Empirical Investigation of an Equity Pairs Trading Strategy

Huafeng (Jason) Chen*

Shaojun (Jenny) Chen**

Feng Li***

Abstract

We show that an equity pairs trading strategy generates large and significant abnormal returns. We find that this return is not driven purely by the short-term reversal of returns. The evidence related to the cross-sectional variation, the time-series variation, and the persistence of the pairs trading profits, and the determinants of return correlations is consistent with the delay in information diffusion as the driver for the pairs trading strategy. Evidence from the liquidity factor and the recent financial crisis suggests that the short-term liquidity provision is not the main cause of the pairs trading strategy.

Keywords: pairs trading, information diffusion, liquidity provision
JEL classifications: G12, G14

September 2012

* Assistant Professor of Finance, Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC V6T 1Z2, Canada. ** Quantitative Analyst, Connor, Clark and Lunn Investment Management, 2200-1111 West Georgia Street, Vancouver, BC V6E 4M3, Canada. *** Associate Professor of Accounting, the Stephen M. Ross School of Business, University of Michigan, 701 Tappan St, Ann Arbor, MI 48109, USA. Huafeng Chen acknowledges the support from the Social Sciences and Humanities Research Council of Canada and the UBC Bureau of Asset Management. We thank Ilia Dichev, Evan Gatev, Ron Giammarino, and the workshop participants at UBC, the Connor, Clark, and Lunn Investment Management, University of Oregon, and PanAgora Asset Management for comments. We also thank Mace Mateo and Sandra Sizer for copy-editing an earlier version of the paper. Please email inquiries to jason.chen@sauder.ubc.ca, jchen@cclgroup.com, or feng@umich.edu.

1 Introduction

Pairs trading strategy is a market neutral strategy that involves the following two steps. The first step is to identify pairs, which are trading instruments (stocks, options, currencies, bonds, etc.) that show high correlations, i.e., the price of one moves in the same direction as the other. In the second step, pairs traders look for divergence in prices between a pair. When a divergence is noticed, traders take opposite positions for instruments in a pair. In this study, we examine a pairs trading strategy based on publicly traded common equity. In equity pairs trading, the trader takes long position for underperforming stock and short position for overperforming stock. The trader then profits from the correction of the divergence.

We test an equity pairs trading strategy that uses historical return correlations to determine pairs. We first estimate the pairwise stock return correlations for all the CRSP firms for each year using return data from the previous five years. For each stock, we identify a set of pair stocks that tend to move most closely with that stock in the last five years. If a given stock's return deviates from its pair stocks in a given month, we examine whether its return converges to its pair stocks in the future and provides potential trading opportunity. We find that a trading strategy that bets on this convergence generates five-factor (market, size, book-to-market, momentum, and short-term reversal) alphas of up to 11% annually for a value-weighted self-financing portfolio, and 36% for an equal-weighted portfolio. We find that short-term reversals explain part of the pairs trading strategy returns. However, there are still substantial returns to the pairs trading strategy even when there is no significant movement in a firm's stock but its pairs have experienced significant price changes, i.e., a stock converges to its pairs even when there is no short-term reversal per se. The magnitude of the return to the pairs trading strategy also represents a significant improvement over that of the short-term reversal strategy documented by Jegadeesh (1990).

We then examine the economic drivers of this abnormal trading profit. The high trading profits provide us higher statistical power when analyzing the sources of the profit. We document several pieces of evidence that are consistent with the delay in information diffusion as the driver of our pairs trading strategy. The information delay explanation posits that when a firm and its peer deviate in stock prices, there is likely news related to the fundamentals of the pair; however, it takes time for the news to disseminate to the pair and this creates trading opportunity. This explanation is a natural candidate for explain the pairs trading profits and is one of the favored explanations in Engelberge, Gao, and Jagannathan (2009). We provide four pieces of evidence that are consistent with the delay in information diffusion explanation. First, we find that the pairs trading strategy is more profitable in firms that are small, without media coverage, lower investor recognition firms, and firms with lower analyst coverage, suggesting that the pairs trading strategy works better in environment with slower information diffusion. Second, we investigate the pairs trading returns over time. We find that the returns to the equity pairs trading strategy have diminished over time. This suggests that the exploitation of the pairs trading strategy by the statistical arbitrageurs might have reduced the effectiveness of the simple pairs trading strategy. This evidence is consistent with the joint hypothesis that information delay is the driver for pairs trading profits and that information efficiency has improved over time (consistent with the adaptive market efficiency hypothesis proposed by Lo (2004, 2005)). Third, we find that the pairs trading profits do not persist beyond the first month, suggesting that persistent fundamental risk is unlikely to explain the pairs trading strategy.

Fourth, we investigate whether the pairs trading profits are driven by known determinants of the return comovement, given that prior studies have identified a large set of variables that are correlated with stock return comovement (e.g., Fama and French (1992, 1993); Sloan (1996)). We start by exploring the characteristics of the pairs and find that there are substantial variations in the pairwise correlations: while the average Pearson correlation coefficient of two stocks' monthly returns in a five-year window is 0.22, the standard deviation is about 0.18. We then examine an array of variables that are likely determinants of the cash-flow correlations, discount-rate correlations, and systematic behavioral biases to explain the variations in pairwise return correlations. Specifically, to explain the return correlation between a pair of stocks, we include the similarity in the following variables between the stocks: earnings movement, industry membership, sales growth, size, book-to-market ratio, accruals, firm location, firm age, exchange membership, S&P Index membership, price level, durations of the expected future cash flows, financial leverage, upstream-downstream industry relation, common analyst coverage, and abnormal trading volume. Most of these variables have been shown by prior studies to be able to predict future returns or return comovement.

We find that the pairs return correlations that can be explained by these common factors are the drivers of the pairs trading strategy. Specifically, we decompose the pairs correlations by regressing pairs correlations on the list of variables discussed in the previous paragraph. The fitted values from the regression capture the pairwise returns correlations explained by these common variables ("common-factor pairs correlations") and the residuals capture the components that are orthogonal to those variables ("residual pairs correlations"). We find that a pairs trading strategy based on the common-factor (residual) pairs correlations can (cannot) generate significant abnormal returns.

Another potential explanation of the pairs trading strategy is the short-term liquidity provision. The short-term liquidity provision explanation posits that the trading profits are compensation for market makers who buy the shares of a particular stock when there is liquidity shock that leads to selling the stock relative to its peers. Engelberg, Gao, and Jagannathan (2009) find evidence consistent with both of the information delay and the short-term liquidity provision explanations. We find evidence inconsistent with short-term liquidity provision as the main driver of the pairs trading profits. In particular, we find that the pairs trading strategy is, if anything, negatively related to the traded Pastor-Stambaugh (2003) liquidity factor, and accounting for the liquidity factor increases the alpha of our trading strategy. Also, our pairs trading strategy performs poorly in the recent liquidity crisis period. The evidence suggests that short-term liquidity provision is not the main reason for the pairs trading returns.

Our paper extends the findings in Gatev, Goetzmann, and Rouwenhorst (2006), who show that there are abnormal returns from a return-based pairwise relative value trading strategy. We confirm their findings on pairs trading strategy. The advantage of our empirical approach is that it provides tradable portfolio opportunities every month, so that the capital can be fully invested all the time. Our approach also produces equal-weighted returns of 1.70% and five-factor alphas of 1.36% per month for the same universe of stocks between July 1962 and December 2002 that Gatev, Goetzmann, and Rouwenhorst (2006) use. Their method produces raw returns of 0.72% and five-factor alphas of 0.51% per month. The improvement of our strategy is partially due to the construction of the diversified pairs portfolios of 50 stocks.

Our main contribution is the examination of the economic drivers of the pairs trading strategy. Building on Engelberg, Gao, and Jagannathan (2009), who provide two explanations for the pairs trading profits, we narrow down to one. We find evidence consistent with the information diffusion explanation but inconsistent with the short-term liquidity provision explanation. The large trading profits we document help us in analyzing the sources of profits. Our paper builds on a large literature that examines the determinants of stock returns and the return comovement: Shiller (1989), Fama and French (1993), Campbell and Mei (1993), Sloan (1996), Barberis, Shleifer, and Wurgler (2005), Kumar and Lee (2006), Greenwood (2008), Pirinsky and Wang (2006), Israelsen (2009), Green and Hwang (2009), Boyer (2010), Gao (2010), and Hameed, Morck, Shen and Yeung (2010). Most of these papers rely primarily on the portfolio approach to identify the return comovement factors, and so must limit their analyses to one or two factors that may affect return comovement. Because we focus on firm-level pairwise return comovement, we can examine a comprehensive list of factors that may drive the comovement. By doing so, we identify the common factor pairs correlations as the return comovement that drives the pairs trading profits. Our paper also builds on and enriches the results in Engelberg, Gao, and Jagannathan (2009) in our common goal to uncover the economic drivers of the pairs trading profits.

The paper proceeds as follows. In Section 2, we devise a new pairs trading strategy and focus on the abnormal future returns. In Section 3, we examine the economic drivers of the abnormal returns to the strategy and find evidence consistent with the delay in information diffusion explanation. Section 4 shows two pieces of evidence that are inconsistent with the short-term liquidity provision as the main driver of the pairs trading strategy. Section 5 concludes.

2 Profitability of a Pairs Trading Strategy

2.1 A Pairs Trading Strategy

In this section, we propose and test an equity pairs trading strategy based on the historical pairwise return correlations. Essentially, this test examines whether the information contained in stock comovement is fully impounded into the prices.

We identify the pairs portfolio as follows. For each stock i in year $t+1$, we compute the Pearson correlation coefficients between the returns of stock i and all other stocks in Center for Research in Security Prices (CRSP) using monthly data from January of year t-4 to December of year t. We then find 50 stocks that have the highest correlations with stock i as its pairs.[1] In each month in year $t+1$, we compute the pairs portfolio return as the equal-weighted average return of the 50 pairs stocks, *Cret*. Our pairwise trading hypothesis is that if in any given month in year $t+1$, a stock's return, *Lret,* deviates from its pairs portfolio returns, *Cret*, then in the following month this divergence should be reversed. For example, if a stock significantly underperforms its pairs portfolio, that stock should experience abnormally higher returns in the next month.

Specifically, for stock i in a month in year $t+1$, we construct a new variable, *RetDiff*, to capture the return divergence between i's stock return and its pairs-portfolio return:

$$RetDiff = beta^C*(Cret\text{-}Rf) - (Lret - Rf),$$

where Rf is the risk free rate and $beta^C$ is the regression coefficient of firm i's monthly return on its pairs-portfolio return using monthly data between year t-4 and t.[2] The use of $beta^C$ addresses the issue of different return volatilities between the stock and its pairs portfolio.

[1] We conduct robustness tests by using 10 and 20 stocks and the empirical inferences are similar.
[2] Alternatively, we can construct the simple return difference as *Cret-Lret*. The empirical results based on this specification are similar, with comparable magnitude.

For *n* stocks, there are *n*(n-1)/2* correlations to be computed. Because the number of observations for the correlations grows exponentially with the number of stocks, the estimation is computationally intensive. To reduce the computation burden, we require that all firms have 60 monthly stock returns data from year *t--4* to year *t*.

Table 1 reports the returns of the portfolios sorted on *RetDiff*. In each month, we form ten portfolios, Decile 1 through Decile 10, based on the previous month's *RetDiff* and the holding period is one month. Our sample period is from January 1931 to December 2007. In Panel A, we report raw returns, Fama-French three-factor (market, size, and book-to-market) alphas, and five-factor (the three factors plus momentum and short-term reversal factors) alphas for the value-weighted portfolios. We use the short-term reversal to examine the pairs trading strategy returns because by construction, the sorting variable *RetDiff* contains information from a stock's lagged returns.[3]

An examination of the raw returns and alphas of the decile portfolios shows that stocks with high *RetDiff* have higher subsequent returns. For the value-weighted portfolios, the zero-cost portfolio Decile 10 – Decile 1 (i.e., longing Decile 10 and shorting Decile 1) generates a return of 1.40% per month (*t*=9.28). The hedge portfolio has a three-factor adjusted alpha of 1.23% with a t-value of 8.32 and a five-factor adjusted alpha of 0.91% (*t*=6.61). In addition to the significant hedge portfolio alphas, the alphas increase almost monotonically from Decile 1 to Decile 10, indicating that sorting on *RetDiff* systematically drives the hedge portfolio returns.

