ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Master Specialization Financial Economics

Pairs-Trading:

Testing the Consistency of a Statistical Arbitrage Investment Strategy

Author: R.L. Cummins
Student number: 299180
Thesis supervisor: Dr. L.A.P. Swinkels
Finish date: November 2010

PREFACE AND ACKNOWLEDGEMENTS

There are a number of people that I would like to thank for assisting me with the realization of my thesis. First of all, I would like to thank Dr. L.A.P. for his kind and supportive supervision of my thesis. Furthermore, I would like to thank M. Kretschmer, M. Voermans and L. Heij for their comments and suggestions. In addition, I would like to express my gratitude to Pelargos Capital for giving me the opportunity and resources to write my thesis. Last but not least I would like to thank my friends, family and especially my girlfriend for their support.

ABSTRACT

We test the statistical arbitrage investment strategy 'pairs-trading', in the U.S., Japan, Hong Kong and China, using daily data between January 2004 and December 2009. Results show that the pairs-trading algorithm designed by Gatev, Goetzmann and Rouwenhorst (2006), does not yield consistent positive excess returns. Neither throughout time, nor in different markets. Analysis of the pairs-trading returns suggests that pairs-trading is extremely profitable during periods of stock market turmoil. We try to increase pairs-trading performance by using proxies of market inefficiencies as a timing instrument. The increase is modest and not consistent in all four markets. We also try to improve performance by adjusting the pairs-trading algorithm. We set up sector restricted pairs and combine different sector pairs into one portfolio. Although the adjusted algorithm does increase returns in the U.S., Japan and China, the increase is modest. We believe that more research is needed to discover factors that drive pairs-trading returns in different markets and different times.

Keywords:

Pairs-Trading, Market Efficiency, Investment Decision, Asset Pricing

JEL Classifications: G11, G12, G14

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

1. Introduction

In this paper we test the pairs-trading strategy designed by Gatev, Goetzmann and Rouwenhorst (2006). They find that pairs-trading in the U.S. yields an monthly average excess return in the order of 1%, between 1962 and 2002. We test the same strategy from January 2004 until December 2009 in the U.S., Japan, Hong Kong and China. Since the publication of the Gatev, Goetzmann and Rouwenhorst paper, several other papers have been written about pairs-trading. Since Gatev, Goetzmann and Rouwenhorst claim that the pairs-trading algorithm is disarmingly simple, most papers try to why pairs-trading yieds excess returns and what factors drive those returns. The consensus is that pairs-traders get rewarded for enforcing the relative pricing of two similar stocks. Pairs-traders search for identical stocks and sell the winner stock and buy the loser stock. The biggest risk of pairs-trading is the so-called divergence risk. There is always a probability that after initial divergence, the stocks in a pair do not converge, but keep on diverging. This leads to potentially large losses for pairs-traders. Recently, Do and Faff (2010) test if pairs-trading stood the test of time. They implement the same pairs-trading algorithm as Gatev, Goetzmann and Rouwenhorst and provide an out-of-sample test by analyzing pairs-trading over the period 2003-2009. Although Do and Faff still find an positive average return, they report a declining trend in pairs-trading. Furthermore, they show that pairs-trading is especially profitable in periods of stock market turmoil. Do and Faff also proved that making some adjustments to the pairs-trading algorithm, performance can be increased. By testing pairs-trading in the U.S., Japan, Hong Kong and China we try to answer the question if pairs-trading yields positive returns in different markets. Consequently, we look for differences between pairs-trading statistics and return in those four countries. Building on the conclusion of Do and Faff that pairs-trading is the most profitable during market turbulence, we provide a sub-period analysis of the returns. By doing so we also look for factors that drive pairs-trading returns and which might serve as a timing instrument. We analyze if by timing pairs-trading opportunities, the performance can be increased. We also test if we can increase performance by altering the pairs-trading algorithm. The adjustments are a logical corollary of the results we find. By making only straightforward adjustment, we try to nip in the bud any data-snooping criticism. The remainder of the paper is organized as follows. Section 2 describes some of the more influential papers about pairs-trading. Section 3 describes the data and is followed by the section that explains the methodology of the paper. The empirical results are described in section 5. Section 6 analyzes the timing

strategy of pairs-trading and section 7 describes the adjusted pairs-trading algorithm. Section 8 concludes.

2. Related Literature

Pairs-trading has been a commonly used quantitative trading strategy, especially by hedge funds. Nonetheless, not much literature has been devoted to the subject. One of the first thorough and influential papers about pairs-trading is Gatev, Goetzmann and Rouwenhorst (2006) (GGR). Since then several other papers have been published, which built to a larger extend on the GGR paper. In this section we review the most important and relevant papers about pairs-trading. Since most papers are based on the paper written by GGR, the results make a good comparison. First, we give a brief summary of the different papers and the intended goal of research. Second, we describe the datasets that are used by the different papers. Third, we describe the methodology used by GGR when setting-up their pairs-trading strategy, since all papers built on the methodology of GGR. We also highlight the most important differences between the methodology used by the other papers and those used by GGR. Fourth, we discuss the main results and findings of each paper. Fifth, we describe the equilibrium model of convergence trading, developed by Kondor (2009) (Kondor).

2.1. Research Questions

GGR provide one of the first discussions on pairs-trading. In their article they set up a relatively easy pairs-trading algorithm of which the risk and return characteristics are discussed.

Engelberg, Gao and Jagannathan (2009) (EGJ) expand on GGR. They try to answer the questions: 'Why some pairs are more profitable than others? What causes the prices of pairs to diverge and how that affects subsequent convergence?'

Andrade, Di Pietro and Seasholes (2005) (APS) try to link uninformed demand shocks with the risk and returns of pairs-trading. The goal is to understand 'why prices of similar securities diverge'.

Do and Faff (2010) (DF), continue with the pairs-trading algorithm designed by GGR and extend the sample to July 2009. They investigate if pairs-trading still show significant returns after 2003. Additionally, they analyze if the pairs-trading algorithm could be improved to generate higher returns.

2.2. Data

The sample GGR use, consists of daily U.S. equity market stock data over the period 1962 through 2002, which is collected from the CRSP database. GGR completed the first draft of

their paper in 1999, using data until 1998. After the first draft they used the sample period 1999-2002 as an out-of-sample test of their strategy. By doing so, they try to deal with data snooping criticism and show that the initial pairs-trading returns are not just an historical artifact. Furthermore, it gives them the opportunity to test if public dissemination of their results affects pairs-trading returns in the period thereafter.

EGJ obtain their data from the CRSP database. The pairs-trading universe consists of all common shares traded on NYSE, AMEX or NASDAQ for the period January 1992 to June 2006.

APS want to investigate the relation between uninformed demand shocks and pairs-trading. Therefore they use all listed stocks on the Taiwan Stock Exchange. It is not possible to identify aggregate uninformed shocks on the NYSE, because such orders are filtered out by brokers and sent to regional exchanges.[1] Using the Taiwan Stock Exchange enables them to identify those uniformed buys and sells. APS collect their data from the Taiwan Economic Journal. Their sample consists of a total of 647 stocks over the period January 1994 through August 2002.

DF use daily CRSP data over the period July 1962-June 2009. The data is restricted to ordinary shares. DF also filtered out the companies that issued more than one equity security, since those securities represented the most frequently repeated pairs.

2.3. Methodology

First we discuss the methodology used by GGR. All papers that are discussed use the same methodology as GGR and so do we. The methodology of GGR forms the cornerstone of all papers. All paper make some adjustments to the methodology of GGR, so we discuss these modifications. A more thorough discussion of the pairs-trading methodology can be found in the methodology part of our paper.

Gatev, Goetzmann and Rouwenhorst

GGR implement a two-stage pairs-trading strategy. The first stage is the formation period of 12-months in which the pairs are matched. The second stage is the trading period of six months, in which the matched pairs are actually traded. GGR have arbitrarily chosen these periods, but do use them consistently throughout the paper. GGR start a new trading cycle every month, this leads to six one-month staggering pairs portfolios every month.

---

[1] Regional exchanges are exchanges in the U.S. that are not the New York Stock Exchange

In the formation period stocks with missing data are excluded. This serves to identify relatively liquid stocks as well as facilitating pairs formation. With the remaining stocks they construct a cumulative total return index for each stock. Pairs are formed by matching two stocks that have a minimum sum of squared deviation (SSD) between the two normalized price series, where prices include reinvested dividends. GGR claim that this matching method resembles the matching strategy of practitioners the most.

GGR make a distinction between unrestricted pairs and restricted pairs. Unrestricted pairs are pairs that are formed of two stocks, without any restriction. Restricted pairs are formed of two stocks that are in the same industry. They use the four broad industries classified by Standard and Poor's to form the restricted pairs: Utilities, Transportation, Financial and Industrials.

The closing and opening of the pairs is based on a standard deviation metric. Pairs are opened when the spread between the normalized price series reaches two historical[2] standard deviations and pairs are closed when the prices cross again. If the pair is still open at the end of the trading period, the pairs is closed and the gains or losses are calculated. Similarly, if a stock is delisted, the delisting or last available price is used to calculate the return.

When examining risk and return, GGR review four portfolios of pairs. The first portfolio consists of the top-5 pairs, the second portfolio of the top-20 pairs, the third portfolio of pairs 101 until 120 and the last portfolio consists of all pairs.

Pay-offs are calculated by going one dollar short in the higher-priced stock and one dollar long in the lower-priced stock. Consequently, the pay-offs are excess returns. This excess return is computed as the reinvested payoffs during the trading interval. GGR distinguish between two measures of excess return: the return on committed capital and the return on fully invested capital. The difference between the two is the opportunity cost of capital; when calculating the return on committed capital, they commit a dollar to a pair, even if the pair is not trading. When calculating the fully invested capital return, they only commit a dollar if the pair is actually trading. GGR argue that 'to the extent that hedge funds are flexible in their sources and uses of funds, computing excess return relative to the actual capital employed may give a more realistic measure of the trading profits'. For comparison; the monthly fully invested average excess return for all pairs over the full sample period is around 1.1%, compared to around 0.6% on committed capital.

---

[2] The historical standard deviation is the standard deviation of the spread during the formation period.

Engelberg, Gao and Jagannathan

For setting up the pairs-trading strategy EGJ use the same algorithm as described by GGR. However, they have made some adjustments which are described below. First, they limit the stocks available for pairing by only allowing for pairs of stocks which are within the same industry. They use the industry classification given by Fama and French (1997), who classified 12 different industries. Second, they identify the 200 pairs that have the smallest average normalized price difference that are 'eligible' for trading. Third, they extend the trading period to 12 months. Fourth, they apply two additional rules for closing pairs. EGJ close positions which have diverged, but did not convergence within six months and call this 'no convergence'. They also consider the same rule with a 10 days horizon, calling it a 'cream-skimming' strategy. Returns are calculated in the same manner as GGR did.

EGJ study news events surrounding the divergence date of the pairs. Companies are matched with news events which are covered by the Factiva database to retrieve the necessary information.

EGJ use two types of regression when analyzing pairs-trading. First, they use a calendar-time regression, which analyzes the risk and return characteristics on a month-to-month basis. Second, pairs are analyzed with an event-time cross sectional regression. Instead of regressing the returns on a month-to-month basis, the returns are now regressed over the event period. This has the advantage that they can add several control variables. Sorting along several dimensions would have created portfolios with only a very few pairs in the calendar-time approach. Furthermore, the event-time approach gives a more complete picture of the cycle of pairs-trading: the opening of the pairs, trading of the pairs and the closure of the pairs. In our paper we are not using a event-time approach since we analyze calendar-time characteristics and not the pairs-trading cycle.

Andrade, Di Pietro and Seasholes

The same pairs-trading strategy is followed as described by GGR. The two stocks that have the closest co-movement –based on the SSD- are ranked as the first pair, forming the twenty closest pairs. Like GGR APS have a 12-month formation period and a 6-month trading period. Pairs are opened when the spread reaches the two standard deviation trigger value. Pairs are closed when they converge or at the end of the trading period. They repeat the strategy every six months, ending with a total sixteen non-overlapping trading periods. APS thus do not apply the same 1-month overlapping strategy as GGR do.

APS follow Andrade, Chang and Seasholes (2004) for determining uninformed trading shocks. From the Taiwan Stock Exchange they gather the number of shares that are held

long margin.[3] APS argue that shares that are held on margin, are shares held by uninformed investors. So changes in the number of shares that are held long on margin identifies uninformed selling or buying. If the number of shares held on margin increase, there is uninformed buying and vice versa.[4]

Do and Faff

Because DF wanted to provide an out-of-sample test of the GGR paper, they used almost the exact same pairs-trading algorithm. Pairs are matched using the SSD of the normalized price series, which are scaled to $1 at the beginning of the trading period. At the beginning of the trading period, DF normalize prices back to $1 again. It is not perfectly clear from GGR whether they normalize prices back to 1 at the beginning of the trading period, or if they continue with the normalized price series of the formation period. DF also try to increase the pairs-trading returns by adjusting the pairs-trading algorithm. They argue that 'good' pairs are not only two stocks that track each other well, but also two stocks that cross frequently. They also state that pairs of stocks that are within the same industry perform better than pairs of stocks that are in different industries. With these insights DF constructed an alternative algorithm: For each of the four major industry groups described by GGR they form the 50 pairs with the lowest SSD. From these 50 pairs they select the 20 pairs which had the most crossings of prices during the formation period.

2.4. Empirical Results

Gatev, Goetzmann and Rouwenhorst

GGR find a monthly average excess return of 1.31% and 1.44% (full invested) for the top-5 and top-20 pairs respectively. Using committed capital, they still find a -in statistical and economical sense- large return of 0.78% and 0.81%, respectively. Other interesting observations are that the standard deviation of a portfolio falls as the number of pairs increases. Moreover, the minimum realized return of the portfolio increases as the number of pairs increases, whereas the maximum realized return remains stable.

Following Jegadeesh (1990), Jegadeesh and Titman (1995) and Conrand and Kaul (1989), the returns might be biased upwards due to the bid-ask bounce. Pairs-trading is a contrarian strategy, so when selling the higher-priced stock one will most likely receive the ask price. However, when buying the lower-priced stock one will probably pay the bid price. Consequently, the opposite is true when closing the pair. This leads to a so-called bid-ask

---

[3] Buying a share with money on loan from a broker
[4] See Andrade, Chang and Seasholes (2004) for the full discussion on uninformed trading.

spread, biasing the returns upwards. GGR circumvent this bias by waiting one day after the prices have diverged to set-up the pair and waiting one day after the prices crossed before closing the positions. By waiting one day, the returns on the fully invested portfolios and committed capital portfolios drop by an average of about 0.30-0.55% and 0.20-0.35%, respectively. From the paper it is not clear which prices GGR use[5] and what they mean by one day waiting. In our view one day waiting could mean that if a pair diverges at day 0, they open the pair at day 1, or it means that if prices diverge at day 0, they open the pair at day 2. This question cannot be answered based on the information GGR give us.

GGR find that on average, pairs are open 3.75 months during the trading period of six months. They argue that pairs-trading is thus an medium-term investment strategy. GGR also make an estimation for the transactions costs of the strategy. They estimate a conservative 3.24%[6] transaction costs every six months, for every pair. This reduces net profits, now ranging from 1.13% to 2.25% over the six-months trading period. However, concluding that the net profits remain economically and statistically significant. GGR point out two possible reasons why they are actually trading too much: First they argue that they are underestimating the standard deviation[7] of the pairs and thus open pairs too soon; second they still open pairs which are close to the end of the trading interval, which might not be the most desirable strategy.

When making cross-sectional analyses of the pairs-trading returns, GGR show that the returns are different for every industry. Utilities is the best performing pairs sector and Transportation the worst, with a mean excess return of 1.08% and 0.58% respectively for the top 20 pairs.

Dividing the stocks in size deciles using CRSP breakpoints, they find that about two-third of the pairs consist of stocks in different deciles. However, 74% of the stocks in the top 20 pairs belong to the top three deciles. Furthermore, 71% of the stocks in the top 20 pairs are stocks in the utility sector and 22% of the pairs consist of stocks which are in different sectors.

GGR also discuss the risk characteristics of the pairs-trading strategy. Over the 1963-2002 period they find an excess return that is about twice as large as the return of the S&P 500. Interestingly enough the risk –measured as standard deviation- is about half as large,

---

[5] They could use closing, opening, intra-day, highest or lowest pricing for instance
[6] Waiting one day before trading leads to a fall in monthly excess return of 0.54 %. GGR argue that this is the average bid-ask spread and use this as a measure for transactions costs. This leads to a total of 6 X 0.54%= 3.24% transactions costs for each trading period of six months.
[7] The historical standard deviation of the pair is used to calculate the two standard deviation trigger. However, pairs are by construction the ones that have the lowest standard deviation; hence the standard deviation is most likely to be underestimated.

leading to a Sharpe Ratio which is four to six times larger than the Sharpe Ratio of the market.

