Title
The Cross-Section of Analyst Recommendations

Authors
Sorescu, Sorin
Subrahmanyam, Avanidhar

Publication Date
2004-01-09

January 9, 2004

The Cross-Section of Analyst Recommendations

Sorin Sorescu[*] and Avanidhar Subrahmanyam[**]

[*] Mays Business School, Texas A&M University.

[**] Anderson Graduate School of Management, University of California at Los Angeles.

The authors acknowledge the use of I/B/E/S data, a service of Thomson Financial, in the preparation of this study. This data has been provided as part of a broad academic program to encourage research on analysts' expectations.

Abstract

The Cross-Section of Analyst Recommendations

We analyze the relation between analyst attributes (years of experience, reputation of the analysts' brokerage houses) and the short- and long-term price reactions to recommendations made by the analysts. We find that in the long-term, the recommendation changes of highly experienced analysts outperform those of low-experience ones. In addition, investors appear to overreact to dramatic upgrades of low-ability analysts, and underreact to small upgrades by high-ability analysts. These results are consistent with the Griffin and Tversky (1992) argument that agents place too much emphasis on the strength of the signal (the dramatic nature of the event) and insufficient emphasis on the weight (the ability of the analyst making the recommendation). The study helps promote an understanding of the analyst industry and its interaction with the investing population.

Introduction

A principal way in which information is disseminated to financial market participants is through the opinions of brokerage house analysts. Considerable amounts of resources are expended in employing analysts, and the popular press awaits the stock recommendations of the relatively better-known analysts with keen anticipation. These recommendations are also posted prominently on popular finance websites such as finance.yahoo.com and moneycentral.msn.com. Conflicts of interest and bias allegations, however, have plagued the analyst industry in recent years. As an anecdotal example of this phenomenon, in October 2001, ten of the 17 analysts following Enron had a"strong buy" rating and five others had a "buy" rating on this stock. The optimism persisted in spite of a 50% loss in the firms market capitalization during the quarter preceding bankruptcy.[1] In a rigorous analysis, Michaely and Womack (1999) find that while analyst recommendations do have some return forecasting ability, they appear to be tainted with conflicts of interest and/or biases, in that analysts tend to favor companies whose equity offerings are marketed by the investment banks to which they are affiliated.

The above issues concerning the analyst industry notwithstanding, prior research (e.g., Irvine, 2003) shows significant return reactions to changes in analyst following, indicating that analysts, in spite of their biases, can influence the cost of capital for a company. In addition, Brennan and Subrahmanyam (1995) show that analysts influence

---

[1] *Wall Street Journal*, "Most Analysts Remain Plugged in to Enron," October 26, 2001, page C1.

liquidity, which, in turn, affects expected returns (Amihud and Mendelson, 1986). Thus, obtaining a sound understanding of the analyst industry and its linkage to financial market prices is important both from an academic as well as a practical viewpoint.

Considerable research has analyzed the information content of analysts' earnings forecasts as well as their summary recommendations about stock investments. Studies such as Barber and Loeffler (1993), Dimson and Marsh (1984), and Stickel (1995) all analyze the predictive content of analyst pronouncements for future stock returns, and, in a careful analysis, Womack (1995) finds evidence of significant abnormal returns following shifts in analysts' opinions about stocks. Barber, Lehavy, McNichols, and Trueman (2001) confirm the return forecasting power of analyst recommendations but argue that after accounting for transaction costs, the abnormal returns obtainable by following analyst recommendations are negligible. Hong and Kubik (2003) suggest that the accuracy with which an analyst forecasts earnings is associated with career moves (e.g., the possibility of being employed at a high-reputation brokerage house).

While the importance of analyst recommendations as a principal avenue of information production is clear from the literature, the notion that analysts attributes vary in the cross-section appears to have received less emphasis than it deserves. As in every profession, there is heterogeneity among agents who become analysts, and this likely influences their avenue of employment as well as investor perceptions. One would expect investors to be aware of at least some characteristics of analysts (such as their experience) when reacting to the opinions rendered by these agents. Our work is concerned with the functional form of the cross-sectional relation between analyst attributes and return reactions in the short- and long-term. More specifically, we explore whether high values of indicators that proxy for analyst ability, such as years of experience and the reputation of the analyst's brokerage house, are associated with superior return forecasting ability in the long-term.

Of course, our analysis of the cross-sectional variation in long-term return performance of analyst recommendations is relevant only if the market does not immediately incorporate information contained in analyst recommendations. In this regard, Womack's (1995) paper suggests that investor reaction to analyst recommendations takes a period of at least twelve months to be completely incorporated into stock prices. This suggests the possibility that investors face cognitive constraints in making investment decisions in response to analyst recommendations. Our hypotheses are therefore motivated by appealing to recent research in the psychology of decision-making. In particular, the analysis of Griffin and Tversky (1992) indicates that the degree of confidence varies across different types of information. Agents place undue emphasis on the strength of information (e.g., the warmth of a recommendation letter), and too little emphasis on its weight (the credibility of the letter writer). In the context of the market for financial analysis, we propose that "strength" translates into the dramatic nature of a pronouncement (e.g., changing of a recommendation from a strong sell to strong buy) whereas “weight” translates into the experience of the analyst or the reputation of the brokerage house to which the analyst belongs.

Our results show a clear pattern that more experienced analysts show a greater ability to forecast stock returns by way of their recommendations. Therefore, there is reliable evidence experience counts in the analyst industry, which validates its use as a proxy for “weight” in our empirical analysis. In addition, the market appears to underreact to small recommendation changes by experienced analysts and overreact to large recommendation changes by inexperienced analysts. These results support the Griffin and Tversky (1992) hypotheses.

Our analysis sheds light on the notion that the quality of information signals can vary in the cross-section of analysts, so that high-ability analysts appear to draw signals that are more strongly associated with future performance of the stock. Our results not only provoke thought on how resources should be allocated to compensate analysts, but also have implications for designing investment strategies. Our analysis also sheds light on the specific types of events that are likely to lead to underreaction (relatively mundane, high-weight events) vs. overreaction (dramatic, low-weight events). Finally, since our analysis tests ex ante implications derived from psychological evidence, it addresses Fama’s (1998) statement that behavioral literature is often subject to “model mining” by providing ex post theoretical explanations of market phenomena.

The remainder of this paper is organized as follows. Section 1 details our hypotheses while Section 2 describes the data. Section 3 discusses the results. Section 4 concludes.