The equal-weighted portfolios generate even higher dispersion in returns. Panel B of Table 1 reports the raw returns, three-factor alphas and five-factor alphas for equal-weighted portfolios sorted by *RetDiff*. For the equal-weighted portfolios, the zero-cost portfolio Decile 10 – Decile 1 (i.e., longing Decile 10 and shorting Decile 1) generates a return of 3.59% per month (*t*=18.69). The

[3] The short-term reversal factor (*ST_Rev*) is provided by Kenneth French and is constructed as follows. Six value-weight portfolios are formed on size and prior (month t-1) returns. The portfolios, which are formed monthly, are the intersections of two portfolios formed on size (market equity, ME) and three portfolios formed on prior (t-1) return. The monthly size breakpoint is the median NYSE market equity. The monthly prior (t-1) return breakpoints are the 30th and 70th NYSE percentiles. *ST_Rev* is the average return on the two low prior return portfolios minus the average return on the two high prior return portfolios, *ST_Rev=1/2 (Small Low + Big Low) - 1/2(Small High + Big High).*

three-factor alpha for the self-financing portfolio is 3.17% per month (t=18.30). The five-factor alpha is 3.00% (t=17.76). Overall, the results in Table 1 suggest that the pairs trading strategy generates significant abnormal returns.[4]

2.2 Is Pairs Trading Driven by Short Term Reversal?

While Table 1 shows that the pairs trading strategy returns survive the five-factor model and hence are unlikely to be explained by the short-term reversal effect, we investigate whether they can be explained by the short-term reversal phenomenon more systematically in this section.

We first examine the factor-loadings of the pairs-based decile portfolios to investigate how the pairs portfolios correlate with these common factors. Table 2 reports the loadings of the pairs portfolios with respect to the five factors: market, size, book-to-market, momentum, and short-term reversal mimicking portfolios. For the value-weighted portfolios (Panel A), the self-financing portfolio (Decile 10 – Decile 1) loads positively and significantly on the market excess returns and negatively and significantly on the momentum factor, but its loadings on SMB and HML are both economically and statistically insignificant. The loading on the short-term reversal factor (ST Rev) is positive and significant (0.58 with a t-statistic of 15.10) and the magnitude is larger and more significant than the loadings on the other factors. The results based on the equal-weighted portfolios (Panel B) are similar: the self-financing portfolio loads positively on the market, SMB, HML, and especially short-term reversal and loads negatively on the momentum factor. However, the beta on the short-term reversal factor is larger and more significant compared with the betas on the other factors.

The positive loading of the pairs trading hedge portfolio on the short-term reversal mimicking portfolio suggests that the strategy partially captures the short-term reversal phenomenon. However, the fact that the pairs trading portfolios still generate significant alphas after controlling for the short-term reversal factor and the other common factors (Table 1) suggests that pairs trading strategy is not completely driven by the short-term reversal of a firm's stock returns.

---

[4] In unreported tables, we also find that consumption CAPM does not explain the profitability of this pairs trading strategy.

We directly examine this conjecture in Table 3, where we report the value- and equal-weighted portfolio returns based on a double sort of the previous month's stock return and pairs portfolio return. The holding period is also one month. Consistent with the findings in Fama (1965) and Jegadeesh (1990), stock returns exhibit a short-term reversal: stocks with low (high) returns in the previous month have high (low) returns in the current month. In all the columns, the portfolio returns in the next month decrease monotonically from the first row (i.e., firms with low lag returns) to the fifth row (i.e., firms with high lag returns). For instance, the average return of firms with low lag returns is 0.93% (t=3.61) per month, but that of firms with high lag returns is only 0.42% (t=1.38).

We also find that the pairs trading returns are incremental to the short-term reversal effect: given a stock's lagged return, if its pairs portfolio has a higher return in the previous month, then this stock is likely to have a high return this month and this effect holds within each quintile of firms sorted on lag returns. In each row of Table 3, the portfolio returns increase almost monotonically from Column 1 to Column 5 when the lag pairs-portfolio returns increase. A portfolio that longs stocks with high lag pairs-portfolio returns and shorts those with low lag pairs-portfolio returns generates positive and significant returns for each row (i.e., within each quintile portfolios sorted on lag firm returns). This shows that pairs trading abnormal returns persist even after the lagged returns are controlled for.

We further test this using a regression approach. Table 4 reports the Fama-MacBeth regressions of monthly returns on the previous month's pairs-portfolio return, *Cret*; the firm's own return in the previous month, *Lret*; and other control variables, such as size and log book-to-market ratios. For returns between July of year *t+1* and June of year *t+2*, we match with size and book-to-market equity at the fiscal year end in year *t*. For the market value of equity, we use Compustat total shares outstanding multiplied by the fiscal year-end price (data item 25*199). Size is the logarithm of the market value of equity. We construct the book value of equity as total assets minus total liabilities (Compustat data item 6-181). Book-to-market equity is then the logarithm of the ratio of the book equity to the market value of equity. Because of the data availability in Compustat, these regressions are for the sample period July 1951 to December 2007.

Columns 1 and 2 in Table 4 show that, consistent with the portfolio results in Table 1, *RetDiff* positively predicts next month's return, and that the effect is highly statistically significant, even after we control for size and book-to-market (the coefficient on *RetDiff* in Column 2 is 0.082 with a *t*-statistic of 18.72.) To examine whether the pairs trading abnormal returns are incremental to those of the short-term reversal strategy, we split *RetDiff* into its two components (*Cret* and *Lret*) and include them directly in the regressions. We find that *Cret* predicts returns positively and *Lret* predicts returns negatively. In Column 3, the coefficient on *Cret* is 0.228 (t=12.93) and that on *Lret* is -0.069 (t=-17.46). The fact that *Cret* is statistically significant even when *Lret* is included in explaining future returns suggests that there is information contained in the pairs stocks that is not driven by just the short-term reversal phenomenon. In Column 4, we add size and book-to-market to the regression. The coefficients on *Cret* and *Lret* both remain statistically significant. To conclude, we find that the short-term reversal partially, but not fully, explains the pairs trading strategy abnormal returns.

2.3 Comparison with Gatev, Goetzmann and Rouwenhorst (2006)

Our trading strategy is related to the pairs trading strategy in Gatev, Goetzmann, and Rouwenhorst (2006, in the rest of this section referred to as GGR). GGR identify pairs by minimizing the sum of squared deviations between two normalized price series in the previous 12 months. A position is opened within the next six months the day after the prices diverge by more than two historical standard deviations. The position is unwound at the next crossing of the prices, or at the end of the six-month period. They report 0.72% per month for the hedge portfolio, with a five-factor alpha of 0.51% per month.[5]

[5] GGR's hedge portfolios are essentially equal weighted. Their weights are $w_{i,t} = (1+r_{i,1})\cdots(1+r_{i,t-1})$. The weights start as equal weights and then by the cumulative stock returns. To the extent that small stocks tend to have higher returns relative to large stocks, this procedure may overweight small stocks, relative to the equal-weighting scheme.

Our trading strategy is an improvement of the GGR strategy in that we provide tradable opportunities at any given point in time. In GGR, at any given point of time, the number of pairs traded is unknown in advance and it is not clear how much capital should be invested in each pair. In addition to ease of implementation, our strategy provides much higher trading profits.

In Panel A of Table 5, we implement our trading strategy on GGR's sample period to make these two strategies comparable. Following GGR, we trade only those stocks that have positive trading volumes on each day of the previous year. We use the sample period of July 1962 to December 2002. We also skip a day between the one-month formation period and the one-month holding period. We find that our equal-weighted hedge portfolio generates raw average returns of 1.70% per month, and a five-factor alpha of 1.36%, more than twice than those in GGR.

We believe that one of the reasons that our trading strategy performs better is that we use a diversified pairs portfolio of 50 stocks, but GGR rely on one matching stock to carry out their pairs trading strategy. In Panel B of Table 5, we test whether using only one stock to form the pairs portfolio is sufficient for the pairs trading strategy. We return to our sample of January 1931 to December 2007 and value-weighted method to be comparable to our earlier results. The results indicate that our trading strategy is still profitable with a five-factor alpha of 0.43% (t=3.70). However, it is less than half of our baseline strategy's alpha of 0.91%. This result is likely due to the fact that the idiosyncratic component of one pairs stock return is too high for it to provide a better benchmark for what a stock's return should have been.

Another possible reason for our improved performance may lie in how we identify pairs. In our strategy, we use the stocks with high correlations. In GGR, pairs are determined by minimizing the sum of squared deviations between two normalized price series in the previous 12 months, and thus all pair prices are co-integrated. Our approach allows us to extract useful information from pairs with high correlations, but perhaps with very different volatilities. For example, if two stocks are perfectly correlated, but one stock return is always twice that of the other stock (perhaps driven by difference in financial leverage ratios), then GGR would not consider these two stocks as a pair. In our trading strategy we would consider them to be a pair.

2.4 Transaction Costs

Does the higher profit relative to GGR reflect higher transaction costs? To answer this question, we follow a procedure in GGR to assess the impact of transaction costs on trading profits. To do so, GGR basically compares trading profits of the strategy with and without skipping a day between the formation period and the holding period.

The logic is as follows. Consider when the trading strategy calls to buy stock *L* (with high return difference) and sell stock *S* (with low return difference) for a month and then close the position. Suppose the extreme case where, at opening the positions, stock *L* trades at bid price and stock *S* trades at ask price. If the next day's prices are equally likely to be at bid or ask, then delaying trade by one day will reduce the excess returns by half of the bid-ask spread of stock *L* plus the half of the bid-ask spread of stock *S*. If at closing the positions, stock *L* trades at the ask and stock *S* trades at the bid, waiting one day will reduce the excess returns again by one half of the sum of the bid-ask spreads of stock *L* and stock *S*. In this extreme case, waiting a day before trading reduces the return of the strategy by the sum of bid-ask spreads of the two stocks.

Table 6 reports the trading strategy's performance if we wait for one day to open and close the position. Comparing Table 6 with Table 1, we see that for the value-weighted portfolios, the average Decile 10 - Decile 1 return reduces from 1.40% per month to 0.92% per month. Interestingly, these numbers are very similar to what GGR report in their equal-weighted trading strategy in the modern sample (1963-2002). We therefore reach similar conclusions as they do, that the trading profits are both economically and statistically significant after trading costs.

The returns of the portfolios imply a value-weighted average of bid-ask spread of (1.40%-0.92%)/2=0.24%. If by waiting a day, the prices used to compute the excess return of 0.92% are equally likely to be at bid or ask, which seems a reasonable assumption, we have to correct these excess returns to reflect that in practice we buy at the ask and sell at the bid prices. In other words, we can subtract the bid-ask spreads to get an estimate of the profits after transaction costs. If we subtract 2 times the bid-ask spread of 0.24% from 0.92%, we get the 0.44% per month, which is a conservative estimate of the trading profits after trading costs.

For our equal-weighted strategy, our estimates of the trading costs are as follows. The average Decile 10 – Decile 1 return is 3.59% per month in Table 1. In Table 6, we see that, after waiting for one day, the average Decile 10 – Decile 1 return is 2.45% per month. Going through the same calculation as above, the equal weighted bid-ask spread is (3.59%-2.45%)/2 = 0.57%. Assuming that the original trading profits of 3.59% are from buying stocks at bid prices and selling short stocks at ask prices, and upon closing, selling stocks at ask prices and buying stocks back at bid prices, then we should subtract 4 times the bid-ask spreads to arrive at an estimate of the trading profits after transaction costs (that is, buying stocks at ask prices and selling short stocks at bid prices, and upon closing, selling stocks at bid prices and buying stocks back at ask prices). This means that the estimated profit after transaction costs is 3.59%-4*0.57%=1.31% per month. This is economically and statistically significant.

We note that the above analysis is conservative. The above analysis assumes that the difference between the returns in Table 1 and Table 6 are all due to bid-ask spreads. In reality, the difference may be due to movements in true prices (bid-ask midpoints) partly. Furthermore, the above analysis assumes a 100% turnover each month. In reality, although the turnover is high for our strategy, we find that for an average of 13.2% of the stocks each month, there is no change in position. This should reduce the actual transaction cost.

