12th International Conference on Credit Risk Evaluation, Venice, Italy, 26-27 Sept 2013

Empirical inference of related trading between two securities: Detecting pairs trading, merger arbitrage, and strategy rules[*]

Keith Godfrey

The University of Western Australia

The traditional approach to studying pairs trading is to simulate profitability using ex-post historical prices. I study the actual trades reported anonymously in security pairs and build statistical inferences of related trading. The approach is based on the time differences between trades. It can distinguish intrinsically related securities from pseudo-random sets, find stocks involved in merger arbitrage in massive sets of paired index constituents, and infer dominant trading rules of mean reversion algorithms. Empirical inference of related trading can enable further studies into pairs trading, strategy rules, merger arbitrage, and insider trading.

Keywords: Inferred trading, empirical inference, pairs trading, merger arbitrage.

JEL Classification Codes: G00, G10, C10, C40, C60

The availability of intraday trading or "tick" data with time resolution of a millisecond or finer is opening many avenues of research into financial markets. Analysis of two or more streams of tick data concurrently is becoming increasingly important in the study of multiple-security trading including index tracking, pairs trading, merger arbitrage, and market-neutral strategies.

One of the greatest challenges in empirical trading research is the anonymity of reported trades. Securities exchanges report the dates, times, prices, and volumes traded, without identifying the traders. In studies of a single security, this introduces uncertainty of whether each market order that caused a trade was the buy or sell order, and there are documented approaches of inference such as Lee and Ready (1991). Studies of two or more securities are significantly more complex.

---

[*] This working paper has developed from my Ph.D. dissertation "Can pairs trading act like arbitrage to enforce the law of one price?", The University of Western Australia, 2012. I am indebted to Philip Brown, Raymond Da Silva Rosa, Arnold Zellner, Robert Faff, Caroline Commerton-Forde, Pamela Moulton, Terry Walter, Doug Foster, Richard Heaney, and Dan Galai for helpful comments and discussions. All remaining errors are my own.

Aside from needing to make these inferences for each security separately, there is additional uncertainty when analysing them together to connect related trades. Researchers face a total lack of information about whether a particular trade in one security was executed by the same trader as a trade in another security.

The innovation in this paper is a statistical approach to inferring related trading. It relies on the time differences between trades and analyses the cumulative distribution functions of these time differences. It is found to be robust to calendar periods and different reporting delays between exchanges, and capable of locating security pairs involved in pairs trading and merger arbitrage.

The remainder of this paper is organised as follows. Section 1 reviews the characteristics of trading in multiple securities and the challenges of empirical inference. Section 2 develops the methodology for inferring related trading, initially for detecting any kind of related trading, then for distinguishing kinds of related trading, and finally for testing the consistency of trading rules with the empirical data. Section 3 presents results for identifying fundamentally related pairs from pseudo-random sets, finding pairs trading in massive sets formed from index constituent pairs, and determining the dominant trading rules in pairs of dual-listed (or Siamese twin) securities.

1. Introduction

The idea of analysing two stocks together for profit has developed over many years. Long before the era of computers, Livermore (1940) described a method of analysing two related stocks to determine their common price trend. It was a laborious process calculated by hand, and the focus at the time was on the common movement rather than the difference. Half a century later the emergence of computer power helped to automate the pair trending analysis, and it was in this environment the idea of pairs trading evolved. Bookstaber (2007) describes how Gerald Bamberger, a young programmer at Morgan Stanley, started to think of the pairs not as a block to be executed but as two sides of a trading strategy. By going long in one and short the other, the net position was market-neutral. Morgan Stanley allowed Bamberger to test his strategy and it made six million dollars in the first year. Nunzio Tartaglia took control of the trading group and reportedly made 50 million dollars for the firm in 1987 (Gatev et al. 2006).

Pairs trading involves the purchase of an under-priced security and the simultaneous sale or short of an over-priced security in such proportions as to maintain a market-neutral position. The combined position is held until either the price difference converges to a target level, or diverges to exceed a stop loss. Being a market-neutral strategy, it aims to make a profit irrespective of the direction of market movement.

A market-neutral equity strategy can involve any number of long and short positions in any combination of securities, provided the overall portfolio has no expected net exposure to risk. Jacobs and Levy (2005) describes several such strategies including market-neutral equity, convertible bond arbitrage, government bond arbitrage, and merger arbitrage, as well as pairs trading. The focus here is on pairs trading and merger arbitrage which involve just two securities.

How much market-neutral trading goes on in practice? At this point the literature becomes silent.

1.1 Simulations of Trading Strategies in Two Securities

Trading simulations estimate the profitability of particular trading strategies, and there are many documented studies involving security pairs. Using daily prices from 1990 to 2001, Alexander and Dimitriu (2002) simulate pairs trading among the 30 stocks in the Dow Jones Industrial Average and estimate annual profits of around 10% with 2% volatility and negligible correlation with the market. Hong and Susmel (2003) simulate pairs trading between 64 Asian ADRs and their underlying stocks from 1991 to 2000 and calculate annualised profits of over 33% if investors were to hold the positions for a year. Gatev et al. (2006) simulate pairs trading in U.S. stocks from 1962 to 2002 and calculate annualised excess returns of around 11% before trading costs, or between 2.6% and 4.5% after costs. Chen et al. (2010) conduct a long-run simulation of pairs trading in U.S. stocks using daily and monthly data from 1931 to 2007 and find average returns of 11% to 36% annually before trading costs. The overall message is the strategy could have been successful. Chen et al. (2010) also document the returns are diminishing over time, suggesting the market is adapting to pairs trading and becoming more efficient.

One way to improve the profitability of a trading algorithm is to trade with higher frequency information using intraday rather than daily data. Any large price discrepancies are likely to be short-lived, and Suarez (2005) points out these will be mostly invisible to observers with daily sampling. There is plenty of intraday price data available, and many recent studies use this. Nath (2003) simulates pairs trading in the secondary market for U.S. government debt using trade and quote data from 1994 to 2000 and finds positive excess returns relative to a duration-matched benchmark. Dunis et al. (2010) simulate pairs trading amongst the Eurostoxx 50 index constituents using five-minute prices and calculate pairs trading underperforms the index after trading costs. Bowen et al. (2010) simulate high-frequency pairs trading on a sample of FTSE 100 constituents during 2007 and find the excess returns of the strategy are sensitive to transaction costs, the entry trigger, and delays in execution: a 15-minute delay in execution can eliminate the returns. They also suggest the time of day can be important, noting the majority of returns occur from positions opened in the first hour of trading.

The literature on trading simulations is extensive and growing. Its weakness is it documents only paper-trade or ex-post simulations, and these prices are not tradeable. Real trading uses ex-ante bid-ask prices and the trades have real market impact. The profitability of real trading is unlikely to agree with the simulations. The problems in using ex-post simulation as a proxy for real trading include:

- The use of historical prices often overlooks the fact these prices may be efficient having already incorporated any profitable trades;
- Simulations do not account for the price impact of the simulated trades being traded;
- Simulations may inadvertently include unrealistic processes such as using forward data; and
- There is an inaccuracy when simulations use traded prices rather than the bid and ask prices.

The first two problems are endemic in any simulation study. The third and fourth are avoidable by experimental design although there are many documented studies that make these mistakes.

Overall the simulated profits may bear little resemblance to the profits or losses that may be incurred by real trading, and many authors are forthright in acknowledging this limitation.

A more subtle impairment to the literature on trading simulations is the conflict of interest between publishing and profiting from such information. Altucher (2004 page xi) puts the dilemma bluntly: “If these systems are so good, why not just use them to print money all day long? Why write about them?” Altucher goes on to explain valid reasons for publishing, the main one being that the trading systems must evolve continuously, requiring constant research and development. Nonetheless the literature is likely to suffer from a selection bias towards the less profitable or unprofitable algorithms, or a delay before publication while profits are exploited. Morgan Stanley kept silent about its evolving pairs trading strategies in the 1980s, but by the 2000’s an abundance of articles on pairs trading had emerged with comments the simulated profits were decreasing.

It would be interesting to know whether arbitrageurs are executing trades similar to those being simulated, and if so how much of that trading goes on, and how closely the profits from real trading match the simulations. These kinds of research question are difficult to answer because of the anonymity of trading. The refereed literature is sparse on such topics.

1.2 Empirical Observations of Trading in Two Securities

There is plenty of anecdotal evidence of pairs trading occurring in practice, but little reference to empirical trading in the refereed literature. Several broad literature searches failed to find any journal articles focussed on the detection of pairs trading or the amount and types of trading being undertaken by practitioners. These included searches of databases such as Business Source Premier, JSTOR, ProQuest 5000 International, ScienceDirect, and the Social Sciences Research Network (SSRN). There are papers that mention the topic in passing and others that attempt small-scale tests as part of another study, but studying the actual trading in detail in two or more securities appears to be a space wide open for research.

