JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS VOL. 41, NO. 1, MARCH 2006
COPYRIGHT 2006, SCHOOL OF BUSINESS ADMINISTRATION, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

The Cross Section of Analyst Recommendations

Sorin Sorescu and Avanidhar Subrahmanyam*

Abstract

We analyze the price reaction to analysts' revisions by testing the Griffin and Tversky (1992) hypothesis that agents place emphasis on the strength of the signal (the dramatic nature of the event) and may de-emphasize the weight (the ability of the analyst making the recommendation). Two attributes, namely, years of experience and the reputation of the analysts' brokerage houses form proxies for analyst ability (or weight) that we validate by documenting that revisions by high ability analysts outperform those by low ability ones. We find evidence of return persistence following small (low strength) revisions by high ability analysts and the opposite return pattern following large (high strength) revisions of low ability analysts, consistent with the arguments of Griffin and Tversky (1992). Our study provides an empirical link between evidence on individual decision making and stock market returns, and also helps promote an understanding of the analyst industry as well as its interaction with the investing population.

I. 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. Irvine (2003) demonstrates that there are 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

*Sorescu, ssorescu@mays.tamu.edu, Mays Business School, Texas A&M University, College Station, TX 77843, and Subrahmanyam, subra@anderson.ucla.edu, The Anderson School, University of California at Los Angeles, 110 Westwood Plaza, Los Angeles, CA 90095. The authors acknowledge the use of IBES data, a service of Thomson Financial, in the preparation of this study. The data have been provided as part of a broad academic program to encourage research on analysts' expectations. We are grateful to Brad Barber (the referee) and Hendrik Bessembinder (the editor) for insightful and constructive suggestions. We also thank Shlomo Benartzi, Sushil Bikhchandani, Bill Bradford, Michael Brennan, Kathryn Dewenter, Laura Frieder, Avi Kamara, Jennifer Koski, Jun Liu, Kevin McCardle, Ed Rice, and seminar participants at UCLA and the University of Washington for helpful comments.

(1995) show that analysts influence 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.

Previous research has extensively 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, more recently, Womack (1996) and Michaely and Womack (1999) both find evidence of significant abnormal returns following shifts in analysts' opinions about stocks.[1] 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.[2] 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).[3]

Studies on the analyst industry generally confirm the notion that stock price movements, on average, are in the direction of analysts' pronouncements. While this emphasizes the importance of analyst recommendations as a principal avenue of information production, there is a question as to whether analysts' attributes influence investor reaction to their recommendations in the cross section. 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 or reputation) when reacting to the opinions rendered by these agents.

In this paper, we investigate hypotheses on investor reactions to analyst recommendations by appealing to recent research in the science of decision making. In particular, the analysis of Griffin and Tversky (1992) indicates that the degree of confidence varies across different types of information. In forming expectations, agents emphasize the strength of information (e.g., the warmth of a recommendation letter), and under-emphasize the 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 revision (e.g., changing of a rec-

[1] Boni and Womack (2006) in this issue document that consolidating analyst recommendations by industry and taking positions in the resulting portfolios leads to highly significant abnormal average profits.

[2] Jegadeesh, Kim, Krische, and Lee (2004) examine the predictive power of consensus recommendations relative to those of recommendation changes, and find that changes are more valuable than consensus recommendations. Jegadeesh and Kim (2003) consider the predictive power of analyst recommendations in an international context. Eames, Glover, and Kennedy (2002) examine the link between analyst recommendations and earnings forecasts.

[3] Hong, Kubik, and Solomon (2000) suggest that inexperienced analysts face a greater risk of job loss as a consequence of inaccurate forecasts than experienced ones, and Clement and Tse (2005) find that the tendency for analysts' forecasts to herd is decreasing in analyst experience. Malloy (2005) addresses the link between geographical proximity of an analyst to a company and the accuracy of earnings forecasts for that company, while Easterwood and Nutt (1999) and Lim (2001) consider rationales for biases in analysts' forecasts. Brav and Lehavy (2003) analyze the relation between analysts' target price revisions and future returns.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

ommendation 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.

We first 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. The results show a clear pattern that analysts with more experience or better reputation show a greater ability to forecast stock returns by way of their recommendations. Therefore, there is reliable evidence that both experience and reputation count in the analyst industry.

In our empirical analysis, we find evidence of drift following small revisions by experienced and reputable analysts and reversal following large revisions by analysts that are either inexperienced or employed in less reputable brokerage firms. These results support the Griffin and Tversky (1992) hypotheses. Reversals following revisions of low weight analysts are strongly evident in upgrades (as opposed to downgrades), suggesting that short-selling constraints mute the initial overreaction to such revisions. Our results not only provoke thought on how resources should be allocated to compensate analysts, but also have implications for designing investment strategies.

