Statistical Arbitrage and Its Exposure to an Idiosyncratic Return Reversal Factor

Binh Do[†]

Bong-Soo Lee

October 2014

Preliminary draft – Please do not quote

Abstract:

Statistical arbitrage is a zero-cost trading strategy that offers positive expected payoffs while still allowing for negative outcomes. Amongst the hedge fund circle, it refers to a class of strategies that seek to exploit short-term mispricing between two or more assets. This paper seeks to explain the widely documented decline in the profitability of such strategies when applied to daily stock prices in the U.S. market over the period of 1980-2013. We find that the profit decline is due to increased divergence instead of reduced mispricing. Importantly, we find that statistical arbitrage returns are highly related to a systematic factor, which we call *idiosyncratic return reversal factor*. Whereas loadings on common risk factors are either statistically or economically insignificant, the loading on this newly constructed factor is highly statistically and economically significant. Variation in this factor subsumes the effect of intra-industry reversals, explains up to 60% of the decline in pairs trading returns, and increases R^2 of the risk-adjustment regression by up to three times. The findings are robust to various strategies including statistical arbitrage on triplets of stocks.

JEL Classification: G11, G12, G14.

Keywords: Convergence trading, statistical arbitrage, pairs trading, cointegration.

[†] Email: binh.do@monash.edu. Tel: +613 9903 1399.

1 Introduction

Arbitrage is a fundamental mechanism that helps to enforce the law of one price and maintain efficiency of capital markets. In the equity market, however, it is an elusive concept because either perfect substitution does not exist (like a currency traded on different exchanges) or there is no definite settlement date where cash flows from the arbitraged assets must converge (as the case of stock index arbitrage).[1]

Statistical arbitrage (SA) is a much looser notion of arbitrage but arguably more relevant for equity markets. SA is zero-cost strategies that promise positive expected payoffs by the virtue of the statistical behaviour of the pricing relations being exploited. Among the hedge fund community, the term *stat arb* encompasses trading strategies that seek to exploit short-term mispricing among securities while maintaining little exposure to market risks. This is achieved by identifying stocks that have moved "together" in the past according to some models, take long-short positions when they diverge from the historical trend and unwind when the prices re-converge. Gatev, Goetzmann and Rouwenhorst (2006) study the most basic form of SA known as pairs trading and find the strategy profitable and insensitive to market risk. A follow-up study by Do and Faff (2010) finds that the strategy performs strongly during market turbulence as relative mispricing appears to be more frequent. However, both studies observe a decline in pairs trading profitability since 1990. In their study of SA using factor analysis, Avellaneda and Lee (2010) also report declining profitability over two periods 1997-2002 and 2003-2007. According to Pole (2007), many practitioners declared SA dead in 2002, but then started hoping of a revival following a

[1] In the "negative stub values" arbitrage studied in Mitchell, Pulvino and Stafford (2002), the mispricing situation involves a company whose market value is less than its ownership stake in a publicly listed subsidiary. Whereas perfect substitution exists in this case, there is absence of a settlement date thus making the arbitrage risky (unless the parent company sells the stake and distribute to its shareholders on a given future date). In the case of merger arbitrage studied in Mitchell and Pulvino (2001), the spread between the offer price and the target's stock price represents the mispricing to be arbitraged. Whilst there is a settlement date, that date is by no means certain as deals may not proceed, again making the arbitrage risky.

resurgence in the performance of SA in 2007. Figure 1 (discussed in more detail later in the paper) plots the monthly returns of a pairs trading strategy and unfortunately confirms that the downward trend resumes and persist through December 2013.

This paper studies the fall of statistical arbitrage in the US equity market over period 1980-2013. Specifically, we examine what mechanism drives this fall.[2] Gatev et al (2006) suspect a latent systematic factor is responsible for the profit decline. We pick up where they left off, seeking to identify the systematic factor. An idiosyncratic return reversal factor is constructed daily that captures price behaviours post price shocks attributable to firms' fundamentals net of impact from market and industry returns. We find that this systematic factor is highly positive and significant for period 1980-1995, at 1.26% per month (t-statistic=15.63). The factor materially declines in period 1996-2013, averaging 0.54% per month (t-statistic=4.99). The decline over the two periods is statistically significant. Importantly, we find that this factor is highly related to monthly pairs trading returns. When employed to risk-adjust these returns, common factors are either statistically insignificant (loadings on Fama and French (1997) factors are insignificantly different from zeros) or economically insignificant (loading on momentum is highly statistically significant but is very small in magnitude, at -0.03). In contrast, the loading on the idiosyncratic reversal factor is both statistically significant (t-statistics=6.63) and economically significant (estimated coefficient = 0.24). The decline in the newly constructed factor is found to explain about 40% of the decline in the raw return of pairs trading over our sample period. R^2 of the risk factor regression increases from 0.09 to 0.25 when the new factor is augmented with the common risk factors.

[2] It should be pointed out that this study considers only price-based SA strategies that are implemented using daily closing prices. It is possible that using high frequency data and/or combining price with other potentially relevant information such as volume and fundamental variables may lead to different outcomes.

We extend the analysis beyond the traditional pairs trading strategy to investigate statistical arbitrage on baskets of stocks using cointegration. Cointegration is a natural framework to model statistical arbitrage, yet to date has received little empirical attention. We formulate an implementation approach that is general enough to accommodate any arbitrary size of baskets (only limited by computing power). Empirical evidence continues to point to a declining pattern in the profitability when the approach is implemented on pairs as well as triplets. The idiosyncratic reversal factor continues to be highly related to the returns from these strategies and explain up to 60% of the decline in the mean returns between the two periods. We also find that the new factor is unique from Hameed and Mian's (2014) intra-industry reversal factor. The former subsumes the effect of the latter in the risk-adjustment regression.

Our main contribution is to resolve the puzzle regarding the demise of statistical arbitrage. A by-product of this investigation is the construction of a new systematic factor, the idiosyncratic reversal factor, which can be readily used by equity market arbitrageurs in benchmarking their performance. Hedge fund type returns are notoriously difficult to risk adjust (see for example Fung and Hsieh, 2001). Existing risk factors do a poor job in explaining the variation in statistical arbitrage performance. The robustness of the idiosyncratic reversal factor in explaining different SA strategies means that it is a strong candidate to consider when assessing whether a certain SA strategy is generating alphas.

In the process of identifying this new factor, we are able to rule out a number of "suspicions" regarding what may cause the profit decline in SA. We find that returns conditional on convergence are not lower in the second period, if not higher. In addition, trades tend to take longer to converge than in the first period of the sample. Furthermore, the convergence rate declines significantly in the second period. These pieces of evidence suggest that reduced mispricing especially via increased competition is unlikely to be behind the demise of SA. We confirm Engelberg, Gao and Jaganathan's (2008) finding that unconditionally, pairs trading returns are positively related to illiquidity of the member stocks. Although we find illiquidity reduces in the second period of our sample, thus confirming findings from studies such as Cheng, Hameed, Subrahmanyam and Titman (2014) who find that liquidity improves in recent years and explain away contrarian returns, we cannot find a link between this reduced illiquidity and the profit decline. In fact, returns on converged trades are even higher in the second period despite the stocks underlying those trades experience improved liquidity. We also rule out changes in the information environment as a driver of the profit decline in SA. Using public news augmented with quarterly earnings announcements, we find that the proportion of trades that open following a price sensitive news event (which according to Engelberg et al, 2008 and confirmed in our data, tends to diverge subsequently), does not vary between the two periods. Continuous disclosure does not seem to affect the strategies being examined.

Our study is related to, and provides empirical evidence in support of, the body of theoretical literature on arbitrage risks. Shleifer and Vishny (1997) argue that fundamental risk is a key risk facing equity-market arbitrageurs. It can be seen that our idiosyncratic reversal factor is an empirical attempt to capture this fundamental risk. Strong price drift (or negative reversal) following idiosyncratic shocks) is evidence of fundamental information content in the price action, which negatively affects SA. Abreu and Brunnermeier (2002) develop a model where arbitrageurs face a synchronization risk, or the uncertainty that other arbitrageurs may not simultaneously trade on the mispricing opportunity, leading to delayed arbitrage. Xiong (2001) shows that when convergence traders, which is essentially statistical arbitrageurs, suffer losses, their liquidation will amplify the original shock. We see these externalities in our results as divergence risk appears to increase over the sample period. Stein (2009) shows that theoretically, increased participation of sophisticated investors such as hedge funds may not necessarily make the market more efficient because such investors are faced with a "crowding effect" (similar to Abreu and Brunnermeier, 2002) and a "leverage effect" (similar to Xiong, 2001). While our study does not directly investigate the state of market efficiency over time, our evidence seems to rule out increased competition as the source of the profit decline in SA.

Section 2 documents the declining profitability in a simple pairs trading strategy and analyzes trading statistics to help uncover a systematic factor. Section 3 reports main empirical results. Section 4 perform robustness tests, in particular, extending from Gatev et al's (2006) distanced based pairs trading to cointegration based statistical arbitrage among baskets of stocks. Section 5 concludes.

2 The fall of statistical arbitrage and possible explanations

Discovered in the early 1980s, statistical arbitrage was executed in the form of pairs trading where stocks were bought and sold in pair combinations (see Pole, 2007 for an account who may have been the creator of the idea). Pairs trading is simple and intuitive: find pairs of stocks that exhibit similar historical price behaviours and take positions when their relative prices depart from the historical norm. However, underlying this simple idea is a battery of techniques and decision rules including how to define "similarity", over what horizon, what is the trigger point, how long to hold the position for, what are risk management rules etc. In addition, when pairs trading is extended to multiple assets, as statistical arbitrage takes the form of baskets trading, the level of complexity has gone up another level. Whereas industry publications suggest sophisticated techniques such as cointegration and factor analysis

(Vidyamurthy, 2004 and Pole, 2007), and in the latter, applied to baskets trading, academic studies such as Gatev et al (2006) have focused on simple, non-parametric models of pairs trading. A recent exception is Avellaneda and Lee (2010) who test statistical arbitrage using the principal component analysis (PCA) approach. Since our interest is to understand what is causing the decline in statistical arbitrage's profitability, we start off with the simple strategy in Gatev et al (2006) and then consider other methods as a robustness check. In the following section, we document the profit decline associated with Gatev et al (2006) strategy and formulate possible explanations for this decline.

2.1 The distance-based pairs trading strategy

First tested in Gatev et al (2006) and further investigated in Engelberg, Gao and Jaggannathan (2008) and Do and Faff (2010), the distance-based strategy identify pairs by minimizing the sum of the squared differences (*SSD*) of the normalized prices:

$$SSD_{i,j} \equiv min_j \sum_{t=1}^{B} (p_{i,t} - p_{j,t})^2 \quad (1)$$

where p is the total return index scaled to start at \$1 at the beginning of the "training period". In other words, the partner to each stock i is one whose Euclidean distance with stock i is the smallest amongst all possible pair combinations. The resulting pairs are subjected to a trading rule that takes a dollar neutral long-short position in a pair whenever its spread $s_t=p_{1t} - p_{2t}$ exceeds or falls below a pre-specified threshold. A traded pair is closed when the spread crosses zero, or the end of the trading period is reached, whichever occurs first. A converged pair can be re-opened if the trading trigger occurs again for the remaining time. This trading rule is applied to a pre-specified set of pairs to generate a diversified portfolio of pairs. Key implementation variables are the length of the training period, the length of the trading period and the threshold to trigger. We respect the convention in the literature, using 1-year training periods, 6-month trading periods and a 2-standard deviation threshold. One may impose intuitive risk management constraints such as requiring the spread to start narrowing *x* days following the trigger, or unwinding the trade *y* days into the trade irrespective of the convergence status (Engelberg et al, 2008). Whilst those constraints may optimize profits, they may add noise to the analysis of the profit trend; as the result, we opt to ignore this risk management aspect and stick with the basic structure.

Data is drawn from the daily CRSP files over period January 1979-December 2010. Following asset pricing studies that use U.S. data (for example, Jegadeesh and Titman, 1993) only stocks with share codes 10 and 11 are included. For each implementation cycle, stocks with prices less than $5 or that do not trade in any day over the formation period are excluded. For delisted stocks, their delisting returns are incorporated in return calculation. To minimize unnecessary computation costs, we match pairs within industry groups that are based on the SIC code. We also restrict the candidate pool to the top 100 pairs with the lowest SSD statistics. Our findings about the profit trend as well as the explanations do not change when we match pairs using Fama and French (1997) 48 industry groups, or when all pairs are admitted to the trading pool.

Pairs matching and trading are implemented on a monthly basis. For the first implementation cycle, the training period starts in January 1979 and end in December 1979, with the trading period starting in January 1980 and end in June 1980. For the second cycle, the training period starts in February 1979 and end in January 1980, with the trading period starting in February 1980 and end in July 1980, and so on. Over the sample period, there are 367 overlapping implementation cycles, each producing six monthly returns that are constructed from daily marked-to-market payoffs to each traded pair. These returns are then aggregated across the overlapping cycles in the same manner as in Jegadeesh and Titman (1993) to give rise to a time series of 372 monthly returns, based on which we perform a trend analysis.