3 Evidence Consistent with Information Delay

One natural candidate for the explanation of the pairs trading strategy is the information delay. For example, suppose news should affect both a stock and its pairs stocks. However, if for some reason, the price is slow to adjust in that stock relative to its pairs stock, then one might expect that stock to catch up with its pairs stocks later on. Engelberg, Gao, and Jagannathan (2009) also argue that the delay in information may be an important driver in their pairs trading profits. We now provide further evidence consistent with the information delay explanation of the pairs trading strategy.

3.1 Cross-Sectional Variation in Relation to Information Environment

To test whether our pairs trading profits are driven by the delay in information diffusion, we examine the performance of our strategy as a function of the information environment: firm size, media coverage, investor recognition, and number of analyst coverage. If the delay in information diffusion is the main cause, then we expect the pairs trading profits to be larger when there is less information.

Table 7 reports the abnormal returns to the pairs trading strategy by dividing the sample into two parts based on size, media coverage, investor recognition, and analyst coverage, respectively. We measure size using the market value of equity at the portfolio formation date in the portfolio formation month. We measure media coverage as the number of news articles in three major newspapers (*Wall Street Journal*, *The New York Times*, and *USA Today*) for each firm in the 12 months before the portfolio formation date. Due to the high cost of collecting news articles from Factiva, we focus only on the three newspapers, rather than the universe of news outlets. We focus on the 1998-2007 period for the same consideration. However, given the wide influence of these three major newspapers (Soltes (2009)), we believe that this should not be an issue for the empirical tests. We also follow Lehavy and Sloan (2008) and use the breadth of ownership using the most recent 13-F data prior to the portfolio formation date to capture investor recognition. The argument is that more broad institutional ownership translates into more investor recognition. In addition, we obtain the number of analysts covering a firm in the most recent month prior to portfolio formation from I/B/E/S. Everything else equal, firms with more analyst coverage tend to have more efficient information environment.

We divide the sample into two subsamples based on each of the information environment variables and Column 1 of Table 7 shows the equal-weighted hedge portfolio returns, calculated as the difference in the 5-factor (market, size, book-to-market, momentum, and short-term reversal) model alphas of Decile 10 and Decile 1 portfolios sorted on the firm-pairs return difference, for each subsample.

The results in column 1 of Table 7 show that firms that are small, without much media coverage, and firms with low investor recognition and low analyst coverage tend to have more significant pairs trading returns. For instance, during the 1931-2007 period, the pairs trading strategy generates a hedge return of 5.08% (t=18.61) for small firms and 1.48% (t=12.33) for large firms; the difference between the two hedge returns is statistically significant. Also, the strategy generates an average monthly equal-weighted hedge return of 3.62% (t=9.00) for firms with low investor recognition. On the other hand, for firms with high investor recognition, this number is 1.90% (t=6.58) and the difference between the two groups is statistically significant.

In Column 2, we adopt a cross-sectional regression approach for each of the two subsamples by regressing next month's returns on *RetDiff* and calculate the time-series average coefficient and *t*-statistics. The results are consistent with those in Column 1, i.e., the coefficients on *RetDiff* are more positive for firms that are small, without media coverage, lower investor recognition firms, and firms with lower analyst coverage, suggesting that the pairs trading strategy works better for these firms.[6]

Therefore, the overall empirical evidence in Table 7 is consistent with the hypothesis that the pairs trading strategy abnormal returns are concentrated in firms with noisier information environment. We interpret the evidence as supporting the delay in information diffusion as the main cause of pairs trading profits.

3.2 Time Series Variation of the Pairs Trading Profits

We examine the abnormal returns to the pairs trading strategy over time. Figure 1 plots the annual returns of the value-weighted (top panel) and equal-weighted (bottom panel) hedge portfolios based on the pairs trading strategy from 1931 to 2007. With the growth of quantitative funds and

[6] In unreported results, we find that in value-weighted portfolios, there is often no statistically significant difference in trading profitability in stocks in different information environment. This is because the value-weighted trading strategy is not statistically significant to begin with after 1981. We thus focus on equal-weighted portfolios in this test.

statistical arbitraging activities over time, we expect the information diffusion to improve over time and the abnormal returns to the simple equity pairs trading strategy to decrease over time. The value-weighted hedge portfolio generates negative returns in 12 years (1941, 1957, 1973, 1981, 1993, 1996, 1999, 2000, 2001, 2003, 2005, and 2007). In contrast, the equal-weighted hedge portfolio generates returns that are greater and only lost money in one year (-9.35% in 2007). However, returns from both the value-weighted and equal-weighted portfolios appear to be smaller over time. Table 8 confirms the message from Figure 1, using a regression approach. The table presents the regressions of the annual value-weighted (Column 1) and equal-weighted (Column 2) of the hedge portfolio returns on *Year*, the calendar year. The results indicate that the coefficients on *Year* are significant and negative, suggesting that the effectiveness of the simple pairs trading strategy is diminishing. Overall, we conclude that, the strategy generates significant abnormal returns for our sample period and the returns begin to diminish over time. This piece of evidence is consistent with the joint hypothesis that the delay in information diffusion drives the pairs trading profits and information diffusion has improved over time (the adaptive market efficiency theory, Lo (2004, 2005)).

3.3 Long-Horizon Returns

To explore the persistence of the pairs trading strategy, Table 9 reports the long horizon returns for hedge portfolios (Decile 10 – Decile 1) sorted by the return difference. Panel A examine value-weighted portfolios. In the first month after portfolio formation, the pairs trading profit is 1.40% with a five-factor alpha of 0.91% (the same as Table 1). Starting in the 2nd month, the pairs trading strategy generates a loss of -0.39% with a five-factor alpha of -0.40%. In each month between the 3rd month and the 6th month, this loss persists. By the end of the six months, the loss from the pairs trading strategy exceeds the profit in the first month.

Panel B examines the equal-weighted portfolios. In the first month, the pairs trading profit is 3.59% with a five-factor alpha of 3%. In the second month, the profit reduces sharply to 0.16% and is not statistically significant. Starting in the third month, the pairs trading strategy generates a loss, although the loss by the end of the sixth month does not exceed the profit in the first month.

The results in Table 9 show that the pairs trading profits are short lived and do not persist beyond the first month. This evidence also suggests that fundamental risk-based explanation is unlikely to explain the pairs trading strategy since the fundamental risk is likely to persist longer than just one month. This piece of evidence is consistent with the information delay explanation of the pairs trading strategy.

3.4 Evidence from the Determinants of the Pairwise Stock Return Correlations

Prior studies find that many economic variables are related to the return comovement. In this section, we ask the following question: are the abnormal returns to the pairs trading strategy driven by the pairs correlations that are related to the known determinants of return comovement?

3.4.1 Determinants of Pairwise Stock Return Correlations

We start out by examining an array of variables that are likely to determine pairs correlations. Specifically, we consider variables that are potentially related to the comovement of firms' cash flows and discount rates. For example, accruals (Sloan 1996) can predict cross-sectional future stock returns. By definition, the variable is related to the comovement of stock returns. Firms with low (high) accruals tend to experience more positive (negative) stock returns in the future. This suggests that firms with similar past accruals tend to comove together in the future and form "pairs". We also consider variables that are related to investors' trading behavior that could cause stock comovement (Barberis, Shleifer, and Wurgler (2005)). To reduce computation burden, we require that firms have market equity of at least $500 million in this section. The variables we use include the following.

Earnings correlation. If the earnings of two stocks show a strong correlation, then their stock returns are more likely to be correlated. We compute the earnings correlations as follows using data from the Compustat monthly Price, Dividends, and Earnings file. For each quarter, we construct the return on equity, *ROE,* as ratio of the 12-month earnings per share to the book value of equity per share (*ERN* over *BKV*). To mitigate the influence of outliers, we set *ROEs* that are greater than 10 or less than -10 to be 10 and -10, respectively. We then compute the correlation between 20 quarterly *ROE* series of any two stocks for the period between year *t-4* and year *t*. We call this variable earnings correlation, *Earncorr*. To be included in our sample, firms must have all 20 quarterly *ROE* data between year *t-4* and year *t*. Alternatively, we construct correlations for changes in *ROEs*, and we label that variable as *Earncorr_ch*.

Earn_surprise_corr. To capture the correlations of two stocks' cash flow news, we calculate the correlation coefficient of their earnings surprises.[7] For each quarter, we measure the earnings surprise for a stock as the I/B/E/S actual quarterly earnings minus the most recent analyst forecast of the earnings divided by the book value of equity. *Earn_surprise_corr* is then computed as the Pearson correlation of the quarterly earnings surprises between a pair of stocks. We require firms to have 20 quarters of earnings surprises between year *t-4* and year *t*.

Revision_corr. We also compute the correlation between two stocks' analyst revisions of future earnings. If two stocks tend to experience similar analyst revisions around the same time (i.e., their earnings forecasts are revised upward or downward around the same time and are of similar magnitude), then they are more likely to be pairs stocks. Every month, we calculate the revision of one-quarter-ahead earnings forecasts by analysts as the current consensus earnings forecast minus last month's consensus forecast divided by the book value of equity. *Revision_corr* is the Pearson correlation of the forecast revision between a pair of stocks calculated using data from year *t-4* to year *t* for firms with at least 36 months of data in this period.

Growth. While earnings correlation captures the way in which historical earnings comove between two stocks, firm growth captures the difference in expected future earnings and therefore the difference in the growth expectation could capture the difference in the movement of stock prices. To measure the growth difference of two firms, we compute the sales growth using data in the last five years for each firm-year and calculate the absolute value of the difference in log growth rates for a pair of stocks.

[7] Earnings correlation and earnings change correlation can be viewed as parsimonious measures of earnings surprise correlations, when the expected earnings are assumed to be a constant, or the lagged earnings, respectively.

Industry. If firms are in the same industry, then they face similar business conditions, and therefore they are likely to experience the same cash-flow and discount-rate shocks. We construct dummy variables for a stock pair that equals 1 if they are from the same industry and 0 otherwise. Bhojraj, Lee, and Oler (2003) document that the Global Industry Classifications Standard (GICS) system is significantly better at explaining stock return comovements than the Standard Industrial Classifcation (SIC) or the North American Industry Classification System (NAICS). Therefore, we use GICS to construct this variable. For any two stocks that have the same eight-digit GICS industry code, we construct a dummy variable *Sgics8* that equals one, and zero otherwise. The prediction is that a pair of stocks with *Sgics8*=1 is more likely to be a pair than two stocks with *Sgics8*=0. We also set dummy variables *Sgics6*, *Sgics4*, *Sgics2* to one for stocks that have the same six-, four-, or two-digit GICS codes.

Size. If firms are similar in size, they may have similar exposures to risk factors (Fama and French (1993)). Therefore, they may have similar expected returns. Alternatively, investors may categorize assets into different styles and move funds among these styles, depending on the styles' relative performance, as suggested by Barberis and Shleifer (2003). Because size is a common style, investors' trading behavior can induce stocks of similar size to move together. Regardless of the interpretation of size as a risk factor or a mispricing factor, prior literature suggests that firms with similar size tend to comove in stock returns. We construct the absolute value difference in size for a pair of stocks as *Dsize*, which is the absolute difference in the logarithm of the market value of equity.

Book-to-market ratios. Similarly, if two firms have similar book-to-market ratios, then they may have similar exposure to some fundamental risks; alternatively, investors may view stocks with similar book-to-market ratios as having the same style (Fama and French (1993) and Boyer (2010)). We therefore construct the absolute value difference in log book-to-market ratios *Dlogbtm* and predict that a pair of firms are more likely to comove together in prices when *Dlogbtm* is smaller.

Accruals. Sloan (1996) shows that firms with low (high) accruals tend to have high (low) stock returns in the future. We construct accruals as earnings minus cash flows from operating activities divided by lagged book value of assets using Compustat annual data following Sloan (1996). We then calculate *Daccrual* as the absolute value of the difference in accruals between two firms and firms with smaller *Daccrual* should be more likely to be a pair.

