Cross-Firm Information Flows and the Predictability of Stock Returns

Anna Scherbina*
UC Davis

Bernd Schlusche†
Federal Reserve Board

This draft: October 11, 2014

ABSTRACT

We identify leader stocks based on their ability to Granger-cause returns of other stocks and show that thus-identified leaders can reliably predict their followers' returns out of sample. Leaders' predictive ability is robust to firm- and industry-level controls and works at the level of individual stocks rather than industries. Many leaders cannot be easily detected using ex-ante firm characteristics: They are often small, belong to a different industry than their followers, and exhibit only a short-lived leadership. We find support for the conjecture that leaders tend to be at the center of important news developments that also affect their followers by showing that, all else equal, firms with greater news coverage have a larger number of followers. We furthermore find that, consistent with the view that equilibrium mispricing is related to arbitrage costs, more heavily traded stocks react to their leaders' signals with a shorter delay. Finally, we present evidence that sophisticated investors trade on leaders' signals.

JEL classification: G10, G12, G14, G17

Keywords: Information Leadership, Lead-Lag Effect, Corporate News Announcements, Limited Attention, Market Efficiency

*Address: Graduate School of Management, University of California, Davis, One Shields Avenue, Davis, CA 95616. E-mail: ascherbina@ucdavis.edu. Phone: (530) 754-8076.

†Address: Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue NW, Washington, DC 20551. E-mail: bernd.schlusche@frb.gov. Phone: (202) 452-2591.

The views expressed in this paper are those of the authors and not necessarily those of the Board of Governors, other members of its staff, or the Federal Reserve System. Anna Scherbina acknowledges the support of the Q Group. We are grateful to Alex Baptista, Dong Lou, Rajnish Mehra, Andreas Neuhierl, and Raghu Rau, seminar participants at George Washington University, ETH Zürich, Johns Hopkins University, the Securities and Exchange Commission, Virginia Tech, and the University of Georgia, as well as the participants of the Luxembourg 2nd Asset Management Summit and the City Research Quantitative Conference and our discussant, Yuliya Plyakha, for their helpful comments.

Cross-Firm Information Flows and the Predictability of Stock Returns

ABSTRACT

We identify leader stocks based on their ability to Granger-cause returns of other stocks and show that thus-identified leaders can reliably predict their followers' returns out of sample. Leaders' predictive ability is robust to firm- and industry-level controls and works at the level of individual stocks rather than industries. Many leaders cannot be easily detected using ex-ante firm characteristics: They are often small, belong to a different industry than their followers, and exhibit only a short-lived leadership. We find support for the conjecture that leaders tend to be at the center of important news developments that also affect their followers by showing that, all else equal, firms with greater news coverage have a larger number of followers. We furthermore find that, consistent with the view that equilibrium mispricing is related to arbitrage costs, more heavily traded stocks react to their leaders' signals with a shorter delay. Finally, we present evidence that sophisticated investors trade on leaders' signals.

JEL classification: G10, G12, G14, G17

Keywords: Information Leadership, Lead-Lag Effect, Corporate News Announcements, Limited Attention, Market Efficiency

I. Introduction

In early 1994, six African-American employees of Texaco Inc. filed a racial discrimination lawsuit against their employer claiming that they were discriminated against in salaries and promotions. In an attempt to speed up a resolution, Reverend Jesse Jackson called for a national boycott of Texaco Inc. The lawsuit was eventually settled in late 1996 for over $140 million, making it the largest settlement for a racial discrimination case at the time. As described in a November 17, 1996, *New York Times* article, the settlement potentially affected other companies as well.[1] In particular, Rev. Jackson announced not only that the Texaco boycott would continue but also that his organization, the Rainbow PUSH Action Network, would study the affirmative action policies of other companies that shared directors with Texaco Inc., such as Gillette, Johnson & Johnson, and Campbell Soup. The article also quoted a lawyer representing firms in discrimination lawsuits as saying, "If you are a consumer-product company, you are quite vulnerable. If you're an Exxon, or an American Express, or a Texaco, it's a big exposure."

One can easily think of other examples for when news about a firm at the center of a valuation-relevant issue could affect other firms that stand to gain or lose depending on how that issue is resolved. A discovery of questionable accounting practices at one firm can cause investors to question financial statements of other firms that apply similar accounting techniques. Labor scandals or product safety concerns may negatively impact other firms with comparable production processes. When a firm expands to a new country with a yet unproven track record of dealing with foreign businesses, news about that firm will be relevant for other firms also seeking to expand to that country. In this paper, we demonstrate the existence of a collection of "bellwether" stocks for many individual stocks that are able to forecast the stocks' returns.[2]

[1] "Size of Texaco Discrimination Settlement Could Encourage More Lawsuits," by Steven A. Holmes, *New York Times*, November 17, 1996.

[2] Of course, the bellwether stocks do not have to be limited to the firms in the news. They may include firms with high levels of investor attention, single-segment firms, or firms in the same supply chain, as per the earlier literature reviewed later in the section.

The direction of a stock's return leadership may be positive or negative, depending on the setting. For example, bankruptcy rumors will have a negative impact on customers, suppliers, and providers of capital, but a positive impact on the firm's competitors. Labor scandals in developing countries that involve U.S. corporations may spread to other U.S. firms that use cheap labor abroad, but corporations based entirely in the United States could benefit by attracting socially-minded investors and consumers. Similarly, some firms stand to lose and some to win depending on how a patent infringement lawsuit is resolved.[3]

These examples illustrate that information can flow in unexpected directions. Firms at the center of a valuation-relevant issue can lead the returns of stocks that are larger, operate in a different industry, and are not tied to the leader by supply chain links. Moreover, the leadership can be short-lived and disappear once the issue is resolved.

This paper contributes to the literature on slow information diffusion. There is ample evidence that prices react slowly to a firm's own news (e.g., the post-earnings announcement drift). Prices may react slower still when the relevant news is announced by a different firm, especially when that news is of non-routine nature, making it difficult to immediately assess its effect on the firm's value.

Every day, a large number of firms release new information. Neuhierl, Scherbina, and Schlusche (2013) have obtained a nearly complete list of corporate press releases issued between April 2006 and August 2009, which allows a glimpse at the volume of information

[3]This is illustrated by a recent copyright infringement lawsuit that was initiated by publisher John Wiley & Sons and was tried by the Supreme Court. The case was determining whether it is allowed to purchase a copyrighted item in one market and then undersell the copyright owner's local price in another, more expensive, market. In that case, the petitioner in the Supreme Court case, Surap Kirtsaeng, resold John Wiley & Sons' foreign-edition textbooks at a higher price in the United States than the price he paid for them elsewhere and was subsequently sued by John Wiley & Sons, the respondent in the Supreme Court case. A diverse set of firms, spanning several industries, filed amicus briefs in this case. In particular, the Association of American Publishers, the Motion Pictures Association of America, the Business Software Alliance, and the Software and Information Industry Association, among others, which prefer that goods may be sold at different prices in different markets without anyone engaging in price arbitrage, filed amicus briefs in support of John Wiley & Sons, while Ebay, Costco, Google, the American Library Association, the Association of Art Museum Directors, Powell's Books Inc., the Association of Service and Computer Dealers International, and other organizations that prefer goods to be purchased and resold freely across markets filed amicus briefs in support of the opponent.

flow.[4] After discarding what is deemed to be inessential announcements, their sample still contains about 218 press releases a day, of which only 19.61% announce financial news (such as earnings, sales, dividends, plans to raise or return capital, etc.). The rest of the sample consists of less routine news about products, partnerships, strategic plans, corporate lawsuits, changes in management, and so on. These news announcements have the potential to affect valuations of other firms, but assessing the degree of their relevance may be difficult in real time.

We show that news indeed travels slowly across stocks and that trading strategies can be devised to exploit delays in stock price reactions at monthly and weekly frequencies. We do not attempt to identify return leaders using ex-ante firm characteristics. Rather, we rely on the statistical ability of leader stocks to Granger-cause their followers' returns. Specifically, in every month (week) and for each combination of stocks i and j, we regress monthly (weekly) returns of stock i on the lag of its own return, the lag of stock j's return, and the lag of the market return, using rolling regression windows that are at least one year long. Stock j is said to Granger-cause the return of stock i if the absolute value of the t-statistic on stock j's lagged return exceeds 2.00 (or 2.57 in a robustness check). Having run these rolling regressions for all stock pairs, we are able to identify a set of leaders for each stock in each month (week), if such leaders exist. We hypothesize that the leaders' ability to forecast the return of their followers will persist for at least another month (week). Hence, we proceed to calculate an aggregate predictive signal from all leaders for a follower's return. We do so by multiplying the estimated regression coefficient on a leader's lagged return by its current-month's return and summing these signals across all leaders. We show that stocks with high aggregate leader signals earn high returns and stocks with low aggregate leader signals earn

[4]Since the adoption of Regulation Fair Disclosure in 2002, corporations must disclose all news that may affect their stock prices to all investors simultaneously and with minimal delay. Issuing press releases via major newswire services has become a widespread way to comply with the new requirements. This development made it possible to assess the sheer volume of important firm-level news.

low returns in the subsequent month (week), which indicates that these signals indeed have an out-of-sample predictive ability for followers' returns.[5]

Our methodology relates this paper to the literature on the lead-lag effect in stock returns. In that literature, stock prices of certain firms (followers) are shown to react with a delay to price innovations of other firms (leaders). Lo and MacKinlay (1990) document that leaders are large firms and followers are small firms by showing that large firms predict returns of small firms, but not vice versa. Although non-synchronous trading or time-varying expected returns could give rise to the lead-lag effect, Lo and MacKinlay (1990), Chordia and Swaminathan (2000), and Anderson, Eom, Hahn, and Park (2012) determine that only a small fraction of the effect can be attributed to these explanations. Subsequent studies have shown that ex-ante stock characteristics other than size are also positively associated with information leadership. These characteristics are analyst coverage (Brennan, Jegadeesh, and Swaminathan (1993)), institutional ownership (Badrinath, Kale, and Noe (1995)), and trading volume (Chordia and Swaminathan (2000)). Taken together, this evidence suggests that the lead-lag effect can be largely ascribed to slow diffusion of common information from stocks that enjoy high levels of investor attention to those that do not.

We take steps to ensure that the predictive ability of leaders cannot be attributed to non-synchronous trading. We limit the sample of followers to only the stocks that traded on the last day of the previous month (or on the last day of the previous week when weekly portfolios are formed), thus largely eliminating the concern about non-synchronous trading. Additionally, for the portfolio results, we require that all followers be priced above $5 per share, which ensures that portfolios are comprised of rather liquid stocks. Moreover, the predictive ability of the monthly strategy survives skipping one month before portfolio formation for equal-weighted portfolios, and the weekly strategy survives skipping up to three weeks for equal-weighted and up to two weeks for value-weighted portfolios.

[5] Moreover, we show that leaders identified at a monthly frequency and leaders identified at a weekly frequency have an independent forecasting ability.

As one may expect, leaders' predictive power is stronger for smaller followers. Yet, in contrast to the results in the lead-lag literature, the strategy works better when leaders are small rather than large stocks: Equal-weighting signals across leader stocks results in a stronger predictive power for the followers' returns than value-weighting signals across leaders. This finding suggests that information flowing from large firms is incorporated into the followers' prices faster than information flowing from small firms. This is not surprising. While large firms may be quicker to react to common macro or industry-wide news, small firms can themselves be the originators of relevant news, as indicated by the above examples. Yet investors initially are more likely to underreact to small-firm news due to limited attention. We further illustrate that leaders may be small stocks by restricting the set of leaders to stocks that are smaller than their followers and showing that the strategy works almost as well. All told, leaders may be small firms, and followers may be large firms.

Leaders' documented ability to predict their followers' returns is unlikely to be explained by data snooping. To illustrate that, we destroy the time dimension in our data. Specifically, we scramble our panel data along the time dimension while preserving each cross section. The leaders that we identify using this dataset are all false leaders that should not possess any predictive ability for their followers' returns, and we show that they indeed do not. Moreover, the anomaly exhibits the properties of other documented anomalies: Its magnitude declines over time and it is stronger for smaller, more neglected stocks. Finally, we show that short sellers increase their shorting demands for stocks that receive low leader signals, which suggests that sophisticated investors trade on leader signals.

Consistent with our conjecture about the impetus for information leadership, we show that the scope of leadership is indeed positively related to the intensity of news developments at the firm level. For this purpose, we have obtained the Thomson-Reuters News Analytics dataset that covers the period from April 1996 to December 2011. Using this dataset, we are able to confirm our conjecture that, controlling for other characteristics, firms with more news stories written about them tend to lead returns of a larger number of other firms.

However, the relation between leadership and news coverage is nonlinear. With very high levels of news coverage, the number of followers begins to decline, as followers' prices begin to react to the leaders' news more quickly.

Another important distinction from the lead-lag literature, besides our finding that small firms can lead returns of larger firms, is that we are able to make within-industry long-short bets. In contrast, Hou (2007) documents that large firms in a particular industry lead small firms in that industry, but not small firms in a different industry. Relying on this kind of large-firm signal would preclude making long-short bets *within* industries, as all stocks in the same industry will receive the same signal. Moreover, in a robustness check, we require that leaders reside in a different industry than their followers and show that the strategy still works. From an investor's perspective, intra-industry long-short bets offer a better hedge than long-short bets made over the entire stock sample without regard for portfolios' industry composition because such bets are industry-neutral, ensuring lower volatility of the long-short return differential.

Recent papers have uncovered new channels of cross-firm information flows. In particular, Menzly and Ozbas (2010) document that information travels between supplier and customer industries, and Hong, Torous, and Valkanov (2007) present evidence that some industries even have the ability to lead the entire market. The information transfer literature in accounting shows that early earnings announcers predict earnings surprises of late announcers within the same industry.[6] Again, these signals will be correlated for all followers within an industry, precluding within-industry long-short bets. Cohen and Lou (2012) show that information diffuses slowly from single-segment firms to multi-industry conglomerates. In this setting, the signals would be also correlated within an industry. Cohen and Frazzini (2008) find that information travels slowly through the supply chain; in that setup, followers in the same

[6] In contrast to the information transfer literature, the leaders' predictive ability documented here is not tied to their earnings announcement activity: When we limit the set of leaders to those that are not announcing earnings in the current month, they still reliably predict their followers' returns in the following month.

industry may receive uncorrelated signals, but as in the lead-lag literature, leaders would tend to be larger firms.

This aforementioned papers assume that the set of leaders for a given firm is predetermined by its customer/supplier ties or by the industry affiliation of its segments. The advantage of the Granger-causality methodology used in this paper is its ability to identify both stable (or recurring) leaders, such as those determined by supply-chain links, and transitory (or non-recurring) leaders, whose leadership for a given firm may be short-lived and disappear once a news development is resolved. Another advantage is that this methodology does not rely on grueling data collection and is not limited by data availability (for example, firms are required by the SEC to report only the identity of any customer that comprises more than 10% of a firms consolidated sales revenues, and hence smaller customers will be missing from the dataset). Finally, by identifying leaders directly from the data, this methodology allows to uncover new channels of information leadership.

This paper presents novel evidence of slow information diffusion by showing that stock prices are slow to incorporate information originating at other firms. Our setting lends itself well to investigating the speed of price discovery. Our analysis of break-even transaction costs suggests that the economic benefit of trading on this form of delayed information processing is not very high. In order to take advantage of the slow information diffusion, the long-short portfolios need to be formed quickly. As a result, the price impact of trade is likely to be high due to the inability to spread trades over time, which limits the dollar amount that could be profitably invested in the strategy. For more heavily traded stocks, price impacts will be lower. However, we show that stocks with high turnover incorporate their leader signals with a shorter delay, weakening the profitability of the trading strategy. These results support the view that the market, although not be perfectly efficient, is competitive, and that stock prices tend to fall within no-arbitrage bounds around the fair value (see, e.g., Shleifer and Vishny (1997) and Lo (2004)).

The paper proceeds as follows: Section II explains the methodology used to identify information leaders. Section III documents the ability of leaders to predict the returns of their followers out-of-sample. Section IV provides evidence that sophisticated investors trade on the strategy described in this paper. Section V investigates the determinants of leadership. Section VI concludes.

II. Identifying Information Leaders

We identify information leaders for each stock i based on its leaders' ability to Granger-cause stock i's return. Specifically, using a rolling window of 12 months (or 36 months) including the current month τ, we run the following monthly regression for each combination of stocks i and j:

$$Ret_t^i = b_0^{ij} + b_1^{ij} Ret_{t-1}^{mkt} + b_2^{ij} Ret_{t-1}^i + b_3^{ij} Ret_{t-1}^j + \epsilon_t^{ij}, \tag{1}$$

where we require that both stocks i and j have 12 (36) monthly return observations available. Stock j is assumed to Granger-cause the return of firm i if the absolute value of the t-statistic on the estimated regression coefficient $\hat{b}_3^{ij}$ is greater than 2.00 (or 2.57 in a robustness check). Furthermore, if the estimated coefficient $\hat{b}_3^{ij}$ is positive, we say that stock j is a positive leader of stock i, and if negative, a negative leader.

When choosing the length of the estimation window, two considerations need to be balanced. On the one hand, it is beneficial to have a longer regression period to reduce noise. On the other hand, making the rolling window overly long will prevent us from uncovering relatively short-lived leader-follower pairs. We therefore settle for two rolling window lengths, 12 months and 36 months.[7] For the weekly-frequency results, we estimate regression (1) with 52 weekly return observations, which adds up to about 12 months.

[7]In the Texaco example from the introduction, during the period from January 1994 to December 1997 when the lawsuit was ongoing, Texaco is identified as a positive leader for Gillette in January 1994 and from July to October 1994, as a positive leader for Campbell Soup from February to April 1994 and again in January 1996, as a negative leader for American Express from July to September 1995, and as a positive leader for American Express from April to July 1997 when the 12-month rolling regression window is used.

Many leaders are misidentified as such due to estimation noise. The following quick calculation illustrates how many stocks are likely to be falsely identified as leaders for each stock i. For each potential follower i, the average number of cross-sectional regressions (1) being run every month equals the average size of the monthly cross section of stocks minus one for stock i itself, or 3,304.68-1. Under the assumption that the leaders for stock i are all stocks j for which $|t\text{-statistic}(\hat{b}_3^{ij})| \geq 2.00$, if the distribution of the estimated coefficients $\hat{b}_3^{ij}$ is perfectly normal, the associated likelihood of falsely identifying as leaders stocks whose true coefficient $\hat{b}_3^{ij}$ equals zero is 4.55% (the two-tailed p-value corresponding to a t-statistic with an absolute value of 2.00). On average, this amounts to about 150 false leaders per follower.[8]

Table 1 provides some statistics for leaders and followers. The data are calculated as of January 31 of each year. Leaders are drawn from an unrestricted dataset that includes all stocks in the CRSP universe. Hence, it is larger than the set of stocks from which followers are drawn.[9] The table shows that every stock eligible to be classified as a follower has, on average, 287 leaders (stock-month observations with no leaders are assigned a value of zero). This does not imply that the difference between 287 and 150 equals the number of independent leaders. Many "true" leaders, especially large leaders for small followers, are likely to offer correlated signals by virtue of reacting to common information shocks ahead of the followers. Hence, the number of "independent" leaders is likely significantly smaller. Moreover, a vast majority, 84% of all firm-month observations, have at least one leader.

When focusing on stocks that have at least one leader, the table shows that positive leaders slightly outnumber negative leaders. The absolute value of the coefficient $\hat{b}_3^{ij}$ is about 0.9 for both positive and negative leaders. For a given follower, its leaders do not typically belong in the same industry, but more positive than negative leaders do. Finally, despite the share price restriction on the followers and none on the leaders, the table shows that a follower stock tends to be smaller, to have a lower turnover, and to be younger than its average leader

[8]As will be discussed later in the paper, the actual distribution is more fat-tailed, resulting in somewhat more false leaders.

[9]Our results are only slightly weakened when we limit the set of potential leaders to common stocks of U.S.-incorporated firms.

stock. The last sub-table sorts, every month, all followers into quintiles based on the number of leaders that a follower has. It can be seen that the stocks with the lowest number of leaders tend to be larger and more heavily traded than other stocks; this is consistent with the result form the lead-lag literature that smaller and less liquid stocks typically have more liquid large-stock leaders, which are simply the first to react to common macro news.

Table 2 reports how persistent the leader-follower pairs are through the years. Having identified a leader-follower pair on January 31 of year t, we calculate the probability that this leader-follower pair also existed up to 10 years back—in January of year $t - \tau$, with $\tau \in \{1, ..., 10\}$—conditional on both the leader and the follower being present in the CRSP dataset at least 12 months or 36 months, depending on the length of the rolling regression window used to identify leaders, prior to the January of year $t - \tau$. The results for the 12-month and 36-month rolling regression windows are reported in Panels A and B, respectively. The panels present these probabilities separately for all leaders, independent of the leadership sign in year t, requiring that the leadership sign is preserved in year $t - \tau$, and for positive and negative leaders only, analogously requiring that the positive (negative) leadership sign be preserved in year $t - \tau$. We take the probability that a leader-follower pair also existed 10 years earlier as a baseline, and report, for every year $t - \tau$, the "excess" probability relative to this baseline (probability in $t - \tau$ minus probability in $t - 10$).[10]

It can be seen that the probability of a leader-follower relation also existing up to five years prior is significantly higher than the baseline probability. Moreover, as expected, these probabilities decline smoothly when moving further back in time, since the firm pairs are likely to share fewer valuation factors. In Panel B, the estimated probabilities of leader-follower pairs being identified as such are substantially higher for prior years 1 and 2 than in Panel A because of the overlapping estimation windows. Positive leader-follower pairs are somewhat more persistent than negative leader-follower pairs. When compared to the baseline number of year $t-10$, the persistence of a leader-follower pair disappears around year

[10]We do not want to use the p-value corresponding to a t-statistic with an absolute value of 2.00 under the normality assumption because the underlying distribution of the estimated coefficients b_3 may be non-normal.

5 for all leader-follower pairs, and around year 7 for positive leader-follower pairs when leaders are identified with a 12-month estimation window; in case of a 36-month leader estimation window, the persistence disappears around years 7 and 8, respectively.[11]

III. Return Predictability

Having obtained a set of J_τ^i leaders for each stock i in month τ, if such leaders exist, we proceed to calculate the aggregate leader signal. We do so by simply summing up the products of each current month's (or week's) leader return and the estimated regression coefficient $\hat{b}_3$:

$$Signal_\tau^i = \sum_{j=1}^{J_\tau^i} w_j \hat{b}_{3\tau}^{ij} Ret_\tau^j, \tag{2}$$

where w_j is the weight on leader j's signal. In our main set of results, signals are either equal-weighted across stock i's leaders, in which case $(w_j = 1/J_\tau^i)$. In robustness checks, we value-weight leader signals using the leaders' market capitalization at the end of month $\tau - 1$. Figure 1 illustrates how the aggregate equal-weighted leader signal is computed.

The advantage of the equal- or value-weighted signal aggregation method is that it is simple. However, it could be improved along two dimensions. The first dimension of improvement would be to devise a more efficient weighting scheme that takes into account historical correlations between leaders' signals and the confidence with which coefficients $\hat{b}_3$ are estimated. Leaders could produce perfectly correlated signals when (1) they simply react with a shorter delay than their followers to common economy- or industry-wide shocks or (2) a subset of stocks react with a shorter delay than their followers to the news of a sole original leader. Currently, the weights on leaders' signals are independent of the leaders' return correlations or their relative forecasting ability. A more efficient weighting method would aim to underweight signals that had large prediction errors and high correlations with other signals over the estimation window and overweight signals that were more precise and

[11]Years $t - 6$ through $t - 9$ are omitted due to space constraints but available upon request.

had low correlations with other signals; this can be accomplished by choosing the optimal weights that would minimize the expected variance of the aggregate signal using the parameters estimated over the rolling window. The second dimension of improvement would focus on eliminating misidentified leaders. For example, leaders that lead very few stocks in a given month or week are likely to be "false" leaders, and their signals should be ignored. In the remainder of this section, we will show that our simple weighting schemes work well in predicting followers' returns, and, hence, we will leave the improvements in signal aggregation to future research.