Finally, actual performance needs to take into account practical matters such as trading costs (brokerage fees and market impact – see Do and Faff, 2012) as well as the fact that positions may have to be established at a price point that is worse than one that triggers the trading (the standard treatment is to skip 1 day). Since these matters may be time varying, adding them to the simulation would only bring noise to the analysis and would not help to decipher what is driving the profit pattern.

2.2 The profit decline in pairs trading

Figure 1 plots the monthly portfolio returns from January 1980 to December 2010. In the first half of the sample, the returns are remarkably consistent positive, as would be expected from an arbitrage-type strategy. The strategy is resilient to the 1987 market crash, if anything, the returns jumped throughout this episode. In contrast, the second half experiences more negative return months. Although there are occasions of strong performing months especially during the 2001 dotcom bust and 2007-2008 financial crisis, they are either preceded or followed by negative returns. Overall, there is visually a declining trend as captured by the 12-month lag moving average.

[Insert Figure 1 about here]

Gatev et al (2006) and Do and Faff (2010) observe profit declines as early as 1989 whilst Avelleneda and Lee (2010) document the decline since 2003, which is also echoed in Pole (2007). There is no precise way to anchor the turning point, but guided by the visual depiction in Figure 1, we simply split our sample at the end of 1995 and conduct statistical testing to compare the performance of the two sub-periods.

Table 1 reports key statistics of the return distribution over the two sub-periods, together with Sharpe ratio and statistics associated with Fama-French's (1997) three-factor model and a five-factor model that also includes a momentum factor and a short-term reversal factor. The raw excess return more than halves in the second period, from 0.83% per month to 0.39% per month, and this decline is statistically significant at 1% level. With the standard deviation remaining largely similar for both periods, Sharpe ratio drops more than a half, from 1.26 to 0.54. A similar pattern is observed for the risk-adjusted returns.[3] The fact the beta is zero under both risk models confirms the market neutral nature of the strategy. The fact that the alpha is very similar to the raw returns suggests that the standard factor models may not be a suitable risk model for this type of arbitrage strategy.

3 Possible mechanisms for the profit decline

We now turn our attention to addressing the main question: what is behind this profit decline? Gatev et al (2006) speculate that there is an unidentified systematic factor that influences the profitability of pairs trading over time and that its dormancy over the recent years is behind the reduced profitability. We aim to identify this factor in this paper. This section contains a diagnostic to identify the mechanism that underlines the profit decline. The exercise will help to inform the identification and construction of the risk factor. To help with the formulation of conjectures, we use a simple statistical model to capture the return dynamic of the strategy. Instead of modelling the monthly returns of the pairs portfolio, we choose to model the return at the individual trade level. The advantage of the latter is that it strips out the impact of frequent changes in the composition of the portfolio. It also allows one to relate trading performance to the trading signal as well as convergence of the trade, which is a critical factor determining profitability.

[3] Skipping 1 day from the signal, Gatev et al (2006) find that the excess return on the top 20 pairs is 1.18% over period 1963-1988, dropping to 0.38% over period 1989-2002. Without skipping, Do and Faff (2010) find that the excess return on the top 20 pairs drops from 1.24% over period 1963-1988 to 0.56% over period 1989-2002, to 0.33% over period 2003-2009.

Since pairs trading is to exploit perceived mispricing amongst securities, profit at the individual pair level is made up of the actual mispricing plus a noise:

$$y_i = \mu_i + \epsilon_i \tag{2}$$

if convergence occurs: $\mu_i \approx h_i, \epsilon_i = 0$

if divergence occurs: $\mu_i = 0, \epsilon_i = e_i < 0$

where h_i is the spread threshold that triggers the opening of the position, the approximation sign reflects the fact that converged pairs are unwound just after the spread crosses zero, and the noise reflects the possibility that the trading signal is false alarm, hence this arbitrage strategy is risky arbitrage. Note that the time subscript is suppressed here for simplicity. Convergence occurs when the spread crosses zero within the trading period and divergence occurs if it does not.[4]. Let I be the indicator variable that equals 1 if there is convergence and 0 otherwise. (2) then becomes:

$$y_i = h_i I + e_i(1 - I) \tag{3}$$

The expected return of each pair trade is given by:

$$E(y_i) = E(h_i I) + E[e_i(1 - I)] = E(h_i | I = 1)xE(I) + E(e_i)\big(1 - E(I)\big) =$$

$$E(h_i | I = 1)x(Prob\ of\ convergence) + E(e_i)\big(1 - (Prob\ of\ convergence)\big)$$

(4)

Thus the expected pair return is the probability weighted average of returns on converged and non-converged pairs, which is expected. One can also see that pairs trading profits decrease

[4] Our divergence status also includes cases where the spread narrows but does not cross zero, in which case the profit is a small positive. The proportion of these cases is about 16-17% and does not change significantly over the 2 sub-periods in question, therefore we do not explicitly distinguish between them and trades that diverge with a negative return.

in the average magnitude of mispricing, decrease in the average magnitude of noise and decrease in the probability of convergence (or decrease in the probability of divergence). Two conjectures exist regarding the mechanism through which the profit decline occurs.

Conjecture 1: The decline in pairs trading profits is driven by reduced mispricing (as captured by an increase in $E(h_i|I=1)$.

Conjecture 2: The decline in pairs trading profits is driven by increased divergence, as captured by a fall in either the probability of convergence, or a fall in (e_i) .[5]

Economic drivers that underlie conjecture 1 could be that increased competition especially amongst hedge funds drives away profits. Gatev et al (2006) and Pole (2007) raise this as a possible explanation. Another possibility stems from the fact that pairs trading profits, or more generally, statistical arbitrage profits, arise due to temporary liquidity pressure (i.e. a stock shot up relative to its peer simply because of a big order that is not absorbed by the current order book). Engelberg et al (2008) document a positive relation between pairs profits and illiquidity. Therefore, reduced profits can be a result of improved market liquidity over time due to innovations such as the emergence of algorithmic trading which is found to improve liquidity (see Hendershott, Jones and Menkveld, 2011).[6]

On the second conjecture, divergence, or at least non-convergence, occurs because there is a permanent price shock that either underlies or reinforces the initial divergence (e.g. an earnings announcement in one stock). This is closely related to fundamental risk in Shleifer and Vishny (1997). Along this line, three possible scenarios may be responsible for the profit

[5] There is another possibility where the decline can be driven by both reduced mispricing and increased divergence. This turns out not to be the case.

[6] In their study of short-term return reversals, Cheng, Hameed, Subrahmanyam and Titman (2014) also document a decline in short-term return reversal at both raw and industry-adjusted levels post 2000. They attribute this decline to recent innovations such as algorithmic trading that facilitates the entrance of market making capacity in response to unexpected withdrawal of informed investors.

decline. The first one involves changes in the information environment where firm-specific news arrives more frequently in the recent period. Regulatory changes such as Regulation Fair Disclosure enacted in 2000 help to increase the quantity of voluntary disclosure to the public (see Bailey, Li, Mao and Zhong, 2003). Also, according to Tetlock (2010), media coverage of corporate news has also increased over time (see Figure 1 in that study). This increased information flow in the second period may detrimentally affect our simple trading strategy which does not differentiate news driven divergence from no-news driven divergence.

A second scenario involves increased post-news drift in the second period. Using a public news database, Chan (2003) finds price drifts following headline news, especially bad news. If the drift is stronger in the second period, be it due to worsening behavioural biases in the sense of Daniel, Hirshsleifer and Subrahmanyam (1998) or increased participation of "momentum traders" versus "news watchers" according to Hong and Stein's (1999) model, arbitragers will experience divergence more frequently.

A third scenario that may contribute to greater divergence is related to price behaviours following price shocks that are not associated with news. There are two contrasting pieces of evidence regarding post no-news behaviours. Chan (2003) finds that stocks exhibit price reversals following the month with extreme returns that are not accompanied by news. In contrast, Gutierrez and Kelley (2008) find that extreme weekly returns are followed, initially by price reversals over the next 2 weeks, but then by price continuation thereafter, up to at least 50 weeks. Moreover, this pattern holds for weekly price movements with *and* without news. Using the term "implicit news" to refer to price shocks unaccompanied by public news (and "explicit news" for news driven price shocks), the authors conclude that "caution should be exercised in modelling traders as strictly overreacting to implicit news" (Gutierrez and Kelley, 2008, p. 417). In light of this latter evidence, it is possible that post news drift (either explicit news or implicit news or both) is stronger in the second period causing increased divergence risk in statistical arbitrage.

In summary, conjecture 2 whereby pairs trading profits decline due to increased divergence risk may be the result of either increased frequency of news arrival, increased price drifts following explicit news events or increased price drifts following both explicit and implicit news.

3 Empirical analysis

3.1 The profit decline and reduced mispricing

To test for the various explanations put forward, first, we compute three key statistics: the convergence ratio which is the empirical measure of E(I) in equation (4), the return per converged pair which is (approximately) the empirical measure of E($h_i|I=1$) and the return per non-converged pair, which is (approximately) the empirical measure of E(e_i)[7]. In addition, we also compute the duration of converged pairs and the duration of non-converged pairs. These statistics are computed for 2 sub-periods, January 1980-December 1995 and January 1996-December 2013.

All else the same, a decline in the mean return per converged pair would broadly be consistent with Conjecture 1. If such decline in the mean return is also accompanied by a shortening of the duration of converged pairs, we would interpret it as indirectly supportive of the increased competition story. Intuitively, if many arbitrageurs trade on the same signal, which is identified *ex post* as true mispricing, the expected profits amongst those trades will be lower.

[7] The first "approximately" refers to the fact that the return on a converged trade is slightly greater than the trading trigger by construction and the second "approximately" refers to the fact non-converged pairs include both diverged pairs with negative returns *e* and pairs that fail to converge completely with positive returns.

Panel A of Table 2 reports trading statistics at the pair trade level for each of the two sub-periods. When appropriate, a statistical test is also performed, comparing the trading statistics in the two sub-periods. We rely on these tests to make conclusions about what factors drive the profit decline. The tests are based on a sample of 28,637 pair trades in the first period and 30,080 trades in the second period. The mean return per trade is 1.69% in the second sub-period, which is statistically significantly lower than the mean return per trade of 3.17% in the first period. Although the distribution of returns is more volatile in the second period, evidenced by greater standard deviation and more extreme values, this decline in the mean returns at the trade level is consistent with the decline in the mean returns at the portfolio level reported in Table 1.

Several pieces of important evidence are revealed in Panel A. First, the convergence ratio drops from 54.75% to 46.34% in the second period. Second, the mean return of converged trades is higher in the second period, increasing from 8.37% to 10.29% per trade. Third, the mean duration of converged trades increases from 26 days to 30 days in the second period. All these changes are highly statistically significant (p-values=0.00).[8] The medians of the variables also point to the same direction. It appears that the pairs trading profit decline in the second period is driven by a greater proportion of pairs trades failing to converge which more than offsets the improved return per converged trade. To make matters worse, the higher divergence rate in the second period is also accompanied by significantly more negative returns amongst non-converged trades. The mean return per non-converged pairs is -5.74% in the second period compared to -3.13% in the first period. The evidence so far seems to be against Conjecture 1 in favour of Conjecture 2. At this juncture, we can also rule out increased competition as a potential explanation for the profit decline. Clearly, such story is inconsistent with higher convergence returns and longer time to convergence.

[8] A 1-sided, z-test for proportions is performed on the convergence ratios whilst a 1-sided, 2-sample t-test is performed on the other variables.

[Insert Table 2 about here]

A situation where return performance at the individual trade level may not translate directly to return performance at the portfolio level is that the number of trading opportunities varies between the two periods. That is, although portfolio returns are built up from the same number of pairs per cycle (100 pairs), not all of these pairs necessarily trade in their cycle, and for pairs that do, the number of re-openings (after successful convergence) may differ between pairs and cycles. This frequency aspect is possibly another source of variation in portfolio returns. Panel A reports that on average, a pair opens 1.69 times in the first period and 1.52 times in the second period. However, this reduced trading per pair can be due to the lower convergence rate observed earlier, as well as reduced trading opportunities.[9] Panel B sheds light on this issue by reporting performance at the cycle level. The performance at this level does depend on the frequency of trading, and is more directly comparable to the overall portfolio performance over time. Moreover, by aggregating trades within a cycle, the analysis is less sensitive to extreme variations at the individual pair level as seen in the distribution of returns per pair.

