CVX	CHEVRON CORP	New York Stock Exchange Inc.
CXO	CONCHO RESOURCES INC	New York Stock Exchange Inc.
DVN	DEVON ENERGY CORP	New York Stock Exchange Inc.
EOG	EOG RESOURCES INC	New York Stock Exchange Inc.
HAL	HALLIBURTON	New York Stock Exchange Inc.
KMI	KINDER MORGAN INC	New York Stock Exchange Inc.
MPC	MARATHON PETROLEUM CORP	New York Stock Exchange Inc.
NBL	NOBLE ENERGY INC	New York Stock Exchange Inc.
OXY	OCCIDENTAL PETROLEUM CORP	New York Stock Exchange Inc.
PSX	PHILLIPS	New York Stock Exchange Inc.
PXD	PIONEER NATURAL RESOURCE	New York Stock Exchange Inc.
SLB	SCHLUMBERGER NV	New York Stock Exchange Inc.
VLO	VALERO ENERGY CORP	New York Stock Exchange Inc.
WMB	WILLIAMS INC	New York Stock Exchange Inc.

8 Appendix B

Spearman's ρ correlation matrices of each sector.

Table 20: Financials Spearman's ρ correlation matrix

	AIG	AMT	AXP	BAC	BK	BLK	C	CB	GS	JPM	MA	MET	MS	PNSC	SCHW	SPG	USB	V	WFC
AIG		-0.691	-0.517	0.334	0.950	0.454	0.927	-0.445	0.948	0.824	0.853	0.974	0.824	0.718	0.911	-0.753	0.140	0.627	0.778
AMT			0.587	-0.187	-0.756	-0.279	-0.606	0.585	-0.630	-0.575	-0.580	-0.652	-0.575	-0.617	-0.763	0.688	0.000	-0.390	-0.599
AXP				0.234	-0.569	-0.395	-0.377	0.464	-0.529	-0.346	-0.436	-0.490	-0.346	-0.361	-0.550	0.406	0.158	-0.205	-0.343
BAC					0.292	0.099	0.421	0.058	0.327	0.620	0.359	0.359	0.620	0.237	0.345	-0.492	0.442	0.508	0.403
BK						0.424	0.891	-0.516	0.947	0.843	0.805	0.948	0.843	0.742	0.923	-0.709	0.086	0.578	0.778
BLK							0.408	0.290	0.491	0.362	0.556	0.430	0.362	0.482	0.370	-0.246	0.584	0.572	0.630
C								-0.338	0.876	0.794	0.837	0.950	0.794	0.752	0.898	-0.820	0.264	0.670	0.822
CB									-0.439	-0.362	-0.306	-0.443	-0.362	-0.139	-0.481	0.383	0.600	-0.078	-0.178
GS										0.878	0.827	0.957	0.878	0.649	0.852	-0.645	0.096	0.629	0.761
JPM											0.678	0.669	0.839	0.497	0.749	-0.658	0.125	0.554	0.679
MA												0.867	0.669	0.718	0.826	-0.726	0.332	0.856	0.847
MET													0.839	0.701	0.913	-0.775	0.147	0.649	0.790
MS														0.497	0.749	-0.658	0.12	0.554	0.679
PNSC															0.805	-0.590	0.497	0.621	0.816
SCHW																-0.824	0.205	0.633	0.770
SPG																	-0.223	-0.588	-0.652
USB																		0.557	0.498
V																			0.795
WFC																			

Table 21: Industrials Spearman's ρ correlation matrix

	ADP	BA	CAT	CSX	EMR	FDX	GD	GE	HON	ITW	JCI	LMT	MMM	NOC	RTN	UNP	UPS	UTX
ADP		0.668	0.111	0.304	0.303	0.324	-0.139	0.383	0.353	0.645	0.303	0.156	0.659	0.143	0.537	0.336	0.447	0.539
BA			0.133	0.322	0.269	0.325	-0.188	0.243	0.545	0.754	0.412	0.115	0.676	0.172	0.348	0.380	-0.020	0.599
CAT				0.891	0.917	0.849	-0.189	-0.254	0.300	0.579	0.759	-0.839	0.582	-0.752	-0.563	0.723	0.019	0.747
CSX					0.888	0.871	-0.341	-0.293	0.264	0.700	0.814	-0.741	0.713	-0.688	-0.388	0.869	0.101	0.855
EMR						0.889	-0.320	-0.182	0.348	0.716	0.814	-0.740	0.751	-0.709	-0.348	0.781	0.139	0.841
FDX							-0.169	-0.195	0.426	0.678	0.789	-0.668	0.731	-0.593	-0.382	0.799	0.190	0.877
GD								0.334	0.342	-0.481	-0.482	0.405	-0.476	0.517	0.039	-0.497	0.321	-0.402
GE									0.392	0.004	-0.440	0.413	-0.126	0.562	0.492	-0.486	0.184	-0.205
HON										0.435	0.211	0.064	0.313	0.210	0.143	0.102	0.095	0.397
ITW											0.734	-0.384	0.907	-0.351	0.046	0.715	-0.035	0.886
JCI												-0.616	0.837	-0.652	-0.259	0.922	0.042	0.879
LMT													-0.386	0.902	0.776	-0.638	0.253	-0.569
MMM														-0.404	0.049	0.815	0.153	0.913
NOC															0.662	-0.660	0.183	-0.542
RTN																-0.293	0.404	-0.184
UNP																	0.095	0.888
UPS																		0.066