In the following, we present results based on portfolio sorts and cross-sectional return regressions. Though our leaders can be any stocks, we restrict the set of potential followers to domestically based common stocks with share codes 10 or 11 that had a trade on the last day of the previous month (or on the last day of the previous week for weekly-frequency portfolios).[12] In all portfolio results, we require that followers be priced above $5 per share in 2011 inflation-adjusted dollars at the end of the last month (week). The data used in the paper are described in the Appendix section A1.

A. Monthly portfolio returns

1. Baseline specification

In the baseline specification, we identify leaders with 12-month rolling regression windows and equal-weight signals across leaders. Having estimated signals for each follower stock in month τ, we sort followers into deciles based on the aggregate leader signal within each of the 36 industries that remain after the industry "Irrigation Systems" drops out and the stocks in the industry labeled "Other" are discarded. We form portfolios at the beginning of month $\tau+1$ and hold them for one month. In the following month, new portfolios are formed based

[12] Our results are virtually unchanged when we also require that leaders be common stocks with share codes 10 or 11.

on the new set of leader signals. Figure 1 illustrates the timeline for our regression windows and portfolio formation.

Panel A of Figure 2 plots the value of \$1 invested in February 1929 at a monthly return equal to that earned on the zero-investment strategy of holding a long position in the decile-10 portfolio and a short position in the decile-1 portfolio. The solid line represents the cumulative return for value-weighted portfolios and the dashed line that for equal-weighted portfolios. The initial \$1 investment would have turned into \$2,010.09 by December 31, 2011 for the equal-weighted strategy. For the value-weighted strategy, it would have turned into only \$75.26. For the equal-weighted strategy, the cumulative return reached its peak in July 2008, at which point the initial \$1 investment was worth \$2,515.95; for the value-weighted strategy, the maximum of \$137.86 is reached in November 1998. The equal-weighted strategy experienced seven months of consecutive negative returns from July 1999 to January 2000 and five months of consecutive negative returns from August to December 2008. The value-weighted strategy experienced six months of consecutive negative returns from May to October 1999 and four months of consecutive negative returns from August to November 2008. During the market crashes of October 1929 and October 1987, both sets of returns were highly positive (they were 7.1% and 17.7% in October 1929, and 1.3% and 2.5% in October 1987, for equal-weighted and value-weighted portfolios, respectively).

Table 3 presents average monthly excess returns for various deciles of equal- and value-weighted follower portfolios (Panels A and B, respectively), along with return differentials between the high- and low-signal portfolios.[13] Over the 1929-2011 period, leaders possess significant out-of-sample predictive ability. Low-signal portfolios earn low returns and high-signal portfolios earn high returns, and returns increase smoothly in magnitude with the signal for both return-weighting methods. Moreover, the alphas of the lowest-signal portfolio (decile 1) are significantly negative for both equal- and value-weighted returns, and the alphas for

[13] All t-statistics are adjusted for autocorrelation in returns using the Newey and West (1987) methodology, and, for each specification, the number of lags is determined as the third root of the number of observations in the time series.

the highest-signal portfolio (decile 10) are significantly positive when equal-weighted, but not when value-weighted. The lack of significance of the value-weighted alpha on the high-signal portfolio suggests that positive information is incorporated faster than negative information, at least for larger stocks. This observation is consistent with the evidence of Hong, Lim, and Stein (2000) that bad news diffuses slower than good news. The return differentials between high- and low-signal portfolios are significantly greater than zero for both equal- and value-weighted portfolios and for all return measures (i.e., excess returns, or alphas relative to the market, three- or four-factor models). The monthly four-factor alphas on the return differentials are equal to 0.64%, with a t-statistic of 5.73, and 0.38%, with a t-statistic of 2.98, for the equal- and value-weighted portfolios, respectively. Since our portfolios are constructed to have the same industry loadings, industry-wide movements are canceled out for the return differentials, thereby minimizing their volatility and maximizing the Sharpe Ratio.

Panels C and D present factor loadings on the four-factor model for equal- and value-weighted portfolios, respectively. The panels show that the high-signal portfolios have significantly lower loadings on the market factor, but significantly higher loadings on the size, book-to-market, and momentum factors, indicating that high-signal firms behave like small value winners. Yet, these loading differentials do not subsume the predictive ability of the leader signal.

Finally, Panel E presents portfolio transition probabilities between the current and the future portfolio assignment, one, two, and 12 months ahead. In the calculations, we only consider those stocks that are present in the sample in both time periods and, as before, we form leader-signal-based portfolios within each of the 36 industries. The table shows that, while there is some persistence in portfolio assignments in the next two months, with the somewhat U-shaped transition probabilities indicting that the stocks in the highest- and lowest-signal portfolios have a higher chance of remaining there relative to other portfolio assignments. However, the stickiness in the portfolio assignments disappears 12 months into the future.

Table 4 presents monthly portfolio returns for the specification in which leaders are identified using 36-month rolling windows. Returns are equal-weighted in Panel A and value-weighted in Panel B. With a longer rolling regression window, regression coefficients can be estimated more precisely, but there is a lower chance of identifying short-term leaders. It can be seen that this methodology produces very similar returns to the baseline specification. Some differences between these two methods will be revealed in the robustness checks and the Fama-MacBeth cross-sectional regressions presented below.

In order to check how the leaders' return predictability is related to the followers' size, every month and within each industry, we sort stocks into size terciles. Then, within each size tercile and industry group, we form decile portfolios based on the leader signal in that month. As before, first sorting on industry allows us to explore the within-industry return predictability and has the additional advantage that industry-specific movements are absent from the portfolio return differentials. Table 5 reports four-factor alphas for the low- and high-signal portfolios (deciles 1 and 10) and for the return differentials between portfolios 10 and 1. The results show that the return differentials are significant for all size terciles but the magnitudes steadily decline as the average size of the followers increases and that the positive alphas of portfolio 10 are only significant for the lowest-size terciles. Both results are consistent with the results of the lead-lag literature showing that large stocks, by virtue of having more attention, react faster to new information. However, significantly negative alphas suggest that stocks across all size groups might be slow to react to negative information due to short-sale constraints.

2. How quickly are leader signals incorporated?

We check how long it takes for the leader signals to be incorporated into their followers' prices. We try skipping one month between the month in which the leader signals are computed and the month in which portfolios are formed. The results are presented in Table 6; Panel A presents the results for the 12-month, and Panel B for the 36-month rolling regression windows. The return differential is still significant for the equal-weighted portfolios but is no longer significant for the value-weighted portfolios. Moreover, the significance for the equal-weighted portfolios is largely explained by the significantly negative alphas of the low-signal portfolios. The alphas of the high-signal portfolios are no longer significantly positive. When two months are skipped from the month in which leader signals are calculated, none of the methods produces significant return differentials, suggesting that information is fully transmitted from leaders to followers within one month value-weighed portfolios and within two months for equal-weighted portfolios.

3. Alternative methods for aggregating leader signals

We also try four alternative methods of aggregating leader signals. Unlike the baseline specification (2), these methods do not involve the magnitude of the estimated regression coefficient $\hat{b}_3$, but only its sign: $Signal^i_\tau = \sum_{j=1}^{J^i_\tau} w_j \mathrm{sign}(\hat{b}^{ij}_3) Ret^j_\tau$. Throughout the paper, we will refer to the leader return weighting schemes that do not rely on the magnitude of $\hat{b}_3$ as "non-parametric" weighting schemes. We use four leader return weighting schemes: (1) equal-weighting; (2) weighting by the leaders' market capitalization as of the end of month $\tau - 1$; (3) weighting by the absolute value of the t-statistic of $\hat{b}_3$; and (4) weighting by the absolute value of $\hat{b}_3$.

The results are presented in Panel A of Table 7. A comparison with the results of Table 3 shows that the original specification produces more significant return differentials for the value-weighted portfolios, while weighting by the absolute value of the t-statistics of $\hat{b}_3$ works best for equal-weighted portfolios. Value-weighting leader returns produces the lowest return differentials, which suggests that signals from large leaders that are overweighted in this weighting scheme are incorporated by the followers faster than signals from small leaders, likely because large leaders are more visible.

Panel B of the table skips one month between the month in which leader signals are calculated and the month in which portfolios are formed, as in Table 6. With the exception of the method in which leader signals are non-parametrically value-weighted, the predictive ability of leaders persists for the equal-weighted return differentials but not for the value-weighted return differentials. And, as with the baseline weighting scheme, none of the return differentials are significant when two months are skipped before portfolio construction.

4. Alternative methods of portfolio construction and other robustness checks

The predictive power of leader signals is robust to a number of other variations of how portfolios are constructed or how leader signals are calculated. The results for these alternative specifications are reported in Table 8.

We begin by sorting followers on the leader signal, not within each industry, but over the *entire sample*. Portfolio returns are reported in Panel A of the table for the specification in which leaders are determined using 12-month rolling regressions and in Panel B for the specification that uses 36-month rolling regressions to identify leaders. The returns are similar to those reported for within-industry sorts (Tables 3 and 4). However, here, for a given return magnitude, the t-statistics are somewhat lower because portfolio returns tend to be more volatile. The reason is that the long and short portfolios are likely to have uneven industry loadings, and, as a result, the long-short portfolio has industry exposure.

The next two panels present portfolio returns for value-weighted leader signals, computed according to formula (2). In Panel C, a 12-month rolling regression window is used, and in Panel D, a 36-month window. The results are not as strong as in the specification in which the leader signals are equal-weighted, which implies that signals from large stocks are incorporated more quickly than with a one-month delay. (Incidentally, the lead-lag literature uses weekly return frequencies.)

Panels E and F present results for the 1990-2011 subperiod for both lengths of the rolling regression windows. Leaders identified with 36-month rolling regression windows have more significant predictive power in that time period than leaders identified with 12-month rolling regression windows. However, neither method produces significant four-factor alphas for value-weighted portfolios. As many other return anomalies, our return anomaly diminishes over time, especially for large stocks.

In order to conserve space, the remainder of the robustness tests are presented only for leaders identified with 12-month rolling regressions. In Panel G, signals exclusively from positive leaders are used in portfolio formation, and in Panel H, signals exclusively from negative leaders are used. In Panel G, both equal- and value-weighted portfolio return differentials are significant, suggesting that positive leaders lead returns for both small and large stocks. In Panel H, the return differentials are only marginally significant for equal-weighted portfolios and insignificant for value-weighted portfolios, which implies that the predictive ability of negative leaders is rather weak, at least for intra-industry sorts.

To illustrate that leaders need not belong in the same industry as their followers, we compute signals only from the leaders that belong to a different industry than the follower stock. As shown in Panel I, the signal from this restricted set of leaders works nearly as well as the signal from the unrestricted set of leaders.

In Panel J, in order to further distinguish our results from those in the lead-lag literature, we limit the set of leaders to the stocks that are smaller than the follower. The significantly positive return differentials indicate that smaller leaders can indeed lead returns of larger followers.

Next, we study the predictive ability of recurring and non-recurring leaders. In Panel K, for each follower, we consider only the leaders that were not identified as that follower's leaders in any month over the previous three years (non-recurring leaders). In Panel L, for each follower, we consider only the leaders that were identified as that follower's leaders in at least one month over the previous three years (recurring leaders). Signals from recurring leaders have a higher forecasting power than signals from non-recurring leaders, especially for value-weighted portfolios. However, one needs to be careful drawing definitive conclusions since the set of non-recurring leaders likely contains more noise, i.e., non-leaders that are mistakenly identified as leaders.

In order to make a distinction between our results and those in the information transfer literature and in Cohen and Frazzini (2008), which describe an underreaction to relevant earnings information announced by other firms, we include, in Panel M, only leaders that are *not* announcing earnings in the current month. Hence, the information in the leaders' current returns is likely unrelated to their earnings news. However, these leaders still forecast their followers' returns in the next month (the return differentials are somewhat lower than in earlier tables because the results in Panel M are based on the more recent sample period). In Panel N, we use only leaders that announce their quarterly earnings in the current month. The return differentials in this panel are somewhat lower in magnitude for equal-weighted portfolios than those in Panel M and are insignificant for value-weighted portfolios, probably because firms announcing earnings typically attract news coverage, which would lead follower stocks to react to the leaders' news with a shorter delay.

In Panels O and P, we introduce an alternative cutoff value for the absolute value of the t-statistic on the regression coefficient $\hat{b}_3$ used to identify leaders. Instead of 2.00, we use a 2.57 cut-off, which corresponds to the 1% two-tailed significance level. In Panel O, portfolios are formed within industries, and in Panel P, over the entire sample. It can be seen that the results are very similar to those that use the 2.00 cutoff (see Panels A and B of Table 3 and Panel A of Table 8, respectively).

Finally, in Panels Q and R, we allow some time to pass between the month in which leaders are identified and the month in which these leaders are used to calculate the aggregate leader signal. In Panel Q, we skip one month, which lowers the return differentials by a factor of about 44% compared to those in Panels A and B of Table 3. In Panel R, we skip 60 months, which renders the return differentials insignificant as the leader-follower relation is unlikely to survive such a long period.

5. Scrambling the data along the time dimension

One may become concerned that the return predictability documented here may be driven not by the lead-lag effect in stock returns as we claim but rather by some stock characteristic, such as, say, idiosyncratic volatility, with stocks at the extremes of this characteristic behaving as though they have extreme leader signals.[14] To address this concern, we argue that if the cross-sectional dimension of the data were to be preserved but the time series dimension used for identifying leaders broken by scrambling the dataset along the time dimension, the signals from newly identified (but in this case, definitely false) leaders will stop predicting their followers' returns.

Specifically, we preserve cross sections but scramble the data along the time dimension by assigning each time period a random number and then sorting all cross sections by this random number. We apply this algorithm to the original monthly-frequency sample from January 1990 to December 2011; the sample is kept relatively short to ensure that most stocks are in existence for a large stretch of the considered time period. Since we require that both stocks in every potential leader-follower pair have returns in the prior 12 months, a relatively short sample helps ensure the maximum number of possible leader-follower pairs. Using the scrambled sample, we re-run regression (1) for all pairs of stocks that have return observations for the past 12 months. As before, we select as leader-follower pairs those stocks for which $\mid t\text{-statistic}(b_3) \mid \geq 2.00$.

In 8.62% of the regressions, a leader is identified. This number is higher than the 4.55% p-value corresponding to the t-statistic of 2.00 that one would observe when the distribution of the estimated coefficients is normal, implying that the actual distribution is fat-tailed. In 4.23% of the regressions, a positive leader is identified, and in 4.39%, a negative leader.

Next, we calculate the equal-weighted leader signal according to equation (2) in each month t and use it to predict the follower return in the next calendar month $t + 1$ (note that we use the follower return in the actual month $t + 1$ and not the scrambled month $t + 1$ in

[14] Later in the paper, we run a set of cross-sectional regressions and try our best to include all relevant stock characteristics to mitigate this concern.

order to show that it is not the cross-sectional variation in some omitted stock characteristic that forecasts the cross section of the next month's returns). We then form decile portfolios in month t based on the leader signal in that month and check portfolio returns in month $t+1$.

The average leader signal is equal to -2.87% for the lowest-signal decile and 3.13% for the highest-signal decile when signals are equal-weighted; these numbers are -2.24% and 2.80%, respectively, when signals are value-weighted. The raw return differentials between the extreme leader signal deciles is 0.12% (t-statistic=0.71) for equal-weighted portfolios and 0.31% (t-statistic=1.24) for value-weighted portfolios. The corresponding four-factor alphas are 0.18% (t-statistic=0.98) and 0.33% (t-statistic=1.27) for the equal- and value-weighted portfolios, respectively. Thus, our methodology does not produce any return predicability on the scrambled data.

When 36-month rolling windows are used to identify leaders, the results are similar. A leader is found in 5.55% of the regressions run (in this case, the distribution of the estimated coefficients $\hat{b}_3$ is less fat-tailed than when using 12-month regression windows). In 2.95% of the regressions run, a positive leader is found, and in 2.60%, a negative leader is found. The average leader signal is -1.95% and 2.28% for the bottom and top decile portfolios, respectively, when signals are equal-weighted, and -1.67% and 1.90%, respectively, when value-weighted. The average raw return differential between the high- and low-signal portfolios is -0.17% (t-statistic= -1.08) for equal-weighted portfolios and -0.05% (t-statistic=-0.19) for value-weighted portfolios. The corresponding four-factor alphas are -0.20% (t-statistic=-1.21) and -0.04% (t-statistic=-0.14), respectively.

Our inability to generate return predictability on the scrambled data is in contrast to leaders' significant ability to predict returns on the actual data for the same time period (see the results for equal-weighted portfolios over the same sample period reported in Table 8, Panel E). These results confirm that the success of the strategy hinges on identifying leaders that truly exhibit return leadership for their followers and on this leadership being, to some extent, preserved out of sample.

B. Weekly portfolio returns

As previously discussed, one reason to switch our analysis to higher frequencies is that signals from leaders may be incorporated into their followers' prices faster than with a one-month delay. (Tellingly, the lead-lag literature uses weekly return frequencies to document the delayed price reaction of small relative to large firms.) Additionally, higher frequencies will generate more data points, allowing us to study the interaction between leadership and news coverage since the news coverage dataset, which starts only in April 1996, is relatively short. In this subsection, we therefore work only with weekly return frequencies. Weekly returns are computed as Monday-to-Friday returns using the CRSP Daily Stock file.

The weekly portfolio construction methodology is similar to the monthly one. We run regression (1) with weekly returns using 52-week rolling regression windows. Even though the window length is still about 12 months, we are able to estimate regression coefficients with greater precision. Once the leaders are identified, we form portfolios every Monday using the combined leader signal from the previous week and hold stocks in the portfolios for one week.

Panels A and B of Table 9 present weekly portfolio returns for equal- and value-weighted portfolios, respectively. The results show that the weekly strategy produces highly significant return differentials for both equal- and value-weighted portfolios over the period 1980-2011. These returns are also highly economically significant, amounting to about 28% per year for equal-weighted and 15% per year for value-weighted long-short portfolios.

Panels C and D report four-factor loadings (using weekly factor returns) for equal- and value-weighted portfolios, respectively. The long-short portfolio loads negatively on the market factor, and, when returns are value-weighted, the long-short portfolio additionally has a negative loading on the HML factor but a positive loading on the momentum factor. Overall, the four factors have little explanatory power for the return differentials, and the resulting alphas are close in magnitude to the raw return differential.

Panel E presents portfolio transition probabilities between the current and the future portfolio assignment, one, two, and 52 weeks ahead. As in the case of monthly-frequency transition probability calculations, we only consider those stocks that are present in the sample in both time periods and form leader-signal-based portfolios within each of the 36 industries. Again, the tables show that there is persistence in the portfolio assignments in the next two weeks but that it disappears 52 weeks into the future.

Panel B of Figure 2 plots the value of $1 invested on January 18, 1980 at a weekly return equal to that earned on the zero-investment strategy of holding a long position in the decile-10 portfolio and a short position in the decile-1 portfolio. The solid line represents the cumulative return for value-weighted portfolios and the dashed line that for equal-weighted portfolios. The initial $1 investment would have turned into $5,649.56 on December 30, 2011 for the equal-weighted strategy. For the value-weighted strategy, it would have turned into only $77.74.

The analysis presented in the remainder of this subsection is conducted at weekly frequencies to conserve space and to better match the availability of news and trading cost data, but the results are qualitatively similar when the analysis is performed at monthly frequencies.

1. The interaction between leader signals and followers' concurrent returns

Once a stock's leadership for a follower stock has become apparent to investors, the follower will start to react to the leader's signal with a shorter delay or no delay at all. In the latter case, a leader's signal will lose its ability to forecast returns. Moreover, conditioning future returns on the past leader signal may even become counterproductive due to the return reversal effect, which is strongly present at both monthly and weekly frequencies. If a follower's price has already moved in the same direction as the signal this week, it will likely move in the opposite direction in the subsequent week.

Since the predictive ability of the leader signal should be the strongest among followers whose prices have not yet co-moved with the signal, conditioning on the correlation between the leader signal and the follower's contemporaneous return would improve the leader signals' predictive ability. We check whether this is the case. Every week, all follower stocks are sorted into quintiles based on their leader signal and then, within each leader-signal quintile, into further quintiles based on their return in that week. Table 10 presents four-factor alphas of the subsequent week's portfolio returns. It can be seen that the leader-signal strategy works within each reversal quintile; it generates a return equal to, on average, about 40% of the return of the reversal-based strategy.[15] As expected, the highest-leader-signal/lowest-prior-week return portfolio (portfolio 51) generates the highest return in the subsequent week, and the lowest-leader-signal/highest-prior-week return portfolio (portfolio 15) generates the lowest return in the subsequent week. The four-factor alphas of the return differential between portfolios 51 and 15 is 1.69% per week (t-statistic=20.51) for equal-weighted portfolios and 1.03% per week (t-statistic=13.73) for value-weighted portfolios.[16]These results show that the performance of the leader-signal strategy can be substantially improved by conditioning on whether or not the followers' prices have likely already reacted to the leader signal.

2. Break-even trading costs

Weekly trading, though more profitable in recent years than monthly trading, entails significantly higher trading costs. The reason for the persistence of the abnormal returns of the strategy at weekly horizons is that it takes skill to manage frequent portfolio turnover; the money left on the table after trading costs is likely small, and the documented returns quickly disappear as the investment amounts increase. The trading costs of the weekly strategy are high because portfolios need to be assembled quickly, which leads to large price impacts. Moreover, portfolios need to be turned over frequently, further boosting trading expenses.

[15]Sorting independently on reversals and leader signals produces very similar results.

[16]As before, forming within-industry portfolios helps eliminate industry-wide price movements and thereby achieve higher t-statistics.

Of course, one could lower trading costs by holding stocks in the portfolio for longer than one week (we will discuss this possibility next) or by holding relatively liquid large stocks; both modifications, however, somewhat reduce the strategy's returns.

Here, we estimate break-even trading costs that would set the post-trading-cost return of the one-week holding strategy to zero. For this estimation, we assume that trading costs are identical across stocks and independent of the amount traded (obviously, in reality, trading costs are lower for more liquid stocks and for smaller trade amounts). Because there is some persistence in the leader signal, the weekly portfolio turnover is lower than 100%, which gives a slight advantage to value-weighted portfolios as they are unaffected by rebalancing costs.

We estimate break-even trading costs for the weekly-frequency trading strategies previously considered in the paper. (Trading costs are expressed in units of return, i.e., as the percentage cost per dollar of a stock traded.) For the simple high-minus-low leader-signal-deciles in Table 9, the break-even trading costs are equal to 0.15% for equal-weighted portfolios and 0.09% for value-weighted portfolios. For the differential between the corner portfolios based on leader signals and return reversals in Table 10 (portfolio 51 - portfolio 15), the break-even trading costs are 0.45% for equal-weighted portfolios and 0.27% for value-weighted portfolios. By way of comparison, using the TAQ dataset for the January 1983 to August 2001 time period, Sadka and Scherbina (2007) estimate a 0.25% average effective spread for a typical stock and a typical trade. Hedge funds are more skilled at minimizing trading costs than an average trader in the TAQ dataset, and their trading costs may easily fall below our estimated break-even values. Moreover, our break-even trading costs are estimated for a simple strategy of trading all stocks and holding portfolios for only one week; sophisticated investors may modify the trading strategy by investing only in liquid stocks and holding stocks for longer than one week, which would achieve lower break-even trading costs.

The relatively low values of the estimated break-even trading costs make it clear that the strategies trading on leader signals can support only small investment amounts since large amounts would entail large price impacts. Therefore, the profitability of trading on slow information diffusion that we document here is not very high. The market, though not perfectly efficient, is competitive in a sense that the arbitrage profits left on the table are small.