Geographic location. If firms are located near one another geographically, then they may be subject to common shocks and have similar returns (Pirinsky and Wang (2006)). We use Compustat's state codes to construct a dummy variable *Sstate*, which equals one if two firms are located in the same state, and zero otherwise.

Firm age. Firms of similar ages are likely to be in similar stages of life cycle, and therefore have comovement in stock returns. We measure firm age as one plus the difference between the current year and the first year that the firm appears in the CRSP monthly data file. We then construct the absolute value difference between the logarithm of the ages of two firms *i* and *j*, as *Dage*=$|\log(\text{Age}_i)-\log(\text{Age}_j)|$.

Exchange listing. If two stocks are listed on the same exchange, then similar market microstructure issues may create stock return comovement between the two stocks. We set the variable *Slisting* to one if two stocks are listed in the same stock exchange, and zero otherwise.

S&P Index. Two stocks that belong to the same S&P major Index (e.g., Utilities, Transportation, and Financial Index) are more likely to move together (Barberis, Shleifer and Wurgler (2005)). This comovement could be driven by similar demand from index funds or due to investors' sentiment. We use the historical S&P index file and create the dummy variable *Sindex*, which equals one if two firms belong to the same S&P major index, S&P mid cap index or S&P small cap index in a given year, and zero otherwise.

Price difference. Green and Hwang (2009) show that firms with similar share prices comove with each other. Therefore, we construct a dummy variable *Dprice*, the absolute difference in the log price at the end of a given year, as a determinant of pairwise return correlations.

Duration. Firms with similar equity duration are likely to move in similar fashions because of correlations between expected future cash flows. We follow Dechow, Sloan, and Soliman (2004) and construct a variable that measures the difference of duration of the cash flows of two stocks, *Dduration*.

Financial leverage. Firms with similar financial leverage ratios (Long-Term Debt/Total Assets) may respond to changes in economic conditions in similar ways. We construct the absolute value difference in two firms' financial leverage ratios. We label this variable as *DLeverage*.

Upstream-downstream industry. Industries that have customer-supplier relations often benefit from one another's survival. Successes in the customers should lead to more order flows to the suppliers. If the market fully incorporates this information, then good news for the customer should be interpreted as good news for the supplier as well, thus creating comovement between the customer and the supplier. We follow Menzly and Ozbas (2006) and use the information from the Bureau of Economic Analysis' Input-Output Benchmark Survey to capture the upstream-downstream industry link between a pair of stocks. We construct the dummy variable *Indlink* as one if two firms' industries have customer-supplier relations, and zero otherwise.

Trading volume. Two stocks' returns could comove with each other because of similar liquidity situations. Therefore, we examine trading volume that is likely to be related to liquidity factors and investor trading behavior that may affect return comovement. For instance, Kumar and Lee (2006) find that firms with similar retail investor trading patterns tend to comove. We calculate the abnormal trading volume of a given stock by using the residual from a regression of monthly trading volume on annual trend and monthly dummies with data from the last 36 months. We calculate the correlation between two firms' abnormal trading volumes, *VolumeCorr*, and include it as a determinant of pairwise return correlation.

Common Analyst Coverage. Israelsen (2009) document that stocks with similar analysts tend to exhibit more excess comovement; Hameed, Morck, Shen, and Yeung (2010) find that stocks followed by few analysts comove significantly with firm-specific fluctuations in the prices of highly followed stocks in the same industry, but do not observe the converse. We therefore construct a variable *Rou_an* as the number of common analysts between two stocks divided by the square root of the product of the numbers of analysts that follow the two stocks. This variable captures the standardized level of overlapping analysts between two firms. We also construct a dummy variable *Drou_an* that equals one for any pair of two stocks that do not share any common financial analyst (i.e., when *Rou_an* is zero), and zero otherwise.

Table 10 provides summary statistics for the pairwise return correlations. Our sample consists of almost 1.5 million pair-years of stocks over the sample period. On average, the Pearson correlation between any two stocks' monthly returns is 0.22 and the standard deviation of this correlation is 0.18. The most negative correlation between any two stocks is -0.562, and the most positive is 0.964.

The table also presents the summary statistics of the potential determinants of the pairwise correlations. The mean value of *Sgics2* is 0.13, indicating that 13% of our sample pairs are from the same two-digit GICS code industry. The mean of the *Sstate* variable is 0.054, which suggests that about 5% of the firm-pairs are from the same state. The mean value of *Slisting* (0.74) suggests that about 74% of our sample pairs are from the same exchange. About 49% of the stock pairs belong to the same S&P 500, S&P mid cap, or S&P small cap indexes.

We estimate the regression of return correlations on the variables we hypothesize as predicting pairwise correlations. In Table 11, we present the results for the OLS pooled regression of stock return correlations on the determinants using data from 1987 to 2005. We start in 1987 because many of the determinants variables (e.g., common analyst coverage) are available only after 1982 and we need to five years of data to estimate them. To account for autocorrelation and the correlation between pairs that share a common stock, we adjust the standard errors by three-way clustering by the permno of the first stock, the permno of the second stock, and year. The three-way clustering method is based on Cameron, Gelbach, and Miller (2010).[8]

Most of the variables explain the pairwise return correlations in the ways we expect, and they have statistically significant coefficients. In Table 11, Column 1, all the four variables that capture the

[8] As an alternative statistical method, we estimate Fama-MacBeth regressions with Newey-West procedure to compute the standard errors, the results remain qualitatively similar.

correlation in earnings load up statistically significantly. Earnings correlation has a *t*-statistic of 4.42 and the correlation of changes in earnings has a *t*-statistics of 2.04. Earnings surprise correlation and earnings revision correlations are more significant with *t*-statistics of 6.47 and 7.43. However, the coefficients on these variables are only 0.024, 0.015, 0.017, and 0.091, which seem small in economic magnitude: given that both the dependent variable and independent variable are correlation coefficients, the perfect model would produce coefficients much closer to one. The R^2 of the regression is only 2.48%. This indicates that pairwise earnings correlations and differences in growth explain less than 3% of the total variation in the pairwise return correlations.

In Table 11, Column 2, we include the four same-industry dummies (*Sgics2* to *Sgics8*) in the regression. All the industry dummies show up as positive and statistically significant. For instance, the coefficient on *Sgics8* is 0.057 (t = 5.05), which suggests that if two firms are within the same eight-digit GICS industry, then compared with a pair of firms that are in the same six-digit GICS industry but not in the same eight-digit GICS industry, their return correlation is higher by 0.057. Compared with two firms that do not share any industry membership, a pair of firms that are in the same eight-digit GICS industry have a return correlation coefficient higher by 0.202 (0.057+0.024+0.041+0.080). The R^2 also increases significantly when we include the industry dummies (the R^2 in Column 2 is 7.23%.)

In Column 4, to further explain pairwise return correlations, we include the distances in firm size and book-to-market ratio (*Dsize* and *Dlogbtm*), the distance in accruals (*Daccrual*), the same-state dummy (*Sstate*), the distance in firm age (*Dage*), the same exchange listing dummy (*Slisting*), the same S&P index membership (*Sindex*); the distance in log prices (*Dprice*), durations (*Dduration*), and financial leverage ratios (*Dleverage*); the industrial customer-supplier link (*Indlink*), abnormal volume correlations (*Volumecorr*), common analyst coverage (*Rou_an*), and a dummy variable for missing common analyst coverage (*Drou_an*). Not surprisingly, the further the distance in size, book-to-market equity, accrual, age, and financial leverage, the lower is the correlation in stock returns of any two firms. All else equal, two firms located in the same state, listed in the same exchange list, in the same stock index, with similar trading volume patterns, and with common analyst coverage tend to comove in stock returns. Three factors show up as statistically insignificant: *Dprice*, the absolute value difference in log prices, *Dduration*, the distance in durations, and *Indlink*, the industrial customer-supplier link.

Overall, even when we include all the determinants in Column 4, there is substantial variation in the return comovement that cannot be explained; when we include all the explanatory variables, the R^2 is still less than 12%.

3.4.2 Portfolios Formed on Fitted and Residual Correlations

To explore the economic drivers of the pairs trading profits, we now examine the trading profits based on different criteria in constructing the pairs portfolio. Specifically, we explore constructing the pairs portfolio according to the total pairwise return correlation, the pairwise return correlation that can be explained by the common factors in Column 4 of Table 11 (fitted correlation), or the correlation that cannot be explained (residual correlation). Decomposing the pairs into those that can be explained by common factors (such as size, book-to-market, and accruals) versus those that cannot sheds light on the sources of the abnormal returns.

The results are reported in Table 12. This table reports the value-weighted returns for portfolios that we form on the return difference (*RetDiff*) on the sample of firms that have valid observations to estimate the regression in Column 4 of Table 11. In Panel A, the pairs portfolio is the 50 stocks with highest correlations. In Panel B, the pairs portfolio is the 50 stocks with highest fitted correlations in Column 4 of Table 11. In Panel C, the pairs portfolio is the 50 stocks with highest residual correlations in Column 4 of Table 11.

In Panel A of Table 12, the five-factor alpha of returns based on total return correlation is 0.50%, and is close to being statistically significant. The trading profits are lower than that reported in Table 1, which is mostly due to the earlier sample period.

In Panel B of Table 12, we find that the five-factor alpha of returns based on fitted return correlation is 0.66% and is statistically significant (t=2.13). Interestingly, in Panel C, the five-factor alpha of returns based on the residual return correlation is only 0.25% and is statistically insignificant (t=0.81). These results indicate that the pairs trading profits are driven by the correlation structure that is captured by our correlation prediction model in Column 4 of Table 11.

We make three remarks on the results on the common-factor pairs correlations. First, they imply that our model that explains the return correlation using common factors works reasonably well. Although the R^2 in that model is relatively low, the determinants capture important information contained in the return comovement. If the model were simply noise, then we would not expect the trading strategy based on the fitted correlations to be profitable. Second, the results do not imply that pairs trading profits are driven by the premiums associated with the common factors, for example, the value premium. We explicitly adjust for risk premiums associated with the common factors including the value premium and find that the trading profits survive the common risk adjustment. Finally, the results suggest that "economic" factors help identify peer stocks that are "fundamentally" similar. Any price deviation from the resulting peer (based on "economic" factors) is therefore more likely due to "non-fundamental" factors and thus more likely to be reverted in the near future. In other words, the pairs trading strategy enhances short-term reversal by better isolating past price movement that is due to "non-fundamental" factors. This is consistent with the notion that pairs trading strategy is due to information inefficiency (consistent with the delay information diffusion).

4 Does Short-Term Liquidity Provision Explain the Profitability of the Pairs Trading Strategy?

Engelberg, Gao, and Jagannathan (2009) argue that their pairs trading strategy may be driven by both the delay in information diffusion and the short-term liquidity provision. These two explanations are hard to disentangle. For example, we have argued that evidence in Section 3 is consistent with the delay in information diffusion hypothesis. However, much of the evidence can be consistent with the short-term liquidity provision hypothesis as well. In this section, we document two pieces of evidence that are against the short-term liquidity provision as the *main* cause of the pairs trading strategy.

4.1 Evidence from the Pastor-Stambaugh Liquidity Factor

One way to test the liquidity provision explanation of the pairs trading strategy is to examine the exposure of the trading returns to the liquidity factor. In Table 13, we report alphas and betas of the hedge portfolios (Decile 10 – Decile 1) relative to standard factors (1, 3, and 5 factors) and the Pastor-Stambaugh liquidity factor. We focus on the Pastor-Stambaugh value-weighted traded liquidity factor (PS_VWF) provided by WRDS. Panel A reports results for the value-weighted pairs trading hedge portfolio and Panel B for the equal-weighted hedge portfolio.

Column 1 of Panel A reports the alpha and beta of the value-weighted hedge portfolio relative to 1 factor (the market). Consistent with earlier findings, the hedge portfolio has a large alpha of 0.87% (t=4.22) and a relatively small market beta (0.30). Column 2 adds the Pastor-Stambaugh liquidity factor relative to Column 1. The results show that 1) The pairs trading strategy appears to be largely unrelated to the Pastor-Stambaugh liquidity factor (in fact, they are slightly negatively related). 2) Because the Pastor-Stambugh liquidity factor is positive on average, the slightly negative loading on the liquidity factor means that the liquidity factor makes the alpha slightly larger (increases from 0.87% to 0.88%). These two results are robust to the use of 3 (Columns 3 and 4), 4 (Columns 5 and 6), and 5 (Columns 7 and 8) factors. They are also robust to the use of equal-weighted portfolios (in Panel B).