Anecdotal evidence comes from traders claiming to have had success, books describing how pairs trading had been conducted, and evidence of government and market responses. Reverre (2001) suggests the arbitrage of Royal Dutch – Shell is a popular model on Wall Street because it has characteristics close to those of absolute convergence. The theory is traders assume the observed value of the price ratio is the superposition of a fundamental function and market noise so they can choose a moving average as an estimator of the fundamental function. Wojcik (2005) describes cases of pair trades going wrong, implying traders were caught up in those trades at the time. Paul (2008) explains how Australia’s share market regulator ASIC relaxed short-selling bans on dual-listed stocks as a result of lobbying from traders, implying pairs trades or arbitrage trades were being conducted at the time. In all these reports there is a lack of detail about the types and quantities of trades undertaken. Hedge Fund Research (2011) says merger arbitrage hedge funds have returned on average 1.12% per annum more than the S&P 500 index from 1998 to 2010, but again there are no details of the algorithms employed.

1.3 Empirical Inference of Trading in Two Securities

Inferring trading between two securities requires a proxy or measure. Do and Faff (2009) mention in one paragraph the possibility of detecting arbitraging activities by examining the spread on the day that follows the opening trigger, arguing the spread should narrow if a large number of traders follow the prescribed strategy and act on mispricing. This proposes one possible way to detect trades, but those words were removed in the subsequent journal revision of the paper, perhaps because a narrowing spread can be caused by many other reasons too.

Schultz and Shive (2010) describe in one paragraph how they investigate the trades in dual-class shares from the perspective of studying how prices converge and diverge. They use a process of matching trades as a proxy for arbitrage trades, where a matched trade is defined as a purchase of shares in one class occurring within one minute of a sale of the same number of shares in the other class. They use the algorithm of Lee and Ready (1991) to classify trades into purchases and sales and investigate the matched sales of expensive shares with purchases of cheap shares, concluding the volume increases when a price differential exists. The result appears to confirm the intuitive proposition that arbitrageurs would exploit such differences.

In parallel, Schultz and Shive express surprise to find the change in volume from matched trades is far less than the change from single-sided trades which they proxy by the non-matched trades. They speculate this is due to the single-sided trades being more important in enforcing the prices than round-trip arbitrage trades, a concept described earlier by Deardorff (1979) as “one-way arbitrage”. It is possible to envisage an alternate explanation that any experimental approach of testing whether the amount of single-sided trading exceeds the amount of round-trip trading involves a joint hypothesis test with the choices of proxies. Put simply, the decision to match only on equal volumes in opposite directions of aggression and within one minute of each other means their proxy for the single-sided arbitrage trades may be sub-optimal.

It is hard to analyse trades in two securities when the trades are anonymous. The problems include:

- An algorithm can infer the direction of aggression in one security but cannot match the orders across two securities. Trading in two securities might involve a market order or limit order in one security followed by a market order in the other, but the first two possibilities would be inferred as having opposite directions of aggression, leading to misclassification;
- A securities exchange may split orders into smaller lots during execution, so any matching algorithm based on traded volume may be inaccurate. This too can lead to misclassifications of paired trades as unrelated single-sided trades and vice versa; and
- Fixed-size matching windows introduce artefacts. Changing the 1-minute window in Schultz and Shive to 30 seconds or 2 minutes is likely to change the results dramatically.

We need a more general way of inferring trading in two or more securities, approaches that can be forgiving of the inherent uncertainty in classifying the observed trades. We also need to exercise caution when interpreting the results from using such proxies, and to devise tests that offer confidence they are working.

2. Inferring Related Trading

Studying related trading in two securities is difficult because traders have a financial incentive to keep their best strategies confidential, and security exchanges collaborate by concealing the traders’ identities. The consequences for research include barriers for matching the simultaneous entry or exit trades in those securities and barriers for matching a subsequent exit with the entry in each security. Researchers have to live with the uncertainty of whether any two trades are part of the same strategy from the same trader. Attempt to match such trades without identification will inevitably be statistical and likely to need large quantities of trading data to provide sufficient statistical confidence. This section develops one such approach.

2.1 Empirical Inference of Related Trading based on Times Between Trades

Trading in two or more securities can be executed either to capture a joint trend or to trade the difference for maintaining market neutrality. Trades in the same direction include program trades where baskets of securities are purchased or sold simultaneously. Trades in opposite directions include the market-neutral strategies of pairs trading, index arbitrage, and merger arbitrage. All of these, whether in the same direction or in opposite directions, represent related trading in the securities. The first task is to be able to identify any kind of related trading.

The analysis here will focus on two securities. These may be any two securities, irrespective of any fundamental relations or cointegration characteristics. The approach here will make use only of the time differences between trades, so there is no influence from the price directions. Additional information from price, and subsequent inferences of the directions of trades, shall be used later to infer different types of related trading.

Consider two securities AA and BB which have sets of trades

$$\begin{aligned} \mathbf{AA} &= \{ \mathrm{AA}_1, \mathrm{AA}_2, \mathrm{AA}_3, \ldots, \mathrm{AA}_{\|\mathbf{AA}\|} \} \\ \mathbf{BB} &= \{ \mathrm{BB}_1, \mathrm{BB}_2, \mathrm{BB}_3, \ldots, \mathrm{BB}_{\|\mathbf{BB}\|} \} \end{aligned} \tag{1}$$

If we denote the signed time difference between trades AA_i and BB_j as $\mathbf{\Delta t}(\mathrm{AA}_i, \mathrm{BB}_j)$ we can define the set of all such time differences by $\mathbf{\Delta t}(\mathbf{AA} \times \mathbf{BB})$. This is the set of time differences to be analysed but it is likely to be too large for practical operations Its size is that of the Cartesian product set formed by matching every trade in **AA** with every trade in **BB**:

$$\|\mathbf{\Delta t}(\mathbf{AA} \times \mathbf{BB})\| = \|\mathbf{AA} \times \mathbf{BB}\| = \|\mathbf{AA}\| \times \|\mathbf{BB}\| \tag{2}$$

To enable practical analysis we can define a subset in which the time differences between trades is limited to a domain $[-T, T]$ for suitable choice of T and denoted $\mathbf{AA} \times \mathbf{BB} : |\mathbf{\Delta t}| \leq T$. We can similarly define a subset for any interval of time differences $[T_1, T_2]$ as

$$\mathbf{AA} \times \mathbf{BB} : T_1 \leq \mathbf{\Delta t} \leq T_2 = \left\{ \begin{array}{ll} (\mathrm{AA}_i, \mathrm{BB}_j) \; : & \mathrm{AA}_i \in \mathbf{AA},\ \mathrm{BB}_i \in \mathbf{BB}, \\ & T_1 \leq \mathbf{\Delta t}(\mathrm{AA}_i, \mathrm{BB}_j) \leq T_2 \end{array} \right\} \tag{3}$$

This notation enables us to define an empirical measure or relative frequency of time differences based on the interval $[T_1, T_2]$ with $-T \leq T_1 \leq T_2 \leq T$ as

$$\begin{aligned} \hat{\mathrm{P}}_{\mathbf{AA} \times \mathbf{BB} : |\mathbf{\Delta t}| \leq T} [T_1 \leq \Delta t \leq T_2] &= \frac{\text{number of products with } \Delta \mathrm{t} \in [T_1, T_2]}{\text{number of products with } \Delta \mathrm{t} \in [-T, T]} \\ &= \frac{\| \mathbf{AA} \times \mathbf{BB} : T_1 \leq \mathbf{\Delta t} \leq T_2 \|}{\| \mathbf{AA} \times \mathbf{BB} : -T \leq \mathbf{\Delta t} \leq T \|} \end{aligned} \tag{4}$$

The empirical distribution or cumulative density function (CDF) is

$$\begin{aligned}\hat{\mathrm{F}}_{\mathbf{AA}\times\mathbf{BB}:\,|\mathbf{\Delta t}|\le T}(t) &= \hat{\mathrm{P}}_{\mathbf{AA}\times\mathbf{BB}:\,|\mathbf{\Delta t}|\le T}[-T \le \Delta t \le t] \\ &= \frac{\text{number of products with } \Delta\mathrm{t} \in [-T,t]}{\text{number of products with } \Delta\mathrm{t} \in [-T,T]} \\ &= \frac{\|\,\mathbf{AA}\times\mathbf{BB} : -T \le \Delta t \le t\,\|}{\|\,\mathbf{AA}\times\mathbf{BB} : -T \le \Delta t \le T\,\|}\end{aligned} \tag{5}$$

These measures are readily calculable from the records of anonymously reported trades by counting the number of pairs having time differences within each range, although this may involve a large number of computations.

The insight to detecting related trading between two securities is to recognise the following:

Proposition 1: If the trading in two securities is unrelated, and the opportunity to trade is available continuously, the time differences between trades in the two securities should approximate a uniform distribution.

This idea is illustrated in panel (a) of Figure 1. When there is no related trading between the two securities, the time differences between any two trades selected at random from the two continuously-traded securities should approximate a uniform distribution on $[-T,T]$. Alternatively, the presence of related trading should distort the distribution according to the time differences of the paired trades being executed by the traders.