1 Hypotheses

To develop our hypotheses, we consider overreaction and underreaction in the context of Griffin and Tversky (1992), using the simplest possible setting. Suppose there is a security with a final payoff of θ, where θ is normally distributed with mean $\bar{\theta}$ and variance v_θ. There is a signal, $\theta + \epsilon$, released by an analyst about the company prior to the disclosure of θ. The variable ϵ is also normally distributed with mean zero, and is independent of θ, with a variance v_ϵ. A finite mass of risk-neutral investors act as price-setters and are biased in processing this signal, in a manner clarified below. The posterior update, and the market price, denoted by P takes the form

$$P = (1-k)\bar{\theta} + k(\theta + \epsilon).$$

The rational Bayesian weight k is given by

$$k = \frac{v_\theta}{v_\theta + v_\epsilon}.$$

Since investors are biased, we assume that their calculated noise variance, v_r, differs from its true counterpart, v_ϵ. Thus, the investors' weight k' takes the form

$$k' = \frac{v_\theta}{v_\theta + v_r} + s.$$

Thus, investors miscalculate the weight of the evidence, represented by the variance of the noise in the signal, and also indulge in an additional miscalculation, which, for now, is assumed to be a linear function of the size of the signal, with a slope s. Later in this section, we will generalize this function, and clarify its relation with psychological evidence.

In this setting, the sign of the covariance of price changes with the signal $\mathrm{cov}(\theta - P, \theta + \epsilon)$ is given by

$$\mathrm{sgn}\left[\mathrm{cov}(\theta - P, \theta + \epsilon)\right] = \mathrm{sgn}\left[\frac{v_\theta}{v_\theta + v_r}(v_r - v_\epsilon) - s(v_\theta + v_\epsilon)\right].$$

It can be seen that if there are no biases, i.e., if $v_r = v_\epsilon$ and $s = 0$, then prices exhibit no underreaction or overreaction, so that the covariance is zero. If investors underreact to the weight ($v_\epsilon < v_r$), prices tend to drift. However, a positive s contributes to overreaction. The net effect depends on the magnitude of each type of bias.

In developing our hypotheses, we rely on recent research in psychology that explicitly addresses the updating of agents' beliefs in response to new information. A large literature shows that agents appear to be overconfident (see, for example, Odean, 1998, for a summary of the evidence), but such a phenomenon is not universal (Edwards, 1968, and Lichtenstein, Fischoff and Phillips, 1982). In turn, motivated by the evidence that agents appear to be underconfident at certain tasks and overconfident at others, Griffin and Tversky (1992) present experimental evidence in support of the argument that the degree of confidence depends on the nature of information to which an agent reacts. They propose that agents tend to overly focus on the strength of a signal and not attach enough importance to the weight of a signal. An example is that in assessing whether a coin is fair, agents focus too much on whether the sequence of tosses has an extreme outcome (all heads), and not enough on the sample size (how many tosses were performed). As an analogy closer to the academic world, Griffin and Tversky (1992) propose that agents are prone to attaching undue importance to the enthusiasm in a recommendation letter, and not enough importance to the credibility of the recommendation writer.

To more explicitly connect "strength" and "weight" to investor misreaction and the equilibrium price, and to model the phenomena document by Griffin and Tversky (1992), we consider a situation where s depends on the strength of the signal. More specifically, we assume that if $|\theta + \epsilon| > C$, where $C > 0$, then $s = s_1 > 0$ and $s = s_2 < s_1$ otherwise. This incorporates the notion that agents' expectations depends on the strength (extremeness) of the signal.

In the above case, the price is non-linear in the signal, and so analytic solutions are not possible. We use Monte Carlo simulations (one million pairs of θ and ϵ draws) to calculate the correlation between price changes and the signal realizations (i.e., the analog of the covariance above) for four cases that encompass high and low weight and strength. The base parameter values are $v_\theta = 1$, $v_r = 2$, $C = 1$, $s_1 = 0.4$, and $s_2 = 0.2$. The ex ante mean $\bar{\theta}$ is set to zero for convenience. For the high-weight (low-weight) case, v_ϵ is 0.6 (1.0).[2] The high (low) strength case is represented by $|\theta + \epsilon| > 1$ (< 1). The

---

[2] While the parameter values are chosen solely for illustrative purposes, the results are robust to varying the parameters within 50% of the indicated range.

four cases are denoted by HSHW (high-strength, high-weight), LSHW (low-strength, high-weight), HSLW, and LSLW. The correlations and conditional price changes yielded by the simulation are given in Table 1.

Please insert Table 1 here.

Conforming to intuition, conditional price change patterns match the correlation patterns, in that whenever the correlation is negative (positive), the expected price change is of a sign opposite to (same as that of) the signal. As can be seen from the table, since the investor bias is stronger for more extreme (high-strength) signals, the market overreacts to such signals. At the same time, the market underreacts to the weight (quality) of the signal. The net result is that prices experience reversals following high-strength, low-weight signals and drift following high-weight, low-strength signals.

In our empirical tests, we use number of years of experience and the reputation of the investment bank to which the analyst belongs as proxies for weight. The number of categories the recommendation change spans is used as a proxy for strength. Note that the validity of the weight proxy itself can be tested by considering whether, across all recommendation changes, analysts with high values of the weight proxies exhibit superior return forecasting ability.[3] All of the above arguments then lead to the following

---

[3] In other work that links agents' ability to performance within the finance arena, Chevalier and Ellison (1999) find that indicators of managerial ability (specifically, SAT scores) are positively related to managerial performance in mutual funds. As an indirect measure of differences in CEO ability across sectors, Palia (2000) finds that in his sample, 74% of CEOs in the manufacturing possessed an MBA degree from one of the top nine business schools, whereas only 33% of those in the utility sector

testable hypotheses.

1. The predictive content of recommendation changes by high-weight analysts will be greater than that of low-weight analysts.
2. There will be return drift (underreaction) following recommendation changes by high weight analysts who issue low-strength signals.
3. There will be return reversal (overreaction) following recommendation changes by low-weight analysts who issue high-strength signals.

We use available data on analyst attributes and recommendation changes to test the above hypotheses, as described in the next section.

2 Data

Our primary data come from the Institutional Brokers Estimate System (I/B/E/S) database, now part of Thomson Financial, which, in turn is available on WRDS (Wharton Research Data Services) for the period 1993-2000. The data are obtained from the recommendation detail file and broker translation files. These files include information on the investment bank with which the analyst is affiliated, as well as the recommendation rendered by each analyst. The categories of recommendations are "strong buy," "buy," "hold," "sell," and "strong sell." These opinions are often disseminated in the

---

possessed such a degree. We do not have access to such proxies for ability for security analysts and rely instead on experience and reputation measures.

popular press and are posted on popular finance websites such as those run by CNBC and Yahoo. We assume that the dissemination of the recommendations is broad enough to influence stock prices, and consider *changes* in recommendations from one category to another.

