

Losing sight of the trees for the forest?

Attention allocation and anomalies

Heiko Jacobs and Martin Weber*

June 2013

Internet appendix attached

This paper tests asset pricing implications of the investor attention shift hypothesis proposed in recent theoretical work. We create a novel proxy for the dynamics of inattention towards firm-specific information and explore its impact on prominent return anomalies. As hypothesized and with all else equal, the proxy positively predicts the post-earnings-announcement drift and negatively predicts the success of momentum strategies. Moreover, it has explanatory power for the profitability of pairs trading, a promising yet widely neglected setting concerned with the relative pricing efficiency of economically linked stocks. Taken together, our findings highlight the importance of time-varying investor attention allocation.

Keywords: Investor behavior, return predictability, limited attention, attention shifts, anomalies, momentum, PEAD, pairs trading, relative-value arbitrage

JEL Classification Codes: G12, G14

*Correspondence: Heiko Jacobs, Finance Area, University of Mannheim, L5, 2, 68131 Mannheim, Germany. E-mail: jacobs@bank.bwl.uni-mannheim.de. Martin Weber, Finance Area, University of Mannheim, L5, 2, 68131 Mannheim, Germany. E-mail: weber@bank.bwl.uni-mannheim.de. Financial support from the Julius Paul Stiegler Memorial Foundation (travel grants), DAI (PhD thesis award), Acatis (research award), and BAI (research award) is gratefully acknowledged. We thank Paul Tetlock (FRA discussant), Peter Nyberg, Joey Engelberg, R. David McLean, Richard Deaves, and Volodymyr Zdorovtsov for valuable comments. Moreover, we would like to thank conference participants at the 2011 Financial Research Association Meeting (FRA) in Las Vegas, at the Inaugural Helsinki Finance Summit on Investor Behavior, at the 18th Annual Meeting of the German Finance Association, at the University of Mannheim, at the WHU, at the LMU Munich, at the Centre for European Economic Research Conference 2012, at the Acatis Value Seminar, and at the EFM 2012 Asset Management Symposium for valuable comments.

1. Introduction

Focussing attention on one task necessarily goes hand in hand with a decrease in attention to other tasks (Kahneman (1973)). This well-established psychological constraint forces market participants to be selective in information processing. In this paper we empirically explore asset pricing implications of time-varying attention towards firm-level information. At first, this requires answers to two questions: How does one proxy for the unobservable dynamics of investor attention? And in which settings do attention shifts matter?

A few studies so far have dealt with these questions. They propose simple, intuitive proxies for limited attention and then almost exclusively apply them to explain time-series variation in the magnitude of the post-earnings-announcement drift (PEAD). Hirshleifer et al. (2009) use the number of same day earnings releases that compete for investors' attention. Similarly, Peress (2008) counts the total number of firms featured in the Wall Street Journal. DellaVigna and Pollet (2009) rely on a Friday dummy by arguing that investors are distracted by the upcoming weekend. Hou et al. (2009) use down market periods, during which investors are assumed to "put their heads in the sand". All papers find that the PEAD is stronger if investor attention is low.

Our contribution is twofold. First, we propose a novel proxy, which builds on recent theoretical work on the dynamics of attention allocation (e.g. Peng (2005), Peng and Xiong (2006)). These models argue that investors aim at optimally allocating their effort across several aggregation levels, which they do depending on priority and urgency. News relevant primarily on some aggregated level typically get most attention, as this information tends to be most important for the valuation of the investor's overall portfolio.¹ Remaining capacities are then used to process more disaggregated (e.g. firm-level) news.

This setting has intuitively appealing time-series implications: If the investor needs to focus on the big picture (the "forest"), he has necessarily fewer resources available to devote to details (the "trees"). As Peng and Xiong (2006) summarize: "In severely con-

¹A direct consequence of this form of attention allocation is category learning behavior, for which there is now ample evidence. Theoretical work includes e.g. Mullainathan (2002) or Barberis and Shleifer (2003). On the empirical side, vivid examples of category thinking are for instance given in Cooper et al. (2001) or Cooper et al. (2005). Category learning also leads to excessive return comovement as recently identified in various settings (e.g. Barberis et al. (2005), Greenwood (2008), Boyer (2011), Green and Hwang (2009) or Pirinsky and Wang (2006)).

strained cases, the investor allocates all attention to market- and sector-level information and ignores all the firm-specific data” (p. 565).

Consequently, the proxy aims at quantifying the unexpected daily information load market participants need to process in order to timely assess the overall market situation. Building on the premise that information shocks partly manifest themselves in abnormal returns, we condense the magnitude and dissemination of unanticipated daily return shocks in a broad range of market segments into a single ratio. We perform yearly decile sorts of this variable. This results in about 25 “high distraction days” (decile 10) in each year, during which we expect investors’ full attention to have been directed to the market-level. There are also about 25 “low distraction days” (decile 1), during which we expect sufficient cognitive resources to have been devoted to individual firms. Several conceptually quite diverse tests collectively lend support to the idea that the proxy is indeed able to identify investor attention shifts.

Second, and in order to show its economic implications, we apply the proxy to several puzzling return phenomena. We start with the PEAD (Bernard and Thomas (1989)) which the previous attention literature has focused on. However, we also investigate the predictability of a broad range of momentum strategies (Jegadeesh and Titman (1993)). In addition, we devote special effort to explore the time-varying profitability of pairs trading (Gatev et al. (2006), Engelberg et al. (2009)). This popular relative-value arbitrage approach employs statistical methods to identify economically linked firms and then exploits slow information diffusion between these firms by means of a long-short strategy.

As we outline below, studying these three asset pricing phenomena jointly is a promising and natural approach to gain novel insights into the role of time-varying attention allocation in financial markets. To put it briefly, there are good reasons to believe that each of these long-standing return anomalies might be partly driven by the dynamics of attention. However, behavioral theories suggest that the implications of the amount of attention devoted to affected firms differ across anomalies due to conceptual differences. This rich set of predictions allows us to empirically discriminate the investor attention hypothesis against alternative conjectures about the sources of these anomalies. The essence of our findings is captured in figure 1.

Please insert figure 1

The discovered findings indicate that the attention shift proxy has statistically significant and economically meaningful predictive power for the magnitude of each return anomaly. We discuss anomalies one by one and start with the PEAD.

If investors are temporarily distracted, the information contained in earnings surprises may not immediately be incorporated into prices, giving rise to a predictable return drift over the next weeks. Our findings support this notion, which has been formalized in the models of e.g. DellaVigna and Pollet (2009) or Hirshleifer et al. (2011). The difference in cumulative abnormal returns over days $t=2$ to $t=60$ between firms with positive earnings surprises and those with negative surprises is only 1.86% and insignificant, conditional on the announcement at $t=0$ having taken place on a low distraction day. However, the difference amounts to 4.28% if earnings were announced on a high distraction day, giving rise to a significant difference of differences. As these findings are in line with previous work, they might also be regarded as an important validation check of the novel proxy.

However, as Hou et al. (2009) highlight, attention effects are likely to play a dual role. They are also at the heart of overreaction-driven return anomalies, which are typically considered to be caused by behavioral biases. If attention interacts with or exacerbates these biases, then situations may occur in which a subset of investors does not pay “too little” attention to individual stocks, but “too much”. Several papers show that this phenomenon often results in fundamentally mostly unjustified abnormal returns followed by reversals (e.g. Barber and Odean (2008), Engelberg et al. (2011), Lou (2011), Seasholes and Wu (2007), Yuan (2012)).

Among the anomalies of this type, the momentum effect is arguably among the most persistent (e.g. Fama and French (2008)), widespread (e.g. Asness et al. (2012)), and prominent (e.g. Chui et al. (2010)). Several behavioral theories attribute momentum and its subsequent reversal to systematic biases in information processing, such as extrapolative expectations (Long et al. (1990)) or overconfidence in conjunction with self-attribution (Daniel et al. (1998)). Empirical work provides support for these overreaction theories and moreover shows that excessive attention thereby plays a key role.²

Thus, in the case of overreaction-driven momentum the impact of attention towards

²A non-exhaustive list includes Antoniou et al. (2010), Byun et al. (2013), Chui et al. (2010), Cooper et al. (2004), Da et al. (2011), Hou et al. (2009), Hillert et al. (2013), Lee and Swaminathan (2000), and Stambaugh et al. (2012).

individual firms should work in the opposite direction as compared to its role for underreaction-driven PEAD. Figure 1 shows that this hypothesis proves true. It displays average monthly returns generated by a composite momentum strategy (with overlapping portfolios) constructed from fourteen traditional or enhanced momentum strategies proposed in the literature. The average momentum return over January 1962 to December 2008 is essentially insignificant (0.32%, t-statistic 0.71) following the two months which are characterized as high distraction periods in each year. However, returns are much larger (1.82%, t-statistic 5.25) than usual in the two months defined as periods of low distraction.

Finally, figure 1 shows findings for pairs trading. This anomaly is loosely related to deviations of the law of one price and thus offers another conceptually different and also previously neglected setting. From the overall 200 million eligible sets of two stocks in our baseline US sample, we identify those which have moved together the most historically. Each month, we select the 100 pairs with minimum distance between normalized historical return paths, verify that they are large and liquid, and then trade them over the next six months. Specifically, whenever cumulative daily returns of any of these top pairs sufficiently diverge, we short the relatively overpriced winner and buy the relatively underpriced loser. If the future resembles the past, prices are likely to finally converge again, thereby generating positive returns on zero-cost portfolios of economic substitutes.

Our findings show that the one-month event-time return on pairs opening on high distraction days is more than twice as high as the return generated on pairs diverging on low distraction days. This insight is based on more than 100,000 round-trip trades between January 1962 and December 2008. We analyze the case of pairs trading in particular depth for a number of reasons.

Linking the type of cross-return predictability observable in pairs trading with time-varying investor attention is a novel and in our view appealing approach. Related studies have shown that limited attention can cause autocorrelation in returns (e.g. Hong et al. (2000)) or lead-lag effects between portfolios of stocks (e.g. Hong et al. (2007), Hou (2007)) and along the supply chain (Cohen and Frazzini (2008), Menzly and Ozbas (2010)). Pairs trading, however, is about predicting the relative performance of two typically large stocks, of which neither is the systematic leader. It is a short-term strategy whose profitability almost monotonically declines in event-time (e.g. Section 4) and which is more pronounced

for pairs with sluggish cross-stock information transfer (Engelberg et al. (2009), Chen et al. (2012)). We provide a number of additional findings which support the notion of slow information diffusion due to limited attention being the cause for short-term mispricings. For example, pairs particularly neglected (covered) by the media appear more (less) profitable and exhibit a higher (lower) sensitivity to attention shifts.

Moreover, the setting in pairs trading is related to other pervasive empirical puzzles, in which there are price discrepancies between similar assets.³ In a broader sense, our in-depth analysis of pairs trading might thus help to better understand how limited attention affects the efficiency of fundamentally linked assets in practice.

In addition, and in contrast to the classical anomalies PEAD as well as momentum, comprehensive empirical studies on the drivers of pairs trading are still rare. This is surprising given its “disarmingly simple” concept (Gatev et al. (2006), p. 797), its apparent violation of even the weak form of market efficiency, as well as its popularity among practitioners.⁴ Moreover, very little is known about pairs trading in international markets, even though only few trading strategies have survived the test of time and independent scrutiny. As a consequence, it is still an open question when, where, and why pairs trading is particularly profitable. We address this gap in the literature with a data set of about 25 million firm days from eight major non-US stock markets.

While figure 1 provides a first glance at the findings, we present the full results in the following sections. It should be noted that all asset pricing tests rely on firms with a market capitalization above the first NYSE decile (NYSE/Amex median in the case of pairs trading) and a stock price of at least 5 USD. This results in eliminating almost half of the firm months in the NYSE/Amex/Nasdaq universe. Moreover, to further mitigate

³For instance, Lee et al. (1991), Pontiff (1995), Chay and Trzcinka (1999), and Cherkes et al. (2009) focus on the relationship of the prices of closed-end fund shares and the per share market value of the assets held by the funds. Mitchell et al. (2002) and Lamont and Thaler (2003) study situations in which a firm’s market value is less than the value of its ownership stake in a publicly traded subsidiary. Rosenthal and Young (1990), Froot and Dabora (1999), Scruggs (2007), Jong et al. (2009), and Baker et al. (2012) study price parity deviations of dual-listed companies (“Siamese Twins”). Gagnon and Karolyi (2010) study discrepancies between the prices of US and home-market shares of companies with cross-listed stocks. Smith and Amoako-Adu (1995), Zingales (1995), and Schultz and Shive (2010) study dual class shares issued by the same company, that differ in voting rights, but have equal cash flow rights.

⁴Gatev et al. (2006) refer to pairs trading as being “among the proprietary “statistical arbitrage” tools currently used by hedge funds as well as investment banks” (p. 797). Similarly, in a practitioner’s book, Vidyamurthy (2004) concludes that pairs trading has “increased in popularity and has become a common trading strategy” (p. 74). Andrade et al. (2005) estimate the realized profit from pairs trading for sophisticated investors at “hundreds of millions of dollars” (p. 3).

microstructural effects, all tests based on daily data skip a day before trading is initiated.

Our findings are stable when a comprehensive list of control variables, both in the cross-section and in the time-series, is taken into consideration. A particular focus is on variables deemed to be related to market frictions. In addition to standard firm characteristics such as size, industry membership, or idiosyncratic volatility, we include various proxies for trading costs such as the Amihud (2002) illiquidity ratio or the recently proposed daily bid-ask spread measure of Corwin and Schultz (2012).

Market-level controls include, for instance, the Fama and French (1993) factors and other well-established risk premiums (such as short-term reversal), squared market return, volatility, turnover, VIX, TED spread, interest rates, or average bid-ask spreads. Finally, we control for calendar effects and previously proposed time-series determinants of the return anomalies. Overall, findings are stable across a number of robustness checks, including subperiod results and limits to arbitrage.

2. Attention shift proxy

2.1. *Empirical design*

Inspired by models such as Peng and Xiong (2006), we aim at identifying days on which market participants are likely to be forced to spend more (or less) resources than usual on understanding “the big picture”. Operationalizing this idea leaves many degrees of freedom. We thus start by constructing a baseline proxy designed as simple as possible and later modify its construction extensively.

A four-step procedure is employed. First, we define the 49 Fama and French (1997) industries as market segments. As this classification has been shown to represent economic boundaries (e.g. Chan et al. (2007)) as well as informational boundaries (e.g. Menzly and Ozbas (2010)), the use of industries in the baseline analysis is an intuitive choice. For January 1960 to December 2008, we compute daily value-weighted returns for all 49 industries separately, thereby taking into account all common shares traded on NYSE, Amex or Nasdaq. Second, we rely on a simple OLS market model to decompose returns

$R_{i,t}$ of market segment i on day t . This allows us to identify return shocks $AR_{i,t}$:

$$AR_{i,t} = |R_{i,t} - \hat{\alpha}_{i,t} - \hat{\beta}_{i,t}R_{m,t}| \quad (1)$$

Parameter estimates are obtained from rolling time-series regressions based on daily return data over the previous year. Third, we condense these shocks into a single measure. From the several plausible weighting schemes we choose an approach that takes the expected level and frequency of market segment-specific shocks into account. Each segment weight $w_{i,t}$ is determined by the inverse of the volatility $\sigma_{i,t}$ of the shock variable $AR_{i,t}$ as follows:

$$\sum_{i=1}^{49} w_{i,t} AR_{i,t} \text{ where } w_{i,t} = \frac{\frac{1}{\sigma_{i,t}}}{\sum_{i=1}^{49} \frac{1}{\sigma_{i,t}}} \quad (2)$$

Shock volatilities $\sigma_{i,t}$ are estimated from $AR_{i,t}$ over the previous year. The weighting approach applies to the intuition that a pronounced return shock in a segment for which large shocks are a common occurrence is less likely to unexpectedly demand extra attention than a pronounced shock in a market segment which “usually behaves as expected”. The appendix displays the time-series of the values obtained after this third step. There are several phases (but no general time trend), in which values typically differ substantially from the sample average.

To thoroughly test for attention effects, we follow the methodology of Hirshleifer et al. (2009) by performing yearly sorts of the values from equation (2) as the fourth and final step. For each year individually, we assign a decile rank to each trading day. This approach might be considered conservative in that we identify the same number of high and low distraction days in each year, irrespective of the actual market environment. Pooling the data results in roughly 1,180 days per decile rank.

Figure 2 shows the time-series of the monthly number of high and low distraction days, thereby demonstrating that these days are typically not heavily clustered in certain months within a given year.

Please insert figure 2

Closer inspection reveals that this holds also true if one takes all deciles into account. The 25th (75th) percentile of the monthly average of daily decile ranks is 4.33 (6.51). The 5th (95th) percentile is 3 (8.18). Thus, findings in later tests will not be driven by a few

extreme months. As the appendix verifies, findings are also not driven by a few extreme market segments.

2.2. *Do we capture attention shifts?*

In the following, we present main insights from five validation checks of the attention shift proxy. Details about the design and further tests are reported in the appendix.

Yuan (2012) provides evidence that record levels for the Dow Jones Industrial Average as well as front-page articles about the stock market are market-wide attention-grabbing events. A meaningful attention shift proxy should thus indicate that the amount of attention market participants direct to aggregate information during these days should increase relative to the amount of attention devoted to firm-level news.

Our findings provide support for this line of reasoning. The unconditional probability of observing a stock market article on the front-page of the New York Times over our sample period from January 1980 to December 2008 is 4.52%. In contrast, on the day after a low (high) distraction day, the likelihood is 1.99% (12.16%). Similarly, a new record level of the Dow is observed on 5.59% of all trading days between January 1962 and December 2007. However, the corresponding value for low (high) distraction days is 4.03% (6.15%).

The investor distraction proxy of Hirshleifer et al. (2009) aims at quantifying the number of competing information signals that draw attention away from a given firm. The average decile rank of the Hirshleifer et al. (2009) proxy over July 1971 to December 2008 is 4.76 on low distraction days, but 5.93 on high distraction days. In other words, both proxies tend to partly identify similar situations.

Recent work such as Da et al. (2011) or Mondria and Wu (2012) demonstrate that analyzing shocks in search terms entered into Google is a promising way of quantifying short-term investor attention allocation and information processing. We follow previous literature in measuring the interest in individual firms by the search volume of ticker symbols. Interest in aggregated information is quantified with search terms like “stock market”, “market segments” or “stocks”. Indeed, high (low) distraction days are *ceteris paribus* associated with a relatively increasing demand for aggregate information in relation to the demand for firm-level information. As a rough estimate, a one standard

deviation change of the Google-based proxy is associated with a 0.7 point change in attention shift proxy decile ranks.

Shifting attention towards aggregate information tends to go along with category learning behavior (see footnote 1). This suggests that true firm-specific turnover, i.e. firm-level trading activity orthogonal to market as well as market-sector trading activity, should be lower on high distraction days. Our findings support this conjecture. For each year from 1962 to 2008 separately, we create a decile-based measure of average daily residual firm-level turnover. The average decile rank is 6.23 on low distraction days, but only 4.43 on high distraction days. This is particularly noteworthy as market turnover is positively related to the attention shift proxy, as the following section shows.

2.3. *Time-series characteristics of the baseline proxy*

Figure 2 displays correlation coefficients between attention shift proxy decile ranks and the rank order of market-level variables, expressed both in raw values and yearly decile ranks. The proxy is essentially uncorrelated with standard risk premiums (including the market excess return or the short-term reversal factor) and weakly positively (0.1-0.25) correlated with factors each assumed to capture a specific aspect of unexpected market conditions. The latter include squared market return, market turnover, rolling ten day return volatility from $t-10$ to $t-1$, as well as the average daily bid-ask spread of firms larger than the first NYSE decile, computed as suggested by Corwin and Schultz (2012).

We additionally derive a number of promising proxies for arbitrage risk in the time-series from previous work: VIX, TED spread, Moody’s BAA to AAA yield spread, term spread, interest rate, and a measure building on abnormal stock returns of investment banks.⁵ The appendix gives more details about the construction of all six proxies, for which data availability on a daily basis starts between 1980 and 1990. Figure 2 verifies

⁵Brunnermeier et al. (2008) rely on the Chicago Board Options Exchange Market Volatility Index (VIX) as well as on the TED spread in an attempt to proxy for “speculators’ willingness and ability to put capital at risk” (p. 334, see also e.g. Asness et al. (2012), Ben-David et al. (2012), Brunnermeier and Pedersen (2009), Frazzini and Pedersen (2013), Moskowitz et al. (2012)). The VIX a popular measure of the volatility implied in S&P 500 index options and widely considered to be a forward-looking measure of overall market uncertainty. A high TED spread, i.e. the difference between the three-month LIBOR interbank market interest rate and the risk-free three-month T-Bill rate, signals that capital is becoming scarce and borrowing is becoming difficult. For a motivation of the other proxies see e.g. Engelberg et al. (2009), Pontiff (1996), Buraschi et al. (2012), Ang et al. (2011) or Menzly and Ozbas (2010).

that correlation coefficients are again low with a typical range from 0 to 0.25.

In conclusion, it can be said that the attention shift proxy appears to be at best weakly related to market frictions. We later control for all variables outlined above.

2.4. Modifications of the baseline proxy

To assure robustness, we construct twelve alternative proxies, which modify the original approach in many dimensions. They differ, for instance, with respect to the type and number of market segments used (49 industries, 100 portfolios sorted on book-to-market ratio and firm size, 25 portfolios sorted on size and short-term reversal) and the weighting scheme of return shocks (volatility weighting, equal weighting, value weighting, interquartile range). They also differ with respect to the model of expected return (market model, four factor model), the type of returns used (raw returns, abnormal returns, abnormal returns orthogonal to idiosyncratic firm-level volatility), and the timing of the returns used (yearly sorts, rolling historical values to assure real-time implementability).

The appendix provides detailed information about these proxies. We have made sure that the qualitative nature of our findings for all anomalies remains stable regardless of which specific proxy design we rely on.

3. Application to classical stock market anomalies

3.1. *Underreaction phenomenon: Post-earnings-announcement drift*

We closely follow the methodology of previous work, such as Hirshleifer et al. (2009) or DellaVigna and Pollet (2009). As the recording of many announcements has been shown to be error-prone before 1995, we start by gathering quarterly earnings announcement data from January 1995 to December 2008. Earnings surprises are defined as the difference between the actual announced earnings as reported in the I/B/E/S detail tape and the median estimate from individual analysts. To assure comparability across firms and quarters, differences are normalized by the stock price. In an attempt to exclude stale forecasts, we only consider one- or two-quarter-ahead estimates issued or reviewed no

more than 60 calendar days before the actual announcement. If a given analyst made several forecasts for the same announcement, we only consider his most recent estimate.

The data is then matched with daily stock prices from CRSP. We impose several screens intended to filter out small and illiquid firms and to minimize the impact of possible data errors or outliers.⁶ In total, 101,812 events remain. Next, we perform independent quarterly decile sorts of earnings surprises. Thus, a value of 10 (1) denotes very positive (negative) surprises in a given quarter.

As the major part of the PEAD is attributable to roughly the first 60 trading days (e.g. Bernard and Thomas (1989)), we focus on the period from $t=2$ to $t=60$. The lag after the actual event in $t=0$ aims at minimizing the impact of microstructural effects. Cumulative stock-level abnormal returns are computed as the difference between the raw buy-and-hold return over the event window and the expected beta-adjusted buy-and-hold return over the same period.

Finally, we rely on the following regression framework (e.g. Hirshleifer et al. (2009)). CAR refers to the cumulative abnormal return, and FS (ASP) denotes the decile rank of forecast surprises (attention shifts). Y_i refers to a vector of control variables, which are all interacted with FS . The interaction effect $FS \times ASP$ is the variable of primary interest. A positive coefficient implies a higher sensitivity of post-announcement abnormal returns to earnings surprises on high distraction days, holding all other factors fixed.

$$CAR(2, 60) = \beta_0 + \beta_1 FS + \beta_2 DP + \beta_3 (FS \times ASP) + \sum_{i=1}^n \delta_i \cdot Y_i + \sum_{i=1}^n \gamma_i (FS \times Y_i) + \varepsilon \quad (3)$$

Firm-level controls include size, market-to-book ratio, the Campbell et al. (2008) distress measure, the daily bid-ask spread measure of Corwin and Schultz (2012), the number of analysts covering the firm, idiosyncratic volatility, parameters from the estimation of expected returns (beta, alpha, R^2), turnover, the Amihud (2002) illiquidity ratio, share price, dummies for Nasdaq and S&P 500 membership, dummies for the 49 Fama and French (1997) industries, dummies for lottery and non-lottery stocks as in Kumar (2009),

⁶We only consider firms with a market capitalization larger than the first NYSE decile and a stock price larger than 5 USD at the beginning of the previous month. As in previous literature, we exclude observations from the sample if at least one of the following conditions is met: The announcement takes place on a non-trading day. An earnings surprise cannot be computed, for instance as I/B/E/S information cannot be linked with CRSP data. The actual or forecasted earnings are greater than the stock price. For each observation, at least one firm-level control variable (for details see below) cannot be computed. Finally, in order to be closer to the normality assumption, we winsorize observations above the 99% level or below the 1% level of the distribution of cumulative abnormal returns over $t=2$ to $t=60$.

and the reporting lag. Details regarding the construction are reported in the appendix. Findings are robust if we only use a subset of these variables.

