Financial Analysts Journal
Volume 66 • Number 4


# Does Simple Pairs Trading Still Work?

Binh Do and Robert Faff

*Despite confirming the continuing downward trend in profitability of pairs trading, this study found that the strategy performs strongly during periods of prolonged turbulence, including the recent global financial crisis. Moreover, alternative algorithms combined with other measures enhance trading profits considerably, by 22 bps a month for bank stocks.*

A decade has passed since the publication of the original research by Gatev, Goetzmann, and Rouwenhorst (GGR 1999) in which they documented economically and statistically significant profits from using a very simple pairs trading rule. The algorithm matches pairs by simply minimizing the historical price spreads over 12 months and trades them in the following 6 months by using two historical standard deviations as the opening trigger. The researchers subsequently extended their analysis to a more recent sample and confirmed steady profitability (GGR 2006). With the overall analysis divided into two time-lagged stages, these studies (GGR 1999, 2006) offer a robust solution to the data-snooping problem inherent in this type of backtesting research. Remarkably, despite the simplicity of the trading rule and the attention that GGR have drawn to the strategy over the past several years, material excess returns on the order of 1 percent a month have apparently persisted, albeit at a declining rate.

Very few trading rules have stood the test of time or independent scrutiny; price momentum documented by Jegadeesh and Titman (1993) is probably the best-known exception. Although Papadakis and Wysocki (2007) and Andrade, di Pietro, and Seasholes (2005) tested the GGR trading rules, they did so either on a subset of the U.S. equity market (the former study) or on a minor international market (the latter). Engelberg, Gao, and Jagannathan (2009) used a modified GGR algorithm to examine the impact of liquidity and idiosyncratic news on pairs trading from 1992 to 2006.

*Binh Do is a lecturer in finance at Monash University, Melbourne, Australia. Robert Faff is professor of finance at the University of Queensland, Brisbane, Australia.*

## Constructing the Pairs Trading Algorithm

In accordance with GGR's pairs trading framework, we formed pairs by using 12 months of daily price data and then traded them in the subsequent 6-month interval. In particular, for each stock in the CRSP universe, we formed pairs by finding a "partner" that minimized the sum of squared differences (SSD) in the normalized prices of the two stocks, where the normalized prices were the total return index, inclusive of dividends, both scaled to start at \$1. Once we had identified pairs, we again scaled the price series to \$1 at the beginning of the trading period and took a \$1 long–short position in a pair whenever its normalized price difference, or spread, diverged by more than two standard deviations from the historical spread calculated over the formation period. We unwound the position at the next crossing of the prices, on the day a stock was delisted, or on the last day of the trading period, whichever occurred first. After a pair had completed a round-trip trade, we subjected it to the same trading algorithm again. We repeated the 12 × 6 implementation cycle every month, effectively mimicking a hedge fund of six managers whose implementation cycles were staggered by one month.

We focused on the portfolio of the top 20 pairs in terms of lowest SSDs, which was the main object of GGR's studies. Like GGR, we looked at two excess return measures: the *return on committed capital* (the total mark-to-market payoff for all pairs divided by the number of pairs in the portfolio) and the *return on employed capital* (the total payoff divided by the number of pairs that actually traded). We defined monthly excess returns to the portfolio as the equally weighted average of returns by the six "managers." Finally, we monitored pairs trading under two alternative trading rules: (1) Opening and closing happen on the same day of the trigger and (2) all trades have a one-day delay. Using the latter rule, suggested by GGR, we sought to alleviate concerns regarding the potential upward bias in the reported returns induced by bid–ask bounce.[1]

Using CRSP daily data for July 1962–June 2009, we restricted the download to ordinary shares with share codes of 10 and 11. Similar to GGR, we filtered out stocks that had either no trading data or invalid return data for at least one day. In recent years, several companies have issued more than one equity security, differentiated only by the share class (e.g., Berkshire Hathaway). In checking the data, we noted that such securities represented the most frequently repeated pairs. To the extent of any restrictions on the ability to arbitrage among related securities, we excluded such pairs when forming the portfolio. Overall, our constructed pairs' return distributions are consistent with those reported in both GGR studies, which buttresses the reliability of our results.

## Performance of Pairs Trading in Recent Years

**Table 1** reports the monthly excess return distributions for the portfolio of top 20 pairs for July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009 (the post-GGR sample), as well as for the full sample. Under the delayed trading rule, the mean excess return on employed capital for 1989–2002 was 0.37 percent for the top 20 pairs, in line with GGR's result of 0.38 percent (GGR 2006, Table 8). Although still statistically significant, that number represents a massive 57 percent decline from the comparable 1962–88 number. More importantly, this declining trend continued, with the mean return shrinking to a mere 0.24 percent for January 2003–June 2009, although the *t*-statistic suggests that this return remains statistically significant. To alleviate concerns that outliers might have biased the mean, we also report the median, the Sharpe ratio, and the percentage of months with negative excess returns. Once again, these statistics show a continuing deterioration in the performance of the portfolio of top 20 pairs. The same trend prevails under the "no delay" trading rule.

**Figure 1** depicts the time-series performance of the portfolio of top 20 pairs over the full 1962–2009 sample. The strategy seemed to perform very well during the 1970s and 1980s, including the 1987 crash. Beginning in the 1990s, however, negative return months appeared frequently, which continued into the first decade of the 21st century. The downward trend is most visible in the 12-month moving-average plot, which shows pairs trading profits peaking in the mid-1970s and then declining since the mid-1990s to settle around zero.[2]

One notable exception to this declining trend is a remarkable pickup in profitability during two recent major bear markets: January 2000–December 2002 and July 2007–June 2009. In unreported subperiod results for the first bear market, the mean excess return was 0.92 percent a month, well above the preceding 1989–99 period's 0.22 percent and the subsequent 2003–07 sample's 0.02 percent. Similarly, the median was 0.78 percent, much higher than the medians for the previous and subsequent periods (0.23 percent and 0.00 percent, respectively). This remarkable performance was repeated during the global financial crisis of 2007 to mid-2009, with the portfolio reporting a mean excess return of 0.71 percent and a median of 0.62 percent, decidedly superior to the performance of the preceding period (2003–2007). Associated with this higher level of returns, however, is an elevated volatility in monthly performance, as seen in

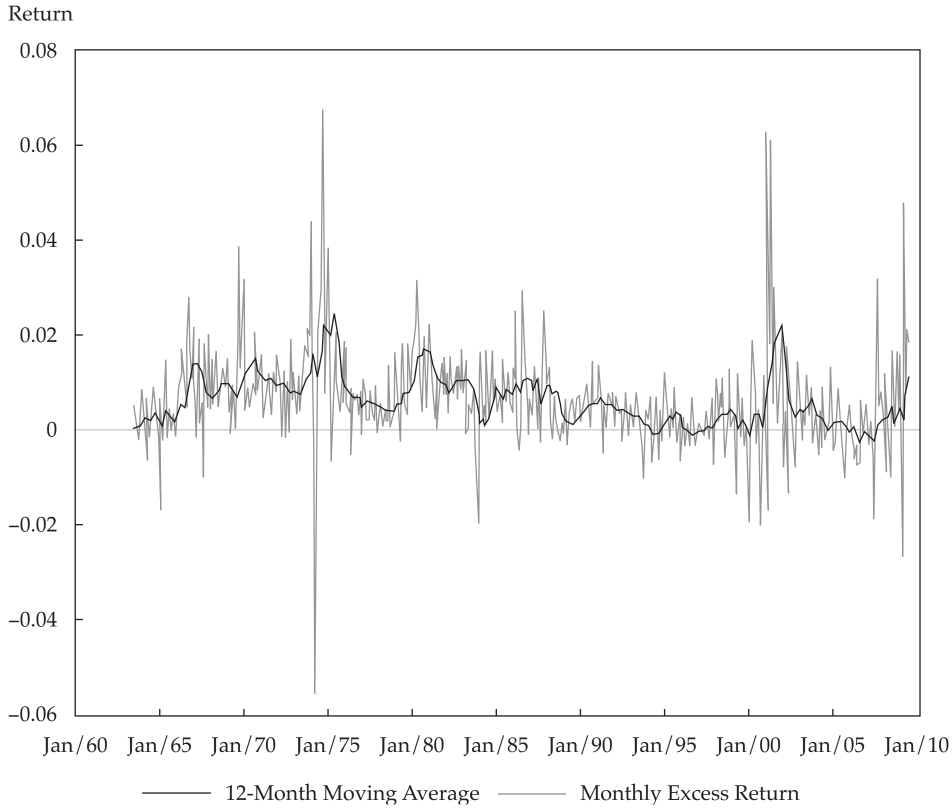
**Table 1. Excess Returns for Portfolio of Top 20 Pairs by Subperiod, July 1962–June 2009**