Our work is related to a series of papers by Mikhail, Walther, and Willis (MWW hereafter) (1997), (2003), and (2004), who also analyze the relation between analyst experience and recommendations as well as earnings forecasts. MWW (1997), (2003) document that more experienced analysts underreact less to earnings forecasts, and generate more accurate forecasts. Further, MWW (2004) report that analysts whose recommendations earn the most returns continue to outperform in the future. While we do succeed in replicating the MWW (2004) results (see footnote 23), we do not use performance persistence as a proxy for weight for two reasons. First, the Griffin and Tversky (1992) hypothesis does not attach any specific role for performance persistence. A second, more relevant issue is that a performance persistence-based measure of weight naturally is skewed toward analysts with greater experience because it omits analysts who have yet to build up a track record. Our result of greater drift for more experienced analysts, however, is consistent with the work of MWW, which suggests that analysts improve with experience. We go beyond MWW's approach by considering the reputation proxy for ability and by considering investor reactions to analysts' signals of varying strength.[4]

While our results, for the large part, are statistically significant, the magnitude of the abnormal returns we document is frequently below 1% per month. Interestingly, about half of the drift following high weight, low strength revisions can be explained by momentum, suggesting that experienced analysts incorporate the effects of momentum into their revisions. Overall, our results suggest that while institutional investors may well find trading strategies based on our evidence to be marginally profitable, transactions costs faced by retail investors would likely render these strategies unprofitable for individuals. Our results do

[4]In other related work, Barber, Lehavy, and Trueman (2005) and Cliff (2005) examine the differences between the predictive ability of recommendations issued by independent analysts and those affiliated with investment banks.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

have implications for the portfolio choices of retail investors. For example, our analysis suggests that investors should avoid being unduly attracted to stocks that have experienced major revisions by high weight analysts, while close attention should be paid to revisions by relatively low weight ones.

Previous research, which applies experimental and survey evidence on agents' behavior in the financial arena, principally consists of two strands. The first involves demonstrating empirically that portfolio choice and trading activity are consistent with such evidence (e.g., Benartzi and Thaler (2001), Odean (1998a), and Hong, Kubik, and Stein (2003)), and the second consists of theoretical models that explain price patterns based on this evidence (e.g., Daniel, Hirshleifer, and Subrahmanyam (1998), Barberis, Shleifer, and Vishny (1998)).[5] Thus, the question of whether results documented in experimental settings empirically affect long-run financial market returns is as yet under-explored.[6] By testing ex ante implications derived from experimental evidence, our study addresses this issue and attempts to mitigate a concern (Fama (1998)) that literature in this area provides post hoc theoretical explanations of market phenomena. In addition, we provide suggestive evidence that categorizing events into different types (relatively mundane, high weight events versus dramatic, low weight events) may lead to sharper predictions about return behavior following the events.

The remainder of this paper is organized as follows. Section II details our hypotheses while Section III describes the data. Section IV discusses the results. Section V conducts some robustness checks, Section VI discusses upgrades and downgrades separately, and Section VII concludes.

II. Hypotheses

To develop our hypotheses, we rely on recent research 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 (1998b), 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 focus on the strength of a signal and attach lesser 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

[5] Odean (1998b) presents a theoretical model of trading activity, but he does not consider the distinction between high and low weight signals. As in that analysis, we do not explicitly study impediments to arbitrage activity, but implicitly rely on recent studies that explore barriers to such activity (Shleifer and Vishny (1997), Abreu and Brunnermeier (2002), and Mitchell, Pulvino, and Stafford (2002)) as rationales for why our predictability is not eliminated by astute arbitrageurs.

[6] See, however, Coval and Shumway (2005) for a treatment of this issue in a short-run (intraday) context.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

overly focus on the enthusiasm in a recommendation letter and focus relatively less on the credibility of the recommendation writer.

In the Appendix, we develop a simple model of a financial market that incorporates information signals of differing strength and weight. We show that the arguments of Griffin and Tversky (1992) imply that return patterns following a high strength, low weight event in financial markets should involve reversals (due to overreaction), whereas there should be return drift following low strength, high weight events (caused by underreaction). We test these implications by using proxies for strength and weight.

We propose that the number of years of experience and the reputation of the investment bank to which the analyst belongs represent the weight of an analyst's recommendation. The number of categories an analyst's revision spans is used as a proxy for strength.[7] These proxies are discussed further in the next section. 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 (as simulations in the Appendix indicate).