When analyzing the risks of pairs-trading, they also consider several different systematic risk factors. First of all they use the three factors as described by Fama and French (1996): *Market, HML* and *SMB.* They also added four more factors: the *Reversal* factor as described by Jegadeesh (1990), among others, the *Momentum* factor as described by Carhart (1997), among others, the U.S. bond *Default Premium* and the bond *Horizon Premium*, as suggested by Ibbotson. Since pairs-trading involves selling the higher-priced stock and buying the lower-priced stock, they expect a negative correlation between momentum and the pairs-trading profits and a positive correlation between the reversal factor and the profits from pairs-trading. When conducting the regression, GGR find an R-squared of only 2%. The market, HML and SMB factors are all statistically insignificant. Although the factors momentum and reversal have the expected sign and more than half are statistically significant, GGR argue that they are not large enough to fully explain the average returns of pairs-trading. They state that pairs-trading returns are fundamentally different from a simple contrarian strategy when looking at the significance of the risk-adjusted returns. When looking at the bond factors they find a positive correlation between both factors and returns, however these factors are insignificant. Overall GGR conclude that the pairs portfolios are almost factor-neutral and that 'this may be expected because they are constructed in a way that should essentially match up economic substitutes'.

Pairs-trading appeared to be very profitable in the 1970s and 1980s. Pairs returns are also much more smooth compared to the returns on the S&P 500. GGR find that the pairs performed well during the period 1969 through 1980, when there was a dramatic real decline in the stock market. When the stock market performed well in the mid-1990s, the profits from pairs-trading were modest. GGR point out several explanations for this phenomenon; increased competition after the 1980s might have diluted profits, as well as the decrease in commissions over the sample period, which led to increased trading activity and hence more aligned prices. Last but not least, it just might be possible that pairs are more profitable in market downturns.

GGR also examined the long and short positions in the pairs portfolios separately. They find that a large part of the pairs risk-adjusted returns comes from the short portfolio. Their conclusion is thus that the returns do not come from a simple mean-reversion strategy. To show that the returns are also not driven by a contrarian effect, they set-up a control portfolio of pairs which have similar prior one-month returns as the actual pairs. The returns of the control portfolio are significantly lower than the return of the actual pairs portfolio, implying that the pairs-trading returns do not come from a contrarian strategy.

To analyze two different sub-periods, GGR spilt the sample into two periods: a post and prior 1988 part. Although they find that the risk-adjusted returns are about two-thirds lower in the second part of the sample, they are still significantly positive. They do find that this decrease in returns is only partly explained by changes in factor exposures and factor volatilities. GGR conclude from this that there is some sort of latent risk factor, which is not captured by the conventional measures of systematic risk that partly explains the profits from pairs-trading.

Finally, GGR perform two tests which examine the possible effects of short-selling constraints on pairs-trading profits. First, they show that pairs-trading profits are not driven by illiquid stocks, as using only the more liquid stocks for the pairs-trading strategy does not affect the returns significantly. Second, they examine if short-recalls diminish the returns on pairs-trading. Stocks which have a high volume recall[8] have somewhat lower profits, but the profits remain positive and significant.

Engelberg, Gao and Jagannathan

One of the first interesting observations of EGJ is the fact that the profitability of the pairs declines over convergence time. Moreover, they show that if pairs do not converge within seven days after the opening, it becomes increasingly unlikely that the stocks will converge at all. From this, EGJ conclude three things: First, they argue that it is important to understand what happens on the divergence date, as the first days after divergence are critical for the profitability of pairs. Second, although pairs generate the most returns in the first days, profits in the subsequent days are still positive and significant. Pairs which are held until 100 days after divergence generate a profit of 2.08% compared to a return of 0.83% for pairs with a holding period of 10 days. This suggests that profits from pairs-trading comes from different sources, as liquidity provision is more a short-term factor and price discovery a longer-term factor, for instance. Third, if traders have to hold on to their positions for a longer period of time, the risks involved need to be examined. In particular the factors that drive the speed of convergence and divergence.

EGJ find that the average NYSE size rank of the pairs is the 65th percentile, from which they conclude that most pairs are relatively large stocks, and consequently relatively liquid stocks. Implementation is thus less a concern. Most of the pairs are stocks in the Financial industry (44.38%), Utility sector (22.52%) and Manufacturing (13.96%).

In the two days prior to divergence, only 6.71% of the companies faced a corporate news event of which quarterly earnings announcement alone attribute to 90% of the total news

---

[8] To simulate recalls on the short position, GGR search for days on which the volume of the stock exceeds the average volume of that stock of the past 18-months by one standard deviation.

events. From the total 27,703 pairs EGJ matched, 69, 23 and 9 pairs experienced index addition or deletion in the 30, 1 and 0 days prior to divergence, respectively.

EGJ find an average risk adjusted return of 70 basis points per month, which is comparable with GGR. The cream skimming strategy earns a monthly alpha of 1.75%, compared to a monthly alpha of 0.70% for the standard strategy.

To examine the relationship between liquidity and returns, they use two variables: pair wise average proportional effective spread (PESPR) and the change is PESPR (dPESPR). They find a strong and positive relationship between the level of liquidity (PESPR) and the profits from the standard (6-months holding period) strategy. This relationship is weaker for the 10-days holding period strategy. Smaller and less liquid pairs tend to outperform the more liquid and larger pairs. Looking at the change in liquidity, they find a positive relationship for the creaming strategy with smaller pairs. This relationship is not present between dPESPR and the standard strategy. Thus, some of the short-term profits are a reward for providing immediate liquidity and pairs which are less liquid have larger long-term profits.

EGJ discussed the relationship between profitability of pairs-trading and idiosyncratic news. They make a distinction between actual news and media coverage. New is former non-public information that has been made public, coverage is the reprinting of previously known public information. For an event to be classified as news, it has to meet two criteria: First, there must be a story about the company in the Dow Jones News Service. Second, the stock had an absolute excess return on the day of divergence of at least two standard deviations[9]. EGJ find an economically and statistically significant difference between pairs with or without news on the day of divergence. Large stocks have a lower monthly alpha of 34 basis points when there has been a news event on the day of divergence. The difference is 30 basis points for smaller stocks. The same is done for common information. Besides idiosyncratic news, news on the whole industry a stock is in also influences the price of that stock. They consider an 'information diffusion rate'. The information diffusion rate captures the difference in adjustment speed of two stocks to industry related news. They show that pairs for which the information diffusion rate is high, returns of that pair are higher compared to pairs which have slower information diffusion rates.

EGJ also discuss the risk aspect of pairs-trading. They test for the same factors as GGR did: *Momentum, Reversal, HML, SMB and Market,* but similar to GGR they find that those factors

---

[9] The two standard deviation criterion is used to make a distinction between 'news' and 'coverage'. Republishing of known information is unlikely to cause a two standard deviation return, whereas news might trigger such a move in the price of the stock.

have little power in explaining pairs-trading returns.[10] EGJ also regress a liquidity risk factor on the pairs-trading returns. They find a negative correlation between the liquidity factors and returns form pair trading. However the additional explanatory value of those factors is low.[11] Next, EGJ regress several macro risk factors on the returns. First of all, they use the U.S. Treasury-Eurodollar (TED) spread as a proxy for funding liquidity risk. Second, they use the AAA/T-bill spread as proxy for the demand-side driven liquidity premium in the economy, to link long-run consumption risk and pairs-trading profits. Third, business cycle risk is captured by using default spreads, using Moody's BAA minus AAA bond yield spreads. They conclude that pairs-trading have little exposure to macroeconomic factors, as only the exposure of pairs-trading to TED is high and significant. The interpretation is that when borrowing is difficult (wide TED spreads), is becomes harder for arbitrageurs to enforce the relative pricing of stocks, hence spreads within the pairs are wider.

Analyzing the event-time regression, EGJ find that most of the results are in line with the result found with the calendar-time approach. On average, pairs with small stocks and growth stocks have higher returns. The same holds for stocks with low past one-month returns and stocks which have a high volatility. Stocks with higher PESPR and low turnover, earn higher profits. Like the calendar-time regression, the event-time regression shows that pairs with idiosyncratic news about one of the two stocks in the pair earn lower returns. The same holds for the information diffusion rate, the higher the information diffusion, the higher the returns.

Using a logistic regression with divergence as binary variable[12], EGJ try to find factors that drive the opening of the pairs, profitability of pairs, time-to-convergence of pairs and the risk of the pairs. Idiosyncratic news about at least one of the stocks in a pair decreases the profitability of that pair. Idiosyncratic news creates opportunities, as the news makes it more likely that the pair diverges. However, it increases the chance that prices keep on diverging and thus lead to a loss for the arbitrageur. The level of liquidity and short-term changes in liquidity lead to an increased profitability of a pair. The probability that a pair opens increases when there is a liquidity shock and the risk to the arbitrageur is smaller, since the higher the liquidity of the stock, the larger the probability that the price difference is traded away. Smaller, less liquid stocks are traded less frequently; hence there is a higher risk that the divergence persists. There is a strong relationship between the relative speed of adjustment to common industry information and profitability of pairs. If there is a difference

---

[10] EGJ find an R-squared of about 30% for the six-months holding period and an R-squared of about 7% for the 10-day holding period.

[11] The R-squared of the regression is about 30% for the six-month holding period and the R-squared of the 10-day holding period is about 8%.

[12] The binary variable is '1' if the pair opens, '0' if it does not open

in adjustment speed, opportunities are created because the prices of the two different stocks adjust to the news at a different speed. Moreover, the risks to the arbitrageur are lower, since the probability that the prices converge again increases. It takes one of the stocks more time to adjust to the news, but probably it will in the end. Size and liquidity are also related to pairs returns, but more in an indirect way. The effect of the previously mentioned factors is larger with size and liquidity. The impact of information diffusion rates is stronger among small, less liquid stocks.

EGJ also performed two robustness checks: default risk and short-sale constraints. In line with GGR, there is no relation between default profitability and pair returns, neither is there for short-sale constraints.

Andrade, Di Pietro and Seasholes

APS find an annual excess return of 10.18%, which is comparable to the 11.28% annual excess return of GGR. In their sample, they have a total of 320 the pairs that are open 70.34% of the time and only five pairs never open. The average pair opens 2.29 times during the half-year trading period.

APS find that on average the stocks diverge by 4.15% on the day of opening. They conclude from this that pairs open mainly because of shock instead of smoothly reaching the trigger value. Furthermore, they show that the rising stock of the pairs account for 71.08% of the openings.

There is a high correlation of about 0.3 between pairs-trading returns and uninformed trading. Most of the uninformed trading of stocks in the pairs comes from uninformed buying. According to APS this is straightforward: uninformed selling shock can easily be absorbed by investors who buy the excess supply. However, when there is an uninformed buying shock, investors need to short that stock if they do not have the stock in possession. This is much harder for most investors, leaving open an statistical arbitrage opportunity.

A survival analysis is performed to assess the relation between the time-to-opening of a pair and uninformed trading. They find that pairs open more quickly when the uninformed buying coefficient is high. There is also a significant positive relationship between market volatility shocks and time-to-opening, indicating that a shock of the market volatility increases the change that the pair will open. There is no significant relationship between the time-to-opening of a pair and the uninformed selling coefficient.

APS regress the returns on the risk factors *Market, SMB, HML* and *Momentum.* Although he loads on the market and HML are significant, the magnitudes are economically insignificant. SMB and momentum have insignificant loadings. They also regress the returns on the uninformed trading factor as described above. They find that this factor is highly significant

and positive. Thus, when a pair experiences uninformed trading in one of the stocks, returns from this pair will be higher than from pairs without uninformed trading. However, APS note that uninformed trading is not a systematic risk factor and hence the constant is not an excess return.

Do and Faff

When analyzing the returns, DF only focus on the top-20 fully invested capital portfolio, since that portfolio is the main object of the GGR study. DF first test if the declining trend in pairs returns observed by GGR continues after 2003. They indeed find persistence in the declining trend. They report a monthly average excess return of 0.24% for the January 2003-June 2009 period. Although being significantly smaller than the returns found by GGR, this returns is still significantly larger than zero. DF find that the declining trend in pairs-trading returns is not present in the two most recent major bear markets: January 2000-December 2002 and July 2007-June 2009. Unreported results show that the average returns during the bear markets were 0.92% and 0.78% a month respectively. In line with the returns, the volatility during the bear markets was also significantly higher than the adjacent periods. However, the Sharpe ratios for these two periods remain superior to those during the other periods.

In an attempt to explain the pair returns and the declining trend in particular, DF investigate two stock market phenomena: arbitrage risk and market efficiency. Basically, arbitrage risk is the risk that in an attempt to arbitrage mispricing in the market, the arbitrageur faces (large) losses, because the mispricing persists. Irrational trading causes prices of similar stock to diverge, leading to opportunities for investors to profit from this mispricing. Increased competition among arbitrageurs will see many arbitrageurs chasing the same opportunities, leading to a decline in profits. DF find that the decreasing trend is mainly caused by an increase of divergence losses, hence arbitrage risk. This is opposite to what GGR suggest, namely that increased competition decreases pairs-trading returns. When analyzing the two bear markets, DF find that the higher profits during these periods has two different drivers. During the first bear market from 2000-2002, the higher performance came from the increased profitability among convergent pairs: 4.84% a month for the 2000-2002 period, compared to 1.51% and 1.69% for the period 1989-1999 and 2003-2007 respectively. The number of converging pairs versus the non-converging pair is comparable during the three periods. The most recent bear market showed an increase in profitability per pair, which was caused by an increase in convergence. The number of pairs which converted multiple times during the trading period increased to 37% from 18% in the period 2003-2007. The number of divergent pairs decreased to 32% from 44%.

DF also test what causes the decrease in profitability of pairs-trading. The declining profits during the 1989-2002 period is for 71% attributable to the increase of arbitrage risk and only 29% to increased market efficiency. This relation is 60/40% for the 2003-2009 period. During the two bear markets the pairs-trading profits increase because the impact of less efficient markets outweighed the increase in arbitrage risk.

As discussed previously, DF also use an alternative pairs-trading algorithm. First they analyze pair returns separately for each industry group. Utilities and Financials are the most profitable industries, which is consistent with GGR. Using the number of crossings during the trading period as a trading rule increases the mean return by 0.06 % and 0.03% respectively for Utilities and Financials. Utilities show a lower divergence loss compared to the financials, which implicates higher homogeneity among utility stocks. In a cross-sectional analysis DF show that there is a positive relationship between pair returns and the homogeneity and number of formation period zero-crossings of a pair. Volatility does not influence the pair returns, since they are insignificant. Industry specific volatility does have a significant relationship with the pair returns, increased volatility, leads to higher returns. However, DF only find a low R^2 of only 0.90%, which indicates that only a minor part of the pair returns are explained by those variables. They conclude that: 'Clearly, much more needs to be learned about pairs-trading as a quantitative arbitrage strategy.'

Literature Overview

Authors	Publication	Research Goal	Period	Country	Data	Methodolody Adjustments	Conclusion
Gatev, Goetzmann and Rouwenhorst	2006	Pairs-Trading Performance	1962-2002	U.S.	U.S. Equity Markets	-	Pairs-trading yields an average excess returns, which cannot be explained by common risk factors
Engelberg, Gao and Jagannathan	2009	Pairs-Trading Profit Sources	1992-2006	U.S.	NYSE, AMEX and NASDAQ	Top-200 Sector pairs, 12-month trading period, 10-days trading period	For each factor trade-off between positive and negative effects on pairs-trading profits
Andrade, Di Pietro and Seasholes	2005	Uninformed Demand shocks vs Pairs-Trading	1994-2002	Taiwan	Taiwan Stock Exchange	Non-overlapping trading periods	Pairs experiencing uninformed trading shocks have higher returns
Do and Faff	2010	Does Pairs-Trading Still work?	1962-2009	U.S.	U.S. Equity Markets	Top-20 sector pairs, lowest SSD, most crossings	Decling trend in pairs-trading profits, mostly explained by increased divergence risk

Table 1: Overview of the discussed literature. Shown are the authors, year of publication, research goal, sample period, country, data, methodology and main conclusion.

Table 1 gives a brief overview of the different papers. The table gives on overview of the different sample periods of the papers, the country in which pairs-trading is tested and which data is used. Since all papers are based on GGR, we only show the modification the papers made to the methodology of GGR. We also include the main conclusion of the papers.

2.5. Equilibrium Model

Kondor developed a equilibrium model of convergence trading and its impact on asset prices. Kondor describes the situation where there are two identical stocks which are traded in different markets. The arbitrageur shorts the expensive stock and buys the cheap stocks. The model captures the opportunity that arbitrageurs face, because of a temporary price difference of the two stocks. The source of the divergence does not matter in this model.

Kondor describes the equilibrium of the model, using four steps. First, at each point in time the arbitrageur has to decide the amount of capital he commits to the arbitrage opportunity. This problem is twofold: If he decides to commit a large part of his capital, he has less capital to invest in an arbitrage opportunity with a larger spread, hence a more profitable opportunity. On the other hand, if he does not commit any capital he faces the risk that the arbitrage opportunity disappears. Second, arbitrageurs are indifferent to how they allocate their capital across time. Third, the market should be cleared. Fourth, the price spread should be picked in such a way that it is consistent with the capital constraints of the arbitrageurs. From this equilibrium Kondor reaches two main conclusions: One, '…as long as the window (arbitrage opportunity) survives, the gap must increase and each arbitrageur must suffer losses…' Kondor emphasizes is not a consequence of the liquidation of positions by arbitrageurs due to the incurred losses, but a consequence of the equilibrium, which requires the gap to widen to provide sufficiently high returns to those who wait. Two, even if prices converge and a arbitrage opportunity exists, arbitrageurs refrain from taking a position. There is always a possibility that the gap widens even further, giving rise to even higher profits. Moreover, it can be very costly to maintain the strategy if the gap keeps widening.