| | Trade on Day of Trigger | | | | Trade One Day after Trigger | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962–1988 | 1989–2002 | 2003–2009 | Full Sample | 1962–1988 | 1989–2002 | 2003–2009 | Full Sample |
| Mean return (employed) | 0.0124 | 0.0056 | 0.0033 | 0.0090 | 0.0086 | 0.0037 | 0.0024 | 0.0062 |
| *t*-Statistic | 9.77** | 3.21** | 2.98** | 9.12** | 8.51** | 2.99** | 2.73** | 8.31 |
| Median return (employed) | 0.0110 | 0.0040 | 0.0008 | 0.0077 | 0.0077 | 0.0029 | 0.0004 | 0.0053 |
| Sharpe ratio | 1.04 | 0.43 | 0.32 | 0.72 | 0.85 | 0.36 | 0.24 | 0.59 |
| Observations with return < 0 | 6% | 25% | 44% | 17% | 11% | 32% | 47% | 22% |
| Mean return (committed) | 0.0122 | 0.0054 | 0.0033 | 0.0089 | 0.0085 | 0.0036 | 0.0024 | 0.0061 |

*Notes*: This table presents distribution statistics for monthly excess returns, over different sample periods, for the portfolio of top 20 pairs with the lowest SSDs. "Employed" refers to the method of computing returns on employed capital; "committed" refers to the method of computing returns on committed capital. The *t*-statistics are computed by using Newey–West standard errors with six lags.

**Significant at the 5 percent level.

**Figure 1. Monthly Excess Returns of the Portfolio of Top 20 Pairs, July 1962–June 2009**




*Notes*: This figure plots the realized monthly excess return time series for the portfolio of the top 20 pairs with the lowest SSDs, as well as the 12-month moving average of the monthly returns. The pairs are opened at the closing price one day after the divergence and are closed at the closing price one day after the convergence.

Figure 1. The Sharpe ratios for these two bear market periods (both at 0.47), however, remain superior to those of the adjacent periods.

## The Declining Trend in Pairs Trading Profitability

We attempted to ascertain what drove the solid performance of pairs trading over the two turbulent periods, as well as the general decline in pairs trading profitability over time.

**Arbitrage Risks vs. Market Efficiency.** Pairs trading is necessarily based on the premise that irrational trading in markets that are less than fully efficient will occasionally cause price divergence among similar securities, which is then exploited by competing arbitrageurs. We refer to this channel, which affects pairs trading profitability, as the "market efficiency effect." This effect also includes the impact of competition among arbitrageurs: Increased competition will see many arbitrageurs exploiting similar opportunities, to the point where even the smallest opportunities might be arbitraged away quickly, leading to small profits. But those arbitrageurs will face several arbitrage risks, such as fundamental risk, noise-trader risk, and synchronization risk. Fundamental risk refers to the possibility of an unexpected disruption in the relative relationship between paired securities (for a discussion of this type of risk—e.g., in the context of a parent–subsidiary arbitrage—see Mitchell, Pulvino, and Stafford 2002). Noise-trader risk arises when irrational trading from noise traders causes further divergence and deters arbitrage activities (see De Long, Shleifer, Summers, and Waldmann 1990). Synchronization risk (see Abreu and Brunnermeier 2002) refers to the uncertainty about when other arbitrageurs will exploit a common mispricing. These risks all work to prevent or delay arbitrage or to inflict losses on the arbitrageurs. We refer to the collective impact of these risks as the "arbitrage risk effect." Thus, pairs trading performance is the combined result of the efficiency effect and the arbitrage risk effect. Which one of these effects dominates depends on whether the diminishing profitability has been brought about more by reduced returns per profitable pair or more by an increased frequency and magnitude of loss-making, nonconvergent trades.

To evaluate the impact of market efficiency and arbitrage risk on pairs trading profits, we started by examining the cross-sectional characteristics of traded pairs across four mutually exclusive groups: (1) pairs that never trade throughout the trading period, (2) pairs that do open but never converge in time during the period, (3) pairs that have one round-trip trade and possibly another nonconvergent trade, and (4) pairs that have multiple round-trip trades and possibly a final nonconvergent trade. Variation over time in the composition of the pairs portfolio, as well as per pair returns and other trading statistics under each of these four pair groups, provides important insights into what drives pairs trading performance.

**Table 2** reports trading statistics for the top 20 pairs pooled across all 547 trading cycles, for a total of 10,940 pairs. Panel A reports results for the three subperiods that were used for the trend analysis, and Panel B partitions the post-1988 sample into four subperiods to focus attention on the two bear markets that were found to produce strong pairs trading performance (i.e., 2000–2002 and 2007–2009). A striking observation from Panel A is that the proportion of loss-making, nonconvergent pairs (Group 2) jumped from 26 percent (1962–1988) to 39 percent (1989–2002)—a statistically significant increase—and to 40 percent for 2003–2009, apparently at the expense of the highly profitable, convergent Group 4 (which dropped from 42 percent for 1962–1988 to 24 percent for both 1989–2002 and 2003–2009), whereas the Group 3 proportion remained relatively steady. In addition, losses from those divergent pairs increased over time (also statistically significantly), as shown by the mean per pair return, whereas the mean returns for the other trading groups stayed relatively constant over the three periods.

These findings suggest that the main driver of the general declining trend in pairs trading profitability is increased rates and magnitude of divergence caused by worsening arbitrage risks and not an increase in market efficiency. This evidence also serves to downplay the role of the hedge fund industry in competing away opportunities, as suggested by GGR (2006). Moreover, other trading statistics do not support the competition effect argument. Specifically, the "diverge-minus-traded-spread" statistic, which measures how much mispricing is corrected the day after the divergence, decreased post-1988 for both Groups 3 and 4 (Table 2, Panel A). The average round-trip lengths, however, increased over time: For Group 3 pairs, 34 days, on average, were needed to correct a two-standard-deviation mispricing for 1962–1988. This number increased to 36.75 days for 1989–2002 and to 38.35 days for 2003–2009. For Group 4 pairs, the corresponding statistics are 19.25, 20.94, and 23.11 days. These movements are statistically significant except for Group 3's change between 1989–2002 and 2003–2009. Obviously, the increased competition among hedge funds in the last couple of decades has not translated into any quickening in the arbitrage process. Although seemingly counterintuitive, this "delayed arbitrage" can be viewed as evidence of "learning" by arbitrageurs: Increased aversion to arbitrage risks, especially synchronization risk (Abreu and Brunnermeier 2002), prompts arbitrageurs to take time to act on any mispricings. Alternatively, Kondor (2009) showed in his model that not immediately trading away mispricings (simply because they may improve tomorrow) may be in the arbitrageur's interest.