3. Alternative signal aggregation methods and the speed of information diffusion at weekly frequencies

As for the monthly-frequency leader signals, we try alternative methods for aggregating weekly leader signals and check their performance for various lags between the week in which the signals are computed and the week in which portfolios are formed. (The four alternative non-parametric methods of aggregating leader returns are described on page 16).

The results are shown in Table 11. As in the case of monthly-frequency signals, the baseline method works best for value-weighted portfolios, while weighting leader returns by the absolute value of the t-statistics of $\hat{b}_3$ works best for equal-weighted portfolios; and, again, value-weighting of leader returns produces the worst return predicability.

This simple analysis shows that leader signals are fully incorporated into the equal-weighted portfolios within the next four weeks, and into the value-weighted portfolios within the next two to three weeks, depending on the specification. However, if trading costs are one of the main impediments to trading on the leader signal, the leader signal will be incorporated more slowly by less liquid stocks. We test for this possibility next. For this test, we consider only our baseline leader-signal aggregation specification (2).

We assume that turnover proxies for liquidity. Therefore, at the end of the week in which the leader signal is computed, we sort stocks into quintiles based on their turnover over the previous 12 months (using the most recent month-end), and then, within each turnover quintile, into leader-signal deciles. Table 12 presents the four-factor alphas of the high-minus-low decile return differentials for equal-weighted portfolios. The rows of the table are organized by the number of weeks skipped after the week in which the leader signal is calculated. The results indicate that the more liquid stocks tend to incorporate their leader signals faster—the long-short strategy's profits are lower for high-turnover stocks across most specifications. In light of the somewhat surprising result that the speed of information diffusion is the slowest among the stocks that are not in the lowest but in the second-lowest turnover quintile, it is helpful to remember that turnover is not a perfect proxy for liquidity. In addition to trading costs, turnover is also influenced by the level of investor disagreement. Overall, the results show that information diffuses more quickly among stocks with lower impediments to trade.

C. Cross-sectional regressions

The finding that the leader signal predicts a follower's return in the subsequent month (week) is further confirmed with a set of Fama and MacBeth (1973) cross-sectional regressions. The regression setting allows us to add various control variables that are known to forecast returns in order to confirm that we have identified an independent source of return predicability. (The control variables are described in detail in Appendix A2.) The regression results are presented in Table 13.

In Panel A, regressions are run for the period 1929-2011 (or the period 1930-2011 when 36-month rolling regression windows are used to identify leaders). In addition to the equal-weighted leader signals, we include the following cross-sectional return predictors that are available over the entire sample period: the previous month's stock return; the previous month's industry return; and the stock's momentum return and market capitalization computed at the end of the previous month. The third column also includes the interaction between the previous month's signal and the previous month's stock return. We expect the coefficient on the interaction variable to be negative because if the follower has already reacted to the leaders' news signal in the previous month (and the interaction variable is high), the magnitude of the reaction in the following month would be lower. In all columns but the last, leaders are identified with 12-month rolling regressions. In the last column, leaders are identified with 36-month rolling regressions. In all columns but the next-to-last, the dependent variable is the follower's return. In the next-to-last column, the dependent variable is the follower's return in excess of the contemporaneous value-weighted return of its industry. Columns four to six include only firms that are above the median in size, turnover, and age, respectively.

In all regression specifications and in all subsamples, the coefficient on the aggregate leader signal is highly statistically significant and equal to just over half of the coefficient on the lagged industry return; it varies in magnitude from 0.080 to 0.240. The highest coefficient estimate is obtained for the specification in which leaders are determined with 36-month regression windows, which is not surprising as the range of aggregate leader signals is much narrower for the 36-month specification than for the 12-month specification (compare the second column of Panel A of Table 4 to the second column of Panel A of Table 3). The reported range of the regression coefficients on the leader signal implies that if two stocks have leader signals that are different by, for example, 0.10, then their next-month's returns would differ by between 0.008 to 0.024. The coefficient on the interaction between the leader signal and previous month's return is indeed negative and significant at the 10% level.

Regressions in Panel B include more controls. These regressions are run for a shorter time period, 1963 to 2011, since Compustat variables and daily return data are not available in the earlier period. The last two columns use signals from leaders that are identified with 36-month rolling regressions. The coefficients on the leader signal are somewhat lower than those in the longer sample, but nevertheless highly significant across all specifications. Consistent with the results in Panels E and F of Table 8 that show that signals from leaders identified with 36-month rolling regressions work better in the later part of our sample for equal-weighted portfolios, the t-statistics on these signals are almost twice as high as those on the signals from leaders identified with 12-month rolling regressions. In unreported results, we included a quarterly earnings announcement dummy interacted with the leader signal, hypothesizing that the coefficient on this interaction term should be negative since earnings announcements typically increase the level of investor attention and may additionally reveal the information embedded in the leader signal. As expected, the regression coefficient is negative; however, it is statistically insignificant.[17]

In Panel C, regressions are run for weekly returns over the period 1980 to 2011. However, in the regression specifications that use analyst coverage and news indicators, as described in the footnotes to the panel, the sample period is shorter. All return-based explanatory variables are computed at weekly frequencies, while all other controls are computed as of the end of the previous month. It can be seen that in all regression models, the coefficients on the weekly leader signal are highly statistically significant, and their magnitudes imply that a difference of 0.10 in the weekly leader signal would produce a difference in followers' returns of between 0.03 to 0.07 in the subsequent week.

Next, we include a number of interactions between the weekly leader signal and various control variables (these interaction variables are also included in the regression as independent controls). As in the monthly regression specification, the coefficient on the interaction variable between the weekly leader signal and the follower's prior-week return is negative and significant. The next interaction is with the quarterly earnings announcement dummy that equals one if the follower made a quarterly earnings announcement in the previous week. As in the monthly regression case, we hypothesize that the coefficient on this interaction variable is negative, and this is what we find; here, the interaction term is significant at the 10% level. We are guided by the same logic when including another interaction with a dummy variable that equals one if the TRNA dataset contains a news story with a relevance score of one written about the follower firm in the previous week. The coefficient on this interaction term is also negative but not significant. In the next four regression models, we include interactions between the weekly leader signal and dummy variables indicating relatively high levels of investor attention (we hypothesize that stocks that rank above the median in terms of institutional ownership, analyst coverage, size, and turnover enjoy higher levels of attention

[17]These results are available upon request.

than stocks that rank below the median on these measures). Stocks with higher levels of investor attention may react to leader signals without a one-week delay, and, hence, we expect the coefficients on these interaction terms to be negative. And indeed, all these coefficients are significantly negative. Next, we include an interaction between the leader signal and a dummy variable for whether the follower's firm age is higher than the median firm age. We hypothesize that the predictive ability of the leader signal may not be as high for followers that have been around longer. Though, as anticipated, the coefficient on the interaction is negative, it is insignificant.

In the last regression model, we include, in addition to the weekly signal, a *monthly* leader signal computed at the end of the previous month to check whether or not it has incremental predictive power for a follower's weekly returns. And indeed it does. Controlling for the weekly aggregate leader signal, as well as other characteristics, a spread of 0.10 in the monthly signal computed at the end of the previous month generates an average difference in the next week's returns of almost 0.002.

All these results confirm that the aggregate leader signal possesses robust predictive ability for followers' returns at both monthly and weekly horizons. Moreover, we document that leader signals work best for followers with lower levels of investor attention. Lastly, we find that monthly- and weekly-frequency leaders have an independent predictive ability at weekly return horizons.

IV. Do Sophisticated Investors Trade on Leader Signals?

If sophisticated investors trade on leader signals, one should observe that stocks receiving low signals experience increased short-selling activity. In order to check whether this is the case, we have obtained data from Markit (formerly, Data Explorers), which collects information on total loanable stock inventory, the amount on loan to short sellers, and loan fees (which are calculated as the average of all applicable loan fees weighted by loan value). The data frequency is daily from July 3, 2006, to present; weekly from August 8, 2004, to June 28, 2006; and monthly from June 19, 2002, to July 21, 2004. Since we are interested in short-selling activity in response to the weekly signal, and Markit's weekly-frequency dates do not align with the dates on which the leader signal is calculated, we will only consider the daily-frequency data sample.

Markit claims to capture stock loan trading information on over 85% of the OTC securities lending market; it is worthwhile to note that its universe of reporting participants (custodians and short sellers, from whom it gathers the information on the number of share available for lending, the number of shares borrowed, and lending fees on borrowed shares) is unstable and tends to grow over time. As a result, short interest, which is defined as the number of shares sold short scaled by the number of shares outstanding, would mechanically increase over time if calculated using Markit's data on loaned shares. To avoid this concern, we use utilization as a measure of short-selling activity. Utilization is calculated by Markit and defined as the percentage of the stock inventory available for lending to short sellers that is currently on loan. This measure of short-selling activity is not mechanically determined by the fluctuations in the number of participating short sellers and lenders.

The average utilization over time is plotted in Figure 3. It is shown to exhibit a sharp drop on September 18, 2008. On that date, the short-selling ban on almost 1,000 financial stocks came into effect, as well as the ban on all naked short selling.[18] Even though the ban on short selling of financial stocks was lifted on October 8, 2008, the utilization number did not rebound. (The ban on naked short selling remains in effect.)

In addition to weekly leader signals, short-selling activity is potentially influenced by a number of slower-moving factors, such as momentum or book-to-market characteristics. Since we would like to isolate the effect of weekly leader signals, we construct our regression to

[18]The ban on naked short selling should not affect the utilization numbers given that naked short sellers do not borrow the stock. (The ban on naked short selling on 19 financial firms came into effect on July 21, 2008, and ended on August 12, 2008 (`https://www.sec.gov/rules/other/2008/34-58166.pdf`).)

explain the week-to-week changes in utilization, $\Delta utilization$, and include controls for other potential weekly-frequency drivers of short-selling demand. The variables of interest are the two indicator variables indicating whether the stock enters or exists the bottom weekly leader-signal decile as of Friday of each week. The indicator variable for entering the bottom signal decile is set to zero if the stock was already in the bottom leader signal decile as of Friday of the previous week. Four other control variables are calculated in a similar fashion. These are indicators for whether a stock enters or exits the bottom weekly industry-return decile and for whether the stock enters or exits the top decile of all weekly returns (this indicator is intended to capture the possible short-selling activity aimed to profit from the weekly-frequency return reversal effect). Thus, we run the following Fama-MacBeth regression at a weekly frequency:

$$\begin{aligned}\Delta utilization_{it} = \alpha &+ \beta_1 \cdot \mathbb{1}_{it}\{\text{Enters bottom signal decile}\} + \beta_2 \cdot \mathbb{1}_{it}\{\text{Exits bottom signal decile}\} \\ &+ \gamma_1 \cdot \mathbb{1}_{it}\{\text{Enters bottom ind. ret. decile}\} + \gamma_2 \cdot \mathbb{1}_{it}\{\text{Exits bottom ind. ret. decile}\} \\ &+ \mu_1 \cdot \mathbb{1}_{it}\{\text{Enters top return decile}\} + \mu_2 \cdot \mathbb{1}_{it}\{\text{Exits top return decile}\} + \epsilon_{it}. \quad (3)\end{aligned}$$

In accordance with the SEC's "t+3" rule, all security transactions must be settled within three business days after the transaction day.[19] Since the weekly leader signal is calculated after market close on Friday, short sellers would be able to trade on the signal on Monday of the following week. In accordance with the t+3 rule, shares sold short on Monday must be borrowed and delivered to the buyers by the close of business on Thursday. Therefore, we calculate the difference in utilization between Thursday the comes six days after the Friday when the leader signal was computed and Thursday of the previous week.

At any given time, a small number of stocks have relatively high lending fees. These stocks are said to be "on special."[20] D'Avolio (2002) reports that at any point of time, 91% of all

[19]See `http://www.sec.gov/investor/pubs/tplus3.htm` for the detailed description.

[20]The lending fee is the difference between the interest rate that is typically earned on a cash collateral and the interest rate that the stock's borrower receives on her cash collateral posted for the short sale.

stocks in the loanable universe have lending fees below 1% per annum, while the remaining 9% have fees above 1% per annum, with the lending fee for this set of stocks averaging 4.3% per annum. Reed (2001) estimates that 5.74% of all loans have fees that exceed the prevailing fee levels by at least 1% per annum. Since short sellers may want to avoid stocks that are on special, in one regression specification, we remove stocks with high lending fees. Our version of Markit does not report the actual average loan fee for each stock but rather provides six loan fee buckets, ranging from 0 to 5, with 0 being the cheapest and 5 the most expensive to borrow.[21] As would be expected, utilization rates increase steadily across the loan fee buckets, with the utilization rate averaging to 15.90% for the zero-bucket, and 31.19%, 36.51%, 42.47%, 48.07%, and 59.24% for the next five fee buckets, that is, buckets 1 to 5, respectively. Bucket zero contains 81.78% of stocks in the sample. The next five fee buckets, 1 to 5, contain 6.23%, 2.89%, 2.03%, 2.01%, and 2.78% of the stocks in our sample, respectively.

We modify our sample in the following ways. In order to control for outliers, we trim the dataset at the 1st and 99th percentiles of $\Delta utilization$ on each date. We start the regression sample period after October 8, 2008, the end of the short selling ban on financial stocks that coincides with the start of the ban on naked short selling, which is still in effect. The sample ends on December 31, 2011. Moreover, as in the portfolio results, we drop all stocks priced at less than $5 per share in the 2011 inflation-adjusted dollars. The average utilization in the resulting sample is 19.93%.

We run the regression on three data samples. The first sample contains all observations. In the second sample, we remove stocks that are expected to announce quarterly earnings in the following week. We hypothesize that short sellers may be more reluctant to sell short these stocks because of the high expected return volatility associated with the price reaction to earnings news. Earnings announcement dates are highly predictable by the previous year's earning announcement dates, and we, therefore, construct this sample by dropping stocks

[21] We drop 2.28% of all observations in our sample that have a missing value for the loan fee bucket assigned. The results are nearly unchanged when these observations are kept.

that made quarterly earnings announcements in the same week last year. Finally, in the third sample, we remove all stock-week observations with the average loan fees in the three highest loan fee buckets, thus dropping 6.98% of stocks that are likely to be "on special" according to the estimates of D'Avolio (2002) and Reed (2001). The average utilization in that sample decreases slightly to 17.60%.

The regression results, reported in Table 14, show that short-selling activity indeed increases after a stock enters the bottom leader signal decile: on Monday following the Friday on which the leader signal is computed, utilization goes up by between 0.075% and 0.084%, depending on the sample restrictions. Though these magnitudes may be economically small, they are statistically significant, indicating that the leader signal is one of the inputs that short-sellers use. The short-selling demand, however, does not significantly decrease following a stock exiting the bottom signal decile. This is consistent with the evidence presented earlier in the paper that leader signals continue to forecast followers' returns for up to four weeks into the future.

V. Leadership and News

In this section, we investigate our conjecture that return leadership is associated with noteworthy news developments at the firm level. For that purpose, we again use the TRNA dataset. For this analysis, we limit the set of potential leaders to common stocks with share codes 10 or 11, since our version of the TRNA dataset covers only U.S.-based firms. We use the first year of the TRNA sample to form the first annual cumulative news count, which reduces the regression sample to the period from April 1997 to December 2011.

The distribution of the number of followers for each stock in our dataset, which is computed using only end-of-year observations and which also includes stocks with zero followers, is plotted in Figure 4. In Panel A, the monthly leadership specification with 12-month rolling regressions is used, and in Panel B, the weekly leadership specification with 52-week rolling regressions is used. Since the number of followers is a count variable, the distributions are non-negative and right-skewed. The requirement that a potential follower traded on the last day of the week, which we impose at weekly frequencies, eliminates more stocks than the requirement that a potential follower traded on the last day of the month, which we impose at monthly frequencies. The average and the median number of followers in Panel A (357.2 and 329, respectively) is, hence, greater than those in Panel B (299.9 and 269, respectively). When computed over the entire sample, the average number of leaders should equal the average number of followers. But the average number of followers reported in Figure 4 is larger than the average number of leaders reported in Table 1. The primary reason is that Table 1 covers the entire sample period, with a lower average number of stocks in the cross section. Additionally, the set of followers in Table 1 does not include stocks priced below $5 per share, which are relatively illiquid and, therefore, should have more large-cap leaders than an average stock. As a result, the average leader count for that sample should be lower than for the entire sample.

We run monthly contemporaneous regressions of the number of followers that a firm has on a set of firm attributes in order to assess whether a firm's capacity to lead is related to the intensity of its news coverage after controlling for other firm characteristics. We estimate our regressions using quasi-maximum likelihood, which is appropriate for a count variable; this estimation method produces consistent and asymptotically normal coefficient estimates even if the underlying distribution is not Poisson (Wooldridge (2002)).[22] Since leadership is determined over a one-year window, we use rolling one-year averages of all explanatory variables; for news, we calculate rolling *total* news counts over the previous year. The regressions are run at a monthly frequency, and the standard errors are clustered by firm.[23] Because news coverage increases over the years and may be uneven across industries, we also include year and industry dummies.

[22]In our case, the underlying distribution is not Poisson because its variance is significantly larger than its mean. We experimented with assuming that the underlying distribution is negative gamma and obtained qualitatively similar results.

[23]Clustered OLS regressions produce qualitatively similar results.

The regression results are reported in Table 15. Pairwise correlations between the control variables are reported in Panel C. We use two measures of news. In regression models (1)-(3), $News$ is the count of all "highly relevant news," or news with a relevance score of one. In regression models (4)-(6), $News$ is the count of only "highly relevant corporate news," or news with a relevance score of one that report on new corporate developments as opposed to reporting on trade order imbalances, expressing opinions about the firm, etc. (more details are provided in Appendix A3). For firms with no news stories in a particular category over the previous year or for firms not covered by the TRNA dataset, we set $News = 0$ (we include the latter set of firms because it is still potentially on the Reuters' radar screen). In Panel A, leaders are identified with monthly regressions, and in Panel B, with weekly regressions. It can be seen that the variable $News$ is highly significant in both panels. However, when the variable $News^2$ is included in the regression, the coefficient on that variable is significantly negative, indicating non-linearity: Even though leaders tend to have more news stories written about them, those firms that receive very intensive news coverage start to drop followers, as followers' prices begin to react to the leaders' news with shorter delays.

The economic interpretation of the regression coefficients on news is as follows. In regression models (1) and (2) of Panel A, the coefficient on news ranges from 0.0056×10^{-2} to 0.0075×10^{-2}. This implies that when a firm moves from the 5th to the 95th percentile of news coverage, or from 0 to 232 highly relevant news items per year (see the last row of Table A3), its number of followers increases by between 1.3% and 1.7%, which amounts to between 4 and 6 additional followers for a median stock (see the box in Figure 4). Redoing these calculations, but now taking the squared news term of model (3) into account, produces an increase of 8 followers. These magnitudes are very similar for models (4)-(6) of Panel A. For regression models (1)-(3) of Panel B, analogous calculations imply a gain of between 2 and 4 followers, and for models (4)-(6), a gain of between 3 and 5 followers. These economic magnitudes are not large, but they could perhaps be increased with a more careful analysis of the contents of the news articles.

The results for the control variables show that, for both monthly and weekly leadership specifications, stocks with high institutional ownership and high analyst coverage tend to have significantly more followers. This result is consistent with the findings of the lead-lag literature that stocks with higher levels of attention react faster to common shocks; hence, they would appear to have more followers. Institutional investors are sophisticated and react to new common information faster than retail investors, while analysts help uncover and publicize the relevant news developments reported by other firms.

Overall, these results indicate that the scope of a firm's leadership is indeed positively related to news developments at the firm level. A more precise news classification could provide better insights into the news categories that matter most for other firms. Such an analysis is, however, beyond the scope of this paper.

VI. Conclusion

This paper documents the existence of a collection of "bellwether" stocks for many individual stocks that reliably predict their followers' returns. We argue that leaders lead the followers' returns along some valuation-relevant dimension and that leaders are not always easily identifiable with ex-ante stock characteristics. Some leader-follower relations may be short-lived or driven by common sentiment rather than economic links and hence not straightforwardly detected. Our methodology of identifying leaders with Granger-causality regressions offers a convenient way to detect delayed information transmission without the need to conduct in-depth research on news developments or on inter-firm links. The return predictability documented here is unlikely to be explained by data snooping. The magnitude of the anomaly diminishes over time, and it is larger for less liquid stocks and stocks with lower levels of investor attention. Moreover, when we scramble the data along the time dimension and identify (false) leaders using the scrambled dataset, these false leaders do not have any return forecasting ability. Furthermore, we show that short selling responds to low leader signals. Finally, we present support for our conjecture that information leaders are likely to be at the center of significant news developments by documenting that a firm's number of followers significantly increases with the number of news articles written about it.

While leaders' predictive ability for followers' returns weakened in recent years at a one-month lag, it continues to be strong at a one-week lag. However, the analysis of break-even trading costs shows that the amount of money that can be profitably invested is small, suggesting that the market, though not perfectly efficient, is competitive, as it competes away most arbitrage opportunities.

Our methodology can be extended and improved along various dimensions. It can be applied to daily- and higher-frequency returns. Switching to higher frequencies would allow to estimate the lead-lag relation with a higher precision and over shorter estimation windows, making it possible to identify very short-lived leader-follower pairs. (However, more frequent trading will entail significantly higher trading costs.) Additionally, more return lags can be introduced in the leadership regression model to best capture the length of the delay in the price reaction of the followers. For our analysis, we use very simple methods for aggregating leaders' signals and show that they work well for forecasting followers' returns. However, in practical applications, the signal aggregation method can be improved by optimally under-weighting returns from a set of leaders with correlated returns and a set of leaders with low forecasting ability, as well as by discarding leaders that are likely misidentified.

Another interesting extension of this analysis would be to aggregate potential followers up to a category index, such as an industry index, a characteristic-based index, or a corporate-image index (e.g., an index for "growth," "value," "dividend paying," "non-dividend paying," "green," "socially responsible," or "well-governed" firms), or even a market-wide index, in order to investigate what types of leaders predict movements of entire stock categories. Hou, Scherbina, Tang, and Wilhelm (2012) do just that for industry indices.[24]

While we have shown here that news about firm fundamentals significantly increase the scope of a firm's leadership, we did not exclude the possibility that some firm-level news may affect prices of other firms mainly through their impact on investor sentiment about the broader stock category. These news are likely to be of the salient type, as described by Shiller (2002). Hence, an in-depth analysis of leaders' news coverage may engender a better understanding of the drivers of investor sentiment.[25]

References

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

Anderson, Robert M., Kyong Shik Eom, Sang Buhm Hahn, and Jong-Ho Park, 2012, Stock return autocorrelation is not spurious, Working paper.

Ang, Andrew, Robert Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Badrinath, S. G., Jayant R. Kale, and Thomas H. Noe, 1995, Of shepherds, sheep, and the cross-autocorrelations in equity returns, *Review of Financial Studies* 8, 401–30.

Brennan, Michael J, Narasimhan Jegadeesh, and Bhaskaran Swaminathan, 1993, Investment Analysis and the Adjustment of Stock Prices to Common Information, *Review of Financial Studies* 6, 799–824.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Chordia, Tarun, and Bhaskaran Swaminathan, 2000, Trading Volume and Cross-Autocorrelations in Stock Returns, *Journal of Finance* 55, 913–935.

Cohen, Lauren, and Andrea Frazzini, 2008, Economic links and predictable returns, *Journal of Finance* 63, 1977–2011.

[24]Using daily return frequencies and five return lags, the authors identify stocks that Granger-cause the returns of their industry's index. Thus identified leaders reliably predict returns of other stocks in the industry, even after including firm- and industry-level controls.

[25]See, e.g., Tetlock (2007), who shows that news media can indeed affect investor sentiment about the stock market.

Cohen, Lauren, and Dong Lou, 2012, Complicated firms, *Journal of Financial Economics* 104, 383–400.