The first eight rows of Panel B portray a very similar picture as that in Panel A. The mean total payoff per cycle significantly drops from 485% to 235% in the second period (and the median dropping from 436% to 200%). In addition, this decline is driven by a lower convergence rate within each cycle, which, combined with worsening performance of non-converged trades, more than offsets the improved performance of converged trades. The new information is the frequency of trading. Although the number of traded pairs per cycle is unchanged at 91 pairs, the total number of trades per cycle significantly drops from 153 to 139. Is this reduced trading purely a by-product of a lower convergence rate in the second period or does it reflect reduced mispricing opportunities? To answer, we compute the mean

[9] Be design, when a pair fails to converge, it cannot be reopened in the same cycle.

total payoff on converged trades per cycle, which is the ultimate proxy for the mispricing amount. This statistic is 664% in the second period, dropping from 702% in the first period, but the drop is insignificantly different from zero. There is no evidence of reduced mispricing opportunity in the second period. The results in Table 2 collectively refute Conjecture 1 in support of Conjecture 2.

Whilst we have rejected Conjecture 1 and in particular ruled out increased competition as an explanation for the profit decline in the recent period, it is still of some interest to examine the state of illiquidity and how it may affect the dynamics of pairs trading profitability. After all, it is a major theoretical driver of profits in statistical arbitrage. To conserve space, we do not tabulate results of this analysis; instead just report the following points. First, in a panel regression with standard errors clustered by industry and cycle, profits per pair trade are significantly and positively related to the pairwise Amihud ratio. This confirms the findings in Engelberg et al (2008). Second, a univariate analysis similar to that in Table 2 shows that the level of Amihud ratios declines in the second period, for all traded pairs, as well as for converged pairs and non-converged pairs. Whilst this finding accords well with the generally accepted wisdom that liquidity has improved in recent years, it casts doubt over the previous finding against reduced mispricing in the second period. Intuitively, if there is a positive relation between a pair's profit and its pairwise Amihud illiquidity ratio and if this latter ratio declines, it is tempting to attribute any profit decline to the decline in the predictive variable. However, it turns out that the positive relation does not hold for the converged trades only sample or the non-converged trades only sample. The positive relation arises in the full sample because on average, the illiquidity ratio is higher for converged pairs than non-converged pairs. In other words, the illiquidity ratio is predictive of whether a pair is likely to converge or not, but is not predictive of the return level conditioned on the status of convergence.

3.2 Profit decline and increased divergence risk

Having refuted reduced mispricing in favour of increased divergence risk as the mechanism explaining the profit decline in the second period, we now examine whether such increased divergence is due changes in the news environment or changes in price reaction to news.

We obtain news events from Dow Jones Newswires as well as four major daily newspapers, the Wall Street Journal, the New York Times, USA Today and the Washington Post. This data is extracted from Factiva.[10] In addition to this media source, we also add quarterly earnings announcements taken from Compustat.[11] Following Engelberg et al (2008), a news event occurs if there is at least one public news item in the Factiva database or an earnings announcement made over window [-1,0] concerning a member of the pair, AND the residual estimated from the market model exceeds 2 standard deviations.[12] To ensure we also observe the negative relation between news and pair returns as documented in Engelberg et al (2008), we regress either returns or the convergence status against the news dummy. The estimated relation is indeed negative and highly significant (not reported). We then proceed to compute the proportion of pairs trades that are accompanied by a news event. If the increased divergence is driven by an increase in the frequency of news arrival, one would expect to see an increase in the proportion of pair trades whose initial divergence is associated with a news event.

Table 3 shows that such statistic increases slightly from 14.98% to 15.33%, but the increase is statistically insignificant (based on a *z*-test). As news occurring after opening may also adversely affect pair profits, we compute the proportion of trades that have at least a news

[10] To identify relevant articles, we perform a headline search similar to Chan (2003), matching company names against headlines and lead paragraphs. Company names are taken from CRSP to ensure the matching is done on the entire history of name changes.

[11] About 60% of earnings announcements are picked up by the public news data, therefore without the Compustat data, one would miss out on 40% of earnings events that are potentially relevant.

[12] We use 60 days to the opening date (date 0) to estimate the residual.

event after the opening and find that statistic to be smaller in the second period. When news both at and after opening are counted, the overall proportion of trades with such news declines as well. We therefore reject the notion that more frequent news arrival underlies the profit decline.

Next, we ask if changes in the price reaction to news drive down pairs profits. We answer this question by computing the probability of convergence amongst news-triggered trades. If such trades are less likely to converge, that would imply price continuation post the news events (or at least, weak reversion). Table 3 shows that conditioned on news, the chance of divergence in the first period is 50.90% which jumps to 60.03%. Similarly, the probability of divergence conditioned on news after also increases significantly in the second period. News does not arrive more often in the second period, but when it does, it more reliably predicts eventual divergence than the first period. Does this finding apply for news triggered trades only? We compute similar statistics for trades that are not accompanied by news events. Conditioned on no-news at the opening, the probability of divergence in the first period is 42.63% (i.e. the probability of successful convergence is 57.37%), jumping to 49.45% (i.e. the probability of successful convergence reduces to 50.55%). This difference is highly statistically significant. Overall, non-convergence is more likely to occur in pairs trading in the second period than in the first period, *regardless* of whether the trade is affected by news or not. The arbitrage is more risky as there is more uncertainty whether the signal represents mispricing or simply a trap. This increased divergence is the mechanism underlying the profit decline in pairs trading.

[Insert Table 3 about here]

3.3 Idiosyncratic return reversals - a systematic factor that is related to pairs trading profits

The above diagnostic shows that increased divergence is responsible for reduced profitability of pairs trading in the second period. More specifically, the second period seems to witness an increase likelihood of getting false signals. A price shock likely to be of idiosyncratic nature tends to be followed by continuation instead of reversals in the recent period, regardless of whether the shock is accompanied by explicit news or implicit news. It could be that the informational content in the news is stronger, thus the price effect is more lasting, at the detriment of statistical arbitrage strategies that formulate signals purely on prices. Whatever the underlying source, the tendency to drift instead of reverse after extreme idiosyncratic price movements must be driving the time series variation of statistical arbitrage.

To capture this effect systematically, we construct a daily portfolio that tracks the return of stocks that experience extreme idiosyncratic returns. We define idiosyncratic shocks as daily returns with absolute values exceeding 2 standard deviations of historical residuals from a daily market model augmented by industry return. We use Fama and French's (1997) system of industry classification and assign a stock to one of 48 industries based on four-digit SIC code. Daily rebalancing is necessary because the pairs trading strategy we try to explain is essentially a daily strategy. To be consistent with the pairs strategy, we use 1 year of daily data to estimate this model. Since convergence of a pair (or otherwise) is typically determined after a few weeks (see Table 2 as well as Engelberg et al 2008), we choose a holding period of 21 days. The strategy is applied to all US stocks that meets the same requirements as those applied to pairs trading (namely, the price must be greater than $5; the stock must trade every day over the training period etc.). The daily implementation means we end up with overlapping portfolios that stagger by 1 day to which we construct a time series of returns in a fashion analogous to Jegadeesh and Titman (1993). To capture the price reversal as opposed to price drift, the portfolio is short stocks that experience positive price shocks (daily return residual being greater than 2 standard deviations of historical residuals) and long stocks that experience negative price shocks (daily return residual being less than negative 2 standard deviations of historical residuals). We name the

Over the sample period, there are 8,577 daily returns which we aggregate to 408 monthly returns to match the time series of pairs trading returns. Figure 2 plot the time series returns of this factor, which we call *idiosyncratic reversal factor.*

[Insert Figure 2 about here]

Figure 2 shows a remarkable resemblance to Figure 1 where pairs trading returns were plotted. The idiosyncratic return reversal is particularly strong and stable for the first half of the sample and more erratic in the second half. The result is a generally downward trend over time, giving an impression of the factor becoming dormant in recent years as described in Gatev et al (2006).

Panel A of Table 4 provides key statistics of the distribution of the newly constructed systematic factor and its correlation with common factors. Panel A reports that over the full sample, 1980-2013, the average monthly return on the factor is 0.88% and highly statistically significant (t-statistic=12.40). The factor is particularly strong in the first half, at 1.26% per month (t-statistic=15.63), and more than halved in the second half, at 0.54% per month. Standard deviation jumps from a remarkably low level of 1.11% to 1.58%. This return profile is very similar to that of pairs trading. The factor is fairly strongly correlated with Jegadeesh's (1990) monthly reversal factor, negatively correlated with momentum, positively correlated with size and negatively correlated with value (although no correlation is detected for the first period). The factor is not correlated with the market risk factor.[13] Whilst its strong correlation with the monthly reversal factor is not surprising, it raises concerns as to whether

[13] There is no correlation with Pastor and Stambaugh's (2003) liquidity risk factor (the correlation coefficient is 0.02 for the full sample).

the new factor captures anything new in the cross section of stock returns or is it merely a variant of what we already know. However, it should be noted that the monthly factor captures return reversals in the month following the extreme performance month whereas the idiosyncratic reversal factor captures reversals following the day with extreme performance. Therefore the latter can be thought of as measuring post news reaction (both explicit and implicit news). In addition, the idiosyncratic factor captures the reversal of extreme performance in the return component that is net of market and industry factors whereas the monthly reversal factor is based on total return.

Nevertheless, the ultimate test is how well it can explain the pairs trading returns over and above the existing factors. Panel B reports the estimated coefficients from regressing the pairs trading returns on the five factors. Panel C reports the risk-adjustment results with the idiosyncratic reversal factor added. Part of Table 1 is reproduced in Panel B for easy reference. It is clear from Panel B that the common risk factors do little in explaining pairs trading profits with the intercept/alpha barely different from the raw return. Although the loading on momentum is negative and significant, it is too small to make a difference to the adjustment process. The loading on the monthly reversal factor, which is most closely related to the mean reversion phenomenon that pairs trading is premised upon, is only significant in the first period and insignificant for the full sample. Loadings on all other factors are both statistically and economically insignificant. In contrast, as shown in Panel C, the idiosyncratic reversal factor is strongly related to pairs trading returns with the loading statistically significant at 1% for the full sample as well as the two sub-periods (t-statistics well above 4). The factor is also economically significant. At 0.24, the loading implies that for a 1 percentage point decrease in the risk factor, pairs trading profits decrease by about 24 bps. Adding the idiosyncratic reversal factor also helps to significantly improve the explanatory power of the risk model. R^2 nearly triples, jumping from 0.09 to 0.25.

To quantify the contribution of the idiosyncratic reversal factor to explaining the pairs trading profit decline, note that for each sample period (as well as the full sample), the following holds statistically:

$$\frac{1}{n}\sum_{i=1}^{n} R_i =$$

$$\alpha + \gamma_1 \frac{1}{n}\sum_{i=1}^{n}\left(R_{m,i} - R_{f,i}\right) + \gamma_2 \frac{1}{n}\sum_{i=1}^{n} HML_i + \gamma_3 \frac{1}{n}\sum_{i=1}^{n} SML_i + \gamma_4 \frac{1}{n}\sum_{i=1}^{n} MOM_i +$$

$$\gamma_5 \frac{1}{n}\sum_{i=1}^{n} MonthlyReversal_i + \gamma_6 \frac{1}{n}\sum_{i=1}^{n} IdioReversal_i \quad (5)$$

Subtract both sides of the equation for the second period from those of the first period, the left hand side is the difference in the mean return between the two periods, which is 0.44%. The right hand side is the sum of the difference in the intercepts and the differences in the factor component (factor loading times the mean realized factor). Over the two periods, the idiosyncratic reversal factor component drops by 0.18%, representing 42% (0.18%/0.44%) of the drop in the average raw return. The alpha drops by 0.27%, representing 61% (0.27%/0.44%) of the drop in the raw return. Changes in the other factor components are minute, thus contributing little to explaining the decline in the raw return. Note that the decomposition is approximate as sample estimates are used in the above equation, thus the contributions from individual factors (the last row in Panel C) do not add up to 1. If the sum of the variations is used to compute the individual factor contribution (the second last row in Panel C), then the idiosyncratic reversal factor accounts for 38% of the variation in the raw return (as opposed to 42%). Either measure suggests that variation in the idiosyncratic reversal factor and the exposure to this factor explain a sizeable portion of the fall in pairs trading profits.

4 Robustness tests

The key findings so far can be summarized as follows. The decline in pairs trading profits is due to increased divergence risk. A systematic factor based on idiosyncratic return reversals appears to capture this divergence risk. Moreover, this new factor explains the time series variation in pairs trading returns as well as the decline in the returns over the sample period reasonably well. The following tests are performed to ensure our results are robust.