Panel A of table 1 displays the main findings. The PEAD is indeed about 50% stronger for firms announcing their earnings on high as opposed to low distraction days. For instance, the sensitivity in specification 4 is $0.675 + 10 \times 0.0372 = 1.047$ for attention shift proxy decile rank 10, but only $0.675 + 1 \times 0.0372 = 0.7122$ for rank 1. It is thus 47% higher on high distraction days.

This result is comparable in terms of magnitude to earlier studies with alternative proxies (e.g. Hirshleifer et al. (2009) or DellaVigna and Pollet (2009)). It holds regardless of whether we consider all earnings surprises (first two columns) or only the extreme deciles (last two columns).

Please insert table 1

Robustness checks Panel B verifies that inferences do not change if we take one or several sets of time-series control variables into account. Calendar effects are controlled for by the inclusion of dummies for the year of the announcement, the month of the year, as well as the day of the week. We consider previously proposed attention proxies by including a dummy for a positive three-year market return (Hou et al. (2009)), a Friday dummy (e.g. DellaVigna and Pollet (2009)), and the decile rank of the number of competing earnings announcements as in Hirshleifer et al. (2009). Market-level controls include all variables outlined in figure 2 (e.g. 10-day rolling volatility, average daily bid-ask spread, VIX, TED spread).

The interaction effect remains highly significant in all cases. Untabulated tests show that it also remains positive once we focus on a subsample of firms which happen to exhibit similar earnings surprises at least once on both a high and a low distraction day. This analysis follows Peress (2008) and is intended to control for unobservable heterogeneity across firms. Finally, further analysis shows that the overall effect is primarily driven by positive earnings surprises. Time-varying short-selling constraints not fully captured by the control variables are thus unlikely to drive the documented findings.

3.2. *Overreaction phenomenon: Momentum*

Since the seminal study of Jegadeesh and Titman (1993), a number of studies has argued that momentum can be enhanced once one considers the interaction of past returns with certain firm characteristics. These variables are often argued to amplify behavioral biases or to proxy for information uncertainty. In an attempt to condense this literature, we construct a composite momentum strategy over the time period of January 1962 to December 2008, which averages the profitability of fourteen individual momentum approaches suggested by previous work.⁷

Most individual momentum strategies are computed using a formation period of six months, a skipped month, and a holding period of six months with overlapping portfolios. Stocks eligible for entering any momentum strategy are required to have a nominal share price of at least 5 USD and a market capitalization larger than the first NYSE decile at the time of portfolio formation. Returns for each momentum approach are value-weighted. Details on the construction are provided in the appendix.

The literature standard is to rely on a monthly frequency for both momentum returns and explanatory variables. While we have made sure that insights also hold when using data on a daily frequency, reported findings are based on a monthly attention shift measure to assure comparability with previous work. First, average monthly values of the baseline (daily) attention shift proxy as described in detail in Section 2.1. are computed. Second, for each year individually, these monthly values are sorted in ascending order and assigned values from 1 to 6. Thus, two months in each year are assigned the same attention shift rank, although findings are not sensitive to this choice. However, this approach assures a sufficiently high number of observations. It might moreover be considered conservative as we do not focus on extreme values and also run the procedure for each year separately.

⁷In addition to traditional momentum (1), we consider strategies enhanced with the following firm characteristics: (2) Age (e.g. Zhang (2006)), (3) turnover (e.g. Lee and Swaminathan (2000)), (4) market-to-book ratio (Asness (1997), Daniel and Titman (1999), Sagi and Seasholes (2007)), (5) credit rating (Avramov et al. (2007)), (6) market capitalization (e.g. Jegadeesh and Titman (1993), Hong et al. (2000), Zhang (2006)), (7) residual analyst coverage (e.g. Hong et al. (2000)), (8) analyst forecast dispersion (Zhang (2006), Verardo (2009)), (9) R^2 (Hou et al. (2006)), (10) formation period return consistency (Grinblatt and Moskowitz (2004)), (11) (idiosyncratic) volatility (Zhang (2006), Jiang et al. (2005)), and (12) nearness to 52-week high (George and Hwang (2004)). In addition, we compute two other modifications of the traditional approach. We consider (13) the extremity of formation period returns (e.g. Bandarchuk and Hilscher (2013)) and (14) the role of intermediate horizon past performance (Novy-Marx (2012)).

We then regress the time series of momentum returns on one-month lagged attention shift proxy ranks. As momentum portfolios are overlapping, we do not intend to predict momentum in the first month of the holding period. Instead, the coefficients of the explanatory variables might be regarded as giving an indication for calendar-time momentum profitability averaged over the whole evaluation period (of six months). Model 1 is univariate. Model 2 includes the Fama and French (1993) factors. Model 3 to 6 augment model 2 with further control variables. Model 3 additionally includes factors for short term reversal, long term reversal, and liquidity (Pástor and Stambaugh (2003)). It further contains monthly versions of those market-level variables covered in detail in figure 2 which are available over the whole sample period. These include, among others, the squared market return, the average daily bid-ask spread, or market volatility computed from daily returns in the month under consideration.

Model 4 to 5 additionally contain all proxies for limits to arbitrage illustrated in section 2.3., some of which have been explored in previous work on momentum (e.g. Asness et al. (2012), Chordia and Shivakumar (2002)). Model 6 additionally contains further variables the literature has linked to time variation in momentum profits. These include market state (Cooper et al. (2004), Hou et al. (2009)), return dispersion (Stivers and Sun (2010)), market volatility (e.g. Wang and Xu (2010)), the Baker and Wurgler (2006) sentiment index (Antonioni et al. (2010), Stambaugh et al. (2012)), and the recently proposed co-momentum measure of Lou and Polk (2012). The appendix provides details on the construction.

Please insert table 2

A pervasive picture emerges. In all models and sample periods, the impact of the lagged attention shift proxy on future momentum profits is negative and highly significant. The effect is also large in economic terms. Depending on the specification, most estimates suggest that the momentum effect in month t is about 100 to 200 basis points lower if investor attention in month $t-1$ was focussed on the market- or sector-level (attention shift proxy rank 6) rather than on individual firms (attention shift proxy rank 1). About 40% of the effect is attributable to the long leg of the portfolio. In essence, the findings uncover that momentum strategies have been largely unprofitable following months during which investors are assumed to have been distracted from processing (and potentially

overreacting to) firm-level information.⁸

Indeed, and as seen from figure 1, the average monthly momentum return from 1962 to 2008 is only 0.32% (Gallant (1987)-adjusted t-statistic 0.71) conditional on the two (lagged) months in each year which are assigned an attention shift proxy rank 6. In contrast, the momentum effect is 1.82% (t-statistic 5.25) following attention shift proxy rank 1, which is assumed to capture months during which investors extensively focus on individual firms. In contrast, the unconditional monthly composite momentum strategy return is 1.08%.

Robustness checks Untabulated tests verify that the impact of lagged attention on future momentum is reasonably stable. For instance, the effect remains qualitatively unchanged if we rely on equal-weighted instead of value-weighted momentum returns. Moreover, we have analyzed each of the fourteen individual momentum strategies in each of the six model specifications separately. This operation yields in more than 85% of all cases findings which are statistically significant at least at the 5% level.

4. Application to pairs trading

For the reasons outlined in the introduction, the analysis of limited attention in the context of pairs trading appears particularly promising and is thus investigated in particular depth. Our analysis thereby adds to the surprisingly small empirical literature on pairs trading so far. Gatev et al. (2006) report economically significant and seemingly anomalous profits between 1962 and 2002. Engelberg et al. (2009) and Chen et al. (2012) further explore cross-sectional patterns. They find that part of the profits seem to stem from differential immediate responses to news that affect both stocks in the pair. Thus, focussing on the dynamics of attention allocation seems an intuitive way to gain deeper insights.⁹

⁸An untabulated analysis of firm-level residual turnover, computed as in Section 2.2., provides additional support for this assumption. Firm-specific turnover for stocks which a typical momentum strategy would bet on is higher (lower) than usual during periods of low (high) distraction.

⁹Do and Faff (2010), Do and Faff (2012), and Chen et al. (2012) report a declining trend in standard pairs trading profitability over the recent past, which is partly driven by a higher fraction of nonconvergent pairs. Again, identifying scenarios in which pairs are ex ante more likely to converge appears crucial to understand the price formation process. In the international context, Andrade et al. (2005) document annual excess returns of about 10% for the Taiwanese stock market between 1994 and 2002. They show that uninformed trading shocks are a major driver of the strategy's profitability.

4.1. *Empirical approach*

We obtain daily stock price data on all common shares traded on NYSE or Amex at any time between January 1960 and December 2008. We impose several restrictions to assure that only large and liquid stocks enter the analysis.¹⁰ To mitigate data mining concerns and to facilitate comparison with previous work, we widely follow the methodology developed in Gatev et al. (2006).

Specifically, we use daily price data to compute a stock-specific cumulative total return index over the whole estimation period. A simple algorithm is then used to determine to what extent two stocks have moved together historically. Let $R_{i,t}$ ($R_{j,t}$) be the normalized return series of stock i (j) in estimation period t , which is comprised of trading days 1 to n . The algorithm, intended to provide a parsimonious, intuitive framework to identify pairs, is then defined as:

$$\frac{1}{n} \sum_{i=1}^n (R_{i,t} - R_{j,t})^2 \quad (4)$$

We compute this distance measure for all possible pair combinations, whose number grows quadratically with the number of eligible stocks. Then we choose at the beginning of each month the top 100 pairs with minimum distance.

In the baseline analysis, we thereby require firms to belong to one of the 49 Fama and French (1997) industries, although we later also study pairs from the same industry. Firms from different industries are interesting candidates for our scenario as industrial boundaries have been shown to go along with informational boundaries induced by specialization of e.g. analysts or fund managers (e.g. Hong et al. (2007), Menzly and Ozbas (2010)). The appendix reveals that the monthly top pairs nevertheless have significantly correlated earnings surprises in the following quarters. Thus, the economic link between two firms in our baseline empirical setting might be thought of as being potentially strong, but simultaneously often also less explicit and transparent. The quantitative mechanism thus offers an intuitive, parsimonious way of identifying pairs, for which cross-stock information transfer is particularly likely to be inhibited in moments of high distraction.

The top pairs are then eligible for trading in the immediately following six month

¹⁰We discard all stocks with at least one missing return or zero trading volume on any day of the twelve-month period during which pairs are matched. Moreover, we only consider stocks with a market capitalization larger than the median of the NYSE/AMEX stock universe at that time.

evaluation period. Prices are again set to equal unity. Following Gatev et al. (2006), if the spread between the cumulative return series of two substitutes exceeds two historical standard deviations, we take a long position in the relatively underpriced stock and a short position in the relatively overpriced stock. The self-financing pair is then held for up to one month. If prices convergence before this cut-off date, the trade is closed with a gain. If prices do not convergence within a month, positions are offset. This results in a loss if prices have diverged even further.

A pair may trade several times during the evaluation period. The money invested in later trades differs depending on whether we report event-time results (the baseline analysis) or calendar-time results. In event-time, we just again go one dollar long (short) into the cheap (expensive) stock. In calendar-time, proceeds from previous trades are reinvested. Nevertheless, overall differences between both methods are small.

Figure 3 illustrates the trading process with examples. Pairs therein open several times during the evaluation period, however not always in the same direction. This is a commonly observed behavior. From the large fraction of pairs that open in total at least three times, roughly 85% hold each stock at least once in both a long and a short position. Therefore, the phenomenon is different from the lead-lag relationship studied in earlier work (e.g. Lo and MacKinlay (1990), Hou (2007), Hong et al. (2007)).

Please insert figure 3

We rely on a conservative return computation scheme. As discussed in Gatev et al. (2006), we skip one day after the divergence and, provided that the pair converges within a month, add one day following the crossing of the prices. This method is intended to account for microstructural effects. Moreover, it works strongly against finding effects attributable to investor attention, as any information overlooked at the day of divergence might be incorporated into prices during the next day without entering our return estimates.

4.2. *Baseline results: US evidence*

We have to ensure that our findings capture the impact of variations in investor inattention rather than variations in other important variables, particularly related to market

frictions. Therefore, table 3 compares selected firm-level and pair-level variables. Among others, these include market capitalization (NYSE/AMEX decile rank), the pre-event Amihud (2002) illiquidity ratio, average pre-event turnover, turnover on the day of divergence, bid-ask spread at the day of divergence computed as in Corwin and Schultz (2012), and idiosyncratic risk (e.g. Pontiff (2006)).

Please insert table 3

The table shows four main findings. First, firms in general tend to be large and liquid. For instance, the median firm belongs to the ninth NYSE/AMEX decile, has an average daily turnover of 0.11%, and is covered by nine analysts (after 1980, untabulated). About 54% of firms are members of the S&P 500. In close to a third of all pair observations, both firms belong to the S&P 500.

Second, there are often only small differences in firm characteristics within pairs and across distraction deciles. There are mixed findings regarding the degree of liquidity on the day of divergence. While mean turnover is higher on high distraction days, mean bid-ask spreads are also about 13 basis points higher. These differences vanish once one considers median values.

Third, both firms and pairs are, in the overall picture, well diversified across industries. However, utility stocks make up close to 30% of all sample firms and are part of all top industry group combinations. We address this issue in later tests.

Fourth, the day of divergence is an interesting date. Pairs on average are opened when cumulative standardized returns have diverged by 6.68%. More than 40% of this difference is attributable to the day of divergence itself. Thus, understanding the underlying drivers of stock price behavior on these days is essential. If prices diverge due to idiosyncratic firm shocks affecting only one constituent, a pairs trading investor is likely to lose money. If the divergence is due to slow information diffusion between both constituents, he is likely to earn money.

In multivariate regressions, we later control for all variables outlined above. However, we start by performing univariate tests. To this end, we compare the average one-month event-time return on zero-cost pairs sorted by attention shift decile ranks as observed on the day of pair divergence. Panel A of table 4 shows findings based on more than 100,000

round trip trades from January 1962 to December 2008. If pairs converge before the month has passed, we assume that the proceeds are held in cash with zero interest.

Please insert table 4

In line with previous literature, traditional pairs trading appears highly profitable with an unconditional monthly return of close to 100 basis points per month. However, pairs opening on low distraction days only yield a return of 53 basis points. In contrast, the estimate for high distraction days is 130 basis points. The difference between decile 10 and decile 1 thus amounts to highly significant and economically meaningful 77 basis points per month.

The effect is not confined to the extreme distraction deciles: Decile 2 to 9 show an almost monotonic increase in profitability. The appendix provides more detailed information about the return distribution. It also verifies that several popular performance measures such as the Sharpe Ratio paint a very similar picture: Trades on high distraction days strongly outperform trades on low distraction days.

To gain more insight into this matter, we study the mechanisms behind these return differences. The limited attention hypothesis predicts that the probability of convergence in the days following divergence should be higher. Imagine, for instance that common news during times of high distraction is released, which clearly and directly affects the first firm in the pair, but has an only indirect and less clear impact on the second firm. If the news does not become instantaneously and fully incorporated into the price of the second stock, then price will temporary diverge and lead to the opening of the pair. Consistent with this line of reasoning, considerably more pairs open on high distraction than low distraction days (see table 4). When investors become fully aware of the link between both firms, relative prices should adjust gradually and the pair is likely to finally converge again.

The second row of panel A in table 4 shows that this prediction is supported by the data. The average fraction of pairs converging within the month after divergence is 36.3%. For pairs opening on low distraction days, however, the convergence rate is only 33.4%. This value almost monotonically increases by distraction deciles, culminating in a convergence rate of 39.7% for decile 10. In other words, simply switching from low distraction to high

distraction days increases the likelihood of convergence by almost 20%.

Figure 4 shows the probability of convergence on a given day in event time. Consistent with the idea of short-lived investor attention constraints being an important driver of pairs trading success, the likelihood of convergence within the first event days is considerably higher (lower) for pairs diverging on high (low) distraction days. After about five days, convergence rates begin to approximate each other more closely until they finally appear indistinguishable.

Please insert figure 4

The higher convergence rate is particularly noteworthy as initial return differences on high distraction days are higher (see also figure 4). However, it is important to mention that it is the attention shift proxy, and not the return difference per se that matters. The latter turns out to be completely insignificant in multivariate regressions. Similar insights can be gained from double-sorts. For instance, we can restrict the analysis to pairs with a return difference of e.g. 2% or less on the day of divergence. Under this constraint, about the same number of pairs opens on high distraction days and low distraction days, while the return difference is very similar in both cases. However, high distraction pairs still outperform low distraction pairs by 73 basis points per month. Thus, the predictive explanatory power of the attention shift proxy is virtually unchanged.

An alternative potential source of profit is that the average return conditioned on non-convergence could be less negative for pairs opening on high distraction days. The limited attention hypothesis does not imply that this should be the case: Non-convergence is (comparatively more) suggestive of idiosyncratic news affecting only one stock in the pair (e.g. Engelberg et al. (2009)). As this type of information is arguably often easier to grasp and process than common news affecting the potentially complex relationship between both firms in the pair, attention constraints should be less binding. Again, table 4 displays findings consistent with this line of reasoning. The difference in returns upon non-convergence between decile 10 and decile 1 is virtually zero and statistically insignificant.

To control for other factors that might partially drive our findings so far, we conduct several multivariate tests. The main results are presented in panel B of table 4. The dependent variable is the pooled one-month event-time return on long-short pairs. The

independent variable of interest is the attention shift proxy, either quantified by the decile rank or by a high/low distraction dummy, which is zero for low distraction days (decile 1) and one for high distraction days (decile 10).

The remaining independent variables comprise up to three control sets. The first set controls for calendar and industry effects (indicator variables for year, month, day of week, as well as pair industry group combinations). The second set controls for nine market-level conditions on the day of divergence (market return, squared market return, market turnover, volatility, factors for daily return premia on size, value, momentum and short-term reversal, average bid-ask spread). The third set includes in total fifteen firm and pair characteristics as outlined in table 3. For instance, controls include the return difference and the average bid-ask spread on the day of pair divergence, the average market capitalization decile and the idiosyncratic risk of the pair, as well as within-pair differences of the before mentioned variables. The appendix displays all regression coefficients.

However, the main coefficient of interest is that of the high/low distraction dummy. Table 4 verifies that the return difference is still estimated to be 50 to 73 basis points per month. The coefficient on the attention shift proxy decile rank suggests that this finding is not solely driven by the extreme deciles. In all specifications, attention proxies remain strongly significant at the one percent level.

4.3. *Baseline results: International evidence*

An effective way to test the validity of the findings is to evaluate the success of the approach in independent samples. This is particular appealing as very little is known about the nature of pairs trading profits in other major international stock markets. Therefore, we study the dynamics of pairs trading profits in Japan, UK, France, Germany, Switzerland, Italy, Netherlands, and Hongkong.¹¹

We gather data from the Compustat Global Daily Stock File. Depending on the avail-

¹¹These markets represent the eight largest non North-American stock markets based on domestic stock market capitalization at the end of 2002, as reported by Datastream. This date roughly represents the middle of the sample period for most of these countries. The number of eight markets is chosen somewhat arbitrarily and can be seen as a compromise between maximizing the sample size and minimizing the fraction of error-prone daily return and volume data as well as the number of months with too few eligible stocks for a reasonable analysis.

ability of reliable trading volume data, the sample period starts at some point in the middle of the 90s and ends, for all markets, in December 2009. The appendix gives more detailed information about the samples. In total, the analysis is based on an initial data set of about 14,000 stocks accounting for 25 million firm days. The analysis closely mirrors the US approach in panel A of table 4. Main findings from in total about 200,000 round-trip trades are displayed in table 5.

Please insert table 5

In the recent past, even traditional pairs trading seems to have been profitable in all countries. Annualized return estimates range from 6% (Italy) to more than 13% (Germany, France). While these results are interesting in their own way, we again focus on the role of investor attention, for which we find strong evidence. With the exception of Japan, the return on pairs opening on low (high) distraction days is smaller (larger) than average sample returns in every single country. The return difference between decile 10 and decile 1 is, with the exception of Italy and Japan, persistently and strongly significant and economically meaningful.

Delving further into this analysis, panel B of table 5 reveals that the major driver of the results is again a higher likelihood of convergence. From low distraction days to high distraction days, the fraction of converging pairs increases by between 12% (from 43.8% to 49.1% in Japan) and 87% (from 15.9% to 29.8% in Hongkong). Remarkably, while pairs trading generates seemingly abnormal returns in all countries, its nature appears to differ substantially e.g. with regard to the number of pairs traded or the (unconditional) fraction of converging pairs. Exploring the sources and consequences of these differences might be an interesting field for further research. In any case, the results strongly confirm the baseline results obtained for the US market.

4.4. *Robustness checks and further insights*

The main insights from a number of sensitivity checks are reported in table 6.

Please insert table 6

Subperiod analysis In panel A, we repeat the baseline analysis for three consecutive subperiods of approximative equal length. In line with results from previous work (Gatev et al. (2006), Do and Faff (2010)), returns to traditional pairs trading seem to decline over time. More importantly though, in all cases, returns originating from divergence on high distraction days are significantly higher than those obtained on low distraction days.

Variations in the data set We control for the impact of the utility sector in two ways. First, we rerun the baseline analysis, excluding utility stocks. Second, we identify the monthly top 100 pairs under the constraint that each firm is only considered once at maximum. This approach not only decreases the fraction of utility stocks from 30% to roughly 17%, but also changes the composition of the data set considerably. Panel B of table 6 verifies that inferences do not change.

Matching procedure We analyze a subsample of trades specifically designed to isolate the impact of variations in attention shifts. We restrict our focus to those pairs which happen to diverge both at least once on a low and on a high distraction day, with the time difference between the average dates of both type of events being less than a year. This leaves 5,488 trades in the data set. This procedure controls for all firm and pair-level variables, including unobserved ones, that do not vary within this time period. Findings displayed in panel C verify that inferences from the baseline analysis still hold.

Limits to arbitrage It is worth noting that many variables controlled for in the multivariate regressions in Section 4.2. already proxy for market frictions. This applies to both cross-sectional controls such as idiosyncratic risk and to time-series controls such as market volatility or average bid-ask spreads.

In addition, the appendix explores the relation between pairs trading profitability, attention shifts, and all measures of time-varying arbitrage constraints outlined in Section 2.3., which are not available over the whole sample period. The difference between high and low distraction days remains highly statistically and economically significant.

These findings also suggest that short-selling constraints do not drive our findings. This is supported by Gatev et al. (2006) who conclude that for institutional investors “pairs trading profits are likely to remain essentially unaffected by potential shorting costs” (p. 825). Also Engelberg et al. (2009) summarize that, due to the fact that the typical firm in pairs trading is large and liquid, “short-sale constraints are not a major friction” (p.

28). In addition, about 60% of the impact of the attention shift proxy are attributable to the long leg of the pairs trading portfolio. Finally, untabulated findings show that the time-series of short interest in the stocks which are part of a pair trade in a given month is essentially uncorrelated with the monthly attention shift proxy.

Yet another approach to control for frictions is to focus on subsamples for which limits to arbitrage are arguably less binding. Panel D exemplarily does so by restricting the analysis to pairs opening on days with positive market returns and to pairs with below median idiosyncratic risk, respectively. Inferences remain unaffected.

Return factor exposure In an attempt to transfer the event-time results of the baseline analysis to calendar time, we extend the maximum holding period from one month to six months. Doing so works strongly against finding differences across attention shift deciles, but assures two smooth time-series of returns on pairs opening on high and low distraction days, respectively. In the first model, we regress the resulting time-series return difference on the Fama and French (1993) factors. The second model additionally includes the traded liquidity factor constructed in Pástor and Stambaugh (2003) as well as factors for short-term reversal, medium-term momentum, and long-term reversal. Untabulated results show that there is no notably loading on any risk premium. However, panel E of table 6 verifies that alphas are persistently statistically significant.

Alternative econometric approach We run Fama/MacBeth-type regressions by first estimating yearly pooled cross-sectional regressions of one-month pairs returns on attention shift decile ranks and then using the time-series of the resulting coefficient to assess its statistical significance. As panel F shows, the coefficient is positive in about 75% of all years, highly statistically significant, and economically meaningful.

Modified proxies The appendix shows that inferences do not change if we rely on any of the twelve modified attention shift proxies briefly described in Section 2.4.. These insights are important in as they represent a generalization of the findings so far.

For instance, results cannot be explained with time-varying behavior of firm-level idiosyncratic volatility (e.g. Brandt et al. (2010), Fink et al. (2010)). Using an orthogonalized proxy returns still yields significant results. Findings are also not attributable to the joint use of industry classifications for both pair construction and market segments. Using portfolios (double-)sorted on size as well as book-to-market or short-term reversal leads

to very similar findings. Note that we later additionally verify that the effect is also significant for pairs constructed solely from firms of the same industry. Thus, findings are not driven by e.g. investors reacting properly to news about one industry but over-reacting to news about another industry during times of distraction.

Further time-series insights into the role of limited attention The appendix explores the role of previously proposed attention proxies on the profitability of pairs trading. Findings show that the overall impact of Fridays, down market periods, the number of competing events, shocks in internet search queries, and holidays reflects our predictions.