In practice trading on most exchanges does not occur continuously. The NYSE opens from 9:30 to 16:00 daily, giving a window of 6.5 hours for trading. If an exchange is open for trading for a window of W seconds per day, and if trades are reported with one-second resolution, there are W ways two trades can be 0 seconds apart, $W-1$ ways they can be 1 second apart in each direction, $W-2$ ways they can be 2 seconds apart, and so on down to 1 way they can be $W-1$ seconds apart. The cumulative density function on $[0,T]$ is then a decreasing square law of the form

$$\mathrm{F}^{**}_{[0,T]}(t) = \frac{Wt - t^2/2}{WT - T^2/2} \tag{6}$$

rather than a uniform distribution. Nonetheless the square law portion is flat around $t=0$, so if we consider a small limit of say $T=100$ seconds which is just 0.43% of the NYSE daily opening window of $W = 23400$ seconds, this approximates a uniform distribution

$$\mathrm{F}^{*}_{[0,T]}(t) = \frac{t}{T} \tag{7}$$

and this means we can continue to work with Proposition 1.

The uniform empirical measure over the interval $t \in [-T,T]$ has CDF

$$F^*(t) = \frac{t+T}{2T} \tag{8}$$

so the observed CDF of time differences will differ from the uniform distribution by

$$[\hat{F} - F^*](t) = \hat{F}_{\mathbf{AA}\times\mathbf{BB}:\,|\Delta \mathbf{t}|\leq T}(t) - F^*(t) \tag{9}$$

This is the curve illustrated in panel (b) of Figure 1.

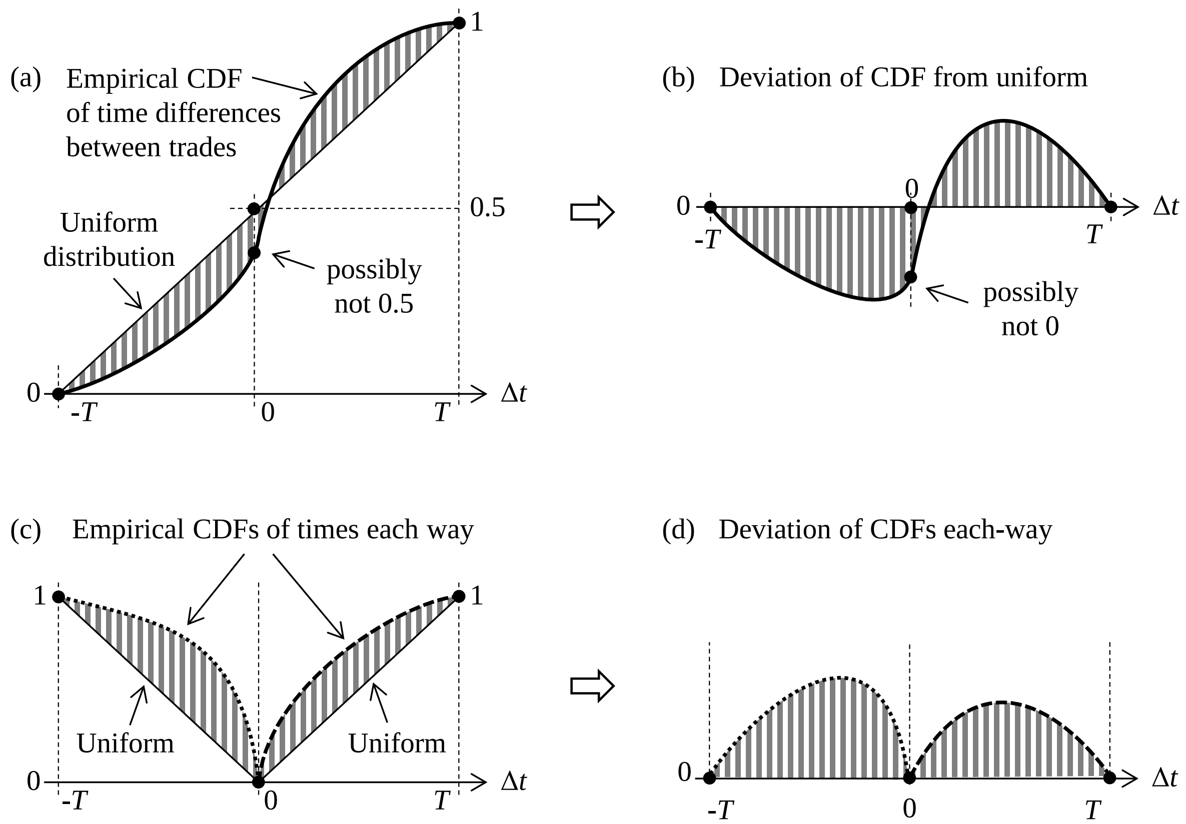

Figure 1 Concept of inferring related trading using the cumulative density functions (CDFs) of time differences between trades in two securities. The horizontal axis in each panel represents the time difference from a trade reported in one security to a trade reported in the other security, within the limits [-*T*, *T*]. Panel (a) illustrates an empirical distribution of such time differences compared with a uniform distribution being the expected distribution when no related trading occurs between the securities. The grey shaded area (the difference between the empirical CDF and the uniform distribution) is a proxy for the amount of related trading between the securities. Panel (b) shows the difference more explicitly. The limitation is the zero-crossing may not be at Δt=0 if there is an excess of trades where one stock leads the other. Panels (c) and (d) illustrate a resolution by computing the CDFs separately for [0, -*T*] and [0, *T*]. The grey regions in panel (d) are proxies for the amount of related trading inferred in each direction. These regions can be measured by the Cramér von Mises criterion and Kolmogorov-Smirnov distance, giving numerical indicators of inferred related trading.

To infer a measure of related trading, we want a single metric to capture the difference: a distance measure between the empirical CDF and the uniform distribution.

Proposition 2: A distance measure between the empirical CDF of time differences between trades in two securities and a uniform distribution can be a proxy for the amount of related trading.

Possible alternatives include the Cramér von Mises (CVM) criterion

$$\omega^2 = \frac{1}{2T}\int_{-T}^{T}\left(\hat{F}(t) - F^*(t)\right)^2 dt \tag{10}$$

and the Kolmogorov-Smirnov (KS) distance

$$D = \sup_{t\in[-T,T]}\left|\hat{F}(t) - F^*(t)\right| \tag{11}$$

The single numerical result from any such measure becomes a Related Trading Indicator (RTI) and we can use subscripts such as CVM or KS to denote the chosen measure. The RTI computations for the CVM and KS measures are summarised as

$$\begin{aligned} \mathrm{RTI}_{\mathrm{CVM}} &= \sqrt{\frac{1}{2T}\int_{-T}^{T}\left(\hat{\mathrm{F}}_{\mathbf{AA}\times\mathbf{BB}:\,|\mathbf{\Delta t}|\le T}(t) - \mathrm{F}^*(t)\right)^2} \\ \mathrm{RTI}_{\mathrm{KS}} &= \sup_{t\in[-T,T]}\left|\hat{\mathrm{F}}_{\mathbf{AA}\times\mathbf{BB}:\,|\mathbf{\Delta t}|\le T}(t) - \mathrm{F}^*(t)\right| \end{aligned} \tag{12}$$

These are the continuous-time equations for the RTIs. In practice the indicators will be computed from the time differences between reported trades, where the trade reporting is resolved to regular discrete time intervals. Denoting the time resolution by τ (which may be a microsecond or smaller) and the maximum time difference of interest $T = N\tau$ for some integer N, the computation becomes

$$\begin{aligned} \mathrm{RTI}_{\mathrm{CVM}} &= \sqrt{\frac{1}{2N+1}\sum_{n=-N}^{N}\left(\hat{\mathrm{F}}_{\mathbf{AA}\times\mathbf{BB}:\,|\mathbf{\Delta t}|\le T}(n) - \mathrm{F}^*(n)\right)^2} \\ \mathrm{RTI}_{\mathrm{KS}} &= \sup_{n\in[-N,N]}\left|\hat{\mathrm{F}}_{\mathbf{AA}\times\mathbf{BB}:\,|\mathbf{\Delta t}|\le T}(n) - \mathrm{F}^*(n)\right| \end{aligned} \tag{13}$$

This is the discrete-time form of the related trading indicators. The RTIs are readily calculable from Equations (5) and (13) although the number of computations may be large.

It should be possible to use either of these RTIs to measure inferred related trading, but there may be subtle differences between them depending on the characteristics of the pair of securities being analysed. The CVM criterion computes a root-mean-square of the grey shaded regions in Figure 1, while the KS distance measures their maximum heights. We can imagine the CVM criterion may be better at inferring related trading when a security pair is traded by many traders having a wide range of time differences in their execution strategies, while the KS distance may be better in pairs where trading is dominated by a single trader executing in a narrow range of time differences.

Each RTI is a relative, not absolute, indicator of related trading. Security pairs with larger indications can be interpreted to have more related trading than pairs with lower indications, ceteris paribus, but the actual value or magnitude has no physical meaning.

By analysing only the empirical distributions of the time differences, we do not require prior knowledge of the trading strategies present in the data. We can expect any related trading to display as an excess relative frequency of trades occurring at particular time differences. We can also imagine traders will want to keep those time differences as short as possible, either when executing a programmed basket of trades simultaneously, or when maintaining market neutrality in the case of pairs trading. A time difference limit of $T = 10$ seconds is likely to be sufficient and $T = 100$ seconds should be more than generous enough to catch the majority of trades of interest.