Table 22: Technology Spearman's ρ correlation matrix

	AAPL	ADBE	AMAT	CRM	CSCO	CTSH	EBAY	FB	HPQ	IBM	INTC	MSFT	NVDA	ORCL	QCOM	TXN
AAPL		-0.228	0.501	-0.096	0.694	0.132	-0.114	-0.353	0.520	0.779	-0.058	-0.180	-0.284	0.700	0.528	0.388
ADBE			-0.636	0.877	-0.112	0.526	0.817	0.890	-0.730	-0.517	0.008	0.683	0.668	-0.618	-0.766	0.014
AMAT				-0.589	0.423	-0.424	-0.593	-0.733	0.848	0.504	0.459	-0.331	-0.407	0.819	0.811	0.573
CRM					0.034	0.564	0.756	0.807	-0.602	-0.348	0.089	0.776	0.610	-0.450	-0.684	0.073
CSCO						0.289	-0.108	-0.305	0.549	0.605	0.255	0.178	-0.116	0.657	0.592	0.441
CTSH							0.487	0.484	-0.295	-0.124	-0.172	0.322	0.459	-0.154	-0.279	0.120
EBAY								0.829	-0.641	-0.401	-0.173	0.591	0.463	-0.511	-0.683	-0.054
FB									-0.819	-0.618	-0.132	0.561	0.642	-0.732	-0.858	-0.051
HPQ										0.626	0.371	-0.288	-0.584	0.887	0.893	0.340
IBM											-0.177	-0.283	-0.644	0.762	0.719	-0.009
INTC												0.449	0.222	0.279	0.188	0.632
MSFT													0.462	-0.197	-0.410	0.133
NVDA														-0.484	-0.614	0.360
ORCL															0.854	0.414
QCOM																0.241

Table 23: Energy Spearman's ρ correlation matrix

	APA	APC	COP	CVX	CXO	DVN	EOG	HAL	KMI	MPC	NBL	OXY	PSX	PXD	SLB	VLO	WMB
APA		0.889	0.958	0.950	0.659	0.929	0.932	0.817	0.872	0.003	0.963	0.889	-0.327	0.909	0.815	-0.535	0.437
APC			0.902	0.832	0.752	0.959	0.896	0.922	0.947	-0.066	0.900	0.798	-0.407	0.785	0.911	-0.618	0.667
COP				0.953	0.702	0.941	0.955	0.830	0.887	-0.052	0.948	0.902	-0.328	0.880	0.827	-0.583	0.434
CVX					0.562	0.881	0.896	0.747	0.849	-0.056	0.931	0.844	-0.355	0.868	0.748	-0.550	0.350
CXO						0.749	0.786	0.826	0.583	0.183	0.702	0.740	0.089	0.694	0.790	-0.249	0.481
DVN							0.907	0.925	0.921	0.031	0.914	0.861	-0.338	0.820	0.935	-0.578	0.614
EOG								0.824	0.850	-0.083	0.949	0.892	-0.268	0.918	0.793	-0.551	0.393
HAL									0.831	0.197	0.833	0.745	-0.168	0.755	0.962	-0.396	0.731
KMI										-0.191	0.868	0.730	-0.503	0.747	0.827	-0.687	0.632
MPC											-0.012	0.007	0.671	0.002	0.214	0.724	0.291
NBL												0.866	-0.340	0.914	0.805	-0.525	0.453
OXY													-0.254	0.785	0.778	-0.533	0.328
PSX														-0.195	-0.182	0.815	-0.138
PXD															0.689	-0.400	0.259
SLB																-0.421	0.756
VLO																	-0.186