Cross-sectional insights into the role of limited attention We expect the return sensitivity to changes in investor attention to be positively related to the degree of informational frictions between the constituents of the pair. We expect these frictions to be lower for pairs consisting of stocks from the same industry or even consisting of value-weighted industry portfolios. Panel A of table 7 shows findings supporting our line of reasoning. Compared to the baseline pairs universe, the impact of the attention shift proxy is about 10% lower, though still highly statistically and economically significant, for top pairs consisting of stocks from the same industry. It is considerably lower, though still significant, for pairs constructed from industry portfolios. The finding of apparently slow information flow between whole industries complements recent insights from studies such as Hong et al. (2007) or Menzly and Ozbas (2010).

Please insert table 7

Panel B shows findings from a test similar in nature. While pairs are based on firms from different industries, some of them operate at the intra-firm level in some common business segments. The economic link for these pairs will arguably be more visible, which makes it comparatively less likely that shocks in limited attention will cause prices to diverge. We follow Berger and Ofek (1995) and Cohen and Lou (2011) to extract data from relevant segment reporting regulations which start in 1977. Findings reported in panel B provide supportive, although not persistently significant results.

Press coverage The extent of a firm’s media coverage appears positively linked to the speed of information diffusion as well as visibility (e.g. Peress (2008), Fang and Peress

(2009), Huberman and Regev (2001)). Therefore, press coverage could help to keep relative prices in line, especially in turbulent moments.

We rely on the Dow Jones News Service (DJNS) database, “the best approximation of public news for traders” (Chan (2003), p. 230), to identify highly and lowly covered pairs.¹² Panel C of table 7 shows that our predictions largely prove to be true. The one-month return difference between pairs receiving disproportionately much coverage and those widely neglected reaches at least 80 basis points. In fact, trading highly covered pairs turns out to be completely unprofitable, whereas trading lowly covered pairs is considerably more profitable compared to trading the average pair. Moreover the sensitivity to changes in the level of investor attention is higher for lowly covered pairs.

5. Conclusion

Understanding how markets process information is one of the major concerns of financial economics. Our insights are broadly in line with the idea of time-varying investor attention allocation being an important source of friction in the price formation process.

The magnitude of phenomena likely to be driven by slow diffusion of information appears negatively related to investor attention. This holds true both for individual stocks, as in the case of the PEAD, and for the relative efficiency of economically linked stocks, as in the case of pairs trading. This mechanism might also be interpreted by referring to Grossman and Stiglitz (1980): in moments in which gathering and correctly processing valuable firm-specific information is particularly costly, the market has to provide higher payoffs as compensation. In contrast, the magnitude of overreaction-driven anomalies such as arguably momentum might be positively related to the extent of investor attention. Attention then may exacerbate behavioral biases, thereby amplifying seemingly abnormal returns. Any competing interpretation faces the hurdle of convincingly explaining these contrasting effects as well as their robustness and economic magnitude.

¹²For each firm that meets the data requirements on pairs trading at some point after 1990, we collect the yearly number of news articles between 1991 and 2008. As this number is strongly positively related to firm size (e.g. Fang and Peress (2009)), we use the top and bottom quintile of residuals from yearly regressions of $\ln(1+\text{number of news})$ on $\ln(\text{average market capitalization})$. Finally, we define a pair as being highly (lowly) covered, if both of its components are firms with high (low) coverage in the year the divergence occurs.

References

- Amihud, Y., 2002, “Illiquidity and stock returns: Cross-section and time-series effects,” *Journal of Financial Markets*, 5, 31–56.
- Andrade, S. C., V. di Pietro, and M. S. Seasholes, 2005, “Understanding the profitability of pairs trading,” *Unpublished working paper, U.C. Berkeley, Northwestern University*.
- Ang, A., S. Gorovyy, and G. B. van Inwegen, 2011, “Hedge fund leverage,” *Journal of Financial Economics*, 102, 102–126.
- Antoniou, C., J. A. Doukas, and A. Subrahmanyam, 2010, “Sentiment and momentum,” *Unpublished working paper, University of Exeter, Old Dominion University, UCLA*.
- Asness, C. S., 1997, “The interaction of value and momentum strategies,” *Financial Analysts Journal*, 61, 29–36.
- Asness, C. S., T. J. Moskowitz, and L. H. Pedersen, 2012, “Value and momentum everywhere,” *Journal of Finance*, forthcoming.
- Avramov, D., T. C. and Gergana Jostova, and A. Philipov, 2007, “Momentum and credit rating,” *Journal of Finance*, 62, 2503–2520.
- Baker, M., and J. Wurgler, 2006, “Investor sentiment and the cross-section of stock returns,” *Journal of Finance*, 61, 1645–1680.
- Baker, M., J. Wurgler, and Y. Yuan, 2012, “Global, local, and contagious investor sentiment,” *Journal of Financial Economics*, 104, 272–287.
- Bandarchuk, P., and J. Hilscher, 2013, “Sources of momentum profits: Evidence on the irrelevance of characteristics,” *Review of Finance*, 17, 809–845.
- Barber, B. M., and T. Odean, 2008, “All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors,” *Review of Financial Studies*, 21, 785–818.
- Barberis, N., and A. Shleifer, 2003, “Style investing,” *Journal of Financial Economics*, 68, 161–199.
- Barberis, N., A. Shleifer, and J. Wurgler, 2005, “Comovement,” *Journal of Financial Economics*, 75, 283–317.
- Ben-David, I., F. Franzoni, and R. Moussawi, 2012, “Hedge fund stock trading in the financial crisis of 20072009,” *Review of Financial Studies*, 25, 1–54.
- Berger, P. G., and E. Ofek, 1995, “Diversification’s effect on firm value,” *Journal of Financial Economics*, 37, 39–65.
- Bernard, V. L., and J. K. Thomas, 1989, “Post-earnings-announcement drift: Delayed price response or risk premium?,” *Journal of Accounting Research*, 27, 1–36.
- Boyer, B. H., 2011, “Style related comovement: Fundamentals or labels?,” *Journal of Finance*, 66, 307–332.
- Brandt, M., A. Brav, J. R. Graham, and A. Kumar, 2010, “Idiosyncratic volatility puzzle: Time trend or speculative episodes?,” *Review of Financial Studies*, 23, 863–899.
- Brunnermeier, M. K., S. Nagel, and L. H. Pedersen, 2008, “Carry trades and currency crashes,” *NBER Macroeconomics Annual 2008*, 23, 313–347.
- Brunnermeier, M. K., and L. H. Pedersen, 2009, “Market liquidity and funding liquidity,” *Review of Financial Studies*, 22, 2201–2238.
- Buraschi, A., E. Sener, and M. Menguturk, 2012, “The dynamics of limits to arbitrage: Evidence from international cross-sectional data,” *Unpublished working paper, London Business School, Ozyegin University*.

- Byun, S. J., S. S. Lim, and S. H. Yun, 2013, "Continuing overreaction and stock return predictability," *Unpublished working paper, KAIST College of Business, DePaul University*.
- Campbell, J. Y., J. Hilscher, and J. Szilagyi, 2008, "In search of distress risk," *Journal of Finance*, 63, 2899–2939.
- Chan, L. K., J. Lakonishok, and B. Swaminathan, 2007, "Industry classifications and return comovement," *Financial Analysts Journal*, 63, 56–70.
- Chan, W. S., 2003, "Stock price reaction to news and no-news: Drift and reversal after headlines," *Journal of Financial Economics*, 70, 223–260.
- Chay, J., and C. A. Trzcinka, 1999, "Managerial performance and the cross-sectional pricing of closed-end funds," *Journal of Financial Economics*, 52, 379–408.
- Chen, H., S. Chen, and F. Li, 2012, "Empirical investigation of an equity pairs trading strategy," *Unpublished working paper, University of British Columbia, University of Michigan*.
- Cherkes, M., J. Sagi, and R. Stanton, 2009, "A liquidity-based theory of closed-end funds," *Review of Financial Studies*, 22, 257–297.
- Chordia, T., and L. Shivakumar, 2002, "Momentum, business cycle, and time-varying expected returns," *Journal of Finance*, 57, 985–1019.
- Chui, A. C., S. Titman, and K. J. Wei, 2010, "Individualism and momentum around the world," *Journal of Finance*, 65, 361–392.
- Cohen, L., and A. Frazzini, 2008, "Economic links and predictable returns," *Journal of Finance*, 63, 1977–2011.
- Cohen, L., and D. Lou, 2011, "Complicated firms," *Journal of Financial Economics*, 104, 383–400.
- Cooper, M. J., O. Dimitrov, and R. Rau, 2001, "A rose.com by any name," *Journal of Finance*, 56, 2371–2388.
- Cooper, M. J., H. Gulen, and P. R. Rau, 2005, "Changing names with style: Mutual fund name changes and their effect on fund flows," *Journal of Finance*, 60, 2825–2858.
- Cooper, M. J., R. C. G. Jr., and A. Hameed, 2004, "Market states and momentum," *Journal of Finance*, 59, 1345–1365.
- Corwin, S. A., and P. Schultz, 2012, "A simple way to estimate bid-ask spreads from daily high and low prices," *Journal of Finance*, 67, 719–759.
- Da, Z., J. Engelberg, and P. Gao, 2011, "In search of attention," *Journal of Finance*, 66, 1461–1499.
- Daniel, K., D. Hirshleifer, and A. Subrahmanyam, 1998, "Investor psychology and security market under- and overreactions," *Journal of Finance*, 53, 1839–1885.
- Daniel, K., and S. Titman, 1999, "Market efficiency in an irrational world," *Financial Analysts Journal*, 55, 28–40.
- DellaVigna, S., and J. M. Pollet, 2009, "Investor inattention and Friday earnings announcements," *Journal of Finance*, 64, 709–749.
- Do, B., and R. Faff, 2010, "Does simple pairs trading still work?," *Financial Analysts Journal*, 66, 83–95.
- Do, B. H., and R. W. Faff, 2012, "Are pairs trading profits robust to trading costs?," *Journal of Financial Research*, 35, 261–287.
- Engelberg, J., P. Gao, and R. Jagannathan, 2009, "An anatomy of pairs trading: The role of idiosyncratic news, common information and liquidity," *Unpublished working paper, University of North Carolina, University of Notre Dame, Northwestern University*.

- Engelberg, J., C. Sasseville, and J. Williams, 2011, "Market madness? The case of mad money," *Management Science*, 58, 351–364.
- Fama, E. F., and K. R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.
- , 1997, "Industry costs of equity," *Journal of Financial Economics*, 43, 153–193.
- , 2008, "Dissecting anomalies," *Journal of Finance*, 63, 1653–1678.
- Fang, L., and J. Peress, 2009, "Media coverage and the cross-section of stock returns," *Journal of Finance*, 64, 2023–2052.
- Fink, J., K. E. Fink, G. Grullon, and J. P. Weston, 2010, "What drove the increase in idiosyncratic volatility during the internet boom?," *Journal of Financial and Quantitative Analysis*, 45, 1253–1278.
- Frazzini, A., and L. H. Pedersen, 2013, "Betting against beta," *Unpublished working paper, AQR Capital Management, New York University*.
- Froot, K. A., and E. M. Dabora, 1999, "How are stock prices affected by the location of trade?," *Journal of Financial Economics*, 53, 189–216.
- Gagnon, L., and G. A. Karolyi, 2010, "Multi-market trading and arbitrage," *Journal of Financial Economics*, 97, 53–80.
- Gallant, R. A., 1987, "Nonlinear statistical models," *Wiley, New York*.
- Gatev, E. G., W. N. Goetzmann, and K. G. Rouwenhorst, 2006, "Pairs trading: Performance of a relative-value arbitrage rule," *Review of Financial Studies*, 19, 797–827.
- George, T. J., and C.-Y. Hwang, 2004, "The 52-week high and momentum investing," *Journal of Finance*, 59, 2145–2176.
- Green, T. C., and B.-H. Hwang, 2009, "Price-based return comovement," *Journal of Financial Economics*, 93, 37–50.
- Greenwood, R., 2008, "Excess comovement of stock returns: Evidence from cross-sectional variation in Nikkei 225 weights," *Review of Financial Studies*, 21, 1153–1186.
- Grinblatt, M., and T. M. Moskowitz, 2004, "Predicting stock price movements from past returns: the role of consistency and tax-loss selling," *Journal of Financial Economics*, 71, 541–579.
- Grossman, S. J., and J. E. Stiglitz, 1980, "On the impossibility of informationally efficient markets," *American Economic Review*, 70, 393–408.
- Hillert, A., H. Jacobs, and S. Müller, 2013, "Media makes momentum," *Unpublished working paper, University of Mannheim*.
- Hirshleifer, D., S. S. Lim, and S. H. Teoh, 2009, "Driven to distraction: Extraneous events and underreaction to earnings news," *Journal of Finance*, 64, 2289–2325.
- Hirshleifer, D., S. S. Lim, and S. H. Teoh, 2011, "Limited investor attention and stock market misreactions to accounting information," *Review of Asset Pricing Studies*, 1, 35–73.
- Hong, H. G., T. Lim, and J. C. Stein, 2000, "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies," *Journal of Finance*, 55, 265–295.
- Hong, H. G., W. Torous, and R. Valkanov, 2007, "Do industries lead stock markets?," *Journal of Financial Economics*, 83, 367–396.

- Hou, K., 2007, "Industry information diffusion and the lead-lag effect in stock returns," *Review of Financial Studies*, 20, 1113–1138.
- Hou, K., L. Peng, and W. Xiong, 2006, "R2 and Price Inefficiency," *Unpublished working paper, Ohio State University, City University of New York, Princeton University*.
- , 2009, "A tale of two anomalies: The implications of investor attention for price and earnings momentum," *Unpublished working paper, Ohio State University, City University of New York, Princeton University*.
- Huberman, G., and T. Regev, 2001, "Contagious speculation and a cure for cancer: A non-event that made stock prices soar," *Journal of Finance*, 56, 387–396.
- Jegadeesh, N., and S. Titman, 1993, "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance*, 48, 65–91.
- Jiang, G., C. M. Lee, and Y. Zhang, 2005, "Information uncertainty and expected returns," *Review of Accounting Studies*, 10, 185–221.
- Jong, A. D., L. Rosenthal, and M. A. V. Dijk, 2009, "The risk and return of arbitrage in dual-listed companies," *Review of Finance*, 13, 495–520.
- Kahneman, D., 1973, "Attention and effort," *Prentice Hall, New Jersey*.
- Kumar, A., 2009, "Who gambles in the stock market?," *Journal of Finance*, 64, 1889–1933.
- Lamont, O. A., and R. H. Thaler, 2003, "Can the market add and subtract? Mispricing in tech stock carve-outs," *Journal of Political Economy*, 111, 227–268.
- Lee, C. M., A. Shleifer, and R. H. Thaler, 1991, "Investor sentiment and the closed-end fund puzzle," *Journal of Finance*, 46, 75–109.
- Lee, C. M., and B. Swaminathan, 2000, "Price momentum and trading volume," *Journal of Finance*, 55, 2017–2069.
- Lo, A. W., and C. A. MacKinlay, 1990, "When are contrarian profits due to stock market overreaction?," *Review of Financial Studies*, 3, 175–205.
- Long, J. B. D., A. Shleifer, L. H. Summers, and R. J. Waldmann, 1990, "Noise trader risk in financial markets," *Journal of Political Economy*, 98, 703–738.
- Lou, D., 2011, "Attraction investor attention through advertising," *Unpublished working paper, London School of Economics*.
- Lou, D., and C. Polk, 2012, "Comomentum: Inferring arbitrage capital from return correlations," *Unpublished working paper, London School of Economics*.
- Menzly, L., and O. Ozbas, 2010, "Market segmentation and cross-predictability of returns," *Journal of Finance*, 65, 1555–1580.
- Mitchell, M., T. Pulvino, and E. Stafford, 2002, "Limited arbitrage in equity markets," *Journal of Finance*, 57, 551–584.
- Mondria, J., and T. Wu, 2012, "Asymmetric attention and stock returns," *Unpublished working paper, University of Toronto, University of California*.
- Moskowitz, T., Y. H. Ooi, and L. H. Pedersen, 2012, "Time series momentum," *Journal of Financial Economics*, 104, 228–250.
- Mullainathan, S., 2002, "Thinking through categories," *Unpublished working paper, MIT*.
- Novy-Marx, R., 2012, "Is momentum really momentum?," *Journal of Financial Economics*, 103, 429–453.

- Pástor, L., and R. F. Stambaugh, 2003, "Liquidity risk and expected stock returns," *Journal of Political Economy*, 111, 642–685.
- Peng, L., 2005, "Learning with information capacity constraints," *Journal of Financial and Quantitative Analysis*, 40, 307–329.
- Peng, L., and W. Xiong, 2006, "Investor attention, overconfidence and category learning," *Journal of Financial Economics*, 80, 563–602.
- Peress, J., 2008, "Media coverage and investors' attention to earnings announcements," *Unpublished working paper, Insead*.
- Pirinsky, C., and Q. Wang, 2006, "Does corporate headquarters location matter for stock returns?," *Journal of Finance*, 61, 1991–2015.
- Pontiff, J., 1995, "Closed-end fund premia and returns implications for financial market equilibrium," *Journal of Financial Economics*, 37, 341–370.
- , 1996, "Costly arbitrage: Evidence from closed-end funds," *Quarterly Journal of Economics*, 1135.
- , 2006, "Costly arbitrage and the myth of idiosyncratic risk," *Journal of Accounting and Economics*, 42, 35–52.
- Rosenthal, L., and C. Young, 1990, "The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever N.V./PLC," *Journal of Financial Economics*, 26, 123–141.
- Sagi, J. S., and M. S. Seasholes, 2007, "Firm-specific attributes and the cross-section of momentum," *Journal of Financial Economics*, 84, 389–434.
- Schultz, P., and S. Shive, 2010, "Mispricing of dual-class shares: Profit opportunities, arbitrage, and trading," *Journal of Financial Economics*, 98, 524–549.
- Scruggs, J. T., 2007, "Noise trader risk: Evidence from the Siamese twins," *Journal of Financial Markets*, 10, 76–105.
- Seasholes, M. S., and G. Wu, 2007, "Predictable behavior, profits, and attention," *Journal of Empirical Finance*, 14, 590–610.
- Smith, B. F., and B. Amoako-Adu, 1995, "Relative prices of dual class shares," *Journal of Financial and Quantitative Analysis*, 30, 223–239.
- Stambaugh, R. F., J. Yu, and Y. Yuan, 2012, "The short of it: Investor sentiment and anomalies," *Journal of Financial Economics*, 104, 288–302.
- Stivers, C., and L. Sun, 2010, "Cross-sectional return dispersion and time variation in value and momentum premiums," *Journal of Financial and Quantitative Analysis*, 45, 987–1014.
- Verardo, M., 2009, "Heterogeneous beliefs and momentum profits," *Journal of Financial and Quantitative Analysis*, 44, 795–822.
- Vidyamurthy, G., 2004, "Pairs trading: Quantitative Method and Analysis," *John Wiley & Sons*.
- Wang, K. Q., and J. Xu, 2010, "Time-varying momentum profitability," *Unpublished working paper, University of Toronto, University of Beijing*.
- Yuan, Y., 2012, "Market-wide attention, trading, and stock returns," *Unpublished working paper, University of Pennsylvania*.
- Zhang, X. F., 2006, "Information uncertainty and stock returns," *Journal of Finance*, 61, 105–137.
- Zingales, L., 1995, "What determines the value of corporate votes?," *Quarterly Journal of Economics*, 110, 1047–1073.

Figure 1: Return anomalies following investor attention shifts

This figure compares the average magnitude of different return anomalies conditioned on lagged values of the baseline attention shift proxy proposed in this paper. The proxy is based on the dispersion of unanticipated daily industry return shocks. Details are provided in Section 2.. For each year separately, a decile-based rank is assigned to each trading day. In the case of the PEAD as well as pairs trading, high (low) distraction days are defined as decile rank 10 (1). In the case of momentum, average monthly values of the daily attention shift proxy are computed. For each year separately, a rank from 1 to 6 is then assigned to two months each. High (low) distraction months are defined as rank 6 (1). Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. In the case of the PEAD as well as pairs trading, standard errors are clustered by day. In the case of momentum, statistical inferences are based on Gallant (1987) standard errors.

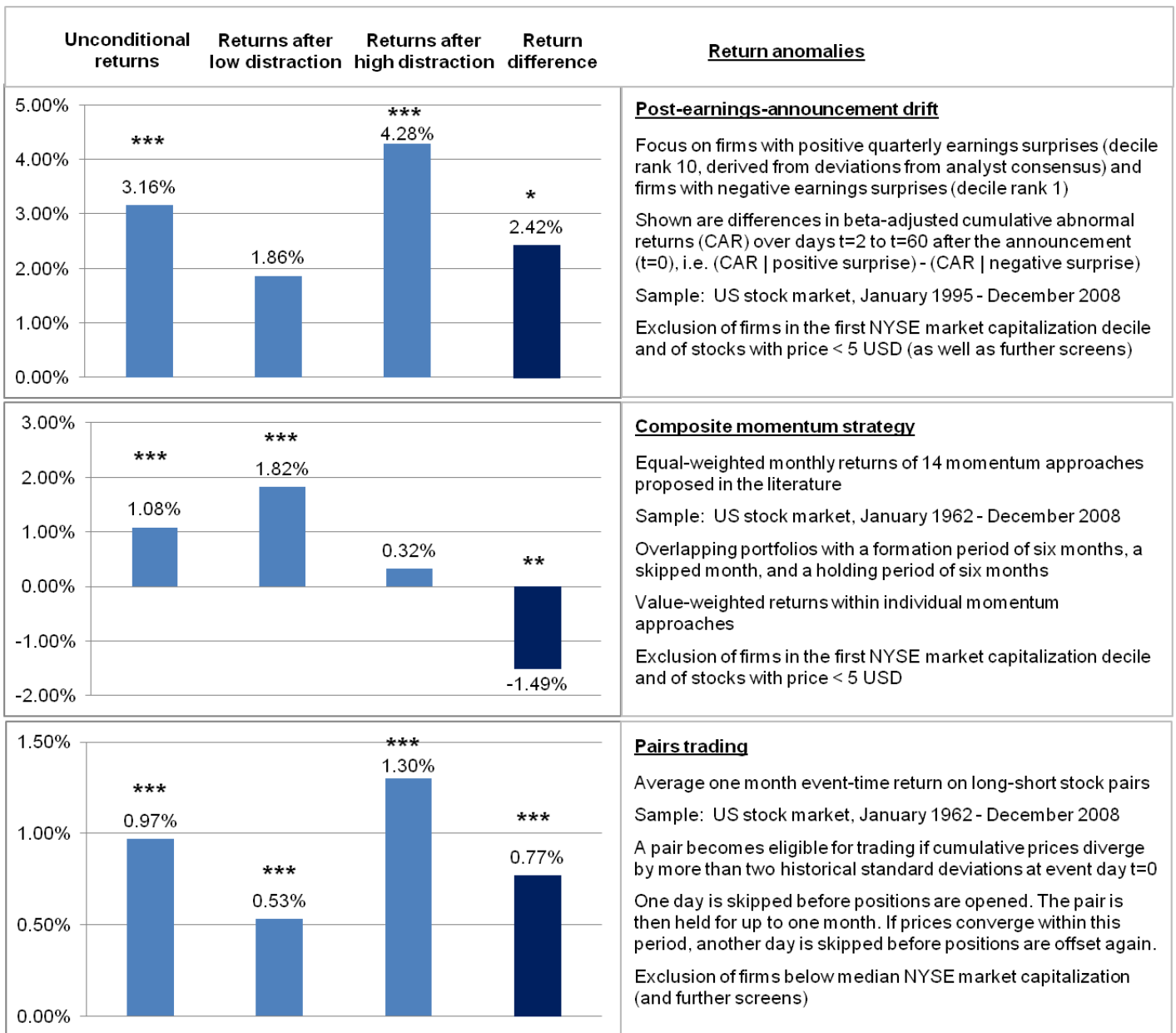
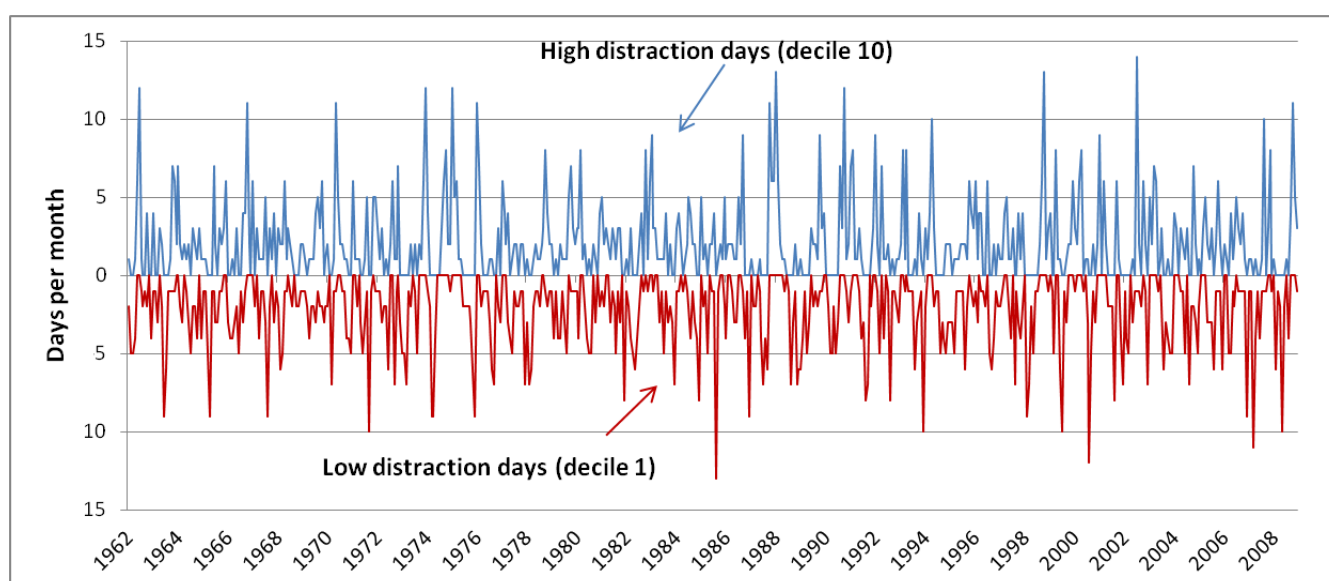


Figure 2: Time-Series Characteristics of the Attention Shift Proxy: High and Low Distraction Days

The graph shows the monthly number of high and low distraction days proxy from January 1962 to December 2008. For the construction of the time-series, a four-step procedure is employed. First, for January 1960 to December 2008, we compute daily value-weighted returns for the 49 Fama and French (1997) industries. Second, we construct daily return shocks defined as the absolute difference between the actual industry return and its expected return as implied by an OLS market model. Parameter estimates are obtained from rolling time-series regressions based on daily return data over the previous year. Third, shocks are condensed into a single ratio. To this end, industry shocks are weighted by the inverse of the volatility of their shock variable over the previous year. Fourth, we assign a decile-based rank to each trading day for each year separately. Days with decile rank 10 (1) are referred to as high (low) distraction days.