4.1 Correlation with the intra-industry reversal

First, our newly constructed idiosyncratic reversal factor seems to bear some resemblance to Hameed and Mian's (2014) intra-industry reversal factor which captures reversal from extreme performance relative to industry peers (as opposed to the entire market, as the case in Jegadeesh's (1990) "market-wide" reversal factor). By benchmarking against industry performance, the intra-industry reversal captures return behaviours following firm specific shocks. Unlike this factor, our idiosyncratic reversal factor tracks return behaviours using event time, i.e. 21 days following the price shock event. We believe this exact tracking is more relevant to the performance of daily price-based statistical arbitrage because by construction, an opening is likely to occur when a price shock event occurs. Nevertheless, it is important to ascertain that our factor is not subsumed in the intra-industry reversal factor. We follow the procedure in Hameed and Mian (2014) to construct the monthly returns of the intra-industry reversal strategy. For each of Fama and French's (1997) 48 industries, we sort industry members into winners and losers based on last month's raw returns using the top and bottom 3 deciles. Industry winners and losers are aggregated. The intra-industry reversal returns are returns to a monthly strategy of buying the industry losers and selling the industry winners. To conserve space, we do not tabulate this set of results, instead only describe pertinent points. Over our sample period, the intra-industry reversal has a mean return of 0.99% per month, dropping from 1.11% per month in the first period 1980-1995 to 0.88% in the second period 1996-201. The magnitude and the decline are very similar to those reported in Cheng et al (2014) who compare the performance of the strategy over 1980-1999 and 2000-2011. The correlation between the idiosyncratic reversal factor and the intra-industry reversal factor is 0.53 for the full sample, 0.46 for the first period and 0.57 for the second period. The pairs trading returns are then regressed on five factors, namely Fama-French factors, momentum and intra-industry reversals, and separately the same five factors augmented with the idiosyncratic reversal factor. The regressions are analogous to those reported to Panels B and C in Table 4, with the monthly reversal factor replaced by the intra-industry reversal. Without the idiosyncratic reversal factor, the intra-industry reversal is positive and significant at 1%. The coefficient is 0.06 for the full sample. When the idiosyncratic reversal factor is introduced to the risk model, the intra-industry reversal becomes insignificant for the full sample, and significant at 10% for the first period. In contrast, the idiosyncratic reversal factor is highly positive and significant. R^2 jumps from 0.13 to 0.24 over the full sample. We conclude that the idiosyncratic reversal factor subsumes the effect of the intra-industry reversal in explaining the variation in pairs trading returns.

4.2 Performance of pairs 101-200

Our results so far are based on a strategy that picks up, for each implementation cycle, the top 100 that have the lowest Euclidean distance amongst all matched pairs. To check whether our results are sensitive to this arbitrary choice, we now examine the next 100 pairs, i.e. pairs 101-200. This subset provides an interesting environment to test if the idiosyncratic reversal factor is truly systematic. Table 5 reports the trading statistics at the individual pair and cycle levels, analogous to Table 2 for the top 100 pairs. In this subset, we again observe a decline in the mean per-pair payoff (Panel A) and the mean per-cycle payoff (Panel B). At both pair and cycle levels, the convergence ratio worsens and the loss per non-converged pair increases over the two periods, with high statistical significance. The mean payoff per converged trade is statistically higher in the second period compared to the first whereas the mean of the mean payoff per converged trade across cycles is not statistically different between the two periods. The evidence again confirms against reduced mispricing and in support of increased divergence as the likely mechanism for the profit decline.

[Insert Table 5 about here]

Table 6 reports the raw and risk adjusted returns of the portfolio, analogous to Table 4. Again, we observe a decline in the raw portfolio return, from 0.81% to 0.54% over the two sample periods. The idiosyncratic reversal factor continues to explain well the variation in the portfolio return of this subset of pairs, with the loading highly significant and R^2 increasing from 0.13 to 0.24 when the factor is added. Using the decomposition in (11), the variation in the idiosyncratic factor is found to explain 40%-50% of the profit decline at the raw level.

[Insert Table 6 about here]

4.3 Cointegration-based statistical arbitrage

This section moves outside the traditional distance-based pairs trading method to explore statistical arbitrage using cointegration. Whilst the distance-based method is simple and involves no estimation risk, a major limitation is that it implicitly imposes a one-to-one restriction on the relationship between two stocks which is not necessarily always the case. In addition, it is unclear how it can be extended to baskets of stocks. The cointegration method described below accommodates any relative relation and can be extended to multiple stocks. Whilst we are certainly not the first to formulate a cointegration-based statistical arbitrage strategy, we are the first to conduct a comprehensive examination of the strategy over a relatively long sample. The ultimate purpose is still to assess if the declining trend in the profitability of statistical arbitrage spreads beyond the simple pairs trading strategy and that the idiosyncratic reversal factor is capable of explaining returns from other statistical arbitrage strategy.

4.3.1 Statistical arbitrage and cointegration method

Statistical arbitrage is based on the idea that prices of n stocks exhibit predictable behaviour over time. In time series analysis, such predictable evolution is the quality of what is known as (*covariance*) *stationary stochastic process*, or a random process in which the distribution at every point in time has the same mean and variance (and the covariance between different points only depends on their distance). Since a stationary process is stable over time, future realizations are predictable in the sense that if an observation falls in the tails of the distribution, it is likely subsequent moves will take the process nearer to the long term mean. For this reason, mean reversion is sometimes used to refer to stationarity.

This mathematical concept is particularly fitting to statistical arbitrage. If one can identify a relation among different asset prices that is stationary, they can formulate a trading strategy that is long (short) on this relation/portfolio if its current value is well below (above) its long run average. The finite variance in a stationary process enables one to use it as a basis to determine whether a given observation is too high or too low relative to the mean.

Cointegration, first introduced by Granger (1981), provides a natural tool to identify stationarity amongst stock prices that are well known to be non-stationary. To define the concept, consider n-dimensional time series $\boldsymbol{p}_t = [p_{1t} \quad p_{2t} \cdots p_{nt}]'$ in which all elements are I(1) meaning they are integrated of order 1 such that the differencing process p_t-p_{t-1} is stationary, or I(0). In general, linear combination $\beta' \boldsymbol{p}_t$ is *I(1)* as well. However, there may exist a special vector β that renders $\beta^{'}\boldsymbol{p}_t \sim I(0)$. In this case, these processes are said to be cointegrated of order (1,1) and β is the cointegrating vector.[14] Intuitively, in a cointegration system, the individual processes may wander unpredictably but they wander together along a common trend. To apply cointegration to SA, note that since $\beta^{'}\boldsymbol{p}_t$ is stationary, one can use it to formulate the trading signal. Specifically, in the same way that the *SSD* approach to pairs trading opens a long-short position whenever the absolute price spread $|p_{1t}-p_{2t}|$ exceeds two historical standard deviations, the cointegration approach also opens positions when the absolute spread $|\beta^{'}\boldsymbol{p}_t|$ exceeds its two standard deviations, which, by virtue of stationarity, are finite. Such method can be applied to any combination of stocks, the viability of which is only limited by the computing power. When applied to pairs of stocks, unlike the previously outlined non-parametric methods, this cointegration approach does not impose a one-for-one relationship between the two stocks since β needs not be $[1\ -1]^{'}$.

4.3.2 Implementation

We look for groups of n stocks that are cointegrated using historical data, estimate the cointegrating vector β, and compute the sample standard deviation for the spread $s_t = \beta^{'}\boldsymbol{p}_t$. The cointegrated baskets are then subjected to a trading rule in the subsequent period in which long/short positions in the baskets are established whenever s_t exceeds a certain threshold.[15] Obviously, central to the strategy is to test for cointegration and estimate the cointegrating vector. Two main tests are Engle and Granger (1987) and Johansen (1988), a brief summary of which are described in Appendix A. In theory, the MLE based method in Johansen (1988) is more rigorous as it, among other things, allows different hypotheses to be tested and does not require the econometrician to make a choice regarding the ordering of the tested series.

[14] As a more general definition, a vector of time series is I(d,b) if each element is I(d) and there exists a vector β such that the linear combination $\beta^T p_t$ is stationary after differencing d-b times.
[15] Here, the implementation is n at the price level. An alternative is to use the logarithm of prices in the test. Empirical results are qualitatively similar.

We choose to implement the Johansen method, but note that the results are qualitatively the same with the Engle-Granger method. In the Johansen method, the cointegrating relation is obtained when the following *Vector Error Correction Model* is estimated:

$$\Delta \boldsymbol{p}_t = C\boldsymbol{p}_{t-1} + B_1 \Delta \boldsymbol{p}_{t-1} + \cdots + B_q \Delta \boldsymbol{p}_{t-q} + \boldsymbol{\varepsilon}_t \qquad (6)$$

where $\Delta \boldsymbol{p}_t = [p_{1t} - p_{1t-1} \quad p_{2t} - p_{2t-1} \cdots p_{nt} - p_{nt-1}]'$, C is a nxn matrix such that:

$$C\boldsymbol{p}_{t-1} = A(\beta' \boldsymbol{p}_{t-1} + \beta_0) \qquad (7)$$

The intercept β_0 in the cointegrating relation is to provide for the case of non-zero mean spreads.

To avoid data snooping criticisms, whenever applicable, we adopt the same empirical setup used in the distance-based pairs trading literature. Thus, we use one year worth of daily closing prices for testing and estimation and conduct trading and measure performance in the subsequent six months. The trading threshold is chosen to be two historical standard deviations of the spread. Specifically, during the six-month trading period, short the basket whenever its spread $s_t = \beta' \boldsymbol{p}_{t-1}$ exceeds (2σ - β_0) (σ is the historical standard deviation of the spread) and long the basket if its spread is below (- 2σ - β_0). To enable comparison with the distance-based strategies, we need to (1) be able to express returns on \$1 investment and (2) achieve dollar neutrality between long and short positions,[16] and (3) achieve self-financing in the setup. To satisfy the first condition, we scale the individual positions so that the aggregate short position (across all stocks in the basket that are to be shorted) is \$1. To satisfy the second condition, we then require that the aggregate (scaled) long position is in the range of [0.9 1.1], so that the net dollar exposure is in the range of [-0.1 0.1]. The allowance

[16] This dollar neutrality condition is important to avoid being unduly exposed to adverse movement in 1 leg. On the other, one can vary this constraint as a way to adjust the "leverage" of the strategy. For example, if a net dollar exposure of up to say \$1 against an aggregate short position of \$1 is permitted, the maximum aggregate long position is \$2. Such strategy would be more sensitive to performance on the long leg. For direct comparison with the traditional pairs trading strategy, we keep the net exposure approximately dollar neutral.

of 0.1 is because it is unlikely to have a situation where the short position exactly offsets the long position. Appendix B contains numerical examples of the trading rule. For the last condition, any non-zero net dollar exposure is offset by a cash account earning and funded at the Fed rate.

Finally, to consistent with the distance-based pairs trading setup, over the trading period, baskets can be opened and closed multiple times and non-converged pairs are to be closed at the end of the period. Basket and portfolio returns are computed in the same way as before. The daily marked-to-market payoff is obtained for each position throughout the trading cycle, and they are summed across positions to obtain a time series of payoffs to the basket. Averaging these payoffs across baskets produces the time-series of payoffs to the portfolio. Buy-and-hold monthly portfolio returns are obtained by compounding daily payoffs. These monthly returns have the interpretation of excess returns to an approximately dollar neutral portfolio with an aggregate short position of $1. The remaining of this section reports the performance of the strategy applied to 2 cases, n=2 (pairs) and n=3 (triplets).

4.3.3 Performance of cointegration-based statistical arbitrage

Figure 3 plots the monthly return of the pairs trading strategy. There is some weak declining trend in the plot. The plot shows severe spikes in the monthly returns in 2000 and, to a lesser extent, in 2009. A closer look at the 2000 spike reveals that the big loss in 2000 occurred in February of that year, just preceding the dot com bust. The price run-up in the tech stocks must have induced these losses, which are quickly reversed in the following month, leading to a huge profit in March 2000. We did observe similar albeit less severe spikes around these time points in the return time series of the distance-based strategy. It turns out the pairs that cause these spikes in Figure 3 are also picked up in the distance-based strategy, however, because we truncated the trading pool to the top 200 pairs, those pairs did not enter into the portfolio. Nevertheless, we need to be conscious of the impact of these outliers on our analysis.

Table 7 reports the trading statistics of the cointegration-based pairs trading strategy over the two periods being investigated. The set of results marked "Raw data" ignores the outliers whereas the second set, marked "Payoff winsorized at [1 99]", is the same as the first set, except that the distribution of pair payoffs is winsorized at the top and the bottom percentile. 29,689 (28,607) cointegrated pairs that meet our trading requirements are traded in the first (second) period. The number repeated trade per pair (Panel A) as well as the number of traded pairs per cycle (Panel B) drop in the second period. With this strategy, we no longer see clear-cut evidence of declining profitability at the pair or cycle levels. If anything, the mean payoff per pair is higher in the second period. The winsorization does not make much difference to the results. However, we continue to observe lower convergence in the second period, dropping from 64.57% to 51.98% at the pair level (Panel A) or from 61.24% to 45.06% if the convergence rate is computed for each cycle and the statistics are averaged across all cycles (Panel B). In addition, we also continue to document increased positive payoffs on converged trades and increased negative payoffs on non-converged trades. Overall, we continue to see evidence that links the overall profit decline to increased divergence as opposed to reduced mispricing.

[Insert Table 7 about here]

Table 8 reports the raw and risk-adjusted returns of the cointegration-based pairs trading strategy. Consistent with the above treatment of outliers, we winsorize the return time series at [1 99]. The raw return drops from 0.72% to 0.49% in the second period, but the difference is no longer significant. Although the decline between the periods is not statistically significant (possibly thanks to positive spikes), it is hard to deny the fact the strategy is increasingly unprofitable towards the end of the sample. Panel A shows that this cointegration strategy is particularly sensitive to momentum and to a lesser degree, monthly reversal. Panel B shows that, the idiosyncratic reversal, when added to the set of explanatory factors, is highly statistically and economically significant. The variation in this component explains about 60% of the drop in the means of the raw returns over the two periods.