Regarding pairs trading in turbulent markets, Panel B reveals that the strong performance during the 2000–02 bear market was brought about by increased profitability among convergent pairs because the proportion of divergent versus convergent pairs was relatively stable—the mean returns were significantly higher than those of the adjacent periods (4.84 percent a month per pair compared with 1.51 percent for 1989–1999 and 1.69 percent for 2003–2007). But the resurgence of profitability in the latest crisis (2007–2009) was driven by a significant jump in the fraction of multiple round-trip pairs (Group 4; to 37 percent from 18 percent for 2003–2007), against a significant decline in the divergent pairs (from 44 percent to 32 percent). This highly desirable Group 4 also saw a statistically significant higher per pair return, which implies a more frequent divergence/convergence within each pair. Such remarkable convergence rates suggest that despite the havoc wrought on the hedge fund community by the sudden evaporation of liquidity and the forced unwinding of leveraged positions, single-strategy, equity-based convergence traders appear to have escaped the liquidity shock (unlike their multistrategy cousins) and successfully synchronized the arbitrage of mispricings that seem abundant in turbulent times.[3]

With respect to the general declining trend in pairs trading profits, which we found to be predominantly driven by increased divergence rates (Table 2, Panel A), a question remains: What is the

**Table 2. Trading Statistics by Pair Group for Top 20 Pairs with Trades One Day after Trigger, July 1962–June 2009**

| | Group 1 Nontraded Pairs | | | | Group 2 Nonconvergent Pairs | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962–1988 | 1989–2002 | 2003–2009 | 1962–2009 | 1962–1988 | 1989–2002 | 2003–2009 | 1962–2009 |
| *A. Three subperiods* | | | | | | | | |
| Proportion | 2% | 3% | 3% | 2% | 26% | 39% | 40% | 32% |
| $z$-Statistic[a] | na | na | na | na | na | 7.04** | 0.55 | na |
| Avg. monthly return per pair | na | na | na | na | –0.0074 | –0.0094 | –0.0108 | –0.0088 |
| $t$-Statistic[b] | na | na | na | na | na | –3.89** | –1.91* | na |
| St. dev.: monthly return per pair | na | na | na | na | 0.0128 | 0.0150 | 0.0151 | 0.0141 |
| Avg. SSD | 0.20 | 0.17 | 0.14 | 0.18 | 0.16 | 0.15 | 0.14 | 0.15 |
| Avg. zero crossings: formation | 43.20 | 33.25 | 30.12 | 37.05 | 44.29 | 38.82 | 34.20 | 40.46 |
| Avg. zero crossings: trading | 21.81 | 13.19 | 14.05 | 17.05 | 5.42 | 5.05 | 3.96 | 5.02 |
| Avg. round-trip length | na | na | na | na | na | na | na | na |
| Diverge minus traded spread | na | na | na | na | 0.0005 | –0.0011 | –0.0015 | –0.0004 |
| | 1989–1999 | 2000–2002 | 2003–2007 | 2007–2009 | 1989–1999 | 2000–2002 | 2003–2007 | 2007–2009 |
| *B. Four subperiods* | | | | | | | | |
| Proportion | 3% | 6% | 3% | 1% | 39% | 40% | 44% | 32% |
| $z$-Statistic[a] | na | na | na | na | 6.33** | 0.36 | 1.09 | –2.76** |
| Avg. monthly return per pair | na | na | na | na | –0.0078 | –0.0151 | –0.0092 | –0.0159 |
| $t$-Statistic[b] | na | na | na | na | –0.83 | –7.47** | 4.47** | –4.89** |
| St. dev.: monthly return per pair | na | na | na | na | 0.0107 | 0.0240 | 0.0129 | 0.0196 |
| Avg. SSD | 0.12 | 0.25 | 0.14 | 0.13 | 0.11 | 0.28 | 0.13 | 0.14 |
| Avg. zero crossings: formation | 34.33 | 31.49 | 30.43 | 28.33 | 40.06 | 34.38 | 34.65 | 32.84 |
| Avg. zero crossings: trading | 16.40 | 7.95 | 15.46 | 5.83 | 5.41 | 3.77 | 4.16 | 3.35 |
| Avg. round-trip length | na | na | na | na | na | na | na | na |
| Diverge minus traded spread | na | na | na | na | 0.0005 | –0.0069 | –0.0008 | –0.0034 |

| | Group 3 Single Round-Trip Pairs | | | | Group 4 Multiple Round-Trip Pairs | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962–1988 | 1989–2002 | 2003–2009 | 1962–2009 | 1962–1988 | 1989–2002 | 2003–2009 | 1962–2009 |
| *A. Three subperiods* | | | | | | | | |
| Proportion | 30% | 34% | 33% | 32% | 42% | 24% | 24% | 34% |
| $z$-Statistic[a] | na | na | na | na | na | na | na | na |
| Avg. monthly return per pair | 0.0053 | 0.0054 | 0.0047 | 0.0053 | 0.0216 | 0.0222 | 0.0208 | 0.0216 |
| $t$-Statistic[b] | na | 0.27 | –1.12 | na | na | 0.61 | –0.67 | na |
| St. dev.: monthly return per pair | 0.0104 | 0.0119 | 0.0131 | 0.0113 | 0.0161 | 0.0383 | 0.0182 | 0.0231 |
| Avg. SSD | 0.16 | 0.15 | 0.14 | 0.15 | 0.14 | 0.15 | 0.14 | 0.14 |
| Avg. zero crossings: formation | 45.05 | 39.66 | 34.78 | 41.75 | 46.93 | 41.68 | 37.25 | 44.80 |
| Avg. zero crossings: trading | 10.68 | 9.46 | 7.57 | 9.82 | 14.86 | 12.61 | 10.96 | 13.97 |
| Avg. round-trip length | 34.00 | 36.75 | 38.35 | 35.55 | 19.25 | 20.94 | 23.11 | 20.01 |
| Diverge minus traded spread | 0.0051 | 0.0035 | 0.0002 | 0.0038 | 0.0077 | 0.0064 | 0.0042 | 0.0071 |
| | 1989–1999 | 2000–2002 | 2003–2007 | 2007–2009 | 1989–1999 | 2000–2002 | 2003–2007 | 2007–2009 |
| *B. Four subperiods* | | | | | | | | |
| Proportion | 34% | 30% | 35% | 29% | 24% | 24% | 18% | 37% |
| $z$-Statistic[a] | na | na | na | na | na | na | na | na |
| Avg. monthly return per pair | 0.0046 | 0.0088 | 0.0060 | 0.0012 | 0.0151 | 0.0484 | 0.0169 | 0.0249 |
| $t$-Statistic[b] | –1.77* | 4.68** | –2.14** | –3.73** | –9.65** | 10.85** | –5.76** | 4.33** |
| St. dev.: monthly return per pair | 0.0080 | 0.0213 | 0.0107 | 0.0177 | 0.0110 | 0.0748 | 0.0097 | 0.0235 |
| Avg. SSD | 0.11 | 0.29 | 0.14 | 0.14 | 0.10 | 0.30 | 0.14 | 0.14 |
| Avg. zero crossings: formation | 40.55 | 35.94 | 34.84 | 34.61 | 42.48 | 38.75 | 37.97 | 36.47 |
| Avg. zero crossings: trading | 10.01 | 7.16 | 8.04 | 6.30 | 13.12 | 10.75 | 10.91 | 11.02 |
| Avg. round-trip length | 36.27 | 38.75 | 39.57 | 35.06 | 21.00 | 20.72 | 23.97 | 22.18 |
| Diverge minus traded spread | 0.0034 | 0.0038 | 0.0005 | –0.0005 | 0.0053 | 0.0105 | 0.0042 | 0.0042 |

na = not applicable.