2.2 Discerning Between Aggressive and Passive Pairs Trading, and Program Trading

The methodology of the previous section is designed to infer related trading between two securities based on the time differences alone. This analysis of the CDFs does not require the prices of those trades nor the volumes traded. The way to apply the methodology to more complex tasks is to recognise it can be applied to any subsets of trades.

Proposition 3: Related trading can be inferred between subsets of trades in two securities that are inferred or constructed from price, volume, and time information to distinguish particular types of related trading.

This means partitioning the set of trades $\mathbf{AA} \times \mathbf{BB} : |\mathbf{\Delta t}| \le T$ into subsets of interest such as those having particular inferred combinations of purchases and sales, or those occurring at times of inferred entries and exits of particular trading algorithms. The choice of partitions determines the information to be deduced.

We begin by incorporating price information to distinguish aggressive and passive pairs trading. The aggressor in each trade can be inferred from the price of the trade and the prevailing bid-ask spread, most simply by bisecting the bid-ask spread, or by more complicated algorithms such as those in Lee and Ready (1991). The simple bisection approach is sufficient here to show the capabilities of the methodology. Alternatives and enhancements can be trialled later. From here on, trades occurring above the midpoint of the prevailing bid-ask spread are classified as initiated by the buyer, and those below the midpoint are classified as initiated by the seller.

When analysing two securities, the relative aggression of any two trades becomes complicated. A pairs trade involves opening or closing opposite positions in the two securities but the style can be either aggressive or passive. An aggressive pairs trade means submitting a pair of market orders in opposite directions, which may be inferred from the trading logs as a pair of trades in opposite directions. A passive pairs trade involves waiting for a limit order to be filled in one security before executing a market order in the other, which may be inferred from the logs as two trades in the same direction (because the direction of aggression inferred from the trade with the limit order is the opposite direction to the passive limit order).

Program or basket trades may also be inferred with aggressions in the same direction because they arise from market orders in the same direction. This leads to potential confusion in distinguishing passive pairs trading from program or basket trading. One possible solution is to partition by time difference as well as by relative direction of aggression. The trades executed from the simultaneous market orders of a program trade should occur closer in time than those requiring a trader to react to news of a passive order being filled before ordering the second trade. Figure 2 illustrates the set of trade combinations $\mathbf{AA} \times \mathbf{BB}$ being reduced initially to the subset $\mathbf{AA} \times \mathbf{BB} : |\mathbf{\Delta t}| \le T$ then subdivided further by time and relative directions of trade.

When a related trading indicator (RTI) is calculated for one of these subsets, it can be called an aggressive pairs trading indicator (APTI), passive pairs trading indicator (PPTI), or basket trading indicator (BTI). In each case the calculation is the same for the general RTI but applied to the particular subset of trades. The APTI, PPTI, or BTI can be suffixed similarly with a subscript for the distance measure employed such as CVM or KS.

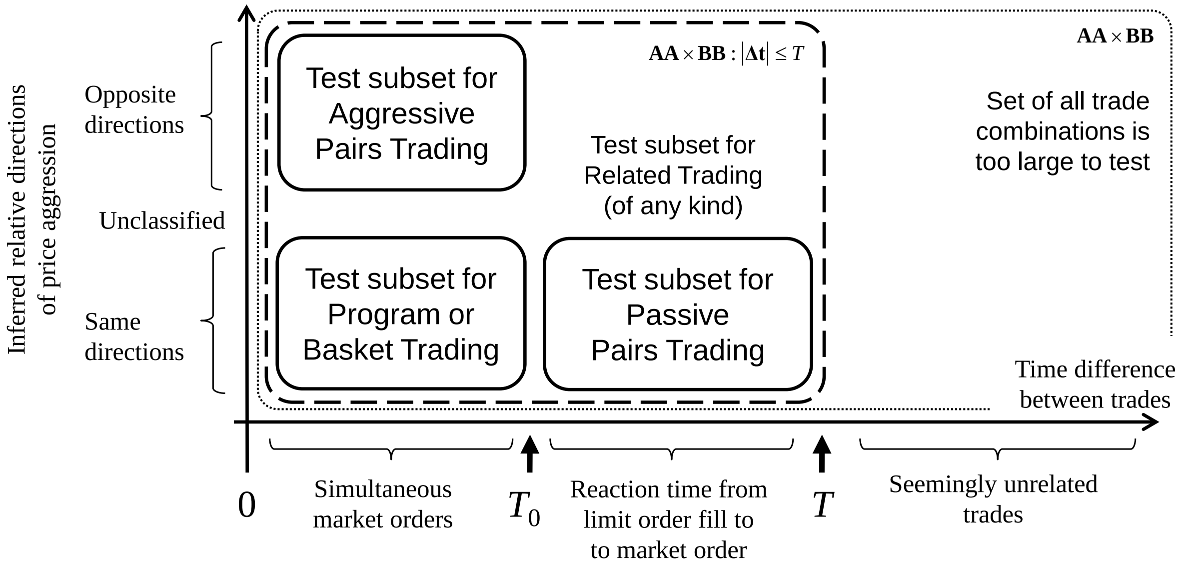

Figure 2 Classification of trading from two securities into subsets based on the inferred relative directions of aggression and the absolute time difference between the trades. Trades in each security are classified as buy or sell according to the trade price and the prevailing bid-ask spread. Trade pairs are then classified in two dimensions with the vertical axis separating those having opposite directions of aggression (buy sell or sell buy) from those with the same directions (buy buy or sell sell), and the horizontal axis being the absolute time difference between the trades.

2.3 Empirical Inference of Trading Rules

Pairs trading is a mean-reverting strategy, relying on the premise the price difference or relative value will return to some kind of mean in the near future. A typical trading rule is to enter a position when the price difference is unusually high or low, and exit when it either returns to become close enough to a target or diverges to a stop loss. Traders must choose several parameters in setting up their trading algorithm. A moving average requires choices such as the period and the type of moving average or the time-weighting (linear or exponential). The entry and exit rules then require choices of how far the price difference or ratio should diverge and converge.

Traders are likely to test several alternatives before committing significant funds to a particular strategy. Similarly any empirical experiment to infer the dominant trading strategies may also need to test several conjectures. The aim of empirical inference is to determine the trading rules most consistent with the empirical trading records.

The ability to infer related trading can help infer the dominant trading strategy when the trades are partitioned according to the entry and exit rules of each strategy under test. We can partition the trade sets further into the buy-sell and sell-buy pairs by inferring directions of trade. If the strategy under test is consistent with the trading activity in the log, there should be an excess of buy-sell pairs at times when the strategy entry rule is at one of its extreme levels and an excess of sell-buy pairs at the other extreme. Alternatively if the strategy is a poor match for the empirical data, there should be no significant excess of either kind observed.

Proposition 4: If real trading in two securities follows a particular algorithm, the empirical inference of related trading should be consistent with the entry and exit signals of that algorithm.

Given a relative valuation rule $\mathbf{v}(\mathrm{AA}_i, \mathrm{BB}_j)$ for a particular trading strategy being tested, we can partition the trade combinations into *K*-quantiles according to the relative valuation rule. With equal partitioning the boundary (or quantising) levels of the *K*-quantiles are determined by

$$P\Big[\ \mathbf{v}(\mathrm{AA}_i, \mathrm{BB}_j)\ \leq\ k\text{-th}\ \ K\text{-quantile}\ \Big] = \frac{k}{K} \tag{14}$$

and we can define subsets $Q_{k(+,-)}$ and $Q_{k(-,+)}$ to represent the trades that may be the buy-AA-sell-BB combinations and sell-AA-buy-BB combinations respectively within the *k*-th *K*-quantile of the strategy being tested.

Applying the related trading methodology to these sets and computing the cumulative density functions of time differences in Equation (5) gives $\hat{\mathrm{F}}_{Q_k(+,-)}(t)$ and $\hat{\mathrm{F}}_{Q_k(-,+)}(t)$. We could apply Equation (9) to difference each from a uniform distribution, but on this occasion it is unnecessary because it would cancel out with the next step. The method of inferring the dominant trading rules is to compute differences between these CDFs. Comparing these CDFs is analogous to the earlier task of comparing a single CDF to a uniform distribution. If the trading rule of the strategy being tested is unrelated to the trades observed, the shape of $\hat{\mathrm{F}}_{Q_k(+,-)}(t)$ should be similar to $\hat{\mathrm{F}}_{Q_k(-,+)}(t)$ within each *k*-quantile partition. Alternatively, if there is an excess of trades consistent with the strategy under test, it should cause a difference between these CDFs.