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 46, 427–466.

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

Fama, Eugene F., and Kenneth R. French, 2000, Characteristics, covariances, and average returns: 1929-1997, *Journal of Finance* 55, 389–406.

Fama, Eugene F., and James MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 51, 55–84.

Hong, Harrison, Terence Lim, and Jeremy Stein, 2000, Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies, *Journal of Finance* 55, 265–295.

Hong, Harrison, Walter Torous, and Rossen Valkanov, 2007, Do industries lead stock markets?, *Journal of Financial Economics* 83, 367–396.

Hou, Kewei, 2007, Industry information diffusion and the lead-lag effect in stock returns, *Review of Financial Studies* 20, 1113–1138.

Hou, Kewei, Anna Scherbina, Yi Tang, and Stefan Wilhelm, 2012, Industry information leaders, Working paper.

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

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Lo, Andrew W ., 2004, The adaptive markets hypothesis, *Journal of Portfolio Management* 30, 15–29.

Lo, Andrew W, and A Craig MacKinlay, 1990, When Are Contrarian Profits Due to Stock Market Overreaction?, *Review of Financial Studies* 3, 175–205.

Menzly, Lior, and Oguzhan Ozbas, 2010, Market segmentation and cross-predictability of returns, *Journal of Finance* 65, 1555–1580.

Neuhierl, Andreas, Anna Scherbina, and Bernd Schlusche, 2013, Market reaction to corporate press releases, *Journal of Financial and Quantitative Analysis* 48, 1207–1240.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive-definite, heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Reed, Adam, 2001, Costly short-selling and stock price adjustment to earnings announcements, Working paper.

Sadka, Ronnie, and Anna Scherbina, 2007, Analyst disagreement, mispricing and liquidity, *Journal of Finance* 62, 2367–2403.

Shiller, Robert J., 2002, Bubbles, human judgment, and expert opinion, *Financial Analysts Journal* 58, 18–26.

Shleifer, Andrei, and Robert Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.

Shumway, Tyler, 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327–340.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

Wooldridge, Jeffrey M., 2002, *Econometric Analysis of Cross Section and Panel Data*. (MIT Press).

Appendix

A1. Data

The data used in this paper are obtained from CRSP monthly and daily files and include all NYSE-, Amex-, and Nasdaq-traded stocks from the CRSP dataset, covering the period from January 1926 to December 2011. We adjust stock returns for delisting in order to avoid survivorship bias (Shumway (1997)).[26]

We do not impose any restrictions on the sample of stocks that are eligible to be identified as leaders. Over the January 1929 to December 2011 period, our sample of potential leaders on average consists of about 3,305 stocks per month. However, we require that the set of follower stocks consist of common shares of U.S.-incorporated firms, or stocks with shares codes 10 or 11. Moreover, we require that these stocks have a trade on the last day of the previous month for the monthly-frequency analysis and on the last day of the previous week for the weekly-frequency analysis. For the portfolio results, we further require that followers be priced above $5 per share in 2011 inflation-adjusted dollars.[27] These restrictions leave us with an average of about 2,175 stocks per month that are eligible to be identified as followers. For the 1929-1960 subsample, this number is 694, and for the 1961-2011 subsample, it is 3,104.

Accounting variables are obtained from the Merged CRSP/Computstat dataset. The tables and figures presented throughout the paper generally cover the period January 1929 to December 2011 (the initial years are used to estimate leadership regressions). However, some variables, such as accounting variables or those calculated using daily return data, are not available for the early part of the sample. Data on analyst coverage are obtains from the I/B/E/S dataset and data on institutional holdings, from the Thompson-Reuters dataset.

[26] Specifically, when a stock is delisted, we use the delisting return from CRSP, if available. Otherwise, we assume the delisting return to be -100%, unless the reason for delisting is coded as 500 (reason unavailable), 520 (went to OTC), 551-573, 580 (various reasons), 574 (bankruptcy), or 584 (does not meet exchange financial guidelines). For these observations, we assume that the delisting return is -30%.

[27] Monthly inflation numbers are obtained from Amit Goyal's web site (`http://www.hec.unil.ch/agoyal`).

The news coverage data are available from the Thompson-Reuters News Analytics (TRNA) dataset for the period April 1996 to December 2011.

Monthly and weekly factor returns and industry classifications are obtained from Kenneth French's web site.[28] The results presented in the paper use 38 industry classifications, but the results are almost unchanged when 12 industry classifications are used instead. The monthly average percentages of firms in our sample per each industry are provided in Table A1. The industry classified as "Irrigation Systems" drops out of our sample after the data restrictions are imposed, reducing the number of industries to 37. Additionally, in the results in which portfolio sorts are performed within industries or in which leaders are required to belong to a different industry than their followers, we drop stocks in the industry identified as "Other" because of the implied heterogeneity (however, as can be seen from the table, this industry has very few stocks).

A2. Variable definition and estimations

This appendix provides detailed descriptions of the variables used in our cross sectional regressions. Unless specified otherwise, all variables are calculated at the month end. Weekly-frequency variables are computed analogously.

Amihud's illiquidity measure *(Illiq)*. Following Amihud (2002), we measure illiquidity for each stock in month t as the average daily ratio of the absolute stock return to the dollar trading volume within the day:

$$Illiq_{i,t} = \text{Avg}_t \left[\frac{|R_{i,d}|}{Volume_{i,d}} \right], \tag{4}$$

where $R_{i,d}$ is the return and $Volume_{i,d}$ is the dollar trading volume for stock i on day d.

Analyst Coverage *(Analyst Cov.)* is defined as the number of analysts issuing annual earnings forecasts for the current fiscal year, computed using the I/B/E/S dataset.

[28] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_48_ind_port.html`.

Beta *(Beta)*. Following Fama and French (1992), the market beta of individual stocks is estimated by running a time-series regression based on the monthly return observations over the prior 60 months if available (or a minimum of 24 months):

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i^1 \left(R_{m,t} - R_{f,t}\right) + \beta_i^2 \left(R_{m,t-1} - R_{f,t-1}\right) + \epsilon_{i,t}, \tag{5}$$

where the market beta of stock i is the sum of the slope coefficients on the current and lagged excess market returns; i.e., $Beta = \hat{\beta}_i^1 + \hat{\beta}_i^2$.

Book-to-market ratio *(Book/Market)*. Following Fama and French (1992, 1993, and 2000), the book-to-market equity ratio is computed at the end of June of each year as the book value of stockholders' equity, plus deferred taxes and investment tax credit (if available), minus the book value of preferred stock, scaled by the market value of equity. Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the book value of preferred stock for the last fiscal-year end. The market value of equity is the product of share price and the number of shares outstanding at the end of December of the previous fiscal year.

Firm age *(Age)* is the number of months since the firm's IPO.

Idiosyncratic volatility *(IVOL)*. Following Ang, Hodrick, Xing, and Zhang (2006), we estimate idiosyncratic volatility of stock i each month as the standard deviation of the daily regression residuals, $\epsilon_{i,d}$, within a month. Specifically, the regression residuals are obtained from the following regression run every month with daily returns:

$$R_{i,d} - R_{f,d} = \alpha_i + \beta_i(R_{m,d} - R_{f,d}) + \eta_i \mathrm{SMB}_d + \delta_i \mathrm{HML}_d + \epsilon_{i,d}, \tag{6}$$

where $R_{i,d}$ is the daily return on stock i on day d, $R_{f,d}$ is the risk-free return (proxied by the return on a one-month T-bill), $R_{m,d}$ is the daily return on the market portfolio (proxied by the return on the CRSP value-weighted index), and SMB_d and HML_d are the daily returns on the size and book-to-market factors. We then convert the idiosyncratic volatility of each stock into a monthly measure by multiplying the estimate by the number of trading days in the month: $IVOL_{i,t} = st.dev._t(\epsilon_{i,d}) \times$ no. of trading days. At least 15 daily return observations in a month are required to estimate IVOL.

Institutional Ownership *(Inst. Ownership)* is defined as the percentage of total shares outstanding owned by institutions, computed using the data in the Institutional Holdings (13F) dataset.

Last month's return *(Ret_{t-1})*. Following Jegadeesh (1990), this short-term reversal predictor is defined as the stock return over the previous month.

Momentum return *(Momentum)*. Following Jegadeesh and Titman (1993), momentum is defined as the cumulative return of a stock over a period of from beginning of month $t-13$ to end of month $t-2$.

Previous month's industry return *(Ind. Ret_{t-1})* is defined as the value-weighted industry return over the previous month.

Size *(Size)*. A stock's size is defined as the product of the price per share and the number of shares outstanding, expressed in thousands of dollars.

Turnover *(Turn)* is the monthly turnover, scaled by the end-of-month number of shares outstanding.

A3. The Thomson-Reuters News Analytics Dataset

The Thomson-Reuters News Analytics dataset (TRNA) is a machine-readable news feed from Thomson Reuters that includes news items from 41 news media outlets and covers the period from April 1996 to December 2011. In addition to news headlines, TRNA contains a variety of quantitative scores for the news computed by Thomson-Reuters, including sentiment (indicating whether the story is positive, negative, or neutral), relevance (measuring how relevant the story is to the firm), and uniqueness (specifying how new or repetitive the story is). In this paper, we use only the relevance score, which ranges between 0 and 1; its usefulness is illustrated in the sample news feed for Cisco Systems Inc. reproduced in Table A2. While stories that concern exclusively Cisco Systems are assigned the maximum relevance score of 1, the last news topic, headlined "Top players in ailing mobile network gear market," is applicable to several firms, and, hence, its relevance to Cisco Systems is calculated to be only about 0.11. For the purposes of this paper, we consider only highly relevant news, which we define to be those with the relevance score of one.

News stories that are deemed to be important or highly anticipated typically start with a sequence of alerts informing readers in the headline of the topic of the forthcoming article while the article is being written. Once the article is posted, it may get updated, appended, overwritten, or corrected. Consequently, news items posted in TRNA are classified as either "Alerts," "Articles," "Appends," or "Overwrites." All news items are further tagged with a news topic code.

A distinct news story for a given firm is tagged by a unique identifier, the primary news access code (PNAC). This identifier allows the reader to keep track of a particular story unfolding with a series of alerts and follow-up reports. Arguably, the more complex or significant the news development, the more items would appear under its assigned PNAC. Unique news counts count each new PNAC as one news item, ignoring possible multiple Alerts, Appends, Overwrites, and Articles that may be contained within one PNAC.

We form a news count for all news items that are written about new corporate developments, as opposed to news that simply mentions a firm in the context of a market or industry review, a report about trade order imbalance, etc. This news count contains stories about corporate bonds, equities, dividends, annual reports, forecasts and estimates of future earnings, corporate insolvencies and bankruptcies, decisions to raise or return capital, strategic decisions, merges, acquisitions, and so on.[29]

Table A3 provides sample statistics for total annual news counts for each of the years 1996-2011 and for the overall sample. The news counts reported are simple count, count of unique news topics (PNACs), count of news with a relevance score of one, and count of corporate news with a relevance score of one.

[29]Specifically, for this count we consider only news tagged with the following topic codes: 'AAA', 'ALLCE', 'BACT', 'BKRT', 'BOSS1', 'BUYB', 'CASE1', 'CLASS', 'COVB', 'CM1', 'DEAL1', 'DIV', 'DVST', 'FIND1', 'FINE1', 'INDX', 'IPO', 'ISU', 'JOB', 'LIST1', 'MEET1', 'MNGISS', 'MONOP', 'MGR', 'NG1', 'NT1', 'PS1', 'RCH', 'REGS', 'RES', 'RESF', 'SL1', 'STAT', 'STK', 'ENV', 'FAKE1', 'ACB', 'CORPD', 'DBT', 'FUND', 'PVE', 'USC', 'INVB', 'INVD', 'INVI', 'INVM', 'INVS', 'INVT', 'ABS', 'LOA', 'BNK', 'CMPNY', 'INV', 'TAX', 'LAW', 'JUDIC', 'FIN', 'FINS', 'FRAUD1', 'DAT', 'CIV', 'CLJ', 'EQB', 'CDM', 'CDV', 'CORPD', and 'DBT'.

Example: Leader stocks B and C for follower stock A

Regression estimated at τ: $Ret_t^A = b_0^{Aj} + b_1^{Aj} Ret_{t-1}^{mkt} + b_2^{Aj} Ret_{t-1}^A + b_3^{Aj} Ret_{t-1}^j + \epsilon_t^{Aj}$, $j = \{B, C\}$
Estimates: $\hat{b}_3^{AB} = 1$ and $\hat{b}_3^{AC} = 1$
Leader returns: $Ret_\tau^B = 1\%$, $Ret_\tau^C = 3\%$
Leader signal: $Signal_\tau^A = \frac{1}{2}(1 \cdot 1\% + 1 \cdot 3\%) = 2\%$

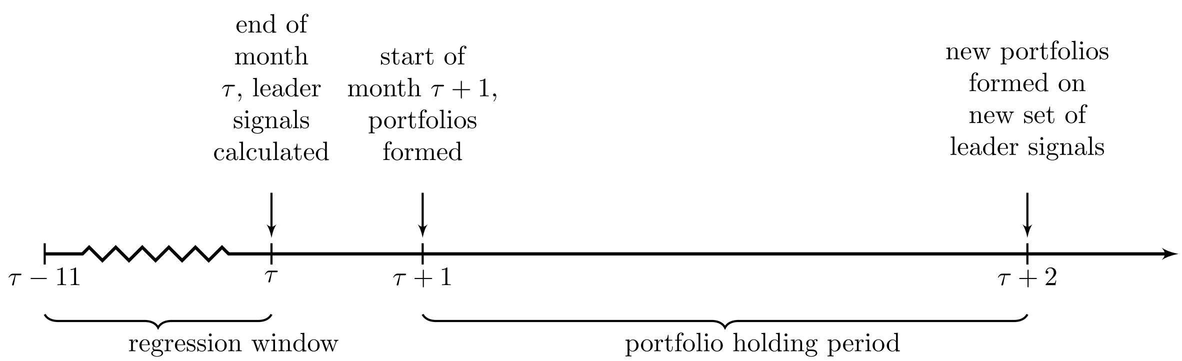
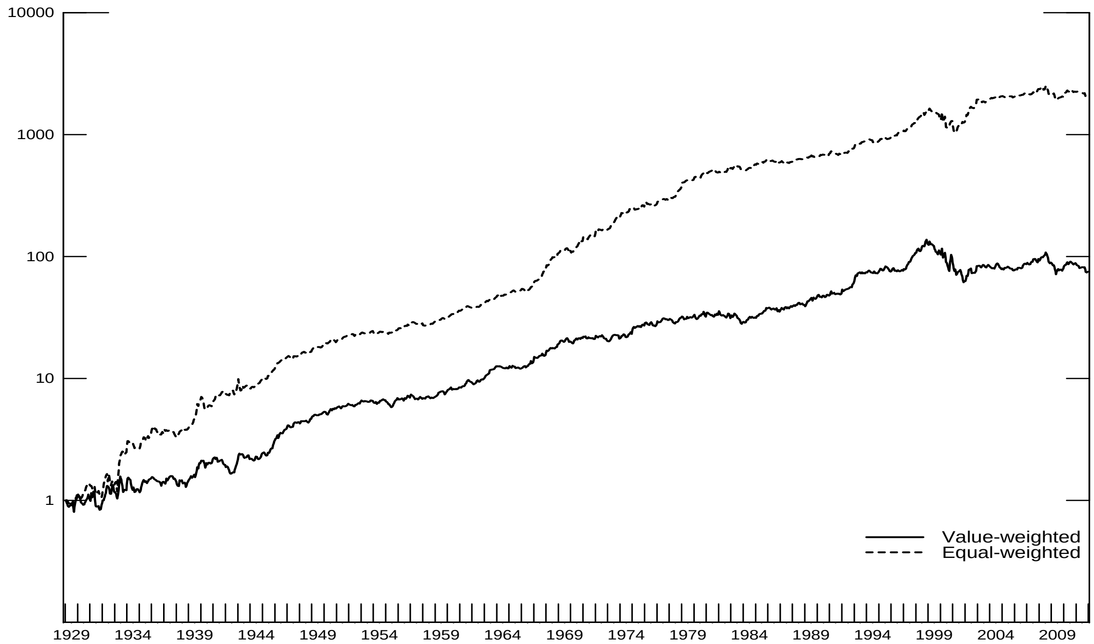


Figure 1. Timeline. This figure presents the timeline for our computations and an example for how an aggregate leader signal is computed.

Panel A: Monthly portfolios



Panel B: Weekly portfolios

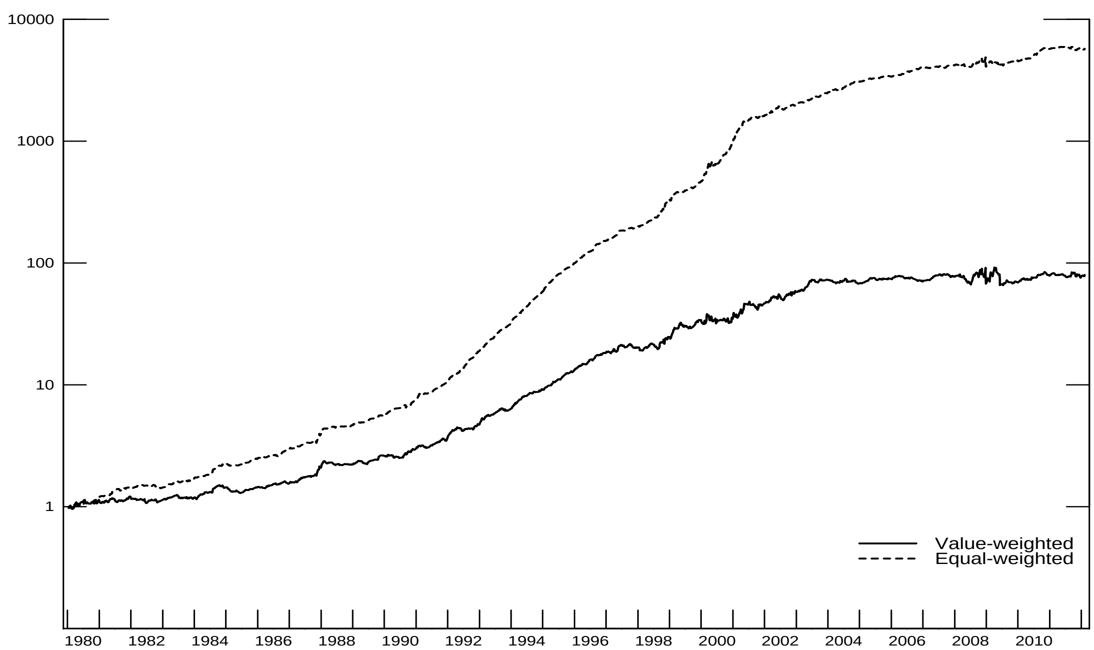


Figure 2. Cumulative returns. The charts plot, for equal- and value-weighted portfolios, the value of $1 invested in the beginning of the period at the return earned on a zero-investment strategy of buying stocks in the top and selling short stocks in the bottom leader signal decile. In Panel A, leaders are calculated with monthly regressions and portfolios are formed monthly. In Panel B, leaders are calculated with weekly regressions and portfolios are formed weekly. The axes are in log-scale. The time periods are February 28, 1929, to December 31, 2011, and January 18, 1980, to December 30, 2011, respectively.

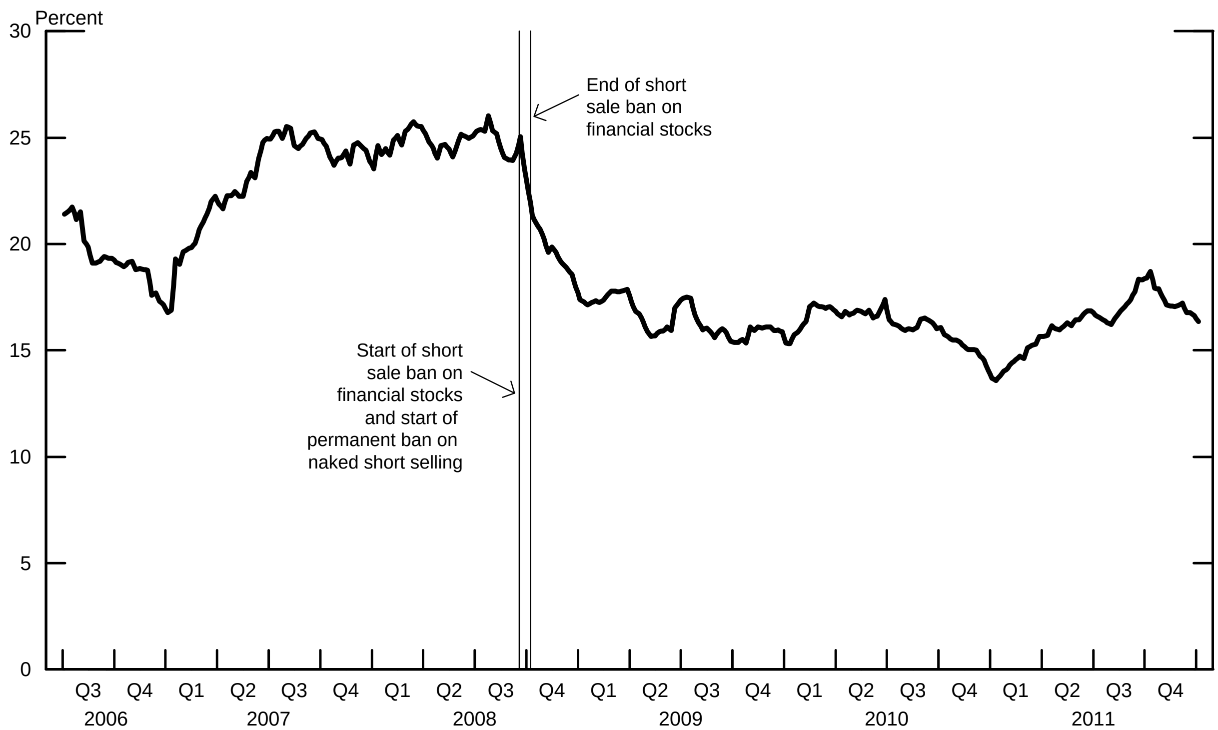


Figure 3. Short-selling activity over time: The average utilization. The figure plots the average utilization across all stocks in the Markit universe. Utilization is defined as the number of shares on loan to short sellers divided by the number of shares available to be loaned out. The sample period is July 3, 2006, to December 31, 2011.