3. Data Description

To test our pairs-trading strategy we use daily data over the period January 2003 to December 2009. For each security we collect both the Total Return Index (RI) and the Trading Volume (VO). The RI is used instead of the regular stock price, because in the RI dividends are reinvested. Since dividends contribute positively to a long stock position and negatively to a short stock position, using the RI gives a better approximation of the returns generated from pairs-trading. VO is used to set-up pairs trading criteria. The VO is used to identify illiquid stocks and delisted stocks. In our study we analyze four different markets: the U.S., Japan, Hong Kong and China. For the U.S. we used all stock included in the Standard and Poor's 500 index (S&P 500) as our investment universe. To account for the 'survivorship bias', we also include all stocks that have delisted during our sample period. The survivorship bias implies that when only stocks that are currently in an index or investment universe are used when back-testing an investment strategy, results can be biased. Only the stocks that 'survive' are included and this can significantly affect the result of the back-test. Consider for instance the situation where we open a pair by buying the stock that went down and selling the stock that went up. If the stock that went down goes bankrupt, we would have a large loss on that pair. By removing the delisted stock ex-ante, we would bias are results upwards, since we eliminate the possibility that we are buying a stock that goes bankrupt. The delisted stocks are retrieved from The Center of Research in Security Prices (CRSP). Including the delisted stocks we have a total of 569 stocks in the U.S. The constituents of the Morgan Stanley Capital International (MSCI) Japan, Hong Kong and China index are provided by MSCI Barra.[13] The RI and VO are gathered from Thomson Reuters Datastream. For Japan we use the MSCI Japan index as our investment universe, leading to a total of 456 stocks. The investment universe in Hong Kong consists of all stocks included in the MSCI Hong Kong index, a total of 62 stocks. The MSCI China index is the investment universe in China, 152 stocks are used. The descriptive statistics are shown in Table 2.

---

[13] *The MSCI data contained herein is the property of MSCI Inc. (MSCI). MSCI, its affiliates and any other party involved in, or related to, making or compiling any MSCI data, make no warranties with respect to any such data. The MSCI data contained herein is used under license and may not be further used, distributed or disseminated without the express written consent of MSCI.*

Descriptive Statistics Sample
Sample Period: *January 2003-December 2009*

	A.U.S.	B. Japan	C. Hong Kong	D .China
Stocks	569	456	62	152
Average Return (Monthly)	1.24%	0.87%	2.50%	3.27%
Standard Deviation (Monthly)	5.11%	5.42%	7.36%	9.61%
Minimum Return (Monthly)	-18.76%	-20.77%	-25.39%	-27.19%
Maximum Return (Monthly)	18.40%	11.40%	25.65%	26.33%
Trading Days	1762	1716	1732	1741

Table 2: Descriptive Statistics of the sample

4. Methodology

The methodology of this paper is based on the methodology of GGR. We try to replicate the methodology of GGR as much as possible. However, GGR are not always specific enough to replicate the methodology one-on-one. Wherever GGR are not perfectly clear we use our own interpretation. Inevitably, this leads to some differences between our methodology and that of GGR. But by using GGR and DF, who use the same methodology, as a reliability check, we make sure that our pairs-trading algorithm does not differ too much. This serves several purposes. First of all, DF use the same methodology. DF provide an out-of-sample test of GGR over the period January 2003-July 2009. This is almost the same sample period we use, which runs from January 2004 until December 2009.[14] If we find similar results as DF this buttresses the reliability of our study. Second, our paper also provides an out-of-sample test of both GGR and DF, by testing the pairs-trading strategy in the U.S., Japan, Hong Kong and China.

4.1. Pairs-trading Cycle

Our pairs strategy consists of two stages. The first stage is called the formation period. The formation period is a 12-month period in which pairs are formed and the historical spread and standard deviation of the pairs are calculated. The second stage is the so-called trading period. During this 6-month period, the previously formed pairs are traded and the returns are generated. A new cycle is started every month, leading to a total of 67 pairs-trading cycles. This strategy can be seen as six independent pairs-trading portfolios with one month staggering. Figure 1 is a graphical explanation of the 1-month staggering strategy.
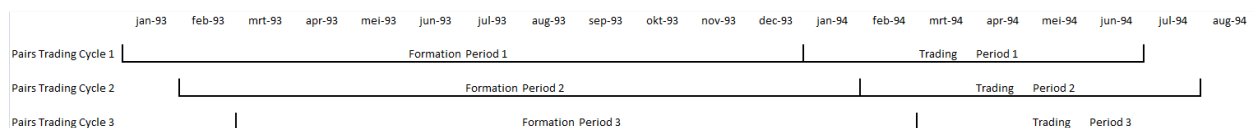
Figure 1: One month staggering trading cycles

4.2. Stock Selection

U.S. pairs are formed from stocks included in the S&P 500 index. For Japan, Hong Kong and China we use the constituents of the MSCI Japan, MSCI Hong Kong and MSCI China index. We consider our universe to be those stocks that are included in the index. Hence, our

---

[14] At the time we started writing the thesis, the paper by Do and Faff was not published yet and we were unaware of the existence of the paper. The publication of the Do and Faff paper decreased the added value of our paper. However, we now use the Do and Faff paper as a reliability check.

universe changes every year as stocks are removed and added to the index. The advantages of this approach are that we include stocks with a relatively high market capitalization and we limit our research space. The disadvantage of this approach is that MSCI decides which stocks are covered and therefore potentially exclude profitable pairs.

All stocks that have one day or more of missing trade information during the formation period are excluded. This is done in order to identify illiquid stocks and facilitate pairs formation. Stocks that have one day or more of missing trade information during the trading period are not excluded ex-ante. In fact, the delisting of stocks forms one of the risks involved in pairs-trading, so removing them would bias our results. Instead, when a stock is delisted, we use the delisting price or the last available price of that stock and close the pair.

4.3. Formation Period

At the beginning of every cycle we normalize all stocks by assigning them the value '1'. We then construct a normalized price series by chain linking the daily returns. Note that we only normalize the prices by setting the starting values to 1, we do not make any volatility adjustments. The normalized price of a stock at any point in time is given by:

$$NP_t = (1 + r_t) \times NP_{t-1} \quad (1)$$

Where NP_t is the normalized price of a stock at time t, r_t is the return of the stock at time t and NP_{t-1} is the normalized price at time t-1. The constructed series is a cumulative total return index[15] starting at the beginning of every formation period and ending at the end of the trading period. Pairs are formed based on the 'sum of squared deviation' (SSD) criterion. For every potential pair we calculate the SSD of the normalized price series during the formation period. This is an exhaustive process as the S&P 500 for instance, has a total of 124.750[16] potential pairs. The first pair consists of the two stocks that have the lowest SSD of all potential pairs. Once a stock is matched into a pair, it cannot be used in another pair, so every stock is used only once . The second pair consists of the two stocks that have the lowest SSD of the remaining potential pairs. We take three different approaches when matching the pairs. First, we match the 10 stocks that form the 5 pairs with the lowest SSD, this is our 'top-5 portfolio'. Second, we construct a 'top-20' portfolio of the 40 stocks that form the pairs with the 20 lowest SSD. Third, we match up all stocks in pairs, calling this the

---

[15] We do not account for taxes. In the real world dividend tax would lower the returns since not all dividend can be reinvested.

[16] 500*500=250.000, 250.000-500=249.500, 249.500/2=124.750

'all-portfolio'. We do this to compare return and pair statistics between the different portfolios. We expect the returns of the all-portfolio to be lower than those of the top-5 and top-20, since the all-portfolio also contain the 'left-over' pairs. These pairs might have a high SSD, but are still matched into a pair just because these are the stocks that are left over. Furthermore, GGR and DF take a similar approach, which makes our results comparable.

We implement both a 'unrestricted' matching strategy and a 'restricted' strategy. The unrestricted strategy does not require two stocks to be in the same industry to form a pair, whereas the restricted strategy does. Of course it is possible to add more restriction to the pair formation process, but this could give rise to data-snooping criticism. Data-snooping criticism usually occurs, when there are a lot of restriction in a model. The researcher could have tried several different restrictions on his model until he found those restriction that lead him to a desirable result. In this way the researcher is building a model around the results, instead of getting results from an ex-ante built model.

After all pairs are formed, the average spread and standard deviation of the spread during the formation period are calculated. These values are used later on when setting-up the trading rules of the pairs-trading strategy.

4.4. Trading Period

After pairing-up all stocks the trading period starts. We use the average spread and standard deviation of the pairs as our trading rules. If the spread of the normalized price series of a pair deviates more than two historical standard deviations from the historical average spread, we open the pair. We use the historical spread and standard deviation because we do not normalize prices to 1 again at the start of our trading period. We continue with the normalized price series which started at the beginning of the formation period. Most pairs will already have a spread larger or smaller than zero at the start of the trading period, so by using the historical spread we prevent that a lot of pairs open right at the start of the trading period. Moreover, we look for pairs which show irregular behavior. A pair which has a average historical spread of 10, with a standard deviation of 2, does not show irregular behavior if the spread is 11 somewhere during our trading period, although the spread is not zero. Figure 2 shows the normalized prices series of the stocks with the lowest SSD during the first trading period. From the picture it is obvious why these two stocks have a low SSD, they move together very closely during the formation period. As described we normalize the price series at the beginning of the formation period and use this series until the end of the trading period. A pair is opened by taking on a one dollar short position in the higher-priced stock and a one dollar long position in the lower-priced stock.

The pair is closed when the spread converged back to the average historical spread. Pairs are also closed when one of the stocks in a pair is delisted. The delisting price or last available price is then used to calculate the return of that pair. Furthermore, if on any particular day a stock is not trading[17], whereas its counterpart is, we close that pair. This is the same practice as during the formation period, only with two important differences. First of all, we might already be trading in a stock that we want to remove from our sample. Therefore we do not remove those stocks ex-ante, but instead we close those pairs using the delisting price or last available price. Second of all, we do not only do this to indentify illiquid stocks in which we do not want to trade, but mainly to cope with stocks that are
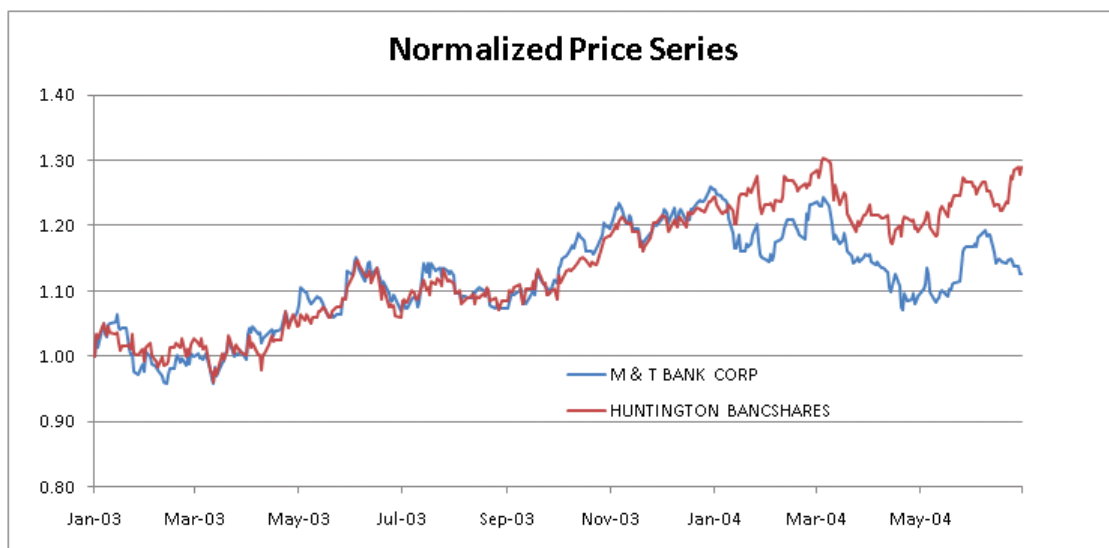
Figure 2: Normalized Price Series of M&T Bank Corp and Huntington Bancshares in the U.S.

being delisted, which could happen due to bankruptcy or mergers and acquisitions. The delisting of stocks is one of the risks involved in pairs-trading, since delisting might prevent a pair from converging[18]. For the US our total sample consists of 569 stocks[19], of which 75 are being delisted during our sample period. The total sample in Japan consists of 456 stocks of which 124 are delisted, there are 30 stocks delisted in Hong Kong on a total sample of 62 and in China 21 stocks are delisted on a total of 152 stocks in our sample. The number of stocks that are delisted is relatively high in Hong Kong and China. This might be

---

[17] We define not trading as: A day on which the stock has no (zero) volume, but the stock market is open.
[18] It could also close the spread instantly if it concerns for instance a takeover of a stock which is long in a pair, and where the takeover price lies (high) above the current market price.
[19] These are all the stocks that are used during our sample period. This number is higher than 500, since stocks are removed an added to the index throughout the sample period. Suppose we have an index of only 2 stocks. If one of those stocks is removed from the sample and replaced by another, brings the total of stocks used to 3.

because Hong Kong and China are more dynamic markets compared to Japan and the US and thus have a higher company turn-over.[20]

In their paper, GGR address the issue of the bid-ask bounce. They argue that when the stocks diverge, the price of the higher priced stock is most likely to be an ask quote, whereas the price of the lower priced stock is probably the bid quote. The opposite is true when the stocks converge again. When closing the pair the price of the short is probably the bid quote and the price of the long is most likely the ask quote. This means that a part of the returns generated from the pairs-trading strategy comes from the bid-ask bounce. GGR address this issue by also calculating pairs returns when waiting one day before opening the pair after the spread hit the two standard deviation mark and closing one day after they converged again. In our paper we use the closing prices of stocks, so there is always a full trading day between the signal and the opening of the pair. We thus expect the bid-ask bounce to be less of an issue in our research, but since we do not know which prices GGR use, we cannot judge if this is different from their paper. So, to be sure we also test the pairs-trading strategy with one day extra waiting before opening.

4.5. Return Calculation

We open a pair when the spread diverges by more than two historical standard deviations. We close the pair again when the spread converges back to the historical average. This means that during our trading period pairs might open and close multiple times, generating multiple cash flows. Pair returns are calculated on a daily basis and are marked-to-market daily. For each pair we calculate the return of both the long and short position separately and add them to get the total pair return.

At the day we get the two standard deviation signal we give the return series of both the long and short position value 1. This is done to be able to chain link the returns. We enter the trade the day after the signal. First we calculate the daily stock returns of the normalized price series. For stock A of pair N, we chain link the returns in the following way:

$$NPP_{at} = (1 + r_{at}) \times NPP_{at-1} \ (2)$$

NPP_{at} is the normalized price of stock A at time t, r_{at} is the return of stock A at time t and NPP_{at-1} is the normalized price of stock A at time t-1. We now get a cumulative chain linked return for both sides of the pair. We subtract one from both series to get the cumulative returns. Daily returns are calculated by subtracting the cumulative return of the day before

---

[20] With company turn-over we mean the number of companies that are being added en removed from the index.

from the cumulative return today. For each pair we now have two series of daily returns. By adding those two daily returns series together we get the daily returns of the pair. When the pair closes, we set the chain link series of each stock back to zero. Since pairs can open and close again several times, note that we enter each pair trade with one dollar every time it opens. So we do not continue with the previously generated return of that pair. The following table shows an example of the return calculation of a stock A of pair N.

Return Calculation Example									
Time (Days)	1	2	3	4	5	6	7	8	9
Standard Deviation of the Spread	1.70	2.10	2.30	1.20	0.00	0.80	2.20	2.10	1.80
Pair Open	no	no	yes	yes	yes	no	no	yes	yes
Daily Return Stock A (%)	1.00%	1.50%	2.00%	-1.00%	1.50%	3.00%	0.50%	-1.00%	2.00%
Long/Short/Flat	F	F	L	L	L	F	F	S	S
NPP A	0.00	1.00	1.02	1.01	1.02	0.00	1.00	1.01	0.99
Cumulative Return A(%)	0.00%	0.00%	2.00%	0.98%	2.49%	0.00%	0.00%	1.00%	-1.02%
Daily Return A (%)	0.00%	0.00%	2.00%	-1.02%	1.51%	0.00%	0.00%	1.00%	-2.02%

Table 3: Return Calculation Example

Time (Days) shows the time in days, *Standard Deviation of the Spread* gives the spread of the pair in standard deviations. *Pair Open* indicates whether the pair is open or not, *Daily Return Stock A (%)* is the percentage daily return of the normalized prices series of stock A in pair N. *Long/Short/Flat* indicates whether the position of stock A in the pair is long, short or flat. *NPP A* is the normalized prices series of the returns of stock A in pair N. This is done to chain link the pair returns, hence this value is zero when the pair is closed. *Cumulative Return A (%)* is the cumulative return in percentage of stock A in pair N and *Daily Return A (%)*, is the daily percentage return of stock A in pair N.