Spearman rank order correlation coefficients between the decile ranks of the attention shift proxy and the rank order of market-level variables based on daily data are as follows. The first number refers to the raw value of the market-level variables, the second to their yearly decile rank (computed as for the attention shift proxy). The sample period comprises all trading days between January 1962 and December 2008.

Market excess return	Small firm factor	Value/growth factor	Momentum factor	Short-term reversal factor
0.04/0.07	-0.08/-0.06	-0.00/-0.00	-0.12/-0.10	0.03/0.00
Average bid-ask spread	Squared market return	Equal-weighted turnover	Rolling 10 day volatility	
0.15/0.08	0.14/0.23	0.14/0.25	0.22/0.27	

Details on the construction of the following variables are given in the appendix.

TED Spread (since 1/1986)	VIX (since 1/1990)	Moody's BAA-AAA (since 1/1986)
0.12/0.12	0.23/0.31	0.08/0.15
LIBOR 3M (since 1/1986)	Term Spread (since 1/1982)	IB Abnormal Return (since 1/1982)
-0.01/-0.05	0.02/0.02	0.05/0.04

Figure 3: Illustration of Pairs Trading Process

This figure illustrates the trading process with two examples. Each month, the 100 pairs with minimum distance between normalized twelve-month daily return indices are selected. They are then eligible for trading in the immediately following six-month evaluation period. At the beginning of this period, prices are set to equal unity. If the spread between the cumulative return series of the two stocks exceeds two historical standard deviations (as estimated in the estimation period), we take a long position in the relatively underpriced stock, which is financed by short-selling the relatively overpriced stock. The self-financing pair is then held for up to one month. If prices converge before this cut-off date, the trade is closed with a gain. If prices do not converge within a month, positions are offset, which results in a loss if prices diverge even further. A pair may open several more times during the trading period. In this case, the trading process is repeated as outlined above.

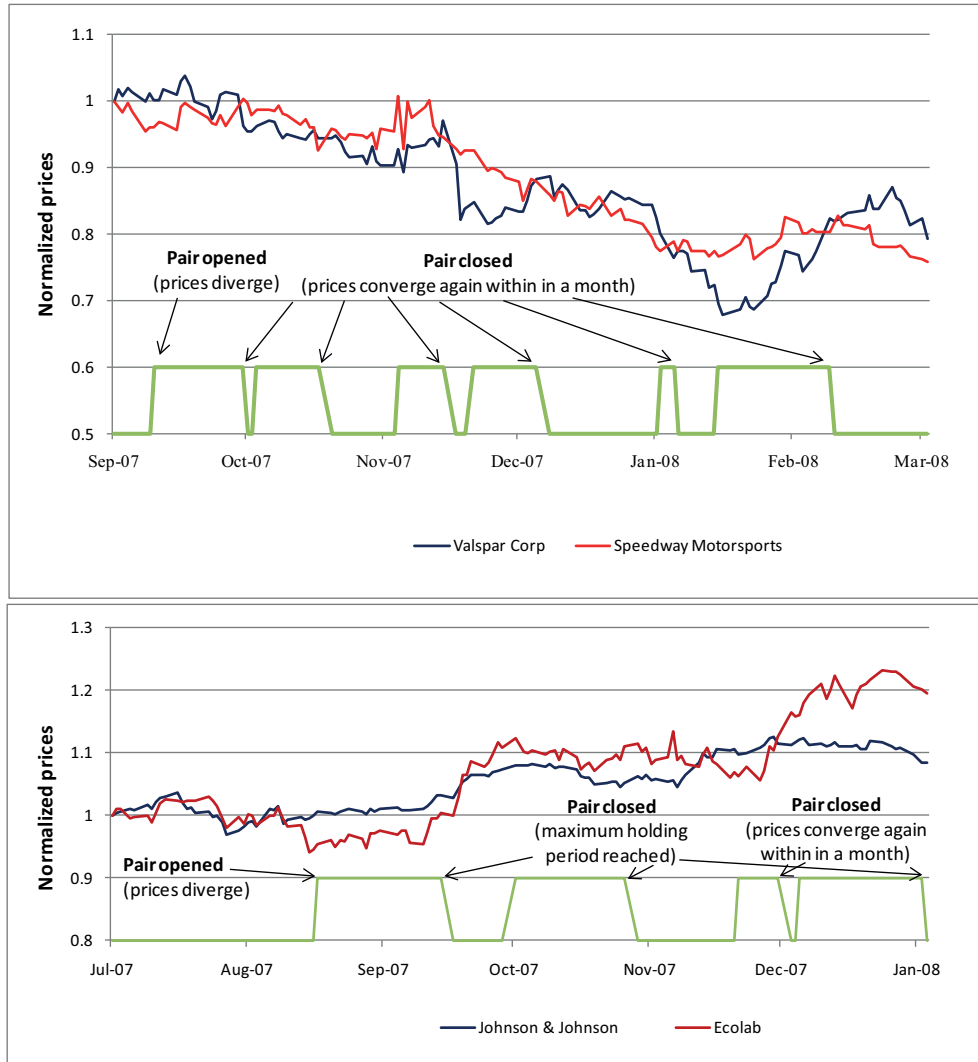


Figure 4: Probability of Convergence and Average Daily Return by Event Day

The upper graph shows the empirical probability of US stock pairs converging on a given event day after divergence. See Section 4.1. for a definition of divergence and convergence. The lower graph shows the average daily return of open pairs in event time. Both figures are based on more than 100,000 round-trip trades between January 1962 and December 2008 (see Section 4.2.).

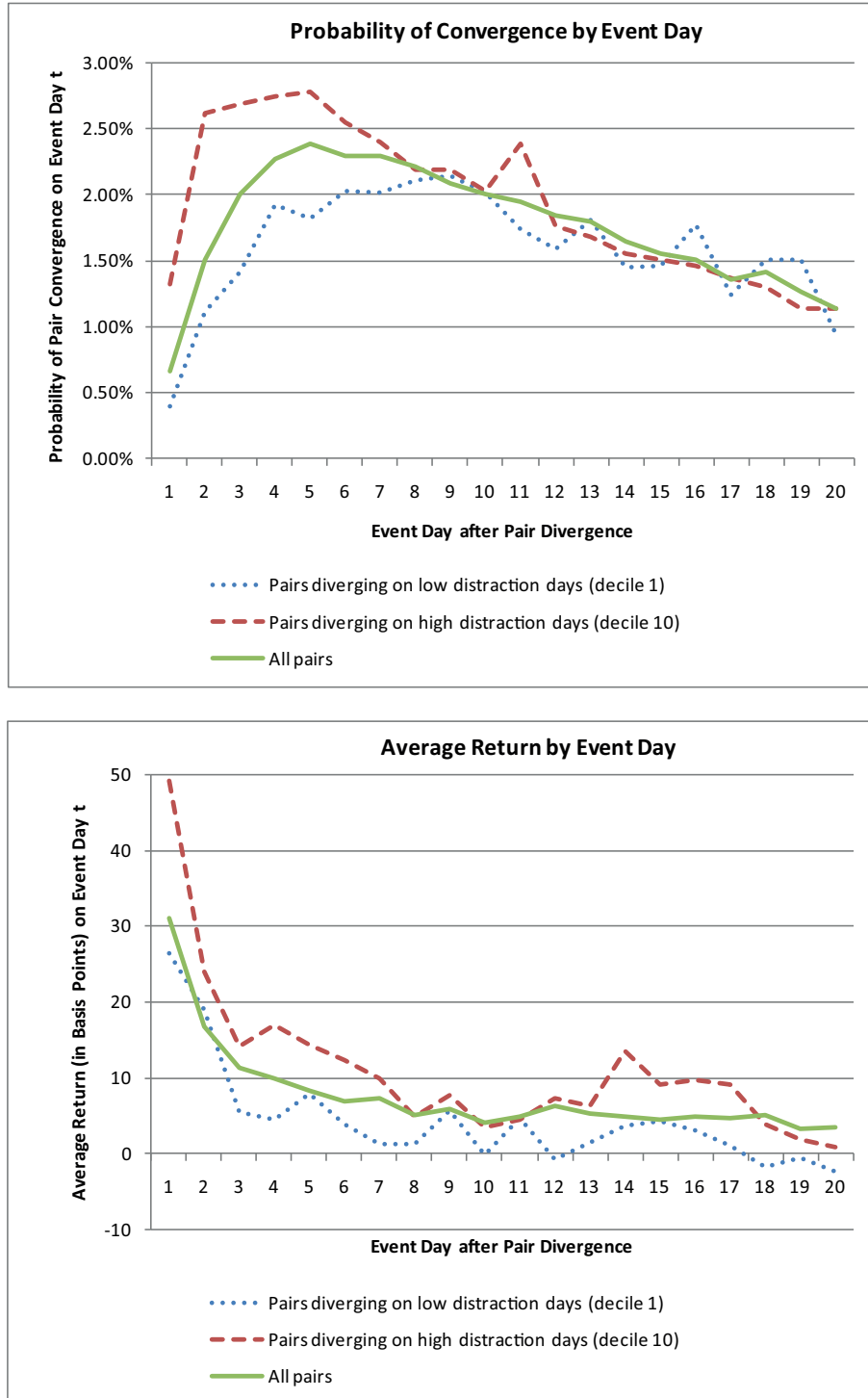


Table 1: Post-earnings Announcement Drift (PEAD) and Attention shifts

This table summarizes the main findings from an analysis intended to explore the ability of the attention shift proxy decile ranks to predict the PEAD. Displayed are selected coefficients from multivariate regressions of pooled cumulative abnormal returns over days $t=2$ to $t=60$ after earnings announcements ($t=0$) on quarterly computed earnings surprise decile ranks (ES), attention shift proxy decile ranks, the interaction effect as the main variable of interest, and up to five sets of control variables. All controls are also interacted with earnings surprise decile ranks (see equation 3). Quarterly earnings announcement data from the I/B/E/S details tape and Compustat are obtained for the period between January 1995 and December 2008. We only consider firms with a market capitalization larger than the first NYSE decile and a stock price larger than 5 USD at the beginning of the previous month. Please refer to Section 3.1. for a detailed description of further filter rules aiming at minimizing the impact of possible data errors or microstructural effects. Earnings surprises are based on scaled deviations of actual earnings from the analyst consensus, as described in detail in Section 3.1.. Cumulative abnormal returns are computed as the raw buy-and-hold return of the stock over the event window minus its expected buy-and-hold beta-adjusted return. In panel A, we consider the following *firm-level controls*: market capitalization, industry membership, market-to-book ratio, the Campbell et al. (2008) distress measure, the daily bid-ask spread measure of Corwin and Schultz (2012), the number of analysts, parameters from the estimation of expected returns (beta, alpha, coefficient of determination), turnover, the Amihud (2002) illiquidity ratio, share price, dummies for Nasdaq and S&P 500 membership, dummies for lottery and non-lottery stocks as in Kumar (2009), idiosyncratic volatility, and the reporting lag. Details regarding the construction are reported in the appendix. In panel B, we (additionally) control for time-series variables. *Previous attention proxies* contain a Friday dummy, deciles for the number of same-day earnings announcements, and a dummy for a positive cumulative three-year market return. Calendar effects include dummies for the year of the announcement, the month of the year, as well as the day of the week. *Market-level controls* and controls for *limits to arbitrage* are explained in detail in Section 2.3.. They include the Fama and French (1993) factor, a momentum factor, a factor for short-term reversal, the squared market return, market turnover, rolling 10-day volatility, a measure for the market-wide bid-ask spread, investment bank abnormal returns, the 3-month LIBOR, the TED spread, a credit spread factor, the term spread, and the VIX. Standard errors are adjusted for heteroscedasticity and clustered by day of pair divergence. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: The interaction between attention shifts and the PEAD, cross-sectional controls				
Model specification	All earnings surprises		Only surprise deciles 1 and 10	
	Univariate	All firm controls	Univariate	All firms controls
	(1)	(2)	(3)	(4)
Attention shift proxy decile rank	0.0433 (0.64)	0.000361 (0.01)	0.00660 (0.06)	-0.0349 (-0.32)
Earnings surprise decile rank (ES)	0.249*** (4.63)	0.364 (0.53)	0.191** (2.36)	0.675 (0.70)
$ES \times$ Attention shift proxy decile rank	0.0147* (1.68)	0.0270*** (3.04)	0.0269** (2.05)	0.0372*** (2.82)
Constant	-2.066*** (-4.78)	-10.45** (-2.24)	-1.571** (-2.30)	-17.08** (-2.35)
Observations	101,812	101,812	20,342	20,342
R-squared	0.003	0.023	0.006	0.031
Panel B: The interaction between attention shifts and the PEAD, further controls				
Model specification	All earnings surprises			
	(1)	(2)	(3)	(4)
$ES \times$ Attention shift proxy decile rank	0.0286*** (3.00)	0.0207** (2.36)	0.0317*** (2.98)	0.0374*** (3.54)
Previous attention proxies	no	no	yes	yes
Calendar effects	yes	no	yes	yes
Market-level controls	no	no	yes	yes
Limits to arbitrage	no	yes	yes	yes
Firm-level controls	no	no	no	yes

Table 2: Momentum Returns and Attention Shifts

This table reports the main findings from time-series regressions of the monthly returns of a composite momentum strategy on one-month lagged attention shift ranks (1-6) and on a number of control variables. Attention shift ranks are computed in two steps. First, average monthly values of the baseline (daily) attention shift proxy as described in section 2.1. are computed. Second, for each year separately, these monthly values are sorted in ascending order and assigned values from 1 to 6. Thus, two months in each year are assigned the same attention shift rank. The returns of the composite momentum strategy are expressed in percent per month and determined as the equal-weighted returns of fourteen (enhanced) momentum strategies, as described in detail in Section 3.2. and the appendix. Individual momentum strategies are computed using a formation period of six months, a skipped month, and a holding period of six months with overlapping portfolios. Stocks eligible for entering any momentum strategy are required to have a nominal share price of at least 5 USD and a market capitalization larger than the first NYSE decile at the time of portfolio formation. Returns of each individual momentum-strategy are value-weighted. Model 1 does not contain any further variables. Model 2 includes the Fama and French (1993) factors. Model 3 to 6 augment model 2 with further control variables. In model 3, these comprise factors for short term reversal, long term reversal, and liquidity. They moreover include the squared market return, market turnover, market volatility, and average bid-ask spread as in Corwin and Schultz (2012). In model 4, these comprise the 3 month LIBOR, the TED spread, a credit spread based on differences in corporate bond rates, the term spread, and a measure building on abnormal stock returns of investment banks. Model 5 is model 4 plus the average monthly value of the VIX. Model 6 includes controls for market state, past market volatility, return dispersion, co-momentum, and investor sentiment, as described in detail in Section 3.2. and the appendix. T-statistics (in parentheses) are adjusted for heteroscedasticity and autocorrelation (Gallant (1987)). Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Model specification	(1)	(2)	(3)	(4)	(5)	(6)
Observations	564	564	492	276	156	522
Sample Start	Jan-62	Jan-62	Jan-68	Jan-86	Jan-93	Jul-65
Sample End	Dec-08	Dec-08	Dec-08	Dec-08	Dec-05	Dec-08
Lagged attention shift rank (1-6)	-0.279*** (-2.78)	-0.319*** (-3.05)	-0.282** (-2.32)	-0.642*** (-4.69)	-0.772*** (-4.82)	-0.379*** (-3.63)
Attributable to long leg	32%	40%	38%	42%	43%	44%
R-squared	0.013	0.074	0.174	0.162	0.145	0.083

Table 3: Descriptive Statistics for Stock and Pair Characteristics by Distraction Deciles

In panel A, *NYSE/AMEX macap decile* refers to the firm's market capitalization decile rank computed at the beginning of the pair's six-month trading period. *Amihud illiquidity ratio* is computed as the average of a stock's absolute daily return divided by its total daily trading volume in millions of dollars. The estimation period for the illiquidity ratio, for average daily turnover as well as for idiosyncratic risk is the twelve-month period ending at the beginning of a pair's trading period. *Idiosyncratic risk* is computed as the standard deviation of the residual obtained from time series regressions of a stock's daily return on factors for the market premium, size, value and momentum. *Bid – ask spread* is computed following the methodology in Corwin and Schultz (2012). *Maximum industry weight* denotes the largest fraction of sample firms belonging to a specific industry group (out of the 49 Fama/French industries). *Industry concentration* is computed as the sum of squared industry weights. In panel B, the first four rows report within-pair differences of stock characteristics, which are computed as in panel A. The last column reports differences in mean characteristics between decile 10 and decile 1. Standard errors are adjusted for heteroskedasticity and clustered by day of pair divergence. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Variable		All	Decile 1	Decile 10	10-1 (Mean)
Panel A: Firm Characteristics by Distraction Deciles					
NYSE/AMEX macap decile	Mean	8.86	8.85	8.86	0.0189
	Median	9	9	9	
Amihud illiquidity ratio	Mean	0.0578	0.0567	0.0566	-0.0002
	Median	0.01	0.0107	0.0094	
Average daily turnover	Mean	0.16%	0.16%	0.17%	0.008%
	Median	0.11%	0.11%	0.11%	
Idiosyncratic risk	Mean	1.12%	1.10%	1.13%	0.032%**
	Median	1.06%	1.05%	1.07%	
Turnover on day of divergence	Mean	0.25%	0.21%	0.30%	0.09%***
	Median	0.08%	0.11%	0.15%	
Bid-ask spread on day of divergence	Mean	0.44%	0.39%	0.52%	0.13%***
	Median	0.32%	0.31%	0.32%	
No. industry groups		49	49	49	
Maximum industry weight	Fraction	29.14%	29.20%	29.53%	
	Industry		Utilities		
Industry concentration		0.114	0.115	0.115	
S&P 500 dummy		0.54%	0.54	0.54	0.00
Panel B: Pair Characteristics by Distraction Deciles					
Macap decile difference	Mean	1.25	1.25	1.24	-0.0052
	Median	1	1	1	
Average daily turnover difference	Mean	0.072%	0.072%	0.073%	-0.00%
	Median	0.040%	0.040%	0.039%	
Amihud illiquidity ratio difference	Mean	0.063	0.061	0.061	-0.000
	Median	0.013	0.0130	0.0130	
Idiosyncratic risk difference	Mean	0.248%	0.237%	0.252%	0.015%***
	Median	0.20%	0.19%	0.20%	
Turnover difference on day of divergence	Mean	0.23%	0.05%	0.26%	0.21%***
	Median	0.08%	0.07%	0.09%	
Bid-ask spread difference on day of divergence	Mean	0.62%	0.54%	0.76%	0.22%***
	Median	0.43%	0.41%	0.45%	
Cumulative price difference upon divergence	Mean	6.68%	6.44%	7.04%	0.60%***
	Median	6.28%	6.05%	6.58%	
Return difference at day of divergence	Mean	2.84%	2.38%	3.59%	1.21%***
	Median	2.29%	1.93%	2.89%	
No. industry group combinations		931	623	697	
Maximum industry group weight	Fraction	15.89%	16.04%	15.13%	
	Industries		Utilities/Communication		
Industry group concentration		0.039	0.041	0.037	
S&P 500 dummy difference		0.44	0.44	0.44	
No. round-trip trades		103,386	8,187	14,011	5,824

Table 4: Pairs trading and Attention shifts: US evidence

This table explores the impact of attention shifts (as measured on the day of divergence) on event-time one-month returns on zero-cost portfolios of US stock pairs. Breakpoints for the deciles are determined separately for each year. Calculations are based on daily data from January 1962 to December 2008. Trading positions in each pair are initiated on the day following the convergence and liquidated on the day following convergence or after one month has passed, respectively. Panel A presents univariate findings. Panel B displays findings from multivariate regressions. The dependent variable of interest is either the attention shift proxy decile rank or a high/low distraction dummy, which is zero for low distraction days (decile 1) and one for high distraction days (decile 10). We consider up to three control sets. The first set controls for calendar and industry effects (indicator variables for year, month, day of week, and pair industry group combinations). The second set controls for market-level conditions on the day of divergence (market return, squared market return, market turnover, 10-day rolling volatility, average bid-ask spread, factors for daily return premia on size, value, momentum and short-term reversal). The third set controls for a number of pair characteristics computed as outlined in table 2 (average firm market capitalization decile rank, \ln (average pre-event turnover), \ln (average turnover on day of divergence), \ln (average pre-event Amihud illiquidity ratio), average idiosyncratic risk, \ln (bid-ask spread as in Corwin and Schultz (2012), S&P 500 dummy, within-pair differences of the just mentioned variables, and return difference attributable to the day of divergence. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. T-statistics (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence.

Panel A: Average One-Month Return on US stock pairs (1/1962-12/2008), Univariate Regressions												
Attention shift decile	All	1	2	3	4	5	6	7	8	9	10	Diff 10-1
Return on pairs	0.97%***	0.53%***	0.86%***	0.77%***	0.87%***	0.95%***	0.94%***	1.02%***	1.10%***	1.09%***	1.30%***	0.77%***
t-statistic	(30.41)	(5.14)	(9.00)	(7.89)	(8.81)	(9.50)	(9.04)	(9.62)	(12.00)	(11.25)	(13.13)	(5.28)
% of convergence	36.3%	33.4%	34.3%	34.3%	34.7%	36.2%	36.1%	35.8%	37.6%	38.1%	39.7%	6.3%***
(Return convergence)	6.42%***	6.14%***	6.29%***	6.33%***	6.24%***	6.33%***	6.37%***	6.53%***	6.39%***	6.51%***	6.71%***	0.56%***
(Return no convergence)	-2.08%***	-2.24%***	-1.94%***	-2.11%***	-1.95%***	-2.08%***	-2.09%***	-2.01%***	-2.02%***	-2.17%***	-2.13%***	0.11%
Observations	103,386	8,187	8,679	9,146	9,398	10,048	10,079	10,595	11,019	12,224	14,011	5,824
Panel B: Average One-Month Return on US stock pairs (1/1962-12/2008), Multivariate Regressions												
Dependent variable of interest	Attention Shift Proxy Decile Rank						High/Low Distraction Dummy					
Model specification	(1)	(2)				(3)	(4)	(5)				(6)
Observations	103,386	103,386				103,386	22,198	22,198				22,198
R ²	0.0355	0.0368				0.0375	0.089	0.091				0.094
Coefficient	0.065***	0.042***				0.039***	0.732***	0.534***				0.501***
t-statistic	(6.18)	(3.49)				(3.33)	(5.19)	(3.27)				(3.03)
Calendar and industry controls	yes	yes				yes	yes	yes				yes
Market-level conditions	no	yes				yes	no	yes				yes
Firm and pair characteristics	no	no				yes	no	no				yes

Table 5: Pairs Trading and Attention Shifts: International Evidence

This table reports monthly event-time returns from pairs trading in international stock markets. The computation of the country-specific attention shift proxies relies on the 10 GICS industry sectors. For the country-specific monthly top 100 pairs, we discard all stocks with at least one missing return or at least two zero/missing trading volume days within the twelve-month estimation period. Apart from that, the methodology is the same as in the baseline analysis for the US market (see panel A table 4). Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. T-statistics (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence.