The evidence in Table 13 show that the pairs trading strategy has little to do with the Pastor-Stambaugh liquidity factor. If anything, controlling for the Pastor-Stambaugh liquidity factor increases the alpha of the pairs trading strategy. This evidence does not support liquidity provision as the main cause of the pairs trading profits.

4.2 Evidence from the Recent Liquidity Crisis

We now provide another test for the short-term liquidity provision hypothesis. If the pairs trading profits are primarily driven by short-term liquidity provision, then the profits should be higher in times with low liquidity. The recent financial crisis in 2008 potentially provides an interesting setting to examine this issue. We therefore examine the performance of our strategy in this sample period.

The results of our trading profits between 2008 and 2010 are reported in Table 14. Because of the short sample, we find that decile portfolios exhibit nonmonotonicity in returns, we therefore focus on quintile portfolios. For the value-weighted portfolios, the zero-cost portfolio Quintile 5 – Quintile 1 generates a return of 1.18% per month (t=2.55). The hedge portfolio has a three-factor adjusted alpha of 1.08% with a t-value of 2.46 and a five-factor adjusted alpha of 0.97% (t=2.29). For the equal-weighted portfolios, the zero-cost portfolio Quintile 5 – Quintile 1 generates a return of 2.09% per month (t=3.83). The three-factor alpha for the self-financing portfolio is 1.99% per month (t=3.62). The five-factor alpha is 1.67% (t=4.10). Overall, the results in Table 14 suggest that the pairs trading strategy generates significant abnormal returns between 2008 and 2010. They also provide an out-of-sample test for our trading strategy as the first draft of this paper predates much of this sample period.

Figure 2 plots the cumulative hedge portfolio (Quintile 5 – Quintile 1) return and Amihud (2002) illiquidity between 2008 and 2010. To mitigate the effects of different stocks exchanges on liquidity measures, we use all stocks in our sample that are listed on NYSE. The Amihud illiquidity measure is computed as follows. We first compute the Amihud measure for each stock in each month. It is calculated as the daily ratio of the absolute value of the return and dollar volume, averaged over all trading days during a month. For ease of disposition, we multiply the measure by 10^6. We then average the measure across stocks in each month. Note that a higher value of the Amihud measure signifies higher illiquidity, as a particular dollar volume traded is associated with a relatively high price movement.

The Amihud measure reaches its highest levels in November 2008 and March 2009. This is consistent with the anecdotal evidence that the market liquidity dries up significantly after the Lehman Brother bankruptcy in September 2008. The top panel of Figure 2 plots the value-weighted cumulative hedge portfolio return and the Amihud measure. Interestingly, the pairs trading strategy performs poorly between August 2008 and March 2009. The equal-weighted cumulative hedge return in the bottom panel also shows that the pairs trading profits are somewhat lower in the periods when liquidity is the lowest.[9] We believe the evidence suggests that the short-term liquidity provision is not the main driver of the pairs trading profits.

5. Conclusion

In this paper, we first extend the results in Gatev, Goetzmann, and Rouwenhorst (2006) by showing that a pairs trading strategy can generate significant abnormal returns. We show that our pairs trading strategy is partly, but not fully driven by the short-term reversal. We also show that the strategy survives the consideration of transaction cost.

We then explore the economic drivers of the abnormal returns to an equity pairs trading strategy. We provide several pieces of evidence that are consistent with the delay in information diffusion explanation of the pairs trading. We show that the profitability of the pairs trading strategy is higher in poorer information environment, has declined over time, does not persist beyond one month, and the strategy is more profitable when we identify stocks with the highest fitted correlations (rather than residual correlations) as pairs.

Finally, we find evidence that suggests short-term liquidity provision is not the main cause of the pairs trading strategy. Our pairs trading strategy is slightly negatively related to the standard liquidity factor. Our pairs trading strategy performs poorly in the recent financial crisis. The evidence suggests that short-term liquidity provision is not the main driver of the pairs trading profits.

[9] Results are qualitatively the same if we look at three-factor alphas or five-factor alphas.

References

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.

Barberis, Nicholas, and Andrei Shleifer, 2003, Style investing, *Journal of Financial Economics* 68, 161-199.

Barberis, Nicholas, Andrei Shleifer, and Jeffery Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283-317.

Bhojraj, Sanjeev, Charles M. Lee, and Derek K. Oler, 2003, What's my line? A comparison of industry classification schemes for capital market research, *Journal of Accounting Research* 41, 745-774.

Boyer, H. Brian, 2010, Style related comovement: fundamentals or labels?, *Journal of Finance*, forthcoming.

Cameron, A. Colin, Jonah Gelbach and Douglas Miller, 2010, Robust inference with multi-way clustering, *Journal of Business and Economic Statistics*, forthcoming.

Campbell, John Y., and Jianping Mei, 1993, Where do betas come from? Asset price dynamics and the sources of systematic risk, *Review of Financial Studies* 6(3), 567-592.

Dechow, Patricia, Richard Sloan, and Mark Soliman, 2004, Implied equity duration: A new measure of equity risk, *Review of Accounting Studies* 9, 197-228.

Engelberg, Joseph, Pengjie Gao, Ravi Jagannathan, 2009, An anatomy of pairs trading: The role of idiosyncratic news, common information and liquidity, Working Paper, UNC.

Fama, Eugene F., 1965, The behavior of stock market prices, *Journal of Business* 38, 34-105.

Fama, Eugene F., and Kenneth R. French, 1992, The Cross Section of Expected Stock Returns, *Journal of Finance* 47, 427-465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Gao, Pengqin, 2010, Characteristic-based covariances and cross-sectional expected returns, working paper, Cornell University.

Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst, 2006, Pairs trading: Performance of a relative value arbitrage rule, *Review of Financial Studies* 19, 797-827.

Green, T. Clifton, and Byoung-Hyoun Hwang, 2009, Price-based return comovement, *Journal of Financial Economics* 93, 37-50.

Greenwood, Robin, 2008, Excess comovement: Evidence from cross-sectional variation in Nikkei 225 weights, *Review of Financial Studies* 21, 1153-1186.

Hameed, Allaudeen, Randall Morck, Jianfeng Shen, and Bernard Yeung, 2010, Information, analysts, and stock return comovement, Working Paper, National University of Singapore.

Israelsen, Ryan, 2009, Does correlated analyst coverage explain excess comovement? Working Paper, Indiana University.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *Journal of Finance* 45, 881-898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: implications for stock market efficiency, *Journal of Finance*, 65-91.

Kallberg, Jarl and Paolo Pasquariello, 2007, Time-series and cross-sectional excess comovement in stock indexes, *Journal of Empirical Finance 15*, 481-502.

Kumar, Alok, and Charles M.C. Lee, 2006, Retail investor sentiment and return comovements, *Journal of Finance* 61, 2451-2486.

Lo, Andrew, 2004, The adaptive markets hypothesis: market efficiency from an evolutionary perspective. *Journal of Portfolio Management* 30, 15–29.

Lo, Andrew, 2005, Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. *Journal of Investment Consulting* 7, 21–44.

Menzly, Lior, and Oguzhan Ozbas, 2006, Cross-industry momentum, Working Paper, University of Southern California.

Pastor, Lubos, and Robert Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.

Pirinsky, Christo A. and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, *Journal of Finance* 61, 1991-2015.

Roll, Richard, 1988, R^2, *Journal of Finance* 43, 541-566.

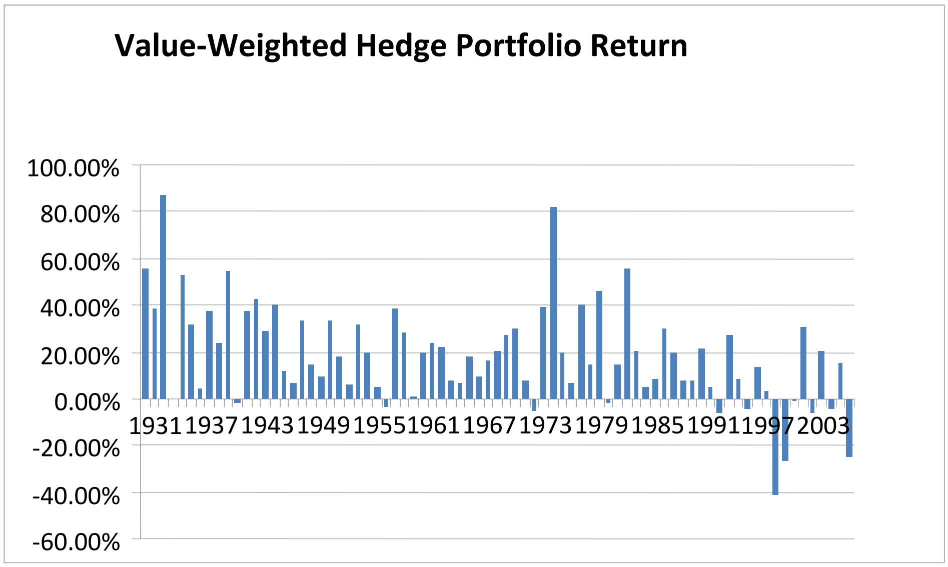
Sloan, Richard, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings, *The Accounting Review* 71, 289-315.

Soltes, Eugene F., 2009, News dissemination and the impact of the business press, Ph.D. dissertation, University of Chicago.

Shiller, Robert, 1989, Comovements in stock prices and comovements in dividends, *Journal of Finance* 44, 719-729.

Figure 1: Hedge Portfolio Return between 1931 and 2007

This figure plots the value-weighted (top panel) and equal-weighted (bottom panel) self-financing portfolio (Decile 10 – Decile 1) returns for the portfolios that are formed on return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year t+1, the pairs portfolio is the equal weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year t-4 and year t. *Lret* is the previous month's stock return. *RetDiff* is $beta^{C}$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^{C}$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The sample period is 1931 to 2007.



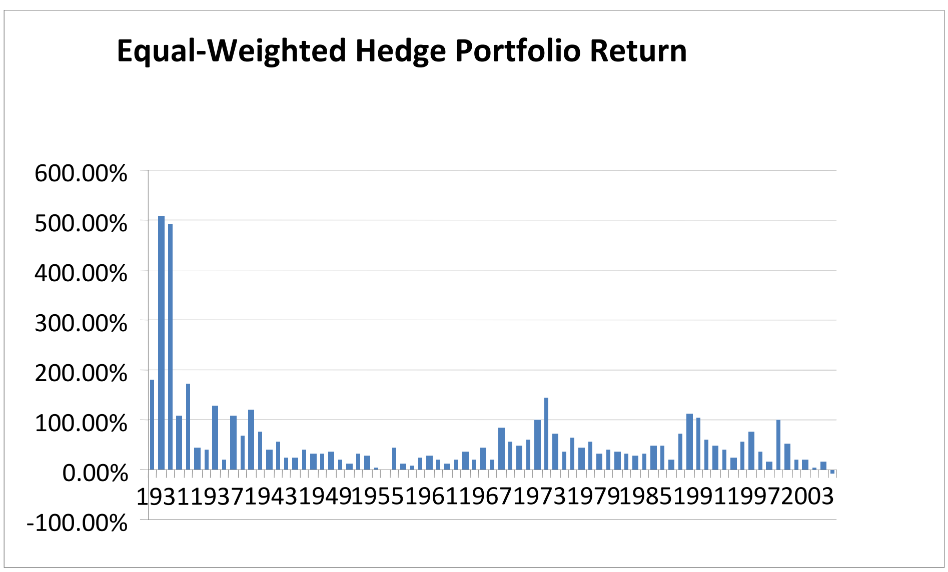


Figure 2: Cumulative Hedge Portfolio Return and Illiquidity Between 2008 and 2010

This figure plots the cumulative value-weighted (top panel) and equal-weighted (bottom panel) self-financing portfolio (Quintile 5 – Quintile 1) returns for the portfolios that are formed on return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year t+1, the pairs portfolio is the equal weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year t-4 and year t. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The sample period is 2008 to 2010. We also plot the average Amihud illiquidity measure (multiplied by 10^6) across all NYSE stocks in our sample.