As detailed in the introduction, our goal is to consider the functional form of the stock price response to analyst attributes as well as the nature of the recommendation change. To reiterate, we use the Griffin and Tversky (1992) framework, which argues that agents place undue emphasis on the strength of information (e.g., the warmth of a recommendation letter), and too little emphasis on its weight (the credibility of the letter writer). As argued in the previous section, in the context of the market for financial analysis, we propose that "strength" translates into the dramatic nature of a switch in signal (e.g., strong sell to strong buy) whereas "weight" translates into the experience of the analyst or the reputation of the brokerage house to which the analyst belongs.

More specifically, our proxy for strength is the change in number of categories corresponding to each recommendation. A single category change is classified as "low strength" and all other changes are denoted "high strength." The notion is that a two-category or greater jump is more likely to be vivid and dramatic, and thereby attract the attention of agents.[4]

---

[4] We also experimented with three classes: a one-level change, a two-level change, and three-or-more-level changes. We did not see any differences in return reactions across the "two" and the

The two specific proxies for "weight" are the number of years of experience (as proxied by the number of days the analyst is present in the I/B/E/S detail file), and a numerical measure of the reputation of the investment bank that employs the analyst. As in any other profession, the supposition is that more senior analysts provide more reliable recommendations; this assumption is further examined below. Furthermore, it seems reasonable that the labor market for analysts will slot high-ability agents to the more prestigious investment banks.

Our reputation measure is a modified form of the Carter and Manaster (1990) and Carter, Dark, and Singh (1988) measure maintained by Jay Ritter on his website (http://bear.cba.ufl.edu/ritter/Rank.HTM). The measure assigns a rank from 1 through 9 to investment banks based principally on the amount of business generated by the bank in terms of handling equity offerings, as well as various other attributes; the website provides further details.

In Table 2 we provide the distribution of the variables that proxy for weight and strength.

Please insert Table 2 here.

Panel A of Table 2 reports the numbers of each type of recommendation changes. The bulk of the upgrades and downgrades are concentrated in the move from a relatively

---

"three-or-more" classes, but due to lower power caused by the smaller number of observations in the three-or-more class, we decided to pool it with the "two-level" class.

neutral category ("hold" and "buy") to the higher or lower categories. The number of downgrades is greater than the number of upgrades, which is interesting and counterintuitive, given the bull market over our sample period. Two-category recommendation changes are fairly common – there are more than 14,000 (22,000) such upgrades (downgrades). Three- and four-class upgrades and downgrades are far less common and range from 300 to 600.

Summary statistics on the weight proxies are presented in Panel B of Table 2. The mean analyst experience is over 1,000 trading days, and the median is very close to the mean, suggesting little skewness. The mean and median reputation of the analysts' host firms are also close to each other. The number of observations is quite healthy; more than 125,000 individual observations of each type of proxy are available.

Since our analysis considers the how strength and weight interact in the stock price reactions to analyst recommendations, we provide a two-way categorization of our observations in Panels C and D of Table 2.

Panel C contains the interaction between the two weight proxies. We find that for each value of the reputation measure, the mean level of analyst experience is fairly steady and ranges from 1,138 to 1,234. There is no evidence of a monotonic pattern of increasing experience as one moves from high-reputation to low-reputation investment banks. Thus, the two measures of weight appear to have substantial orthogonal variation. Indeed, the correlation between the two proxies is only about 0.05.

The interaction between the weight proxies and the strength proxies are reported in Panel D. A feature that stands out is that there are more downgrades by analysts working for high-reputation banks. This may be because downgrades may be less costly for high-reputation investment banks. In particular, with an established track record of the affiliated bank, and a steady brokerage clientele, an unpopular downgrade may be more credible and better-received by the market. There appear to be enough individual observations in each cell of the table to allow inferences about how investor reactions to upgrades and downgrades vary with the strength and weight proxies.

3 Results

We first describe the procedure for measuring abnormal returns. We divide the data into different portfolios based on the strength and weight proxies, as follows. Our first categorization is based on the strength proxies:

- Low-strength (a rating change of only one level)
- High-strength (a rating change of more than two levels)

The second classification is based on the level of experience of the analyst, measured in terms of his "experience" in the I/B/E/S database, defined as follows:

- Low-weight (less than five years experience)
- High-weight (more than five years experience)

Finally, the third category is based on the reputation of the investment bank to which the analyst belongs:

- Low-weight (investment bank ranking less than 8)
- High-weight (investment bank ranking equal to or greater than 8)

For each portfolio, we measure abnormal returns with daily data for the period $[t-2, t+19]$ days (t=day of recommendation change), and with monthly data for the 48-month period that follows day $t+19$. We also adjust for the "leftover" period that begins with t+20 and ends on the last trading day of that current month. This provides a complete picture of the price reaction to each type of event, including the initial event-window reaction and the subsequent over/under reaction.

Thus, during the time period $[t-2, t+19]$, measured in days (t =recommendation change day), we measure the cumulative abnormal returns using the market-adjusted model with daily data from CRSP (i.e., the method of Brown and Warner, 1985). Rather than adding up daily abnormal returns, we compound them beginning with day $t-2$. On any day ($t*$), the abnormal return is computed as:

$$[1+R_s(t-2)][1+R_s(t-1)]*[1+R_s(t)]\dots[1+R_s(t*)]$$

$$-[1+R_m(t-2)]*[1+R_m(t-1)]*[1+R_m(t)]\dots[1+R_m(t*)],$$

where R_s is the raw return on the stock, and R_m is the CRSP equally-weighted index return, including dividends.

For the time period $t + 20$ days until the last trading day of that calendar month, we compute a "leftover" abnormal return for each firm, which we later lump with the monthly abnormal return of the subsequent month. This allows us to "zoom in" inside the event window using CRSP daily data, while preserving the ability to measure longer-term returns with monthly data, as is customary. For example, for an event that occurs on Feb. 10 (Monday), $t = 0$ is Feb. 10, $t = 19$ is March 7 (Friday). Abnormal returns are computed daily for the period between Feb. 10 and March 7. We then compute a residual (or leftover) abnormal return using daily data for the period March 10 (Monday) to March 28 (Friday), and use this to adjust the intercepts of the Fama and French model for the longer-term horizons that are measured with monthly data and begin with the month of April, as described below.