	Japan	UK	France	Germany	Switzerland	Italy	Netherlands	Hongkong
Panel A: One Day Waiting, Full Sample Period (1962-2008)								
Time Period	1/1995-12/2009	1/1995-12/2009	1/1996-12/2009	1/1996-12/2009	6/1997-12/2009	6/1995-12/2009	1/1995-12/2009	1/1995-12/2009
Observations	36,992	27,185	25,833	26,006	16,841	25,596	21,320	20,894
Panel A: One-month Pairs Trading Abnormal Returns								
Pairs Trading Returns: All	1.03%*** (17.02)	0.73%*** (9.58)	1.11%*** (13.03)	1.09%*** (10.38)	0.54%*** (4.15)	0.48%*** (4.10)	0.82%*** (7.39)	1.00%*** (4.48)
Pairs Trading Returns: Decile 1	1.19%*** (6.96)	0.28% (1.39)	0.79%*** (2.93)	0.61%* (1.84)	-0.65% (-1.52)	0.32% (0.91)	0.37% (1.18)	-0.99% (-1.46)
Pairs Trading Returns: Decile 10	1.38%*** (7.01)	1.42%*** (5.80)	1.73%*** (6.34)	2.04%*** (6.30)	1.71%*** (3.35)	0.66% (1.44)	2.03%*** (5.94)	2.63%*** (4.38)
Diff 10-1	0.19% (0.00260)	1.14%*** (0.00316)	0.94%** (0.00383)	1.14%*** (0.00463)	2.37%*** (0.00664)	0.34% (0.00577)	1.66%*** (0.00463)	0.0362*** (0.00906)
Panel B: Fraction of Convergence								
% of Convergence: All	44.8%	28.9%	29.6%	29.6%	19.5%	26.7%	23.2%	23.1%
% of Convergence: Decile 1	43.8%	25.6%	26.9%	27.0%	14.8%	25.0%	20.9%	15.9%
% of Convergence: Decile 10	49.1%	34.1%	32.8%	32.2%	24.4%	30.0%	30.1%	29.8%
Diff 10-1	5.37%***	8.50%***	5.89%***	5.14%***	9.65%***	5.00%***	9.20%***	13.9%***

Table 6: Robustness Checks

This table presents results from various robustness checks. Panel A displays subperiod results from the baseline approach. Standard errors (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence. In Panel B, the process of identifying the monthly top 100 pairs is modified. *Excluding utility firms* means we do not consider any pair with at least one firm belonging to Fama/French (1997) industry group 31. *Only different firms* means we do not select a pair if at least one of the firms is already a component of any higher-ranked pair. Standard errors (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence. Panel C shows results for *same pairs*, i.e. a subsample of pairs that diverge both at least once on a high distraction day and on a low distraction day. Standard errors (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence. Panel D shows findings for the subsample of pairs that diverge on days with positive market returns or have below median idiosyncratic risk. Cut-offs for idiosyncratic risk are determined separately for each year, and are based on all pairs that start trading in the year under consideration. Panel E shows alphas obtained from calendar time-series regressions of the difference in pairs trading returns on well-established risk factors. The 3 *factor model* (1/1962-12/2008) includes the Fama and French (1993) factors, the 7 *factor model* (1/1968-12/2008) additionally controls for short-term reversal, long-term reversal, momentum, and liquidity. Panel F reports results from Fama/MacBeth-type regressions. We first estimate yearly pooled cross-sectional regressions of one-month pairs returns on distraction decile ranks and then use the time-series of resulting coefficients to assess the statistical significance of the attention shift proxy. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Subperiod Analysis				
	Attention Shift Decile: All	Attention Shift Decile: 1	Attention Shift Decile: 10	Difference: 10-1
Subperiod: 1962-1977	1.49%*** (29.16)	1.04%*** (6.01)	1.77%*** (11.57)	0.73%*** (3.16)
Subperiod: 1978-1993	1.09%*** (20.64)	0.71%*** (4.06)	1.29%*** (7.54)	0.59%** (2.41)
Subperiod: 1994-2008	0.24%*** (4.01)	-0.25% (-1.44)	0.79%*** (4.22)	1.04%*** (4.08)
Panel B: Variations in the Data Set (Monthly Top 100 Pairs)				
	Attention Shift Decile: All	Attention Shift Decile: 1	Attention Shift Decile: 10	Difference: 10-1
Excluding utility firms	1.20%*** (38.71)	0.59%*** (5.78)	1.35%*** (13.92)	0.76%*** (7.45)
Only different firms	1.08%*** (37.24)	0.79%*** (8.68)	1.44%*** (16.55)	0.65%*** (5.00)
Panel C: Limitation to Firms that Diverge both at Least Once on High and Low Distraction Days				
	Attention Shift Decile: All	Attention Shift Decile: 1	Attention Shift Decile: 10	Difference: 10-1
Same pairs	2.02%*** (33.06)	1.40%*** (7.87)	2.13%*** (13.31)	0.73%*** (3.05)
Panel D: Limitation to Subsamples with Low Limits to Arbitrage				
	Attention Shift Decile: All	Attention Shift Decile: 1	Attention Shift Decile: 10	Difference: 10-1
Positive market return	0.97%*** (22.03)	0.53%*** (3.73)	1.38%*** (11.75)	0.85%*** (4.56)
Below median idiosyncratic risk	0.82%*** (19.84)	0.33%*** (3.37)	1.17%*** (9.71)	0.74%*** (3.37)
Panel E: Alphas from Calendar Time Regressions on Risk Factors				
3 factor model		0.30%** (2.19)		
6 factor model		0.29%** (2.04)		
Panel F: Alternative Regression Approach				
Coefficient on attention shift decile rank		0.57%*** (37.77)		

Table 7: Cross-sectional Tests: Attention Shifts and Pairs Trading Profitability

Panel A shows the sensitivity of pairs trading returns to attention shift proxy decile ranks (as observed on the day of divergence) for several samples of top pairs: The monthly top 100 pairs each consisting of firms from different industries, the monthly top 100 pairs each consisting of firms from the same industries, and the monthly top 20 pairs each consisting of two value-weighted industries. In all cases, we use the Fama/French (1997) classification with 49 industries. The first column shows the return difference between pairs diverging on high distraction days (decile 10) and pairs diverging on low distraction days (decile 1), as quantified by including the high/low distraction dummy. The approach resembles the methodology used in panel A of table 4. The second column shows findings from regressing pairs returns on the attention shift proxy decile rank. Panel B reports results from a test similar in fashion. It compares the return sensitivity for pairs of which constituent firms share (do not share) at least one business segment, as described in detail in Section 4.4.. Panel C compares returns from pairs consisting only of firms with high residual media coverage and pairs consisting only of firms with low residual media coverage, as described in detail in Section 4.4.. The first column compares average event-time one-month pairs trading returns. The second column shows findings from regressing pairs returns on the attention shift proxy decile rank. In all panels, statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. Standard errors (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence.

Panel A: Impact of Investor Distraction on Pairs Consisting of Alternative Assets		
	High/Low Distraction Dummy	Distraction Decile Rank
Top 100 pairs with stocks	0.77***	0.065***
from different industries (N=103,386)	(5.28)	(9.15)
Top 100 pairs with stocks	0.71***	0.057***
from the same industry (N=100,726)	(5.92)	(6.13)
Top 20 industry-level pairs	0.39**	0.033**
(N=14,180)	(2.15)	(2.36)
Panel B: Pairs With and Without Common Industry Segments (since 1977)		
	High/Low Distraction Dummy	Distraction Decile Rank
No shared industry segment	0.97***	0.073***
	(4.22)	(4.06)
Shared industry segment	0.84**	0.019
(about 18% of observations)	(2.15)	(0.63)
Difference	0.13	0.054*
	(0.34)	(1.86)
Panel C: Pairs with High and Low Residual Media Coverage		
	Return: All Deciles	Attention Shift Decile Rank
Low residual media coverage	0.77***	0.010*
	(4.81)	(0.17)
High residual media coverage	-0.01	-0.12
	(-0.03)	(-1.20)
Difference	0.82%**	0.22%*
	(2.41)	(1.83)

Appendix for “Losing sight of the trees for the forest?”

Attention allocation and anomalies”

April 2013

This appendix contains test, figures, and tables that supplement the analysis in the paper.

A. Data sets and variable descriptions Table 1 gives an overview of the data samples used in the study. Table 2 describes the construction of selected variables relied on in the paper.

B. Attention Shift Proxy In appendix B.1, we explore the relationship between the attention shift proxy proposed in the paper and alternative measures of attention allocation. The latter include firm-level residual turnover, the Hirshleifer et al. (2009) distraction measure, New York Times front page articles about the stock market, new record levels of the Dow, as well as shocks in internet search queries.

Subsequently, the appendix displays several figures intended to give additional information on the baseline attention shift proxy used in the paper. The proxy relies on the Fama and French (1997) industries as a classifications scheme for market segments. Figure 1 shows the time-series of weighted abnormal industry-level returns, i.e. the result after step three of the construction process for the attention shift proxy. Figures 2 to 4 illus-

trate the relationship between specific characteristics of abnormal industry-level returns and resulting attention shift decile ranks.

Table 3 gives an overview over the twelve modified attention shift proxies. Table 4 provides a transition matrix between baseline attention shift decile ranks and the decile ranks of a modified proxy available in real time.

C. Momentum strategies

This section provides more information on the fourteen individual momentum approaches suggested by previous work, which are relied on in the paper.

D. Pairs Trading

Appendix D.1 verifies that the future earnings surprises of the monthly top 100 pairs are significantly correlated, suggesting that these firms are not only statistically, but also fundamentally linked.

Appendix D.2, which includes table 5, provides further time-series insights into the role of limited attention. It explores the role of previously proposed attention proxies on the profitability of pairs trading. These proxies include Fridays, down market periods, the number of competing events, shocks in internet search queries consistently, and holidays.

Appendix D.3 explores the impact of time-varying arbitrage risk on pairs trading profitability. To this end, it starts with a short description of the arbitrage risk proxies employed. Table 6 then displays findings from univariate regressions of the one-month return on zero-cost US stock pairs on these proxies. Table 7 shows pairwise correlations between these arbitrage proxies and attention shift proxy decile ranks. Tables 8 and 9 show findings from multivariate regressions of pairs trading profits on arbitrage risk proxies, the high/low distraction dummy, and a set of control variables.

Figure 5 shows the empirical cumulative distribution function of the return on US stock pairs sorted by attention shift proxy decile ranks as observed on the day of divergence. Figure 6 shows results from a replication of the baseline analysis (see figure 1 and table 4 in the paper) when relying on in total twelve modified distraction proxies.

Table 10 reports distribution details of the return on US stock pairs sorted by attention shift proxy decile ranks as observed on the day of divergence. Table 11 displays corresponding performance measures. Table 12 shows the impact of modified attention shift proxies on pairs trading profitability. Table 13 shows differences between the return on pairs diverging on high and low distraction days when conditioned on the level of the return differential on the day of divergence. Table 14 shows the full list of regression coefficients for pair-level and market-level variables from table 4 in the paper. Table 15 provides descriptive statistics for pairs trading samples in international stock markets.

References are provided on the last three pages.

A Data Sets and Variable Descriptions

Table 1: Overview of Data Sets Used in the Study

Data Source			Description	Sample Period
CRSP			Daily and monthly stock market data for all US firms trading on NYSE, AMEX or NASDAQ	January 1960 - December 2008
Compustat Stock File	Global	Daily	Daily stock market data for firms of eight large non-North American stock markets (Japan, UK, France, Germany, Switzerland, Italy, Netherlands, Hongkong)	Middle of the 1990ies - December 2009
Compustat Segment Files			Financial information about firm segments based on yearly data	1977-2008
Compustat Quarterly and Annual	Fundamentals		Earnings announcement dates, quarterly earnings, and fundamental data	Varying - 2008
Compustat constituents	Index	Con-	Index membership	Varying - 2008
Compustat Short Interest File	Supplemental		Short Interest	1973 - 2008
Factiva			Yearly number of Dow Jones News Services articles about each firm that meets data requirements on pairs trading in some period after 1990	1991-2008
IBES (Detail and summary tape)			Analyst earnings forecasts/revisions, earnings announcement dates (detail tape), number of analysts following (summary tape)	Varying - December 2008
Federal Reserve Board H15 release			Several interest rates (such as Moody’s corporate bond rates)	1982 - December 2008
Kenneth R. French’s Library			Several portfolio returns and risk factors	Middle of the 1960ies - December 2008
Thomson Datastream and Bloomberg	Reuters	Datas-	Various control factors (such as the VIX)	Varying - December 2008
Lexisnexis			Front-page articles of the New York Times	January 1980 - December 2008
Dow Jones			Dow Jones Industrial Average on a daily basis	January 1960 - December 2007
NBER			US business cycle expansions and contractions based on monthly data	January 1960 - December 2008
Google			Standardized weekly time-series of search frequencies for ticker symbols and other terms, constructed from Google’s application “Insights for Search”	January 2004 - December 2008

Table 2: Description of selected variables

Distress risk	Our measure of distress risk is based on Campbell et al. (2008), who use a dynamic logit model based on a broad set of accounting and market variables to empirically quantify a firm’s failure probability. We follow the implementation used in Stambaugh et al. (2012) and described in detail in the appendix of Chen et al. (2011).
(Non-)lottery stocks	We follow Kumar (2009) in defining a (non)-lottery type stock as one with above (below) median idiosyncratic volatility, above (below) median idiosyncratic skewness, and below (above) median nominal stock price.
Reporting lag (PEAD)	The reporting lag is computed as $\ln(1 + \text{number of day from the quarter-end to the actual earnings announcement date})$.
Daily bid-ask spread	We follow the computation proposed in Corwin and Schultz (2012) (see equations 14 and 18 in their paper).
Market to book ratio	We rely on quarterly computed book value as described in detail in Chen et al. (2011).
Idiosyncratic volatility	We measure idiosyncratic volatility as the standard deviation of the residual obtained from regressing a stock’s excess return on a Carhart (1997) factor model. We rely on daily rolling regressions over the previous month and require at least 20 valid observations.
Market state (momentum analysis)	The market state is a dummy variable that takes on a value of 1 if the cumulative market return over the previous 36 months was negative.
Market volatility (momentum analysis)	Market volatility is estimated from the previous 36 monthly returns.
Co-momentum measure	The co-momentum measure is inspired by recent work of Lou and Polk (2012). In each month, we sort stocks into deciles based on their cumulative return over month $t-7$ to month $t-2$. We require stocks to have a market capitalization greater than the first NYSE decile as well as stock price of at least 5 USD at the time of portfolio formation. We then restrict the sample to those stocks that a typical momentum trading strategy would bet. We do so by excluding firms contained in deciles 2 to 9. For each of the remaining stocks (i.e. those contained in deciles 1 and 10), we compute daily residual returns over months $t-7$ to $t-2$ relative to a Fama and French (1993) three-factor model. For decile 1 and 10 separately, we measure the average correlation of the residual return of a given stock with all other stocks in the momentum decile in question. This leads to two separate co-momentum measures (in month t) for winner stocks as well as for loser stocks.

B Attention Shift Proxy

B.1 Attention shift proxy and alternative measures of attention allocation

The goal of the following tests is to further explore whether the novel attention shift proxy proposed in the paper is able to capture short-term investor attention shifts. To this end, we investigate the interaction of the proxy with the following variables: Average daily residual firm-specific turnover, Hirshleifer et al. (2009) distraction measure, New York Times front-page articles about the stock market, record levels for the Dow, and shocks in internet search queries. We describe these tests one by one.

Residual firm specific turnover For January 1960 to December 2008, we compute daily turnover shocks at the firm-level, thereby taking into account all common shares traded on NYSE, Amex or Nasdaq. Turnover shocks ($TS_{i,t}$) for firm i at day t are defined as the difference between actual turnover ($T_{i,t}$) and expected turnover. Expected turnover is based on the intercept and coefficients obtained from regressing a stock’s daily turnover on the value-weighted market turnover and the relevant value-weighted Fama and French (1997) industry turnover. Note that this approach is an extension of the turnover market model brought forward in e.g. Tkac (1999) or Lo and Wang (2000). The sample period for this regression are the twelve months ending one month prior to day t , i.e. we skip one month between estimation and evaluation period. In sum, turnover shocks $TS_{i,t}$ are computed for each firm and day separately as follows:

$$TS_{i,t} = T_{i,t} - \hat{\alpha}_{i,t} - \hat{\beta}_{1,t} Turnover_{market,t} - \hat{\beta}_{2,t} Turnover_{industry,t} \quad (1)$$

Based on this estimate, we equally weight all measures for a given day to obtain average firm-level residual turnover. We expect a negative relation between the attention shift proxy proposed in the paper and average firm-level residual turnover: If investors focus primarily on aggregate information (high distraction days, decile rank 10) and thus exhibit category learning behavior, average firm-level residual turnover, which controls for market-level and sector-level turnover, should be low. In contrast, if investors devote much attention to individual firms (low distraction days, decile rank 1), average firm-level

residual turnover should be high.¹

We explore this relationship in different ways. To keep the analysis as intuitive and simple as possible, we start by performing independent yearly decile sorts of the firm-level residual turnover measure. This approach is identical to the construction of the attention shift proxy. It turns out that the average decile rank of the firm-level turnover measure is 6.23 on low distraction days, but only 4.43 on high distraction days. In other words, truly firm-specific turnover is higher on low distraction days, as expected. As the following table shows, these findings become stronger if we restrict the focus to high firm-level turnover days (decile 10) and low firm-level turnover days (decile 1).

Attention shift proxy	Turnover measure		
	Low firm-level turnover	High firm-level turnover	Total
Low distraction day	29.02%	70.98%	100%
High distraction day	75.25%	24.75%	100%
Total	60.40%	39.60%	100%

While low distraction days tend to be at the same time high firm-level turnover days as opposed to low-firm turnover days (71% to 29% likelihood), the opposite is true for high distraction days (25% to 75% likelihood).

Finally, we verify these insights in a more formal regression approach. We regress the attention shift decile ranks or the high/low distraction dummy on raw values of average firm-specific residual turnover or on its decile ranks. In some specifications, we also include year dummies and the market return. Inferences remain unchanged if we add further market-level controls, such as volatility or squared market returns. In all regressions, the sample period is January 1962 to December 2008.

¹Note that the focus on the residual measure is crucial here. The market turnover itself is positively related to attention shift decile ranks, as described in detail in the paper.

Model specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Attention shift decile ranks				High/low distraction proxy			
Firm-specific residual turnover (raw)	-493.24*** (-9.86)		-620.05*** (-9.87)		-135.65*** (-7.03)		-170.90*** (-6.71)	
Firm-specific residual turnover (deciles)		-0.18*** (-19.98)		-0.19*** (-19.18)		-0.05*** (-15.71)		-0.05*** (-15.18)
Year fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Market return	No	No	Yes	Yes	No	No	Yes	Yes
Adjusted R ²	1.07%	3.32%	2.92%	2.92%	0.14%	0.18%	2.16%	2.16%

Across all specifications, there is a significant negative relationship between attention towards firm specific information, as quantified with the novel proxy proposed in the paper, and average firm-level residual turnover. To sum up, these findings are consistent with the idea that the attention shift proxy picks up actual investor attention allocation.

Hirshleifer et al. (2009) distraction measure

We closely follow the approach proposed in Hirshleifer et al. (2009). From July 1971 on, we gather the daily number of earnings announcements as accessible via Compustat. We then perform quarterly independent decile sorts to obtain a measure of inattention towards firm-level information. A high value (decile rank 10) suggests that only few resources are available to process the information contained in each individual earnings announcement.

We run similar tests as outlined above for the case of residual firm-specific turnover. The average decile rank of the Hirshleifer et al. (2009) distraction measure is 4.76 on low distraction days, but 5.93 on high distraction days. In other words, both distraction proxies tend to identify similar situations. These findings become stronger if we restrict our focus to Hirshleifer (2009) et al. high distraction days (decile 10) and low distraction days (decile 1). This can be seen from the following table.

Attention shift proxy	Hirshleifer (2009) et al. distraction measure		
	Low distraction day	High distraction day	Total
Low distraction day	76.53%	23.47%	100%
High distraction day	38.62%	61.38%	100%
Total	56.21%	43.79%	100%

Low distraction days computed as in Hirshleifer et al. (2009) tend to be at the same time low distraction days, as quantified with the novel attention shift proxy proposed in the paper. The combination (low distraction day as in the paper and low distraction day

as in Hirshleifer et al. (2009)) is more than three times as likely as the combination (low distraction day as in the paper and high distraction day as in Hirshleifer et al. (2009)). Similar insights are gained for high distraction days.

Finally, we verify these insights in a more formal regression approach. We regress the attention shift decile ranks or the high/low distraction dummy on the Hirshleifer et al. measure. In some specifications, we also include year dummies and the market return. Inferences remain unchanged if we add further market-level controls. In all regressions, the sample period is July 1971 to December 2008.

Model specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Attention shift decile ranks				High/low distraction proxy			
Hirshleifer et al. (2009) decile ranks	0.11*** (10.37)		0.11*** (10.38)		0.03*** (8.81)		0.03*** (8.89)	
Hirshleifer et al. (2009) high/low distraction day		1.26*** (9.25)		1.26*** (9.13)		0.38*** (8.93)		0.38*** (8.42)
Year fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Market return	No	No	Yes	Yes	No	No	Yes	Yes
Adjusted R ²	1.16%	4.38%	0.86%	4.54%	3.78%	14.52%	2.02%	14.03%

Across all specifications, there is a significant positive relationship between attention towards firm specific information, as quantified with the novel proxy proposed in the paper, and inattention towards valuable firm-level information, as quantified with the Hirshleifer et al. (2009) distraction measure. Inferences remain unchanged if we restrict our focus to the time period after 1995, when earnings announcement dates are less likely to suffer from slight imperfections. To sum up, these findings are consistent with the idea that the attention shifts proxy picks up actual investor attention allocation.

New York Times Front-Page Articles about the Stock Market

Our goal is to identify articles which satisfy two criteria, as suggested by Yuan (2012): First, they need to be published on a trading day at the front-page of the New York Times. Second, they need to deal with the stock market.

We rely on the LexisNexis database, which provides access to the New York Time archive from 1980 on. Consequently, our sample period is January 1980 to December 2008. We restrict our analysis to articles articles which have been published in section A/section 1, page 1. The goal of identifying articles that deal with the stock market

leaves some degrees of freedom. We operationalize this idea by only considering articles whose headline contains at least one of the words “stocks”, “stock market”, “Dow”. The use of these three keywords is also motivated by the application “Google correlate”, which identifies words with similar time-series behavior in search query frequencies. Our findings are robust to the use of only a subset of these terms. We have also experimented, with similar findings, with a broader set of keywords, as in e.g. the “Google trends” analysis explained below. If anything, the focus on just three keywords is likely to add additional noise to the analysis, making it more difficult to identify statistically significant relationships.

We exclude articles that are published on a non-trading day. Articles written on day t are then matched with trading day $t-1$. However, findings are also unchanged if we attribute articles published on day t to trading day t . In total, we are able to identify 331 trading days with stock market articles on the front-page of the New York Times.

The unconditional likelihood of observing such an article on a given day from January 1980 to December 2008 is 4.52%. In contrast, following low (high) distraction days, the likelihood is 1.99% (12.16%). These insights are also confirmed in the following regressions. Findings hold if we control for other market-level variables, such as e.g. volatility or squared market turnover.

Model specification	(1)	(2)	(3)	(4)
	Attention shift decile ranks		High/low distraction proxy	
Dummy for NYT front-page article	1.49*** (9.15)	1.49*** (9.21)	0.39*** (10.50)	0.41*** (10.44)
Year fixed effects	No	Yes	No	Yes
Market return	No	Yes	No	Yes
Adjusted R ²	1.16%	0.79%	3.98%	2.17%

To sum up, findings again are consistent with the idea that the attention shifts proxy picks up actual investor attention allocation.

Record levels for the Dow

The Dow Jones Industrial Average is available from CRSP until the end of 2007. Our sample period is therefore January 1962 to December 2007. As in Yuan (2012), we identify days on which the Dow reaches a new record level. In total, this happens on 690 days.

On average, a new record level is observed on 5.59% of all trading days. However, the corresponding values for low and high distraction days are 4.03% and 6.15% respectively. To the extent that a new record level is a distracting event, these findings thus support the line of reasoning behind the novel attention shift proxy. These insights are also confirmed in the following regressions. Findings hold if we control for other market-level variables, such as e.g. volatility or squared market turnover.

Model specification	(1)	(2)	(3)	(4)
	Attention shift decile ranks		High/low distraction proxy	
Dummy for new record level of the Dow	0.40*** (3.66)	0.39*** (3.17)	0.11** (2.40)	0.11** (2.19)
Year fixed effects	No	Yes	No	Yes
Market return	No	Yes	No	Yes
Adjusted R ²	0.24%	0.02%	0.23%	-0.03%

To sum up, findings again are consistent with the idea that the attention shifts proxy picks up actual investor attention allocation.

Shocks in internet search queries

We are interested in whether high (low) distraction days tend to go along with a comparatively higher (lower) demand for market-level or segment-level information in relation to firm-specific information. In the following analysis, our measure of information demand is based on indices of weekly internet search activity obtained from the online tool Google Insights. In doing so, we build on findings of an emerging literature, which highlights the ability of such an approach to proxy for (in particular retail) investor attention allocation (e.g. Da et al. (2011) or Mondria and Wu (2012)).