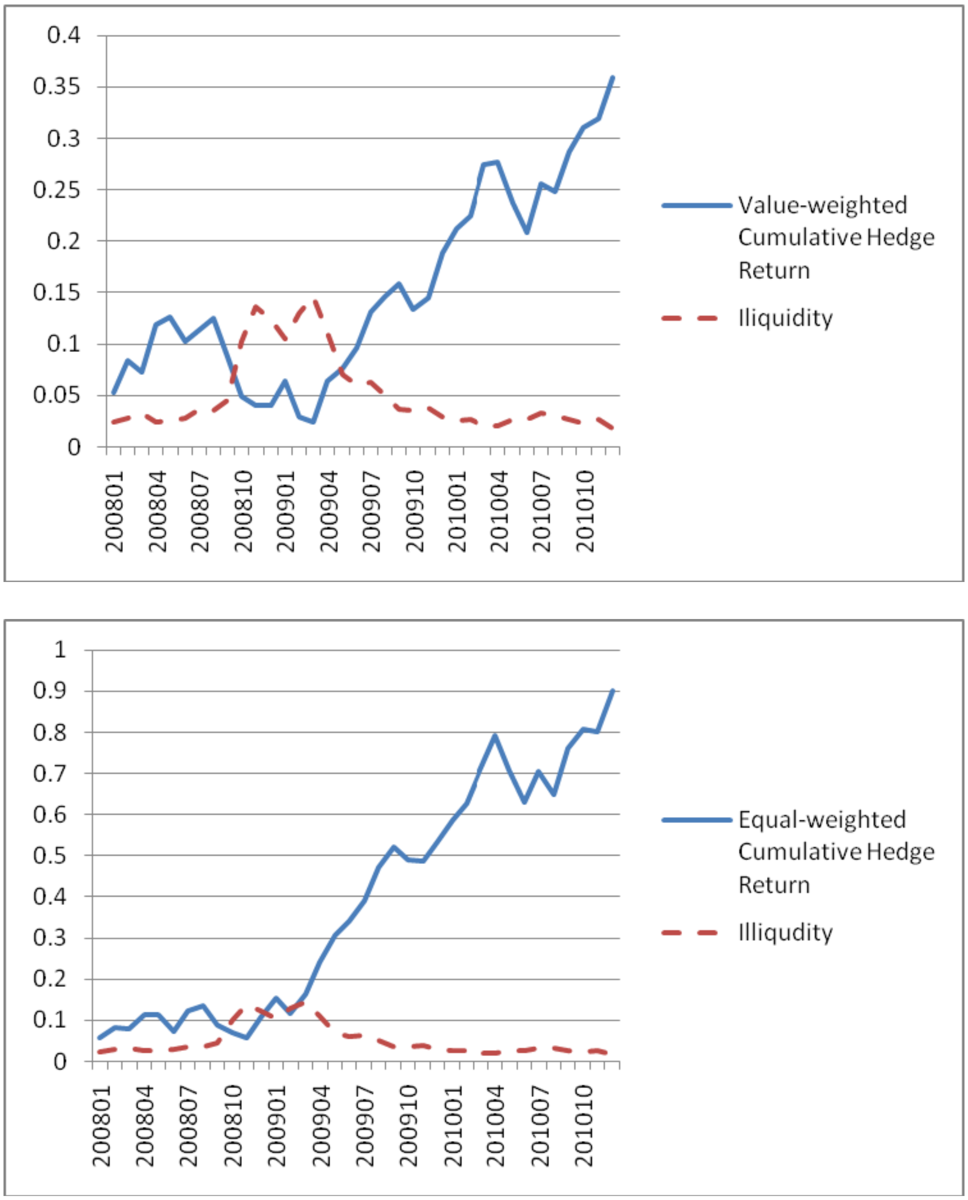


Table 1: Portfolios Formed on Return Difference

This table reports the value- and equal-weighted returns for portfolios that we form on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The three factors are excess market return, SMB, and HML. The five factors are the three factors, plus the momentum factor, and the short-term reversal factor. Panel A reports value-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. Panel B reports equal-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. The sample period is January 1931 to December 2007.

Panel A: Value-weighted portfolios

	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.45%	-0.70%	-0.45%
2	0.65%	-0.43%	-0.17%
3	0.74%	-0.26%	-0.13%
4	0.93%	-0.03%	0.04%
5	0.97%	0.04%	0.04%
6	1.17%	0.24%	0.18%
7	1.16%	0.18%	0.10%
8	1.35%	0.33%	0.26%
9	1.53%	0.39%	0.39%
Decile 10	1.86%	0.52%	0.46%
Decile 10-Decile 1	1.40%	1.23%	0.91%
t-statistics	(9.28)	(8.32)	(6.61)

Panel B: Equal-weighted portfolios

	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.00%	-1.45%	-1.15%
2	0.61%	-0.74%	-0.53%
3	0.94%	-0.36%	-0.22%
4	1.07%	-0.20%	-0.14%
5	1.22%	-0.03%	-0.01%
6	1.37%	0.11%	0.05%
7	1.56%	0.25%	0.16%
8	1.78%	0.41%	0.32%
9	2.22%	0.69%	0.68%
Decile 10	3.59%	1.72%	1.85%
Decile 10-Decile 1	3.59%	3.17%	3.00%
t-statistics	(18.69)	(18.30)	(17.76)

Table 2: Time-Series Test

This table reports the factor loadings for portfolios that we form on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The five factors are the excess market return, SMB, HML, the momentum factor, and the short-term reversal factor. The sample period is January 1931 to December 2007.

Panel A: Value-weighted portfolios						
	Alpha	Rm-Rf	SMB	HML	MOM	ST Rev
Decile 1	-0.45%	1.13	0.28	0.08	0.00	-0.31
2	-0.17%	1.10	-0.01	0.09	-0.04	-0.27
3	-0.13%	1.01	-0.09	0.10	-0.03	-0.12
4	0.04%	0.98	-0.12	0.09	-0.02	-0.08
5	0.04%	0.96	-0.15	0.05	-0.02	0.02
6	0.18%	0.94	-0.08	0.05	-0.01	0.08
7	0.10%	0.98	-0.10	0.07	-0.01	0.12
8	0.26%	1.01	0.01	0.01	-0.04	0.15
9	0.39%	1.10	0.09	0.04	-0.12	0.17
Decile 10	0.46%	1.25	0.32	0.07	-0.14	0.27
Decile 10 - Decile 1	0.91%	0.12	0.04	-0.02	-0.14	0.58
t-statistics	(6.61)	(4.36)	(0.96)	(-0.46)	(-4.39)	(15.10)

Panel B: Equal-weighted portfolios						
	Alpha	Rm-Rf	SMB	HML	MOM	ST Rev
Decile 1	-1.15%	1.02	1.01	0.31	-0.14	-0.20
2	-0.53%	1.04	0.62	0.33	-0.10	-0.13
3	-0.22%	1.01	0.52	0.32	-0.09	-0.06
4	-0.14%	0.99	0.43	0.35	-0.06	0.01
5	-0.01%	0.96	0.41	0.33	-0.09	0.09
6	0.05%	0.93	0.52	0.33	-0.06	0.16
7	0.16%	0.99	0.54	0.35	-0.06	0.19
8	0.32%	1.01	0.64	0.36	-0.09	0.23
9	0.68%	1.06	0.90	0.41	-0.16	0.22
Decile 10	1.85%	1.13	1.40	0.57	-0.37	0.32
Decile 10 - Decile 1	3.00%	0.11	0.39	0.26	-0.23	0.52
t-statistics	(17.76)	(3.34)	(7.86)	(5.49)	(-5.97)	(11.09)

Table 3: Portfolio Returns Sorted on Previous Month's Return and Previous Month's Pairs Portfolio Return

This table reports the average returns for 25 portfolios formed on lagged return and lagged pairs portfolio return between 1931 and 2007. For each month in year *t+1*, we form a portfolio based on the previous month's stock return and pairs portfolio return, and held for one month. The pairs portfolio is the equal-weighted portfolio of 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *t*-statistics are reported below the average returns.

Panel A: Value weighted

	Low Lag Pairs Return	2	3	4	High Lag Pairs Return	L/S
Low Lag Return	0.93%	1.54%	1.90%	2.35%	2.47%	1.71%
	(3.61)	(6.02)	(7.12)	(8.12)	(8.27)	(7.64)
2	0.84%	1.15%	1.39%	1.70%	2.10%	1.26%
	(3.83)	(5.52)	(6.87)	(7.42)	(8.21)	(6.19)
3	0.57%	1.00%	0.91%	1.37%	1.69%	1.12%
	(2.61)	(5.00)	(4.50)	(7.14)	(8.00)	(6.45)
4	0.35%	0.76%	0.98%	1.18%	1.34%	0.99%
	(1.61)	(3.71)	(4.87)	(5.76)	(5.91)	(5.54)
High Lag Return	0.42%	0.62%	0.70%	0.86%	0.95%	0.56%
	(1.38)	(2.37)	(2.96)	(3.76)	(4.07)	(2.47)

Panel B: Equal weighted

	Low Lag Pairs Return	2	3	4	High Lag Pairs Return	L/S
Low Lag Return	1.96%	2.43%	2.70%	3.09%	3.20%	1.73%
	(6.23)	(7.96)	(8.36)	(8.95)	(9.16)	(8.00)
2	0.79%	1.43%	1.77%	1.90%	2.51%	1.76%
	(3.25)	(5.82)	(7.13)	(7.52)	(8.55)	(9.25)
3	0.66%	1.08%	1.43%	1.60%	2.12%	1.46%
	(2.86)	(4.74)	(6.32)	(7.32)	(8.54)	(9.74)
4	0.33%	0.96%	1.13%	1.33%	1.80%	1.47%
	(1.36)	(4.01)	(4.87)	(5.68)	(7.00)	(9.25)
High Lag Return	0.10%	0.43%	0.54%	0.59%	0.86%	0.73%
	(0.30)	(1.56)	(2.11)	(2.38)	(3.27)	(2.97)

Table 4: Fama-MacBeth Regressions of Monthly Returns

This table reports the Fama-MacBeth regressions of monthly returns on lagged variables. *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C*(Cret-Rf) - (Lret - Rf)$, where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. For returns between July of year *t+1* and June of year *t+2*, we match with *Size* and book-to-market equity at the fiscal year end in year *t*. The market value of equity is Compustat total shares outstanding multiplied by the fiscal year-end price (25*199). *Size* is the logarithm of the market value of equity. The book value of equity is the total assets minus total liabilities (6-181). *Logbtm* is the logarithm of the ratio of the book equity to the market value of equity. All the regressions are for the sample period July 1951 to December 2007. *t*-statistics are reported below the coefficients.

	1	2	3	4
RetDiff	0.082	0.082		
	(18.14)	(18.72)		
Cret			0.228	0.191
			(12.93)	(12.45)
Lret			-0.069	-0.073
			(-17.46)	(-18.04)
Size		-0.001		-0.001
		(-2.66)		(-2.81)
Logbtm		0.002		0.002
		(3.98)		(4.19)
Avg. Obs.	1994	1994	1994	1994
Avg. R^2	0.011	0.038	0.025	0.047

Table 5: Returns to Pairs Trading Strategy (Comparison with GGR strategy)

This table reports the returns for portfolios that we form on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks (Panel A) or one stock (Panel B) that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. In Panel A, we form equal-weighted portfolios on stocks that have positive trading volumes every day in the previous 12 months. We skip one day between the one-month formation period and the one-month holding period. In Panel B, we form value-weighted portfolios on all stocks that have 60 monthly returns in the most recent five years. The three factors are excess market return, SMB, and HML. The five factors are the three factors, the momentum factor, and the short-term reversal factor. The sample period is July 1962 to December 2002 for Panel A, and January 1931 to December 2007 for Panel B.