For the time period $t + 2$ months to $t + 49$ months, we measure abnormal returns using the intercept of Fama and French (FF) portfolios.[5] For each horizon (1 month, 2 months,..., 48 months) we form a separate calendar time portfolio, and obtain the intercepts corresponding to the groups defined above. Suppose that for a given group the intercept of the 1-month FF calendar time portfolios is 0.5, the two-month intercept is 0.3, and the thirty-six month intercept is 0.10, then the evolution of the abnormal returns through time would be represented as follows:

---

[5] As in Boehme and Sorescu (2002), we estimate this regression using weighted least-squares, with the square root of the number of observations in each month being used as weights.

For the first 20 trading days, we would calculate the daily market-adjusted CARs, constructed as described above. (We assume 20 trading days represent one month.) For the second month, we calculate [1+CAR(20)]*[1+leftover CAR]*[1+0.5]−1, where

- CAR(20) is the value of the cumulative abnormal returns at the end of the 20-trading day window,
- the residual CAR is the additional abnormal return leading to the end of that calendar month (on average this occurs 10 trading days after the end of the 20-day window), and
- 0.5 is the intercept of the calendar time portfolio regression formed by choosing only firms that had a recommendation change event exactly two months ago.

Similarly, for the third month, we would calculate, $[1+\text{CAR}(20)]*[1+\text{leftover CAR}]*[1+0.3]^2 - 1$, and so on, and finally, for the last (forty-ninth) month, we would calculate $(1+\text{CAR}(20))*(1+\text{leftover CAR})*(1+0.10)^{48} - 1$.[6]

3.1 Short-Term Return Reactions

Table 3 presents the short-term abnormal (market-adjusted) returns corresponding to various types of recommendations.

Please insert Table 3 here.

---

[6] We also tried an alternative procedure which simply added the Fama-French intercepts, as opposed to compounding them, and found that the results were largely unchanged.

Generally, the reactions conform to intuition with large downgrades leading to greater abnormal returns. In addition, recommendations by low-weight analysts lead to smaller short-term reactions. Indeed, the return differential between analysts that rank high on both measures of weight and those that rank low is substantial. Specifically, the cumulative abnormal returns for the period [-2,20] vary from +1.80% for one-class upgrades issued by low-weight analysts to -10.66% for two-or-more class downgrades issued by high-weight analysts. Each of the abnormal return numbers are statistically significant at the 1% level, and are also statistically different from each other at that same significance level. Overall, these results support the notion that analysts' reputation as well as their experience are considered by agents in their stock investing decisions.

To construct a summary measure of return performance as a function of strength and weight, we construct a portfolio in each time-interval (day or month) which is long one share on upgrades and short one share on downgrades. We then obtain the cumulative abnormal returns for this "hedge portfolio" using the Brown and Warner (1985) method, and report these abnormal returns for the short-term horizons of [-2,2] and [-2,20] in Panel B of Table 3. Again, we find that the abnormal returns in every case are significant, and the high weight recommendations receive stronger reactions than low weight ones. For the low weight group, the difference between reactions for high strength and low strength recommendations is negligible. But, the corresponding differential reaction for the high-weight group is about 4%, suggesting that the market does indeed attach greater meaning to high strength recommendations from high weight analysts as opposed to those from low-weight ones.

3.2 Long-Term Return Reactions

We next turn to a longer-term assessment of market reactions to analyst upgrades and downgrades. Our first set of tests considers whether our stimuli for weight are justified. That is, across all types of recommendation changes, do high-weight analysts possess greater return forecasting ability than low-weight ones? To address this question, we consider the long-term returns following recommendation changes by high- and low-weight analysts. We construct an equally-weighted portfolio that takes a long position in upgrades and a short position on downgrades, and consider two types of windows: including and excluding the announcement period.[7] Our long-term performance measurement period ends on the last trading day of the month following the next recommendation revision by the same analyst, or at the end of the estimation horizon (four years), whichever comes first. This is to make sure that we mimic a strategy that could actually be implemented.[8] We then obtain the cumulative abnormal returns for this "hedge portfolio" using the Fama and French (1993) factors, applying the procedure described above and report these abnormal returns in Table 4.

Please insert Table 4 here.

---

[7] Our method is similar to that used by Mitchell and Stafford (2000), and Boehme and Sorescu (2002).

[8] We also tried simply keeping all post-event horizons at four years, and the results were unchanged.

In both cases, the point estimates indicate that the information content of experienced analysts' recommendations is much larger than of inexperienced ones. For example, the cumulative 48-month return for high-experience analysts is about 33% larger than that for low-experience ones. Including only the post-announcement period, the return differential is about 22%. The return on the hedge portfolio formed by the high-weight analyst recommendations is significant at the 10% level for the post-announcement period, and at the 5% level when one includes the ostensible leakage two days prior to the change in recommendations. Indeed, inexperienced analysts, appear to have limited forecasting power, since the total abnormal return (including the event period) for this group is statistically indistinguishable from zero. Thus, while Table 3 suggests that the market, in the short term, believes that there is valuable information in these analysts' forecasts, subsequent resolution of uncertainty reveals that this is not the case. This further strengthens the case for using experience as a proxy for GT weight.

Owing to the controversy about the interpretation of the Fama and French (1993) factors as capturing systematic risk (see, e.g., Daniel and Titman, 1997), we also report results which perform return adjustments using a one-factor market model. In this case, we find that the returns for the high-weight case are significant at the 5% level in all case. The return differential also increases in magnitude for the case where the two days prior to the announcement are included, and reaches as high as 57% for recommendations issued by high-experience analysts relative to those issued by low-experience ones.

Overall, the results in Table 4 present a compelling picture that experience does count in the analyst industry: the information content of upgrades and downgrades issued by experienced analysts is greater than those issued by the relatively inexperienced ones.[9] The evidence in favor of the reputation proxy for weight is less compelling. Indeed, when one uses both weight measures to form high-weight and low-weight categories, the return differential between these categories actually narrows relative to the case where one uses experience as the sole proxy for weight.

3.3 Return Reactions as Functions of Strength and Weight

We now turn to tests of the Griffin and Tversky (1992) hypothesis. We first plot the holding-period abnormal return on hedge portfolios that represent the high-weight, low-strength, and low-weight, high-strength cases (where weight is based on experience).

Please insert Figure 1 here.