We use the individual pair returns to calculate the total pairs-trading strategy returns. There are two ways to calculate the total pairs-trading strategy returns. The first way is what GGR call the 'return on committed capital'. GGR argue that it might be possible that portfolio managers have to allocate funds to a pair, whether or not the pair is trading. They have to reserve this fund, being able to open a pair when the spread hits the trigger. Consider the situation when we have identified 10 potential pairs. If we enter every pair with 1$, under the committed capital rule we have to allocate 10$ to our strategy. Even if only 5 pairs open, we still divide our total pay-off by 10$. This method of calculating returns is fairly conservative. Since we allocate a dollar even if the pair is not open, there are opportunity costs for the dollar that is not invested. GGR argue that this might not be the most realistic approach. Since hedge funds are relatively flexible in their sources and fund allocation, it is more realistic that they only allocate a dollar to a pair that is actually open. The second way

to calculate total strategy returns is the 'return on fully invested capital'. We assume that hedge fund allocate their fund only to the pairs that trade and use the remaining resources for different strategies. Returns are now calculated as the average return of the open pairs.

Note that in both cases the strategy return is calculated by dividing the sum of all individual pair returns divided by N pairs.[21] This implies that all pairs are equally weighted. In practice this would imply that we have to rebalance on a daily basis. Not only does this lead to a high turnover and thus high transaction cost, it is also a rather unrealistic assumption of how the pairs-trading strategy is implemented in the 'real world'. If any, this would bias our results downwards. First of all, by rebalancing we subtract some funds from our winning pairs and add some funds to our losing pairs. The effect of the winning pairs on total returns is thus decreased and the effect of the losing pairs is increased. Second, rebalancing leads to increased transaction costs. When rebalancing, one is buying and selling stocks on a daily basis, increasing costs of implementing the pairs-trading strategy and thus lowering the profits.

We start a new 6-month trading period every month, leading to a total of 67 trading periods. This results in a daily return series of six overlapping trading periods. Daily strategy returns are calculated by taking the average return of those six overlapping periods. In this way we also correct for correlation caused by the overlap in the same fashion as Jegadeesh and Titman (1993). The monthly average return and standard deviation are calculated in the following way:

$$Average\ Monthly\ Return = \left((\textstyle\sum_{t=1}^{x} R_{pt})^{\frac{1}{x}} \right)^{20} \quad (3)$$

$$Monthly\ Standard\ Deviation = \sigma_{DSR} \times \sqrt{20} \quad (4)$$

Where R_{pt} is the return of the pairs-trading strategy at time t and σ_{DSR} is the daily standard deviation of the daily pairs-trading strategy returns vector.

---

[21] N being either all potential pairs, or pairs that are actually open

5. Empirical Results

In the following section we discuss the empirical results of our pairs-trading strategy. We discuss the return characteristics of the strategy, both with unrestricted pairs formation and restricted pairs formation. First, we test the reliability of our model by comparing our results with those of GGR and DF. Second, we compare the results of pairs-trading in the U.S. with pairs-trading in Japan, China and Hong Kong to see if they are different. Third, we analyze the returns of the restricted strategy. Again the U.S. results are compared with GGR and DF and with the results in Japan, Hong Kong and China. Fourth, we perform a sub-period analysis to see if there is a pattern in pairs-trading returns. Fifth, we make a risk analysis by regressing the pairs-trading returns on several risk factors.

5.1. U.S. Results versus GGR and DF

Table 4 shows the monthly return characteristics of the pairs-trading strategy in the U.S. of our study, GGR and DF. We compare our results with the one day waiting results of GGR and DF. We do this because we use closing prices and hence wait one full trading day after the trigger signal before entering a trade. As a check we also implemented the strategy with one day extra waiting.[22] In the last row of panel A we can see that the average returns are not significantly affected if we wait one day extra.

Return Characteristics	A. U.S.			B. GGR			C. DF
Sample Period: *January 2004-December 2009*				*July 1963-December 2002*			*January 2003-June 2009*
Pairs Portfolio	Top-5	Top-20	All	Top-5	Top-20	All	Top-20
Average Excess Return *(fully invested)*	-0.16%	0.32%	0.22%	0.75%	0.90%	0.72%	0.24%
t-Statistic	-0.59	1.99	1.12	6.26	9.29	7.92	2.73
Standard Deviation	2.38%	1.41%	1.71%	2.10%	1.53%	1.58%	-
Sharpe Ratio	-0.07	0.23	0.13	-	-	-	0.24
Skewness	-1.16	0.42	1.31	0.34	1.45	3.32	-
Kurtosis	18.99	2.27	12.69	10.64	16.13	25.66	-
Minimum	-8.29%	-2.73%	-4.37%	-12.63%	-8.22%	-2.30%	-
Maximum	7.87%	8.15%	11.14%	14.35%	13.49%	16.33%	-
Negative Returns	46%	50%	54%	35%	23%	32%	-
Average Excess Return *(committed capital)*	-0.02%	0.19%	0.10%	0.46%	0.52%	0.39%	0.24%
Average Excess Return *(1d waiting)*	-0.17%	0.33%	0.20%	-	-	-	-

Table 4: Monthly U.S. Return Characteristics of our study, GGR and DF

Panel A shows the return characteristics of the U.S. pairs-trading strategy. In row 1 we find the monthly average excess return of the fully invested pairs-trading strategy. If we

---

[22] This means that we receive a signal at the closing of day 0 and start trading at the closing price of day 2.

compare our results with GGR we see that our returns are significantly lower for all-portfolios. For the top-5 portfolio we find an insignificant return of -0.16% compared to 0.75% for GGR. The top-20 portfolio of GGR generates a return of 0.90%, which is significantly higher than the return of 0.32% we find. The returns on the all-portfolio are 0.22% and 0.72% for us and GGR respectively. The standard deviations are comparable, for GGR: 2.10%, 1.53% and 1.58% for the top-5, top-20 and all-portfolio, respectively. We find: 2.38%, 1.41% and 1.71% for the top-5, top-20 and all-portfolio, respectively. The returns for the 2003-2009 period are significantly smaller than those for the 1963-2002 period. Since DF use almost the same sample period as we do, comparing our results with those of DF would also be a good reliability check. DF only report the results of the top-20 fully invested unrestricted strategy, with one day waiting. DF report a declining trend in pairs-trading. The returns they find for the 2003-2009 period are lower than the returns in the preceding periods. DF find an average excess return of 0.24%, with a Sharpe ratio of 0.24. We find an average excess return which is slightly higher: 0.32%. However, apparently this higher return comes at a cost: we find a Sharpe ratio of 0.23, implying that the volatility of returns is higher compared to DF. Furthermore we see that the percentage of negative returns is almost the same in our study as in DF, 50% compared to 47%.

A. U.S. Pair Characteristics				B. GGR		
Sample Period: January 2004-December 2009				July 1963-December 2002		
Pairs Portfolio	Top-5	Top-20	All	Top-5	Top-20	All
Percentage of pairs traded per period	97.3%	97.9%	94.7%	96.2%	96.5%	-
Average trip of pairs in days	68.7	68.7	73.1	75.0	75.2	79.4
Mixed Sector Pairs	30%	47%	70%	20%	22%	33%

Table 5: Pairs statistics of our study and GGR

Table 5 compares some pairs statistics of GGR and our study. Almost all our pairs open during the trading period, about 97% for the top-5 and top-20 pairs and about 95% for all pairs. These results are similar as GGR, who find that on average 4.81 (96.2%) and 19.30 (96.5%) pairs trade of the top-5 and top-20 portfolio. The period that a pair is actually open is called the 'trip' of a pair. The average trip of the pairs is around 70 days for all pairs, which is close to the 75 days[23] GGR find for the top-5 and top-20 portfolio. There is a difference between the number of mixed sector pairs we find and the number of mixed sector pairs GGR find. We find an average ranging from 30% to 70% from the top-5 to all pairs portfolio. GGR find an average of 20% mixed sector pairs for the top-5 portfolio and an

---

[23] GGR find an average of 3.75 months, which is 75 days based on 20 day month.

average of 33% mixed sector pairs for all pairs. This difference might be explained by the fact that we use 9 different sectors classifications and GGR use only 4 different sectors. Two stocks that are in the same industry according to GGR might be in a different industry according to our industry classifications.

The main objective of testing pairs-trading in the U.S. over the period January 2004-December 2009, is to check if our results are reliable. First of all, our study confirms the declining trend in pairs-trading returns observed by DF. The pair returns we find are significantly smaller than those of GGR, but comparable to those of DF. Second, we see that the pairs statistics of our study are comparable to those of GGR. The comparison of the return and pair characteristics of our study with GGR and DF buttresses the reliability of our results.

5.2. Return Characteristics Japan, Hong Kong and China

Pairs-trading returns show a declining trend in U.S. Would pairs-trading in other countries yield different returns? Is pairs-trading a profitable investment strategy in Japan, Hong Kong or China?

Panel B of table 6 shows the return characteristics of pairs-trading in Japan, which we consider to be a developed market like the U.S. The first thing we see is that the monthly average excess returns in Japan are significantly higher than in the U.S.; 0.63%, 0.76% and 0.77% for the top-5, top-20 and all-portfolio respectively. These returns are close to GGR: 0.75%, 0.90% and 0.72% for the top-5, top-20 and all-portfolio, respectively. The standard deviations in Japan are slightly higher than those in the U.S. The Sharpe ratio of the top-20 portfolio in Japan is about twice as large as in the U.S.: 0.43 versus 0.23. Pairs-trading in Japan appears to be much more profitable than in the U.S. over the 2004-2009 period. If we look at the minimum, maximum and percentage of negative returns, we see that there are some differences between Japan and the U.S. Interestingly enough we see that both the minimum and maximum return of the top-20 portfolio in Japan are lower than in the U.S., so this does not explain the higher profitability in Japan. The difference might be explained by looking at the percentage of months with a negative return. In the U.S. 50% of the months show a negative return for the top-20 portfolio. In Japan this is only 31%. Apparently, there is much more consistency in positive pairs-trading returns in Japan than in the U.S. In Japan, pairs-trading shows a positive return 70% of the time, which is close to the 77% of positive returns GGR find and much larger than the 53% DF find during the 2003-2009 period.

Pairs Portfolio	A. U.S.			B. Japan			C. Hong Kong			D. China		
	Top-5	Top-20	All	Top-5	Top-20	All	Top-5	Top-20	All	Top-5	Top-20	All
Average Excess Return *(fully invested)*	-0.16%	0.32%	0.22%	0.63%	0.76%	0.77%	0.27%	0.43%	0.43%	0.63%	0.03%	-0.09%
t-Statistic	-0.59	1.99	1.12	1.91	3.66	3.53	0.58	0.98	0.98	1.07	0.07	-0.19
Standard Deviation	2.38%	1.41%	1.71%	2.84%	1.79%	1.85%	4.01%	3.74%	3.74%	5.05%	3.85%	3.87%
Sharpe Ratio	-0.07	0.23	0.13	0.22	0.43	0.41	0.07	0.11	0.11	0.12	0.01	-0.02
Skewness	-1.16	0.42	1.31	0.02	0.25	0.80	0.26	0.08	0.08	0.19	0.12	0.18
Kurtosis	18.99	2.27	12.69	2.80	1.49	3.67	3.57	2.09	2.09	2.43	1.85	1.92
Minimum	-8.29%	-2.73%	-4.37%	-5.84%	-3.79%	-3.49%	-7.78%	-10.06%	-10.06%	-8.82%	-9.05%	-8.98%
Maximum	7.87%	8.15%	11.14%	10.07%	6.49%	7.73%	15.63%	12.62%	12.62%	16.24%	10.37%	12.93%
Negative Returns	46%	50%	54%	38%	31%	39%	53%	43%	43%	42%	47%	51%
Average Excess Return *(committed capital)*	-0.02%	0.19%	0.10%	0.26%	0.47%	0.40%	0.19%	0.13%	0.13%	0.47%	0.02%	-0.11%
Average Excess Return *(1d waiting)*	-0.17%	0.33%	0.20%	0.39%	0.74%	0.73%	0.23%	0.56%	0.56%	0.54%	0.03%	-0.04%

Table 6: Return Characteristics in the U.S., Japan, Hong Kong and China

Hong Kong and China form the other markets of our study. The first thing we see is that all average returns in Hong Kong and China are insignificant, with *t-statistics* well below the 5% significance threshold-value of 1.645. If we look at the top-20 portfolios in both countries we see that Hong Kong shows an average return of 0.43% and China a return of only 0.03% per month. Note that the average returns in Hong Kong are higher than in the U.S., but still insignificant, which is caused by the higher standard deviation of the pair return in Hong Kong. This is also reflected in the Sharpe ratios of Hong Kong and China. The Sharpe ratios are well below those of Japan and the U.S., ranging from a maximum of 0.12 for the top-5 portfolio in China, to a minimum of -0.02 for the all-portfolio in China.

Clearly, pairs-trading is much riskier in Hong Kong than in Japan and the U.S., using volatility as a risk measure. Obviously, the minimum and maximum returns in Hong Kong and China are also larger than in the U.S. and Japan. The minimum returns are -10.06% and -9.05% in Hong Kong and China, compared to -2.73% and -3.79% in the U.S. and Japan. The maximum returns in Hong Kong and China are 12.62% and 10.37% and 8.15% 6.49% in the U.S. and Japan. The percentage of negative returns in Hong Kong is 43% and 47% in China. These percentages are higher than in Japan, and slightly lower than in the U.S.

Comparing pairs-trading in the four different markets, there are some apparent differences. Most notably is the difference in volatility of the returns. Pairs-trading in Hong Kong and China is much riskier than in Japan and the U.S. If we look at the monthly stock level volatility of the different countries, wee that this does not explain the differences. The average monthly stock level volatility is calculated by:

$$\sigma_c = \frac{1}{N} \, X \sum_{i=1}^{N} \sigma_i \quad (5)$$

Where σ_c is the average monthly stock level volatility in country c, N is the number of stocks and σ_i is the monthly volatility of stock i. The average monthly stock level volatility in the U.S. is 6.49%, in Japan 6.26%, in Hong Kong 6.50% and China 8.15%. Obviously the volatility is the highest in China, which might explain the higher volatility of the pairs-trading returns in China, but there appears to be no significant difference between Hong Kong and Japan and the U.S. Despite the differences in returns and volatilities we do see that there are diversification benefits in all four markets. In all four markets, the top-20 portfolio has a lower standard deviation than the top-5 portfolio. However, these diversification benefits disappear when more pairs with higher SSD's are added. In all four markets the all-portfolio has a higher standard deviation than the top-20 portfolio. This is because pairs with higher SSD's consist of stocks which are less homogenous than the top pairs. These higher SSD pairs have a higher probability of large returns –either negative or positive- since there is a possibility that these pairs will have a large divergence, which is somewhat less the case for more homogenous pairs. Homogenous pairs probably share more fundamentals, forcing the prices to move together more closely. The increased volatility of the all-portfolio implies that there is a number of pairs which maximizes the diversification benefits and hence, minimizes the standard deviation of the returns. If more pairs with lower SSD's are added, the benefits of diversification are reversed and the standard deviation of the pairs-trading returns increases again.

5.3. Pairs Statistics Japan, Hong Kong and China

To get a better understanding of the differences between pairs-trading in different markets, we also compare the pair statistics of the U.S. with those of Japan, Hong Kong and China. Results are shown in table 5. If we compare the pair statistics of the U.S. with those of Japan -which are shown in Panel B- we see a high degree of similarity. Almost all pairs open during the 6-months trading interval: around 98% in both Japan and the U.S. for the top-20 portfolio. The average trip of a pair, the average number of times a pair opens and the percentage of mixed sector pairs is almost equal in Japan and the U.S. We also looked at the percentage of pairs that are still open at the end of the trading interval, this is shown in row 2. At the end of the trading interval all pairs are closed. It is at least questionable that a hedge fund manager would close all of his pairs at the end of a pre-defined trading horizon. There are two main effects when closing the pairs at the end of the trading period. On the one hand, forced closure of pairs prevents pairs to converge, which causes our profits to be lower. On the other hand forced closure also closes pairs which do not converge at all. In

this case it works as a stop-loss function and improves our returns. Since we start a new trading cycle every month, both effects are limited. If we close down a potentially profitable pair, this pair would also show up in the trading period which starts one month later. If we close down a potential losing pairs, because it does not converge and hence, stops being a pair, this pair would not show up in later trading periods, since the relationship between the two has broken down. The best way to test the effect of forced closure is by removing the forced closure restriction and look at the result. However, this is beyond the scope of this paper, but it might be an interesting topic for further research. The average SSD of the pairs is somewhat higher in Japan than in the U.S. for the top-5 and top-20 pairs. If a pair has a low SSD it does not necessarily means that it will generate high returns. In fact, looking at the U.S. and Japan we see that Japan has a higher average return, but the average SSD is also higher.