*Notes*: This table presents trading statistics by pair group, over different sample periods, for the portfolio of top 20 pairs with the lowest SSDs. Group 1 pairs never trade during the trading period. Group 2 pairs open but never converge in time during the trading period. Group 3 pairs have one round-trip trade and possibly another nonconvergent trade. Group 4 pairs have multiple round-trip trades and possibly a final nonconvergent trade. Pairs are opened at the closing price one day after the divergence and closed at the closing price one day after the convergence.

[a]The $z$-statistic for the null that the proportion of nonconvergent pairs in the current period equals that of the previous period.
[b]The $t$-statistic for the null that the mean return per pair in the current period equals that of the previous period.

*Significant at the 10 percent level.
**Significant at the 5 percent level.

extent of the impact of increased competition among arbitrageurs, or more generally, what is the magnitude of the market efficiency effect? That the per pair returns for both Groups 3 and 4 were lower for 2003–2009 than for 1989–2002 suggests that the effect is nontrivial. Similarly, Panel B results show that either of the two effects can be positive. To further distinguish the risk effect from the efficiency effect and quantify the individual contribution of each effect to pairs trading profits, we used a simple attribution scheme. The change in the average six-month return per pair between periods $t$ and $t + 1$ is given by

$$R_{t+1} - R_t = \Delta R_{t+1} = \sum_{i=1}^{4} p_{i,t+1} r_{i,t+1} - \sum_{i=1}^{4} p_{i,t} r_{i,t}, \quad (1)$$

where $p_{i,t}$ and $r_{i,t}$ are the proportion of pairs traded in period $t$ that fall in pair group $i$ and the average monthly trading return per pair, respectively. Analogous to fund performance attribution analysis, this return increment can be rewritten as

$$\begin{aligned}\Delta R_{t+1} = &\sum_{i=1}^{4} \left( r_{i,t+1} - r_{i,t} \right) p_{i,t} \\ &+ \sum_{i=1}^{4} \left( p_{i,t+1} - p_{i,t} \right) r_{i,t+1},\end{aligned} \quad (2)$$

where the first term represents the effect from changes in average returns per pair group and the second term represents the effect from changes in pair composition. Because Group 2 is the manifestation of the arbitrage risk, we can isolate the arbitrage risk impact from the efficiency impact by capturing the aggregate change in the average per pair return caused by changes in both the proportion of Group 2 pairs and the loss per pair in that group. Thus, we can express Equation 2 as

$$\begin{aligned}\Delta R_{t+1} = &\left( r_{2,t+1} - r_{2,t} \right) p_{2,t} + \left( p_{2,t+1} - p_{2,t} \right) r_{2,t+1} \\ &+ \sum_{i=(1,3,4)} \left( r_{i,t+1} - r_{i,t} \right) p_{i,t} \\ &+ \sum_{i=(1,3,4)} \left( p_{i,t+1} - p_{i,t} \right) r_{i,t+1}.\end{aligned} \quad (3)$$

Because changes in the proportions of Groups 1, 3, and 4 may be partly induced by changes in the proportion of Group 2, we can further isolate the risk impact by assuming that arbitrage risk affects Groups 1, 3, and 4 evenly and using the following decomposition:

$$\begin{aligned}\Delta R_{t+1} = &\sum_{i=1}^{4} \left( p^{*}_{i,t+1} - p_{i,t} \right) r_{i,t+1} + \left( r_{2,t+1} - r_{2,t} \right) p_{2,t} \\ &+ \sum_{i=1}^{4} \left( p_{i,t+1} - p^{*}_{i,t+1} \right) r_{i,t+1} \\ &+ \sum_{i=(1,3,4)} \left( r_{i,t+1} - r_{i,t} \right) p_{i,t},\end{aligned} \quad (4)$$

where $p^*_2 = p_2$ and $p^*_1$, $p^*_3$, and $p^*_4$ are such that the relative proportions among the four pair groups stay the same between two periods. The first two terms capture the impact from changes in the rate and magnitude of divergence brought about by arbitrage risks, and the last two terms capture the impact from changes in market efficiency over two periods. We conducted this attribution analysis for July 1962–December 1988, January 1989–December 2002, and January 2003–June 2009 (**Table 3**, Panel A), as well as for July 1962–December 1988, January 1989–December 1999, January 2000–December 2002, January 2003–June 2007, and July 2007–June 2009 (Table 3, Panel B). Panel A shows that of the 51 bp decline between 1962–1988 and 1989–2002, 36 bps, or 71 percent, is attributable to worsening arbitrage risk, whereas the remaining 29 percent is attributable to improved market efficiency, part of which is possibly caused by increased competition among hedge funds. The mix of these effects is approximately 60 percent/40 percent for 2003–2009. Panel B shows negative risk effects and positive efficiency effects during the two bear markets (2000–2002 and 2007–2009). Thus, in these markets, pairs trading profits increased because the impact of a less efficient market outweighed an increase in arbitrage risk.[4]

**Table 3. Arbitrage Risk Effect and Market Efficiency Effect on Pairs Profitability, July 1962–June 2009**

| | Panel A | | | Panel B | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962–1988 | 1989–2002 | 2003–2009 | 1962–1988 | 1989–1999 | 2000–2002 | 2003–2007 | 2007–2009 |
| Monthly return per pair | 0.0086 | 0.0035 | 0.0021 | 0.0086 | 0.0022 | 0.0082 | 0.0011 | 0.0045 |
| Change between periods | | –0.0051 | –0.0014 | | –0.0064 | 0.0060 | –0.0072 | 0.0034 |
| Arbitrage risk effect | | –0.0036 | –0.0008 | | –0.0024 | –0.0033 | 0.0016 | –0.0001 |
| Market efficiency effect | | –0.0015 | –0.0006 | | –0.0040 | 0.0093 | –0.0088 | 0.0035 |

*Notes*: This table decomposes the change in monthly return per pair between subperiods into an arbitrage risk effect and a market efficiency effect by using Equation 4. The arbitrage risk effect captures the impact of changes in such arbitrage risks as fundamental risk, noise-trader risk, and synchronization risk. The market efficiency effect captures the impact of changes in the state of market efficiency induced by, for example, increased competition among arbitrageurs. Results are based on a trading scenario in which pairs are opened at the closing price one day after the divergence and closed at the closing price one day after the convergence.