The figure suggests significant drift for the high-weight, low-strength portfolio and an overreaction to low-weight, high-strength portfolio. The drift is substantial; the cumulative abnormal returns for the high-weight, low-strength portfolio amount to more 60% over the 48 months following the recommendation shift. The overreaction suggests

---

[9] The reader might wonder why our results are so different from Barber, Lehavy, McNichols, and Trueman (BLMT) (2001). BLMT build an aggregate recommendation measure across all analysts, and re-balance their portfolio daily, each time there is a change in recommendation of any of the analysts who follow the stock. Because of the implicit daily re-balancing, reasonable estimates of transaction costs drastically reduce potential abnormal profits. By contrast, we follow changes in recommendation by the same analyst, and assume monthly instead of daily re-balancing.

a cumulative negative return of about 14% (measured from the peak) following a recommendation for the high-strength, low-weight portfolio, and an almost complete price reversal.

Table 5 presents statistical tests for the qualitative features represented in Figure 1.

Please insert Table 5 here.

In particular, we document the long-run cumulative abnormal returns for various strength and weight combinations, using windows of different sizes for up to 48 months following the recommendation shift.

Panel A provides the evidence for high-weight, low-strength changes in recommendations. While there is not much evidence of nonzero abnormal returns in the shorter run, the results suggest underreaction in the longer term. In particular, the total returns from thirteen to 48 months following the event are as high as 29% and 53.5% for the analysts whose experience is high, though conditioning on reputation as an additional proxy for weight does not enhance the magnitudes appreciably. For the total period from two days prior to the recommendation change to 48 months after, the cumulative abnormal return is about 40% for the high-experience, high reputation analysts, and this number is significant at the 1% level.

In Panel B, we document the abnormal returns for the high-strength, low-weight case. There is evidence of overreaction in this case in the shorter term (up to 12 months). For the low experience analysts, the cumulative abnormal return from one to twelve months following the recommendation shift is about 9% for the Fama and French (1993) factors, and is statistically significant at the 1% level. The case for overreaction is weakened when one conditions on both experience and reputation. This possibly indicates that the market responds more quickly and rationally to high-strength recommendations by analysts who belong to reputable investment banks, and that reputation of the investment bank is not the ideal proxy for the "weight" of an analyst's recommendation.

Overall, the results suggest substantial cross-sectional variation in the response of stock prices to analyst recommendations. The results support the notion that more experienced analysts render more useful recommendations. In addition, the market appears to underreact to low-strength recommendations issued by experienced analysts and, to a lesser extent, overreact to high-strength recommendations issued by inexperienced ones. By and large, these results are consistent with the arguments and evidence presented by Griffin and Tversky (1992).

4 Conclusion

In this paper, we consider how the heterogeneity of analyst attributes in the cross-section influences market reaction to upgrades and downgrades issued by analysts. Our analysis demonstrates that there is considerable variation in the information content of analyst recommendations as well as in the market reactions to such recommendations.

We first explore if high-ability analysts (as proxied by years of experience or the reputation of the investment bank that houses the analyst) exhibit superior ability to forecast returns. Our results are largely consistent with the notion that experience counts in the analyst industry, in that more experienced analysts offer more informative recommendations, though the role of investment bank reputation is modest at best. We formulate additional hypotheses based on recent psychological research on how agents process information. In particular, Griffin and Tversky (1992) indicate that agents place too much emphasis on the strength of a signal (e.g., the warmth of a recommendation letter) and too little emphasis on its weight (the credibility of the letter writer). This implies that agents underreact to high-strength (vivid), low-weight signals, and overreact to high-weight, low-strength signals. In our study, the proxy for strength is the number of categories covered by an upgrade or downgrade with a two-or-more category change in recommendation being classified as a high-strength event. The number of years of experience (as measured by the amount of time the analyst is present in the I/B/E/S database) and the reputation of the investment bank with which the analyst is affiliated serve as alternative proxies for weight. Our hypotheses are formulated ex ante, and hence address Fama's (1998) concern that behavioral research is often conducted by obtaining ex post explanations for phenomena.

Our analysis indicates that market reactions to analyst recommendations do not fully account for the experience (weight) of the analyst making the recommendation. Specifically, the market appears to overreact to large recommendation changes made by relatively inexperienced analysts and underreact to small recommendations by experienced ones. The results are strongly significant and provide support to the prediction of Griffin and Tversky (1992).

The study not only furthers our understanding of whether analyst ability to forecast returns varies in the cross-section, but also demonstrates that behavioral biases do indeed manifest themselves in agents' reactions to clear information signals such as analyst recommendations. In addition, the study demonstrates that it is possible to adapt psychological literature to the analysis of financial market agents in a manner that yields clear, testable implications.

References

Amihud, Y., and H. Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223-249.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman, 2001, Can investors profit from the prophets? Security analyst recommendations and stock returns, *Journal of Finance* 56, 531-563.

Barber, B., and D. Loeffler, 1993, The DartBoard column - 2nd-hand information and price pressure, *Journal of Financial and Quantitative Analysis 28*, 273-284.

Boehme R., and S. Sorescu, 2002, The long-run performance following dividend initiations and resumptions: Underreaction or product of chance?, *Journal of Finance* 57, 871-900.

Brennan, M., and A. Subrahmanyam, 1995, Investment analysis and price formation in securities markets, *Journal of Financial Economics* 38, 361-381.

Brown, S., and J. Warner, 1985, Using daily stock returns: The case of event studies, *Journal of Financial Economics* 14, 3-31.

Carter R., F. Dark, and A. Singh, 1998, Underwriter reputation, initial returns, and the long-run performance of IPO stocks, *Journal of Finance* 53, 285-311.

Carter, R., and S. Manaster, 1990, Initial public offerings and underwriter reputation, *Journal of Finance* 45, 1045-1067.

Chevalier, J., and G. Ellison, 1999, Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance, *Journal of Finance* 54, 875-899.

Daniel K., and S. Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance* 52, 1-33.

Dimson, E., and P. Marsh, 1984, An analysis of brokers' and analysts' unpublished forecasts of U.K. stock returns, *Journal of Finance* 39, 1257-1292.

Edwards, W., 1968, Conservatism in human information processing, in *Formal representation of human judgment* (B. Klienmuntz, ed.), John Wiley and Sons, New York, NY, pp. 17-52.

Fama, E., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283-306.

Fama, E. and K. French, 1993, Common risk factors in returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Griffin, D., and A. Tversky, 1992, The weighing of evidence and the determinants of confidence, *Cognitive Psychology* 24, 411-435.

Hong, H., and J. Kubik, 2003, Analyzing the analysts: Career concerns and biased earnings forecasts, *Journal of Finance* 58, 313-351.

Irvine, P., 2003, The incremental impact of analyst initiation of coverage, *Journal of Corporate Finance* 9, 431-451.