There are clear differences between pairs-trading in the U.S. and Japan and China and Hong Kong when looking at the pairs statistics. The percentage of pairs that open lies around 90% for both China and Hong Kong for the top-20 pairs. This is somewhat lower than the 98% we find for Japan and the U.S. The number of pairs that are still open at the end of the trading horizon is also significantly lower in Hong Kong and China. This number lies around 60% in the U.S. and Japan, compared to around 80% in Hong Kong and China. Of course, this difference is partially explained by the lower percentage of pairs that open in the first place. However, this difference is too small to explain the full difference in pairs that are open at the end of the trading interval. The average trip is similar in all four markets. The average number of times that a pair opens is lower in China and Hong Kong than in Japan and the U.S. The SSD's are also higher than in the U.S. and Japan, especially in China. These results show that in Hong Kong and China less pairs open, they open less frequently and at the end of the trading interval less pairs are still open.

Pair Characteristics
Sample Period: *January 2004-December 2009*

	A. U.S.			B. Japan			C. Hong Kong			D. China		
Pairs Portfolio	Top-5	Top-20	All	Top-5	Top-20	All	Top-5	Top-20	All	Top-5	Top-20	All
Percentage of pairs traded per period	97.3%	97.9%	94.7%	97.0%	97.2%	94.0%	92.0%	88.1%	88.1%	94.0%	89.0%	87.0%
Percentage of pairs still open at end of trading period	79.7%	79.4%	77.3%	78.8%	77.9%	75.7%	69.3%	61.4%	61.4%	60.6%	60.2%	58.5%
Average trip of pairs in days	68.7	68.7	73.1	63.5	67.1	71.2	70.4	71.5	71.5	61.4	68.0	68.9
Average times pairs open during trading period	1.73	1.72	1.50	1.74	1.72	1.52	1.47	1.38	1.38	1.63	1.45	1.41
Average SSD	0.16	0.23	5.79	0.21	0.33	2.25	0.60	0.60	4.99	1.17	4.24	10.23
Mixed Sector Pairs	30%	47%	70%	34%	51%	69%	61%	69%	69%	73%	78%	79%

Table 7: Pair Statistics in Japan, Hong Kong and China

The percentage of mixed sector pairs is significantly higher in China and Hong Kong: 69% for the top-20 portfolio in Hong Kong and 78% for the top-20 portfolio in China. This could be an explanation for the higher volatility we find. It might also explain some of the difference we see between the returns on the all-portfolios and the top-portfolios. DF argue that homogenous pairs reduce the nonconvergence risk. They argue that: 'Fundamentally similar assets are likely to converge, and if they do not, they are not likely to drift apart'.

In the next section we only analyze the top-20 portfolio. The top-20 portfolio is the main portfolio analyzed by GGR and is the only portfolio analyzed by DF. Moreover we feel that the top-20 portfolio is the best for pairs-trading, since it combines having homogeneous pairs, when defining pairs with a low SSD as homogeneous, with diversification benefits.

5.4. Industry Specific Pairs-trading

GGR and DF also analyze industry specific pairs-trading. They formed pairs from stocks that are within the same industry and analyzed the returns of the different industries. This strategy is referred to as the 'restricted' pairs-trading strategy. Both GGR and DF tested the restricted pairs-trading strategy by dividing all stock into four industry groups: Utilities, Financials, Transportation and Industrials. GGR find that Utilities is the best performing sector with an average monthly excess return of 1.08% for the top-20 pairs. Financial is the second-best with a return of 0.78%, followed by Industrials (0.61%) and Transportation (0.58%). DF confirm these results. They find that for the full sample, Utilities, Financials, Transportation and Industrials generate returns of 0.64%, 0.75%, 0.50% and 0.46% respectively, for the full sample.

We repeated the restricted pairs-trading strategy in our study. However, we used the 9 industry sectors specified by Bloomberg: Basic Materials, Communications, Cyclical Consumer, Non Cyclical Consumer, Energy, Financial, Industrial, Technology and Utilities to divide our stocks into sectors. This serves two purposes. First, the four major industry groups used by GGR and DF are S&P industry groups, which are not available for the Asian markets. Second, dividing our investment universe into only four different industry groups still allows for a lot of heterogeneity within those groups. Dividing our universe into 9 groups, creates a higher degree of homogeneity within the groups. Figure 3 shows the industry sector breakdown of the different pair portfolios.[24] Since the all-portfolio contains all stocks of the market, it gives us an overview of the relative size[25] of the different sector in

---

[24] This is the sector breakdown for the unrestricted pairs. The graph tells us how many stocks of a specific sector are matched into a pair, whether or not with this is with a stock in the same sector.
[25] The percentage is based on the number of stocks, not on the market cap of the stocks.

each market. We see that especially the stocks in the U.S. and China are spread relatively equal over all sectors. In Japan the Consumer Cyclicals, Consumer Non-Cyclicals, Financials and Industrials have a high presence. The Financial sector is by far the most dominating sector in Hong Kong.

In the top pairs, Utilities have a high presence in the top pairs in all markets. This is consistent with the findings of GGR, who show that 71% of the top-20 portfolio consists of Utility stocks. Financials are also matched into the top pairs relatively often. Furthermore we see that Technology and Industrial stocks are rarely matched into the top pairs, especially in the U.S. and Japan.
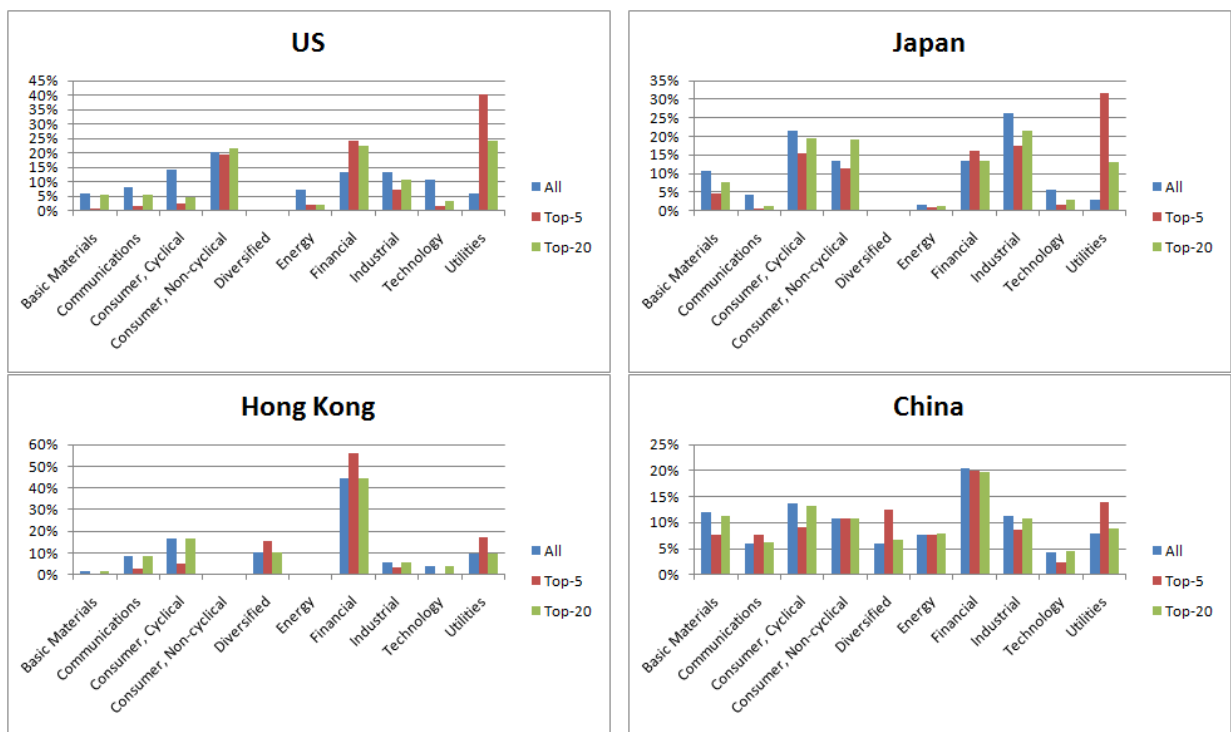
Figure 3: Sector breakdown of the pairs portfolios

Some industry groups do not have enough stocks to match 20 pairs, for these groups we matched all stocks into pairs and show the results. We did not analyze every industry group in all the four markets. Especially Hong Kong and China do not always have enough stocks in a sector to set up pairs, as can be seen from the number of stocks per industry in row 1 of Panel A, B, C and D of table 8. If there are less than 16 stocks in a sector we did not report the results. This number is somewhat arbitrary, but we feel that if we use even less stocks the results are not representative any more. If there are only a few potential pairs, it is possible that during parts of the sample period there are no pairs at all. The exception is Utilities in Japan, which we included although there are only 13 Utilities. This is because in

the GGR and DF study, Utilities is among the best performing sectors and because Utilities have a high presence in the top-pairs. We do not find the same results as GGR and DF did. Financials is the best performing industry with a return of 0.93%. Utilities however, is one of the worst performing sectors with an average excess return of only 0.08% per month. This is lower than the average of 0.26% DF find over the same period. Noteworthy is the standard deviation of 3.93% of the Financials, compared to the 1.79% of GGR. Of course this increased volatility can be explained by the recent financial crisis.

Return Statistics
Sample Period: *January 2004-December 2009*

A. US

Sector	Basic Materials	Communications	Conumer, Cyclical	Consumer, Non Cyclical	Energy	Financial	Industrial	Technology	Utilities
Stocks	34	52	78	119	41	80	70	60	31
Average Excess Return *(top- 20 fully invested)*	-0.19%	0.07%	0.09%	0.33%	0.08%	0.93%	0.73%	0.49%	0.08%
t-Statistic	-0.37	0.16	0.32	1.80	0.21	2.07	2.99	1.51	0.29
Standard Deviation	4.43%	3.47%	2.48%	1.64%	3.49%	3.93%	2.12%	2.83%	2.28%
Sharpe Ratio	-0.04	0.02	0.04	0.20	0.02	0.24	0.35	0.17	0.03

B. Japan

Sector	Basic Materials	Communications	Conumer, Cyclical	Consumer, Non Cyclical	Energy	Financial	Industrial	Technology	Utilities
Stocks	45	23	100	63	9	71	106	25	13
Average Excess Return *(top- 20 fully invested)*	0.85%	1.44%	0.93%	0.97%	-	1.04%	0.89%	1.66%	0.87%
t-Statistic	2.10	2.03	3.20	3.61	-	2.98	3.24	3.24	2.31
Standard Deviation	3.48%	6.04%	2.48%	2.29%	-	2.97%	2.35%	4.35%	3.15%
Sharpe Ratio	0.25	0.24	0.37	0.42	-	0.35	0.38	0.38	0.27

C. Hong Kong

Sector	Basic Materials	Communications	Conumer, Cyclical	Consumer, Non Cyclical	Energy	Financial	Industrial	Technology	Utilities
Stocks	3	7	11	1	0	22	5	2	3
Average Excess Return *(top- 20 fully invested)*	-	-	-	-	-	0.67%	-	-	-
t-Statistic	-	-	-	-	-	1.00	-	-	-
Standard Deviation	-	-	-	-	-	5.70%	-	-	-
Sharpe Ratio	-	-	-	-	-	0.12	-	-	-

D. China

Sector	Basic Materials	Communications	Conumer, Cyclical	Consumer, Non Cyclical	Energy	Financial	Industrial	Technology	Utilities
Stocks	16	10	24	18	12	32	22	4	8
Average Excess Return *(top- 20 fully invested)*	-0.67%	-	0.26%	-0.25%	-	-0.08%	0.08%	-	-
t-Statistic	-0.47		0.20	-0.20		-0.08	0.07	-	-
Standard Deviation	12.09%	-	11.13%	11.03%	-	7.61%	11.03%	-	-
Sharpe Ratio	-0.06	-	0.02	-0.02	-	-0.01	0.01		-

Table 8: Return characteristics of industry specific pairs

One of the other interesting things we find is the high degree of dispersion between the different industries. The worst performing industry is Basic Materials with a return of -0.19%. The best performing industry is Financial, with a return of 0.93%. If we take the Sharpe ratio as a performance measure, Industrials is the best performing industry in our sample with a Sharpe ratio of 0.35, against a Sharpe ratio of 0.24 of the Financials. Using the restricted pairs-trading algorithm does not necessarily increase the returns in the U.S. Although some industries have higher average returns than the unrestricted strategy, there are as many industries that generate lower returns.

In Japan the best performing sector is Technology and the worst is Basic Materials with returns of 1.66% and 0.85% respectively. The increase in returns of the restricted pairs strategy is significant. The monthly average excess return of the unrestricted strategy is 0.76% for the top-20 portfolio. Hence, the worst performing sector in Japan still outperforms the unrestricted strategy. Considering the Financials and Utilities again we see the same picture as we did for the US. The Financials belong to the top performers with a return of 1.04%, whereas Utilities is the second worst performing sector with a return of 0.87%. Looking at the Sharpe ratios we see that the Non Cyclical Consumer stocks is the best performing industry, with a Sharpe ratio of 0.42. Interestingly enough we see that the unrestricted pairs portfolio shows a superior Sharpe ratio of 0.43. Clearly there are huge diversification benefits in Japan. The standard deviation of the unrestricted pairs portfolio is only 1.79%, compared to 2.29% for the Non Cyclical Consumer stocks, which has to lowest standard deviation of all sectors.

Financials is the only sector in Hong Kong that has enough stocks to implement a reasonable pairs strategy. Still we consider this sector to see if the Financials outperform the unrestricted pairs. The Financials have a monthly average excess return of 0.67%, with a standard deviation of 5.70%, leading to a Sharpe ratio of 0.12. The average return, standard deviation and Sharpe ratio of the unrestricted strategy are 0.43%, 3.74% and 0.11 respectively. Again there are clearly diversification benefits, as the standard deviation of the unrestricted strategy is lower, as is the average return. The Sharpe ratio is about the same for the unrestricted strategy and the Financials.

For China we analyzed 5 different sectors: Basic Materials, Consumer Cyclical, Non Cyclical Consumers, Financials and Industrials. Basic Materials is again the worst performing sector with a return of -0.67%. The best performing sector is Cyclical Consumers, with a return of 0.26%. Only Consumer Cyclical and Industrials outperform the 0.03% average return of the unrestricted strategy. Moreover we see that all industries have high standard deviations, leading to low Sharpe ratios. The benefits of diversifications are especially high in China, since the standard deviation of the unrestricted strategy is 3.85%, compared to a standard deviation of 7.61% for Financials, which is the sector with the lowest standard deviation.

As opposed to DF and also GGR, our results do not suggest that restricted pairs-trading gives superior results compared to unrestricted pairs-trading. We do find sectors with a higher average excess return than the unrestricted pairs. However, these higher returns come at a cost; the standard deviation of the restricted strategy is higher than the standard deviation of the unrestricted pairs. Consequently we do not find much sectors with a superior Sharpe ratio compared to the unrestricted strategy. Clearly there are large

diversification benefits in pairs-trading. Japan is the only country of which each sector return is higher than the unrestricted return. Picking the right sectors when setting up pairs, is crucial when applying the restricted pairs-trading strategy. This might seem harder than it is. When analyzing the different sector of the different countries, it is not so obvious which sectors to pick. According to GGR and DF, Utilities and Financials are the best sectors. We do find that Financials is one of the best performing sectors in the US, Japan and Hong Kong, however it is not in China. Furthermore we see that Utilities is one of the worst performing sector in both the US and Japan. Industrials is one of the best performing sector in the US, Japan and China, measuring performance both as average return and Sharpe ratio. Basic Materials appears to be the worst sector to apply the pairs-trading strategy, since it is the single worst performing sector in the US, Japan and China.

We also calculate the average return, standard deviation and Sharpe ratio of all sectors combined. The results are shown in table 9. We do not include Hong Kong, since we only applied the restricted strategy for the Financials in Hong Kong. The average return in the U.S. is 0.32%, with a standard deviation of 1.50%. This leads to a Sharpe ratio of 0.21. This is a higher Sharpe ratio, than the Sharpe ratio of the all-portfolio of the restricted strategy (0.13). In Japan, an average return of 1.13% and standard deviation of 1.73% leads to a Sharpe ratio of 0.65. Again, this is higher than the Sharpe ratio of the all-portfolio of the restricted strategy, which is 0.41. In China we find an Sharpe ratio of 0.07, compared to a Sharpe ratio of -0.02 for the unrestricted all-portfolio. Setting up pairs with stocks that are in the same industry group, do increase the performance of the pairs-trading strategy. Further on in this paper we will test if we can increase the performance even more, by selecting the top-20 pairs with the lowest SSD form the restricted pairs.

Return Statistics Sample Period: *January 2004-December 2009*			
	A.U.S.	B. Japan	C. China
Average Return	0.32%	1.13%	0.38%
Standard Deviation	1.50%	1.73%	5.24%
Sharpe Ratio	0.21	0.65	0.07

Table 9: Return Statistics of all pairs from the restricted strategy in the U.S. (A), Japan (B) and Hong Kong (C).