The excess of trades should appear as either positive or negative according to the direction of the excess. The difference between the CDFs can be denoted $\hat{\mathrm{X}}_{\mathbf{AA}\times\mathbf{BB}(+-,k)}$ being the excess of buy-AA-sell-BB pairs to sell-AA-buy-BB pairs in the *k*-th *K*-quantile subset and calculated as

$$\begin{aligned} \hat{\mathrm{X}}_{\mathbf{AA}\times\mathbf{BB}(+-,k)}(t) &= \left(\hat{\mathrm{F}}_{Q_k(+,-)}(n) - F^*(t)\right) - \left(\hat{\mathrm{F}}_{Q_k(-,+)}(n) - F^*(t)\right) \\ &= \hat{\mathrm{F}}_{Q_k(+,-)}(t) - \hat{\mathrm{F}}_{Q_k(-,+)}(t) \end{aligned} \tag{15}$$

This generates a set of *K* difference curves corresponding to the *K*-quantiles, each curve representing the excess of inferred strategy trading for the corresponding *K*-quantile of the relative value rule. An ideal result for a perfect match of trading strategy would be an ordered alignment the *K* curves, with the greatest inferred trading in one direction when the relative valuation rule is most favourable one way, and the greatest reverse trading occurring when the rule is at its opposite extreme. The curves can be plotted with the CDFs calculated each way to assist visual inspection, so the two sides have the *K*-quantiles oriented the same way up.

It is convenient to compute a single numerical score for the goodness of fit of a trading strategy, just as it is to compute a numerical indicator for related trading. The aim of the numerical score is to measure how well the K curves are ordered. If the K-quantiles are oriented so the first K-quantile corresponds to an expected purchase of AA and sale of BB, and if the inferred trading is consistent with the strategy, the curves would ideally be found ordered with any two curves satisfying

$$j < k \;\Rightarrow\; \hat{\mathrm{X}}_{\mathbf{AA}\times\mathbf{BB}(+-,j)}(t) \;>\; \hat{\mathrm{X}}_{\mathbf{AA}\times\mathbf{BB}(+-,k)}(t) \tag{16}$$

for all $t \in [-T, T]$. The degree to which the curves obey this empirically is an indication of how well the strategy fits the trading data. We can define a consistency of ordering for the j-th and k-th K-quantiles as a function on j, k, and t as

$$\hat{\mathrm{C}}(j,k,t) = \begin{cases} 1 \text{ if } \hat{\mathrm{X}}_{\mathbf{AA}\times\mathbf{BB}(+-,j)}(t) > \hat{\mathrm{X}}_{\mathbf{AA}\times\mathbf{BB}(+-,k)}(t) \\ -1 \text{ otherwise} \end{cases} \tag{17}$$

then compute a single average over all j, k, and t as a K-quantile ordering score (KQOS) for the strategy under test as

$$\mathrm{KQOS} = \frac{2}{K^2 - K} \sum_{j=1}^{K-1} \sum_{k=j+1}^{K} \frac{1}{2T} \int_{-T}^{T} \hat{\mathrm{C}}(j,k,t) \tag{18}$$

Instead of calculating this on [-T, T] the time differences can be separated by sign, leading to separate KQOSs for trades in AA leading BB and trades in BB leading AA, analogous to the each-way CDFs lead to separate RTIs. The KQOSs for these can be denoted $\mathrm{KQOS}_{\mathrm{AA}\rightarrow\mathrm{BB}}$ and $\mathrm{KQOS}_{\mathrm{AA}\leftarrow\mathrm{BB}}$ for calculation on [0, T] and [0, -T] respectively.

Each K-quantile ordering score has the range $[-1,1]$ with an intuition similar to a correlation coefficient. The maximum value of 1 can occur if the inferred trading is ordered perfectly with the test strategy for all time lags. A value around zero suggests the test strategy is inconsistent with or unrelated to the inferred trading. The minimum of -1 can occur if the inferred trading appears perfectly ordered in the reverse way to the strategy, which would suggest the K-quantiles were defined the wrong way for the for the strategy. The absolute value |KQOS| provides an overall score for the test strategy and is forgiving of inadvertent reversal.

The |KQOS| is a relative indicator of the goodness of fit of various strategies rather than an absolute measure for any particular strategy. When several alternative strategies are tested and one is observed to stand out with a |KQOS| closer to 1 than the others, it can be labelled the dominant strategy or the rule most prevalent amongst those tested.

3. Empirical Explorations of Related Trading

3.1 Validation of the Methodology

One way to validate the related trading inference methodology is to test it on security pairs where pairs trading is anticipated and compare the inferred levels with randomly chosen pairs having no such prior expectation. This requires choices of security pairs where pairs trading is anticipated and choices for the comparison pairs. In an ideal test the comparison pairs can be chosen from pseudo-random sets having trading characteristics similar to the main pairs.

The NYSE contains several pairs of closely-related securities trading as stocks or American Depository Receipts (ADRs). The following four pairs are selected for testing here:

- Tickers (Reuters instrument codes) RDSa and RDSb are the twin ADRs of Royal Dutch Shell Plc, a global petrochemical company. Royal Dutch Shell was studied previously by Rosenthal and Young (1990) and Froot and Dabora (1999) when it had a dual-company structure prior to 2005. The post-2005 structure comprises two share classes a single parent company derived from Dutch and British origins, and they are practically identical.
- Tickers UN and UL are the twin ADRs of Unilever NV and Plc respectively, a global nutrition and hygiene product manufacturer, with a dual-listed Dutch and British company structure.
- Tickers CCL and CUK are the twin stock and ADR respectively of Carnival Corporation, a global cruise line and holiday provider. These codes derive from a Panamanian company Carnival that took over a UK company P&O Cruises in 2003.
- Tickers BHP and BBL are the twin ADRs of BHP Billiton, the world's largest diversified resources group, representing the Australian and British limbs of its dual-listed company structure formed in 2001.

These four pairs represent four economically significant entities, and each has twin securities that can be expected to be traded together in strategies such as pairs trading.

To test the methodology we need to compare with similar pairs not anticipated to be involved in pairs trading. Comparison sets can be constructed from pairs of securities having similar trading characteristics to each of the main pairs. Picking the four closest securities to RDSa by number of trades in the period, plus four more by trading volume, and four by dollar volume, gives a set of 13 securities with characteristics similar to RDSa (including RDSa itself). Building a set similarly of 13 securities around RDSb enables a set of $13 \times 13 = 169$ security pairs to be constructed from all combinations of the RDSa and RDSb sets, of which one pair is expected to show pairs trading and 168 are the pseudo-random comparison pairs. This approach is repeated for each of the four pairs of interest and for each time period to be tested.

Trading data are obtained from Thomson Reuters Tick History with access through SIRCA. In 2008 there were 54 009 328 trades in the 13 securities in the set constructed around RDSa and 16 063 792 in the set around RDSb. Pairing all these trades gives a Cartesian product set of size 822 418 826 468 768 which is unwieldy for analysis. Limiting the time difference between trades to $T = 100$ seconds reduces the number of trade pairs for analysis to 37 305 542 257. This is still a large set but it is manageable.

Figure 3 shows the empirical distributions or CDFs of the 169 pairs in the Royal Dutch Shell test set in 2008 computed from Equation (5). One pair stands out a long way from the other 168 and it is RDSb-RDSa as predicted. A similar result occurs in the pairs for Unilever, Carnival, and BHP Billiton. The results for all four pairs are robust to calendar year when repeated for sets generated separately for 2009 and 2010. It is unlikely these perfect results are obtained by chance. The 37 305 542 257 trade combinations analysed for the Royal Dutch Shell set in 2008 are independent of the 43 696 415 909 combinations in 2009 and the 32 688 816 870 in 2010, and again from the Unilever, Carnival and BHP Billiton sets in 2008 to 2010 which range from

17 954 524 703 to 171 083 917 733 combinations. These numbers are large and come from 12 disjoint sets. The methodology appears to be working for these securities.

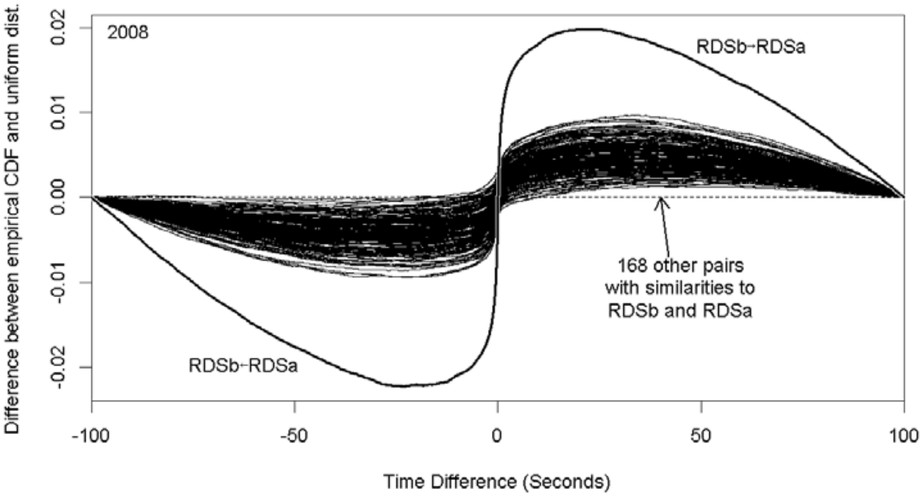

Figure 3 Deviations from the uniform distribution of the CDFs of time differences between trades in Royal Dutch Shell pair RDSb-RDSa and 168 pseudo-random pairs in 2008. The set of pairs is constructed by combinations of RDSb and 12 similarly-traded securities with RDSa and 12 similarly-traded securities. The curve for RDSb-RDSa stands out from all the other pairs by a significant distance, indicating a much greater presence of related trading. Repeating this test for Unilever, Carnival and BHP Billiton in three disjoint time periods validates the methodology of computing numerical related trading indicators (RTIs) by measurement of these curves.