[Insert Table 8 about here]

Finally, we report the results on the triplets trading strategy using cointegration. The main tenor of these results is very similar to that of the pairs trading case. Figure 4 plots the monthly return of the triplet case and demonstrates strong resemblance with Figure 3. Table 9 reports the trading statistics of the strategy at the trade and cycle level. More than a million trades are examined in each period (Panel A) compared to less than 30,000 trades for the pairs trading strategies. The median number of unique triplets traded per cycle is 2,773 in the first period, and 2,590 in the second period. These compare to less than 100 pairs per cycle in the base of pairs trading (see Table 7). This highlights the point that as the basket size increases, the number of trading opportunities increases exponentially.[17] This aspect aside, the patterns in the distribution of payoffs in the case of pairs trading continue to feature in this triplets trading. That is, we continue to see reduced convergence, increased payoff per converged trade and worsening negative payoff per non-converged trade.

[Insert Table 9 about here]

Table 10 reports the raw and risk-adjusted returns for the triplets portfolio. The average raw return reduces from 0.71% per month to 0.35% per month. Unlike the cointegration-based pairs trading results, the decline is highly statistically significant. Like the pairs trading

[17] Needless to say, the implementation on the triplets strategy is highly computationally intensive. We employ parallel computing to handle the task, and with 40 nodes, we are able to reduce the run time from several weeks to a day or so.

results throughout this paper, the idiosyncratic reversal factor is positively and significantly related to the triplets portfolio returns. The newly constructed factor also subsumes the effect of the monthly reversal factor (except for the second period where the loading on the monthly reversal factor is significant at 10%). Variation in this factor explains about 17%-20% of the drop in the raw returns.

5 Conclusion

This study examines and explains the time series variation in the simulated profitability of various statistical arbitrage strategies that are implemented on the U.S. equity market over the period 1980-2013. Using closing daily prices to profit from price divergence among stocks has become increasingly unprofitable since late 1990, early 2000. Increased competition, reduced mispricing or changes in the information environment do not appear to be responsible for this well-documented trend in statistical arbitrage. Instead, we find strong evidence of increased divergence, which is linked to the idiosyncratic return reversal factor. This newly constructed systematic factor, which captures price behaviours following idiosyncratic return shocks, is found to be highly related to statistical arbitrage returns and explain a great part of the fall in the strategy's profitability. The findings extend from the traditional distance-based pairs trading strategy to cointegration-based methods involving baskets of stocks. The idiosyncratic reversal factor can be used to evaluate performance of statistical arbitrage strategies, which may still be profitable outside the daily, price-based implementation setting. Investigation of high frequency statistical arbitrage or strategies that combine price with microstructure and/or fundamental data to better differentiate between noise and signal is an interesting topic for future research.

APPENDIX A –Cointegration Tests

Engle and Granger (1987) propose a two-stage method in which the first stage involves regressing one price series on the rest:

$$p_{1t} = \beta_0 + \beta_2 p_{2t} + \beta_3 p_{3t} + \cdots + \beta_n p_{nt} + \epsilon_t \quad \text{(A.1)}$$

In the second stage, a unit root test is performed on the regression residual $\widehat{\epsilon_t}$. Rejection of a unit root on $\widehat{\epsilon_t}$ means the prices are cointegrated of order (1,1) with cointegrating vector $\beta = [1 - \beta_2 - \beta_3 - \ldots - \beta_n]^T$. Since the stationarity test is performed on the regression residual $\widehat{\epsilon_t}$, inference is made based on a special table of critical values (provided in Engle and Granger, 1987). The econometric literature on cointegration highlights several shortcomings under the Engle-Granger test. For instance, the procedure requires making arbitrary choice as to which series is normalized to unity and thus appears in regression (A.1) as the dependent variable. This normalization choice leads to different estimates of the cointegrating relation and possibly different conclusions about cointegration. Another complication arises when there is more than one cointegration relation when n>2. The Engle-Granger test just picks up one of such relations.

An alternative to the Engle-Granger test for cointegration is a maximum-likelihood (MLE) framework introduced by Johansen (1988, 1991). Granger representation theorem (Engle and Granger, 1987) shows that if *n* price series $p_{1t}, p_{2t}, \ldots, p_{nt}$ are cointegrated, their dynamics can be represented by a *vector error correction model* VEC:

$$\Delta \boldsymbol{p}_t = C\boldsymbol{p}_{t-1} + B_1 \Delta \boldsymbol{p}_{t-1} + \cdots + B_q \Delta \boldsymbol{p}_{t-q} + \boldsymbol{\varepsilon}_t \quad \textbf{(A.2)}$$

Where $\boldsymbol{p}_t = [p_{1t} \quad p_{2t} \cdots p_{nt}]'$, $\Delta \boldsymbol{p}_t = [p_{1t} - p_{1t-1} \quad p_{2t} - p_{2t-1} \cdots p_{nt} - p_{nt-1}]'$, C is a nxn matrix such that:

$$C\boldsymbol{p}_{t-1} = A\beta' \boldsymbol{p}_{t-1} \quad \textbf{(A.3)}$$

Under this dynamic, changes in each series are related to past changes in that series itself and past changes in the other series, as well as the scalar $\beta'\boldsymbol{p}_{t-1}$, $B = [\beta_1 \quad \beta_2 \ldots \beta_n]'$. That is, the *levels* of $p_2, p_3, \ldots p_n$ contain information that is useful for forecasting p_1 beyond that contained in a finite number of lagged changes in $p_1, p_2 \ldots p_n$. When C is zero, the system collapses to a *vector autoregressive model* (VAR) which means the variables may be relevant to forecasting one another but they do not move together along some common stochastic trend.

Of interest to the SA application is that $B'\boldsymbol{p}_{t-1}$ is stationary and B is the cointegrating vector which is an *n-by-r* matrix that contains r independents cointegrating vectors. Whereas the Engle-Granger test is for the null hypothesis of no cointegration against the alternative of cointegration, Johansen (1988, 1991) provides procedures to test for the number of cointegrating relations, and in the process, obtain MLE estimates of the coefficients in Eq. (12).

Two hypothesis tests are possible under the Johansen "approach". The first test, known as "Trace" test, is for the null of r cointegrating relations against the alternative of n cointegrating relations, $r<n$. Note that the presence of n cointegrating vectors amongst n I(1) series means the series themselves are already stationary. The second test, using "L-max" statistics, allows one to test the null of r cointegrating relations against the alternative of r+1 relations. For the SA application, the interest is whether or not there is at least 1 cointegrating relation. Therefore, in our opinion, the "L-max" approach is more relevant as we can test the null of no cointegration against the alternative of 1 cointegration relation. We use this test in our implementation, in particular, we use the result from the test of r=0 against r=1.

Finally, we include 1 lag term in the error term in (A.2) to account for any remaining autocorrelation. We also include a constant in the cointegration relation. That is, (A.3) becomes:

$$C\boldsymbol{p}_{t-1} = A(\beta'\boldsymbol{p}_{t-1} + \beta_0) \tag{A.4}$$

APPENDIX B – Numerical examples of basket trading rules

Example 1: Spread exceeds threshold but there is no trade					
Upper threshold	24				
Lower threshold	5				
	beta	Current price	Position	Scaled position	
Stock 1	1	100	100	1.00	
Stock 2	-0.5	45	-22.5	-0.23	
Stock 3	-0.8	65	-52	-0.52	
Spread			25.5	>24: violates the upper threshold	
Aggregate short				1	
Aggregate long				-0.75	
Net dollar exposure				0.26	>0.1: there is no trade
Example 2: Spread exceeds threshold and there is a trade					
Upper threshold	1				
Lower threshold	-15				
	beta	Current price	Position	Scaled position	
Stock 1	0.8	150	120	1.00	
Stock 2	0.3	60	18	0.15	
Stock 3	-1.5	90	-135	-1.13	
Spread			3	>1: violates the upper threshold	
Aggregate short				1.00	
Aggregate long				-0.98	
Net dollar exposure				0.03	>0.1 and <-0.1: there is a trade

REFERENCES

Abreu D and Brunnermeier MK. (2002) Synchronization risk and delayed arbitrage. *Journal of Financial Economics* 66(2-3): 341-360.

Andrew P. (2007) *Statistical Arbitrage: Algorithmic Trading Insights and Techniques,* New Jersey: John Wiley & Sons.

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

Bailey W, Haitao LI, Mao CX, et al. (2003) Regulation Fair Disclosure and Earnings Information: Market, Analyst, and Corporate Responses. *Journal of Finance* 58(6): 2487-2514.

Chan W. (2003) Stock Price Reaction to News and No-News: Drift and Reversals after Headlines. *Journal of Financial Economics* 70: 223-260.

Cheng S, Hameed A, Subrahmanyam A, et al. (2014) Short-Term Reversals and the Efficiency of Liquidity Provision. *Working Paper.*

Daniel K, Hirshleifer D and Subrahmanyam A. (1998) Investor Psychology and Security Market under- and Overreactions. *Journal of Finance* 53(6): 1839-1885.

Do B and Faff R. (2010) Does Simple Pairs Trading Still Work? *Financial Analysts Journal* 66(4): 1-13.

Do B and Faff R. (2012) Are Pairs Trading Profits Robust to Trading Costs? *Journal of Financial Research* 35(2): 261-287.

Engelberg J, Gao P and Jagannathan R. (2008) An Anatomy of Pairs Trading: the Role of Idiosyncratic News, Common Information and Liquidity. *www.ssrn.com*.

Engle RF and Granger CWJ. (1987) Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55(2): 251-276.

Fung W and Hsieh D. (2001) The risk in hedge fund strategies: Theory and Evidence from Trend Followers. *Review of Financial Studies* 14(2): 313-341.

Gatev E, Goetzmann W and Rouwenhorst KG. (2006) Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Review of Financial Studies* 19(3): 797-827.

Granger CWJ. (1981) Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*: 121-130.

Gutierrez R and Kelley EK. (2008) The Long-Lasting Momentum in Weekly Returns. *Journal of Finance* 63(1): 415-447.

Hameed A and Mian GM. (2014) Industries and Stock Return Reversals. *Journal of Quantitative and Financial Analysis* Forthcoming.

Hendershott TJ, Jones CM and Menkveld AJ. (2011) Does Algorithmic Trading Improve Liquidity? *Journal of Finance* 66(1): 1-33.

Hong H and Stein JC. (1999) A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *Journal of Finance* 54(6): 2143-2184.

Jegadeesh N. (1990) Evidence of Predictable Behavior of Security Returns. *Journal of Finance* 45(3): 881-898.

Jegadeesh N and Titman S. (1993) Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48(1): 65-91.

Johansen S. (1988) Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12: 231-254.

Johansen Sr. (1991) Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 56: 1551-1580.

Mitchell M and Pulvino T. (2001) Characteristics of Risk and Return in Risk Arbitrage. *Journal of Finance* 56(6): 2135-2175.

Mitchell M, Pulvino T and Stafford E. (2002) Limited Arbitrage in Equity Markets. *Journal of Finance* 57(2): 551-584.
Pástor Lu and Stambaugh RF. (2003) Liquidity Risk and Expected Stock Retums. *Journal of Political Economy* 111(3): 642-685.
Schleifer A and Vishny R. (1997) The Limits of Arbitrage. *Journal of Finance* 52(1): 35-55.
Stein JC. (2009) Presidential Address: Sophisticated Investors and Market Efficiency. *Journal of Finance* 64(4): 1517-1548.
Tetlock PC. (2010) Does Public Financial News Resolve Asymmetric Information? *Review of Financial Studies* 23(9): 3520-3557.
Vidyamurthy G. (2004) *Pairs Trading: Quantitative Methods and Analysis*: Wiley Finance.
Xiong W. (2001) Convergence trading with wealth effects: an amplification mechanism in financial markets. *Journal of Financial Economics* 62: 247-292.

Table 1 – Excess returns to the distance-based pairs trading strategy

This table reports raw and risk-adjusted returns to a simple pairs trading strategy in which pairs are matched amongst stocks of the same SIC code and by minimizing the sum of squared differences (SSD) in total return indices scaled to $1 at the beginning of the 1-year training period. Each month over period January 1980 to December 2013, 100 pairs with the lowest SSD are subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever its absolute spread exceeds 2 standard deviations of historical spreads, with the position unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. The returns are monthly excess returns that average across 6 overlapping portfolios that stagger by 1 month. The five-factor model is Fama-French's (1997) three-factor model augmented with the momentum factor and short-term reversal factor. t-statistics (reported the second row for each period) are computed using Newey-West standard errors with 6 lags. *,**,*** indicate statistical significance at 10%,5%, and 1% levels, respectively.