Lichtenstein, S., B. Fischoff, and L. Phillips, 1982, Calibration of probabilities: The state of the art to 1980, in *Judgment under uncertainty: Heuristics and Biases* (D. Kahneman, P. Slovic, and A. Tversky, eds.), Cambridge University Press, Cambridge, MA, pp. 306-334.

Michaely, R., and K. Womack, 1999, Conflict of interest and the credibility of underwriter analyst recommendations, *Review of Financial Studies* 12, 653-686.

Mitchell M., and E. Stafford, 2000, Managerial decisions and long-term stock price performance, *Journal of Business* 73, 287-329.

Odean, T., 1998, Volume, volatility, price, and profit when all traders are above average, *Journal of Finance* 53, 1887-1934.

Palia, D., 2000, The impact of regulation on manager labor markets, *Rand Journal of Economics* 31, 165-179.

Stickel, S., 1995, The anatomy of the performance of buy and sell recommendations, *Financial Analysts Journal* 51, 25-39.

Womack, K., 1995, Do brokerage analysts recommendations have investment value?, *Journal of Finance* 50, 1005-1005.

Table 1

Simulations of Overreaction and Underreaction to Signals with Varying Strength and Weight (S, W denote strength and weight, and H, L denote high and low, respectively).

Case	Correlation between information signal and future price change	Expected future price change conditional on a positive signal	Expected future price change conditional on a negative signal
HSHW	-0.31	-0.19	0.19
LSHW	0.17	0.04	-0.04
HSLW	-0.55	-0.43	0.43
LSLW	0.05	-0.01	0.01

Table 2: Descriptive Statistics

The number of recommendation changes (upgrades and downgrades) is shown as a function of the strength of the change, the experience of the analyst (measured in trading days) and the reputation of the analyst's brokerage house.

Panel A: Distribution of *strength* variables

Upgrades			Downgrades		
Strength of the change	Description	Number of observations	Strength of the change	Description	Number of observations
One-class upgrades	Strong sell to Sell	319	One-class downgrades	Strong buy to Buy	24,819
	Sell to Hold	3,018		Buy to Hold	31,545
	Hold to Buy	22,215		Hold to Sell	4,538
	Buy to Strong buy	22,675		Sell to Strong sell	395
Two-class upgrades	Strong sell to Hold	2,074	Two-class downgrades	Strong buy to Hold	19,499
	Sell to Buy	774		Buy to Sell	1,377
	Hold to Strong Buy	12,972		Hold to Strong sell	2,424
Three-class upgrades	Strong sell to Buy	434	Three-class downgrades	Strong Buy to Sell	548
	Sell to Strong buy	303		Buy to Strong Sell	598
Four-class upgrades	Strong sell to Strong buy	401	Four-class downgrades	Strong buy to Strong sell	620
Total upgrades		65,185	Total downgrades		86,323

Panel B: Distribution of *weight* proxies

Weight Proxy	N. Obs.	Mean	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Analyst *experience* measured by number of days in the IBES file	151,508	1,164	0	492	991	1,701	3,315
Reputation of the brokerage house, measured as in Carter and Manaster (1990)	125,945	7.21	1	6	8	9	9

Table 2, continued

Panel C: Interaction among *weight* proxies

Value of the broker *reputation* measure	N. Obs.	Analyst *experience*		Value of the broker *reputation* measure	N. Obs.	Analyst *experience*	
		Mean	Median			Mean	Median
9	32,917	1,234	1,072	4	2,739	994	757
8	32,711	1,138	924	3	2,313	964	855
7	25,603	1,213	1,083	2	1,088	878	708
6	11,941	1,188	1,058	1	748	1,028	842
5	15,885	1,151	968				

Correlation between *experience* and *reputation* ➔ 0.05219 ($p<0.001$)

Panel D: Interaction between *weight* proxies and *strength* variables – Number of observations on the event date

	Strength Variables ➔	Upgrades		Downgrades	
		One-class (**small**)	Two or more classes (**large**)	One-class (**small**)	Two or more classes (**large**)
Weight Variables ↓					
Experience	Reputation				
High		19,605	5,387	26,530	8,889
Low		28,622	11,571	34,767	16,137
	High	29,632	8,983	38,722	13,894
	Low	10,355	4,778	12,740	6,841
High	High	12,118	2,918	17,331	5,282
High	Low	9,184	1,485	5,313	2,259
Low	High	17,514	6,065	2,139	8,612
Low	Low	6,171	3,293	7,427	4,582

Table 3: Short-term abnormal returns

Panel A shows the mean *market-adjusted* cumulative abnormal returns (CARs) surrounding each analyst recommendation change. The CARs are computed using the Brown and Warner (1985) methodology, for two different windows: [t-2, t+2] and [t-2, t+19], where t is the date of the recommendation change. CARs are classified according to weight and strength variables. Panel B shows CARs of a hedge portfolio strategy involving long positions on stocks being upgraded and short positions on stocks being downgraded. The event window in Panel B is [t-2, t+19] and the CARs are market-adjusted as in Brown and Warner (1985).

Panel A: Short-Term Cumulative Abnormal Returns as a function of Weight and Strength

		Strength Variables ➔	Upgrades		Downgrades	
			One-class (**small**)	Two or more classes (**large**)	One-class (**small**)	Two or more classes (**large**)
Weight Variables ↓		**Measured Dependent Variable**				
Experience	Reputation					
High	High	**Number of Observations**	11590	2806	16628	5041
High	Low		3976	1428	5040	2151
Low	High		16976	5815	20752	8262
Low	Low		5947	3141	7124	4420
High	High	**CAR [-2, +2]**	3.23% *[30.1]****	4.19% *[17.3]****	-5.46% *[-44.6]****	-7.92% *[-32.0]****
High	Low		2.98% *[15.5]****	2.72% *[9.67]****	-4.84% *[-23.3]****	-6.87% *[-17.28]****
Low	High		2.48% *[32.7]****	1.97% *[14.6]****	-4.67% *[-51.8]****	-5.52% *[-34.9]****
Low	Low		1.80% *[12.6]****	1.71% *[10.0]****	-4.40% *[-26.8]****	-4.58% *[-21.8]****
High	High	**CAR [-2, +19]**	3.65% *[19.8]****	4.43% *[11.7]****	-7.47% *[-45.4]****	-10.66% *[-32.1]****
High	Low		3.17% *[9.13]****	3.06% *[6.01]****	-7.13% *[-23.5]****	-8.95% *[-17.0]****
Low	High		2.44% *[18.7]****	1.86% *[8.64]****	-6.60% *[-48.8]****	-7.65% *[-33.9]****
Low	Low		1.78% *[7.26]****	1.50% *[5.07]****	-6.26% *[-26.0]****	-6.64% *[-23.0]****

***, **, * Statistically significant at the 1%, 5% and 10% levels, respectively.

Table 3, continued

Panel B: Zero-investment (hedge) returns during days [-2, +19] as a function of Weight and Strength

			Strength Dimension	
			High	Low
			A recommendation change by two or more levels	A recommendation change by one level
Weight Dimension	High	Analysts from firms with high reputation and with more than five year experience	15.1% *[t=43.1]******	11.1% *[t=64.2]******
	Low	Analysts from firms with low reputation and less than five year experience	8.13% *[t=27.9]******	8.05% *[t=33.1]******

*** Statistically significant at the 1% level.

Note: The abnormal returns in each cell are statistically significant from each other at the 1% level, except for the difference between low-weight-high-strength and low-weight-low-strength, which is not significant at the 10% level.