Besides validating the methodology, the tests also discovered a few pairs in the pseudo-random comparison sets that appear to be traded. Figure 4 shows the RTIs calculated for the 169 security pairs around Unilever (UN and UL) in 2009. The UL-UN pair stands out by a long distance from the main clump of pseudo-random pairs, verifying the methodology, but there is an additional feature of interest. The pair SWK-BDK also stands out albeit by a smaller margin. This pair comprises Stanley Works and Black & Decker, two competing manufacturers of tools and hardware which merged on 12 March 2010 to form the Stanley Black and Decker Corporation. The empirical inference of related trading here appears to have found merger arbitrage in the pre-merger securities (the 2009 data is at least two months prior to the merger date). This deserves further study. The ability to infer such related trading may open more avenues for empirical research into mergers and acquisitions, including empirical inference of possible insider trading activity ahead of particular announcement dates.

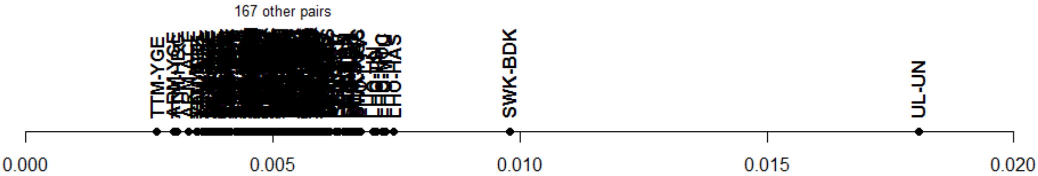

Figure 4 Related Trading Indicator RTI $_{CVM}$ for the Unilever pair UL-UN and 168 pseudo-random pairs in 2009. The set of pairs is constructed by all combinations of UL and 12 similarly-traded securities with UN and 12 similarly-traded securities. The related trading methodology works as anticipated by distinguishing the UL-UN pair from the others, but a second pair SWK-BDK also stands out to a lesser extent. This pair is Stanley Works and Black & Decker, two competing manufacturers of tools and hardware which merged on 12 March 2010 to form the Stanley Black and Decker Corporation. Related trading (in this case merger arbitrage) is inferred in those securities in the 2009 data more than two months ahead of the merger.

Additional testing finds the methodology is robust to the pairs of security exchanges on which the trades execute. The BHP Billiton sets from each of 2008 to 2010 were partitioned by each of the 25 combinations of the five most frequently traded exchanges, and the methodology continued to identify the BBL-BHP pair regardless of which security traded first. This is important because different exchanges may have different delays in their trade reporting mechanisms and these could introduce errors into any analysis of time differences between trades. Slight patterns did emerge with this partitioning, suggesting there may be differences between exchange reporting mechanisms, but these were insufficient to affect the main result for the pairs tested. The differences in the CDFs for different combinations of exchanges may even be able to help infer relative differences in the exchange reporting delays, and order the exchanges by reporting delay. This may deserve further investigation.

3.2 Exploration for Pairs Trading amongst Index Constituents

With more than 40 000 equity securities listed on exchanges around the world, there is a potential for related trading in more than 800 million security pairs. In practice pairs trading can be performed only when one or both of securities can be shorted so we can restrict the universe to pairs of short-sellable securities. Interactive Brokers lists 15 383 of these in March 2013, so the practical universe still contains more than 118 million pairs. Random searches for pairs trading in this space are unlikely to have success. The explorations here shall be restricted to four sets of index constituents: the S&P 500, NASDAQ 100, S&P MidCap 400, and FTSE 100. The numbers of trade combinations to analyse in these sets are still large, as listed in Table 1.

Index	Chain RIC	Time period	Codes	Code pairs	Trade products within *T*=100 seconds
S&P 500	0#.spx	2011 H1	513	131 328	87 529 904 232 756
NASDAQ 100	0#.ndx	2010	107	5 671	15 127 152 881 204
S&P MidCap 400	0#.mid	2010	451	101 475	3 637 794 060 531
FTSE 100	0#.ftse	2010	117	6 786	151 046 067 927

Table 1 Summary of the four equity indices explored for pairs trading among constituents. RIC means Reuters Instrument Code in Thomson Reuters Tick History, and Chain RICs listed here provide the constituent lists of each underlying indices. The periods of analysis are the 12 months of 2010 except for the S&P 500 which is analysed during the first six months of 2011. The S&P period was reduced to six months because there are many more trades amongst its constituents than in all the other indices combined, and for robustness its period was chosen not to overlap with 2010. The numbers of codes analysed in each case exceeds the size of the index because of additions and deletions during the periods and the presence of additional codes for different classes of shares. In general for N codes there are $(N^2-N)/2$ pairs. Even after reducing the trade combinations to those executed within T=100 seconds, the numbers of trade combinations for analysis remain large.

Figure 5 shows the result of computing the aggressive pairs trading indicator APTI_{CVM} on the pairs of index constituents from the S&P 500 index with time limit T=10 seconds. The S&P 500 data set was obtained from Thomson Reuters Tick History using chain 0#.spx which returned 513 codes including those added, deleted, and changed during the period as well as additional codes for distinct classes of shares. The 1 450 933 036 trades reported in the first six months of 2011 create a Cartesian product set of size 1 048 326 510 036 240 846 from the 131 328 security pairs. With the time difference limited to 100 seconds, the number of trade products to analyse is reduced to 87 529 904 232 756. This is still a large number and took a month to process on an eight-core laptop. The analysis was also performed with limits of 1, 2, 5, 10, 20 and 50 seconds for robustness. We can envisage improvements in performance over time as computer hardware evolves because the methodology developed here and experimental sequence are both entirely parallelisable.

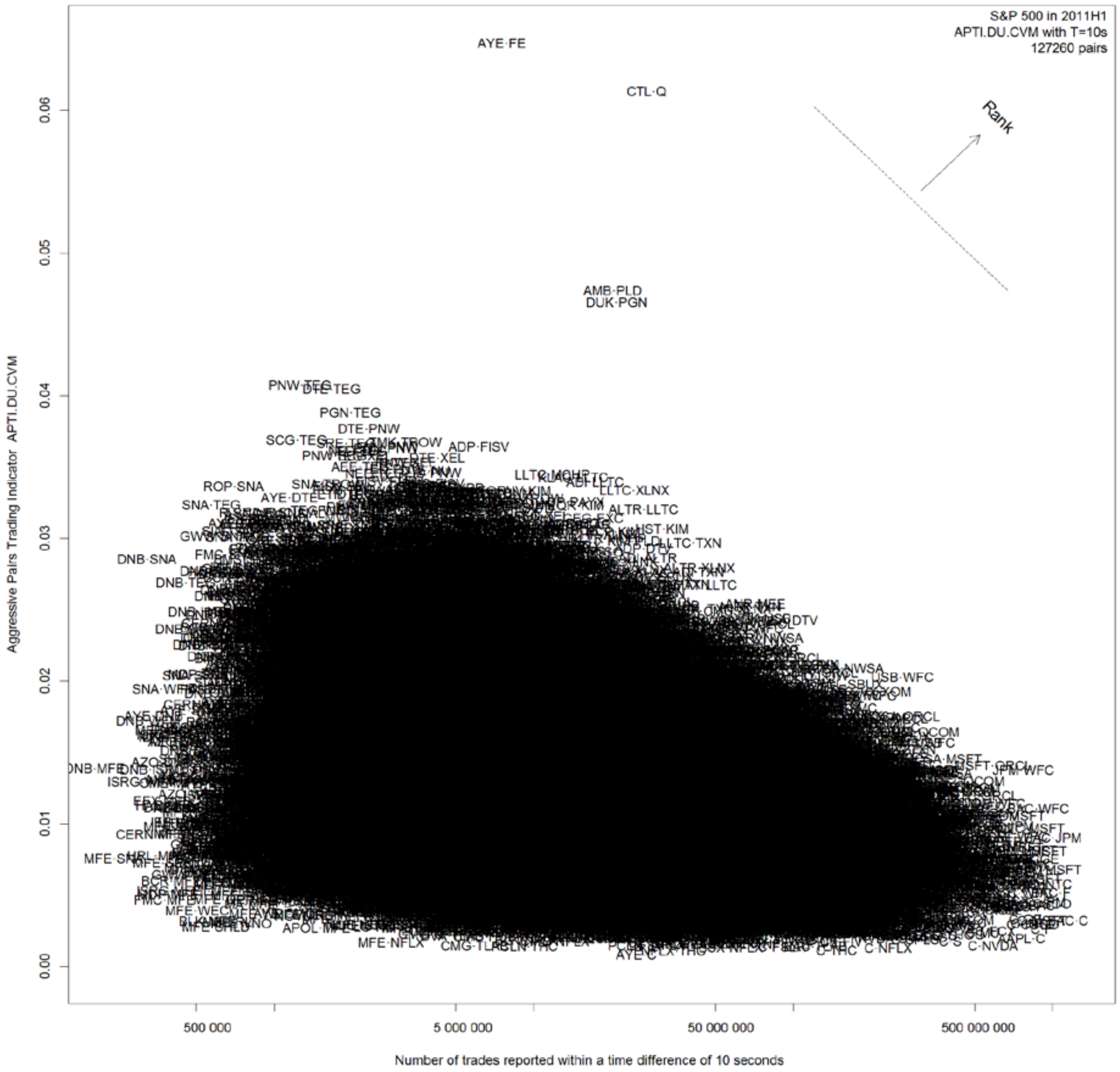

Figure 5 Scatterplot of aggressive pairs trading inferred empirically by indicator $APTI_{CVM}$ in 127 260 pairs of S&P 500 constituent combinations in the first six months of 2011. Pairs are formed between all combinations including index additions and removals during the period, provided the securities trade at least 1% as frequently as the most frequently traded security. When plotted in two dimensions with inferred related trading vertically and the frequency of trading horizontally, a frontier of pairs with the greatest inferred related trading faces the upper right corner. Examples of outstanding security pairs near the top of the diagram include AYE-FE and CTL-Q which were candidates for merger arbitrage trading during this period. Pairs along the leading (upper right) frontier tend to comprise related industries such as ALTR-LLTC (semiconductors), MSFT-ORCL (software), and JPM-WFC (banking). Detailed analysis of these pairs is listed in Table 2.