Subperiod	Raw Return	Standard Deviation	Skewness	Sharpe Ratio	Alpha Fama-French	Beta Fama-French	Alpha Five-Factors	Beta Five-Factors
Full sample	0.60%	0.73%	1.06	0.82	0.59%	0.00	0.63%	-0.01
	11.03				10.11	0.29	10.54	-0.59
Jan1980-Dec1995	0.83%	0.66%	1.77	1.26	0.83%	-0.01	0.86%	-0.01
	11.12				9.36	-0.50	9.55	-0.58
Jan1996-Dec2013	0.39%	0.72%	0.99	0.54	0.38%	0.02	0.41%	-0.01
	6.88				6.31	1.13	6.97	-0.40

Table 2 – Performance statistics at the individual pair and cycle levels

This table reports the return distribution and composition of converged and non-converged pairs at the individual trade (Panel A) and trading cycle (Panel B) levels for a simple pairs trading strategy. In this strategy, each month over period January 1980 to December 2013, pairs are matched amongst stocks of the same SIC code and by minimizing the sum of squared differences (SSD) in total return indices scaled to $1 at the beginning of the 1-year training period. 100 pairs with the lowest SSD are subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever its absolute spread exceeds 2 standard deviations of historical spreads, with the position unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. p-values are from one-sided t-test of two samples except for the test on convergence ratio which is a z-test of proportions. The tail of the test is based on the relative magnitude of the means of the variable in two periods. For example, the mean payoff per trade in the second period, 1.69%, is less than the mean payoff per trade in the first period, 3.17%. In this case, the alternative hypothesis is that the mean payoff per trade in the second period is less than the first period (i.e. left tail test). *,**,*** indicate statistical significance at 10%,5%, and 1% levels, respectively.

	1980-1995	1996-2013	p-value
Panel A: Performance statistics at the individual trade level			
Number of trades	28,637	30,080	na
Mean number of trades per pair	1.6864	1.5244	na
Payoff per trades Mean	3.17%	1.69%	0.00
Median	5.38%	4.66%	na
Standard deviation	8.55%	12.25%	na
Min	-132.06%	-174.01%	na
Max	68.78%	116.07%	na
Convergence ratio	54.75%	46.34%	0.00
Mean payoff on converged trades	8.37%	10.29%	0.00
Median payoff on converged trades	7.80%	9.20%	na
Mean payoff on non-converged trades	-3.13%	-5.74%	0.00
Median payoff on non-converged trades	-1.33%	-2.96%	na
Mean duration of converged trades	26	30	0.00
Median duration of converged trades	19	22	na
Mean duration of non-converged trades	60	64	0.00
Median duration of non-converged trades	60	66	na
Panel B: Performance statistics at the trading cycle level			
Mean payoff per cycle	485%	235%	0.00
Median payoff per cycle	436%	200%	na
Mean convergence ratio	53.30%	44.87%	0.00
Median convergence ratio	51.97%	45.54%	na
Mean payoff per converged trade	8.40%	10.08%	0.00
Median payoff per converged trade	8.19%	9.36%	na
Mean payoff per non-converged trade	-3.10%	-5.65%	0.00
Median payoff per non-converged trade	-2.89%	-4.84%	na
Mean number of unique traded pairs per cycle	91	91	0.13
Median number of unique traded pairs per cycle	92	93	na
Mean number of trades	153	139	0.00
Median number of trades	148	135	na
Mean total payoff on converged trades	702%	664%	0.89
Median total payoff on converged trades	634%	563%	na

Table 3 –The frequency of idiosyncratic news and pairs trading performance

This table compares the impact of news arrival on pairs trading over two periods January 1980-December 1995 and January 1996-December 2013. A news event is said to occur with a pair if at least one of its stocks has a news event being a (1) news article from five sources of public news, Dow Jones Newswire, the Wall Street Journal, the New York Times, USA Today and the Washington Post, or (2) a quarterly earnings announcement. p-values are from one-sided z-test of proportions. *,**,*** indicate statistical significance at 10%,5%, and 1% levels, respectively.

	1980-1995	1996-2013	p-value
Proportion of trades with news at opening	14.98%	15.33%	0.17
Proportion of trades with news after opening	59.06%	57.01%	0.00
Proportion of trades with news at or after opening	63.88%	62.49%	0.00
Proportion of trades with news at opening that did not converge	50.90%	60.03%	0.00
Proportion of trades with news after opening that did not converge	52.93%	59.99%	0.00
Proportion of trades with news at or after opening that did not converge	51.83%	59.17%	0.00
Proportion of trades with no news at opening that did not converge	42.63%	49.45%	0.00
Proportion of trades with no news after opening that did not converge	30.80%	39.25%	0.00
Proportion of trades with no news at or after opening that did not converge	29.79%	37.58%	0.00

Table 4 –Idiosyncratic return reversals as a systematic factor

Over the period from January 1980 to December 2013, each day, the residual from model $r_{i,t} = \alpha_i + \beta_0 \ R_{M,t} + \delta_0 R_{I,t} + \omega_{i,t}$ is estimated for each stock using its prior 252 daily returns. R_M denotes the corresponding market return and R_I denotes the stock's industry return, based on Fama and French's (1997) 48 industry classification. A stock is classified as an *idiosyncratic winner* (*idiosyncratic loser*) if its residual return on the day is greater (less) than 2 standard deviations of the stock's historical residuals. A portfolio is formed that is long idiosyncratic losers and short idiosyncratic winners and is held for 21 days. The strategy is rolled over daily, giving rise to 21 overlapping portfolios. Daily idiosyncratic reversal factor returns are computed by averaging the returns on the same day that are produced by the 21 overlapping portfolios (i.e. the method is analogous to that in Jegadeesh and Titman, 1993). The idiosyncratic return reversal factor is the monthly return time series which aggregates from the daily returns. Panel A reports the mean and standard deviation of this monthly return time series as well as its correlation with six well-known risk factors. *Mkt-Rf, HML* and *SML* are respectively the market, value/growth and size factors from Fama and French (1997). *Momentum* is the 6-month momentum factor (Jegadeesh and Titman, 1993). *Monthly Reversal* is Jegadeesh's (1990) short –term reversal factor. Panel B(C) reports the risk-adjustment of the pairs trading strategy described in Table 1, with the risk model comprising the six common factors (the six common factors plus the idiosyncratic reversal factor). ***,**,* indicate statistical significance at 1%, 5% and 10% respectively.

Panel A : Idiosyncratic reversal and other risk factors

			Correlation				
	Mean	Standard Deviation	Mkt-Rf	HML	SML	Momentum	Monthly Reversal
Full sample	0.0088	0.0143	0.01	-0.24	0.21	-0.25	0.44
	12.40						
Jan1980-Dec1995	0.0126	0.0111	-0.03	-0.03	0.10	-0.30	0.39
	15.63						
Jan1996-Dec2013	0.0054	0.0158	0.02	-0.32	0.27	-0.26	0.48
	4.99						

Panel B: Risk-adjustment of pairs trading returns using existing risk factors

	Raw return	Alpha	Mkt-Rf	HML	SML	Momentum	Monthly Reversal	R-squared
Full sample	0.0060	0.0063	-0.0088	-0.0083	-0.0063	-0.0414	0.0222	0.09
	11.03	10.54	-0.59	-0.72	-0.52	-4.65	1.50	
Jan1980-Dec1995	0.0083	0.0086	-0.0138	-0.0349	-0.0024	-0.0416	0.0456	0.12
	11.12	9.55	-0.58	-1.45	-0.10	-2.80	2.11	
Jan1996-Dec2013	0.0039	0.0041	-0.0058	0.0042	-0.0101	-0.0427	0.0147	0.12
	6.88	6.97	-0.40	0.34	-0.67	-4.91	0.94	

Panel C: Risk-adjustment of pairs trading returns with the idiosyncratic reversal factor added

	Alpha	Mkt-Rf	HML	SML	Momentum	Monthly Reversal	Idiosyncratic Reversal	R-squared
Full sample	0.0042	-0.0069	0.0181	-0.0209	-0.0345	-0.0227	0.2432	0.25

	8.03	-0.53	1.40	-1.68	-4.39	-1.67	6.63	
Jan1980-Dec1995	0.0057	-0.0125	-0.0255	-0.0089	-0.0323	0.0072	0.2342	0.25
	8.83	-0.61	-1.19	-0.39	-2.41	0.36	5.29	
Jan1996-Dec2013	0.0030	-0.0015	0.0294	-0.0241	-0.0356	-0.0242	0.2056	0.24
	5.48	-0.12	2.34	-1.72	-4.35	-1.48	4.59	
Change in loading *x* factor mean	-0.0027	0.0000	0.0001	-0.0001	-0.0002	-0.0003	-0.0018	
Contribution based on								
Approximated change in raw return	55%	-1%	-3%	1%	4%	6%	38%	
Actual change in raw return	61%	-1%	-3%	2%	4%	7%	42%	

Table 5 – Performance statistics at the individual pair and cycle levels – pairs 101-200

Each month over period January 1980 to December 2013, pairs are matched amongst stocks of the same SIC code and by minimizing the sum of squared differences (SSD) in total return indices scaled to $1 at the beginning of the 1-year training period. Pairs 101 to 200 based on the ascending ranking of SSD are subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever its absolute spread exceeds 2 standard deviations of historical spreads, with the position unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. This table reports the return distribution and composition of converged and non-converged pairs at the individual trade (Panel A) and trading cycle (Panel B) levels *,**,*** indicate statistical significance at 10%,5%, and 1% levels, respectively.

		1980-1995	1996-2013	p-value
Panel A: Performance statistics at the individual trade level				
Number of trades		23,337	27,230	na
Mean number of trades per pair		1.4903	1.4515	na
Payoff per trades	Mean	3.57%	2.45%	0.00
	Median	7.83%	5.64%	na
	Standard deviation	14.02%	15.55%	na
	Min	-124.63%	-151.66%	na
	Max	60.85%	155.96%	na
Convergence ratio		47.35%	43.59%	0.00
Mean payoff on converged trades		13.72%	14.12%	0.00
Median payoff on converged trades		12.81%	12.55%	na
Mean payoff on non-converged trades		-5.56%	-6.57%	0.00
Median payoff on non-converged trades		-2.52%	-3.17%	na
Mean duration of converged trades		30	30	0.51
Median duration of converged trades		23	23	na
Mean duration of non-converged trades		61	63	0.00
Median duration of non-converged trades		62	65	na
Panel B: Performance statistics at the trading cycle level				
Mean payoff per cycle		446%	308%	0.00
Median payoff per cycle		400%	241%	na
Mean convergence ratio		45.91%	41.68%	0.00
Median convergence ratio		44.68%	41.03%	na
Mean payoff per converged trade		13.87%	13.78%	0.37
Median payoff per converged trade		13.71%	12.69%	na
Mean payoff per non-converged trade		-5.44%	-6.46%	0.00
Median payoff per non-converged trade		-5.33%	-5.77%	na
Mean number of unique traded pairs per cycle		84	87	0.00
Median number of unique traded pairs per cycle		86	88	na
Mean number of trades		125	126	0.67
Median number of trades		120	120	na
Mean total payoff on converged trades		811%	776%	0.20
Median total payoff on converged trades		717%	630%	na

Table 6 – The portfolio returns of pairs 101-200 and risk adjustment

Each month over period January 1980 to December 2013, pairs are matched amongst stocks of the same SIC code and by minimizing the sum of squared differences (SSD) in total return indices scaled to $1 at the beginning of the 1-year training period. This table reports the raw and risk adjusted returns to the portfolio that comprises pairs 101 to 200 in terms of lowest SSD. These pairs are subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever its absolute spread exceeds 2 standard deviations of historical spreads, with the position unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. Panel A report risk adjustment results using five common risk factors. *Mkt-Rf, HML* and *SML* are respectively the market, value/growth and size factors from Fama and French (1997). *Momentum* is the 6-month momentum factor (Jegadeesh and Titman, 1993). *Monthly Reversal* is Jegadeesh's (1990) short –term reversal factor. Panel B reports risk adjustment results with an idiosyncratic reversal factor added. The description of this factor and its statistics are provided in Table 4.