Panel A: Monthly-frequency leaders

Panel B: Weekly-frequency leaders

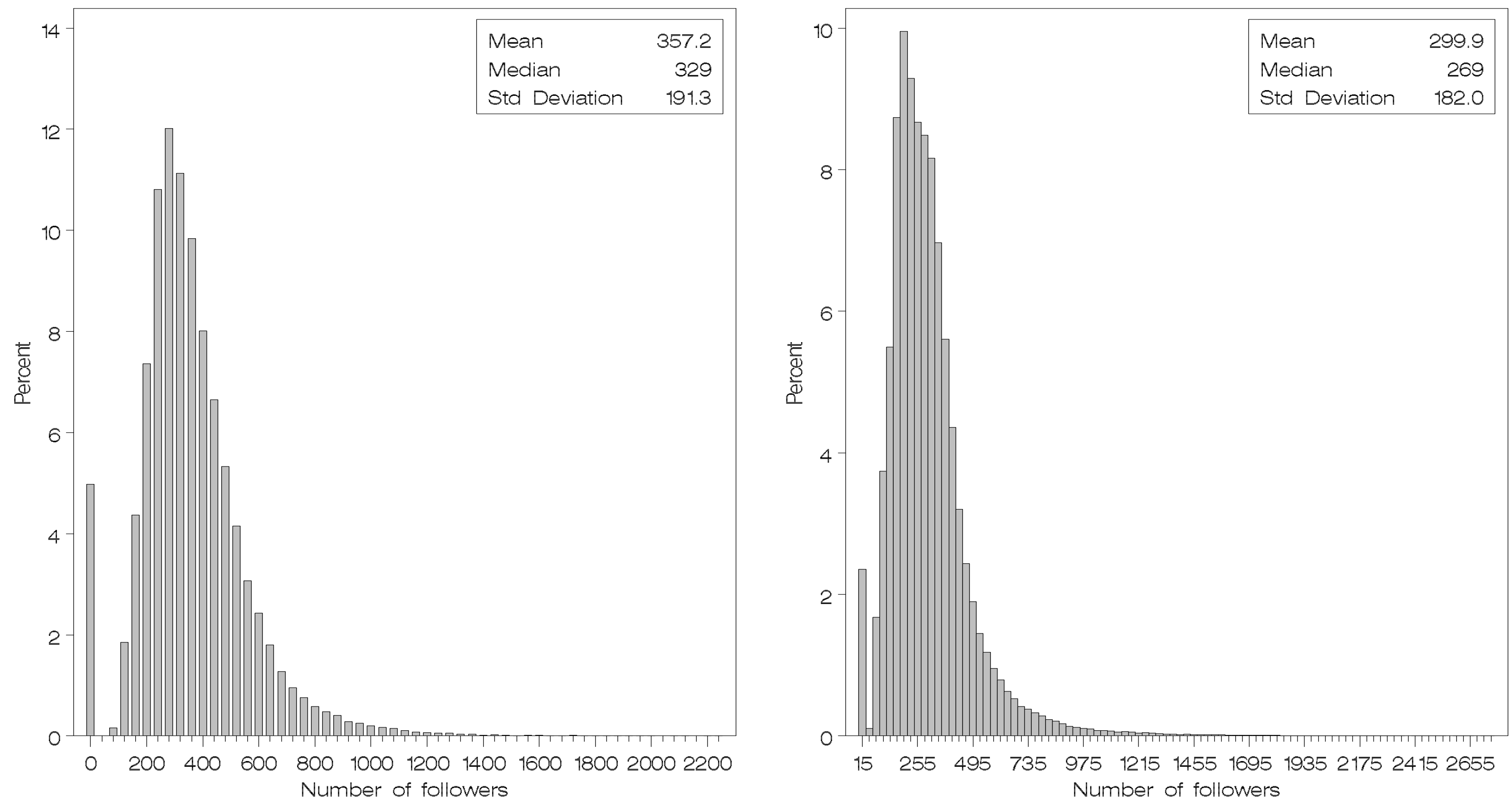


Figure 4. Distribution of the number of followers. The charts plot the distributions of the number of followers computed across all common shares of U.S.-incorporated firms (CRSP stocks with share codes 10 or 11). Leaders are the stocks that are found to Granger-cause monthly (weekly) returns of their followers in one-year monthly-frequency (weekly-frequency) rolling regressions. The numbers of followers are sampled at the end of January of each year. The sample period for both plots is April 1997 to December 2011, corresponding to the sample period of Table 15.

Table 1
Descriptive statistics on followers

This table presents characteristics of follower stocks. Followers are stocks whose returns were shown to be Granger-caused by their leaders' returns, as described in the text. The set of possible followers is limited to stocks that traded on the last day of the previous month and were priced above \$5 per share in 2011 inflation-adjusted dollars. Both leaders and followers are limited to common shares of U.S.-incorporated firms. The statistics are calculated as of January 31 of each year. The sample period is 1929-2011.

The entire sample

Average number of leaders (including observations with no leaders)	286.89
Fraction of stock-month observations with at least one leader	84.00%

The sample limited to stocks with existing leaders (followers)

Fraction of leaders that are positive leaders	53.03%
Average regression coefficient on a positive leader's lagged return	0.89
Average regression coefficient on a negative leader's lagged return	-0.91
Average fraction of a followers' leaders in the same industry, using 12 ind. classifications[†]	
– positive leaders	15.28%
– negative leaders	13.77%
Average fraction of a followers' leaders in the same industry, using 38 ind. classifications[†]	
– positive leaders	8.25%
– negative leaders	7.27%
Fraction of followers larger than its average leader	22.07%
Fraction of followers with greater turnover than its average leader	37.94%
Fraction of followers older than its average leader	44.94%

[‡]The industry classification "Other" is excluded.

	Number of leaders				
	1 (low)	2	3	4	5 (high)
Avg. number of leaders in quintile	234.37	268.14	299.82	338.71	427.90
Market capitalization (in \$, mil)	158.81	154.48	153.19	154.74	138.79
Share turnover over the past 12 months	1.03	1.00	0.99	0.99	1.00
Firm age (in years)	16.86	17.00	17.00	16.66	16.08

Table 2
Persistence of leadership

This table presents probabilities, and the associated standard errors, of a leader-follower pair in January of year t having also been identified as a leader-follower pair of the same sign in January of year $t-\tau$, $\tau \in \{1, ..., 10\}$, provided that both stocks were present in the CRSP dataset for 12 months (Panel A) or 36 months (Panel B) prior to January 31 of year $t-\tau$. Panel A presents the results for leaders identified using 12-month rolling regression windows, and Panel B, using 36-month rolling regression windows.

Panel A: Leaders are identified using 12-month rolling regression windows

	All leaders			Positive leaders			Negative leaders		
Number of years prior (τ)	prob.	excess prob. relative to $\tau = 10$	std. err. of excess prob.	prob.	excess prob. relative to $\tau = 10$	std. err. of excess prob.	prob.	excess prob. relative to $\tau = 10$	std. err. of excess prob.
1	5.982%	2.565%	0.006%	3.526%	1.578%	0.004%	2.457%	0.986%	0.004%
2	3.538%	0.120%	0.005%	2.051%	0.104%	0.004%	1.487%	0.017%	0.004%
3	3.468%	0.051%	0.005%	2.017%	0.070%	0.004%	1.451%	-0.019%	0.004%
4	3.461%	0.044%	0.006%	2.010%	0.063%	0.004%	1.451%	-0.019%	0.004%
5	3.416%	-0.001%	0.006%	1.984%	0.037%	0.004%	1.432%	-0.038%	0.004%
...									
10	3.417%	0.000%	-	1.947%	0.000%	-	1.470%	0.000%	-

Panel B: Leaders are identified using 36-month rolling regression windows

	All leaders			Positive leaders			Negative leaders		
Number of years prior (τ)	prob.	excess prob. relative to $\tau = 10$	std. err. of excess prob.	prob.	excess prob. relative to $\tau = 10$	std. err. of excess prob.	prob.	excess prob. relative to $\tau = 10$	std. err. of excess prob.
1	30.244%	27.366%	0.009%	18.321%	16.461%	0.007%	11.923%	10.904%	0.006%
2	12.654%	9.776%	0.008%	7.936%	6.076%	0.006%	4.717%	3.698%	0.005%
3	4.416%	1.538%	0.006%	2.878%	1.018%	0.005%	1.538%	0.519%	0.004%
4	2.944%	0.066%	0.006%	1.922%	0.062%	0.005%	1.022%	0.003%	0.004%
5	2.912%	0.034%	0.006%	1.897%	0.037%	0.005%	1.015%	-0.004%	0.004%
...									
10	2.878%	0.000%	-	1.860%	0.000%	-	1.019%	0.000%	-

Table 3

Portfolios sorted on the equal-weighted leader signal within 36 industries, 1929-2011

This table presents monthly abnormal returns of leader-signal-sorted portfolios. Leaders for each stock are identified using 12-month rolling regressions, as described in the text. At the beginning of each month, all stocks that traded on the last day of the prior month, that were priced above $5 per share in 2011 inflation-adjusted dollars, and that had leader stocks are sorted into decile portfolios within each of the 36 industries based on the last month's equal-weighted aggregate leader signal, computed as described in the text. Portfolio returns are equal-weighted in Panel A and value-weighted in Panel B. The second column reports the weighted-average leader signal, which is equal-weighted across followers in each portfolio in Panel A and value-weighted across followers in Panel B; the third column reports the average portfolio return in excess of the risk-free rate; the fourth column reports the market alpha; the fifth column reports the alpha of the Fama and French (1993) three-factor model; and the sixth column reports the alpha of the four-factor model that also includes the Carhart (1997) momentum factor. The last row reports the return differential between the high- and the low-signal portfolios (deciles 10 and 1). Panels C and D report the four-factor model factor loadings for the equal- and value-weighted portfolios, respectively. Newey-West-adjusted t-statistics are reported in parentheses. Panel E reports portfolio transition probabilities.

Panel A: Equal-weighted portfolios

Decile	Leader signal	Excess return	Market alpha	3-factor alpha	4-factor alpha
1	-3.58%	0.52%	-0.27%	-0.47%	-0.37%
		(1.93)	(-2.25)	(-5.81)	(-4.61)
2	-1.99%	0.71%	0.00%	-0.16%	-0.10%
		(3.02)	(0.03)	(-2.78)	(-1.62)
3	-1.26%	0.80%	0.10%	-0.08%	0.02%
		(3.44)	(1.13)	(-1.75)	(0.44)
4	-0.72%	0.87%	0.19%	0.02%	0.10%
		(3.89)	(2.40)	(0.47)	(2.37)
5	-0.28%	0.86%	0.20%	0.04%	0.09%
		(3.97)	(2.61)	(0.94)	(1.81)
6	0.14%	1.01%	0.33%	0.15%	0.22%
		(4.50)	(4.05)	(3.32)	(4.71)
7	0.63%	1.05%	0.35%	0.14%	0.23%
		(4.49)	(3.82)	(2.89)	(4.69)
8	1.14%	1.13%	0.40%	0.18%	0.26%
		(4.62)	(4.06)	(3.53)	(5.14)
9	1.88%	1.21%	0.45%	0.21%	0.29%
		(4.68)	(4.07)	(3.62)	(5.04)
10	3.54%	1.35%	0.52%	0.25%	0.27%
		(4.68)	(3.82)	(3.29)	(3.52)
10-1		0.83%	0.79%	0.71%	0.64%
		(7.35)	(7.03)	(6.48)	(5.73)

Panel B: Value-weighted portfolios

Decile	Leader signal	Excess return	Market alpha	3-factor alpha	4-factor alpha
1	-2.85%	0.34%	-0.36%	-0.39%	-0.34%
		(1.46)	(-4.37)	(-4.67)	(-4.07)
2	-1.70%	0.49%	-0.12%	-0.13%	-0.10%
		(2.50)	(-2.02)	(-2.14)	(-1.68)
3	-1.09%	0.53%	-0.04%	-0.05%	-0.02%
		(2.88)	(-0.85)	(-0.88)	(-0.29)
4	-0.65%	0.60%	0.04%	0.05%	0.03%
		(3.34)	(0.80)	(0.93)	(0.61)
5	-0.25%	0.56%	-0.01%	-0.01%	0.01%
		(3.08)	(-0.20)	(-0.31)	(0.29)
6	0.10%	0.61%	0.07%	0.06%	0.06%
		(3.55)	(1.60)	(1.28)	(1.37)
7	0.51%	0.62%	0.06%	0.06%	0.06%
		(3.45)	(1.18)	(1.25)	(1.15)
8	0.99%	0.71%	0.12%	0.08%	0.07%
		(3.71)	(1.98)	(1.28)	(1.17)
9	1.58%	0.77%	0.14%	0.09%	0.10%
		(3.77)	(2.21)	(1.45)	(1.67)
10	2.81%	0.85%	0.16%	0.07%	0.04%
		(3.68)	(1.76)	(0.83)	(0.48)
10-1		0.52%	0.52%	0.45%	0.38%
		(4.08)	(4.09)	(3.60)	(2.98)

Panel C: Factor loadings for the equal-weighted portfolios

Decile	alpha	MKT	SMB	HML	UMD	R^2
1	-0.37%	1.16	0.83	0.22	-0.092	91.65%
	(-4.61)	(72.58)	(33.28)	(9.33)	(-4.98)	
2	-0.10%	1.08	0.60	0.23	-0.066	94.14%
	(-1.62)	(91.84)	(33.14)	(13.32)	(-4.94)	
3	0.02%	1.04	0.62	0.26	-0.104	96.28%
	(0.44)	(113.7)	(43.27)	(18.80)	(-9.89)	
4	0.10%	1.03	0.54	0.26	-0.084	96.42%
	(2.37)	(118.6)	(40.13)	(19.84)	(-8.43)	
5	0.09%	1.00	0.51	0.25	-0.043	95.24%
	(1.81)	(104.0)	(33.86)	(17.51)	(-3.93)	
6	0.22%	1.02	0.56	0.29	-0.066	96.06%
	(4.71)	(112.1)	(39.27)	(21.50)	(-6.26)	
7	0.23%	1.03	0.62	0.34	-0.088	95.86%
	(4.69)	(105.6)	(41.23)	(23.60)	(-7.90)	
8	0.26%	1.06	0.71	0.37	-0.082	96.06%
	(5.14)	(106.5)	(45.85)	(24.87)	(-7.14)	
9	0.29%	1.10	0.81	0.39	-0.082	95.50%
	(5.04)	(97.47)	(46.01)	(23.37)	(-6.36)	
10	0.27%	1.17	1.03	0.43	-0.024	93.37%
	(3.52)	(77.20)	(43.61)	(18.79)	(-1.39)	
10-1	0.64%	0.01	0.20	0.20	0.067	7.21%
	(5.73)	(0.41)	(5.81)	(6.12)	(2.64)	

Panel D: Factor loadings for the value-weighted portfolios

Decile	alpha	MKT	SMB	HML	UMD	R^2
1	-0.34%	1.20	0.15	-0.01	-0.042	87.56%
	(-4.07)	(71.69)	(5.68)	(-0.58)	(-2.18)	
2	-0.10%	1.06	-0.01	0.02	-0.025	91.14%
	(-1.68)	(89.21)	(-0.79)	(1.07)	(-1.84)	
3	-0.02%	1.00	-0.03	0.00	-0.030	92.23%
	(-0.29)	(96.15)	(-1.63)	(0.11)	(-2.51)	
4	0.03%	1.00	-0.08	0.02	0.015	91.89%
	(0.61)	(95.89)	(-4.73)	(1.29)	(1.30)	
5	0.01%	1.01	-0.09	0.04	-0.029	93.31%
	(0.29)	(104.8)	(-5.72)	(2.62)	(-2.63)	
6	0.06%	0.95	-0.03	0.05	-0.005	93.60%
	(1.37)	(106.6)	(-1.95)	(4.12)	(-0.53)	
7	0.06%	1.01	-0.08	0.03	0.004	92.32%
	(1.15)	(98.45)	(-5.13)	(1.72)	(0.32)	
8	0.07%	1.02	0.05	0.12	0.005	90.66%
	(1.17)	(84.84)	(2.62)	(6.74)	(0.34)	
9	0.10%	1.08	0.02	0.14	-0.016	91.20%
	(1.67)	(87.36)	(1.25)	(7.68)	(-1.14)	
10	0.04%	1.14	0.38	0.14	0.027	88.23%
	(0.48)	(69.72)	(15.13)	(5.80)	(1.45)	
10-1	0.38%	-0.06	0.24	0.16	0.069	4.78%
	(2.98)	(-2.31)	(5.99)	(4.11)	(2.37)	

Panel E: Portfolio transition probabilities

Between month t and month $t+1$

Portfolio in	Portfolio in month $t+1$									
month t	1	2	3	4	5	6	7	8	9	10
1 (low signal)	0.24	0.13	0.09	0.07	0.06	0.06	0.06	0.07	0.08	0.13
2	0.12	0.15	0.12	0.10	0.09	0.08	0.08	0.08	0.09	0.08
3	0.08	0.12	0.13	0.12	0.11	0.10	0.10	0.09	0.08	0.06
4	0.07	0.10	0.12	0.13	0.12	0.12	0.11	0.10	0.08	0.05
5	0.06	0.09	0.11	0.12	0.13	0.13	0.12	0.10	0.08	0.05
6	0.05	0.08	0.10	0.12	0.13	0.14	0.13	0.11	0.09	0.06
7	0.05	0.08	0.10	0.11	0.11	0.13	0.13	0.12	0.10	0.06
8	0.06	0.08	0.09	0.10	0.10	0.11	0.12	0.13	0.12	0.08
9	0.08	0.09	0.09	0.08	0.08	0.09	0.10	0.12	0.15	0.12
10 (high signal)	0.12	0.08	0.07	0.06	0.06	0.06	0.07	0.09	0.13	0.25

Between month t and month $t+2$

Portfolio in	Portfolio in month $t+2$									
month t	1	2	3	4	5	6	7	8	9	10
1 (low signal)	0.18	0.12	0.09	0.07	0.07	0.07	0.07	0.08	0.10	0.16
...										
10 (high signal)	0.15	0.10	0.08	0.07	0.06	0.07	0.07	0.09	0.12	0.19

Between month t and month $t+12$

Portfolio in	Portfolio in month $t+12$									
month t	1	2	3	4	5	6	7	8	9	10
1 (low signal)	0.12	0.11	0.10	0.09	0.09	0.09	0.09	0.09	0.10	0.11
...										
10 (high signal)	0.12	0.11	0.10	0.09	0.09	0.09	0.09	0.10	0.11	0.12

Table 4

Portfolios sorted on the equal-weighted leader signal within 36 industries; leaders are identified using 36-month rolling regressions, 1930-2011

This table presents monthly abnormal returns of leader-signal-sorted portfolios. Leaders for each stock are identified using 36-month rolling regressions, as described in the text. At the beginning of each month, all stocks that traded on the last day of the prior month, that were priced above $5 per share in 2011 inflation-adjusted dollars, and that had leader stocks are sorted into decile portfolios within each of the 36 industries based on the last month's equal-weighted aggregate leader signal, computed as described in the text. Portfolio returns are equal-weighted in Panel A and value-weighted in Panel B. The second column reports the weighted-average leader signal, which is equal-weighted across followers in each portfolio in Panel A and value-weighted across followers in Panel B; the third column reports the average portfolio return in excess of the risk-free rate; the fourth column reports the market alpha; the fifth column reports the alpha of the Fama and French (1993) three-factor model; and the sixth column reports the alpha of the four-factor model that also includes the Carhart (1997) momentum factor. The last row reports the return differential between the high- and the low-signal portfolios (deciles 10 and 1). Panels C and D report the four-factor model factor loadings for the equal- and value-weighted portfolios, respectively. Newey-West-adjusted t-statistics are reported in parentheses.

Panel A: Equal-weighted portfolios

Decile	Leader signal	Excess return	Market alpha	3-factor alpha	4-factor alpha
1	-2.13%	0.55%	-0.29%	-0.53%	-0.38%
		(2.01)	(-2.34)	(-6.58)	(-4.80)
2	-1.20%	0.77%	0.03%	-0.15%	-0.11%
		(3.26)	(0.36)	(-2.32)	(-1.64)
3	-0.75%	0.82%	0.10%	-0.09%	0.01%
		(3.57)	(1.12)	(-1.80)	(0.11)
4	-0.40%	0.90%	0.21%	0.04%	0.10%
		(4.14)	(2.75)	(0.82)	(2.14)
5	-0.12%	0.94%	0.22%	0.02%	0.09%
		(4.13)	(2.64)	(0.50)	(1.72)
6	0.14%	1.01%	0.30%	0.11%	0.16%
		(4.51)	(3.69)	(2.47)	(3.65)
7	0.46%	1.14%	0.41%	0.20%	0.26%
		(4.92)	(4.75)	(4.41)	(5.54)
8	0.79%	1.20%	0.45%	0.22%	0.25%
		(4.94)	(4.53)	(3.86)	(4.37)
9	1.25%	1.36%	0.58%	0.32%	0.39%
		(5.35)	(5.36)	(6.01)	(7.22)
10	2.24%	1.45%	0.58%	0.27%	0.36%
		(4.98)	(4.23)	(3.59)	(4.71)
10-1		0.89%	0.86%	0.80%	0.74%
		(8.18)	(7.86)	(7.40)	(6.72)

Panel B: Value-weighted portfolios

Decile	Leader signal	Excess return	Market alpha	3-factor alpha	4-factor alpha
1	-1.69%	0.37%	-0.33%	-0.37%	-0.30%
		(1.67)	(-3.98)	(-4.48)	(-3.56)
2	-1.02%	0.56%	-0.09%	-0.13%	-0.11%
		(2.75)	(-1.45)	(-2.12)	(-1.74)
3	-0.65%	0.43%	-0.17%	-0.16%	-0.15%
		(2.33)	(-2.98)	(-2.97)	(-2.68)
4	-0.35%	0.53%	-0.07%	-0.06%	-0.06%
		(2.90)	(-1.28)	(-1.26)	(-1.12)
5	-0.12%	0.54%	-0.05%	-0.05%	-0.05%
		(2.99)	(-1.02)	(-1.02)	(-1.05)
6	0.11%	0.62%	0.05%	0.05%	0.05%
		(3.59)	(1.21)	(1.09)	(1.18)
7	0.39%	0.60%	0.02%	0.02%	0.03%
		(3.39)	(0.47)	(0.35)	(0.50)
8	0.67%	0.84%	0.23%	0.19%	0.19%
		(4.44)	(4.02)	(3.44)	(3.35)
9	1.04%	0.83%	0.18%	0.12%	0.13%
		(4.11)	(2.84)	(1.97)	(2.12)
10	1.76%	0.94%	0.21%	0.10%	0.08%
		(4.03)	(2.34)	(1.26)	(0.99)
10-1		0.56%	0.54%	0.47%	0.38%
		(4.45)	(4.21)	(3.71)	(2.94)

Table 5

Double sort on size and leader signal

This table presents monthly abnormal returns of size- and leader-signal-sorted portfolios. The sample consists of stocks that traded on the last day of the prior month, that were priced above $5 per share in 2011 inflation-adjusted dollars, and that had leaders. Within each industry, stocks are first sorted into terciles based on market capitalization as of the end of the previous month and then, within each industry and size tercile, into deciles based on the equal-weighted leader signal computed at the end of the previous month. This sorting method ensures an near-equal industry representation within each portfolio. In Panel A, leaders are identified with 12-month rolling regressions and in Panel B, leaders are identified with 36-month rolling regressions. The tables report four-factor alphas for equal-weighted portfolios and, in the last row, the return differentials between the high- and low-signal portfolios. Newey-West-adjusted t-statistics are reported in parentheses.

Panel A: Leaders are identified with 12-month rolling regressions

Leader signal	Size group		
decile	Small	Medium	Large
1 (low)	-0.49%	-0.28%	-0.29%
	(-3.08)	(-2.69)	(-4.21)
...			
10 (high)	0.33%	0.06%	0.07%
	(2.02)	(0.56)	(0.94)
10-1	0.82%	0.34%	0.36%
	(4.18)	(2.38)	(3.53)

Panel B: Leaders are identified with 36-month rolling regressions

Leader signal	Size group		
decile	Small	Medium	Large
1 (low)	-0.37%	-0.32%	-0.17%
	(-1.96)	(-3.76)	(-1.73)
...			
10 (high)	0.62%	-0.02%	0.04%
	(3.13)	(-0.19)	(0.42)
10-1	0.99%	0.30%	0.20%
	(4.40)	(2.03)	(1.76)

Table 6
One-month waiting period before portfolio formation

This table presents monthly abnormal returns of leader-signal-sorted portfolios. The sample consists of stocks that traded on the last day of the prior month, that were priced above $5 per share in 2011 inflation-adjusted dollars, and that had leaders. One month is skipped between the month in which leader signals are calculated and the time of portfolio formation. In Panel A, leaders are identified with 12-month rolling regressions and in Panel B, leaders are identified with 36-month rolling regressions. Portfolios are formed within 36 industries based on the equal-weighted leader signal computed at the end of the previous month. The tables report excess returns and four-factor alphas for equal- and value-weighted portfolios and, in the last row, the return differentials between the high- and low-signal portfolios. Newey-West-adjusted t-statistics are reported in parentheses.