5.5. Sub-period Analysis

DF demonstrate that pairs-trading returns are especially high during bear markets. During the January 2000-December 2002 bear market the average return is 0.92%, compared to 0.22% in the preceding 1989-1999 period and 0.02% in the period 2003-2007. The mean excess return during the recent bear market of July 2007-June 2009 was 0.71%. Associated

with these higher returns is the increase in volatility of the pairs-trading returns. To see if we find a same pattern during the most recent global bear market and to analyze if pairs-trading returns are trending, mean-reverting, stable or a random-walk, we make a sub-period analysis.

Figure 4.I and Figure 4.II show the total cumulative return series (TCR) in panel A, panel B shows the market index value, panel C the 20-day market index volatility and panel D shows the percentage of open pairs in the U.S. and Japan. The percentage of open pairs is calculated as the average number of pairs that are open, divided by the number of pairs during the six overlapping trading periods. We take the 20-day moving average of the percentage open pairs to show the trending behavior. Moreover by taking the moving average we adjust for the forced closure of the pairs at the end of every month. Because all pairs are closed at the end of the trading interval, there is a dip in percentage of pairs open at every month-end. The number of new pairs that opens at the start of a new trading period is usually smaller than the number of pairs that close at the end of the month. In this way the graph represents the trend in percentage of pairs open caused by price movements, not by forced closure. First we look at the TCR of the U.S. The TCR is slightly negative throughout most of the sample period. From the beginning of 2008 onwards we see an upward trend, which continues until the end of our sample. The TCR becomes positive at the start of 2009. This has some important implications for the results we found earlier. We found a positive average excess return for the full sample period in the U.S. However, this positive return is only caused by a huge pick-up in pairs-trading return during the last two years of our sample. During the preceding 2004-2008 period the TCR is negative. Clearly, pairs-trading returns in the U.S. pick-up during the S&P 500 downturn. This can also be seen in the 50-day volatility of the S&P 500. As the index falls, volatility increases. The 20-day volatility shows a mean-reverting pattern, which can also be seen in the TCR and S&P 500. Interestingly enough we see that pairs-trading returns increase the most after the stock markets reached the lowest point and after the stock market volatility peaked. Looking at the percentage of open pairs, we see that near the end of our sample period the number of open pairs gradually declines, which causes a steep increase in pairs returns, since pairs that are closing generate positive returns.[26]

The TCR of Japan shows a stable increasing trend as can be seen in panel A of figure 4.II, with the exception of 2005, which appears to be a bad year for pairs-trading. Comparable to the U.S. we see that the TCR rises the steepest in the last 1.5 year of our sample period.

---

[26] This is of course not necessarily the case with forced closure, but in this case the downtrend is not only caused by the forced closure of pairs.

However, without the last 1.5 year, pairs-trading in Japan still generates a positive return. The returns are almost always positive throughout the sample period, except for 2005. Like the U.S. the TCR starts rising at the beginning of the July 2007-June 2009 bear market. Again this is accompanied by an increased volatility. As opposed to the U.S. the steep increase in pairs-trading returns is not accompanied by a declining trend in open pairs in Japan. The percentage of pairs open in Japan is quite stable throughout the whole sample period.

Figure 4: Total Cumulative Return (A), Market Index Value (B) , Market Index 20-day Volatility (C) and Percentage of Open Pairs (D) in the U.S. (I) and Japan (II)

Figure 5.III, panel A shows the TCR of Hong Kong. This chart shows a lot of resemblance with the U.S. The TCR is stable around zero until January 2009. From 2009 onwards we see a sharp increase in the TCR. This increase is the main driver of the positive average excess return we find over the full sample period. We see that the TCR of Hong Kong reaches is trough when the MSCI Hong Kong index reaches its peak at the end of 2007. At the start of
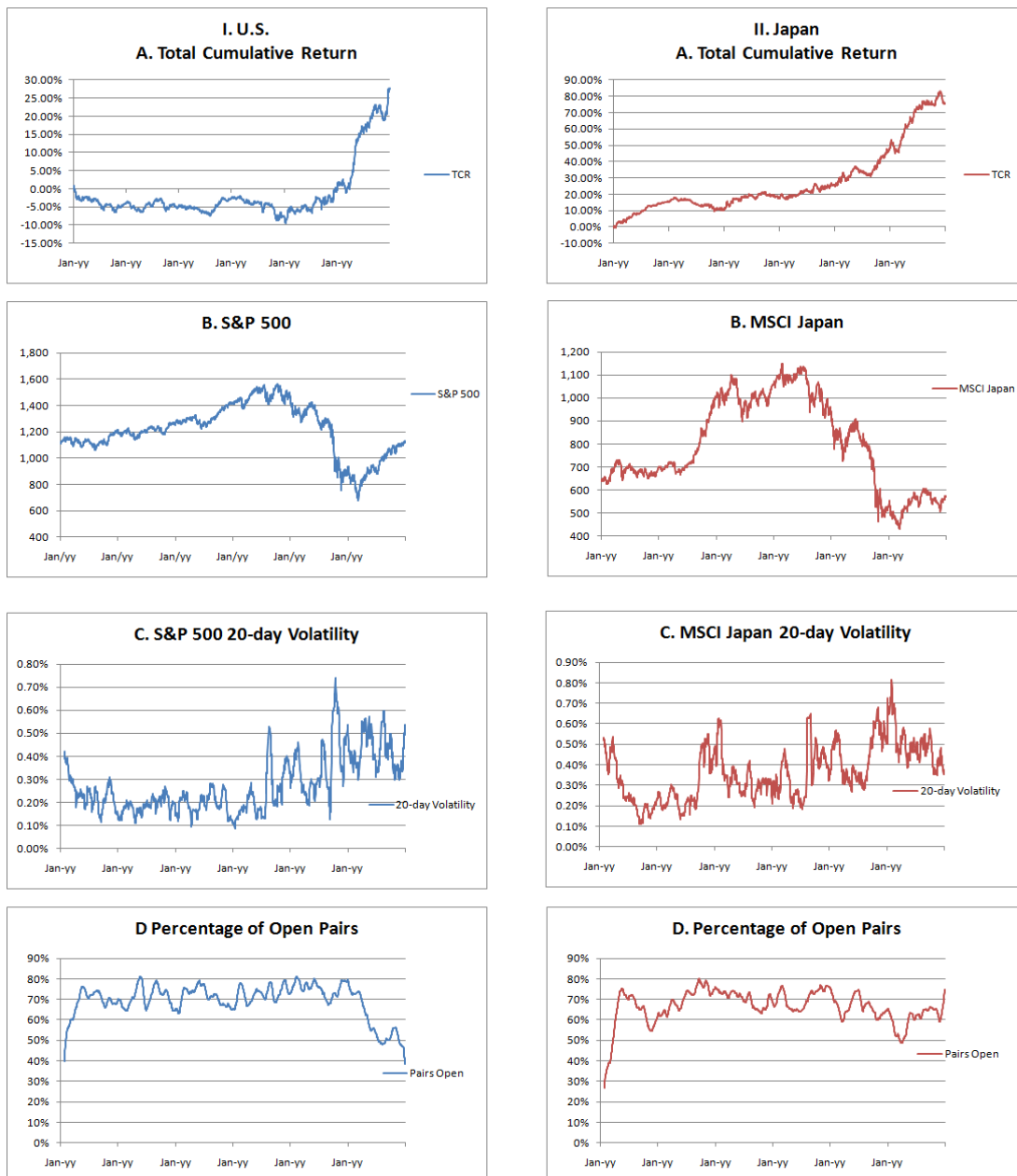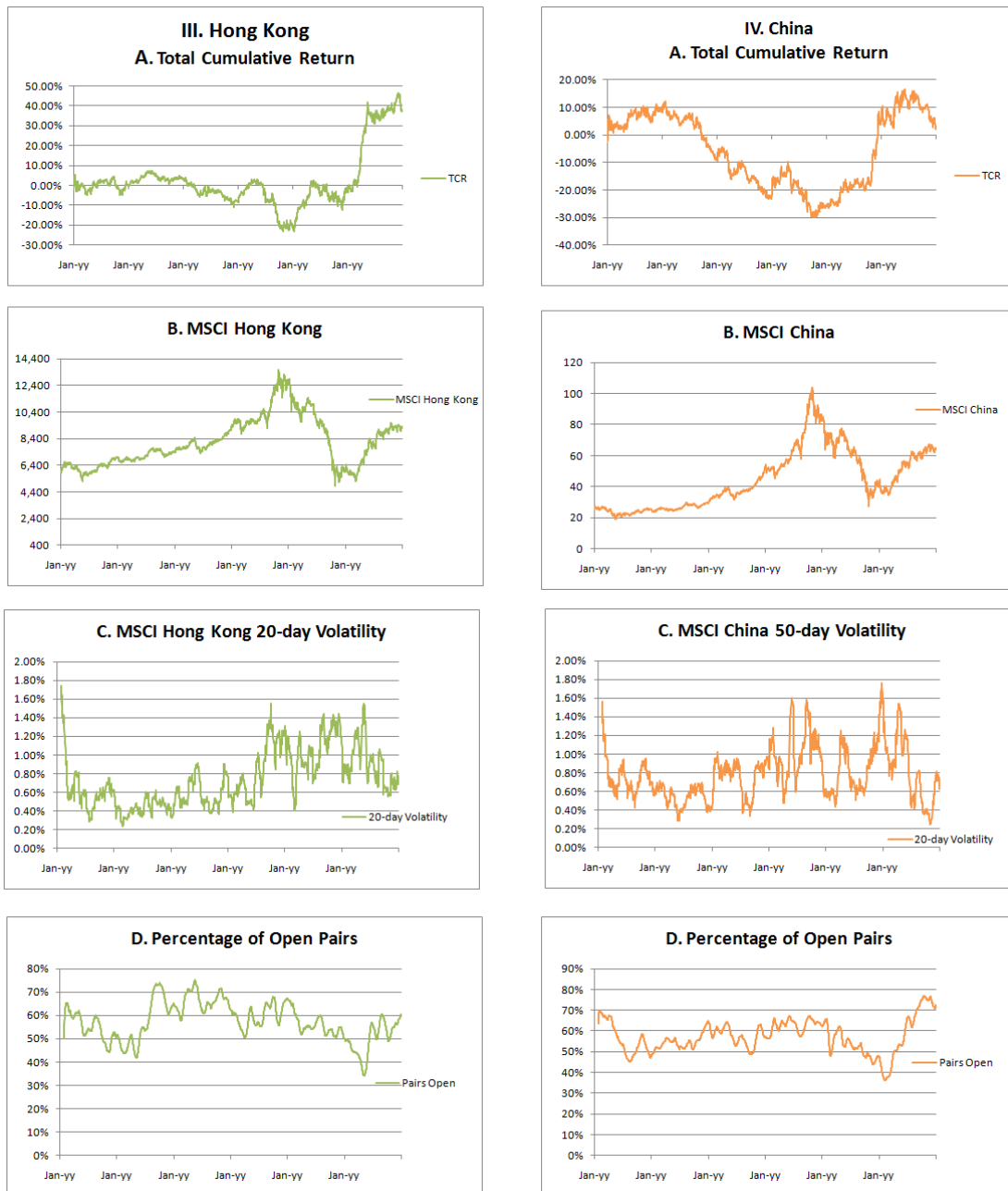

Figure 5: Total Cumulative Return (A), Market Index Value (B), Market Index 20-day Volatility (C) and Percentage of Open Pairs (D) in Hong Kong (III) and China (IV)

2008 we see that the MSCI Hong Kong index falls, corresponding with an increase of the TCR. When the MSCI Hong Kong reaches its low in January 2009, the TCR reaches its peak. The TCR shows a cyclical pattern which appears to be correlating with the stock market volatility. We see that an increased volatility leads initially to a decrease in pair returns, but when the volatility decreases again, pairs-trading returns increase. There is an increase in open pairs in 2005, but this does not lead to a decrease in pairs-trading returns. Similarly, the increase in pairs-trading returns at the end of the sample period is not accompanied by a decrease in open pairs.

Figure 5.IV shows the same charts for China. The pattern of the TCR, MSCI China index and 20-day volatility is almost the same as in Hong Kong. We see a downward trend in pairs-trading returns up until January 2007. From 2007 onwards we see a cyclical pattern in the TCR, which corresponds with a cyclical pattern in the stock market volatility. There are some differences however. The TCR shows a more negative trend until 2007, reaching its low at the end of 2006. From that point onwards, the stock market falls, corresponding with an increase in stock market volatility. Without the last financial crisis, pairs-trading in China probably would have shown a negative average return over the full sample period. Between 2005 and 2007 we see an upward trend in the percentage of pairs open, which is also reflected by a downward trend in the pairs-trading returns. Interestingly we see that at the end of the sample the number of open pairs increase sharply as do the returns.

Looking at the TCR charts for all four countries we see that pairs-trading generates the highest average return in Japan. But even more important: Japan is the only country in which pairs-trading is profitable throughout the whole sample period. In all four countries pairs-trading returns increased during the last financial crisis. Pairs-trading in the U.S., Hong Kong and China probably would not have been profitable without the recent financial crisis. Furthermore we see a cyclical pattern in Hong Kong, China and to a lesser extent in Japan. This cyclical pattern seems to be corresponding with the stock market volatility. Pairs-trading returns are correlated with changes in stock market volatility, not the level of stock market volatility. In all four countries we see that the stock market volatility is the lowest in the first years, but still there are differences in pairs-trading returns. However, in all four countries pairs-trading returns are the largest in periods of high stock market volatility. Pairs-trading appears to be especially profitable after periods of high volatility in the stock markets, when the volatility is declining. This pattern might be explained by the irrationality of investors during periods of high volatility. If certain investors behave irrational, this creates opportunities for hedge fund managers. Irrational behavior of investors may cause prices to be misaligned. However, pairs-trading does not generate profits when prices are misaligned per se, but when these misalignments are 'corrected'. This might explain the

pattern we observe. During periods of increased volatility, opportunities are created for the hedge fund managers, who then set-up pairs. When the volatility decreases again, the price misalignments in the market are corrected, leading to profits for the pairs-traders.

5.6. Regression Analysis

To asses to what extend pairs-trading returns can be explained by several systematic risk factors, we conduct a regression analysis. We carry-out two different regressions: The first regression is done with the same five factors GGR use in their analysis: *Market, Size, Value, Momentum* and *Reversal.* This is the standard Fama and French regression, with *Momentum* and *Reversal* as additional factors. We add those factors because pairs-trading is an investment strategy based on the price movement of stocks. Since both *Momentum* and *Reversal* are also investment strategies based on price movements we can investigate if pairs-trading is just a modified version of those two investment strategies, or if pairs-trading is indeed a different strategy. *Market* is the stock market return minus the risk-free rate. We take the one-month U.S. T-bill rate as proxy for the risk-free rate, since we assume to be a U.S. based investor. Since pairs-trading is a market-neutral strategy we except to see an insignificant relationship between *Market* and pairs-trading returns. It should be noted that this strategy is not completely market neutral since we do not adjust for market exposure (beta). *Size* is the Fama-French risk factor that captures the subtracts the returns of large stocks from the returns of small stocks. We are not size neutral since we match companies using SSD, which does not take size into account. It is hard to say whether *Size* should have a positive or negative beta. *Size* captures the average outperformance of small caps over large caps. EGJ find that liquidity is positively correlated with pairs-trading returns. This is caused by the fact that highly liquid stocks have a higher probability of opening and closing. Since large caps are usually more liquid than small caps, we expect that the more large caps we use in our pairs-trading strategy, the higher the returns. If we are buying large caps, this would imply a negative beta for *Size.* However, we are also shorting large caps, implying a positive *Size* beta. GGR find a positive, insignificant *Size* beta. *Value* captures the difference in return between growth and value stocks. Value stocks are stocks with a low price-to-book ratio, growth stocks are stocks with a high price-to-book ratio. Again, we are not completely value neutral. If we assume that we match two stocks that are identical based on price movements, we might see a positive sign for the *Value* beta, since we expect to short the higher price-to-book company and buy the lower price-to-book company. *Momentum* is a factor thoroughly described by Jegadeesh and Titman (1993). They assume that stocks that have performed well in the recent period continue to do so. Of course this would lead to a

negative sign in our regression because we are selling stocks that went up and buying stocks that went down. *Reversal* is described by Jegadeesh (1990) and Lehmann (1990). Reversal is an investment strategy which buys loser stocks and sells winner stocks, which is the contrary of the momentum strategy.[27] Obviously we would expect a positive sign for the reversal coefficient, since we are buying the underpriced stock and selling the overpriced stock. Besides the fact that we test for the estimated coefficient signs, we test in general if these factors can explain the returns from pairs-trading. The estimated regression has the following form:

$$R_p = \alpha + \beta_m R_m + \beta_{smb} R_{smb} + \beta_{hml} R_{hml} + \beta_{mom} R_{mom} + \beta_{rev} R_{rev} \ (6)$$

α Is the intercept, R_p is the return of our pairs-trading strategy, R_m is the return of the market, R_{smb} is the return of the size portfolio, R_{hml} the return of the value portfolio, R_{mom} is the return of the momentum strategy and R_{rev} is the return of the reversal strategy. Panel I of table 10 shows the results from this regression in all four markets. Panel A shows the regression results of the U.S. *Market* has a negative sign, but is insignificant as expected. This means that the stock market does not influence our pairs-trading returns, which is also found by GGR. *Size* is positive and significant. This implies that the stocks we sell have on average a larger market capitalization than the stocks we are buying. Since in periods that small caps outperform large caps, our pairs-trading returns are also higher. *Value* and *Momentum* are both negative and significant. GGR find a insignificant beta for *Value* and a significant, negative beta for *Momentum.* The negative *Value* beta implies that on average we are buying the higher price-to-book stocks and selling the lower price-to-book stocks. When the low price-to-books stocks outperform the high price-to-book stocks (when *Value* is positive), pairs-trading returns are negative and vice versa. As expected *Momentum* has a negative beta, confirming that we are selling winner stocks and buying loser stocks. *Reversal* is positive but insignificant. GGR also find a positive but insignificant *Reversal* beta. This means that our pairs-trading strategy is not the same as a reversal strategy. The adjusted- R^2 of the regression is 47%, which is about twice as high as the 24% GGR find. Looking at the regression results of Japan, which can be found in panel B, we see that the model has a much lower adjusted- R^2 of only 26%. This can also be seen by looking at the coefficients. *Value* and *Reversal* have the opposite beta sign as we predicted, but only *Momentum* is significant.