Panel A: Portfolios are for stocks traded every day, skip one day, equal-weighted

	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.22%	-0.93%	-0.60%
2	0.48%	-0.70%	-0.51%
3	0.64%	-0.53%	-0.39%
4	0.83%	-0.33%	-0.26%
5	1.03%	-0.16%	-0.06%
6	1.21%	0.00%	0.06%
7	1.33%	0.10%	0.18%
8	1.46%	0.21%	0.30%
9	1.69%	0.41%	0.53%
Decile 10	1.91%	0.54%	0.76%
Decile 10-Decile 1	1.70%	1.47%	1.36%
t-statistics	(8.32)	(7.20)	(8.52)

Panel B: Only one stock in the pairs portfolio, value-weighted

	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.59%	-0.58%	-0.24%
2	0.59%	-0.43%	-0.27%
3	0.80%	-0.18%	-0.04%
4	0.82%	-0.16%	-0.04%
5	0.89%	0.00%	0.00%
6	1.04%	0.10%	0.09%
7	1.19%	0.24%	0.13%
8	1.27%	0.28%	0.10%
9	1.40%	0.35%	0.18%
Decile 10	1.54%	0.34%	0.19%
Decile 10-Decile 1	0.95%	0.92%	0.43%
t-statistics	(7.22)	(6.95)	(3.70)

Table 6: Portfolios Formed on Return Difference (Skip a Day)

This table reports the value- and equal-weighted returns for portfolios that we form on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. We skip one day between the one-month formation period and the one-month holding period. The three factors are excess market return, SMB, and HML. The five factors are the three factors, plus momentum factor, and short-term reversal factor. Panel A reports value-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. Panel B reports equal-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. The sample period is January 1931 to December 2007.

Value weighted	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.68%	-0.46%	-0.15%
2	0.75%	-0.31%	-0.02%
3	0.78%	-0.19%	-0.06%
4	0.95%	0.02%	0.09%
5	0.97%	0.07%	0.04%
6	1.12%	0.21%	0.15%
7	1.12%	0.16%	0.09%
8	1.26%	0.29%	0.24%
9	1.39%	0.28%	0.27%
Decile 10	1.60%	0.32%	0.27%
Decile 10-1	0.92%	0.78%	0.42%
t-statistics	(6.11)	(5.26)	(2.98)

Equal weighted	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.40%	-1.05%	-0.71%
2	0.82%	-0.52%	-0.27%
3	1.06%	-0.22%	-0.05%
4	1.15%	-0.09%	-0.01%
5	1.28%	0.06%	0.08%
6	1.37%	0.14%	0.10%
7	1.48%	0.20%	0.13%
8	1.69%	0.35%	0.30%
9	1.99%	0.49%	0.55%
Decile 10	2.84%	1.05%	1.24%
Decile 10-1	2.45%	2.11%	1.95%
t-statistics	(15.70)	(14.97)	(14.08)

Table 7: Pairs Trading Strategy for Firms with Different Size, Media Coverage, Investor Recognition, and Analyst Coverage

This table reports the pairs trading strategy return as a function of firm media coverage, investor recognition, and analyst coverage. Column 1 shows the average monthly hedge portfolio returns (Decile 10 – Decile 1) for the equal-weighted portfolios. Column 2 reports the average coefficients on *RetDiff* in the monthly cross-sectional regression of returns on *RetDiff*, where *RetDiff* is defined as in Table 4. Small (large) firms are those firms with a market value of equity below (above) median value in a cross-section. Firms with (without) media coverage are those firms that have at least one (do not have any) coverage by *Wall Street Journal, The New York Times,* and *USA Today* in the portfolio formation month. Low (high) investor recognition firms are those that have investor recognition below (above) the median value in a cross-section, where investor recognition is calculated following Lehavy and Sloan (2008). Firms with low (high) analyst coverage are those firms with the number of analysts following below (above) median in a cross-section based on the I/B/E/S data; if a firm is not included in I/B/E/S, it is assumed to have zero analyst coverage. T-statistics are in parentheses. The "Yes/No" indicator indicates whether the difference in hedge returns or coefficient on *RetDiff* is statistically significant at the 0.10 level.

	1 Pairs trading hedge portfolio return (EW)	2 Cross-sectional regression coefficient (EW)
Small firms (1931-2007)	5.08%	0.105
	(18.61)	(21.04)
Large firms (1931-2007)	1.48%	0.057
	(12.33)	(14.49)
Difference statistically significant?	Yes	Yes
Firms without media coverage in the portfolio formation month (1998-2007)	3.80%	0.024
	(3.53)	(3.02)
Firms with media coverage in the portfolio formation month (1998-2007)	0.07%	0.004
	(0.05)	(0.56)
Difference statistically significant?	Yes	Yes
Firms without media coverage in the portfolio formation month (1998-2007)	3.80%	0.024
	(3.53)	(3.02)
Firms with media coverage in the portfolio formation month (1998-2007)	0.07%	0.004
	(0.05)	(0.56)
Difference statistically significant?	Yes	Yes
Low investor recognition firms (1981-2007)	3.62%	0.055
	(9.00)	(10.57)
High investor recognition firms (1981-2007)	1.90%	0.043
	(6.58)	(7.79)

Difference statistically significant?	Yes	Yes
Firms with low analyst coverage (1982-2007)	3.80%	0.057
	(9.06)	(10.35)
Firms with high analyst coverage (1982-2007)	1.70%	0.038
	(5.75)	(6.57)
Difference statistically significant?	Yes	Yes

Table 8: Regressions of the Hedge Portfolio Returns on Year

This table presents the regression results of the annual value-weighted (Column 1) and equal-weighted (Column 2) hedge portfolios based on the pairs trading strategy on the calendar year. The sample period is 1931 to 2007. T-statistics are in parentheses.

	1	2
Intercept	8.620	25.278
	(4.41)	(3.24)
Year	-0.004	-0.013
	(-4.31)	(-3.17)
Obs.	77	77
Adjusted R^2	0.19	0.11

Table 9: Long Horizon Returns of Hedge Portfolios Sorted by Return Difference

This table reports the value- and equal-weighted returns for hedge portfolios (Decile 10 – Decile 1) that we form on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C*(Cret\text{-}Rf) - (Lret - Rf)$, where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The three factors are excess market return, SMB, and HML. The five factors are the three factors, plus the momentum factor, and the short-term reversal factor. Month 1 is the first month after portfolio formation and Month 2 is the second month after portfolio formation, etc. Panel A reports value-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. Panel B reports equal-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. The sample period is January 1931 to December 2007.

Panel A: Value-weighted portfolios			
Month	Raw Return	3-Factor Alpha	5-Factor Alpha
1	1.40%	1.23%	0.91%
	(9.28)	(8.32)	(6.61)
2	-0.39%	-0.51%	-0.40%
	(-2.64)	(-3.43)	(-2.67)
3	-0.50%	-0.71%	-0.37%
	(-3.20)	(-4.78)	(-2.42)
4	-0.21%	-0.44%	-0.39%
	(-1.20)	(-2.63)	(-2.25)
5	-0.18%	-0.44%	-0.13%
	(-1.04)	(-2.65)	(-0.75)
6	-0.54%	-0.80%	-0.49%
	(-3.04)	(-4.76)	(-2.79)
Panel B: Equal-weighted portfolios			
Month	Raw Return	3-Factor Alpha	5-Factor Alpha
1	3.59%	3.17%	3.00%
	(18.69)	(18.30)	(17.76)
2	0.16%	0.04%	0.09%
	(1.06)	(0.24)	(0.55)
3	-0.52%	-0.61%	-0.40%
	(-4.10)	(-4.85)	(-3.09)
4	-0.14%	-0.40%	-0.29%
	(-0.96)	(-3.03)	(-2.13)
5	-0.06%	-0.20%	-0.16%
	(-0.43)	(-1.52)	(-1.23)
6	-0.59%	-0.75%	-0.45%
	(-4.96)	(-6.52)	(-3.85)

Table 10: Summary Statistics

This table reports the summary statistics for the determinants of stock return correlations between 1987 and 2005. Each observation represents a stock pair year. *Corr* is the stock return correlation between two stocks using monthly returns between year *t-4* and *t*. We calculate earnings correlation (*Earncorr*) from the Compustat Price, Dividends, and Earnings data set as follows. In each quarter, we construct the return on equity *ROE* as the ratio of the 12-month earnings per share to the book value of equity per share (ERN over BKV). To mitigate outliers, we set *ROEs* that are greater than 10 or less than -10 to be 10 and -10, respectively. We then compute the correlation between 20 quarterly *ROEs* of any two stocks A and B, between year *t-4* and year *t*. *Earncorr_ch* is the correlation between two firms' changes in *ROE*s. *Earn_surprise_corr* is the Pearson correlation of the quarterly earnings surprises, measured as the I/B/E/S actual quarterly earnings minus the most recent analyst forecast of the earnings divided by the book value of equity. *Revision_corr* is the pairwise correlation between the monthly revisions in analyst forecasts of next quarter's earnings, where revisions are calculated as the change in mean consensus earnings forecasts scaled by lagged book value of equity. *Dgrowth* is the absolute value difference in five-year log sales growth rates. For any two stocks A and B, we construct a dummy variable *Sgics8* to be one if they have the same eight-digit GICS industry code, and zero otherwise. We set dummy variables *Sgics6*, *Sgics4*, *Sgics2* to one for stocks that have the same six-digit GICS codes, four-digit GICS codes, two-digit GICS codes, respectively. Market equity is the product of Compustat total shares outstanding and the fiscal year-end price (25*199). Size is the logarithm of the market equity at the fiscal year end in year *t*. *Dsize* is the absolute value difference in size. The book value of equity is the total assets minus total liabilities. Book-to-market equity is then the ratio of the book equity to the market value of equity at the fiscal year end in year *t*. *Dlogbtm* is the absolute value difference in log book-to-market ratios. *Daccrual* is the absolute value of the difference in accruals between a pair of stocks, where accruals are calculated as operating income after depreciation scaled by lagged book value of assets. *Sstate* is a dummy variable that is one if two firms are located in the same state, and zero otherwise. Firm age is the difference between the current year and the first year that a firm appears in CRSP, plus one. *Dage* is the absolute value difference between logarithm of ages of two firms *i*, and *j*, $|\log(Age_i)-\log(Age_j)|$. *Slisting* is one if two stocks are listed on the same exchange, and zero otherwise. *Sindex* is one if two stocks belong to the same S&P major, mid cap, or small cap indexes, and zero otherwise. *Dprice* is the absolute value difference in log prices per share. *Dduration* is the absolute value difference in cash flow durations. *Dleverage* is the absolute value difference in financial leverage ratios (long-term debt/total assets). *Indlink* is one if two firms belong to industries that have customer-supplier links, and zero otherwise. *Volumecorr* is the correlations between two firms' abnormal trading volumes. We include a stock in the sample only if it has 60 valid monthly returns, 20 quarterly earnings, and market equity of at least $500 million. *Rou_an* is the number of common financial analysts between two stocks divided by the square root of the product of the numbers of analysts that follow the two stocks. *Drou_an* is a dummy variable that equals one for any pair of two stocks that do not share any common financial analyst (i.e., when *Rou_an* is zero), and zero otherwise. There are 1,407,466 observations.