Panel A: Risk-adjustment of pairs trading returns using common risk factors

	Raw return	Alpha	Mkt-Rf	HML	SML	Momentum	Monthly Reversal	R-squared
Full sample	0.0067	0.0068	0.0004	-0.0072	-0.0057	-0.0466	0.0423	0.13
	11.94	10.69	0.02	-0.51	-0.37	-4.96	2.87	
Jan1980-Dec1995	0.0081	0.0083	0.0133	0.0017	0.0021	-0.0619	0.0647	0.18
	12.04	10.69	0.45	0.06	0.08	-2.99	3.51	
Jan1996-Dec2010	0.0054	0.0055	-0.0067	-0.0094	-0.0095	-0.0453	0.0336	0.12
	6.73	6.07	-0.28	-0.49	-0.48	-3.96	2.00	

Panel B: Risk-adjustment of pairs trading returns with the idiosyncratic reversal factor added

	Alpha	Mkt-Rf	HML	SML	Momentum	Monthly Reversal	Idiosyncratic Reversal	R-squared
Full sample	0.0049	0.0022	0.0180	-0.0196	-0.0399	-0.0008	0.2330	0.24
	8.12	0.13	1.49	-1.26	-4.12	-0.06	5.33	
Jan1980-Dec1995	0.0058	0.0145	0.0101	-0.0037	-0.0535	0.0304	0.2096	0.26
	9.59	0.56	0.36	-0.15	-2.91	1.42	3.51	
Jan1996-Dec2010	0.0044	-0.0021	0.0176	-0.0246	-0.0376	-0.0082	0.2211	0.22
	5.50	-0.10	1.06	-1.25	-3.19	-0.46	3.72	
Change in loading *x* factor mean	-0.0014	-0.0001	0.0000	-0.0001	-0.0001	-0.0005	-0.0014	
Contribution based on								
Approximated change in raw return	38%	3%	0%	3%	3%	14%	40%	
Actual change in raw return	49%	4%	0%	3%	3%	18%	52%	

Table 7 – Performance statistics at the individual pair and cycle levels – cointegration-based strategy

Each month over period January 1980 to December 2013, cointegrated pairs are identified amongst stocks of the same SIC code, with a Johansen cointegration test (Johansen, 1988) performed on the price level with 1 lag term. These pairs are then subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever the demeaned cointegration spread exceeds 2 standard deviations of historical spreads, provided the resulting net dollar exposure does not exceed $0.1 with the short position scaled at $1. The trade is unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. This table reports the return distribution and composition of converged and non-converged pairs at the individual trade (Panel A) and trading cycle (Panel B) levels. The last three columns represent the set of results with payoffs winsorized at [1 99] to minimize the effect of the spikes in February-March 2010. *,**,*** indicate statistical significance at 10%,5%, and 1% levels, respectively.

		Raw data			Payoff winsorized at [1 99]		
		1980-1995	1996-2013	p-value	1980-1996	1996-2014	p-value
Panel A: Performance statistics at the individual trade level							
Number of trades		29,689	28,607	na	29,689	28,607	na
Mean number of trades per pair		1.5797	1.2837	na	1.5797	1.2837	na
Payoff per trades	Mean	3.14%	3.76%	0.01	3.16%	4.49%	0.00
	Median	5.33%	9.63%	na	5.33%	9.63%	na
	Standard deviation	13.55%	41.59%	na	13.16%	32.52%	na
	Min	-286.60%	-1554.07%	na	-103.66%	-103.66%	na
	Max	125.09%	394.70%	na	68.39%	68.39%	na
Convergence ratio		64.57%	51.98%	0.00	64.57%	51.98%	0.00
Mean payoff on converged trades		8.68%	24.09%	na	8.66%	23.30%	na
Median payoff on converged trades		6.68%	19.50%	na	6.68%	19.50%	na
Mean payoff on non-converged trades		-6.94%	-18.23%	0.00	-6.86%	-15.86%	0.00
Median payoff on non-converged trades		-3.65%	-9.10%	na	-3.65%	-9.10%	na
Mean duration of converged trades		24	26	0.02	24	26	0.02
Median duration of converged trades		16	18	na	16	18	na
Mean duration of non-converged trades		75	76	0.00	75	76	0.00
Median duration of non-converged trades		74	76	na	74	76	na
Panel B: Performance statistics at the trading cycle level							
Mean payoff per cycle		5.02	5.95	0.29	4.99	4.99	0.50
Median payoff per cycle		4.09	1.68	na	4.09	1.64	na
Mean convergence ratio		61.24%	45.06%	0.00	61.24%	45.06%	0.00
Median convergence ratio		62.17%	44.13%	na	62.17%	44.13%	na
Mean payoff per converged trade		9.93%	19.47%	0.00	9.96%	20.33%	0.00
Median payoff per converged trade		9.02%	17.84%	na	9.02%	17.97%	na
Mean payoff per non-converged trade		-7.27%	-11.59%	0.00	-7.36%	-12.66%	0.00
Median payoff per non-converged trade		-6.72%	-10.30%	na	-6.72%	-10.54%	na
Mean number of unique traded pairs per cycle		101	103	0.41	101	103	0.41
Median number of unique traded pairs per cycle		77	68	na	77	68	na
Mean number of trades		159	132	0.07	159	132	0.07
Median number of trades		116	81	na	116	81	na
Mean total payoff on converged trades		8.88	16.04	0.01	8.89	16.58	0.01
Median total payoff on converged trades		6.77	6.58	na	6.77	6.74	na

Table 8 – The portfolio returns of a pairs trading strategy using cointegration

Each month over period January 1980 to December 2013, cointegrated pairs are identified amongst stocks of the same SIC code, with a Johansen cointegration test (Johansen, 1988) performed on the price level with 1 lag term. These pairs are then subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever the demeaned cointegration spread exceeds 2 standard deviations of historical spreads, provided the resulting net dollar exposure does not exceed \$0.1 with the short position scaled at \$1. The trade is unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. Panel A report risk adjustment results using five common risk factors. *Mkt-Rf, HML* and *SML* are respectively the market, value/growth and size factors from Fama and French (1997). *Momentum* is the 6-month momentum factor (Jegadeesh and Titman, 1993). *Monthly Reversal* is Jegadeesh's (1990) short – term reversal factor. Panel B reports risk adjustment results with an idiosyncratic reversal factor added. The description of this factor and its statistics are provided in Table 4. Due to the spikes in February-March 2000, a winsorization at [1 99] is applied to the portfolio return time series.

Panel A: Risk-adjustment of pairs trading returns using existing risk factors

	Raw return	Alpha	Mkt-Rf	HML	SML	Momentum	Monthly Reversal	R-squared
Full sample	0.0060	0.0065	-0.0200	-0.0412	0.0998	-0.1274	0.0487	0.27
	6.4665	6.81	-1.00	-2.05	3.13	-7.47	2.02	
Jan1980-Dec1995	0.0072	0.0080	-0.0042	0.0189	0.0132	-0.0884	-0.0078	0.12
	8.77	9.95	-0.25	0.73	0.49	-5.27	-0.31	
Jan1996-Dec2013	0.0048	0.0054	-0.0717	-0.0352	0.1419	-0.1579	0.0775	0.37
	3.08	3.87	-2.07	-1.14	3.78	-7.94	2.85	

Panel B: Risk-adjustment of pairs trading returns with the idiosyncratic reversal factor added

	Alpha	Mkt-Rf	HML	SML	Momentum	Monthly Reversal	Idiosyncratic Reversal	R-squared
Full sample	0.0044	-0.0181	-0.0142	0.0849	-0.1203	0.0025	0.2497	0.31
	4.36	-0.94	-0.65	2.71	-6.84	0.09	4.12	
Jan1980-Dec1995	0.0053	-0.0030	0.0275	0.0073	-0.0797	-0.0430	0.2149	0.19
	5.70	-0.21	1.15	0.28	-5.07	-1.70	3.60	
Jan1996-Dec2013	0.0043	-0.0674	-0.0102	0.1279	-0.1508	0.0388	0.2044	0.39
	3.20	-2.08	-0.30	3.36	-7.60	1.12	2.12	
Change in loading *x* factor mean	-0.0010	-0.0008	-0.0001	0.0005	-0.0010	0.0009	-0.0016	
Contribution based on								
Approximated change in raw return	33%	26%	4%	-17%	32%	-29%	52%	
Actual change in raw return	41%	32%	5%	-22%	40%	-37%	66%	

Table 9 – Performance statistics of a triplets trading strategy based on cointegration

Each month over period January 1980 to December 2013, cointegrated baskets of 3 stocks are identified amongst stocks of the same SIC code, with a Johansen cointegration test (Johansen, 1988) performed on the price level with 1 lag term. These baskets are then subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a basket whenever the demeaned cointegration spread exceeds 2 standard deviations of historical spreads, provided the resulting net dollar exposure does not exceed $0.1 with the short position scaled at $1. The trade is unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. This table reports the return distribution and composition of converged and non-converged pairs at the individual trade (Panel A) and trading cycle (Panel B) levels. The last three columns represent the set of results with payoffs winsorized at [1 99] to minimize the effect of the spikes in February-March 2010. *,**,*** indicate statistical significance at 10%,5%, and 1% levels, respectively.

		Raw data			Payoff winsorized at [1 99]		
		1980-1995	1996-2013	p-value	1980-1996	1996-2014	p-value
Panel A: Performance statistics at the individual trade level							
Number of trades		1,220,932	1,460,407	na	1,220,932	1,460,407	na
Mean number of trades per basket		1.68	1.32	na	1.68	1.32	na
Payoff per trades	Mean	3.01%	3.19%	0.00	3.01%	4.10%	0.00
	Median	4.81%	10.22%	na	4.81%	10.22%	na
	Standard deviation	10.94%	43.66%	na	10.76%	34.23%	na
	Min	-353.93%	-2612.76%	na	-116.00%	-116.00%	na
	Max	173.20%	355.85%	na	66.61%	66.61%	na
Convergence ratio		67.33%	54.19%	-	67.33%	54.19%	-
Mean payoff on converged trades		7.39%	23.77%	na	7.38%	23.23%	na
Median payoff on converged trades		5.92%	20.08%	na	5.92%	20.08%	na
Mean payoff on non-converged trades		-6.03%	-21.14%	-	-5.99%	-18.52%	-
Median payoff on non-converged trades		-3.43%	-10.26%	na	-3.43%	-10.26%	na
Mean duration of converged trades		23	26	0.00	23	26	0.00
Median duration of converged trades		15	17	na	15	17	na
Mean duration of non-converged trades		74	76	0.00	74	76	0.00
Median duration of non-converged trades		73	77	na	73	77	na
Panel B: Performance statistics at the trading cycle level							
Mean payoff per cycle		196.21	218.90	0.43	196.67	281.19	0.25
Median payoff per cycle		121.61	43.31	na	121.58	41.26	na
Mean convergence ratio		64.32%	48.19%	0.00	64.32%	48.19%	0.00
Median convergence ratio		64.93%	49.15%	na	64.93%	49.15%	na
Mean payoff per converged trade		8.18%	17.53%	0.00	8.17%	17.08%	0.00
Median payoff per converged trade		7.49%	15.25%	na	7.49%	15.25%	na
Mean payoff per non-converged trade		-6.34%	-13.39%	0.00	-6.29%	-12.25%	0.00
Median payoff per non-converged trade		-5.51%	-9.50%	na	-5.51%	-9.39%	na
Mean number of unique traded baskets per cycle		3892	5213	0.06	3892	5213	0.06
Median number of unique traded baskets per cycle		2773	2590	na	2773	2590	na
Mean number of trades		6529	6856	0.60	6529	6856	0.60
Median number of trades		4433	3294	na	4433	3294	na
Mean total payoff on converged trades		324.71	883.02	0.01	324.51	862.90	0.01
Median total payoff on converged trades		216.05	235.13	na	216.05	232.34	na

Table 10 – The portfolio returns of a triplets trading strategy using cointegration

Each month over period January 1980 to December 2013, cointegrated baskets of three stocks are identified amongst stocks of the same SIC code, with a Johansen cointegration test (Johansen, 1988) performed on the price level with 1 lag term. These triplets are then subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever the demeaned cointegration spread exceeds 2 standard deviations of historical spreads, provided the resulting net dollar exposure does not exceed $0.1 with the short position scaled at $1. The trade is unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. Panel A report risk adjustment results using five common risk factors. *Mkt-Rf, HML* and *SML* are respectively the market, value/growth and size factors from Fama and French (1997). *Momentum* is the 6-month momentum factor (Jegadeesh and Titman, 1993). *Monthly Reversal* is Jegadeesh's (1990) short –term reversal factor. Panel B reports risk adjustment results with an idiosyncratic reversal factor added. The description of this factor and its statistics are provided in Table 4. Due to the spikes in February-March 2000, a winsorization at [1 99] is applied to the portfolio return time series.