---

[27] Momentum and Reversal are not the exact opposite. They have different horizons, and the momentum strategy usually waits one month before buying the winner stocks to account for one-month reversal, which is in fact the phenomenon the reversal strategy is based on.

Regression Analysis
Sample Period: *January 2004-December 2009*

	A.U.S.		B. Japan		C. Hong Kong		D .China	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
I. Factor Model: *Fama-French, Momentum, Reversal*								
Intercept	0.0030	0.06	0.0067	0.00	0.0059	0.15	0.0019	0.62
Market	-0.0420	0.11	-0.0410	0.47	-0.0007	0.99	-0.0290	0.55
Size	0.2882	0.01	-0.0523	0.61	0.2342	0.21	0.1995	0.38
Value	-0.1249	0.01	-0.0668	0.35	-0.0531	0.68	0.0876	0.60
Momentum	-0.2064	0.00	-0.2160	0.00	-0.6735	0.00	-0.3386	0.01
Reversal	0.0191	0.62	-0.0528	0.35	-0.0796	0.49	-0.1562	0.10
Adj R sq.	47%		26%		30%		15%	
II. Factor Model: *Fama-French, Momentum, Reversal Dispersion, Volatility, SSD*								
Intercept	-0.0084	0.10	-0.0058	0.53	-0.0067	0.39	-0.0021	0.75
Market	-0.0461	0.14	-0.0377	0.43	0.0164	0.82	-0.0312	0.50
Size	0.2684	0.00	-0.1095	0.23	0.3391	0.10	0.1427	0.57
Value	-0.0825	0.10	-0.0999	0.13	-0.0635	0.60	0.0248	0.89
Momentum	-0.1892	0.00	-0.2116	0.00	-0.7564	0.00	-0.3621	0.01
Reversal	0.0065	0.88	-0.0699	0.16	-0.0903	0.43	-0.1468	0.17
Dispersion	-0.0161	0.05	0.0195	0.11	0.0066	0.29	0.0036	0.85
Volatility	0.0026	0.52	0.0124	0.01	-0.0207	0.14	0.0067	0.42
SSD	0.0498	0.04	0.0360	0.20	0.0213	0.11	0.0009	0.35
Adj R sq.	52%		33%		33%		13%	

Table 10: Regression analysis of several risk factors on pairs-trading returns in the U.S. (A), Japan (B), Hong Kong (C) and China (D). The U.S. factors are collect from the Kenneth French website. The Japan, Hong Kong and China factors are provided by JPMorgan. The returns are for each factor are calculated by constructing 5 portfolios and subtracting the return of the fifth portfolio from the first portfolio. We correct for possible autocorrelation and heteroskedasticity as described by Newey and West (1987)

We also find a positive and significant alpha, implying that pairs-trading in Japan does generate a risk-free returns. In Hong Kong, *Reversal*, *Size* and *Value* do not have the expected coefficient signs, but are also insignificant. *Momentum* is the only significant variable, with the expected negative coefficient. The adjusted- R^2 of the regression is 30%. The model has an adjusted- R^2 of only 15% in China. *Momentum* is the only significant variable and has a negative sign. *Size* and *Value* have a positive sign and *Reversal* a negative sign; however these variables are all insignificant.

Since pairs-trading is an investment strategy which tries to profit from market inefficiencies, we also include some variables that might give an indication about market inefficiencies and thus pairs-trading opportunities. The first variable we include is *Dispersion*. For every stock in our investment universe we calculate the average monthly return. We then calculate the average return of the stocks that belong to the top-10% best performing stocks and the average return of the bottom 10%. We subtract both average returns to get the return dispersion in the market. We expect that in markets with low return dispersion there are relatively few profitable pairs-trading opportunities, since pairs-trading opportunities arise

when there is a difference in the relative pricing of two stocks. We use the month-to-month change in the dispersion as variable in the regression. If the dispersion increases, we expect pairs-trading opportunities to arise. However, this does not lead to positive pairs-trading returns, since an increasing spread of a pairs does create an opportunity, but initially leads to a loss. If the market dispersion decreases and prices get more aligned, we expect positive pairs-trading returns, since the spreads of the pairs we opened are now closing and hence, generate a profit. We thus expect a negative sign for the dispersion coefficient. The second variable we add is *Volatility. Volatility* is the monthly change in the 20-day volatility of the market index. As we saw earlier, there appeared to be a correlation between the stock market volatility and pairs-trading returns. Furthermore we believe that in times of high stock market volatility investors show more irrational behavior than in periods with low stock market volatility. This irrational behavior creates pairs-trading opportunities, since there is a higher probability that prices are misaligned. Again we use the change in volatility since we believe that increasing market volatility creates opportunities and decreasing volatility leads to pairs-trading profits. When investors start acting rational again, the misaligned prices will be aligned again. The third variable we add is *SSD*. *SSD* is the sum of squared deviations between the normalized price series of a pair. The *SSD* is the average SSD of the pairs in the portfolio during the formation period. Since every month has six overlapping portfolios, the SSD is calculated is the average of those six portfolios. We include *SSD* since it is the metric on which we base our pair formation. By including *SSD* we analyze if there is a positive or negative correlation between pairs-trading returns and the relative co-movement of prices. It is not obvious what the relationship is beforehand. On the one hand, pairs with low SSD's are pairs which prices move closely together and thus are the best pairs in our model. On the other hand, prices that move together but never divergence and converge again do not generate returns. So like DF concluded, the best pairs are probably those who move quite closely together, but also diverge and converge from time to time. *SSD* is calculated as the average of the six overlapping SSD's during the formation period.

Panel II of Table 7 shows the result from the following regression:

$$R_p = \alpha + \beta_m R_m + \beta_{smb} R_{smb} + \beta_{hml} R_{hml} + \beta_{mom} R_{mom} + \beta_{rev} R_{rev} + \beta_d C_d + \beta_i C_i + \beta_{ssd} SSD \ (7)$$

C_d is the change in stock market dispersion, C_i is the change in the 20-day stock market volatility and SSD is the average sum of squared deviation. The added variables are not the returns of a portfolio. This means that the constant of the regression is not an excess return. The three additional variables do have some additional explanatory value in the U.S. The

adjusted- R^2 of the new regression is now 52%, compared to 47% of the base case regression. *Dispersion* has indeed a negative sign and is significant. When the market dispersion decreases, pairs-trading returns increase. *Volatility* is positive and insignificant. There is a positive and significant relationship between *SSD* and pairs-trading returns. Apparently the pairs-trading returns increase as the *SSD* of the pairs increase. This relationship breaks down if the number of pairs increases, since the all-portfolio –of which the SSD's are higher- have a lower average return than the top-20 portfolio. In Japan the adjusted- R^2 of the new regression increased to 33% from 26%. Furthermore we see that *Dispersion*, *Volatility* and *SSD* have a positive sign, but only *Volatility* is significant at the 5% level. So in Japan increasing stock market volatility leads to higher pairs-trading returns. In Hong Kong *Volatility* and *Dispersion* have the expected signs. *SSD* has a positive sign, but all added coefficients are insignificant. This adjusted- R^2 of the regression is somewhat higher; 33%. *Dispersion, Volatility* and SSD are all insignificant in China. The adjusted- R^2 decreased from 15% to 13%.

In table 11 we provide the correlation matrix of all variables. Panel A shows the correlation matrix for the U.S. As can be seen from the matrix, the correlation between the different factors is relatively low. *Market* and *Momentum* have the highest correlation: -0.47. Remarkable are the low correlations between *Dispersion*, *Volatility* and *SSD*. The low correlation between *SSD* and *Volatility* and *Dispersion,* might be explained by the fact that the *SSD* is calculated over the preceding 12 months and *Volatility* and *Dispersion* over the current month. We would expect a higher correlation between *Volatility* and *Dispersion*, since we expect more dispersion in volatile markets. The highest correlation in Japan, which is shown in panel B, is -0.74 between *HML* and *Momentum*. Again we see low correlations between *DIS, VOL* and *SSD*. In panel C, we find high correlations between *SMB* and *HML* and between *MOM* and *SMB* in Hong Kong. In China (panel D), -0.77 is the highest correlation and is between *MOM* and *HML*. Looking at the correlation matrices for all four countries, we find relatively low correlation between all factors. Thus, it is rational to use all these variables, since none of the variables is extremely correlated with one of the others.

All in all we see that there are differences between the adjusted- R^2 of the regression models in the four markets. We do see that the proxies for market inefficiencies do add some explanatory power to the model, but the increase is modest. In all markets, there is still a great percentage of the pairs-trading returns that cannot be explained by the risk factors we used. Interestingly enough *SSD* is positive and significant in the U.S. and China. This means that the best pairs performing pairs are not the pairs with the lowest SSD, but pairs with a

somewhat higher SSD. Furthermore we see that *Dispersion* and *Volatility* have some explanatory value in Japan and the U.S., but do not in Hong Kong and China.

Correlation Matrix
Sample Period: *January 2004-December 2009*

A.US	Mkt	SMB	HML	MOM	REV	DIS	VOL	SSD
Mkt								
SMB	0.35							
HML	0.40	0.21						
MOM	-0.47	-0.15	-0.41					
REV	0.37	0.24	0.20	-0.11				
DIS	-0.01	0.04	0.27	-0.19	-0.12			
VOL	-0.06	0.18	0.24	-0.06	0.06	0.27		
SSD	0.15	0.14	0.04	-0.34	0.02	-0.05	0.01	

B. Japan	Mkt	SMB	HML	MOM	REV	DIS	VOL	SSD
Mkt								
SMB	-0.08							
HML	0.18	-0.54						
MOM	-0.09	0.52	-0.74					
REV	-0.23	0.23	-0.18	0.13				
DIS	0.04	0.30	-0.09	0.07	0.22			
VOL	-0.02	0.20	-0.02	0.10	0.04	0.08		
SSD	-0.14	-0.34	0.30	-0.34	-0.03	-0.09	-0.18	

C. Hong Kong	Mkt	SMB	HML	MOM	REV	DIS	VOL	SSD
Mkt								
SMB	-0.27							
HML	0.22	-0.74						
MOM	-0.09	0.74	-0.69					
REV	-0.30	0.59	-0.48	0.47				
DIS	0.15	-0.16	0.17	0.02	-0.05			
VOL	-0.04	0.06	0.04	-0.03	0.00	0.12		
SSD	-0.29	0.03	-0.01	0.09	0.05	-0.11	-0.04	

D. China	Mkt	SMB	HML	MOM	REV	DIS	VOL	SSD
Mkt								
SMB	-0.37							
HML	0.15	-0.71						
MOM	0.03	0.53	-0.77					
REV	-0.10	0.06	0.08	-0.04				
DIS	0.30	-0.23	0.21	0.03	0.12			
VOL	0.10	-0.04	0.07	0.06	0.11	0.27		
SSD	-0.19	0.29	-0.05	0.02	-0.19	-0.05	-0.11	

Table 11: Correlation matrix of the factors Market(Mkt), Size(SML), Value(HML), Momentum(MOM), Reversal(REV), Dispersion(DIS), Volatility(VOL) and SSD. Panel A shows the U.S., panel B Japan, panel C Hong Kong and panel D China.

6. Timing Pairs-Trading

In the sub-period analysis we observed that pairs-trading did not generate consistent positive returns throughout the sample period. In this section we analyze if we can increase pairs-trading performance by timing when to invest in pairs-trading. In the previous section we argued that market inefficiency creates pairs-trading opportunities. It would therefore be interesting to test, if using the variables that are a proxy for market inefficiency as timing indicators improve performance. To analyze the relationship between those indicators and pairs-trading returns, we first conduct a regression. In this regression we regress the variables *Dispersion*, *Volatility* and *SSD* on the monthly pairs-trading returns. However, *Dispersion* and *Volatility* are regressed with a one-month lag. *SSD* is already lagged, since it is the SSD over the formation period. In this way we can use those variables as indicators for our pairs-trading strategy.

Regression Analysis Sample Period: *January 2004-December 2009*								
	A.U.S.		B. Japan		C. Hong Kong		D .China	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Lagging Factor Model: *Dispersion (t-1), Volatility (t-1), SSD*								
Intercept	-0.0220	0.00	-0.0126	0.21	-0.0008	0.94	-0.0056	0.52
Dispersion (t-1)	0.0240	0.02	0.0024	0.85	-0.0026	0.60	0.0008	0.96
Volatility (t-1)	-0.0009	0.87	-0.0134	0.01	0.0057	0.77	0.0020	0.85
SSD	0.1139	0.00	0.0652	0.05	0.0104	0.45	0.0015	0.20
Adj R sq.	24%		10%		-4%		-3%	

Table 12: Regression analysis of the three market inefficiency proxies, for the U.S. (A), Japan (B), Hong Kong (C) and China (D).

As can be seen from table 12, results are mixed. Looking at the U.S. in panel A, we see that the adjusted- R^2 of the regression is 24%. This is only 10% for Japan (panel B) and even negative in Hong Kong and China (panel C and D). In the U.S. we see that if stock market dispersion in the previous month increased, pairs-trading returns in this month will be higher. The same holds for *SSD*. *Volatility* is negative and insignificant. In Japan we find significant coefficients for *Volatility* and *SSD*. If the stock market volatility dropped in the previous month, pairs-trading returns in this month increase. A higher SSD during the formation period leads to higher returns. In Hong Kong and China, all coefficients are insignificant; this is also reflected by the adjusted- R^2, which is negative. Looking at the results, there is no factor that shows a consistent relationship with pairs-trading returns. In the U.S. and Japan, the *SSD* beta is positive and significant, but insignificant in Hong Kong and China. Furthermore, we see that the *Volatility* beta is negative in the U.S. and Japan,

but only significant in Japan. The *Volatility* beta is positive in Hong Kong and China, albeit insignificant. However, since we test the *consistency* of pairs-trading, we do include the same timing indicators in all four markets.

Looking at the regression we find a positive, significant relationship between pairs-returns and *Dispersion* in the U.S. So we give the *Dispersion* indicator the value '1' if the month-to-month stock market dispersion change is positive and '0' otherwise. There is a negative, significant relationship between *Volatility* and returns in Japan. The *Volatility* indicator is assigned the value '1' if the month-to-month change in volatility is negative and '0' otherwise. In all four markets there is a positive relation between *SSD* and returns. If the *SSD* value is above the sample average we set the indicator to '1' and '0' otherwise. We realize that we do not know the average SSD ex-ante, but if we would implement pairs-trading in the future, we could use the historical average. We set up two different timing strategies. Based on these indicators we now decide whether to invest in the pairs-trading strategy or not.

We consider two different strategies. Timing strategy 1 (TS1) only invests in pairs-trading if two or more indicators signal '1'. Timing strategy 2 (TS2) invests if one or more indicators signal '1'. The results of this strategy are shown in table 13.I, panel A shows the results for TS1 and panel B for TS2. If we first look at TS1, we see that in the U.S. the standard strategy yields an average excess return of 0.34%, with a standard deviation of 1.81%.[28] The timing strategy yields an average excess return of 0.31%, with a standard deviation of 1.44%. The Sharpe ratio of the timing strategy is a bit higher; 0.21 versus 0.19 for the standard strategy. In Japan the timing strategy yields a return of 0.78%, compared to 0.75% for the standard strategy. Again the standard deviation of the timing strategy is lower, leading to a higher Sharpe ratio. In Hong Kong, the average return of the timing strategy is 0.20%, 0.24% lower than the standard strategy. The Sharpe ratio of the timing strategy is 0.09, compared to 0.11 for the standard strategy. In China we find almost equal average returns for the standard and timing industry: 0.03% and 0.02%, respectively. The standard deviation of the standard strategy is higher: 3.44% versus 1.22% for the timing strategy. If we now look at TS2 (panel II), we see that the results in the U.S. do not change much. Both the average return and standard deviation are a bit higher compared to TS1, but the Sharpe ratio of the timing strategy is still higher than that of the standard strategy. In Japan we find a lower return for TS2, compared to the standard strategy. However, the standard deviations are almost equal, leading to a lower Sharpe ratio for TS2. In Hong Kong,

---

[28] Note that these results are somewhat different form the result we reported earlier. This is because we now use the monthly returns, whereas the previous results are based on daily returns.

the average return of TS2 is higher than the return of the standard strategy and the for the old strategy. In China TS2 yields a negative excess return, of course leading to a negative standard deviation.