Variable	Mean	Std Dev	Minimum	Maximum
Corr	0.216	0.182	-0.562	0.964
Earncorr	0.078	0.453	-0.991	0.998
Earncorr_ch	0.062	0.390	-0.996	0.996
Earn_surprise_corr	0.015	0.248	-0.995	0.998
Revision_corr	0.018	0.153	-0.970	1.000
Dgrowth	0.409	0.373	0	5.607
Sgics8	0.015	0.123	0	1
Sgics6	0.029	0.167	0	1
Sgics4	0.059	0.236	0	1
Sgics2	0.126	0.332	0	1
Dsize	1.324	1.050	0	6.906
Dlogbtm	0.712	0.585	0	6.394

Daccrual	0.063	0.062	0	1.159
Sstate	0.054	0.226	0	1
Dage	0.681	0.505	0	2.315
Slisting	0.744	0.436	0	1
Sindex	0.494	0.500	0	1
Dprice	0.619	0.564	0	10.448
Dduration	1.555	1.541	0	28.810
Dleverage	0.155	0.118	0	0.887
Indlink	0.744	0.436	0	1
Volumecorr	0.094	0.200	-0.778	0.852
Rou_an	0.016	0.074	0	1
Drou_an	0.896	0.306	0	1

Table 11: Determinants of Stock Return Correlations

This table reports OLS regression results of the pairwise stock return Pearson correlation coefficient (*Corr*) on the determinants between 1987 and 2005. Each observation represents a stock pair year. *Corr* is the stock return correlation between two stocks using monthly returns between year *t-4* and *t*. We calculate earnings correlation (*Earncorr*) from the Compustat Price, Dividends, and Earnings dataset as follows. In each quarter, we construct the return on equity *ROE* as the ratio of the 12-month earnings per share to the book value of equity per share (*ERN* over *BKV*). To mitigate outliers, we set *ROEs* that are greater than 10 or less than -10 to be 10 and -10, respectively. We then compute the correlation between 20 quarterly *ROEs* of any two stocks A and B, between year *t-4* and year *t*. *Earncorr_ch* is the correlation between two firms' changes in *ROE*s. *Earn_surprise_corr* is the Pearson correlation of the quarterly earnings surprises, measured as the I/B/E/S actual quarterly earnings minus the most recent analyst forecast of the earnings divided by the book value of equity. *Revision_corr* is the pairwise correlation between the monthly revisions in analyst forecasts of next quarter's earnings, where revisions are calculated as the change in mean consensus earnings forecasts scaled by lagged book value of equity. *Dgrowth* is the absolute value difference in five-year log sales growth rates. For any two stocks A and B, we construct a dummy variable *Sgics8* to be one if they have the same eight-digit GICS industry code, and zero otherwise. We set dummy variables *Sgics6*, *Sgics4*, *Sgics2* to one for stocks that have the same six-digit GICS codes, four-digit GICS codes, two-digit GICS codes, respectively. Market equity is the product of Compustat total shares outstanding and the fiscal year-end price (25*199). Size is the logarithm of the market equity at the fiscal year end in year *t*. *Dsize* is the absolute value difference in size. The book value of equity is the total assets minus total liabilities. Book-to-market equity is then the ratio of the book equity to the market value of equity at the fiscal year end in year *t*. *Dlogbtm* is the absolute value difference in log book-to-market ratios. *Daccrual* is the absolute value of the difference in accruals between a pair of stocks, where accruals are calculated as operating income after depreciation scaled by lagged book value of assets. *Sstate* is a dummy variable that is one if two firms are located in the same state, and zero otherwise. Firm age is the difference between the current year and the first year that a firm appears in CRSP, plus one. *Dage* is the absolute value difference between logarithm of ages of two firms *i*, and *j*, $|\log(Age_i)-\log(Age_j)|$. *Slisting* is one if two stocks are listed on the same exchange, and zero otherwise. *Sindex* is one if two stocks belong to the same S&P major, mid cap, or small cap indexes, and zero otherwise. *Dprice* is the absolute value difference in log prices per share. *Dduration* is the absolute value difference in cash flow durations. *Dleverage* is the absolute value difference in financial leverage ratios (long-term debt/total assets). *Indlink* is one if two firms belong to industries that have customer-supplier links, and zero otherwise. *Volumecorr* is the correlations between two firms' abnormal trading volumes. *Rou_an* is the number of common financial analysts between two stocks divided by the square root of the product of the numbers of analysts that follow the two stocks. *Drou_an* is a dummy variable that equals one for any pair of two stocks that do not share any common financial analyst (i.e., when *Rou_an* is zero), and zero otherwise. We report three-way clustered *t*-statistics by permno of stock *i*, permno of stock *j*, and year in parentheses below the coefficients.

	1	2	3	4
Earncorr	0.024	0.022	0.020	0.020
	(4.42)	(4.24)	(3.97)	(3.56)
Earncorr_ch	0.015	0.013	0.014	0.014
	(2.04)	(1.79)	(1.99)	(1.89)
Earn_surprise_corr	0.017	0.015	0.015	0.014
	(6.47)	(5.70)	(5.74)	(5.82)
Revision_corr	0.091	0.078	0.076	0.071
	(7.43)	(6.96)	(6.71)	(5.91)
Dgrowth	-0.044	-0.041	-0.039	-0.028
	(-4.14)	(-3.88)	(-3.71)	(-3.35)
Sgics8		0.057	0.055	0.007
		(5.05)	(4.92)	(0.76)
Sgics6		0.024	0.025	-0.011

		(2.76)	(2.78)	(-1.36)
Sgics4		0.041	0.038	0.016
		(4.48)	(4.23)	(1.88)
Sgics2		0.080	0.079	0.069
		(11.20)	(11.07)	(8.91)
Dsize			-0.010	-0.008
			(-3.84)	(-3.78)
Dlogbtm			-0.018	-0.019
			(-4.80)	(-4.26)
Daccrual				-0.051
				(-1.96)
Sstate				0.010
				(2.47)
Dage				-0.014
				(-3.49)
Slisting				0.035
				(3.81)
Sindex				0.016
				(2.67)
Dprice				-0.004
				(-1.0)
Dduration				0.003
				(1.19)
Dleverage				-0.034
				(-2.78)
Indlink				0.003
				(0.58)
Volumecorr				0.072
				(3.53)
Rou_an				0.180
				(8.33)
Drou_an				-0.046
				(-3.68)
Obs.	1,407,466	1,407,466	1,407,466	1,407,466
Adjusted R^2	2.48%	7.23%	8.01%	11.67%

Table 12: Returns to Portfolios on Fitted and Residual Correlations

This table reports the value-weighted returns for portfolios that we form on the return difference (*RetDiff*) on the sample of firms that have valid observations to estimate the regression in Column 4 of Table 11. The sample period is January 1988 to December 2006. *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations (total, fitted, or residual correlations) with a given stock between year *t-4* and year *t*. In Panel A, the pairs portfolio is the 50 stocks with highest correlations. In Panel B, the pairs portfolio is the 50 stocks with highest fitted correlations in Column 4 of Table 11. In Panel C, the pairs portfolio is the 50 stocks with highest residual correlations in Column 4 of Table 11. In all three panels, *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The three factors are excess market return, SMB, and HML. The five factors are the three factors, plus the momentum factor, and the short-term reversal factor.

Panel A: Stocks with highest correlations as comovers

	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.77%	-0.17%	-0.12%
2	0.62%	-0.41%	-0.44%
3	1.10%	0.16%	0.20%
4	1.07%	0.07%	-0.02%
5	0.95%	-0.07%	-0.06%
6	1.17%	0.25%	0.26%
7	1.81%	0.85%	0.74%
8	1.44%	0.48%	0.42%
9	1.36%	0.33%	0.40%
Decile 10	1.19%	0.10%	0.37%
Decile 10-1	0.42%	0.26%	0.50%
t-statistics	(1.29)	(0.78)	(1.65)

Panel B: Stocks with highest fitted correlations as comovers

	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.66%	-0.30%	-0.22%
2	0.71%	-0.26%	-0.29%
3	0.92%	-0.09%	-0.08%
4	1.14%	0.21%	0.15%
5	0.87%	-0.09%	-0.05%
6	1.26%	0.27%	0.21%
7	1.59%	0.65%	0.70%
8	1.42%	0.43%	0.27%
9	1.24%	0.28%	0.32%
Decile 10	1.37%	0.23%	0.44%
Decile 10-1	0.71%	0.53%	0.66%
t-statistics	(2.27)	(1.66)	(2.13)

Panel C: Stocks with highest residual correlations as comovers

	Raw Return	3-Factor Alpha	5-Factor Alpha
Decile 1	0.96%	0.02%	0.10%

2	0.83%	-0.16%	-0.18%
3	0.88%	-0.11%	-0.07%
4	1.06%	0.11%	0.00%
5	1.06%	0.11%	0.12%
6	1.35%	0.38%	0.42%
7	1.58%	0.66%	0.69%
8	1.28%	0.22%	0.13%
9	1.23%	0.19%	0.20%
Decile 10	1.24%	0.20%	0.35%
Decile 10-1	0.28%	0.17%	0.25%
t-statistics	(0.80)	(0.47)	(0.81)

Table 13: Liquidity Factor

This table reports alphas and factor loadings for the hedge portfolios (Decile 10 – Decile 1) that we form on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The six factors are the excess market return, SMB, HML, the momentum factor, the short-term reversal factor, and the Pastor-Stambaugh liquidity factor. The sample period is January 1968 to December 2007 (480 monthly returns).

Panel A: Value-weighted portfolio

	1	2	3	4	5	6	7	8
Alpha	0.87%	0.88%	0.80%	0.81%	1.19%	1.23%	0.71%	0.75%
	(4.22)	(4.21)	(3.82)	(3.82)	(5.93)	(6.03)	(3.96)	(4.18)
Rm-Rf	0.30	0.30	0.30	0.30	0.25	0.25	0.14	0.13
	(6.66)	(6.62)	(5.80)	(5.80)	(5.09)	(5.09)	(3.15)	(3.13)
SMB			0.16	0.16	0.15	0.15	0.09	0.08
			(2.42)	(2.41)	(2.48)	(2.45)	(1.58)	(1.54)
HML			0.10	0.10	0.00	0.01	-0.02	-0.01
			(1.31)	(1.34)	(0.03)	(0.15)	(-0.36)	(-0.17)
MOM					-0.40	-0.40	-0.22	-0.23
					(-8.40)	(-8.47)	(-5.04)	(-5.17)
STRev							0.71	0.72
							(12.31)	(12.41)
PS_VWF		-0.02		-0.02		-0.07		-0.09
		(-0.28)		(-0.34)		(-1.13)		(-1.83)
Adj. R2	8.30%	8.12%	9.18%	9.01%	20.75%	20.80%	39.83%	40.13%

Panel B: Equal-weighted portfolio

	1	2	3	4	5	6	7	8
Alpha	3.24%	3.25%	3.04%	3.07%	3.48%	3.53%	2.97%	3.03%
	(15.32)	(15.21)	(14.42)	(14.42)	(17.45)	(17.58)	(17.06)	(17.40)
Rm-Rf	0.24	0.23	0.27	0.27	0.21	0.21	0.09	0.09
	(5.04)	(4.99)	(5.19)	(5.19)	(4.41)	(4.40)	(2.26)	(2.23)
SMB			0.27	0.27	0.26	0.26	0.19	0.18
			(4.00)	(3.98)	(4.22)	(4.19)	(3.57)	(3.51)
HML			0.31	0.32	0.20	0.22	0.18	0.19
			(4.04)	(4.12)	(2.82)	(3.01)	(2.88)	(3.17)
MOM					-0.44	-0.45	-0.25	-0.26
					(-9.40)	(-9.56)	(-6.00)	(-6.22)
STRev							0.75	0.76
							(13.41)	(13.61)
PS_VWF		-0.04		-0.06		-0.11		-0.14
		(-0.57)		(-0.92)		(-1.87)		(-2.78)
Adj. R2	4.85%	4.71%	9.63%	9.60%	23.64%	24.04%	44.52%	45.30%

Table 14: Pairs Trading Profits between 2008 and 2010

This table reports the value- and equal-weighted returns for quintile portfolios that we form on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret*-*Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The three factors are excess market return, SMB, and HML. The five factors are the three factors, plus the momentum factor, and the short-term reversal factor. Panel A reports value-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. Panel B reports equal-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. The sample period is January 2008 to December 2010.

Panel A: Value weighted

	Raw Return	3-Factor Alpha	5-Factor Alpha
Quintile 1	-0.46%	-0.63%	-0.75%
2	0.04%	0.02%	0.04%
3	0.04%	-0.07%	0.03%
4	0.58%	0.50%	0.45%
Quintile 5	0.72%	0.45%	0.22%
Quintile 5-1	1.18%	1.08%	0.97%
t-statistics	(2.55)	(2.46)	(2.29)

Panel B: Equal weighted

	Raw Return	3-Factor Alpha	5-Factor Alpha
Quintile 1	0.11%	-0.45%	-0.68%
2	0.18%	-0.23%	-0.37%
3	0.54%	0.17%	-0.01%
4	1.05%	0.54%	0.30%
Quintile 5	2.20%	1.54%	0.99%
Quintile 5-1	2.09%	1.99%	1.67%
t-statistics	(3.83)	(3.62)	(4.10)