Panel A: Leaders are identified with 12-month rolling regressions

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.71%	-0.22%	0.55%	-0.12%
	(2.64)	(-2.77)	(2.39)	(-1.38)
2	0.87%	0.03%	0.63%	-0.01%
	(3.48)	(0.51)	(3.02)	(-0.16)
3	0.87%	0.06%	0.61%	0.03%
	(3.68)	(1.11)	(3.22)	(0.55)
4	0.88%	0.08%	0.65%	0.09%
	(3.85)	(1.62)	(3.44)	(1.55)
5	0.83%	0.06%	0.62%	0.11%
	(3.70)	(1.15)	(3.44)	(2.21)
6	0.82%	0.05%	0.58%	0.05%
	(3.76)	(1.21)	(3.36)	(0.98)
7	0.88%	0.10%	0.59%	0.04%
	(4.01)	(2.21)	(3.35)	(0.79)
8	0.85%	0.03%	0.52%	-0.03%
	(3.74)	(0.69)	(2.85)	(-0.58)
9	0.90%	0.03%	0.52%	-0.14%
	(3.72)	(0.50)	(2.63)	(-2.36)
10	1.05%	-0.00%	0.73%	-0.06%
	(3.68)	(-0.03)	(3.23)	(-0.78)
10-1	0.33%	0.21%	0.18%	0.05%
	(2.89)	(1.81)	(1.43)	(0.42)

Panel B: Leaders are identified with 36-month rolling regressions

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.78%	-0.18%	0.58%	-0.14%
	(2.85)	(-2.49)	(2.49)	(-1.61)
2	0.88%	0.00%	0.68%	0.04%
	(3.46)	(0.06)	(3.22)	(0.65)
3	0.88%	0.05%	0.65%	0.05%
	(3.75)	(0.96)	(3.38)	(0.96)
4	0.85%	0.02%	0.61%	0.05%
	(3.81)	(0.54)	(3.41)	(1.04)
5	0.92%	0.07%	0.60%	0.01%
	(4.09)	(1.40)	(3.27)	(0.17)
6	0.92%	0.11%	0.58%	0.04%
	(4.19)	(2.55)	(3.46)	(0.80)
7	0.94%	0.11%	0.54%	-0.04%
	(4.30)	(2.30)	(3.02)	(-0.72)
8	0.99%	0.08%	0.76%	0.13%
	(4.20)	(1.49)	(4.03)	(2.38)
9	1.00%	0.07%	0.61%	-0.06%
	(4.19)	(1.12)	(3.19)	(-0.97)
10	1.10%	0.01%	0.67%	-0.16%
	(4.01)	(0.15)	(3.04)	(-1.96)
10-1	0.32%	0.19%	0.09%	-0.03%
	(3.03)	(1.80)	(0.73)	(-0.22)

Table 7

Alternative methods for aggregating leader signals

This table presents monthly abnormal return differentials between the highest- and lowest-signal decile portfolios. The sample consists of stocks that traded on the last day of the prior month, that were priced above $5 per share in 2011 inflation-adjusted dollars, and that had leaders. Leader signals are computed non-parametrically, by multiplying the past-months leader returns by the sign($\hat{b_3}$) and weighting them as described above each set of results. In Panel B, one month is skipped between the time that leader signals are calculated and portfolio formed. Leaders are identified with 12-month rolling regressions. Portfolios are formed within 36 industries based on the equal-weighted leader signal. Newey-West-adjusted t-statistics are reported in parentheses.

Panel A: Portfolios are formed based on month $t-1$ leader signals, corresponding to Table 3

Return Differentials (Portfolio 10−Portfolio 1)			
Equal-weighted		Value-weighted	
Excess return	4-factor alpha	Excess return	4-factor alpha
Non-parametric signals are equal-weighted			
0.83%	0.66%	0.35%	0.26%
(8.35)	(6.65)	(3.31)	(2.38)
Non-parametric signals are value-weighted			
0.29%	0.24%	0.23%	0.22%
(4.07)	(3.34)	(2.75)	(2.50)
Non-parametric signals are weighted by $\lvert t\text{-stat}(\hat{b_3})\rvert$			
0.87%	0.70%	0.37%	0.28%
(8.79)	(7.13)	(3.24)	(2.37)
Non-parametric signals are weighted by $\lvert\hat{b_3}\rvert$			
0.74%	0.56%	0.21%	0.12%
(7.67)	(5.85)	(1.90)	(1.05)

Panel B: Portfolios are formed based on month $t-2$ leader signals, corresponding to Table 6

Return Differentials (Portfolio 10 - Portfolio 1)			
Equal-weighted		Value-weighted	
Excess return	4-factor alpha	Excess return	4-factor alpha
Non-parametric signals are equal-weighted			
0.31%	0.25%	0.04%	-0.01%
(3.76)	(2.95)	(0.44)	(-0.11)
Non-parametric signals are value-weighted			
0.07%	-0.03%	0.03%	-0.05%
(0.95)	(-0.48)	(0.34)	(-0.56)
Non-parametric signals are weighted by $\lvert t\text{-stat}(\hat{b_3})\rvert$			
0.29%	0.23%	0.11%	0.05%
(3.52)	(2.75)	(1.03)	(0.49)
Non-parametric signals are weighted by $\lvert\hat{b_3}\rvert$			
0.28%	0.20%	0.08%	0.01%
(3.20)	(2.25)	(0.80)	(0.11)

Table 8

Alternative specifications and robustness checks

This table presents monthly abnormal returns of leader-signal-sorted portfolios. The sample consists of stocks that traded on the last day of the prior month, that were priced above $5 per share in 2011 inflation-adjusted dollars, and that had leaders. In the baseline specification, leaders are identified with 12-month rolling regressions and portfolios are formed within 36 industries based on the equal-weighted leader signal computed at the end of the previous month. Variations on this baseline specification are described in each panel heading. Each panel reports excess returns and four-factor alphas for equal- and value-weighted portfolios and, in the last row, the return differentials between the high- and low-signal portfolios. Newey-West-adjusted t-statistics are reported in parentheses.

Panel A: Stocks are sorted over the entire sample and *not* within each industry

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.54%	-0.36%	0.31%	-0.41%
	(1.91)	(-4.13)	(1.26)	(-4.03)
2	0.66%	-0.11%	0.49%	-0.11%
	(2.74)	(-1.74)	(2.29)	(-1.48)
3	0.80%	0.04%	0.49%	-0.03%
	(3.56)	(0.73)	(2.70)	(-0.48)
4	0.84%	0.10%	0.52%	-0.03%
	(3.94)	(1.92)	(2.99)	(-0.48)
5	0.90%	0.14%	0.61%	0.07%
	(4.17)	(2.76)	(3.47)	(1.35)
6	0.97%	0.20%	0.60%	0.04%
	(4.47)	(4.40)	(3.45)	(0.84)
7	1.08%	0.24%	0.69%	0.09%
	(4.72)	(4.56)	(3.80)	(1.71)
8	1.13%	0.28%	0.74%	0.12%
	(4.70)	(4.97)	(3.81)	(1.90)
9	1.28%	0.32%	0.91%	0.19%
	(4.78)	(4.93)	(4.21)	(2.50)
10	1.34%	0.23%	0.97%	0.07%
	(4.39)	(2.69)	(3.77)	(0.74)
10-1	0.80%	0.59%	0.66%	0.48%
	(6.45)	(4.82)	(4.40)	(3.16)

Panel B: Stocks are sorted over the entire sample and *not* within each industry; leaders are identified with 36-month rolling regressions

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.58%	-0.37%	0.43%	-0.29%
	(1.99)	(-3.79)	(1.63)	(-2.62)
2	0.74%	-0.10%	0.42%	-0.18%
	(3.02)	(-1.59)	(2.01)	(-2.28)
3	0.75%	-0.03%	0.41%	-0.20%
	(3.43)	(-0.46)	(2.13)	(-3.26)
4	0.90%	0.09%	0.53%	-0.06%
	(4.07)	(1.61)	(2.96)	(-1.09)
5	0.96%	0.15%	0.57%	0.01%
	(4.51)	(2.74)	(3.33)	(0.20)
6	1.02%	0.16%	0.65%	0.04%
	(4.64)	(3.27)	(3.69)	(0.69)
7	1.12%	0.24%	0.71%	0.10%
	(4.91)	(4.51)	(3.86)	(1.63)
8	1.19%	0.28%	0.83%	0.17%
	(5.11)	(5.30)	(4.31)	(2.55)
9	1.41%	0.40%	0.94%	0.17%
	(5.36)	(6.04)	(4.48)	(2.37)
10	1.50%	0.38%	1.03%	0.09%
	(4.99)	(4.28)	(3.88)	(0.86)
10-1	0.93%	0.75%	0.60%	0.38%
	(6.96)	(5.49)	(3.61)	(2.25)

Panel C: Leader signals are value-weighted

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.82%	-0.11%	0.45%	-0.22%
	(2.97)	(-1.55)	(2.01)	(-2.95)
2	0.84%	0.01%	0.53%	-0.07%
	(3.40)	(0.28)	(2.67)	(-1.37)
3	0.89%	0.09%	0.59%	0.03%
	(3.74)	(1.85)	(3.18)	(0.49)
4	0.90%	0.12%	0.61%	0.08%
	(3.99)	(2.65)	(3.38)	(1.56)
5	0.97%	0.19%	0.68%	0.08%
	(4.34)	(4.18)	(3.69)	(1.78)
6	0.97%	0.17%	0.61%	0.09%
	(4.27)	(3.86)	(3.48)	(1.93)
7	1.01%	0.19%	0.58%	0.03%
	(4.47)	(3.73)	(3.21)	(0.49)
8	1.02%	0.16%	0.57%	-0.07%
	(4.30)	(3.16)	(2.97)	(-1.15)
9	1.07%	0.16%	0.71%	0.06%
	(4.32)	(2.61)	(3.62)	(1.04)
10	1.05%	0.03%	0.82%	0.08%
	(3.75)	(0.48)	(3.62)	(0.90)
10-1	0.23%	0.14%	0.37%	0.30%
	(2.82)	(1.69)	(3.25)	(2.54)

Panel D: Leader signals are value-weighted; leaders are identified with 36-month rolling regressions

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.81%	-0.19%	0.53%	-0.21%
	(2.88)	(-2.65)	(2.38)	(-2.70)
2	0.88%	-0.05%	0.50%	-0.17%
	(3.49)	(-0.87)	(2.50)	(-2.77)
3	0.98%	0.13%	0.61%	-0.02%
	(4.11)	(2.60)	(3.25)	(-0.33)
4	0.94%	0.12%	0.50%	-0.05%
	(4.19)	(2.48)	(2.86)	(-1.10)
5	0.96%	0.14%	0.60%	0.01%
	(4.39)	(3.16)	(3.21)	(0.21)
6	1.07%	0.26%	0.63%	0.08%
	(4.82)	(5.51)	(3.75)	(1.71)
7	1.05%	0.19%	0.63%	0.06%
	(4.68)	(3.76)	(3.59)	(1.21)
8	1.16%	0.23%	0.77%	0.15%
	(4.80)	(4.48)	(4.22)	(2.85)
9	1.15%	0.18%	0.76%	0.09%
	(4.66)	(3.51)	(3.87)	(1.51)
10	1.17%	0.09%	0.85%	0.01%
	(4.22)	(1.37)	(3.59)	(0.08)
10-1	0.36%	0.28%	0.31%	0.21%
	(4.30)	(3.33)	(2.72)	(1.80)

Panel E: 1990-2011 sample period

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.33%	-0.31%	0.24%	-0.32%
	(0.80)	(-2.37)	(0.60)	(-1.97)
2	0.61%	0.01%	0.40%	-0.13%
	(1.78)	(0.07)	(1.25)	(-1.10)
3	0.70%	0.11%	0.48%	-0.00%
	(2.18)	(1.42)	(1.63)	(-0.01)
4	0.66%	0.09%	0.47%	0.03%
	(2.21)	(1.25)	(1.71)	(0.31)
5	0.73%	0.16%	0.57%	0.21%
	(2.50)	(2.33)	(2.03)	(2.00)
6	0.75%	0.18%	0.57%	0.13%
	(2.58)	(2.65)	(2.18)	(1.36)
7	0.80%	0.21%	0.46%	-0.02%
	(2.66)	(2.93)	(1.60)	(-0.18)
8	0.80%	0.20%	0.63%	0.13%
	(2.59)	(2.59)	(2.23)	(1.16)
9	0.90%	0.25%	0.48%	-0.09%
	(2.65)	(2.92)	(1.50)	(-0.65)
10	0.81%	0.08%	0.53%	-0.17%
	(2.00)	(0.61)	(1.37)	(-0.92)
10-1	0.48%	0.38%	0.30%	0.15%
	(2.37)	(1.88)	(1.12)	(0.57)

Panel F: 1990-2011 sample period; leaders are identified with 36-month rolling regressions

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.32%	-0.32%	0.37%	-0.18%
	(0.78)	(-2.45)	(0.99)	(-1.16)
2	0.51%	-0.10%	0.40%	-0.14%
	(1.53)	(-1.05)	(1.27)	(-1.07)
3	0.58%	-0.02%	0.32%	-0.12%
	(1.84)	(-0.20)	(1.09)	(-1.06)
4	0.63%	0.05%	0.30%	-0.15%
	(2.15)	(0.70)	(1.06)	(-1.54)
5	0.74%	0.16%	0.43%	-0.04%
	(2.58)	(2.23)	(1.56)	(-0.38)
6	0.80%	0.22%	0.58%	0.11%
	(2.78)	(3.08)	(2.23)	(1.28)
7	0.94%	0.34%	0.54%	0.10%
	(3.20)	(4.79)	(2.08)	(1.00)
8	0.97%	0.36%	0.58%	0.09%
	(3.18)	(4.68)	(1.99)	(0.82)
9	1.11%	0.46%	0.69%	0.12%
	(3.37)	(4.96)	(2.19)	(0.89)
10	1.07%	0.34%	0.84%	0.15%
	(2.70)	(2.80)	(2.30)	(0.90)
10-1	0.75%	0.66%	0.47%	0.32%
	(3.75)	(3.28)	(1.96)	(1.34)

Panel G: Only signals from positive leaders are used

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.50%	-0.35%	0.26%	-0.46%
	(1.95)	(-3.56)	(1.16)	(-4.72)
2	0.69%	-0.11%	0.49%	-0.13%
	(2.91)	(-1.37)	(2.42)	(-1.80)
3	0.76%	-0.01%	0.54%	-0.07%
	(3.33)	(-0.23)	(2.74)	(-1.07)
4	0.87%	0.10%	0.65%	0.09%
	(3.77)	(1.92)	(3.49)	(1.40)
5	0.96%	0.14%	0.62%	0.07%
	(4.06)	(2.91)	(3.20)	(1.24)
6	1.03%	0.23%	0.62%	0.06%
	(4.35)	(5.11)	(3.37)	(1.19)
7	1.11%	0.26%	0.76%	0.12%
	(4.63)	(5.17)	(3.96)	(2.30)
8	1.17%	0.29%	0.82%	0.21%
	(4.77)	(4.98)	(4.16)	(3.57)
9	1.19%	0.26%	0.88%	0.21%
	(4.63)	(3.57)	(4.11)	(2.82)
10	1.21%	0.20%	0.84%	0.09%
	(4.31)	(1.99)	(3.48)	(0.80)
10-1	0.71%	0.55%	0.58%	0.54%
	(4.23)	(3.20)	(3.48)	(3.17)

Panel H: Only signals from negative leaders are used

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.71%	-0.23%	0.48%	-0.22%
	(2.54)	(-2.48)	(2.01)	(-2.17)
2	0.92%	0.02%	0.65%	-0.01%
	(3.61)	(0.23)	(3.02)	(-0.18)
3	0.96%	0.11%	0.58%	-0.02%
	(4.06)	(2.27)	(2.89)	(-0.28)
4	1.01%	0.19%	0.57%	-0.00%
	(4.20)	(3.78)	(2.98)	(-0.02)
5	0.99%	0.19%	0.69%	0.12%
	(4.25)	(3.80)	(3.55)	(2.30)
6	1.01%	0.21%	0.69%	0.08%
	(4.37)	(4.64)	(3.60)	(1.54)
7	1.03%	0.24%	0.78%	0.20%
	(4.50)	(4.59)	(4.13)	(3.42)
8	0.94%	0.15%	0.58%	-0.01%
	(4.02)	(2.47)	(3.03)	(-0.09)
9	0.98%	0.11%	0.65%	-0.01%
	(3.90)	(1.45)	(3.13)	(-0.12)
10	0.95%	0.03%	0.51%	-0.24%
	(3.62)	(0.29)	(2.25)	(-2.75)
10-1	0.24%	0.26%	0.03%	-0.02%
	(1.51)	(1.66)	(0.20)	(-0.11)

Panel I: Only signals from leaders in a different industry are used

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.54%	-0.36%	0.40%	-0.31%
	(1.97)	(-4.42)	(1.68)	(-3.49)
2	0.72%	-0.08%	0.47%	-0.13%
	(3.07)	(-1.37)	(2.39)	(-2.09)
3	0.77%	-0.01%	0.54%	-0.02%
	(3.37)	(-0.26)	(3.01)	(-0.48)
4	0.91%	0.16%	0.64%	0.11%
	(4.20)	(3.27)	(3.70)	(2.24)
5	0.87%	0.09%	0.57%	0.02%
	(3.90)	(1.75)	(3.06)	(0.38)
6	0.99%	0.20%	0.56%	0.01%
	(4.40)	(4.65)	(3.24)	(0.30)
7	1.08%	0.24%	0.68%	0.10%
	(4.54)	(4.88)	(3.70)	(1.97)
8	1.13%	0.27%	0.70%	0.10%
	(4.59)	(4.92)	(3.70)	(1.75)
9	1.18%	0.27%	0.74%	0.06%
	(4.59)	(4.97)	(3.61)	(0.98)
10	1.33%	0.25%	0.86%	0.10%
	(4.62)	(3.17)	(3.74)	(1.11)
10-1	0.79%	0.61%	0.46%	0.41%
	(7.00)	(5.38)	(3.59)	(3.09)

Panel J: Only signals from leaders that are smaller than followers are used

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.71%	-0.26%	0.36%	-0.31%
	(2.44)	(-3.39)	(1.55)	(-3.57)
2	0.81%	-0.05%	0.53%	-0.07%
	(3.25)	(-0.95)	(2.72)	(-1.16)
3	0.88%	0.08%	0.48%	-0.08%
	(3.77)	(1.73)	(2.69)	(-1.48)
4	0.87%	0.08%	0.63%	0.09%
	(3.83)	(1.72)	(3.55)	(1.78)
5	0.92%	0.13%	0.62%	0.09%
	(4.18)	(2.83)	(3.48)	(1.80)
6	0.95%	0.19%	0.48%	-0.05%
	(4.35)	(4.47)	(2.77)	(-1.20)
7	1.00%	0.21%	0.64%	0.07%
	(4.48)	(4.53)	(3.49)	(1.40)
8	1.04%	0.22%	0.75%	0.13%
	(4.47)	(4.67)	(3.93)	(2.32)
9	1.07%	0.17%	0.70%	0.04%
	(4.22)	(2.98)	(3.49)	(0.70)
10	1.32%	0.28%	0.79%	0.02%
	(4.60)	(3.50)	(3.36)	(0.26)
10-1	0.61%	0.54%	0.42%	0.33%
	(6.07)	(5.21)	(3.38)	(2.62)

Panel K: Only signals from first-time leaders in three years are used

	EW portfolios		VW portfolios	
	Excess	4-factor	Excess	4-factor
Decile	return	alpha	return	alpha
1	0.80%	-0.14%	0.49%	-0.16%
	(2.94)	(-1.75)	(2.26)	(-2.16)
2	0.89%	0.04%	0.59%	-0.05%
	(3.64)	(0.66)	(2.94)	(-0.91)
3	0.84%	0.03%	0.53%	-0.05%
	(3.69)	(0.57)	(2.86)	(-0.95)
4	0.90%	0.12%	0.57%	0.05%
	(4.03)	(2.80)	(3.34)	(0.96)
5	0.91%	0.12%	0.51%	-0.04%
	(4.04)	(2.60)	(2.91)	(-0.87)
6	0.94%	0.16%	0.65%	0.13%
	(4.17)	(3.66)	(3.76)	(2.65)
7	1.00%	0.20%	0.59%	0.07%
	(4.31)	(4.33)	(3.39)	(1.59)
8	1.04%	0.21%	0.69%	0.10%
	(4.40)	(4.24)	(3.81)	(2.00)
9	1.08%	0.17%	0.73%	0.08%
	(4.22)	(3.16)	(3.63)	(1.37)
10	1.14%	0.13%	0.73%	0.02%
	(4.10)	(2.12)	(3.28)	(0.28)
10-1	0.34%	0.27%	0.24%	0.19%
	(3.93)	(3.03)	(2.38)	(1.75)

Panel L: Only signals from leaders that also led the follower at some time in the previous three years are used

	EW portfolios		VW portfolios	
	Excess	4-factor	Excess	4-factor
Decile	return	alpha	return	alpha
1	0.54%	-0.37%	0.35%	-0.31%
	(1.99)	(-4.84)	(1.58)	(-3.69)
2	0.73%	-0.07%	0.52%	-0.11%
	(3.05)	(-1.11)	(2.56)	(-1.81)
3	0.80%	0.03%	0.55%	0.03%
	(3.50)	(0.67)	(3.11)	(0.66)
4	0.89%	0.12%	0.57%	-0.00%
	(4.03)	(2.63)	(3.21)	(-0.03)
5	0.87%	0.11%	0.56%	0.01%
	(3.95)	(2.33)	(3.04)	(0.16)
6	1.00%	0.19%	0.66%	0.10%
	(4.33)	(3.99)	(3.78)	(2.22)
7	1.05%	0.22%	0.60%	0.02%
	(4.47)	(4.50)	(3.26)	(0.40)
8	1.12%	0.25%	0.66%	0.05%
	(4.66)	(5.27)	(3.54)	(0.87)
9	1.25%	0.31%	0.82%	0.14%
	(4.82)	(5.41)	(4.02)	(2.20)
10	1.33%	0.27%	0.84%	0.03%
	(4.64)	(3.50)	(3.57)	(0.34)
10-1	0.79%	0.64%	0.49%	0.34%
	(7.59)	(6.11)	(3.79)	(2.59)

Panel M: Only signals from leaders that are *not* announcing quarterly earnings in the current month are used (sample period: 1972-2011)

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.30%	-0.35%	0.17%	-0.27%
	(0.98)	(-4.22)	(0.63)	(-2.52)
2	0.57%	-0.08%	0.44%	-0.07%
	(2.13)	(-1.29)	(1.89)	(-0.79)
3	0.65%	0.03%	0.43%	0.00%
	(2.62)	(0.59)	(2.00)	(0.02)
4	0.73%	0.10%	0.52%	0.11%
	(2.99)	(1.83)	(2.49)	(1.64)
5	0.72%	0.09%	0.55%	0.14%
	(2.97)	(1.59)	(2.62)	(2.05)
6	0.79%	0.15%	0.56%	0.13%
	(3.26)	(2.66)	(2.74)	(2.17)
7	0.83%	0.18%	0.45%	0.02%
	(3.37)	(3.01)	(2.13)	(0.25)
8	0.87%	0.21%	0.54%	0.07%
	(3.43)	(3.28)	(2.51)	(0.98)
9	0.90%	0.18%	0.53%	-0.02%
	(3.32)	(2.66)	(2.25)	(-0.23)
10	0.85%	0.05%	0.53%	-0.14%
	(2.69)	(0.56)	(1.87)	(-1.13)
10-1	0.55%	0.40%	0.36%	0.13%
	(4.50)	(3.20)	(2.06)	(0.76)

Panel N: Only signals from leaders that *are* announcing quarterly earnings in the current month are used (sample period: 1972-2011)