**Pairs Trading and Earnings Events.** Papadakis and Wysocki (2007) found that 9–14 percent of pair openings are triggered by accounting events (i.e., earnings announcements and analyst forecasts). They also found that trades triggered by accounting events perform worse than nonevent trades (–0.12 percent to 0.07 percent, compared with 0.62 percent a month for "unconditional" trading) because of investor underreaction, which is well documented in the earnings literature. To the extent that earnings events are becoming more frequent, the decline in pairs trading profitability and the increase in divergence are conceivably related to this earnings news effect. We tested this conjecture by examining whether a trading algorithm that conditions on earnings events affects profitability. Specifically, we modified the GGR algorithm so that if the threshold is exceeded during the five-day window after an accounting event, no trading occurs for fear of further divergence. Following Papadakis and Wysocki (2007), we matched CRSP securities with the report dates of quarterly earnings in Compustat and with the analyst estimate dates in I/B/E/S (the latter are available post-1970s only).[5]

Our results (unreported) were very similar to those generated by the "unconditional" algorithm: The mean excess return was 61 bps (compared with 62 bps), and the proportion of negative return months dropped from 47 percent to 44 percent for 2003–2009 and remained the same for the other periods. Trading statistics show that the composition of pair types (in the spirit of the classification scheme discussed previously) was also largely unchanged. This result suggests that the earnings effect in Papadakis and Wysocki (2007) applies to too small a sample to affect the whole portfolio's performance.[6] It also implies that the increased divergence risk in pairs trading is not explained by an increase in the frequency of accounting events over the more recent period.[7]

## Alternative Pairs Trading Algorithms

The pairs trading strategy that we have discussed thus far matches pairs on the basis of one single metric, the SSD, or the historical closeness in pricing between securities. We also examined the cross-sectional characteristics of traded pairs to investigate the usefulness of this metric and to motivate experimenting with alternative algorithms that might improve profitability by minimizing the fraction of nonconvergent trades.

**Cross-Sectional Characteristics of GGR Pairs.** Table 2 reports several trading statistics that provide interesting insights with regard to the characteristics of "good pairs" versus "bad pairs" and how effective the GGR pair-matching scheme is in differentiating these two groups, *ex ante*. First, the SSD measure, the basis of the GGR pair-matching scheme, is uniform across all pair groups, especially the last three groups. The average SSDs for the traded pairs are all about 0.14–0.16 for 1962–1988, 0.15 for 1989–2002, and 0.14 for 2003–2009. Clearly, one cannot use such a metric to identify and exclude potentially divergent pairs in order to improve trading performance. Second, the number of zero crossings in the formation period does have some usefulness in predicting future convergence. Defined as the number of times the normalized spread crosses the value zero, this statistic measures the frequency of divergence and convergence between two securities, a desirable property for pairs trading. Intuitively, good candidates are those that not only track each other well (captured by the SSD metric) but also exhibit frequent deviations that are subsequently reversed under the force of arbitrage. For all subperiods in Panels A and B of Table 2, the average number of zero crossings in the formation period is largest for Group 4 pairs and diminishes as one moves back to Group 1 pairs. In other words, for any given period, securities whose prices closely track each other and most frequently cross each other in the past year appear most likely to go on to be profitable pairs in the subsequent six-month period. As expected, the number of zero crossings in the trading period, which is an *ex post* measure, is higher for Groups 1, 3, and 4 pairs than for the divergent Group 2 pairs. That the nontraded pairs (Group 1) have a higher number of zero crossings without trading means that the paired securities are too close a substitute such that the magnitude of deviation is too small to trigger trading.

**Pairs Trading by Industry Group.** GGR (1999, 2006) performed pairs trading within four Standard & Poor's major industry groups: Utilities, Financials, Transportation, and Industrials. Although they documented statistically significant profits across all four groups, they found greater profits in Utilities and Financials. This finding is entirely expected: Utility companies face rather stable demand, their products have low differentiation, and electricity suppliers are generally subject to some form of rate regulation; therefore, a great deal of homogeneity exists among Utilities. Financials, however, are sensitive to such common macroeconomic factors as interest rates and unemployment shocks; hence, their share prices are likely to move together. We tested whether this industry-based pattern held for the extended period—in particular, whether Utilities and Financials pairs remained profitable.

We assigned each of the securities in the CRSP files to the four major industry groups by using the first two digits of their SIC codes (49 for Utilities, 60–67 for Financials, 40–47 for Transportation, and 15–17 and 30–39 for Industrials). In an unreported analysis, we found that the profit decline is pervasive across the industry groups (the average excess returns for the full sample are 64 bps, 75 bps, 50 bps, and 46 bps, respectively). Clearly, restricting pair matching to either Utilities or Financials results in material profit improvement. Consistent with GGR (1999, 2006), Transportation and Industrials are inferior, with excess returns dropping to zero for 2003–2009.

Next, using previously gained insights, we tested whether incorporating the time-series dimension of the historical data in the form of the number of zero crossings can further improve pairs selection and performance. Specifically, we tested the following algorithm: Pairs were matched within the same industry groups, first on the basis of the SSD statistic. For each of the four major sectors, out of the top 50 pairs (in terms of lowest SSDs), the 20 pairs that had the highest number of zero crossings during the 12-month formation period were admitted to the portfolio. The original trading rule was then applied to this portfolio. Because the results for Transportation and Industrials are again considerably inferior to those for Utilities and Financials (consistent with previous findings), we report the results for Utilities and Financials only.

**Table 4** shows the excess return distribution, by subperiod, for the portfolio of the top 20 pairs for Utilities and Financials, which reveals an increase in the mean return of another 6 bps and 3 bps, to 70 bps and 78 bps, respectively. **Table 5** provides a breakdown of performance, by pair group, for Utilities pairs (Panel A) and Financials pairs (Panel B). For most subperiods, the Utilities portfolio has a poorer pair composition (i.e., lower proportions of Groups 3 and 4 pairs and a higher proportion of the nonconvergent Group 2 pairs) than do the Financials portfolio and the unconditional portfolio (Table 2). The nonconvergent group in the Utilities portfolio, however, has a much lower loss per pair than do the other portfolios. For example, for 2003–2009, Group 2 pairs in the Utilities portfolio returned, on average, –0.91 percent a month, compared with –1.59 percent for the Financials portfolio and –1.08 percent for the original portfolio (the last statistic reported in Table 2, Panel A). We found similar patterns for the other periods.

This lower divergence loss in the Utilities portfolio implies a greater degree of homogeneity within the major group, which could simply be a result of the particular industry classification scheme chosen. Indeed, using, for example, the Fama–French (1997) finer industry classification, in which companies are grouped into 48 industries on the basis of their SIC codes, Utilities as a major group remains an industry among the 48, whereas Financials is broken down into Banks, Real Estate, and Trading. To further test the impact of homogeneity on pairs trading, we examined an algorithm that matches pairs within Banks (SIC codes: 6000–6199). **Table 6** reports the excess return distributions for the portfolio of the top 20 pairs with the lowest SSDs (Panel A) and the portfolio of the

**Table 4. Excess Returns for Portfolio of Top 20 Pairs within Same Industry Group, July 1962–June 2009**

| | Utilities | | | | Financials | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962–1988 | 1989–2002 | 2003–2009 | Full Sample | 1962–1988 | 1989–2002 | 2003–2009 | Full Sample |
| Mean return (employed) | 0.0102 | 0.0032 | 0.0026 | 0.0070 | 0.0088 | 0.0078 | 0.0041 | 0.0078 |
| *t*-Statistic | 9.62** | 4.54** | 3.42** | 9.30** | 7.84** | 7.23** | 1.93** | 10.00** |
| Median return (employed) | 0.0090 | 0.0038 | 0.0018 | 0.0062 | 0.0081 | 0.0072 | 0.0011 | 0.0062 |
| Sharpe ratio | 1.02 | 0.41 | 0.28 | 0.70 | 0.53 | 0.71 | 0.34 | 0.54 |
| Observations with return < 0 | 8% | 29% | 40% | 18% | 23% | 24% | 40% | 26% |
| Mean return (committed) | 0.0100 | 0.0031 | 0.0025 | 0.0068 | 0.0084 | 0.0076 | 0.0041 | 0.0076 |

*Notes*: This table presents distribution statistics for monthly excess returns, over different sample periods, for the portfolio of top 20 pairs with the highest number of zero crossings chosen from 50 pairs with the lowest SSDs. The pairs are formed from either Utilities stocks or Financials stocks. "Employed" refers to the method of computing returns on employed capital; "committed" refers to the method of computing returns on committed capital. The *t*-statistics are computed by using Newey–West standard errors with six lags. Pairs are opened at the closing price one day after the divergence and closed at the closing price one day after the convergence.