Google Insights allows to download time-series quantifying the relative importance of specific search queries entered in google.com from within the United States. This search

volume index is scaled in the sense that it represents the number of searches for specific terms relative to all other search terms in Google in the same period. It is moreover normalized in the sense that the period with the highest relative search volume is assigned a value of 100. Search volume for the remaining periods can take on values of 0 to 100, depending on their relative importance within the whole time-series under consideration. In the following analysis, we use these raw search volume index values. We have verified, however, that our findings are very similar if we base our tests simply on the rank order of the index values. The latter procedure tends to give less weight to extreme observations. The sample period for our analysis is January 2004 to December 2008. The start date is determined by the availability of Google Insights data. The end date corresponds to the end of the sample period of our study.

To construct a measure for firm-specific information demand over time, we widely follow the approach developed in Da et al. (2011). First, we randomly select 100 S&P 500 stocks. Second, for each of these firms, we download the weekly search volume index of their ticker symbols. Da et al. (2011) argue that people entering ticker symbols into Google pay direct attention to the firm under consideration and are most likely interested in gaining more information about its stock. Third, to construct a measure of average demand for firm-specific information in a given week, we equally weight all non-zero firm-level observations.

Constructing a sensible measure for sector-level or market-level information demand over time is arguably more ambiguous. As there appears to be little guidance by previous literature, we address this issue in two ways. As a parsimonious first approach, we simply use the search volume index of a term which arguably comes directly to mind: “stock market”. As a second approach, our main method, we equally weight the search volume data for seven terms: “stock market”, “Dow”, “market”, “S&P”, “stocks”, “market index”, “market segments”. We report results for both approaches separately.

To construct a measure of the relative importance of the demand for information primarily relevant at some aggregate level, we simply divide our weekly proxy for market-level information by the corresponding value for firm-specific information. Finally, values are standardized so that the mean is zero and the standard deviation 1. We refer to the resulting time-series as “internet-based attention allocation proxy”.

We then regress the time-series of attention proxy decile ranks on this alternative measure of attention shifts. The following table shows the main findings.

Model specification	(1)	(2)	(3)	(4)	(5)	(6)
Internet-based attention	0.71***	0.86***	0.61***	0.72***	0.83***	0.57***
allocation proxy	(4.83)	(5.71)	(4.30)	(4.58)	(4.32)	(3.84)
Sample start	1/2004	1/2004	1/2004	1/2004	1/2004	1/2004
Sample end	12/2008	12/2008	12/2008	12/2008	12/2008	12/2008
N	251	251	251	251	251	251
Year-fixed effects	no	yes	yes	no	yes	yes
Adjusted R ²	0.11	0.12	0.21	0.11	0.12	0.19

In Model 1 to 3, we rely on the search volume index for “stock market”, in model 4 to 6, we make use of our richer alternative method to quantify market-level information demand. Depending on the regression specification, we add no further controls (model 1 and 4), year dummies (model 2 and 5), or additionally controls for the absolute weekly market return and the average weekly VIX value (model 3 and 6). We have experimented with further controls, which however did not affect the qualitative nature of our findings. The displayed coefficients suggest that a one standard deviation change of the internet-based attention allocation proxy tends to go along with a 0.57 to 0.86 point change in attention shift proxy decile ranks. These estimates are persistently and highly statistically significant. To sum up, the findings are consistent with the idea that the proxy picks up actual investor attention shifts.

Figure 1: Time-Series of Weighted Abnormal Industry-Level Returns

The graph shows the time-series of weighted abnormal daily industry-level returns, i.e. the result after step three of the construction process for the attention shift proxy. For its construction, a three-step procedure is employed. First, for January 1960 to December 2008, we compute daily value-weighted returns for the 49 Fama/French (1997) industries. Second, we construct daily return shocks defined as the absolute difference between the actual industry return and its expected return as implied by an OLS market model. Parameter estimates are obtained from rolling time-series regressions based on daily return data over the previous year. Third, shocks are condensed into a single ratio. To this end, industry shocks are weighted by the inverse of the volatility of their shock variable over the previous year.

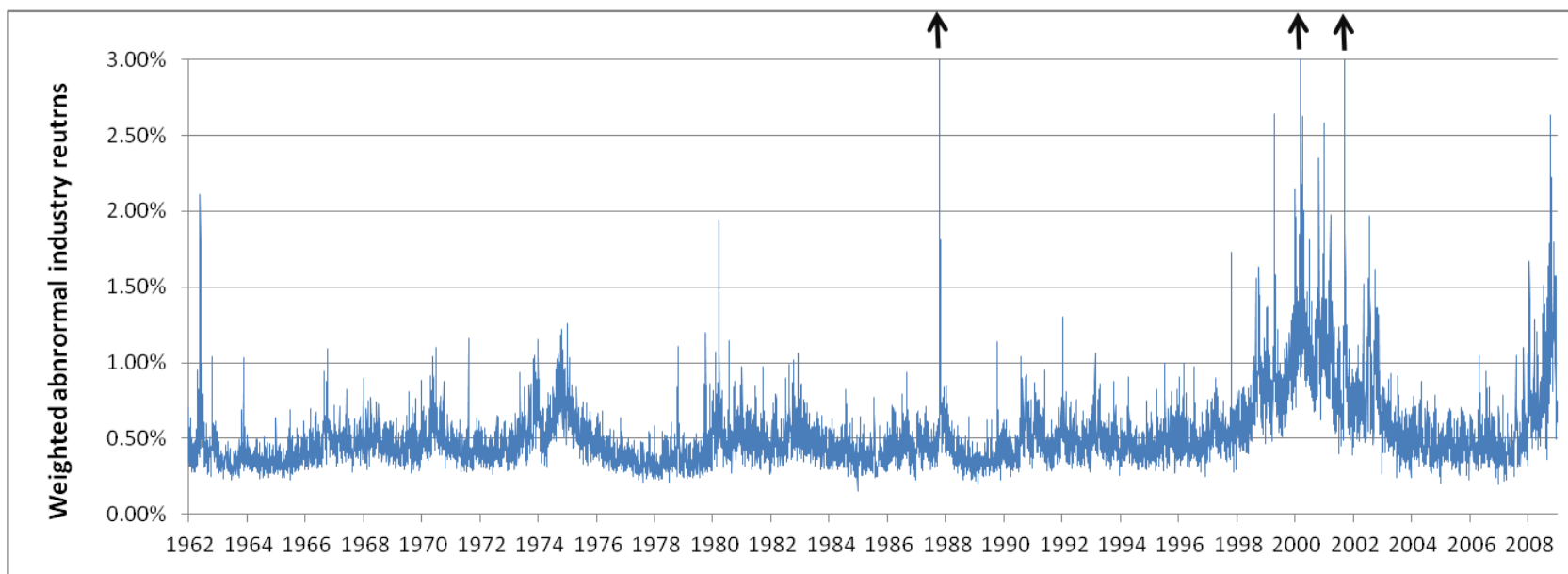


Figure 2: Mean of Abnormal Industry Returns by Attention Shift Proxy Decile Ranks

The following figures are intended to illustrate the relationship between specific characteristics of abnormal industry-level returns and resulting attention shift proxy decile ranks. To this end, we compute, at each day, the mean of different types of abnormal industry returns, where industries are represented by the 49 Fama/French (1997) segments. The upper graph uses (raw) abnormal returns, the middle graph uses absolute abnormal returns, and the lower graph uses weighted absolute abnormal returns, where industry weights are determined by the inverse of the volatility of their shock variables over the previous year. See section 2 of the paper for a detailed description of how abnormal returns and weights are computed. By attention shift proxy decile ranks, the boxes illustrate the 25th, 50th, and 75th percentile of the time-series of the mean of these abnormal returns. The adjacent values in the box plot are the most extreme values within 1.5*interquartile range (75th percentile-25th percentile) of the nearer quartile.

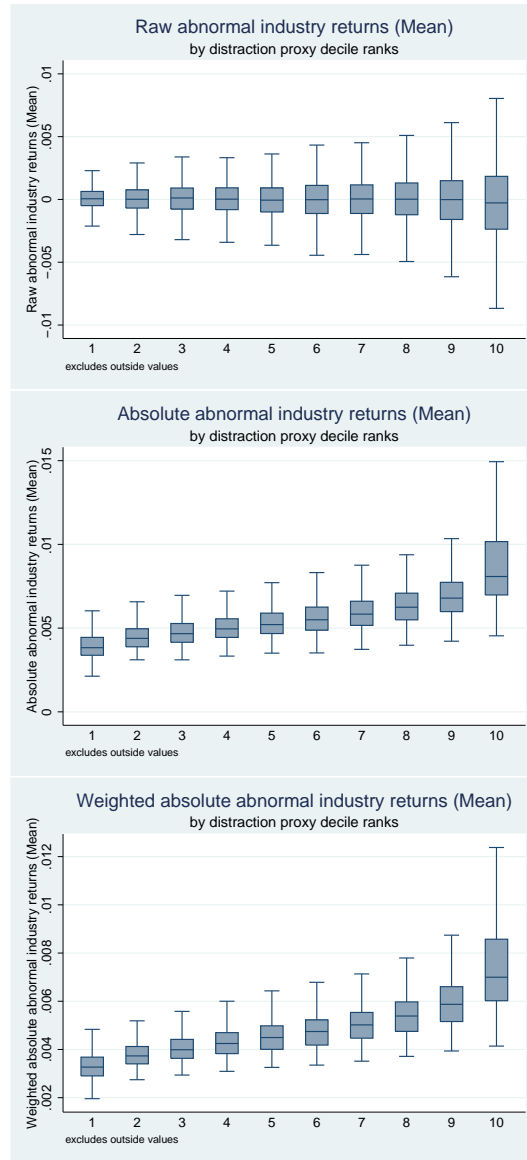


Figure 3: Standard Deviation of Abnormal Industry Returns by Attention Shift Proxy Decile Rank

The following figures are intended to illustrate the relationship between specific characteristics of abnormal industry-level returns and resulting attention shift proxy decile ranks. To this end, we compute, at each day, the standard deviation of different types of abnormal industry returns, where industries are represented by the 49 Fama/French (1997) segments. The upper graph uses (raw) abnormal returns, the middle graph uses absolute abnormal returns, and the lower graph uses weighted absolute abnormal returns, where industry weights are determined by the inverse of the volatility of their shock variables over the previous year. See section 2 of the paper for a detailed description of how abnormal returns and weights are computed. By attention shift proxy decile ranks, the boxes illustrate the 25th, 50th, and 75th percentile of the time-series of the standard deviation of these abnormal returns. The adjacent values in the box plot are the most extreme values within 1.5*interquartile range (75th percentile-25th percentile) of the nearer quartile.

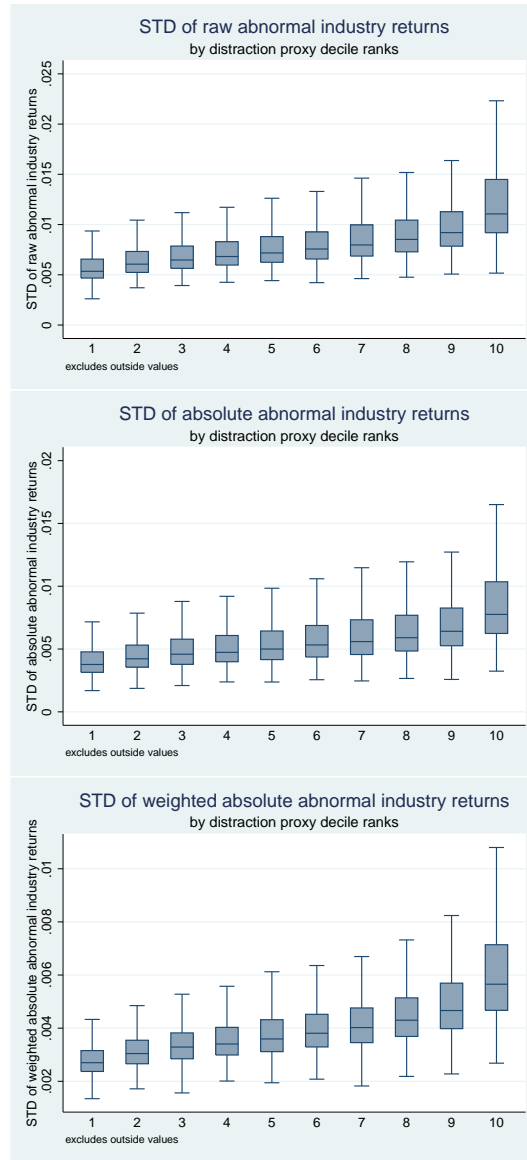


Figure 4: Maximum Fraction of a Single Industry Return Shock by Attention Shift Proxy Decile Rank

The following figure is intended to illustrate the relationship between specific characteristics of abnormal industry-level returns and resulting attention shift proxy decile ranks. To this end, we first compute the baseline attention shift proxy as outlined in section 2 of the paper. In short, the proxy is constructed as the sum of weighted absolute abnormal industry returns, where industries are represented by the 49 Fama/French (1997) segments. For each day, we then identify the maximum fraction a single industry return shock accounts for. By attention shift proxy decile ranks, the boxes illustrate the 25th, 50th, and 75th percentile of the time-series of this maximum weight. The adjacent values in the box plot are the most extreme values within 1.5*interquartile range (75th percentile-25th percentile) of the nearer quartile.

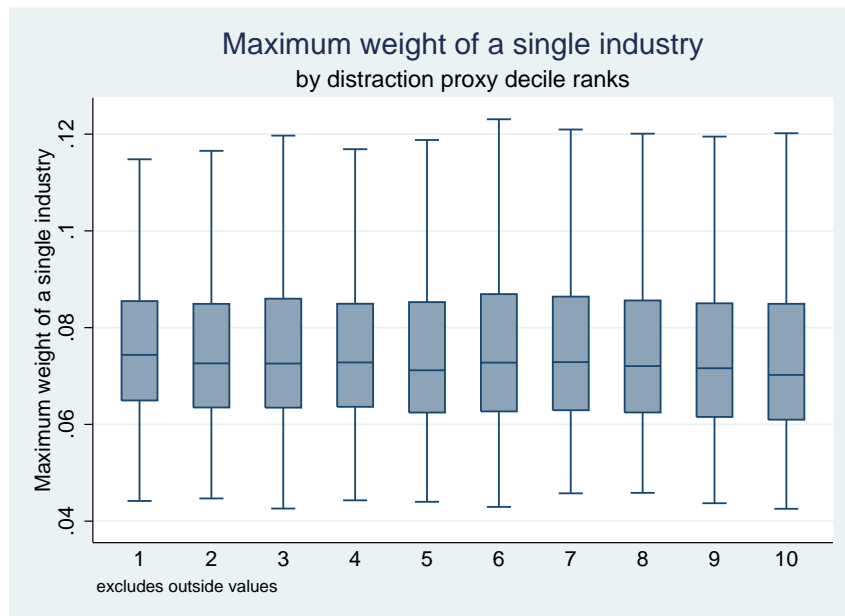


Table 3: Modified Attention Shift Proxies: Descriptions

This table reports descriptive statistics of the baseline and 10 modified attention shift proxies. Factor loadings with respect to the model of expected returns are estimated from time-series regressions based on daily data over the previous year. In the value-weighted case, abnormal industry returns are weighted by the relative market capitalization of industries. The interquartile range is computed as the 75th percentile of the cross-section of daily abnormal industry returns minus the 25th percentile. In model 9, the time-series of the baseline distraction proxy is orthogonalized with respect to the time-series of the average idiosyncratic volatility of the constituents of the pairs that diverge on a given day. Using the same method as in the baseline approach, attention shift proxy decile ranks are then computed from the residual. Returns on the 100 portfolios formed on book-to-market and size as well as on the 25 portfolios formed on size and short-term reversal are taken from Kenneth French's data library. Mode 12 is based on rolling historical raw proxy data. Specifically, for a given day, we use the preceding 250 trading days to compute the distraction decile the day belongs to. This implies that the beginning of the sample period is one year later (January 1963) than in the baseline case.

Proxy Label	Proxy availability	Computation Scheme	Expected returns	Market Segments
Baseline Attention Shift Proxy	1/1962-12/2008	Volatility-weighting of absolute abnormal returns	OLS Market Model	49 Fama French Industries
Alternative Attention Shift Proxy 1	1/1961-12/2008	Equal-weighting of absolute abnormal returns	OLS Market Model	49 Fama French Industries
Alternative Attention Shift Proxy 2	1/1961-12/2008	Value-weighting of absolute abnormal returns	OLS Market Model	49 Fama French Industries
Alternative Attention Shift Proxy 3	7/1964-12/2008	Equal-weighting of absolute abnormal returns	Four-Factor Model	49 Fama French Industries
Alternative Attention Shift Proxy 4	7/1964-12/2008	Value-weighting of absolute abnormal returns	Four-Factor Model	49 Fama French Industries
Alternative Attention Shift Proxy 5	7/1965-12/2008	Volatility-weighting of absolute abnormal returns	Four-Factor Model	49 Fama French Industries
Alternative Attention Shift Proxy 6	1/1961-12/2008	Interquartile range of raw abnormal returns (Q3 - Q1)	OLS Market Model	49 Fama French Industries
Alternative Attention Shift Proxy 7	7/1964-12/2008	Interquartile range of raw abnormal returns (Q3-Q1)	Four-Factor Model	49 Fama French Industries
Alternative Attention Shift Proxy 8	1/1960-12/2008	Regression approach with raw returns	None	49 Fama French Industries
Alternative Attention Shift Proxy 9	1/1964-12/2008	Baseline approach orthogonal to firm-level idiosyncratic risk	OLS Market Model	49 Fama French Industries
Alternative Attention Shift Proxy 10	7/1965-12/2008	Volatility-weighting of absolute abnormal returns	OLS Market Model	100 book-to-market x size
Alternative Attention Shift Proxy 11	7/1964-12/2008	Volatility-weighting of absolute abnormal returns	OLS Market Model	25 size x short-term reversal
Alternative Attention Shift Proxy 12	1/1963-12/2008	Volatility-weighting of absolute abnormal returns (real-time availability)	OLS Market Model	49 Fama French Industries

Table 4: Transition Matrix: Baseline Attention Proxy Decile Ranks and Modified Ranks Available in Real Time (1/1963-12/2008, N=102,259)

This table shows the transition matrix for baseline attention shift proxy decile ranks and modified decile ranks, which are based on rolling historical raw proxy data. Specifically, for a given day, we use the preceding 250 trading days to compute the distraction decile the day belongs to. This implies that the beginning of the sample period is one year later (January 1963) than in the baseline case. Computations are based on all 102,259 sample pair trades, which are opened between January 1963 and December 2008. This implies that attention shift proxy ranks at days where more (less) pairs diverge have a stronger (weaker) impact on the results displayed.

[illegible]

C Momentum Approaches

The next page provides an overview over the fourteen long-short momentums strategies, which are derived from previous work and relied on in this analysis in the paper. *ID* is a running number to identify anomalies, *Momentum Anomaly* offers a corresponding name. If not noted otherwise, all approaches are computed using a formation period of six months, a skipped month, and a holding period of six months with overlapping portfolios. If momentum is interacted with other firm variables, than this is done based on an independent 5*5 sorting in each month.

ID	Anomaly	Short Description	Reference Papers	Type of Portfolios	Computation
1	Standard momentum	Winners of the recent past outperform losers of the recent past	Jegadeesh and Titman (1993)	5 Past Returns (5-1)	Formation period return are computed as the cumulative return in months t-7 to t-1
2	Age-enhanced momentum	Momentum is stronger in younger firms	Zhang (2006)	5 Past Returns * 5 Age (5,1 - 1,1)	Age is computed as the number of months since min (IPO Date from Compustat, PERMCO appears in CRSP). The youngest firms are in quintile 1.
3	Turnover-enhanced momentum	Momentum is stronger among high turnover firms	Lee and Swaminathan (2000)	5 Past Returns * 5 Turnover (5,5 - 1,5)	We use NASDAQ-adjusted turnover as proposed in Anderson and Dyl (2005).
4	Market-to-book ratio-enhanced momentum	Momentum is stronger among firms with high market-to-book ratios	Asness (1997), Daniel and Titman (1999), Sagi and Seasholes (2007)	5 Past Returns * 5 Market-to-Book (5,5 - 1,5)	Book equity is computed on a quarterly basis following the approach of Chen et al. (2011).
5	Credit rating-enhanced momentum	Momentum is stronger among firms with low credit rating	Avramov et al. (2007)	5 Past Returns * 5 Credit Rating (5,5 - 1,5)	We rely on the S&P domestic long term issuer credit rating, which uses 22 ratings from AAA to D. Portfolio 1 denotes the firms with the best rating.
6	Size-enhanced Momentum	Momentum is stronger among small firms	Jegadeesh and Titman (1993), Hong et al. (2000), Zhang (2006)	5 Past Returns * 5 Size (5,1 - 1,1)	Size is computed based on NYSE market capitalization quintiles (i.e. stocks from the first NYSE deciles are not dropped though the 5\$ barrier remains). Portfolio 1 denotes the smallest firms.
7	(Residual) analyst coverage-enhanced momentum	Momentum is stronger among firms with low residual analyst coverage	Hong et al. (2000)	5 Past Returns * 5 Residual Analyst Coverage (5,1 - 1,1)	Residual analyst coverage is based on monthly cross-sectional regressions of $\ln(1 + \text{number analysts})$, 0 if missing) on one month lagged $\text{macap} + \text{nasdaq dummy}$. As in Hong et al. (2000), coverage is measured stale, i.e. in t-12. Portfolio 1 denotes the firms with the lowest coverage.
8	Forecast dispersion-enhanced momentum	Momentum is stronger among firms with high analyst forecast dispersion	Zhang (2006), Verardo (2009)	5 Past Returns * 5 Analyst Forecast Dispersion (5,5 - 1,5)	Forecast dispersion is based on the screens for residual analyst coverage. At least two analysts need to cover the firm. Portfolio 1 denotes the firms with the lowest forecast dispersion.
8	R^2 enhanced momentum	Momentum is stronger among low firms with low R^2	Hou et al. (2006)	5 Past Returns * 5 $R^2(5,1 - 1,1)$	R^2 is determined by regressing monthly firm excess returns on a Carhart (1997) model over the preceding 60 months. Values are then sorted in quintiles based on NYSE stock cut-offs. Portfolio 1 denotes the firms with the lowest R^2 .
10	Return consistency-enhanced momentum	Momentum is stronger for firms with consistently positive (or negative) returns during the formation period	Grinblatt and Moskowitz (2004)	5 Past Returns conditioned on winner/loser consistency dummy=1	The winner (loser) consistency dummy takes on a value of 1 if at least 5 out of the 6 formation period returns were positive (negative). Only these stocks are considered in the trading strategy.
11	(Idiosyncratic) volatility-enhanced momentum	Momentum is stronger for firms with high (standard/idiosyncratic) volatility	Zhang (2006), Jiang et al. (2005)	5 Past Returns * 5 Idiosyncratic volatility (5,5 - 1,5)	Idiosyncratic volatility is estimated as root mean squared error obtained from time-series regressions of monthly firm returns on a Carhart (1997) model over the months t-62 to t-2. Portfolio 1 denotes firms with low idiosyncratic volatility.
12	52 week high-enhanced momentum	Momentum is stronger among winners (losers) close (far away) from the 52 week high	George and Hwang (2004)	5 Past Returns * 5 Nearness to 52 week high (5,5 - 1,1)	Nearness to the 52 week high is computed as in George and Hwang (2004)
13	Formation period return-enhanced momentum	Momentum is stronger for firms with extreme formation period returns	Bandarchuk and Hilscher (2013)	10 Past Returns (10-1)	Formation period returns are computed as the cumulative return in months t-7 to t-1.
14	Intermediate momentum	Intermediate returns (i.e. t-12 to t-7) cause momentum	Novy-Marx (2012)	10 Past Returns (10-1)	Formation period returns are computed as the cumulative return in months t-12 to t-7

D Pairs Trading

D.1 Standardized unexpected earnings and top 100 pairs

To assess whether firms identified with the pairs trading techniques used in the paper are also fundamentally related, we start by pooling all monthly top 100 pairs from January 1990 on.² We then construct standardized unexpected earnings (SUE) for each firm during the following year. More specifically, we consider all cases in which both firms announce same-quarter earnings within the 12 months period immediately following pairs formation.³

We rely on Compustat Fundamental Quarterly to compute earnings per share excluding extraordinary items. Following standard practice (see e.g. Chordia and Shivakumar (2006)), SUE are constructed relying on a seasonal random walk model of quarterly earnings. Unexpected earnings are defined as the difference between the most recently announced earnings per share and the earnings per share in the same quarter of the previous year. Unexpected earnings are then standardized by the standard deviation of unexpected earnings computed over the previous eight quarters. For each quarter under consideration, we then compute SUE not only for the top 100 pairs as outlined above, but additionally for each firm with non-missing data in the CRSP/Compustat merged database. This allows us to take market-wide trends in earnings growth into account. As in e.g. Hirshleifer et al. (2009), we then perform quarterly decile sorts of all stocks based on their SUE.⁴ Thus, firms with the most negative (positive) earnings surprise in a given quarter obtain a value of 1 (10). The mean value is 5.5.

We then focus our attention on the in total 44,820 observations of same quarter SUEs for the top 100 pairs in the year following pair formation. More specifically, we are inter-

²The start date is chosen to ensure (earnings and earnings announcement) data quality and to match the start date used in several other tests in the paper (e.g. on the role of the VIX or the media). However, our inferences remain unchanged if we rely on earlier or later starting dates. The end year is 2009 as the last year of pairs formation is 2008.