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.37%	-0.30%	0.34%	-0.24%
	(1.18)	(-3.32)	(1.24)	(-2.35)
2	0.53%	-0.09%	0.29%	-0.20%
	(1.99)	(-1.45)	(1.21)	(-2.44)
3	0.59%	-0.03%	0.43%	0.01%
	(2.34)	(-0.42)	(1.96)	(0.16)
4	0.70%	0.09%	0.50%	0.03%
	(2.82)	(1.44)	(2.35)	(0.43)
5	0.74%	0.11%	0.66%	0.32%
	(3.06)	(1.96)	(3.11)	(4.49)
6	0.79%	0.15%	0.50%	0.08%
	(3.24)	(2.62)	(2.38)	(1.24)
7	0.80%	0.15%	0.47%	0.03%
	(3.27)	(2.67)	(2.22)	(0.48)
8	0.82%	0.17%	0.47%	-0.03%
	(3.20)	(2.62)	(2.15)	(-0.36)
9	0.92%	0.23%	0.58%	0.07%
	(3.39)	(3.44)	(2.49)	(0.87)
10	0.80%	0.05%	0.50%	-0.07%
	(2.59)	(0.60)	(1.91)	(-0.68)
10-1	0.42%	0.35%	0.15%	0.17%
	(3.47)	(2.80)	(1.03)	(1.10)

Panel O: Leaders are determined using a cutoff t-statistic $(\hat{b}_3) \geq 2.57$

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.58%	-0.32%	0.34%	-0.33%
	(2.12)	(-4.57)	(1.43)	(-4.02)
2	0.74%	-0.08%	0.56%	-0.06%
	(3.07)	(-1.50)	(2.88)	(-0.98)
3	0.80%	0.02%	0.53%	-0.02%
	(3.50)	(0.37)	(2.92)	(-0.38)
4	0.89%	0.15%	0.59%	0.06%
	(4.08)	(3.07)	(3.41)	(1.17)
5	0.85%	0.05%	0.62%	0.05%
	(3.76)	(1.09)	(3.33)	(1.04)
6	0.97%	0.19%	0.59%	0.05%
	(4.33)	(4.24)	(3.46)	(1.07)
7	1.04%	0.24%	0.65%	0.12%
	(4.54)	(5.04)	(3.58)	(2.43)
8	1.12%	0.25%	0.66%	0.07%
	(4.72)	(5.27)	(3.62)	(1.30)
9	1.21%	0.26%	0.73%	0.05%
	(4.59)	(4.19)	(3.55)	(0.85)
10	1.32%	0.27%	0.87%	0.06%
	(4.58)	(3.57)	(3.72)	(0.84)
10-1	0.74%	0.59%	0.53%	0.40%
	(7.20)	(5.69)	(4.46)	(3.24)

Panel P: Leaders are determined using a cutoff t-statistic $(\hat{b}_3) \geq 2.57$; stocks are sorted over the entire sample and *not* within each industry

	EW portfolios		VW portfolios	
Decile	Excess return	4-factor alpha	Excess return	4-factor alpha
1	0.55%	-0.35%	0.31%	-0.43%
	(1.96)	(-4.40)	(1.26)	(-4.49)
2	0.73%	-0.05%	0.49%	-0.09%
	(3.01)	(-0.78)	(2.38)	(-1.26)
3	0.80%	0.04%	0.56%	0.01%
	(3.54)	(0.72)	(3.10)	(0.24)
4	0.83%	0.07%	0.55%	0.01%
	(3.85)	(1.37)	(3.12)	(0.21)
5	0.90%	0.14%	0.53%	-0.02%
	(4.16)	(3.01)	(3.07)	(-0.33)
6	0.94%	0.16%	0.61%	0.05%
	(4.39)	(3.40)	(3.47)	(1.02)
7	1.07%	0.27%	0.70%	0.12%
	(4.77)	(5.45)	(3.89)	(2.24)
8	1.16%	0.28%	0.77%	0.12%
	(4.82)	(4.96)	(3.99)	(1.97)
9	1.21%	0.27%	0.73%	0.02%
	(4.66)	(4.50)	(3.38)	(0.26)
10	1.34%	0.25%	0.90%	0.04%
	(4.35)	(3.02)	(3.51)	(0.44)
10-1	0.79%	0.61%	0.58%	0.47%
	(6.52)	(5.04)	(4.30)	(3.37)

Panel Q: Skip one month between the end of the rolling regression window and the estimation of the leader signal

	EW portfolios		VW portfolios	
	Excess	4-factor	Excess	4-factor
Decile	return	alpha	return	alpha
1	0.76%	-0.24%	0.51%	-0.22%
	(2.78)	(-3.14)	(2.29)	(-2.71)
2	0.85%	-0.02%	0.60%	-0.03%
	(3.58)	(-0.38)	(3.11)	(-0.43)
3	0.90%	0.06%	0.60%	0.01%
	(3.86)	(1.25)	(3.31)	(0.28)
4	0.90%	0.09%	0.55%	-0.02%
	(4.05)	(2.01)	(3.16)	(-0.38)
5	0.95%	0.13%	0.60%	0.05%
	(4.28)	(2.91)	(3.32)	(0.94)
6	1.02%	0.19%	0.56%	-0.00%
	(4.55)	(4.16)	(3.23)	(-0.09)
7	1.06%	0.19%	0.64%	0.02%
	(4.52)	(3.89)	(3.47)	(0.45)
8	1.15%	0.23%	0.70%	0.07%
	(4.66)	(4.46)	(3.71)	(1.35)
9	1.25%	0.28%	0.79%	0.09%
	(4.96)	(4.99)	(3.87)	(1.50)
10	1.22%	0.09%	0.81%	0.00%
	(4.27)	(1.15)	(3.42)	(0.04)
10-1	0.45%	0.33%	0.30%	0.23%
	(4.49)	(3.19)	(2.33)	(1.72)

Panel R: Skip five years between the end of the rolling regression window and the estimation of the leader signal

	EW portfolios		VW portfolios	
	Excess	4-factor	Excess	4-factor
Decile	return	alpha	return	alpha
1	0.89%	-0.07%	0.70%	-0.02%
	(3.93)	(-0.96)	(3.76)	(-0.28)
2	0.93%	0.03%	0.66%	-0.03%
	(4.47)	(0.56)	(3.73)	(-0.54)
3	0.92%	0.05%	0.66%	0.02%
	(4.76)	(1.01)	(4.11)	(0.38)
4	0.95%	0.14%	0.61%	0.01%
	(5.06)	(3.14)	(3.94)	(0.31)
5	0.91%	0.09%	0.57%	-0.07%
	(4.97)	(2.16)	(3.70)	(-1.34)
6	0.98%	0.15%	0.70%	0.07%
	(5.36)	(3.95)	(4.64)	(1.51)
7	0.96%	0.12%	0.65%	0.05%
	(5.14)	(2.99)	(4.24)	(1.10)
8	1.06%	0.20%	0.73%	0.07%
	(5.57)	(4.46)	(4.65)	(1.36)
9	0.99%	0.09%	0.66%	0.02%
	(4.98)	(1.94)	(4.13)	(0.31)
10	0.94%	-0.03%	0.56%	-0.21%
	(4.31)	(-0.54)	(2.99)	(-3.18)
10-1	0.05%	0.03%	-0.15%	-0.19%
	(0.60)	(0.44)	(-1.53)	(-1.87)

Table 9

Weekly portfolios sorted on the equal-weighted leader signal within 36 industries, 1980-2011

This table presents weekly abnormal returns of leader-signal-sorted portfolios. Leaders for each stock are identified using 52-week rolling regressions, as described in the text. At the beginning of each week, all stocks that traded on the last day of the prior week, that were priced above $5 per share in 2011 inflation-adjusted dollars, and that had leader stocks are sorted into decile portfolios within each of the 36 industries based on the previous week's equal-weighted aggregate leader signal, computed as described in the text. Portfolio returns are equal-weighted in Panel A and value-weighted in Panel B. The second column reports the weighted-average leader signal, which is equal-weighted across followers in each portfolio in Panel A and value-weighted across followers in Panel B; the third column reports the average weekly portfolio return in excess of the risk-free rate; the fourth column reports the market alpha; the fifth column reports the weekly alpha of the Fama and French (1993) three-factor model; and the sixth column reports the weekly alpha of the four-factor model that also includes the Carhart (1997) momentum factor, using weekly factor returns. The last row reports the return differential between the high- and the low-signal portfolios (deciles 10 and 1). Panels C and D report the four-factor model factor loadings for the equal- and value-weighted portfolios, respectively. Panel E presents transition probabilities. Newey-West-adjusted t-statistics are reported in parentheses.

Panel A: Equal-weighted portfolios

Decile	Excess return	Market alpha	3-factor alpha	4-factor alpha
1	-0.15%	-0.28%	-0.31%	-0.28%
	(-1.74)	(-6.73)	(-12.2)	(-11.3)
2	0.02%	-0.11%	-0.14%	-0.11%
	(0.24)	(-3.32)	(-7.52)	(-6.56)
3	0.09%	-0.03%	-0.06%	-0.04%
	(1.28)	(-0.92)	(-3.54)	(-2.29)
4	0.15%	0.04%	0.01%	0.03%
	(2.33)	(1.46)	(0.73)	(1.98)
5	0.17%	0.06%	0.03%	0.05%
	(2.61)	(2.22)	(2.20)	(3.31)
6	0.21%	0.10%	0.07%	0.09%
	(3.17)	(3.49)	(4.62)	(5.59)
7	0.25%	0.14%	0.11%	0.13%
	(3.72)	(4.55)	(6.87)	(7.70)
8	0.29%	0.18%	0.15%	0.17%
	(4.13)	(5.57)	(9.45)	(9.83)
9	0.32%	0.21%	0.18%	0.20%
	(4.34)	(5.74)	(9.78)	(10.63)
10	0.38%	0.26%	0.23%	0.25%
	(4.43)	(5.54)	(8.92)	(9.48)
10-1	0.53%	0.54%	0.55%	0.53%
	(12.63)	(12.79)	(12.83)	(12.54)

Panel B: Value-weighted portfolios

Decile	Excess return	Market alpha	3-factor alpha	4-factor alpha
1	-0.04%	-0.19%	-0.23%	-0.18%
	(-0.56)	(-6.11)	(-7.11)	(-6.05)
2	0.02%	-0.12%	-0.13%	-0.12%
	(0.27)	(-5.85)	(-6.05)	(-5.82)
3	0.09%	-0.04%	-0.04%	-0.04%
	(1.50)	(-2.01)	(-2.43)	(-1.91)
4	0.12%	0.00%	0.00%	0.00%
	(2.20)	(0.19)	(0.09)	(0.11)
5	0.11%	-0.00%	0.00%	-0.00%
	(2.03)	(-0.26)	(0.07)	(-0.15)
6	0.15%	0.03%	0.04%	0.04%
	(2.65)	(2.25)	(2.76)	(2.57)
7	0.14%	0.02%	0.02%	0.02%
	(2.57)	(1.40)	(1.40)	(0.93)
8	0.22%	0.10%	0.10%	0.10%
	(3.68)	(5.15)	(5.45)	(5.00)
9	0.24%	0.11%	0.11%	0.11%
	(3.54)	(4.79)	(4.73)	(4.73)
10	0.24%	0.10%	0.10%	0.11%
	(3.17)	(3.23)	(3.24)	(3.44)
10-1	0.28%	0.29%	0.32%	0.29%
	(6.27)	(6.50)	(6.96)	(6.30)

Panel C: Factor loadings for the equal-weighted portfolios

Decile	alpha	MKT	SMB	HML	UMD	R^2
1	-0.28%	1.09	0.89	0.14	-0.156	91.85%
	(-11.3)	(40.72)	(24.90)	(3.09)	(-5.82)	
2	-0.11%	1.01	0.71	0.18	-0.115	94.35%
	(-6.56)	(52.09)	(18.43)	(6.18)	(-5.27)	
3	-0.04%	0.95	0.64	0.20	-0.103	94.52%
	(-2.29)	(52.49)	(15.55)	(6.74)	(-4.22)	
4	0.03%	0.92	0.59	0.19	-0.095	95.05%
	(1.98)	(65.74)	(15.10)	(7.59)	(-4.16)	
5	0.05%	0.91	0.58	0.19	-0.090	95.31%
	(3.31)	(61.41)	(16.39)	(7.28)	(-3.92)	
6	0.09%	0.90	0.59	0.19	-0.089	95.44%
	(5.59)	(63.84)	(19.58)	(7.30)	(-4.21)	
7	0.13%	0.91	0.62	0.21	-0.087	94.99%
	(7.70)	(58.38)	(20.91)	(6.59)	(-3.82)	
8	0.17%	0.91	0.64	0.17	-0.083	94.58%
	(9.83)	(56.15)	(24.17)	(5.63)	(-3.75)	
9	0.20%	0.95	0.74	0.16	-0.086	94.29%
	(10.63)	(60.55)	(32.00)	(5.10)	(-4.44)	
10	0.25%	1.01	0.92	0.07	-0.100	91.23%
	(9.48)	(49.00)	(38.42)	(1.97)	(-5.09)	
10-1	0.53%	-0.08	0.03	-0.07	0.057	3.30%
	(12.54)	(-1.71)	(0.55)	(-0.95)	(1.41)	

Panel D: Factor loadings for the value-weighted portfolios

Decile	alpha	MKT	SMB	HML	UMD	R^2
1	-0.18%	1.23	0.27	0.22	-0.200	83.54%
	(-6.05)	(27.39)	(7.62)	(2.61)	(-4.49)	
2	-0.12%	1.09	0.06	0.08	-0.045	90.03%
	(-5.82)	(50.56)	(2.13)	(2.06)	(-1.93)	
3	-0.04%	1.01	-0.04	0.05	-0.029	91.36%
	(-1.91)	(66.65)	(-1.57)	(2.29)	(-1.57)	
4	0.00%	0.99	-0.07	0.03	-0.002	92.22%
	(0.11)	(72.96)	(-2.80)	(1.32)	(-0.14)	
5	-0.00%	0.96	-0.08	-0.04	0.018	91.77%
	(-0.15)	(106.5)	(-4.03)	(-1.77)	(1.18)	
6	0.04%	0.96	-0.09	-0.05	0.010	92.48%
	(2.57)	(107.2)	(-5.55)	(-2.77)	(0.82)	
7	0.02%	0.97	-0.04	0.02	0.038	90.79%
	(0.93)	(85.60)	(-2.14)	(1.01)	(2.79)	
8	0.10%	0.96	0.00	-0.02	0.037	90.12%
	(5.00)	(60.47)	(0.17)	(-0.72)	(2.34)	
9	0.11%	1.04	0.06	-0.01	0.004	88.06%
	(4.73)	(44.77)	(2.78)	(-0.23)	(0.19)	
10	0.11%	1.12	0.29	-0.03	-0.055	82.66%
	(3.44)	(45.12)	(6.11)	(-0.66)	(-1.79)	
10-1	0.29%	-0.10	0.02	-0.25	0.146	6.37%
	(6.30)	(-1.59)	(0.24)	(-2.12)	(2.28)	

Panel E: Portfolio transition probabilities

Between week t and week $t+1$

Portfolio in	Portfolio in week $t+1$									
week t	1	2	3	4	5	6	7	8	9	10
1 (low signal)	0.21	0.12	0.08	0.06	0.06	0.06	0.06	0.07	0.10	0.18
2	0.11	0.13	0.11	0.09	0.09	0.09	0.09	0.09	0.11	0.09
3	0.08	0.11	0.12	0.11	0.10	0.11	0.10	0.10	0.10	0.07
4	0.06	0.09	0.11	0.12	0.12	0.12	0.11	0.10	0.09	0.06
5	0.06	0.09	0.11	0.12	0.13	0.13	0.12	0.10	0.08	0.05
6	0.05	0.08	0.11	0.12	0.13	0.14	0.13	0.11	0.09	0.05
7	0.06	0.09	0.10	0.11	0.12	0.13	0.12	0.11	0.09	0.06
8	0.07	0.10	0.10	0.11	0.10	0.11	0.11	0.12	0.11	0.08
9	0.09	0.11	0.10	0.09	0.08	0.09	0.10	0.11	0.13	0.11
10 (high signal)	0.17	0.10	0.07	0.06	0.06	0.06	0.07	0.08	0.12	0.22

Between week t and week $t+2$

Portfolio in	Portfolio in week $t+2$									
week t	1	2	3	4	5	6	7	8	9	10
1 (low signal)	0.20	0.12	0.08	0.07	0.06	0.06	0.06	0.07	0.10	0.17
...										
10 (high signal)	0.17	0.10	0.08	0.06	0.06	0.06	0.07	0.08	0.11	0.21

Between week t and week $t+52$

Portfolio in	Portfolio in week $t+52$									
week t	1	2	3	4	5	6	7	8	9	10
1 (low signal)	0.11	0.11	0.10	0.09	0.09	0.09	0.09	0.10	0.11	0.11
...										
10 (high signal)	0.11	0.11	0.10	0.09	0.09	0.09	0.09	0.10	0.11	0.11

Table 10
Weekly portfolios sorted on the equal-weighted leader signal and the past week's return, within 36 industries, 1980-2011

This table presents weekly four-factor alphas of portfolios sorted every week and within each of the 36 industries first into leader-signal quintiles and then into further quintiles based on the past week's return. Leaders for each stock are identified using weekly 52-week rolling regressions, as described in the text. The set of stocks is limited to those that have traded on the last day of the previous week, that were priced above \$5 per share in 2011 inflation-adjusted dollars, and that had leaders. Portfolio returns are equal-weighted in Panel A and value-weighted in Panel B. Newey-West-adjusted t-statistics are reported in parentheses.

Panel A: Equal-weighted portfolios

Signal	Prior week's return quintile						
quintile	1	2	3	4	5	5-1	
1	0.41%	-0.04%	-0.19%	-0.35%	**-0.78%**	-1.19%	
	(11.43)	(-1.80)	(-8.43)	(-13.2)	(-18.4)	(-18.74)	
2	0.52%	0.10%	-0.02%	-0.14%	-0.43%	-0.95%	
	(18.43)	(5.03)	(-1.08)	(-7.05)	(-15.4)	(-20.41)	
3	0.55%	0.17%	0.05%	-0.05%	-0.33%	-0.87%	
	(18.70)	(7.71)	(2.57)	(-2.93)	(-14.2)	(-20.07)	
4	0.66%	0.26%	0.11%	0.03%	-0.28%	-0.94%	
	(19.75)	(11.98)	(5.52)	(1.35)	(-12.4)	(-20.74)	
5	**0.91%**	0.37%	0.19%	0.05%	-0.34%	-1.25%	
	(18.52)	(13.22)	(7.56)	(2.08)	(-11.3)	(-19.99)	51-15
5-1	0.50%	0.41%	0.38%	0.39%	0.44%		**1.69%**
	(10.90)	(11.88)	(10.84)	(11.40)	(10.63)		(20.51)

Panel B: Value-weighted portfolios

Signal	Prior week's return quintile						
quintile	1	2	3	4	5	5-1	
1	0.25%	0.01%	-0.13%	-0.26%	**-0.53%**	-0.78%	
	(5.74)	(0.26)	(-3.87)	(-8.16)	(-11.18)	(-11.98)	
2	0.32%	0.09%	0.01%	-0.14%	-0.32%	-0.64%	
	(9.16)	(3.48)	(0.49)	(-5.50)	(-9.37)	(-11.74)	
3	0.41%	0.16%	0.04%	-0.12%	-0.31%	-0.71%	
	(10.27)	(5.97)	(1.66)	(-4.65)	(-9.68)	(-12.47)	
4	0.41%	0.20%	0.04%	-0.10%	-0.24%	-0.65%	
	(10.42)	(7.26)	(1.40)	(-4.20)	(-7.92)	(-12.58)	
5	**0.49%**	0.27%	0.11%	-0.01%	-0.27%	-0.77%	
	(10.43)	(7.65)	(3.72)	(-0.25)	(-6.81)	(-12.30)	51-15
5-1	**0.24%**	0.26%	0.25%	0.26%	0.26%		**1.03%**
	(4.16)	(5.18)	(5.14)	(5.84)	(4.35)		(13.73)

Table 11
Alternative methods for aggregating weekly leader signals, various weekly lags

This table presents weekly four-factor alphas of the return differentials between the top- and bottom-decile portfolios formed based on month $t - Lag$ leader signals within each of the 36 industries. In the first two columns, the leader signal is aggregated as in Table 9. In the next six columns, the leader signals is aggregated non-parametrically by multiplying the past-months leader returns by the sign($\hat{b}_3$) from weekly leader regressions and weighting them as described above each set of results, corresponding to Table 7. The number of weeks skipped before portfolios are formed is indicated in each row heading. Newey-West-adjusted t-statistics are reported in parentheses.

Four-Factor Alphas of the Return Differential (Portfolio 10−Portfolio 1)

How the leader signal is computed									
Parametric		Non-parametric							
equal-weighted		equal-weighted		value-weighted		$\|t$-stat$\|$-weighted		$\|\hat{b}_3\|$-weighted	
Weighting method for portfolio returns									
EW	VW	EW	VW	EW	VW	EW	VW	EW	VW
$Lag = 0$ weeks									
0.53%	0.29%	0.56%	0.24%	0.22%	0.15%	0.57%	0.26%	0.50%	0.22%
(12.54)	(6.30)	(14.73)	(6.61)	(9.39)	(4.26)	(14.92)	(6.97)	(13.40)	(6.32)
$Lag = 1$ week									
0.21%	0.11%	0.22%	0.15%	0.11%	0.11%	0.22%	0.14%	0.21%	0.12%
(7.30)	(2.28)	(8.83)	(4.03)	(5.63)	(3.14)	(8.58)	(3.75)	(8.39)	(3.40)
$Lag = 2$ weeks									
0.13%	0.07%	0.14%	0.06%	0.02%	0.02%	0.14%	0.05%	0.13%	0.06%
(5.10)	(1.53)	(6.63)	(1.69)	(1.28)	(0.68)	(6.45)	(1.21)	(6.51)	(1.70)
$Lag = 3$ weeks									
0.07%	0.08%	0.08%	0.06%	0.04%	0.03%	0.08%	0.04%	0.07%	0.05%
(2.49)	(1.47)	(3.46)	(1.78)	(2.22)	(1.04)	(3.85)	(1.03)	(3.06)	(1.53)
$Lag = 4$ weeks									
0.03%	0.01%	-0.02%	0.00%	-0.02%	0.00%	0.03%	0.03%	0.03%	0.03%
(0.89)	(0.11)	(-0.84)	(0.14)	(-0.84)	(0.14)	(1.05)	(0.74)	(1.03)	(0.75)

Table 12
Portfolios double sorted on turnover and the weekly leader signal, at various weekly lags

This table presents weekly four-factor alphas of equal-weighted return differentials between top and bottom weekly-signal decile portfolios. Within each of the 36 industries, stocks are first sorted into quintiles based on their turnover over the past 12 months and then on the leader signal. The number of weeks skipped between the week in which turnover and leader signals are calculated and the week in which portfolios are formed is indicated in each row heading. Portfolios are formed based on month $t-Lag$ leader signals, corresponding to Table 9. Newey-West-adjusted t-statistics are reported in parentheses. All cells insignificant at the 10% level are shaded in grey.