Panel A: Risk-adjustment of baskets trading returns using existing risk factors

	Raw return	Alpha	Mkt-Rf	HML	SML	Momentum	Monthly Reversal	R-squared
Full sample	0.0052	0.0055	-0.0252	-0.0534	0.1074	-0.0974	0.0739	0.26
	5.7484	6.19	-1.41	-2.11	2.90	-5.03	2.80	
Jan1980-Dec1995	0.0071	0.0075	-0.0079	0.0047	0.0202	-0.0488	0.0138	0.07
	8.88	9.14	-0.52	0.22	0.84	-2.62	0.55	
Jan1996-Dec2013	0.0035	0.0040	-0.0826	-0.0432	0.1496	-0.1331	0.1059	0.37
	2.33	3.32	-2.60	-1.27	3.26	-5.84	3.81	

Panel B: Risk-adjustment of baskets trading returns with the idiosyncratic reversal factor added

	Alpha	Mkt-Rf	HML	SML	Momentum	Monthly Reversal	Idiosyncratic Reversal	R-squared
Full sample	0.0034	-0.0233	-0.0257	0.0921	-0.0901	0.0266	0.2556	0.30
	3.58	-1.36	-1.05	2.60	-4.75	0.97	4.67	
Jan1980-Dec1995	0.0056	-0.0071	0.0108	0.0159	-0.0426	-0.0115	0.1546	0.11
	7.72	-0.52	0.54	0.69	-2.39	-0.45	3.31	
Jan1996-Dec2013	0.0028	-0.0779	-0.0154	0.1342	-0.1252	0.0630	0.2265	0.39
	2.40	-2.65	-0.41	2.90	-5.66	1.93	2.87	
Change in loading *x* factor mean	-0.0028	-0.0009	-0.0001	0.0005	-0.0009	0.0006	-0.0007	
Contribution based on								
Approximated change in raw return	66%	21%	2%	-12%	22%	-15%	17%	
Actual change in raw return	77%	24%	2%	-14%	25%	-18%	20%	

Figure 1 – Monthly portfolio returns to a distance-based pairs trading strategy over 1980-2013

This figure plots monthly excess returns to a distanced-based pairs trading strategy. Each month over period January 1980 to December 2013, pairs are matched amongst stocks of the same SIC code and by minimizing the sum of squared differences in total return indices scaled to $1 at the beginning of the 1-year training period. 100 pairs are subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever its absolute spread exceeds 2 standard deviations of historical spreads, with the position unwound when the spread crosses zero or at the end of the trading period, whichever occurs first.

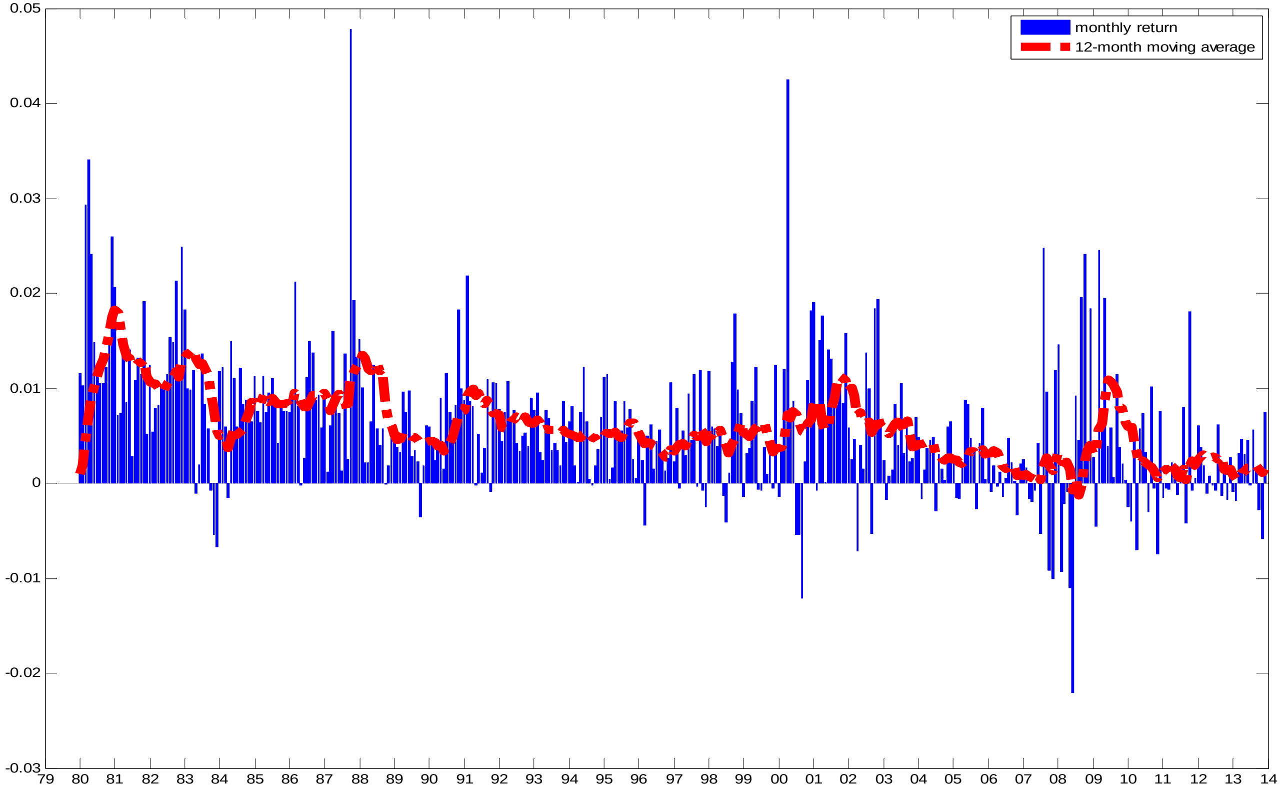

Figure 2 – Monthly returns to an idiosyncratic return reversal factor

This figure plots monthly excess returns on a systematic factor that captures return reversals from idiosyncratic shocks that are unexplained by daily returns of the market and industries. Over the period from January 1980 to December 2013, each day, the residual from model $r_{i,t} = \alpha_i + \beta_0 R_{M,t} + \delta_0 R_{I,t} + \omega_{i,t}$ is estimated for each stock using its prior 252 daily returns. R_M denotes the corresponding market return and R_I denotes the stock's industry return, based on Fama and French's (1997) 48 industry classification. A stock is classified as an *idiosyncratic winner* (*idiosyncratic loser*) if its residual return on the day is greater (less) than 2 standard deviations of the stock's historical residuals. A portfolio is formed that is long idiosyncratic losers and short idiosyncratic winners and is held for 21 days. The strategy is rolled over daily, giving rise to 21 overlapping portfolios. Daily idiosyncratic reversal factor returns are computed by averaging the returns on the same day that are produced by the 21 overlapping portfolios (i.e. the method is analogous to that in Jegadeesh and Titman, 1993). The idiosyncratic return reversal factor is the monthly return time series which aggregates from the daily returns.

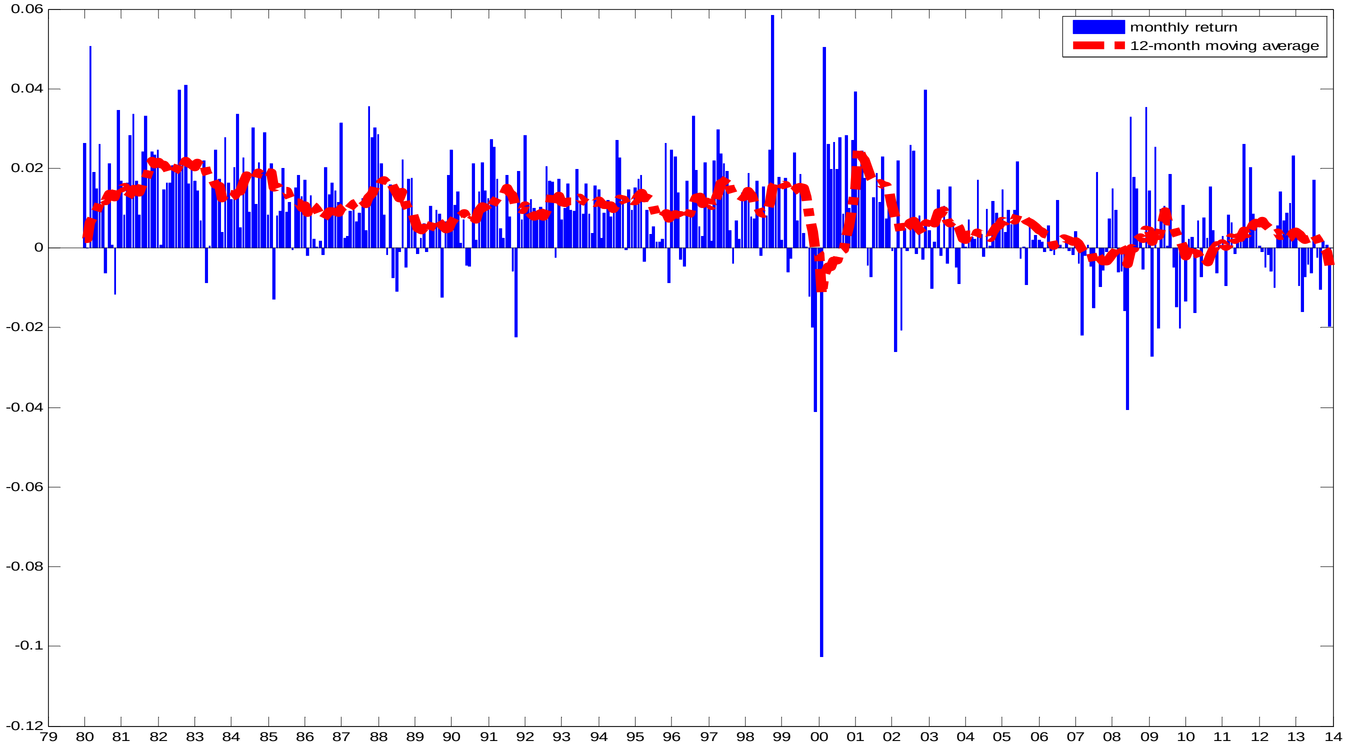

Figure 3 – Monthly portfolio returns to a cointegration-based pairs trading strategy over 1980-2013

Each month over period January 1980 to December 2013, cointegrated pairs are identified amongst stocks of the same SIC code, with a Johansen cointegration test (Johansen,1988) performed on the price level with 1 lag term. These pairs are then subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever the demeaned cointegration spread exceeds 2 standard deviations of historical spreads, provided the resulting net dollar exposure does not exceed $0.1 with the short position scaled at $1. The trade is unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. This figure plots the monthly portfolio returns to the strategy over the sample period.

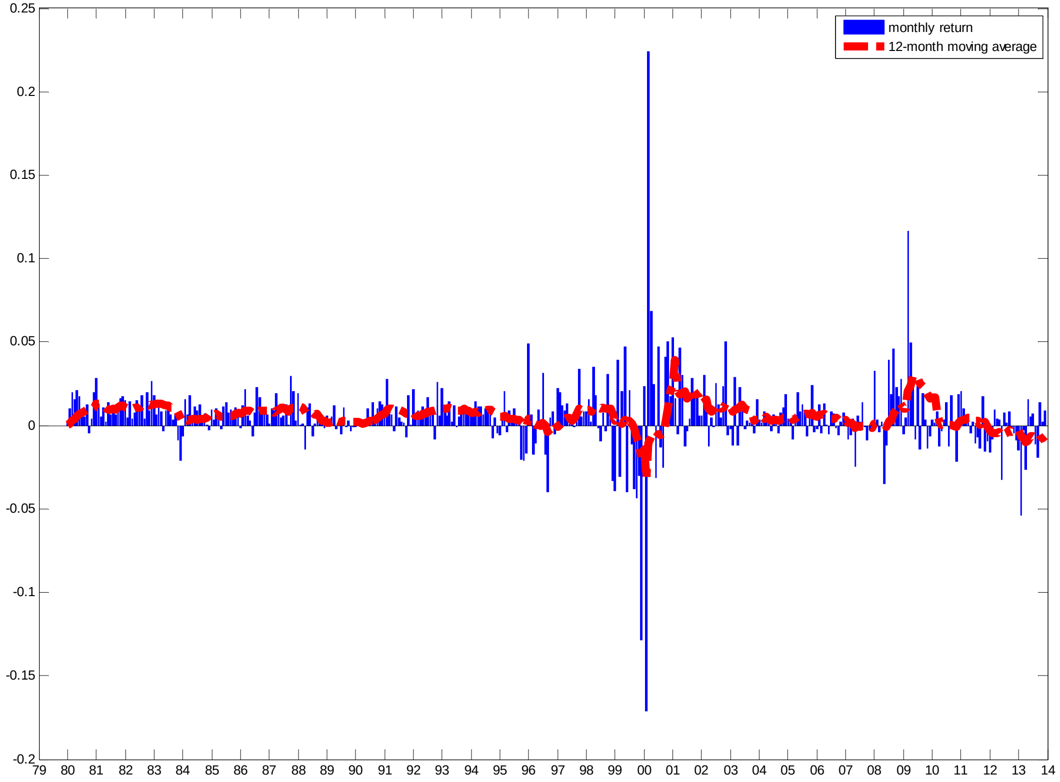

Figure 4 – Monthly portfolio returns to a cointegration-based triplet trading strategy over 1980-2013

Each month over period January 1980 to December 2013, cointegrated baskets of three stocks are identified amongst stocks of the same SIC code, with a Johansen cointegration test (Johansen,1988) performed on the price level with 1 lag term. These baskets are then subjected to a trading rule over the subsequent 6 months, calling for a long-short position in a pair whenever the demeaned cointegration spread exceeds 2 standard deviations of historical spreads, provided the resulting net dollar exposure does not exceed $0.1 with the short position scaled at $1. The trade is unwound when the spread crosses zero or at the end of the trading period, whichever occurs first. This figure plots the monthly portfolio returns to the strategy over the sample period.