³For instance, assume a top 100 pair whose formation period runs from January 2000 to December 2000. We then consider all quarterly earnings announced over January 2001 to December 2001. Findings are similar if we include a lag of one quarter between both periods.

⁴Findings are even slightly stronger if we rely on yearly as opposed to quarterly sorts.

ested whether knowing the first pair constituent's SUE for a given quarter helps to predict the second pair constituent's SUE for the same quarter. A significantly positive relation would be suggestive of a fundamental link between both firms. To answer this question, we run a pooled tobit regression of the 44,820 SUEs of the later announcing firms on the SUE of the corresponding earlier announcing firms. Standard errors are clustered by quarter. The variable of interest, the marginal effect of the earlier announcing firm's SUE, takes on a highly significant (t-statistic: 7.06) value of 0.0606. We obtain similar coefficients in multivariate settings controlling for e.g. differences in firm market capitalization or turnover.

Collectively, these findings suggest that the first constituent's earnings surprise has explanatory power for the second (=later announcing) constituent's earnings surprise. Note that this is not a simple industry effect as both firms are required to belong to different 49 Fama/French industries. Thus, this statistically highly significant and economically non-trivial result lends support to the idea that firms of top pairs are fundamentally related despite operating in different market segments.

D.2 Further time-series insights into the role of limited attention

As a first test, we analyze the role of other investor distraction proxies, which are inspired by previous work. Specifically, we construct five simple alternative dummy variables for limited attention in the time-series.

Following DellaVigna and Pollet (2009) and Peress (2008), we employ a Friday dummy.

Following the idea developed in Hirshleifer et al. (2009) and Peress (2008), we construct a variable based on the number of same-day events competing for investor attention. To this end, we compute the number of pairs opening on a given day. There is considerable variation in each year and no general time trend. The latter is not surprising, as the number of pairs eligible for trading remains constant after the first six months of the sample period, which we exclude. As the number of opening pairs is significantly positively correlated (0.25) with attention shift proxy decile ranks, we rely on the orthogonalized number of diverging pairs. Finally, we use the top and bottom quintile to identify days with an unexpectedly large number (dummy=1) or small number (dummy=0) of newly opening pairs.

With regard to the third and fourth proxy, we follow Hou et al. (2009) and Karlsson et al. (2009), who provide evidence that investors tend to be less attentive during down market periods. We create a dummy variable that takes the value of one if NBER classifies a month as recession and zero otherwise. An alternative dummy takes the value of one if the cumulative three year value-weighted market return is negative and zero otherwise.

Our fifth proxy is inspired by recent studies such as Da et al. (2011) or Mondria and Wu (2012). It uses the main Google-based distraction proxy briefly described in more detail in appendix B.1. Exploiting shocks in internet search volume between January 2004 and December 2008, the approach aims to quantify changes in the relative demand for market-level as opposed to firm-level information. As with the number of competing events, we again use the top and bottom quintile to identify days with a particularly strong demand (dummy=1) or weak demand (dummy=0) for aggregate stock market information.

We then imitate our baseline approach of section 4 in the paper by regressing pair returns separately on each of these limited attention dummies (specification 1) as well as

on the full set of control variables used in panel B of table 4, including the attention shift proxy decile rank (specification 2). As we have five alternative proxies, two regression specifications, and two return computation schemes, we run 20 regressions in total. Each of the attention dummies is constructed in a way that a positive coefficient is expected. The main findings are presented in panel A of table 5.

Table 5: Further Tests (I): Investor Attention and Pairs Trading Profitability

Panel A reports the impact of alternative limited attention dummies on the profitability of pairs trading. Details about the construction of each proxy are given in the text. Specification 1 displays findings from univariate tests. Specification 2 is multivariate (see panel b of table 4 in the paper for details). Due to the limited number of observations, the multivariate regression with the Google-based dummy does not controls for industry groups. In all cases, standard errors are adjusted for heteroscedasticity and clustered by day of pair divergence. T-statistics are reported in parentheses. Panel B shows the implied percentage increase in return difference between enhanced high and low distraction scenarios when benchmarked against the baseline findings in table 4 of the paper. Enhanced high (low) distraction scenarios are defined as days satisfying both distraction proxy decile rank=10 and alternative limited attention dummy=1 (distraction decile=1 + alternative dummy=0). In panel C, we compare mean and median returns of pairs opening on the last trading day before federal holidays with mean and median returns of pairs opening on any other day of the year. The table shows the fraction of years in which returns on pre-holiday pairs trading are higher. P-values (in parentheses) are computed from one-sided binominal probability tests with an assumed yearly success rate of 50%. *Excess probability of convergence* is computed as the difference between the fraction of converging pairs that diverged immediately before the federal holiday and the fraction of converging pairs that diverged on any other day of the year. T-statistics are reported in parentheses. In all panels, statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Impact of Limited Attention Proxies Inspired by Previous Work					
Model specification	Fridays	No. pairs opening	NBER recessions	3Y market return	Google
Specification 1	0.0002	0.0033***	0.0090***	0.0055***	0.0117*
(no further controls)	(0.21)	(3.18)	(8.99)	(5.51)	(1.72)
Specification 2	0.0007	0.0008	0.0041**	0.0003	0.0234*
(full set of controls)	(0.28)	(0.71)	(2.31)	(0.25)	(1.81)
Panel B: Implied Percentage Increase in the Return Difference between High Distraction and Low Distraction Scenarios					
Model specification	Fridays	No. pairs opening	NBER recessions	3Y market return	Google
Specification 1 (no further controls)	10.30%	1.34%	99.35%	71.98%	23.30%
Specification 2 (full set of controls)	42.70%	2.60%	40.86%	41.65%	18.98%
Panel C: Relative Success of Pre-Holiday Pairs Trading					
Pairs Trading Returns	New Year's Day	Washington's Birthday	Memorial Day		
Mean	72%***	61%*	47%		
Median	70%***	46%	45%		
Excess probability of convergence	15.77%***	-0.48%	3.87%		
(baseline: 36.15%-36.3%)	(3.67)	(-0.14)	(0.94)		
Pairs Trading Returns	Independence Day	Labor Day	Thanksgiving	Christmas Day	
Mean	51%	60%	61%*	70%***	
Median	53%	60%	55%	63.%*	
Excess probability of convergence	5.61%	3.87%	0.42%	7.39%*	
(baseline: 36.15%-36.3%)	(1.35)	(1.14)	(0.10)	(1.68)	

As predicted, the coefficient is positive in all 20 cases although some proxies lack significance, in particular if one controls for calendar, industry, market, firm, and pair characteristics. The persistently positive sign is broadly consistent with the idea of limited attention affecting the relative efficiency of linked assets.

Given these insights, we explore whether the sensitivity of pairs trading returns to attention shift proxy decile ranks becomes even higher when considering the proxy’s possible interaction with the attention proxies inspired by previous work.

Specifically, consider the following two extreme scenarios. The first sample consists of all days with attention shift proxy decile rank 10 for which, at the same time, the alternative distraction proxy also identifies a distracting situation (e.g. a Friday or a down market period). In these cases, attention constraints should become particularly binding. The second sample consists of all days with attention shift decile rank 1 for which, at the same time, the alternative distraction proxy also identifies a situation in which attention constraints should be less binding. We expect the difference in returns of pairs opening in one of these two extreme situations to be larger than the return difference between decile 10 and decile 1 in our baseline scenario (see table 4 in the paper).

We regress one-month pairs returns on the attention shift proxy decile rank, on the alternative limited attention dummy variable, and the interaction effect (as well as on a large set of control variables). Again, we have 20 regressions in total. Panel C of table 5 shows the implied percentage change in return difference between high and low distraction scenarios, as outlined above, when benchmarked against our baseline findings. As hypothesized, return difference increase in each case.

Next, we study whether pairs trading is particularly profitable immediately before those seven federal holidays for which the NYSE has been closed over the whole sample period. This line of reasoning is backed up by work on holiday effects (e.g. Frieder and Subrahmanyam (2004), Hong and Yu (2009)) and the distracting nature of even “ordinary Fridays” (DellaVigna and Pollet (2009)). One might be concerned about liquidity-related market frictions on these days. Indeed, market turnover is on average about 10% to 20% lower. However, this does not translate into lower turnover for those stocks which are constituents of pairs diverging on these days. In fact, turnover for these firms is even

slightly higher than in some pre-event period. Also bid-ask spreads are hardly affected. A plausible explanation for this finding is that these firms are likely to be affected by news, which stipulates trading activity even in a low market turnover environment.

We compare returns on pairs that diverge on the last trading day before the holiday with returns on pairs that open on any other day. Specifically, for each year and each holiday separately, we determine whether the mean (median) pre-holiday pairs trading return is larger or smaller than the return over the rest of the year. Panel D of table 5 shows the fraction of years in which pre-holiday pairs trading is more profitable. The fraction is larger than 50% in 11 out of 14 cases, and often also statistically significant.

Before Christmas and New Year’s Day, pairs trading seems particularly successful. For instance, in about 70% of all sample years, mean and median returns from pairs opening on the last trading day of the year are higher than corresponding returns over the rest of the year. This is substantially driven by a considerably higher than usual fraction of converging pairs. While on average less than 37% of pairs converge within a month, more than 50% do if they diverge immediately before New Year’s Day.

D.3 Investor attention shifts, time-varying arbitrage risk, and pairs trading profitability

To study the possible interaction between limits to arbitrage, our measure of investor attention shifts, and the event-time profitability of pairs trading, we start by constructing a set of arbitrage risk proxies derived from the literature. More specifically, we consider the following six time-series measures:

1. The VIX, i.e. the Chicago Board Options Exchange Market Volatility Index. The VIX is based on the implied volatility of S&P 500 index options. It is thus a measure of expected stock market volatility.

2. The TED spread, i.e. the difference between the 3-months LIBOR Eurodollar rate and the 3-months T-Bill rate. The TED spread is a risk-related spread, as short-term US government debt is considered virtually riskless, whereas the LIBOR additionally reflects (perceived) credit risk in interbank loans.

3. An additional credit spread, defined as the difference between Moody's BAA corporate bond and Moody's AAA corporate bond rates

4. The term spread, i.e. the difference between 10 year Treasury bond yield and 3 months T-bills yield.

5. The LIBOR, i.e. the London Interbank Offer Rate (LIBOR).

6. A measure building on the abnormal stock returns of the nine institutions studied in Ang et al. (2011): Bear Stearns, Citigroup, Credit Suisse, Goldman Sachs, HSBC, JP Morgan, Lehman Brothers, Merrill Lynch, and Morgan Stanley. Ang et al. (2011) argue that, at least over their sample period from 2004 to 2009, these firms might be considered representative of the broker/dealer and investment banking activity. Ang et al. (2011) for instance show that the average cost of protection from a default of these institutions rises in times of high arbitrage risk as proxied by the VIX. We construct daily abnormal returns for each of the aforementioned banks defined as the absolute difference between the actual stock return and its expected return as implied by an OLS market model. Parameter estimates are obtained from rolling time-series regressions based on daily return data over the previous year. We then value-weight these abnormal returns to create an index of abnormal daily returns for investment banks. One might argue that large unexpected arbitrage risks might reflect themselves in negative abnormal returns for these institutions, which are likely to be engaged in arbitrage activities.⁵

Databases used for the construction of the six arbitrage proxies are Thomson Reuters Datastream, Bloomberg, CRSP, and the Federal Reserve Board H15 release. On a daily basis, the proxies become available between January 1982 and January 1990 and then, in each case, run until the end of our sample period in December 2008. The tables on the following slides show the sample period for each specification. We both consider the raw values of arbitrage proxies as well as their yearly decile ranks, computed as for the attention shift proxy. In total, this leaves twelve measures of limits to arbitrage.

⁵Note that one also could come up with an alternative hypothesis: If these institutions run trading strategies which tend to work better in times of high investor distraction, and if these times are not characterized by increased arbitrage risk, then the abnormal stock return for these institutions might actually be positive. See for instance Menzly and Ozbas (2010) for evidence on arbitrageurs engaging in trading strategies that profit from the cross-predictability effect in stock prices induced by slow information diffusion. If anything, our findings reported in the following tables support this second hypothesis, rendering explanations based on increased arbitrage risks unlikely.

The analysis then proceeds in four steps, whose main findings are displayed in the following four tables. First, we test the impact of time-varying arbitrage risk on pairs trading profitability by running twelve univariate regressions.

Second, we document Spearman rank order correlation coefficients between attention shift proxy decile ranks and the rank order of the twelve arbitrage proxies.

Third, we run multivariate regressions as in panel B of table 4 of the paper. More specifically, we run pooled multivariate regressions of the one-month return on zero-cost US stock-pairs on the high/low distraction dummy, the full set of controls, and, additionally, on arbitrage risk proxies. We start by considering only one arbitrage proxy in each of the twelve regressions we run. This is motivated by the partly high correlation between these proxies. It moreover helps to distinguish which arbitrage risks really matter for the case of pairs trading.

Finally, in a forth step, we again run multivariate regressions, but now consider more than one arbitrage proxy in each regression. More specifically, we run ten regressions with different sets of those arbitrage proxies which have been turned out to be most relevant in the tests in step 3.

Table 6: The Impact of Time-Varying Arbitrage Risk on Pairs Trading Profitability: Univariate Regressions

This table displays findings from pooled regressions of the one-month return on zero-cost US stock pairs on proxies deemed to measure time-varying arbitrage risk. With regard to these proxies, we either use its raw value or its decile ranks, computed separately for each year. Pairs trading returns are computed under the conservative “one day waiting” return scheme. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. T-statistics (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence.

Model specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample start (End: 12/2008)	1/1990	1/1990	1/1986	1/1986	1/1986	1/1986	1/1986	1/1986	1/1982	1/1982	1/1982	1/1982
Observations	39,666	39,666	48,481	48,481	48,481	48,481	48,481	48,481	57,484	57,484	57,484	57,484
Adjusted R ²	0.12%	0.15%	0.08%	0.14%	0.18%	0.08%	0.01%	0.00%	0.12%	0.00%	0.12%	0.00%
VIX (Raw Value)	0.028*** (3.54)											
VIX (Yearly Decile Ranks)		0.094*** (5.22)										
TED Spread (Raw Values)			0.516*** (7.11)									
TED Spread (Yearly Decile Ranks)				0.071*** (4.33)								
Moody's BAA-AAA (Raw Values)					0.83*** (4.83)							
Moody's BAA-AAA (Yearly Decile Ranks)						0.071*** (4.30)						
LIBOR 3M (Raw Values)							0.037* (1.69)					
LIBOR 3M (Yearly Decile Ranks)								0.007 (0.42)				
Term Spread (Raw Values)									0.207*** (5.45)			
Term Spread (Yearly Decile Ranks)										-0.002 (-0.13)		
IB Abnormal Return (Raw Values)											0.148 (0.34)	
IB Abnormal Return (Yearly Decile Ranks)												0.0142 (0.89)
Controls for calendar and industry effects	no	no	no	no	no	no	no	no	no	no	no	no
Controls for market-level conditions	no	no	no	no	no	no	no	no	no	no	no	no
Controls for pair characteristics	no	no	no	no	no	no	no	no	no	no	no	no

Table 7: Pair-wise correlations of proxies for arbitrage risk and distraction decile ranks

Panel A: Pairwise correlations (raw values of arbitrage proxies)								
		Distraction Deciles	VIX	TED Spread	Moody's BAA-AAA	LIBOR 3M	Term Spread	IB Abnormal Return
Distraction Deciles	Start	1.00						
VIX	1/1990	0.23	1.00					
TED Spread	1/1986	0.12	0.54	1.00				
Moody's BAA-AAA	1/1986	0.08	0.62	0.50	1.00			
LIBOR 3M	1/1986	-0.01	-0.05	0.40	-0.09	1.00		
Term Spread	1/1982	0.02	0.01	-0.05	0.24	-0.56	1.00	
IB Abnormal Return	1/1982	0.05	0.00	0.04	0.01	0.01	0.01	1.00
Panel B: Pairwise correlations (yearly decile ranks of arbitrage proxies)								
		Distraction Deciles	VIX	TED Spread	Moody's BAA-AAA	LIBOR 3M	Term Spread	IB Abnormal Return
Distraction Deciles	Start	1.00						
VIX	1/1990	0.31	1.00					
TED Spread	1/1986	0.12	0.14	1.00				
Moody's BAA-AAA	1/1986	0.15	0.21	0.07	1.00			
LIBOR 3M	1/1986	-0.05	-0.04	0.33	-0.13	1.00		
Term Spread	1/1982	0.02	0.01	0.15	-0.05	-0.15	1.00	
IB Abnormal Return	1/1982	0.04	0.05	0.01	0.04	0.00	0.02	1.00

Table 8: The Impact of Time-Varying Arbitrage Risk and Investor Attention Shifts on Pairs Trading Profitability: Multivariate Regressions (1)

This table displays findings from pooled multi-variate regressions of the one-month return on zero-cost US stock pairs on the high/low distraction dummy, a set of controls as well as on proxies deemed to measure time-varying arbitrage risk. With regard to these arbitrage proxies, we either use its raw value or its decile ranks, computed separately for each year. All calendar-, industry-, market-, and pair-level control variables from panel B of table 4 in the paper are also included. Pairs trading returns are computed under the conservative “one day waiting” return scheme. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. T-statistics (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence.

Model specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample start (End: 12/2008)	1/1990	1/1990	1/1986	1/1986	1/1986	1/1986	1/1986	1/1986	1/1982	1/1982	1/1982	1/1982
Observations	8,372	8,372	10,422	10,422	10,422	10,422	10,422	10,422	12,434	57,484	8,372	8,372
Adjusted R ²	7.07%	7.06%	7.76%	7.66%	7.60%	7.60%	7.60%	7.60%	7.22%	7.22%	7.26%	7.26%
High/Low Distraction Dummy	0.783*** (3.02)	0.702*** (2.71)	0.741*** (3.14)	0.750*** (2.98)	0.714*** (3.17)	0.736*** (2.96)	0.736*** (3.05)	0.647*** (3.07)	0.666*** (3.02)	0.676*** (3.11)	0.612*** (2.83)	0.612*** (2.77)
VIX (Raw Value)	0.038 (1.22)											
VIX (Yearly Decile Ranks)		0.050 (1.09)										
TED Spread (Raw Values)			1.08*** (2.96)									
TED Spread (Yearly Decile Ranks)				0.077** (2.12)								
Moody's BAA-AAA (Raw Values)					0.323 (0.54)							
Moody's BAA-AAA (Yearly Decile Ranks)						0.033 (0.92)						
LIBOR 3M (Raw Values)							-0.118 (-0.68)					
LIBOR 3M (Yearly Decile Ranks)								-0.03 (-0.89)				
Term Spread (Raw Values)									0.318* (1.93)			
Term Spread (Yearly Decile Ranks)										0.06* (1.91)		
IB Abnormal Return (Raw Values)											12.36** (2.29)	
IB Abnormal Return (Yearly Decile Ranks)												0.078** (2.52)
Controls for calendar and industry effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls for market-level conditions	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls for pair characteristics	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Table 9: The Impact of Time-Varying Arbitrage Risk and Investor Attention Shifts on Pairs Trading Profitability: Multivariate Regressions (2)

This table displays findings from pooled multi-variate regressions of the one-month return on zero-cost US stock pairs on the high/low distraction dummy, a set of controls as well as on proxies deemed to measure time-varying arbitrage risk. With regard to these arbitrage proxies, we either use its raw value or its decile ranks, computed separately for each year. All calendar-, industry-, market-, and pair-level control variables from panel B in table 4 in the paper are also included. Pairs trading returns are computed under the conservative “one day waiting” return scheme. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. T-statistics (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence.

[illegible]

Figure 5: Cumulative Distribution Function of Pairs Trading Profits by Attention Shift Deciles

This figure shows the empirical cumulative distribution function of event-time one-month returns on zero-cost portfolios of US stock pairs sorted by attention shift proxy decile ranks as observed on the day of divergence. Breakpoints for the deciles are determined separately for each year. We only consider low distraction days (=decile 1) and high distraction days (=decile 10). In Panel A (B), trading positions in each pair are initiated on the day of divergence (on the day following the convergence) and liquidated on the day of convergence (on the day following the convergence). For better readability, extreme returns (larger than 20% or smaller than -20%) are not shown. Extreme returns account for roughly 1% of all sample observations.

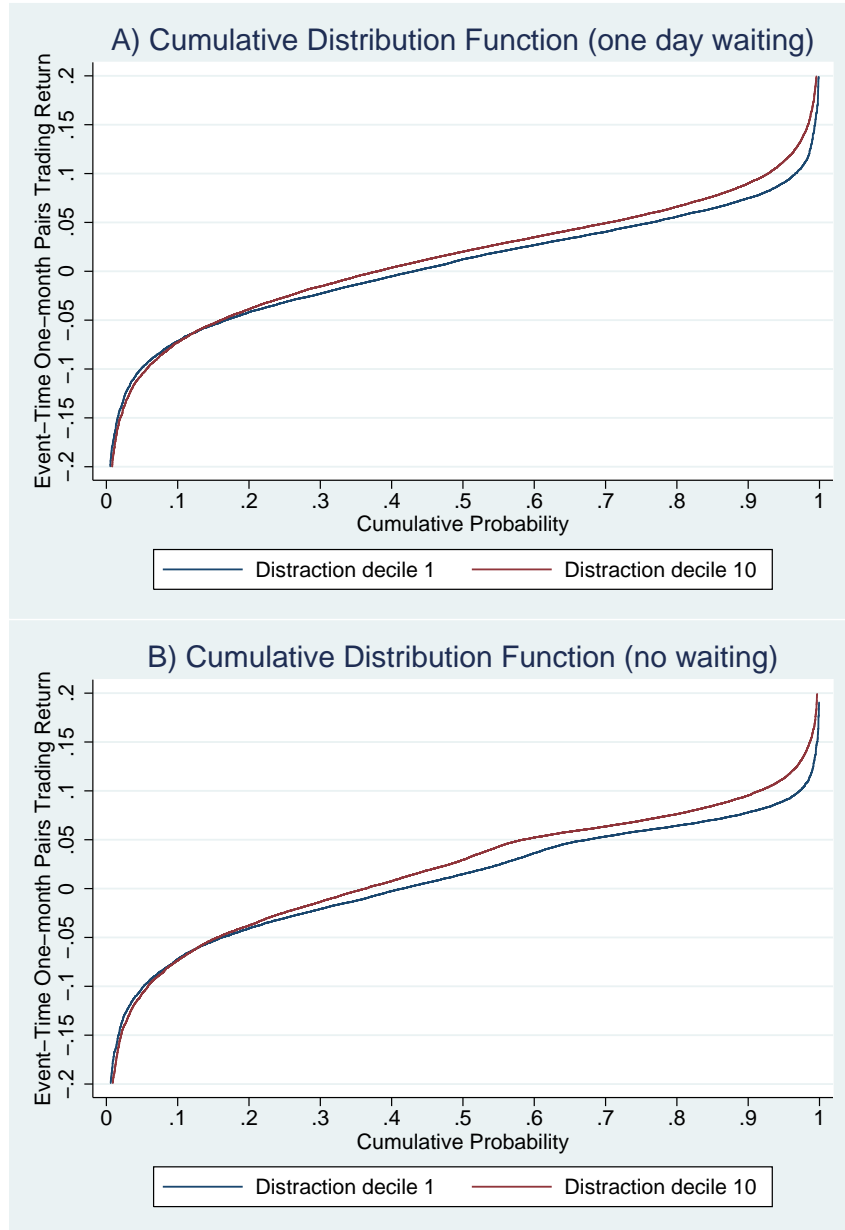


Figure 6: Pairs Trading Returns and Modified Attention Shift Proxies

This figure shows average one month event-time returns on US long-short stock pairs sorted by the decile ranks of the baseline attention shift proxy (model 1) and in total twelve modified attention shift proxies. Tables 3 and 4 in the appendix give more details about the specific modifications. Table 12 in the appendix also provides a statistical analysis of return differences between decile 1 and decile 10.