Table 4: Ability to forecast future stock values as a function of Weight proxies

Abnormal returns are estimated for two different horizons. In Panel A, the horizon includes the announcement of the recommendation change: [t-2 days, t+48 months]. Panel B examines only the post-announcement horizon [t+2 months, t+48 months], where t is the day of the change in analyst recommendation. Abnormal returns in Panel A are computed using the daily Brown and Warner market model for the period [t-2days, t+1month], and using Calendar Time Portfolio approach for the period [t+2 months, t+48 months]. Abnormal returns in Panel B are computed monthly, using exclusively the Calendar Time Portfolio approach. During the period [t+2months, t+48 months] equally weighted (monthly rebalanced) calendar time portfolio returns ($R_{u,t}$) are calculated each month from all firms experiencing a recommendation upgrade in the previous 48 months (excluding the most recent month). These firms are placed in a "long" portfolio. Likewise, a similar, but "short" portfolio is formed of all firms that have experienced a recommendation downgrade during the same period ($R_{d,t}$). The monthly differences in the raw returns on the long and short portfolio are regressed, alternatively, on the three Fama and French (1993) factors, and on the market factor (CAPM). The resulting intercepts from each of the following two models provide a measure of the abnormal performance of the hedge portfolio.

$$R_{u,t}-R_{d,t} = \alpha_p + \beta_p(R_{m,t}-R_{f,t}) + s_pSMB_t + h_pHML_t + e_{p,t}.$$

$$R_{u,t}-R_{d,t} = \alpha_p + \beta_p(R_{m,t}-R_{f,t}) + e_{p,t}.$$

Intercepts are reported on an annualized basis, and t-statistics are show in brackets next to each measure. In Panel A, these intercepts are adjusted with the hedge abnormal returns computed with daily data during [t-2days, t+1month]. $R_{f,t}$ is the return of the one-month T-Bills. ($R_{m,t}$-$R_{f,t}$) is the excess return of the equally weighted CRSP market portfolio. SMB_t is the difference in returns between portfolios of small and big stocks. HML_t is the difference in returns between portfolios of high and low book-to-market ratio stocks. The model estimated is weighted least squares, and monthly returns are weighted by the square root of the number of firms contained in each month.

Table 4, continued

Panel A: Long-term returns including the announcement period

Hedge portfolio abnormal returns during the window [*t*-2days, *t*+48 months] where *t* is the date of the recommendation change. ➔		Model estimated			
		Calendar-time Fama-French three-factor model (the number of months in the calendar equals 108)		Calendar-time CAPM (one-factor model) (the number of months in the calendar equals 108)	
The hedge portfolio takes long positions on upgrades and short positions on downgrades		Abnormal Returns (%)	T-statistic	Abnormal Returns (%)	T-statistic
Weight Variables ↓					
Experience	Reputation				
High		40.0%	[2.53]**	63.3%	[2.94]***
Low		7.31%	[0.87]	5.97%	[0.50]
High	High	41.4%	[2.49]**	64.4%	[2.92]***
Low	Low	11.7%	[1.24]	11.1%	[0.80]

Table 4, continued

Panel B: Long-term returns excluding the announcement period

Hedge portfolio abnormal returns during the window [*t*+2 months, *t*+48 months] where *t* is the date of the recommendation change. ➔		Model estimated			
		Calendar-time Fama-French three-factor model (the number of months in the calendar equals 108)		Calendar-time CAPM (one-factor model) (the number of months in the calendar equals 108)	
The hedge portfolio takes long positions on upgrades and short positions on downgrades		Abnormal Returns (%)	T-statistic	Abnormal Returns (%)	T-statistic
Weight Variables ↓					
Experience	Reputation				
High		24.7%	[1.70]*	45.4%	[2.32]**
Low		-1.77%	[-0.21]	-3.00%	[-0.24]
High	High	25.7%	[1.67]*	46.0%	[2.31]**
Low	Low	2.74%	[0.28]	2.21%	[0.15]

***, **, * Statistically significant at the 1, 5 and 10 percent levels, respectively.

Table 5: Tests of Under-Reaction and Over-Reaction

Abnormal returns are estimated for the long-term horizon [t-2 days, t+48 months], and for various subsets thereof, where t is the day of the change in analyst recommendation. Abnormal returns for the period [t-2days, t+1 month] (or any subset thereof) are computed using the daily Brown and Warner market model. Abnormal returns for the period [t+2 months, t+48 months] (or any subset thereof) are computed monthly, using exclusively Calendar Time Portfolios. For example, when the entire period [t+2months, t+48 months] is of interest, equally weighted (monthly rebalanced) calendar time portfolio returns ($R_{u,t}$) are computed each month from all firms experiencing a recommendation upgrade in the previous 48 months (excluding the most recent month). These firms are placed in a "long" portfolio. Likewise, a similar, but "short" portfolio is formed of all firms that have experienced a recommendation downgrade during the same period ($R_{d,t}$). The monthly differences in the raw returns on the long and short portfolio are regressed, alternatively, on the three Fama and French (1993) factors, and on the market factor (CAPM). The resulting intercepts from each of the following two models provide a measure of the abnormal performance of the hedge portfolio.

$$R_{u,t}-R_{d,t} = \alpha_p + \beta_p(R_{m,t}-R_{f,t}) + s_pSMB_t + h_pHML_t + e_{p,t}.$$

$$R_{u,t}-R_{d,t} = \alpha_p + \beta_p(R_{m,t}-R_{f,t}) + e_{p,t}.$$

Intercepts are reported on an annualized basis, and t-statistics are show in brackets next to each measure. For measurement horizons that begin prior to t=2months, these intercepts are adjusted with the hedge abnormal returns computed with daily data during the [t-2days, t+1month] period. $R_{f,t}$ is the return of the one-month T-Bills. ($R_{m,t}$-$R_{f,t}$) is the excess return of the equally weighted CRSP market portfolio. SMB_t is the difference in returns between portfolios of small and big stocks. HML_t is the difference in returns between portfolios of high and low book-to-market ratio stocks. The model estimated is weighted least squares, and monthly returns are weighted by the square root of the number of firms contained in each month.