Timing Strategy Sample Period: *January 2004-December 2009*								
I. Timing Strategy 1	A.U.S.		B. Japan		C. Hong Kong		D .China	
	Standard	Timing	Standard	Timing	Standard	Timing	Standard	Timing
Average Return	0.34%	0.31%	0.78%	0.75%	0.44%	0.20%	0.03%	0.02%
Standard Deviation	1.81%	1.44%	1.86%	1.58%	3.92%	2.24%	3.44%	1.22%
Sharpe Ratio	0.19	0.21	0.42	0.47	0.11	0.09	0.01	0.01
Percentage of the time invested	100%	43%	100%	60%	100%	39%	100%	39%
II. Timing Strategy 2	A.U.S.		B. Japan		C. Hong Kong		D .China	
	Standard	Timing	Standard	Timing	Standard	Timing	Standard	Timing
Average Return	0.34%	0.37%	0.78%	0.73%	0.44%	0.66%	0.03%	-0.10%
Standard Deviation	1.81%	1.76%	1.86%	1.84%	3.92%	3.56%	3.44%	2.93%
Sharpe Ratio	0.19	0.21	0.42	0.40	0.11	0.19	0.01	-0.03
Percentage of the time invested	100%	83%	100%	89%	100%	83%	100%	83%

Table 13: Monthly Return Characteristics of Timing Strategy 1 (I) and Timing Strategy 2 (II) in all four markets. Standard is the standard pairs-trading strategy, Timing is the pairs-trading with timing indicators

Looking at the overall picture, there seems to be no consistent improvement of pairs-trading performance when using a timing indicator. For the U.S. we find that for both timing strategies the Sharpe ratio improved from 0.19 to 0.21. For Japan, TS1 has a higher Sharpe ratio, but TS2 a lower Sharpe ratio. For Hong Kong, TS1 has a lower Sharpe ratio, TS2 a higher Sharpe ratio. TS1 has a higher Sharpe ratio in China, but the Sharpe ratio of TS2 is lower. All in all, the Sharpe ratios do not change that much. However, because we no use a timing instrument to indicate whether to invest or not, we are not invested all the time. The percentage of the time we are invested under the timing strategies are shown in row 4 of panel I and II. If we assume that hedge fund are flexible in their capital allocation, they can invest in different investment strategies when they do not invest in pairs-trading. Combining pairs-trading with a different investment strategy would be an interesting topic for further research.

7. Adjusted Pairs-Trading Algorithm

Pairs-trading benefits from diversification. The standard deviation of the unrestricted strategy is lower than the standard deviation of the restricted strategy. However, the returns of pairs-trading in some sector of the restricted strategy are higher than those of the unrestricted strategy. In this section we test a pairs-trading strategy which combines these two characteristics. First, we match stocks that are in the same sector like we did with the restricted strategy. Second, we select the 20 pairs with the lowest SSD of all pairs. This means that in the top pairs we combine pairs of different sectors, but all pairs consist of stocks that are in the same sector. In this way we try to maximize the trade-off between diversification benefits and average returns. Table 14 shows the return characteristics of this new and old strategy in all four markets.

Return Characteristics								
Sample Period: *January 2004-December 2009*								
	A. U.S.		B. Japan		C. Hong Kong		D. China	
Pairs Strategy	New	Old	New	Old	New	Old	New	Old
Average Excess Return *(fully invested)*	0.46%	0.32%	0.90%	0.76%	0.17%	0.43%	0.23%	0.03%
t-Statistic	2.97	1.99	4.25	3.66	0.35	0.98	0.46	0.07
Standard Deviation	1.36%	1.41%	1.80%	1.79%	4.15%	3.74%	4.32%	3.85%
Sharpe Ratio	0.34	0.23	0.50	0.43	0.04	0.11	0.05	0.01
Skewness	0.21	0.42	0.27	0.25	0.16	0.08	0.30	0.12
Kurtosis	1.95	2.27	1.51	1.49	2.06	2.09	2.82	1.85
Minimum	-2.43%	-2.73%	-4.12%	-3.79%	-8.36%	-10.06%	-9.04%	-9.05%
Maximum	5.14%	8.15%	6.98%	6.49%	10.56%	12.62%	14.41%	10.37%
Negative Returns	37%	50%	35%	31%	48%	43%	55%	47%

Table 14: Return Characteristics of the new and old pairs-trading strategy in the U.S. (A), Japan (B), Hong Kong (C) and China (D).

To see if the new strategy indeed combines diversification benefits with higher returns, we first look at the returns. The returns of the new strategy are 0.14%, 0.14% and 0.20% higher than the old strategy in the U.S., Japan and China, respectively. Surprisingly enough, the new strategy generates a 0.26% lower return in Hong Kong. Again, as we saw earlier with the unrestricted strategy, average pairs-trading returns in Hong Kong and China are not significant. Looking at the average returns of the different sector of the restricted strategy we see that Financials, Industrials and Technology have higher average returns than our new strategy with returns of 0.93%, 0.73% and 0.49%, respectively. The same holds for Japan, where we see that 4 out of 8 sectors outperform the new strategy, based on average return. The average return of the Financials in Hong Kong is 0.67%, which is higher than the 0.17% of the new strategy. In China only the Cyclical Consumers outperform the new strategy, but as mentioned before are the returns in China insignificant.

The diversification benefits are measured by the standard deviation. The standard deviations of the new strategy is Japan and the U.S. are comparable with the standard deviation of the old strategy. The standard deviation in the U.S. is now 1.36% compared to the 1.41% for the old strategy. The new strategy in Japan has a standard deviation of 1.80%, compared to 1.79% of the old strategy. The standard deviations of China and Hong Kong are higher under the new strategy: 4.32% and 4.15%, compared to 3.85% and 3.74% for the old strategy. At first sight this indicates that the new strategy does not have the diversification benefits we hoped for. But if we now look at the restricted strategy we do see an improvement. In the U.S., Consumer Cyclical is the sector with the lowest standard deviation:1.64%, this is still higher than the 1.36% standard deviation of the new strategy. In Japan, Consumer Cyclicals are again the sector with the lowest standard deviation: 2.29% and again this standard deviation is higher than the standard deviation of the new strategy. The diversification gains are even higher in Hong Kong and China: Financials in Hong Kong has a standard deviation of 5.70%, compared to the 4.15% of the new strategy. The lowest standard deviation in China is 7.61%, which is much larger than the 4.32% of the new strategy.

The Sharpe ratio is a performance measure based on both the return and standard deviation, therefore the Sharpe ratio is the ideal measure to see which of the strategies performs best. The new strategy has a higher Sharpe ratio than the unrestricted strategy in all four markets, except for Hong Kong. The result is even more striking when looking at the restricted strategy. The Sharpe ratios of the U.S., Japan and China of the new strategy are all higher than the Sharpe ratios of the restricted strategy. Again, Hong Kong is the exception, together with U.S. Industrials which has a Sharpe ratio of 0.35, compared to the 0.34 of the new strategy. From this results we can conclude that the new strategy indeed combines the diversification benefits from the unrestricted strategy with the higher returns of the restricted strategy, leading to a higher Sharpe ratio in almost all cases.

These results are confirmed by the cumulative return series of the new and old strategy, which can be seen in figure 6. But despite the fact that the new strategy performs better than the other strategies, we also want to see if the new strategy is an attractive investment strategy. Panel A clearly shows the higher returns of the new strategy in the U.S. In the first three years, the new strategy generates a cumulative excess return of about 10%. Although being significantly higher than the returns of the old strategy, this is still relatively modest, especially if one considers that the returns are calculated without subtracting transaction costs. Like the old strategy, most of the positive returns are generated in the last year of the sample. Noteworthy is also the fact that the spread between the new and old strategy returns appears to be relatively stable. In Japan the difference is much smaller. Up until

2009 the returns of the new and old strategy move closely together, with the new strategy outperforming the old strategy most of the time. The difference is made in the last year of the sample, where the new strategy clearly outperforms the old strategy. This is the opposite of what we see in Hong Kong. Here the old strategy slightly outperforms the new strategy up until 2009 and in the last year of the sample the difference gets bigger in the advantage of the old strategy. In China the new strategy performs visibly better than the old strategy. Especially in the first year or so, the new pairs-trading strategy is proven to be really profitable with an cumulative excess return of about 20%. In the period January 2005-October 2007 the returns fall, reaching the bottom at about -35%, which is even lower than the lowest point of the old strategy. From October 2007 onwards the return from pairs-trading are starting to pick-up quickly, reaching a peak somewhere in October 2009 of 35%, compared to a peak of only 15% for the old strategy.
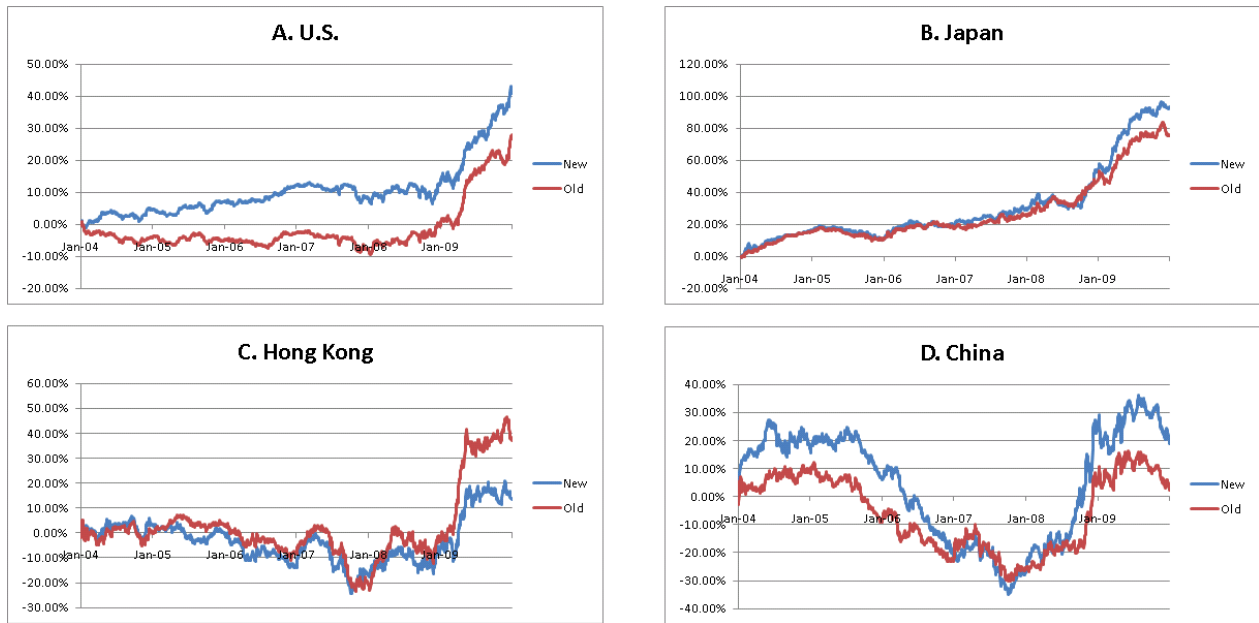

Figure 6: Cumulative Return of both the new and old strategy in the U.S. (A), Japan (B), Hong Kong (C) and China (D)

Clearly Japan is by far the most profitable pairs-trading market, with a cumulative excess return of almost 100% over the six year in our sample. Although 60% of the returns are generated in the last 12 to 18 months, pairs-trading still generates an cumulative excess return of 40% in the period before, which makes pairs-trading still an attractive investment strategy in Japan. The returns in the U.S. are relatively modest. In the first 5 years of our sample, pairs-trading in the U.S. only yields an average excess return of about 2%. Pairs-trading performs good in 2009, with an excess return of about 30%, which is quite

impressive. But looking at the full sample period, pairs-trading is not really a desirable investment strategy in the U.S., since almost all returns are only generated in the last year of the financial crisis, when we saw a global (financial) crisis. The picture is even more depressing in Hong Kong. Up until 2009, pairs-trading would have yielded a slightly negative return and even with the recent financial crisis the cumulative return in Hong Kong does not pass the 20% mark. Additionally the picture shows that the returns from pairs-trading in Hong Kong are much more volatile, making the strategy even more unattractive. Possibly the most fascinating picture is shown by China. In the first years pairs-trading in Japan shows high returns, even higher than Japan. However, this period of flourishing pairs-trading returns is followed by a period of negative returns. The last years of the sample are booming again, with cumulative returns increasing from -30% to 30%. As one might expect from such a pattern, pairs-trading returns in China are relatively volatile compared to Japan and the U.S.

8. Conclusion

This article examines pairs-trading in the U.S., Japan, Hong Kong and China. Using the same pairs-trading algorithm as Gatev, Goetzmann and Rouwenhorst (2006), we confirm the declining trend in pairs-trading profitability in the U.S. during the January 2004-December 2009 period, observed by Do and Faff (2010). We show that pairs-trading in Japan yields a monthly average excess return of about 0.76% for the top-20 pairs. Pairs-trading in Hong Kong and China yields insignificant positive average returns. There is a difference in pairs-trading characteristics between the U.S. and Japan and Hong Kong and China. Most notably is the higher volatility of pairs-trading returns in Hong Kong and China. Continuing on the conclusion of Do and Faff that pairs-trading is especially profitable during market turmoil's, we found that in all four countries pairs-trading has been extremely profitable during the most recent global financial crisis. In fact, pairs-trading in the U.S., Hong Kong and China would not have yielded positive returns without the recent crisis. Pairs-traders exploit statistical arbitrage opportunities that arise when there is a discrepancy in the relative pricing of two stocks. This discrepancy is caused by irrational behavior of certain investors. The proxies we used for market inefficiency: price dispersion in the stock market, SSD and the 20-day volatility of the stock market, do explain some of the pairs-trading profits. We showed that using these proxies as a timing instrument do sometimes increase performance. However, this increase in performance is not consistent for every market and the increase is relatively modest. Comparing the results of the unrestricted pairs-trading, which combines stocks of different sectors into a pair, with the restricted strategy, which only combines stocks that are in the same sector into a pair, we noticed some apparent differences. The unrestricted strategy profits from diversification benefits, that arise because stocks from different sectors are used. The standard deviation of the unrestricted strategy is much lower. The restricted strategy profits from the increased homogeneity of the pairs, leading to higher returns than the unrestricted strategy. In an effort to combine these two effects we created a new pairs-trading algorithm. In this strategy we only use pairs that consist of stocks that are in the same industry, but we combine these pairs into a portfolio consisting of pairs from different sectors. We select the 20 pairs that have the lowest SSD. Again, the results are not consistent. In the U.S., Japan and China we find an superior Sharpe ratio for the new strategy, but the Sharpe ratio of the new strategy is lower in Hong Kong. The increase in performance is also modest.

Our findings suggest that the pairs-trading algorithm designed by Gatev, Goetzmann and Rouwenhorst, does not yield consistent positive excess returns. Neither throughout time, nor

in different markets. Using a timing strategy and adjusted algorithm did not increase performance consistently. We did not find factors that explain or predict pairs-trading performance in all four markets. It is therefore hard to construct an algorithm that yields positive excess return in all markets. Further research on the factors that drive pairs-trading returns in different markets is needed to construct an algorithm that yields consistent positive returns. It would also be interesting to test if combining different investment strategies yield better returns that the strategies do separately.

9. References

Andrade, S., Di Pietro, V. and Seasholes, S., 2005, "Understanding the Profitability of Pairs Trading", *working paper*, University of California, Berkeley, and Northwestern University (February)

Do, B., Faff, R., 2010, "Does Simple Pairs Trading Still Work?", *Financial Analyst Journal*, vol. 66, no. 4 (July/August): 83-95

Engelberg, J., Goa, P. and Jagannathan, R., 2009, "An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity", *working paper*, University of North Carolina at Chapel Hill, University of Notre Dame, and Northwestern University (February)

Gatev, E., Goetzmann W. and Rouwenhorst K., 2006, "Pairs Trading: Performance of a Relative Value Arbitrage Rule", *Review of Financial Studies*, vol. 19, no 3. (Autumn):797-827

Jegadeesh, N., 1990, "Evidence of Predictable Behavior of Security Returns", *Journal of Finance*, vol. 45, no.3 (July): 881-898

Jegadeesh, N., and Titman, S., 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiencies", *Journal of Finance*, vol. 48, no. 1 (March): 631-655

Kondor, P., 2009, "Risk in Dynamic Arbitrage: The Price Effects of Convergence Trading", *Journal of Finance*, vol. 64, no. 2 (April): 631-655.

Lehmann, B., 1990, "Fads, Martingales, and Market Efficiency", *Quarterly Journal of Economics*, vol. 105, no. 1 (February): 1-28

Newey, W. and West, K., 1987, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, vol. 55, no.3 (May): 703-708