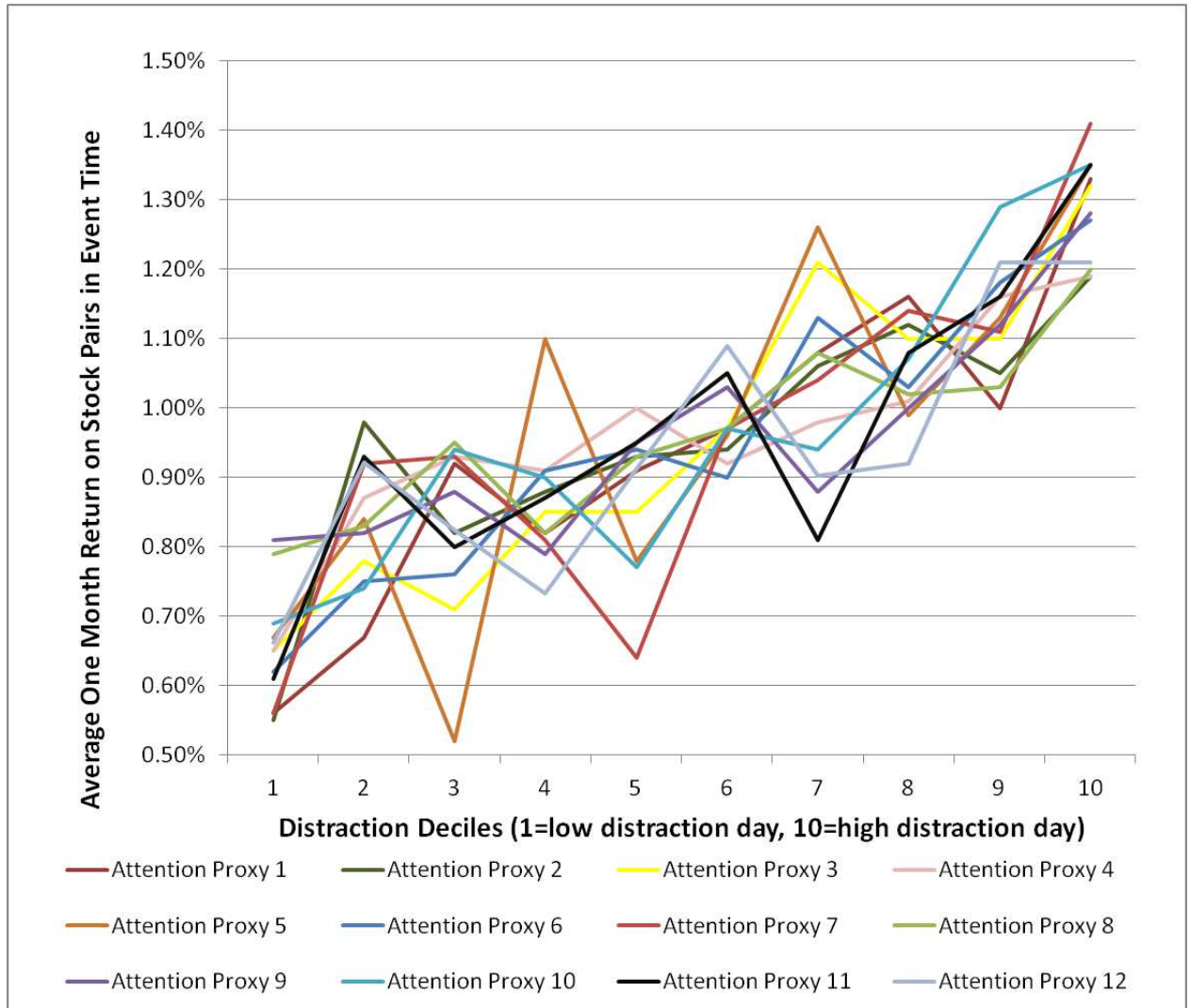


Table 10: Distribution of One-month Pairs Trading Abnormal Returns by Distraction Deciles

This table reports distribution details of event-time one-month returns on zero-cost portfolios of US stock pairs sorted by distraction proxy decile ranks as observed on the day of divergence. Breakpoints for the deciles are determined separately for each year. Calculations are based on daily data from January 1962 to December 2008. In Panel A (B), trading positions in each pair are initiated on the day of divergence (on the day following the convergence) and liquidated on the day of convergence (on the day following the convergence).

Distraction Decile	All	1	2	3	4	5	6	7	8	9	10
Panel A: One Day Waiting, Full Sample Period (1962-2008)											
N	103,386	8,187	8,679	9,146	9,398	10,048	10,079	10,595	11,019	12,224	14,011
mean	0.97%	0.53%	0.86%	0.77%	0.87%	0.95%	0.94%	1.02%	1.10%	1.09%	1.30%
sd	6.57%	6.20%	6.30%	6.53%	6.32%	6.58%	6.56%	6.54%	6.52%	6.73%	7.05%
p1	-17.83%	-17.23%	-17.08%	-18.08%	-17.37%	-17.04%	-18.70%	-17.72%	-17.73%	-18.06%	-18.59%
p10	-7.04%	-7.16%	-6.60%	-7.02%	-6.95%	-7.02%	-6.95%	-6.98%	-6.94%	-7.28%	-7.28%
p25	-2.73%	-3.16%	-2.73%	-2.77%	-2.70%	-2.76%	-2.61%	-2.70%	-2.58%	-2.68%	-2.64%
p50	1.59%	1.24%	1.41%	1.36%	1.52%	1.57%	1.49%	1.62%	1.71%	1.72%	2.01%
p75	5.23%	4.80%	4.97%	5.05%	5.02%	5.17%	5.24%	5.27%	5.33%	5.47%	5.71%
p90	8.21%	7.49%	7.79%	7.85%	7.90%	.0809082	8.16%	8.28%	8.31%	8.54%	9.00%
p99	15.19%	13.49%	14.62%	14.41%	13.81%	15.04%	15.25%	15.46%	14.92%	16.07%	16.77%
Panel B: No Waiting, Full Sample Period (1962-2008)											
N	104,125	8,222	8,738	9,179	9,436	10,094	10,122	10,657	11,146	12,332	14,199
mean	1.38%	0.89%	1.20%	1.08%	1.23%	1.31%	1.35%	1.33%	1.53%	1.56%	1.90%
sd	6.77%	6.40%	6.41%	6.70%	6.57%	6.75%	6.71%	6.76%	6.69%	6.95%	7.30%
p1	-18.55%	-17.21%	-17.65%	-18.74%	-18.33%	-17.98%	-19.22%	-19.17%	-17.93%	-18.61%	-19.29%
p10	-7.11%	-7.20%	-6.65%	-7.13%	-7.06%	-7.05%	-7.00%	-7.10%	-7.13%	-7.31%	-7.36%
p25	-2.62%	-3.03%	-2.58%	-2.74%	-2.61%	-2.63%	-2.44%	-2.75%	-2.49%	-2.63%	-2.45%
p50	2.14%	1.48%	1.80%	1.68%	1.96%	2.03%	2.09%	2.16%	2.35%	2.54%	2.94%
p75	6.33%	5.90%	5.96%	6.03%	6.11%	6.27%	6.29%	6.33%	6.45%	6.62%	6.95%
p90	8.47%	7.79%	7.99%	8.04%	8.18%	8.33%	8.32%	8.45%	8.57%	8.82%	9.54%
p99	13.80%	12.12%	12.40%	13.12%	12.83%	13.12%	12.99%	13.50%	13.63%	14.30%	15.85%

Table 11: Performance Ratios (Monthly) of Pairs Trading by Distraction Deciles

This table reports performance ratios of event-time one-month returns on zero-cost (long-short) portfolios of US stock pairs sorted by distraction proxy decile ranks as observed on the day of divergence. Breakpoints for the deciles are determined separately for each year. Calculations are based on daily data from January 1962 to December 2008. In Panel A (B), trading positions in each pair are initiated on the day of divergence (on the day following the convergence) and offset on the day of convergence (on the day following the convergence). The Sharpe Ratio is computed as the average return of a trade divided by the average standard deviation. The Rachev Ratio refers to the expected tail return divided by the expected tail loss. We compute it as the average return above the median return divided by the (absolute value of the) Conditional VaR at the 95% level. The Rachev Ratio is thus highest for strategies that best manage to simultaneously deliver high returns and provide insurance for high losses. The Sortino Ratio is computed as the average return divided by its downside volatility.

Distraction Decile	All	1	2	3	4	5	6	7	8	9	10
Panel A: One Day Waiting, Full Sample Period (1962-2008)											
N	103,386	8,187	8,679	9,146	9,398	10,048	10,079	10,595	11,019	12,224	14,011
Sharpe Ratio	0.148	0.086	0.136	0.118	0.138	0.144	0.144	0.156	0.168	0.162	0.185
Rachev Ratio	0.390	0.366	0.386	0.366	0.383	0.387	0.376	0.398	0.405	0.400	0.413
Sortino Ratio	0.266	0.187	0.239	0.202	0.243	0.248	0.253	0.262	0.305	0.301	0.343
Panel B: No Waiting, Full Sample Period (1962-2008)											
N	104,125	8,222	8,738	9,179	9,436	10,094	10,122	10,657	11,146	12,332	14,199
Sharpe Ratio	0.204	0.139	0.187	0.162	0.187	0.194	0.201	0.197	0.228	0.224	0.261
Rachev Ratio	0.426	0.400	0.414	0.394	0.412	0.420	0.404	0.422	0.448	0.441	0.465
Sortino Ratio	0.195	0.117	0.178	0.147	0.182	0.187	0.186	0.211	0.224	0.218	0.244

Table 12: Modified Attention Shift Proxies: One-month Pairs Trading Returns by Distraction Deciles

This table reports one-month event-time returns on zero-cost portfolios of stock pairs sorted by attention shift proxy deciles as observed on the day of the divergence. Breakpoints for the attention shifts deciles of each proxy are determined separately for each year. Standard errors (in parentheses) are adjusted for heteroskedasticity and clustered by day of pair divergence. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Attention Distraction Proxy No.	1	2	3	4	5	6	7	8	9	10	11	12
Distraction Decile 1	0.0056*** (0.00100)	0.0055*** (0.000967)	0.0065*** (0.00103)	0.0065*** (0.00104)	0.0067*** (0.00104)	0.0062*** (0.00101)	0.0056*** (0.000979)	0.0079*** (0.00106)	0.0081*** (0.00106)	0.0069*** (0.000970)	0.0061*** (0.000929)	0.0066*** (0.000916)
Distraction Decile 10	0.0133*** (0.000981)	0.0119*** (0.000990)	0.0132*** (0.00101)	0.0119*** (0.000997)	0.0135*** (0.00102)	0.0127*** (0.000997)	0.0141*** (0.00104)	0.0120*** (0.000938)	0.0128*** (0.00102)	0.0135*** (0.00103)	0.0135*** (0.00105)	0.0121*** (0.000935)
Difference (10-1)	0.0077*** (0.00140)	0.0064*** (0.00138)	0.0067*** (0.00145)	0.0053*** (0.00144)	0.0068*** (0.00146)	0.0065*** (0.00142)	0.0085*** (0.00143)	0.0041*** (0.00142)	0.0047*** (0.00147)	0.0066*** (0.00141)	0.00742*** (0.00140)	0.00552*** (0.00131)

Table 13: High and low distraction pairs conditioned on the level of return differential on the day of divergence

To demonstrate that the level of the return differential on the day of divergence as well as the number of opening pairs do not drive our findings, we perform conditional sorts. We first focus on the subsample of pairs that diverge on a high or low distraction day. The first row (“Baseline”) of the following table shows the results. High distraction pairs outperform low distraction pairs by 77 basis points per month. However, there are 14,011 pair trades on high distraction days, while there are only 8,187 trades on low distraction days. In addition, the return difference on the day of divergence is 3.59% on high distraction days and only 2.38% on low distraction days. Thus, it could be the case that our findings were primarily driven by these factors instead of by the attention shift proxy. We therefore restrict our focus to the subsample of high or low distraction pairs which at the same time exhibit a return differential at the day of divergence of less than a given threshold. We start with a cut-off of 5% and then step-wise impose stricter screens until we arrive at a threshold of 1%. As the table shows, this has three consequences: First, the number of opening pairs on high and low distraction days becomes very similar. Under the 1% and 2% constraints, even more pairs open on low than on high distraction days. Second, also the return differential attributable to the day of divergence becomes similar. Under the 1% constraint, the difference is 0.00%. Thus, these two aspects can hardly drive potential differences in pairs trading profitability during high and low distraction days in this setting. It is worth noting that most other firm and pair characteristics are very similar across attention shift decile ranks, as table 3 in the paper verifies. Third, however, the impact of the attention shift proxy remains very stable, significant, and economically large under each constraint. Thus, holding the number of opening pairs as well as the return differential constant, the attention shift proxy matters to about the same extent as in the baseline analysis. It is worth noting that similar insights are obtained from multivariate regressions (table 4 of the paper), in which we include the return difference on the day of divergence as a control variable. Coefficients are shown on the following page.

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	N	N	Return diff (divergence)	Return diff (divergence)	(3)-(4)	Pairs Trading Return	Pairs Trading Return	(6)-(5)
	High distraction	Low distraction	High distraction	Low distraction		High distraction	Low distraction	
Baseline	14,011	8,187	3.59%	2.38%	1.21%	1.30%	0.53%	0.77%***
max. 5% return differential	11,227	7,619	2.50%	1.99%	0.51%	1.20%	0.55%	0.65%***
max. 3% return differential	7,403	6,222	1.79%	1.58%	0.20%	1.14%	0.50%	0.64%***
max. 2% return differential	4,217	4,273	1.26%	1.18%	0.07%	1.18%	0.45%	0.73%***
max 1% return differential	1,275	1,537	0.65%	0.65%	0.00%	1.14%	0.27%	0.87%***

Table 14: Multivariate Analysis: Attention Shifts and Returns on Pairs Trading

This table displays coefficients obtained from model specifications 3 and 6 of panel B, table 4 in the paper. These coefficients are estimated from pooled multivariate regressions of the one-month return on zero-cost US stock pairs (in %) on the attention shift proxy as well as on a number of control variables. The attention shift proxy is quantified by means of the *Attention shift deciles* (left-hand side, model 3 in table 4) or a *High/Low Distraction Dummy* (right-hand side, model 6 in table 4), which is zero for low distraction days (decile 1) and one for high distraction days (decile 10). Displayed are coefficients for all pair characteristics computed as outlined in table 3 and 4 of the paper (average firm market capitalization decile rank, ln (average pre-event turnover), ln (average pre-event Amihud illiquidity ratio), average idiosyncratic risk, bid/ask spread as in Corwin and Schultz (2012), S&P membership dummy, within-pair differences in these variables, return difference attributable to the day of divergence, ln (average turnover on day of divergence) and ln (difference in turnover on day of divergence)). Displayed are also coefficients obtained for controls of market-level conditions on the day of divergence (market return, squared market return, market turnover, 10 day rolling volatility, factors for daily return premia on size, value, momentum and short-term reversal, average bid/ask spread.). Included in the regression (but not shown) are moreover controls for calendar and industry effects (indicator variables for year, month, day of week, and pair industry group combinations). Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively. T-statistics (in parentheses) are adjusted for heteroscedasticity and clustered by day of pair divergence.

Model specification	Model (3), panel B, table 4 in the paper		Model (6), panel B, table 4 in the paper	
Sample period	1/1962-12/2008		1/1962-12/2008	
Observations	103,386		22,198	
R ²	3.75%		9.04%	
Attention shift deciles	0.039***	(3.33)		
High/low distraction dummy			0.501***	(3.03)
Pair-level variables				
Difference in NYSE/AMEX macap decile	0.011	(0.34)	0.039	(0.62)
Difference in (pre-event) Amihud illiquidity ratio	0.012	(0.44)	-0.085	(-1.38)
Difference in (pre-event) daily turnover	-0.087***	(-3.90)	-0.143***	(-2.95)
Difference in (pre-event) idiosyncratic risk	0.186	(1.14)	0.326	(1.03)
Difference in turnover on day of divergence	-0.034	(-1.35)	0.012	(0.23)
Difference in bid/ask spread on day of divergence	0.056*	(1.88)	0.104	(1.44)
Difference in S&P membership	-0.020	(-0.37)	-0.166	(-1.38)
Average NYSE/AMEX macap decile	0.073	(1.11)	0.106	(0.76)
Average (pre-event) Amihud illiquidity ratio	0.209*	(1.68)	0.250	(0.94)
Average (pre-event) daily turnover	0.076	(0.70)	0.596**	(2.53)
Average (pre-event) idiosyncratic risk	0.757***	(4.11)	1.065***	(2.85)
Average turnover on day of divergence	-0.014	(-.20)	-0.135	(-1.04)
Average bid/ask spread on day of divergence	-0.067**	(-2.14)	-0.115	(-1.50)
Average S&P membership	0.140	(1.32)	0.179	(0.77)
Return difference at day of divergence	-0.038	(-0.02)	1.909	(0.46)
Market-level variables at day of divergence				
Average bid-ask spread	0.731***	(3.25)	0.611**	(2.28)
Rolling 10 day volatility	0.158	(1.56)	0.078	(0.49)
Squared market return	-0.002	(-0.61)	0.007	(1.64)
Market turnover	1.14**	(2.32)	0.540	(0.69)
Market premium over risk-free rate	-0.025	(-0.25)	0.052	(0.86)
Small firm factor (SMB)	-0.089*	(-1.70)	-0.238**	(-2.57)
Value firm factor (HML)	-0.046	(-0.54)	0.108	(0.89)
Momentum factor (UMD)	-0.002	(-0.03)	-0.026	(-0.28)
Short-term reversal factor	-0.061	(-1.15)	-0.085	(-1.12)
Controls industry effects	yes		yes	
Controls calendar effects	yes		yes	

Table 15: Descriptive Statistics for Pairs Trading Samples in International Stock Markets

In panel A, *Total market capitalization* refers to the value reported by Thomson Financial Datastream for the total domestic market capitalization at the year-end 2002, which roughly marks the middle of the sample period for most countries . In panel B, *Number of industries* states how many of the 10 Global Industry Classification Standard (GICS) industry sectors are represented in the pairs trading sample. *Maximum industry weight* denotes the largest fraction of sample firms belonging to a specific industry group. *Industry concentration* is computed as the sum of squared industry weights. In panel C, values are computed analogously for within-pair industry group combinations.

		Japan	UK	France	Germany	Switzerland	Italy	Netherlands	Hongkong
Panel A: Overall Market Characteristics									
Sample period		1/1995-12/2009	1/1995-12/2009	1/1996-12/2009	1/1996-12/2009	6/1997-12/2009	6/1995-12/2009	1/1995-12/2009	1/1995-12/2009
Total market cap. (in billion USD)		2100.19	1819.29	877.87	651.57	540.76	448.22	429.80	417.61
Number of sample firms		4,873	4,867	1,424	1,302	387	500	354	521
Number of firm days (in million)		12.00	6.35	2.11	2.54	0.32	0.47	0.68	0.87
Panel B: Firm Characteristics									
Number of industries		10	10	10	10	10	10	9	10
Max. industry weight	Fraction	26.50%	20.88%	25.88%	21.29%	30.13%	31.65%	26.99%	38.11%
	Industry	Industrials	Cons. Discretionary	Financials	Industrials	Financials	Financials	Financials	Financials
Industry concentration		16.16%	14.81%	16.46%	21.29%	18.07%	18.79%	17.46%	22.16%
Panel C: Pair Characteristics									
Total no. of pairs traded		36,992	27,185	25,833	26,006	16,841	25,596	21,320	20,894
Average turnover	Mean	0.16%	0.44%	0.21%	0.35%	0.43%	0.41%	0.50%	0.56%
at day of divergence	Median	0.09%	0.28%	0.12%	0.20%	0.25%	0.26%	0.28%	0.29%
Cumul. return difference	Mean	6.53%	8.17%	9.86%	11.70%	12.41%	10.90%	11.13%	16.50%
upon divergence	Median	5.99%	7.37%	9.45%	11.20%	11.47%	9.96%	10.69%	14.96%
Return difference	Mean	3.70%	3.21%	4.07%	4.73%	4.24%	4.30%	4.18%	5.91%
at day of divergence	Median	2.88%	2.40%	3.30%	3.69%	3.00%	3.34%	3.07%	4.46%
No. industry group comb.		41	45	44	42	42	44	36	45
Max. industry group weight	Fraction	11.55%	13.98%	13.39%	12.40%	14.96%	21.29%	19.20%	24.82%
	Industries	Industrials	Cons. Discretionary	Industrials and	Industrials	Health Care	Con. Discretionary	Industrials	Industrials
		and Financials	and Financials	and Financials	Cons. Discretionary	and Financials	and Financials	and Financials	and Financials
Industry group concentration		6.61%	6.00%	6.71%	6.29%	8.31%	9.26%	7.83%	13.05%

References

- Anderson, A.-M., and E. A. Dyl, 2005, "Market structure and reported trading volume: NASDAQ versus the NYSE," *Journal of Financial Research*, 28, 115–131.
- Ang, A., S. Gorovyy, and G. B. van Inwegen, 2011, "Hedge fund leverage," *Journal of Financial Economics*, 102, 102–126.
- Asness, C. S., 1997, "The interaction of value and momentum strategies," *Financial Analysts Journal*, 61, 29–36.
- Avramov, D., T. Chordia, G. Jostova, and A. Philipov, 2007, "Momentum and credit rating," *Journal of Finance*, 62, 2503–2520.
- Bandarchuk, P., and J. Hilscher, 2013, "Sources of momentum profits: Evidence on the irrelevance of characteristics," *Review of Finance*, 17, 809–845.
- Campbell, J. Y., J. Hilscher, and J. Szilagyi, 2008, "In search of distress risk," *Journal of Finance*, 63, 2899–2939.
- Carhart, M. M., 1997, "On persistence in mutual fund performance," *Journal of Finance*, 52, 57–82.
- Chen, L., R. Novy-Marx, and L. Zhang, 2011, "An alternative three-factor model," *Unpublished working paper, Washington University, University of Rochester, Ohio State University*.
- Chordia, T., and L. Shivakumar, 2006, "Earnings and price momentum," *Journal of Financial Economics*, 80, 627–656.
- Corwin, S. A., and P. Schultz, 2012, "A simple way to estimate bid-ask spreads from daily high and low prices," *Journal of Finance*, 67, 719–759.
- Da, Z., J. Engelberg, and P. Gao, 2011, "In search of attention," *Journal of Finance*, 66, 1461–1499.
- Daniel, K., and S. Titman, 1999, "Market efficiency in an irrational world," *Financial Analysts Journal*, 55, 28–40.
- DellaVigna, S., and J. M. Pollet, 2009, "Investor inattention and Friday earnings announcements," *Journal of Finance*, 64, 709–749.

- Fama, E. F., and K. R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.
- , 1997, "Industry costs of equity," *Journal of Financial Economics*, 43, 153–193.
- Frieder, L., and A. Subrahmanyam, 2004, "Non-secular regularities in returns and volume," *Financial Analysts Journal*, 60, 29–34.
- George, T. J., and C.-Y. Hwang, 2004, "The 52-week high and momentum investing," *Journal of Finance*, 59, 2145–2176.
- Grinblatt, M., and T. M. Moskowitz, 2004, "Predicting stock price movements from past returns: the role of consistency and tax-loss selling," *Journal of Financial Economics*, 71, 541–579.
- Hirshleifer, D., S. S. Lim, and S. H. Teoh, 2009, "Driven to distraction: Extraneous events and underreaction to earnings news," *Journal of Finance*, 64, 2289–2325.
- Hong, H. G., T. Lim, and J. C. Stein, 2000, "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies," *Journal of Finance*, 55, 265–295.
- Hong, H. G., and J. Yu, 2009, "Gone fishin': Seasonality in trading activity and asset prices," *Journal of Financial Markets*, 12, 672–702.
- Hou, K., L. Peng, and W. Xiong, 2006, "R2 and Price Inefficiency," *Unpublished working paper, Ohio State University, City University of New York, Princeton University*.
- , 2009, "A tale of two anomalies: The implications of investor attention for price and earnings momentum," *Unpublished working paper, Ohio State University, City University of New York, Princeton University*.
- Jegadeesh, N., and S. Titman, 1993, "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance*, 48, 65–91.
- Jiang, G., C. M. Lee, and Y. Zhang, 2005, "Information uncertainty and expected returns," *Review of Accounting Studies*, 10, 185–221.
- Karlsson, N., G. Loewenstein, and D. Seppi, 2009, "The ostrich effect: Selective attention to information," *Journal of Risk and Uncertainty*, 38, 95–115.
- Kumar, A., 2009, "Who gambles in the stock market?," *Journal of Finance*, 64, 1889–1933.

- Lee, C. M., and B. Swaminathan, 2000, "Price momentum and trading volume," *Journal of Finance*, 55, 2017-2069.
- Lo, A. W., and J. Wang, 2000, "Trading volume: Definitions, data analysis, and implications of portfolio theory," *Review of Financial Studies*, 13, 257-300.
- Lou, D., and C. Polk, 2012, "Comomentum: Inferring arbitrage capital from return correlations," *Unpublished working paper, London School of Economics*.
- Menzly, L., and O. Ozbas, 2010, "Market segmentation and cross-predictability of returns," *Journal of Finance*, 65, 1555-1580.
- Mondria, J., and T. Wu, 2012, "Asymmetric attention and stock returns," *Unpublished working paper, University of Toronto, University of California*.
- Novy-Marx, R., 2012, "Is momentum really momentum?," *Journal of Financial Economics*, 103, 429-453.
- Peress, J., 2008, "Media coverage and investors' attention to earnings announcements," *Unpublished working paper, Insead*.
- Sagi, J. S., and M. S. Seasholes, 2007, "Firm-specific attributes and the cross-section of momentum," *Journal of Financial Economics*, 84, 389-434.
- Stambaugh, R. F., J. Yu, and Y. Yuan, 2012, "The short of it: Investor sentiment and anomalies," *Journal of Financial Economics*, 104, 288-302.
- Tkac, P. A., 1999, "A trading volume benchmark: Theory and evidence," *Journal of Financial and Quantitative Analysis*, 34, 89-114.
- Verardo, M., 2009, "Heterogeneous beliefs and momentum profits," *Journal of Financial and Quantitative Analysis*, 44, 795-822.
- Yuan, Y., 2012, "Market-wide attention, trading, and stock returns," *Unpublished working paper, University of Pennsylvania*.
- Zhang, X. F., 2006, "Information uncertainty and stock returns," *Journal of Finance*, 61, 105-137.