Four-factor alphas of the return differentials (decile 10−decile 1)

Stock turnover quintile				
1 (low)	2	3	4	5 (high)
$Lag = 0$ weeks				
0.63%	0.59%	0.47%	0.53%	0.38%
(13.38)	(13.29)	(9.65)	(11.20)	(5.98)
$Lag = 1$ week				
0.25%	0.24%	0.24%	0.25%	0.11%
(5.53)	(5.73)	(5.48)	(5.27)	(1.81)
$Lag = 2$ weeks				
0.10%	0.19%	0.07%	0.18%	0.09%
(2.35)	(4.39)	(1.66)	(3.79)	(1.45)
$Lag = 3$ weeks				
0.12%	0.09%	0.08%	0.05%	0.08%
(2.54)	(2.11)	(1.87)	(1.17)	(1.33)
$Lag = 4$ weeks				
0.05%	0.12%	-0.03%	-0.02%	-0.07%
(1.09)	(2.90)	(-0.73)	(-0.44)	(-1.05)
$Lag = 5$ weeks				
0.05%	0.08%	0.10%	0.05%	0.07%
(1.11)	(1.89)	(2.28)	(1.01)	(1.07)
$Lag = 6$ weeks				
-0.01%	-0.03%	-0.05%	0.04%	-0.07%
(-0.12)	(-0.70)	(-1.01)	(0.86)	(-1.11)

Table 13
Cross-sectional regressions

This table presents the results of Fama and MacBeth (1973) regressions of stock returns on a set of explanatory variables lagged by one month in Panels A and B and by one week in Panel C. In Panels A and B, all variables are computed at monthly frequencies. In Panel C, superscript w indicates leader signals and returns computed at weekly frequencies; all other variables are computed at monthly frequencies at the end of the previous month. The explanatory variables are described in Appendix A2. The sample consists of all common stocks of U.S.-incorporated firms that traded at the end of the previous month (the previous week in Panel C) and had leaders. *med* is the median value of each variable. Newey-West-adjusted t-statistics are reported in parentheses.

Panel A: Sample period: January 1929 - December 2011

				Subsamples				Leaders from 36-mo. rolling windows*
				Size>med	Turn>med	Age>med		
×100	Ret_t	Ret_t	Ret_t	Ret_t	Ret_t	Ret_t	$Ret_t - Ind.Ret_t$	Ret_t
Leader Signal $(\text{EW})_{t-1}$	10.959^a	10.464^a	12.190^a	10.111^a	8.026^a	8.49^a	9.408^a	24.000^a
	(6.81)	(6.80)	(5.43)	(5.69)	(4.62)	(4.67)	(6.94)	(8.86)
Ret_{t-1}	-6.653^a	-7.439^a	-5.887^a	-4.399^a	-5.598^a	-7.226^a	-7.367^a	-7.910^a
	(-12.45)	(-13.74)	(-11.21)	(-8.68)	(-12.12)	(-12.81)	(-14.46)	(-13.84)
Momentum	0.110^a	0.106^a	0.060^a	0.062^b	0.093^a	0.056^a	0.101^a	0.124^a
	(2.86)	(2.84)	(2.51)	(2.95)	(3.09)	(2.88)	(2.98)	(3.10)
Ind. Ret_{t-1}		20.251^a	18.421^a	18.070^a	15.634^a	15.412^a		15.762^a
		(13.13)	(13.79)	(11.79)	(13.83)	(14.39)		(15.56)
Size	0.000^c	0.000	0.000^c	0.000	0.000^b	0.000^b	0.000^c	0.000^b
	(-1.77)	(-1.85)	(-1.59)	(-1.60)	(-2.34)	(-2.27)	(-1.86)	(-2.07)
Leader Signal $(\text{EW})_{t-1}$			-15.555^c					
$\times Ret_{t-1}$			(-1.67)					

Panel B: Sample period: August 1963 - December 2011

			Leaders from 36-mo. rolling windows*	
$\times 100$	Ret_t	Ret_t	Ret_t	Ret_t
Leader Signal $(\mathrm{EW})_{t-1}$	6.354^a	7.684^a	18.802^a	22.120^a
	(3.15)	(3.39)	(6.53)	(6.39)
Ret_{t-1}	-6.000^a	-5.063^a	-6.333^a	-5.241^a
	(-11.25)	(-10.09)	(-11.72)	(-10.64)
Momentum	0.050^a	0.047^a	0.052^a	0.052^a
	(4.07)	(3.47)	(4.48)	(4.06)
Ind. Ret_{t-1}	17.469^a		15.212^a	
	(10.19)		(13.02)	
Size	0.000^b	0.000^b	0.000^b	0.000^b
	(-2.49)	(-2.27)	(-2.42)	(-2.17)
Book/Market	0.204^a	0.113^a	0.191^a	0.197^a
	(3.63)	(3.94)	(3.26)	(3.23)
Beta	0.126		0.159	
	(1.20)		(1.42)	
Illiq	0.044^b	0.049^b	0.038^c	0.039^b
	(2.21)	(2.44)	(1.91)	(1.97)
IVOL	-0.119^b	-0.130^b	-0.119^b	-0.106^c
	(-2.57)	(-2.20)	(-2.57)	(-1.73)

* The sample period for this regression is two years shorter.

Panel C: Weekly returns; sample period: January 1980 - December 2011

$\times 100$	Ret_t^w	Ret_t^w	Ret_t^w	Ret_t^w	Ret_t^w	$Ret_t^{w\S}$	Ret_t^w	$Ret_t^{w\P}$	Ret_t^w	Ret_t^w	Ret_t^w	Ret_t^w
Leader Signal $(\text{EW})_{t-1}^w$	70.187^a	31.581^a	42.886^a	34.815^a	33.334^a	32.316^a	38.575^a	38.816^a	44.986^a	53.295^a	33.311^a	29.668^a
	(14.15)	(10.37)	(12.17)	(11.15)	(10.77)	(8.22)	(11.49)	(11.78)	(12.04)	(13.23)	(9.05)	(9.90)
Ret_{t-1}^w		-8.528^a	-8.553^a	-8.700^a	-8.551^a	-5.849^a	-8.535^a	-8.382^a	-8.521^a	-8.521^a	-8.534^a	-8.560^a
		(-28.27)	(-27.23)	(-27.50)	(-28.62)	(-16.13)	(-28.24)	(-25.13)	(-28.24)	(-28.27)	(-28.26)	(-28.38)
Ind. Ret_{t-1}^w		12.714^a	13.327^a	12.754^a	12.693^a	6.726^a	12.678^a	11.592^a	12.691^a	12.661^a	12.711^a	12.576^a
		(15.80)	(16.30)	(15.89)	(15.88)	(6.22)	(15.81)	(13.68)	(15.87)	(15.76)	(15.79)	(15.75)
Leader Signal $(\text{EW})_{t-1}^w$ interacted with:												
$\times$ Ret_{t-1}^w				-172.794^a								
				(-5.17)								
$\times$ $\mathbb{1}\{Qtr.EarnAnn.\}_{t-1}$					-2.871^c							
					(-1.73)							
$\times \mathbb{1}\{News\}_{t-1}$						-2.843						
						(-1.62)						
$\times \mathbb{1}\{Inst.Ownership > med\}_{t-1}$							-33.400^a					
							(-7.32)					
$\times \mathbb{1}\{AnalystCoverage > med\}_{t-1}$								-37.784^a				
								(-7.09)				
$\times \mathbb{1}\{Size > med\}_{t-1}$									-36.928^a			
									(-7.96)			
$\times \mathbb{1}\{Turnover > med\}_{t-1}$										-37.217^a		
										(-6.06)		
$\times \mathbb{1}\{Age > med\}_{t-1}$											-2.591	
											(-0.64)	
Leader Signal $(\text{EW})_{t-1}^{monthly}$												1.779^a
												(3.27)
Extended Controls[†]	Yes	Yes	No[‡]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

[a], [b], and [c] indicate significance at the 1%, 5%, and 10% levels, respectively.

[†] Extended controls include Size, Book/Market, Momentum, Illiq, and IVOL and the dummies in the interaction terms when applicable.

[‡] Controls include Size and Book/Market.

[§] The regression sample period is April 1996 - December 2011.

[¶] The regression sample period is December 1983 - December 2011.

Table 14
Short selling in response to the leader signal

This table presents the results of the Fama-MacBeth regression of the change in utilization (defined as the number of shares on loan relative to the total number of shares available to be loaned out for short selling) on indicator functions of whether a stock enters or exits the bottom decile of the concurrent leader signal, industry return, or the top decile of the concurrent own return on each Friday (relative to the Friday or the week before):

$$\begin{aligned} \Delta utilization_{it} &= \alpha + \beta_1 \cdot \mathbb{1}_{it}\{\text{Enters bottom signal decile}\} + \beta_2 \cdot \mathbb{1}_{it}\{\text{Exits bottom signal decile}\} \\ &+ \gamma_1 \cdot \mathbb{1}_{it}\{\text{Enters bottom ind. ret. decile}\} + \gamma_2 \cdot \mathbb{1}_{it}\{\text{Exits bottom ind. ret. decile}\} \\ &+ \mu_1 \cdot \mathbb{1}_{it}\{\text{Enters top return decile}\} + \mu_2 \cdot \mathbb{1}_{it}\{\text{Exits top return decile}\} + \epsilon_{it} \end{aligned}$$

Assuming that short sellers would set up short positions on the following Monday, utilization changes are calculated between the Thursday of the following week and the preceding Thursday, in order to account for the "t+3" security transaction settlement rule. The sample is trimmed at the top and bottom 1% of utilization, and stocks priced at less than \$5/share in the 2011 inflation-adjusted dollars are dropped. The sample period is October 8, 2008, to December 30, 2011. Newey-West-adjusted t-statistics are reported in parentheses.

Specification	(1)	(2)	(3)
α	-0.025	-0.026	-0.023
	(-0.85)	(-0.88)	(-0.82)
β_1	0.075^b	0.076^b	0.084^b
	(2.05)	(2.06)	(2.49)
β_2	-0.033	-0.032	0.020
	(-0.97)	(-0.95)	(0.56)
γ_1	-0.021	-0.021	-0.014
	(-0.68)	(-0.66)	(-0.45)
γ_2	0.051	0.052	0.078^a
	(1.60)	(1.63)	(2.95)
μ_1	0.464	0.465	0.544
	(1.18)	(1.18)	(1.37)
μ_2	0.263^c	0.263^c	0.171
	(1.74)	(1.74)	(1.15)

Data selection criteria:
(1): All data included.
(2): Excludes stocks with quarterly earnings announcements anticipated next week.
(3): Excludes stocks with average loan fees in the top three fee buckets.

Table 15
Determinants of leadership, April 1997 - December 2011

This table presents the results of regressions of the number of followers (including zeros for the stocks that have no followers) on a set of explanatory variables, which are described in Appendices A1 and A2. The sample consists of all common shares of U.S.-incorporated firms. Panel A reports results for monthly-frequency leaders identified using 12-month rolling regressions and Panel B for weekly-frequency leaders identified using 52-week rolling regressions. In regression specifications (1)-(3), all highly relevant news and, in models (4)-(6), all highly relevant corporate news are counted over the previous 12-month period; the values of all explanatory variables are averaged over the previous 12 months. Panel C reports pairwise correlations between the control variables. Regressions are estimated with quasi-maximum likelihood and standard errors are clustered at the firm level. z-statistics are reported in parentheses.

Panel A: Leadership is determined with monthly rolling regressions

	All highly relevant news			Highly relevant corp. events				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
News ($\times 10^2$)	0.0075^a	0.0056^a	0.0112^a	0.0100^a	0.0074^a	0.0123^a		
	(6.30)	(4.19)	(6.15)	(6.50)	(4.58)	(5.40)		
News^2 ($\times 10^4$)			-0.0002^a			-0.0002^b		
			(-3.81)			(-2.46)		
Inst. Ownership		0.0359^a	0.0351^a		0.358^a	0.0351^a	0.0368^a	
		(6.63)	(6.45)		(6.60)	(6.45)	(6.81)	
An. Coverage		0.0020^a	0.0018^a		0.0020^a	0.0019^a	0.0022^a	
		(6.83)	(5.84)		(6.88)	(6.24)	(7.67)	
Size ($\times 10^6$)		-0.4893^a	-0.5741^a		-0.4730^a	-0.5320^a	-0.2241^b	0.2730^b
		(-3.98)	(-4.64)		(-3.91)	(-4.38)	(-2.09)	(2.55)
Turnover		0.0003	0.0002		0.0002	0.0001	0.0006	
		(0.38)	(0.24)		(0.25)	(0.16)	(0.93)	
Book/Market		0.0031^b	0.0031^b		0.0031^b	0.0031^b	0.0032^b	
		(2.00)	(2.00)		(2.00)	(2.01)	(2.01)	
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Leadership is determined with weekly rolling regressions

	All highly relevant news			Highly relevant corp. events				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
News ($\times 10^2$)	0.0066^a	0.0039^a	0.0070^a	0.0099^a	0.0070^a	0.0117^a		
	(4.84)	(2.66)	(2.96)	(5.24)	(3.65)	(3.59)		
News^2 ($\times 10^4$)			-0.0001^b			-0.0002^b		
			(-2.28)			(-2.28)		
Inst. Ownership		0.0483^a	0.0478^a		0.480^a	0.0472^a	0.0489^a	
		(7.84)	(7.74)		(7.78)	(7.64)	(7.95)	
An. Coverage		0.0020^a	0.0019^a		0.0020^a	0.0018^a	0.0021^a	
		(6.07)	(5.70)		(6.00)	(5.63)	(6.42)	
Size ($\times 10^6$)		-0.4242^a	-0.4742^a		-0.4743^a	-0.5335^a	-0.2399^c	0.2855^b
		(-3.02)	(-3.28)		(-3.24)	(-3.53)	(-1.87)	(2.15)
Turnover		0.0005	0.0004		0.0003	0.0002	0.0007	
		(0.66)	(0.58)		(0.40)	(0.30)	(1.02)	
Book/Market		-0.0004	-0.0005		-0.0005	-0.0005	-0.0004	
		(-0.56)	(-0.58)		(-0.58)	(-0.60)	(-0.50)	
Industry dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

^{a}, ^{b}, and ^{c} indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel C: Correlations between control variables†

	News (all)	News (corp.)	Inst. Own.	An. Cov.	Size	Turnover	Book/Market
News (all)	1.0000	0.94071	0.26187	0.4077	0.5072	0.1870	-0.0194
		(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)
News (corp.)		1.0000	0.2659	0.3890	0.4671	0.1982	-0.0150
			(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)
Inst. Own.			1.0000	0.4773	0.1073	0.2586	-0.0637
				(<.0001)	(<.0001)	(<.0001)	(<.0001)
An. Cov.				1.0000	0.4323	0.2831	-0.1108
					(<.0001)	(<.0001)	(<.0001)
Size					1.0000	0.0180	-0.0399
						(<.0001)	(<.0001)
Turnover						1.0000	-0.0745
							(<.0001)

†p-values reported in parentheses.

Table A1
Industries

This table presents the monthly average percentages of stocks in the industries in our sample. The sample consists of common shares of U.S.-incorporated firms (stocks with share codes 10 or 11) that traded on the last day of the previous month and were priced above $5 per share in 2011 inflation-adjusted dollars. The averages are computed using only the months that have at least one stock observation in a given industry. The sample period is 1929-2011.

Industry	% of stocks
Steam Supply	0.04%
Nonmetalic Minerals, Except Fuels	0.26%
Agriculture, Forestry, and Fishing	0.26%
Other	0.29%
Sanitary Services	0.31%
Public Administration	0.37%
Furniture and Fixtures	0.44%
Lumber and Wood Products	0.55%
Leather and Leader Products	0.64%
Radio and Television Broadcasting	0.76%
Telephone and Telegraph Communication	0.81%
Construction	0.85%
Tobacco Products	1.00%
Miscellaneous Manufacturing Industries	1.11%
Apparel and other Textile Products	1.19%
Printing and Publishing	1.25%
Rubber and Miscellaneous Plastics Products	1.28%
Paper and Allied Products	1.67%
Textile Mill Products	1.68%
Stone, Clay and Glass Products	1.74%
Mining	1.86%
Oil and Gas Extraction	2.18%
Wholesale	2.26%
Petroleum and Coal Products	2.71%
Fabricated Metal Products	2.81%
Instruments and Related Products	2.91%
Primary Metal Industries	4.84%
Food and Kindred Products	5.20%
Transportation	5.40%
Electric, Gas, and Water Supply	5.43%
Transportation Equipment	5.73%
Electrical and Electronic Equipment	5.87%
Chemicals and Allied Products	6.13%
Machinery, Except Electrical	6.38%
Services	6.88%
Retail Stores	7.09%
Finance, Insurance, and Real Estate	10.42%

Table A2
A representative news flow

This table reproduces a subsample of news stories for Cisco Systems Inc. (CSCO) reported in the Thomson-Reuters News Analytics dataset. The Reuters Primary News Access Code (PNAC) is a unique identifier for a distinct news topic for a firm. Item type could be an Article, Alert, Append, or Overwrite. The relevance score is a number between 0 and 1 that quantifies how relevant a news item is for a given firm.

Date	Time	PNAC	Item type	Relevance	Headline
1-Oct-09	6:36:51	nL1456988	ARTICLE	1	UPDATE 1-Cisco to buy Tandberg for $3 billion
1-Oct-09	8:03:30	nL1456988	ARTICLE	1	UPDATE 2-Cisco to buy Tandberg for $3 billion
1-Oct-09	8:03:30	nL1456988	APPEND	1	UPDATE 2-Cisco to buy Tandberg for $3 billion
1-Oct-09	8:14:37	nL1456988	ARTICLE	1	UPDATE 2-Cisco to buy Tandberg for $3 billion
1-Oct-09	8:14:37	nL1456988	APPEND	1	UPDATE 2-Cisco to buy Tandberg for $3 billion
1-Oct-09	10:46:06	nL1456988	ARTICLE	1	UPDATE 3-Cisco to buy video meetings firm Tandberg for $3 bln
1-Oct-09	10:46:06	nL1456988	APPEND	1	UPDATE 3-Cisco to buy video meetings firm Tandberg for $3 bln
1-Oct-09	10:46:06	nL1456988	APPEND	1	UPDATE 3-Cisco to buy video meetings firm Tandberg for $3 bln
1-Oct-09	11:02:21	nL1456988	ARTICLE	1	REFILE-UPDATE 3-Cisco to buy video firm Tandberg for $3 bln
1-Oct-09	11:02:21	nL1456988	APPEND	1	REFILE-UPDATE 3-Cisco to buy video firm Tandberg for $3 bln
1-Oct-09	13:50:02	nL1456988	ARTICLE	1	UPDATE 4-Cisco bets on video again with $3 bln Tandberg buy
1-Oct-09	13:50:02	nL1456988	APPEND	1	UPDATE 4-Cisco bets on video again with $3 bln Tandberg buy
1-Oct-09	18:48:34	nL1456988	ARTICLE	1	UPDATE 5-Cisco bets on video growth with $3 bln Tandberg bid
1-Oct-09	18:48:34	nL1456988	APPEND	1	UPDATE 5-Cisco bets on video growth with $3 bln Tandberg bid
1-Oct-09	19:07:24	nL1456988	ARTICLE	1	REFILE-UPDATE 5-Cisco bets on video growth with Tandberg bid
1-Oct-09	19:07:24	nL1456988	APPEND	1	REFILE-UPDATE 5-Cisco bets on video growth with Tandberg bid
1-Oct-09	21:29:20	nL1456988	ARTICLE	1	UPDATE 6-Cisco bets on video growth with Tandberg bid
1-Oct-09	21:29:20	nL1456988	APPEND	1	UPDATE 6-Cisco bets on video growth with Tandberg bid
5-Oct-09	14:17:38	nL527951	ALERT	1	GENEVA-CISCO CEO SAYS ALMOST NO PRODUCT OVERLAP WITH TANDBERG
5-Oct-09	14:17:38	nL527951	ALERT	1	GENEVA-CISCO CEO: SEES ACQUISITIONS IN THE INDUSTRY ”HEATING UP”
5-Oct-09	14:17:38	nL527951	ARTICLE	1	BRIEF-Cisco CEO sees more acquisitions in industry
2-Nov-09	11:09:36	nL2335867	ARTICLE	1	UPDATE 1-Cisco’s bid for Tandberg ”fair”, says consultant
2-Nov-09	13:54:29	nL2335867	ARTICLE	1	UPDATE 2-Cisco’s bid for Tandberg fair, consultancy says
2-Nov-09	20:23:45	nL2335867	ARTICLE	1	UPDATE 3-Cisco defends Tandberg bid as ”very good price”
3-Nov-09	12:14:20	nL3595357	ARTICLE	1	POLL-Cisco’s bid for Tandberg to fall short-analysts
3-Nov-09	12:50:35	nL3604734	ARTICLE	0.111803	FACTBOX-Top players in ailing mobile network gear market
3-Nov-09	12:50:35	nL3604734	APPEND	0.111803	FACTBOX-Top players in ailing mobile network gear market

Table A3
News count statistics, April 1996 - December 2011

This table presents statistics on news coverage from the Thomson-Reuters News Analytics dataset. News items are defined as highly relevant if the Reuters' relevance score is equal to one. Unique news counts category counts only the number of distinct news strands (identified by PNAC), ignoring Alerts, Appends, Overwrites, and multiple Articles within the same news topic code. Corporate events category includes only news items that cover new corporate developments.

	% firms	All items				Unique news topics				All highly relevant items				Highly relevant corp. events			
Year	covered	mean	med	5%	95%	mean	med	5%	95%	mean	med	5%	95%	mean	med	5%	95%
1996	16.03%	1.74	0.00	0.00	6.00	0.96	0.00	0.00	4.00	0.92	0.00	0.00	4.00	0.05	0.00	0.00	0.00
1997	38.58%	26.94	0.00	0.00	97.00	14.50	0.00	0.00	52.00	14.57	0.00	0.00	59.00	2.27	0.00	0.00	11.00
1998	40.84%	26.21	0.00	0.00	92.00	15.15	0.00	0.00	57.00	13.48	0.00	0.00	53.00	4.86	0.00	0.00	22.00
1999	42.40%	26.96	0.00	0.00	95.00	16.86	0.00	0.00	60.00	13.37	0.00	0.00	53.00	6.35	0.00	0.00	30.00
2000	43.70%	28.60	0.00	0.00	102.00	19.80	0.00	0.00	70.00	13.97	0.00	0.00	59.00	6.62	0.00	0.00	31.00
2001	52.52%	31.52	1.00	0.00	118.00	22.46	1.00	0.00	81.00	16.01	1.00	0.00	68.00	8.78	0.00	0.00	41.00
2002	54.25%	36.63	2.00	0.00	143.00	25.20	1.00	0.00	92.00	18.42	1.00	0.00	79.00	10.33	0.00	0.00	48.00
2003	74.72%	85.05	27.00	0.00	327.50	70.85	25.00	0.00	270.00	53.12	22.00	0.00	202.00	33.86	13.00	0.00	133.50
2004	81.52%	95.77	33.00	0.00	352.00	75.48	29.00	0.00	276.50	62.21	28.00	0.00	226.00	41.67	17.00	0.00	155.50
2005	88.85%	131.36	50.00	0.00	484.00	101.01	40.00	0.00	370.00	89.67	43.00	0.00	324.00	62.58	28.00	0.00	231.00
2006	94.66%	163.25	77.00	0.00	552.00	117.61	48.00	0.00	417.00	113.35	69.00	0.00	367.00	88.13	56.00	0.00	279.00
2007	97.20%	187.84	83.00	6.00	623.00	137.31	53.00	4.00	467.00	121.56	73.00	5.00	379.00	95.39	60.00	3.00	298.00
2008	98.41%	223.27	78.00	8.00	773.00	163.51	49.00	3.00	620.00	133.19	66.00	7.00	460.00	104.98	52.00	5.00	372.00
2009	98.56%	228.90	85.00	10.00	879.00	156.02	50.00	5.00	645.00	138.47	69.00	9.00	492.00	113.32	55.00	7.00	411.00
2010	98.92%	214.23	95.00	15.00	708.50	138.94	54.00	7.00	491.00	137.78	79.00	14.00	432.00	115.75	66.00	13.00	363.50
2011	98.63%	215.11	96.00	16.00	676.00	147.54	60.00	8.00	475.00	131.75	77.00	15.00	389.00	130.51	77.00	15.00	386.00
Avg.	64.75%	92.67	14.00	0.00	370.00	65.70	10.00	0.00	260.00	57.29	12.00	0.00	232.00	43.02	7.00	0.00	179.00