Table 5, continued

Panel A: Underreaction for *High-Weight*, *Low-Strength* Rating changes

	Proxy for "*high weight*" ➔	Experience ≥ 5 years			Experience ≥ 5 years and Reputation ≥ 7		
Measurement Period ↓	Model ➔	Market-Model CARs	Calendar Time Portfolios (N. Months = 108)		Market-Model CARs	Calendar Time Portfolios (N. Months = 108)	
	Measurement Window ↓		Fama-French 3-factor	CAPM (1-factor)		Fama-French 3-factor	CAPM (1-factor)
Event window	T – 2 days, T + 2 days	8.44%			8.63%		
Post-event period	T + 1 month, T + 6 months		-2.08% *[-0.81]*	-0.19% *[-0.07]*		-2.17% *[-0.84]*	-0.24% *[-0.08]*
	T + 1 month, T + 12 months		-3.30% *[-062]*	1.40% *[0.23]*		-3.32% *[0.62]*	1.34% *[0.22]*
	T + 1 month, T + 36 months		12.2% *[1.02]*	29.8% *[1.98]***		13.5% *[1.11]*	30.9% *[2.05]***
	T + 1 month, T + 48 months		22.4% *[1.63]**	42.4% *[2.25]***		23.47% *[1.64]**	42.9% *[2.27]***
	T + 13 months, T + 48 months		29.1% *[2.37]****	53.5% *[3.52]****		28.8% *[2.24]***	52.4% *[3.40]****
Total returns: event window and post-event period[(a)]	T – 2 days, T + 12 months		8.52% *[1.77]***	13.4% *[2.32]***		8.63% *[1.78]***	13.4% *[2.32]***
	T – 2 days, T + 48 months		38.5% *[2.64]****	61.0% *[3.01]****		39.9% *[2.62]****	61.8% *[3.04]****

***, **, * Statistically significant at the 1, 5 and 10 percent levels, respectively (one-tail test).

Table 5, continued

Panel B: Overreaction for *Low-Weight*, *High-Strength* Rating changes

	Proxy for "*low weight*" ➔	Experience < 5 years			Experience < 5 years and Reputation < 7		
Measurement Period ↓	Model ➔	Market-Model CARs	Calendar Time Portfolios (N. Months = 108)		Market-Model CARs	Calendar Time Portfolios (N. Months = 108)	
	Measurement Window ↓		Fama-French 3-factor	CAPM (1-factor)		Fama-French 3-factor	CAPM (1-factor)
Event window	T – 2 days, T + 2 days	7.08%			6.29%		
Post-event period	T + 1 month, T + 6 months		-5.54% *[-2.65]******	-6.58% *[-2.81]******		-4.24% *[-1.63]**	-6.05% *[-2.10]****
	T + 1 month, T + 12 months		-9.06% *[-2.45]******	-10.0% *[-2.30]****		-5.36% *[-1.25]*	-8.32% *[-1.62]**
	T + 1 month, T + 36 months		-7.31% *[-1.06]*	-10.8% *[-1.05]*		-9.03% *[-1.22]*	-13.1% *[-1.24]*
	T + 1 month, T + 48 months		-6.77% *[-0.73]*	-9.46% *[-0.74]*		-10.1% *[-1.00]*	-7.59% *[-0.58]*
	T + 13 months, T + 48 months		1.47% *[0.19]*	0.37% *[0.03]*		-3.57% *[-0.41]*	-0.97% *[-0.09]*
Total returns: event window and post-event period[(a)]	T – 2 days, T + 12 months		0.71% *[0.43]*	0.11% *[0.23]*		3.65% *[1.01]*	0.83% *[0.30]*
	T – 2 days, T + 48 months		3.48% *[0.44]*	0.77% *[0.12]*		-1.12% *[-0.02]*	1.58% *[0.17]*

(a)

***, **, * Statistically significant at the 1, 5 and 10 percent levels, respectively (one-tail test).

Figure 1

The long-term stock price response to recommendation changes is plotted for the period [t-2 days, t+48 months], where t is the day of the change in analyst recommendation. The plot depicts the stock price response of a hedge portfolio containing long position in recommendation upgrades ($R_{u,t}$) and short positions in downgrades ($R_{d,t}$). Abnormal returns for the period [t-2days, t+1 month] (or any subset thereof) are computed using the daily Brown and Warner market model. Abnormal returns for the period [t+2 months, t+48 months] (or any subset thereof) are computed monthly, using Calendar Time Portfolios returns regressed on the market factor (CAPM):

$$R_{u,t}-R_{d,t} = \alpha_p + \beta_p(R_{m,t}-R_{f,t}) + e_{p,t}.$$

where $R_{f,t}$ is the return of the one-month T-Bills, and ($R_{m,t}-R_{f,t}$) is the excess return of the equally weighted CRSP market portfolio. The model estimated is weighted least squares, and monthly returns are weighted by the square root of the number of firms contained in each month.

Each point on the graph represents a holding-period abnormal return of the hedge portfolio, cumulated beginning with day t-2. For any day (t*) during the [t-2days, t=19days] window, the holding period abnormal return is computed as:

(1+Rs(t-2))*(1+Rs(t-1))*(1+Rs(t))*…*(1+Rs(t*)) –
(1+Rm(t-2))*(1+Rm(t-1))*(1+Rm(t))*…*(1+Rm(t*)),

where Rs is the raw return on the stock, and Rm is the Equally Weighted NYSE+AMEX+NASDAQ return from CRSP, including dividends.

For the time period [t+20 days, t+1month] (i.e. until the last trading day of that calendar month), we compute a "leftover" abnormal return for each firm, which we later lump with the monthly abnormal return of the subsequent month.

For the time period [t+2 months, t+48 months], we measure abnormal returns using the calendar time portfolio CAPM intercepts. For each horizon shown on the graph (1 month, 2 months … 48 months), we form a separate calendar time portfolio, and use the intercept of that portfolio, together with the daily abnormal returns measured during the [t-2days, t+1 month] window, to obtain an overall measure of holding period abnormal performance.

(Figure appears on next page.)

Figure 1: Undereaction for High-Weight-Low-Strength events and Overreaction for Low-Weight-High-Strength-Events