The pairs from the S&P 500 index in Figure 5 are plotted in two dimensions with the aggressive pairs trading indicator $APTI_{CVM}$ on the vertical axis and the number of trades within time difference of T=10 seconds on the horizontal axis. The number of pairs plotted is 127 260 after discarding those securities which trade less than 1% as frequently as the most traded security.

The diagram can be described as having a few pairs such as CTL-Q and AYE-FE standing out by a wide margin, and a diagonal frontier of pairs with the most inferred related trading facing the upper right corner.

Pair	Potential explanation for the empirically inferred trading
AYE-FE	FirstEnergy Corp (FE) is an electricity provider which absorbed Allegheny Energy Inc (AYE) on 25 February 2011, two months into the period analysed. The inferred trading is likely to be merger arbitrage.
CTL-Q	CenturyLink Inc (CTL) is an integrated telecommunication company which absorbed Qwest Communications International Inc (Q) on 1 April 2011 in the middle of the period analysed. The inferred trading is likely to be merger arbitrage.
AMB-PLD	AMB Corporation (AMB) and Prologis Inc (PGD) were real estate investment trusts (REITs) that merged on 3 June 2011 toward the end of the analysis period. The inferred trading is likely to be merger arbitrage.
DUK-PGN PNW-TEG DTE-TEG DTE-PNW SCG-TEG DTE-XEL	Duke Energy Corporation (DUK), Progress Energy Inc (PGN), Pinnacle West Capital Corporation (PNW), Integrys Energy Group Inc (TEG), DTE Energy Co (DTE), SCANA Corporation (SCG) and Xcel Energy Inc (XEL) are energy companies with electricity generation, transmission, and/or distribution businesses. Any pair of these is a plausible candidate for pairs trading based on common industry fundamentals.
ADP-FISV	Automatic Data Processing (ADP) and Fisserv Inc (FISV) are providers of human payroll systems and financial services technology. Pairs trading is plausible based common industry fundamentals.
LLTC-MCHP KLAC-LLTC ADI-LLTC LLTC-XLNX ALTR-LLTC LLTC-TXN ALTR-XLNX	Linear Technology Corporation (LLTC), Microchip Technology Inc (MCHP), KLA-Tencor Corporation (KLAC), Analog Devices Inc (ADI), Xilinx Inc (XLNX), Altera Corporation (ALTR) and Texas Instruments Inc (TXN) are all high-tech companies involved in the integrated circuit industry. Any pair of these is a plausible candidate for pairs trading based on common industry fundamentals.
HST-KIM	Host Hotels & Resorts (HST) and Kimco Realty Corp (KIM) are REITs with interests in hotels and shopping centres respectively. Pairs trading is plausible based common industry fundamentals.
BK-USB USB-WFC BK-WFC JPM-WFC BAC-WFC BAC-JPM	The Bank of New York Mellon Corporation (BK), U.S. Bancorp (USB), Wells Fargo & Company (WFC), JPMorgan Chase & Co (JPM) and Bank of America Corporation (BAC) are all financial sector businesses providing banking, insurance, investments, and finance. Any pair of these is a plausible candidate for pairs trading based on common industry fundamentals.
CVX-XOM COP-XOM	Chevron (CVX), Exxon Mobil (XOM) and ConocoPhillips (COP) are oil and gas companies. Any pair of these is a plausible candidate for pairs trading based on common industry fundamentals.
CMCSA-NWSA	Comcast Corporation and News Corporation are providers of entertainment, information and news. Pairs trading is plausible based on the common industry.
MSFT-ORCL	Microsoft (MSFT) and Oracle (ORCL) are two of the world's leading software development companies. Pairs trading is plausible based on the common industry.

Table 2 Analysis and explanations for pairs trading in leading security pairs from Figure 5. The three pairs most outstanding (AYE-FE, CTL-Q, and AMB-PLD) are found to have merged during the period of analysis so trading inferred empirically is likely to have arisen from merger arbitrage. The next 25 pairs identified on the frontier are found to comprise companies in the same industries making them plausible candidates for pairs trading based on common fundamentals. The 28 leading pairs happen to be either involved in mergers or share industry fundamentals, despite comprising just 0.022% of the 127 260 pairs plotted in Figure 5. The fact the leading pairs identified from empirical inference of pairs trading have strong justification for merger arbitrage or pairs trading gives confidence the methodology is working.

Table 2 shows an investigation of the pairs on the diagonal frontier. It finds the top three pairs identified (AYE-FE, CTL-Q, and AMB-PLD) are candidates for merger arbitrage because they merged during the study period. The next 25 listed along the diagonal frontier (such as PGN-TEG and LLTC-XLNX) are all strong candidates for pairs trading because their securities each turn out to be from the same industries. By contrast an inspection of the bottom end finds pairs such as C-NFLX (Citigroup and Netflix: a bank and a movie subscription service) that seem to have no apparent relation. The way the RTI and APTI approaches are finding strong candidates for pairs trading among the 127 260 pairs analysed suggests they are successful at detecting such trading empirically. The methodology appears to be accurate at its task.

The other three indices provide similar kinds of results. Investigation of the NASDAQ 100 constituent pairs in 2010 identifies LLTC-XLNX and MSFT-ORCL which are common with the S&P 500 analysis but offer robustness because the analyses here are for disjoint periods. Several company pairs are identified which turn out to be in the same industries. Investigation of the S&P MidCap 400 (which by construction, unlike the NASDAQ 100, has no overlap with the S&P 500) highlights pairs which turn out to be from the same industry and are strong candidates for pairs trading. Analysis of the FTSE 100 finds two outstanding pairs which each comprise two share classes from the same company (NG-NGn and SDR-SDRt) in addition to several pairs found to be in the same industries. The methodology appears to be working as anticipated and is robust to the different exchanges and calendar periods involved.Further investigations suggest there is more to learn about applying the methodology to other exchanges internationally. Initial explorations of the German DAX 30 constituents and Australian ASX 200 constituents found little inference of pairs trading, compared with the US and UK index constituents. It would be interesting to test the reasons. Perhaps the DAX 30 contains too few constituents and perhaps pairs trading in the Australian market is less well developed than in the US and UK, or alternatively perhaps there is something different about the way those exchanges report their trades. There are many opportunities for subsequent research. It is sufficient here to demonstrate the potential of the methodology by discussing its success with constituent pairs from the S&P 500, S&P Midcap 400, NASDAQ 100, and FTSE 100 indices.

3.3 Emprical Inference of a Dominant Trading Rule

The final experiments re-examine the four closely-related pairs Royal Dutch Shell, Unilever, Carnival, and BHP Billiton to investigate the dominant trading rules in each pair. The law of one price suggests the twins in each pair should trade at the same price with that price being enforced by arbitrage. In practice the twins in some of these pairs trade at significantly different prices, particularly BHP which has traded at a premium to BBL above 25% in 2011.

Four possible candidate strategies will be tested for each of the four pairs. The first trading rule tested is that of price equalisation or reversion to parity, where entry rules linked to the difference between the most recently traded prices. The remaining three trading rules are based on price reversion to moving averages based on the most recent 100 000, 10 000, and 1 000 trades respectively. These choices are somewhat arbitrary and may not match exactly the trading activities in the market, but they turn out to be sufficiently accurate to demonstrate the capability of the approach.

The trades in each set are divided into K=5-quantiles (which could be called quintiles) according to the entry rules of each of the four trading rules, then the differences between CDFs are calculated as per Equation (15) and the 5-quantile ordering scores (the 5QOSs) are calculated from Equation (18). The results are shown in Table 3. Visual inspections of the 5-quantile curves in each case (not plotted here) confirm the 5QOSs are accurate summaries of each case.