**Significant at the 5 percent level.

**Table 5. Trading Statistics for Top 20 Sector-Based Pairs with Lowest SSDs and Highest Number of Zero Crossings, July 1962–June 2009**

| | Group 1 Nontraded Pairs | | | | Group 2 Nonconvergent Pairs | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962–1988 | 1989–2002 | 2003–2009 | 1962–2009 | 1962–1988 | 1989–2002 | 2003–2009 | 1962–2009 |
| *A. Utilities* | | | | | | | | |
| Proportion | 3% | 3% | 5% | 3% | 26% | 40% | 47% | 33% |
| $z$-Statistic[a] | na | na | na | na | na | 7.47** | 3.18** | na |
| Avg. monthly return per pair | na | na | na | na | –0.0060 | –0.0098 | –0.0091 | –0.0080 |
| $t$-Statistic[b] | na | na | na | na | na | –7.01** | 0.86 | na |
| St. dev.: monthly return per pair | na | na | na | na | 0.0124 | 0.0165 | 0.0149 | 0.0146 |
| Avg. SSD | 0.31 | 0.37 | 0.36 | 0.34 | 0.22 | 0.23 | 0.30 | 0.24 |
| Avg. zero crossings: formation | 46.29 | 41.28 | 34.84 | 42.44 | 49.27 | 44.25 | 35.68 | 44.71 |
| Avg. zero crossings: trading | 20.63 | 13.87 | 13.27 | 17.29 | 5.85 | 5.27 | 4.96 | 5.46 |
| Avg. round-trip length | na | na | na | na | na | na | na | na |
| Diverge minus traded spread | na | na | na | na | 0.0030 | 0.0002 | –0.0005 | 0.0013 |
| *B. Financials* | | | | | | | | |
| Proportion | 6% | 2% | 1% | 4% | 39% | 34% | 38% | 38% |
| z-Statistic[a] | na | na | na | na | na | –2.92** | 1.24 | na |
| Avg. monthly return per pair | na | na | na | na | –0.0151 | –0.0159 | –0.0159 | –0.0154 |
| $t$-Statistic[b] | na | na | na | na | na | –0.75 | –0.03 | na |
| St. dev.: monthly return per pair | na | na | na | na | 0.0279 | 0.0240 | 0.0239 | 0.0263 |
| Avg. SSD | 6.65 | 0.50 | 0.28 | 5.39 | 1.29 | 0.44 | 0.23 | 0.90 |
| Avg. zero crossings: formation | 35.75 | 42.67 | 43.11 | 37.17 | 37.63 | 44.16 | 45.37 | 40.56 |
| Avg. zero crossings: trading | 14.77 | 17.36 | 12.21 | 15.08 | 4.92 | 4.90 | 4.45 | 4.85 |
| Avg. round-trip length | na | na | na | na | na | na | na | na |
| Diverge minus traded spread | na | na | na | na | –0.0014 | –0.0006 | –0.0019 | –0.0012 |

| | Group 3 Single Round-Trip Pairs | | | | Group 4 Multiple Round-Trip Pairs | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962–1988 | 1989–2002 | 2003–2009 | 1962–2009 | 1962–1988 | 1989–2002 | 2003–2009 | 1962–2009 |
| *A. Utilities* | | | | | | | | |
| Proportion | 32% | 35% | 33% | 33% | 39% | 23% | 15% | 30% |
| $z$-Statistic[a] | na | na | na | na | na | na | na | na |
| Avg. monthly return per pair | 0.0077 | 0.0064 | 0.0091 | 0.0075 | 0.0238 | 0.0206 | 0.0248 | 0.0231 |
| $t$-Statistic[b] | na | –3.34** | 3.85** | na | na | –4.81** | 3.67** | na |
| St. dev.: monthly return per pair | 0.0098 | 0.0124 | 0.0159 | 0.0117 | 0.0162 | 0.0155 | 0.0153 | 0.0161 |
| Avg. SSD | 0.21 | 0.22 | 0.29 | 0.23 | 0.18 | 0.18 | 0.25 | 0.18 |
| Avg. zero crossings: formation | 49.59 | 44.68 | 36.20 | 46.09 | 51.40 | 45.96 | 36.69 | 49.13 |
| Avg. zero crossings: trading | 11.47 | 9.62 | 8.69 | 10.47 | 15.25 | 12.75 | 11.45 | 14.41 |
| Avg. round-trip length | 33.71 | 36.52 | 37.50 | 35.17 | 19.16 | 21.24 | 22.38 | 19.86 |
| Diverge minus traded spread | 0.0062 | 0.0034 | 0.0022 | 0.0047 | 0.0090 | 0.0065 | 0.0073 | 0.0083 |
| *B. Financials* | | | | | | | | |
| Proportion | 32% | 35% | 31% | 33% | 22% | 29% | 30% | 25% |
| z-Statistic[a] | na | na | na | na | na | na | na | na |
| Avg. monthly return per pair | 0.0144 | 0.0101 | 0.0044 | 0.0116 | 0.0437 | 0.0333 | 0.0280 | 0.0374 |
| $t$-Statistic[b] | na | –5.91** | –5.91** | na | na | –7.08** | –3.54** | na |
| St. dev.: monthly return per pair | 0.0211 | 0.0176 | 0.0180 | 0.0199 | 0.0414 | 0.0233 | 0.0324 | 0.0352 |
| Avg. SSD | 0.89 | 0.43 | 0.21 | 0.65 | 0.72 | 0.38 | 0.24 | 0.52 |
| Avg. zero crossings: formation | 38.54 | 44.98 | 44.84 | 41.48 | 41.01 | 45.88 | 44.86 | 43.35 |
| Avg. zero crossings: trading | 8.94 | 8.94 | 7.90 | 8.80 | 12.33 | 12.29 | 11.52 | 12.18 |
| Avg. round-trip length | 36.65 | 38.16 | 39.63 | 37.54 | 21.38 | 22.52 | 21.93 | 21.87 |
| Diverge minus traded spread | 0.0043 | 0.0044 | 0.0014 | 0.0040 | 0.0103 | 0.0082 | 0.0061 | 0.0089 |

na = not applicable.

*Notes*: This table presents trading statistics by pair group, over different sample periods, for the portfolio of top 20 pairs with the highest number of zero crossings chosen from 50 pairs with the lowest SSDs. The pairs are formed from either Utilities stocks (two-digit SIC code of 49) or Financials stocks (two-digit SIC codes of 60–67). Group 1 pairs never trade during the trading period. Group 2 pairs open but never converge in time during the trading period. Group 3 pairs have one round-trip trade and possibly another nonconvergent trade. Group 4 pairs have multiple round-trip trades and possibly a final nonconvergent trade. Pairs are opened at the closing price one day after the divergence and closed at the closing price one day after the convergence.

[a]The $z$-statistic for the null that the proportion of nonconvergent pairs in the current period equals that of the previous period.
[b]The $t$-statistic for the null that the mean return per pair in the current period equals that of the previous period.