Table 3 shows the results for the 2009 trading data. Several inferences can be made from this table. The first pair, Royal Dutch Shell, shows a significant dominance of RDSb being traded before RDSa. Perhaps the orders are placed in RDSb first because it is the less-frequently traded security. The reasons could deserve further investigation, an opportunity for further research. The second inference from Royal Dutch Shell is the greatest 5-quantile ordering score occurs for the strategy of price equalisation, although all four strategies show some consistency with the empirical trading data. The implication is the dominant trading strategy is reversion to price equality, and this observation seems reasonable because the prices of the two classes of security RDSa and RDSb have traded extremely closely since Royal Dutch Shell formed a single parent company structure in 2005, remaining within about 5% of each other at all times.

The results for Unilever suggest a strategy of mean reversion to a moving average of between 10 000 and 1 000 trades is more common than the strategy of price equalisation. The results for Carnival show a similar rejection of price equalisation, although the magnitudes of the scores are not as strong as for Unilever. These observations make sense because the prices in the Unilever and Carnival pairs have each deviated much further than with Royal Dutch Shell, which means a strategy targeting price parity is less likely to succeed. The twin securities in Unilever reached a price difference of 10% in the year prior to this study, and those in Carnival exceeded 15%.

Security pair and trade sequence / Trading rule being tested	Royal Dutch Shell		Unilever		Carnival		BHP Billiton	
	RDSa then RDSb	RDSb then RDSa	UN then UL	UL then UN	CCL then CUK	CUK then CCL	BHP then BBL	BBL then BHP
Reversion to parity (price equalisation)	0.15	0.87	0.49	0.45	0.53	0.31	-0.13	-0.50
Reversion to 100 000 - trade moving average	0.36	0.75	0.53	0.61	0.66	0.74	0.75	0.63
Reversion to 10 000 - trade moving average	0.43	0.82	0.80	0.90	0.70	0.76	0.76	0.79
Reversion to 1 000 - trade moving average	0.56	0.72	0.73	0.83	0.36	0.77	0.98	0.88

Table 3 Ordering scores for empirically inferred pairs trading in 5-quantiles of trade combinations selected to test four possible trading strategies in four pairs of closely-related securities. The numbers are the 5-quantile ordering scores (5QOSs) calculated by Equations (15) and (18) and are analogous to correlation with a range of [-1, 1]. The results for Royal Dutch Shell show the empirically inferred trading is most consistent with price equalisation although the other strategies are possible fits too, and there is a strong preference for the RDSb security to be traded first. The results for Unilever and Carnival favour mean reversion to moving averages between 10 000 and 1 000 trades. The results for BHP Billiton show strong rejection of a price equalisation strategy and strong consistency with a trading strategy of reversion to a moving average of around 1 000 trades (about 40 minutes for BHP and BBL in 2009).

The most interesting result comes from BHP where the prices of the twin ADRs maintained a difference of around 20% during the period of study. The results for BHP in Table 3 reject the strategy of mean reversion with the poorest match of all four strategies under test, while giving a near-perfect score of 0.98 for the strategy of reverting to a moving average of 1 000 trades (about 40 minutes for BHP-BBL in 2009). This is the highest score of all of the tests. The rejection of empirical evidence of a price equalisation strategy makes sense given the persistent 20% price difference, while the strong compatibility with the 40-minute moving average strategy suggests pairs trading is occurring on a high-frequency basis. It appears the empirically inferred pairs trading is enforcing the high-frequency price difference in BHP Billiton rather than restoring price equality.

These results show it can be possible to infer the dominant pairs trading strategies in closely-related securities despite the anonymity of trading. While the methodology is imprecise, the results seem accurate because they are consistent with the fundamental and technical observations of the twin securities. Of the four sets of security twins studied, the BHP Billiton pair has the greatest and most persistent price disparity and was found to be traded with short-term mean reversion rather than price equalisation. By contrast the single parent company and dual-class security structure of Royal Dutch Shell offers the greatest fundamental match and it was found to be traded accordingly.

There are implications here for our understanding of the law of one price (LOOP). The premise of the LOOP is identical goods should trade at the same price, enforced by arbitrage. In financial markets this means securities offering the same dividend stream should trade at the same price. The empirical inference of trading in Royal Dutch Shell post 2005 is consistent with this idea, while at first sight the empirical inference for BHP Billiton appears not to be. The empirical inference for BHP Billiton suggests a short-term moving average of the price ratio is being enforced by pairs trading. If there are fundamental reasons behind the price ratio of BHP Billiton, such as investor tax heterogeneity, the observable moving average ratio may be a proxy for the unobservable fundamental ratio, in which case the pairs trading can be thought of as enforcing the law of one price.

4. Conclusions

This paper overcomes one of the greatest challenges in empirical trading research – the anonymity of trades – to infer related trading in two securities. Empirical inference of related trading enables insights into the nature of pairs trading in practice, and the dominance of particular trading rules or strategies. The approach works by analysing the statistics of time differences between trades. If the trading in two securities is unrelated, the empirical distribution of time differences should be approximately uniform. The difference measured between the empirical distribution and a uniform distribution becomes the proxy for the amount of inferred related trading.

To verify this approach, the related trading indicators (RTIs) constructed that way are shown to be capable of distinguishing security pairs expected to be pairs traded from pseudo-random sets of similar pairs. Aggressive pairs trading indicators (APTIs) which incorporate buy and sell inferences from price data are shown to be capable of finding securities involved in merger arbitrage and pairs trading from amongst massively large sets of security pairs formed from index constituents. Passive pairs trading indicators (PPTIs) and basket trading indicators (BTIs) defined similarly to infer those kinds of trades respectively. CDF differences and K-quantile ordering scores (KQOSs) are shown to be capable of inferring dominant trading strategies by comparing the empirically inferred related trading for several possible trading rules.

Empirical inference of related trading from anonymous trading records is demonstrably capable of detecting pairs trading, detecting merger arbitrage, and inferring dominant trading rules. These abilities in turn may open more ways to study simultaneous trading empirically in two or more securities, for example to detect insider trading in merger arbitrage ahead of an announcement. Research opportunities in empirical trading appear wide open, and much work remains to be done.

References

Alexander, C. and Dimitriu, A. (2002) The cointegration alpha: Enhanced index tracking and long-short equity market neutral strategies, ISMA Discussion Papers in Finance 2002-08, ISMA Centre, University of Reading, UK.

Altucher, J. (2004) *Trade Like a Hedge Fund*, John Wiley & Sons Inc., Hoboken, New Jersey.

Bookstaber, R. (2007) *A Demon of our Own Design*, John Wiley and Sons, Hoboken, New Jersey.

Bowen, D. A., Hutchinson, M. C. and O'Sullivan, N. (2010) High-frequency equity pairs trading: transaction costs, speed of execution, and patterns in returns, *The Journal of Trading*, 5(3): 31-38.

Chen, H., S. J. Chen and Li, F. (2010) Empirical investigation of an equity pairs trading strategy, *SSRN working paper 1361293*, April 2010.

Deardorff, A. V. (1979) One-way arbitrage and its implications for the foreign exchange markets, *Journal of Political Economy*, 87(2).

Do, B. and Faff, R. (2009) Does simple pairs trading still work? *Working paper*, 10 Feb 2009.

Dunis, C. L., Giorgioni, G, Laws, J. and Rudy, J. (2010) Statistical arbitrage and high-frequency data with an application to Eurostoxx 50 equities, *working paper*, March 2010.

Froot, K.A. and Dabora, E.M. (1999) How are stock prices affected by the location of trade? *Journal of Financial Economics* 53(2), 189-216.

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

Hedge Fund Research (2011) *HFR global hedge fund industry report: Second quarter 2011*, retrieved 29th February 2012 from http://www.hedgefundresearch.com/pdf/hfr_ir_sample.pdf

Hong, G. and Susmel, R. (2003) Pairs-trading in the Asian ADR market, unpublished working paper, Saginaw Valley State University, and Bauer College of Business, University of Houston.

Jacobs, B. I. and Levy, K. (2005) *Market Neutral Strategies*, Hoboken, New Jersey, John Wiley & Sons.

Lee, C.M.C. and Ready, M.J. (1991) Inferring trade direction from intraday data, *Journal of Finance* 46(2), 733-746.

Livermore, J. L. (1940) *How to Trade in Stocks*, Duell, Sloan, and Pearce, New York.

Nath, P. (2003) High frequency pairs trading with U.S. treasury securities: risks and rewards for hedge funds, *SSRN working paper 565441*.

Paul, S. (2008) Australia eases shorting ban for dual-listed stocks, *Thomson Financial News Limited*, 23rd September 2008.

Reverre, S. (2001) The Complete Arbitrage Deskbook, McGraw-Hill, New York, pp 474-480.

Rosenthal, L. and Young, C. (1990) The seemingly anomalous price behavior of Royal Dutch / Shell and Unilever N.V./PLC, *Journal of Financial Economics* 26(1), 123-141.

Schultz, P. and Shive, S. (2010) Mispricing of dual-class shares: Profit opportunities, arbitrage, and trading, *Journal of Financial Economics*, 98(3): 524-549.

Suarez, E. D. (2005) Arbitrage opportunities in the depositary receipts market: Myth or reality? *Journal of International Financial Markets, Institutions and Money* 15: 469-480.

Wojcik, R. (2005) *Pairs trading: a professional approach*, PDF dated 19 November 2005, retrieved in 2012 from http://www.futuresindustry.org/downloads/Audio/Companion/Three-2812.pdf