**Significant at the 5 percent level.

**Table 6. Pairs Trading within Bank Stocks One Day after Trigger, July 1962–June 2009**

| | Top 20 Pairs with Lowest SSDs | | | | Top 20 Pairs with Highest Zero Crossings from 50 Lowest SSDs | | | |
|---|---|---|---|---|---|---|---|---|
| | 1962–1988 | 1989–2002 | 2003–2009 | Full Sample | 1962–1988 | 1989–2002 | 2003–2009 | Full Sample |
| Mean return (employed) | 0.0089 | 0.0069 | 0.0025 | 0.0074 | 0.0090 | 0.0088 | 0.0053 | 0.0084 |
| *t*-Statistic | 7.51** | 23.28** | 1.09 | 9.55** | 7.33** | 5.82** | 2.30** | 9.44** |
| Median return (employed) | 0.0087 | 0.0064 | 0.0011 | 0.0064 | 0.0079 | 0.0070 | 0.0038 | 0.0070 |
| Sharpe ratio | 0.56 | 0.57 | 0.20 | 0.51 | 0.53 | 0.55 | 0.38 | 0.52 |
| Observations with return < 0 | 28% | 26% | 42% | 29% | 26% | 24% | 32% | 26% |
| Mean return (committed) | 0.0079 | 0.0064 | 0.0025 | 0.0067 | 0.0079 | 0.0081 | 0.0053 | 0.0076 |

*Notes*: This table presents distribution statistics for monthly excess returns, over different sample periods, for two portfolios of top 20 pairs formed from bank stocks (SIC codes: 6000–6199). "Employed" refers to the method of computing returns on employed capital; "committed" refers to the method of computing returns on committed capital. The *t*-statistics are computed by using Newey–West standard errors with six lags. Pairs are opened at the closing price one day after the divergence and closed at the closing price one day after the convergence.

**Significant at the 5 percent level.

top 20 pairs with the highest number of zero crossings among the 50 pairs with the lowest SSDs (Panel B). No detectable improvement arises from trading only bank stocks as opposed to financial stocks in general. When the number of zero crossings in the pair-matching scheme (Panel B) is taken into account, however, the mean excess return increases by another 6 bps (to 84 bps), substantially higher than the 62 bp return (Table 1, last column) achieved by the standard top SSD-matched portfolio.

To put in perspective the importance of the zero-crossing statistic and the economic homogeneity between paired securities relative to the SSD metric, we performed a cross-sectional analysis of pair returns with respect to those variables. Specifically, we used the following regression:

$$\begin{aligned} Pair\ return_i &= Constant + a_1 Time\ trend + a_2 SSD_i \\ &+ a_3 SSD_i^2 + a_4 \log\left(Zero\ crossings\right) \\ &+ a_5 SameIndustryFlag + a_6 IndustryVol_i \\ &+ a_7 \left(IndustryVol_i\right)^2 + e_i, \end{aligned} \quad (5)$$

where *Pair return* is the excess return over the six-month trading period (for definitions of the other variables, see notes to **Table 7**). To increase the power of the regression and avoid serial correlation biases, we used the top 200 pairs with the lowest SSDs, pooled over 91 nonoverlapping trading cycles for the entire sample (1962–2009). Table 7 reports the summary statistics and regression results. The estimated coefficients for the zero-crossing and homogeneity variables are statistically significant, whereas those for $SSD$ and $SSD^2$ are not. These results further confirm the usefulness of the first two variables and the limitations of the last two in implementing pairs trading programs. The negative and significant time trend also supports the story of declining profitability documented previously. Interestingly, although marketwide volatility has been found not to affect pairs trading profitability (see Note 7), Table 7 shows that company-specific industry volatilities do, at both linear and quadratic levels. The signs of all the explanatory variables are as expected. Note that the $R^2$ value is very small, reflecting the significant variation among pairs that cannot be explained by the variables considered. Clearly, much more needs to be learned about pairs trading as a quantitative arbitrage strategy.[8]

## Conclusion

This article has examined pairs trading—a simple equity convergence trading strategy that has been found to be profitable over a long period, albeit at a declining rate. Extending the original sample used by GGR (1999, 2006), we found a continuation of the declining trend in recent years. Interestingly, increased competition in the hedge fund industry appears to be only part of the story: We found that worsening arbitrage risks in various pairs portfolios have accounted for as much as 70 percent of the decline, with the rest attributable to increased efficiency. Notably, pairs trading performance is particularly strong during market downturns, including the recent 2007–09 global financial crisis.

We demonstrated that SSD, as a conventional pair-matching criterion, is insufficient to identify close economic substitutes, which is a critical success factor for pairs trading. We proposed—and demonstrated success in—incorporating two additional metrics: industry homogeneity and the frequency of historical reversal in the price spread.

**Table 7. Regression Analysis of Pairs Trading Profitability, July 1962–June 2009**

| Summary Statistics | | Regression Results | | |
|---|---|---|---|---|
| | | | Coefficient | *t*-Statistic |
| Number of pairs[a] | 18,014 | *Constant* | –0.0973 | –4.82** |
| Mean return | 0.0292 | *Time trend* | –0.0004 | –6.56** |
| Standard deviation | 0.1838 | *SSD* | 0.0463 | 1.48 |
| Median excess return | 0.0319 | $SSD^2$ | –0.0245 | –0.72 |
| Mean SSD | 0.32 | log(*Zero crossings*) | 0.0244 | 4.99** |
| Standard deviation | 0.15 | *SameIndustryFlag* | 0.0245 | 7.68** |
| Mean number of zero crossings | 36.58 | *IndustryVol* | 0.4291 | 6.06** |
| Standard deviation | 9.79 | $(IndustryVol)^2$ | –0.5696 | –3.15** |
| Pairs from same industry | 38% | $R^2$ | 0.0090 | |

*Notes*: This table presents summary statistics and estimation results for the regression model specified by Equation 5. The dependent variable is the pair return per trading cycle of the top 200 pairs in terms of lowest SSDs, collected over 91 nonoverlapping six-month trading periods. *Time trend* assigns a trend value to each pair, with pairs traded over the first trading subperiod given a value of 1, those traded over the second subperiod given a value of 2, and so on. *SSD* is the pair's sum of squared price differences in the formation period. $SSD^2$ is the square of *SSD*. *Zero crossings* is the pair's number of zero crossings over the formation period. *SameIndustryFlag* is a dummy variable that equals 1 if the pair comes from the same industry (under the classification scheme in Fama and French 1997) and zero otherwise. *IndustryVol* is the average six-month realized volatility of the two industries to which the securities belong, computed over the relevant six-month period by using daily returns from all stocks in the relevant industry. $IndustryVol^2$ is the square of *IndustryVol*. $R^2$ is the coefficient of determination.

[a]The number of pairs is less than 200 × 91 = 18,200 because we excluded pairs whose stocks have an SIC code not belonging to one of the 48 industries in the classification scheme of Fama and French (1997).

**Significant at the 5 percent level.

Homogeneity involves matching securities within the same and narrowly defined industry groups to ensure close substitution by classification and, hence, lower divergence risk. To some extent, this metric can be viewed as a first step toward incorporating a fundamental aspect in pairs trading, which is traditionally a technical concept. Reversal frequency, computed as the number of zero crossings by the normalized price spread, measures how frequently the two securities crossed each other in the past. A high number of zero crossings signifies a "track record" of frequent mispricings within the pair that were successfully corrected by market participants. When combined with the SSD and homogeneity metrics, this track record measure has been found to enhance trading profits considerably.

Our findings have implications for practitioners who wish to implement pairs trading or, more generally, convergence trading. Profitable pairs trading is as much about identifying and excluding (or risk managing) divergent pairs as it is about identifying and including convergent pairs—in particular, multiple convergent pairs. Therefore, feasible strategies should start by defining fundamentally homogeneous asset groups because doing so helps not only avoid unnecessary search costs but also reduce nonconvergence risk. Fundamentally similar assets are likely to converge, and if they do not, they are likely not to drift apart. The 48-industry classification scheme by Fama and French (1997) has been found to generate immediate improvement as compared with the cruder four major group classifications. Future research and proprietary trading programs should consider even finer classification schemes and alternative portfolio formations that include intra-industry matched pairs from different industries, as opposed to top pairs within a single industry, the focus of this article.

*For their comments and suggestions, we thank participants who attended the following gatherings: the 2009 FMA European Conference in Turin, Italy; a research workshop at the Department of Accounting and Finance, Monash University; and the 2009 Q-Group–Monash Colloquium. Finally, we thank Dr. Rasika Amarasiri for his excellent computing support. This article is part of a project funded by an AFAANZ 2008 research grant and a 2008 department grant from the Department of Accounting and Finance, Monash University.*

*This article qualifies for 1 CE credit.*

# Notes

1. Although the algorithm is rather straightforward, the computational cost is substantial and concentrates on the pair-matching stage, in which the normalized price spread is computed for each possible pair of securities that has met all data requirements. We used a "parallel computing" approach, which enabled us to reduce the computing time to a couple of hours. The MATLAB code is available from the authors upon request.
2. Four additional tests further support this conclusion of declining profitability, the full details of which are available from the authors upon request. The first is a statistical test on the time trend in which the monthly excess returns are regressed against time. Using a Wald-type test statistic (Vogelsang 1998) that is robust to serial correlation, we found that the coefficient is negative and statistically significant. The second test checks whether the pattern remains for a stock sample that is less affected by liquidity and short sale constraints. We performed this test by excluding companies in the lowest size decile (using NYSE breakpoints) and also those priced under $5 (which is consistent with common practice in the momentum trading literature; see Jegadeesh and Titman 1993). Pairs trading performed on this dataset exhibits slightly smoother monthly performance, yet the declining trend, as well as the bear market resilience, remains robust. Third, we also investigated whether the profitability decline and the increased divergence risk apply to pairs portfolios other than that of the 20 pairs with the lowest SSDs. Specifically, we looked at the portfolios of the top 50 pairs and the top 100 pairs, which might offer a greater diversification effect. We found that profits generated from these portfolios also declined over time. Finally, we conducted a test of statistical arbitrage formulated by Hogan, Jarrow, Teo, and Warachka (2004), which is also used in a number of trading strategy studies (see, e.g., Yu 2006). Statistical arbitrage is defined intuitively as a strategy that generates positive monthly profits, on average, in present value terms and in which the variation in the profit series averaged across time is increasingly negligible. Confirming the *t*-test in Table 1, the estimated mean return parameter ($\mu$) is positive and significant. The parameter on the marginal decrease in return volatility ($\lambda$), however, is positive, leading to the rejection of pairs trading as a statistical arbitrage for 1962–2009.
3. In an illuminating account of the August 2007 event in which quantitative traders suffered one of the biggest losses during the financial crisis of 2007–2009, Khandani and Lo (2007) suggested that the losses resulted from the sudden unwinding of large long–short equity portfolios held by multistrategy funds, even though the initial shock seemed to emanate from somewhere else.
4. Note that the monthly returns reported in the first row of Panel A of Table 3 are slightly different from the monthly figures in Table 1 (under the one-day delayed trading rule) because the former are the monthly returns averaged across pairs whereas the latter are the monthly returns averaged across all six portfolios that are staggered by one month.
5. We obtained the CRSP–Compustat matching from the CRSP–Compustat merged database via WRDS (Wharton Research Data Services). We performed the CRSP–I/B/E/S matching by using CUSIP as the secondary identifier. We were able to find a matched I/B/E/S record for all stocks of the group of top 20 pairs.
6. Note that the effect documented in Papadakis and Wysocki (2007) was found for only 9–14 percent of pairs. Also, their results were based on a much narrower database than were ours, with only 2 trading cycles a year (commencing in January and July) instead of the 12 monthly cycles that we used.
7. This analysis is restricted to earnings events only and does not purport to suggest that news in general has no effect on pairs trading. In fact, Engelberg et al. (2009) found that the type of news released around the divergence date plays a significant role in explaining the cross-sectional variations in pairs profits. Another potential explanation for the increased divergence documented previously is elevated market volatility, which could translate into spiraling losses in the sense of Xiong's equilibrium model (2001): When an unfavorable shock inflicts losses on convergence traders, thus eroding their risk-bearing capacity, they are forced to unwind positions, which amplifies the original shock. We tested this possibility by regressing pair returns against market volatility over the same month but did not find any significant effect.
8. The results from our cross-sectional analysis also shed light on why Industrials and Transportation pairs underperform Utilities and Financials pairs. Only 13 percent of the Industrials pairs are made up of stocks from the same industry, as defined in Fama and French (1997), compared with 45.8 percent for Financials pairs and 100 percent for Utilities pairs. Although almost all the Transportation pairs (99.88 percent) are made up of stocks from the same Fama–French industry group, this particular industry is not homogeneously defined because it comprises rather diverse businesses, ranging from railways and trucking to bus charters and travel agencies. This finding is reflected in the statistics for SSDs and the number of zero crossings: The average SSD for Transportation pairs is 2.65, compared with just 0.22 for Utilities pairs and 0.91 for Financials pairs; the average number of zero crossings for Transportation pairs is 30, compared with 46 for Utilities and 41 for Financials.

# References


Abreu, D., and M. Brunnermeier. 2002. "Synchronization Risk and Delayed Arbitrage." *Journal of Financial Economics*, vol. 66, no. 2–3 (November–December):341–360.

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

De Long, J., A. Shleifer, L. Summers, and R. Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, vol. 98, no. 4 (August):703–738.

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

Fama, E., and K. French. 1997. "Industry Costs of Equity." *Journal of Financial Economics*, vol. 43, no. 2 (February):153–193.

Gatev, E., W. Goetzmann, and K. Rouwenhorst. 1999. "Pairs Trading: Performance of a Relative Value Arbitrage Rule." Working paper, Yale School of Management.

———. 2006. "Pairs Trading: Performance of a Relative-Value Arbitrage Rule." *Review of Financial Studies*, vol. 19, no. 3 (Fall):797–827.

Hogan, S., R. Jarrow, M. Teo, and M. Warachka. 2004. "Testing Market Efficiency Using Statistical Arbitrage with Applications to Momentum and Value Strategies." *Journal of Financial Economics*, vol. 73, no. 3 (September):525–565.

Jegadeesh, N., and S. Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, vol. 48, no. 1 (March):65–91.

Khandani, A., and A. Lo. 2007. "What Happened to the Quants in August 2007?" *Journal of Investment Management*, vol. 5, no. 4 (4th quarter):5–54.

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

Mitchell, M., T. Pulvino, and E. Stafford. 2002. "Limited Arbitrage in Equity Markets." *Journal of Finance*, vol. 57, no. 2 (April):551–584.

Papadakis, G., and P. Wysocki. 2007. "Pairs Trading and Accounting Information." Working paper, Boston University School of Management and MIT Sloan School of Management (September).

Vogelsang, T. 1998. "Trend Function Hypothesis Testing in the Presence of Serial Correlation." *Econometrica*, vol. 66, no. 1 (January):123–148.

Xiong, W. 2001. "Convergence Trading with Wealth Effects: An Amplification Mechanism in Financial Markets." *Journal of Financial Economics*, vol. 62, no. 2 (November):247–292.

Yu, F. 2006. "How Profitable Is Capital Structure Arbitrage?" *Financial Analysts Journal*, vol. 62, no. 5 (September/October):47–62.