All of the above arguments then lead to the following testable hypotheses.

H1. The predictive content of recommendation changes by high weight analysts will be greater than that of low weight analysts.

H2. There will be return drift following recommendation changes by high weight analysts who issue low strength signals.

H3. There will be return reversal 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.

III. Data

Our primary data come from the Institutional Brokers Estimate System (IBES) database, now part of Thomson Financial, which, in turn, is available on Wharton Research Data Services (WRDS) for the period December 1993 to December 2002. 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. IBES distinguishes five categories of recommendations, labeled 1 through 5, which we interpret as "strong buy," "buy," "hold," "sell," and "strong sell," respectively.[8] These opinions are often disseminated in the popular press

[7] 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 74% of CEOs in manufacturing possess an MBA degree from one of the top nine business schools, whereas only 33% of those in the utility sector possess such a degree. We do not have access to such proxies for ability for security analysts and rely instead on experience and reputation measures.

[8] Not all investment banks use a five-point scale. However, IBES requires all contributors to report their recommendation directly on a five-point scale, even if they do not normally use a five-point scale.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University 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 revisions (i.e., *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 emphasis on the strength of information (e.g., the warmth of a recommendation letter), and de-emphasize 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, two-category changes are classified as medium strength, and three- or four-category changes are denoted high strength. The notion is that a three- or four-category jump is likely to be vivid and dramatic, and thereby attract the attention of agents.[9]

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 IBES 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. Furthermore, it seems reasonable that the labor market for analysts will slot high ability agents to the more prestigious investment banks.[10] Admittedly, these are imperfect proxies for ability. For example, we only start counting the days in the IBES database from date of commencement of the detail file and not from when the analyst enters the profession. However, it is likely that our measures are significantly correlated with true ability, which is what we desire for our tests. In any case, we validate the efficacy of these measures later in the paper.[11]

Since the analyst firms themselves perform the conversion and have reputational incentives to use an accurate conversion method, there should be a high level of uniformity in interpreting the five-point scale recommendations of IBES across analysts.

[9]In our data, the bulk of the upgrades and downgrades are concentrated in the move from a relatively neutral category (hold and buy) to the higher or lower categories. Three- and four-class upgrades and downgrades are far less common and range from 300 to 600, suggesting that results for high strength revisions could potentially suffer from a lack of power. That we often find significance in this case is noteworthy. Later in the paper, we discuss results using alternative cutoffs for strength.

[10]See Fox (2004) for empirical evidence consistent with this basic notion in the managerial labor market.

[11]An alternative proxy for analyst ability is the track record of the analysts' earnings forecasts or their earlier recommendations. As we mentioned in the Introduction, such an exercise would skew the sample toward high experience analysts because a sufficiently long track record would be required to obtain enough observations for a reliable calculation of ability. In addition, earnings forecast accuracy may not be tied as strongly to analyst incentives as the direct measure of analyst value, namely, return forecastability. Another possibility is to use the All-Star analyst list as an indicator of high ability, but only 15% of analysts fall into this category (viz. Clarke et al. (2004)) suggesting that the usage of this group would exclude a number of analysts who potentially are of high caliber.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

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.[12] 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.

Summary statistics on the weight proxies are presented in Panel A of Table 1. 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 reputations 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.

TABLE 1

Descriptive Statistics

Table 1 presents statistics associated with the experience of the analyst (measured in trading days) and the reputation of the analyst's brokerage house.

Panel A. Distribution of Weight Proxies

Weight Proxy	No. of Obs.	Mean	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Analyst experience measured by no. of days in IBES file	151,508	1,164	1	492	991	1,701	3,315
Reputation of brokerage house, measured as in Carter and Manaster (1990)	125,945	7.21	1	6	8	9	9

Panel B. Interaction among Weight Proxies

Value of Broker Reputation Measure	No. of Obs.	Analyst Experience	
		Mean	Median
9	32,917	1,234	1,072
8	32,711	1,138	924
7	25,603	1,213	1,083
6	11,941	1,188	1,058
5	15,885	1,151	968
4	2,739	994	757
3	2,313	964	855
2	1,088	878	708
1	748	1,028	842
Correlation between experience and reputation		0.05219	($p < 0.001$)

Panel B 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 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.

[12] Ritter's Website is http://bear.cba.ufl.edu/ritter.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

IV. 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. We note that the cutoff for high versus low strength and weight is somewhat of a judgment call since the Griffin and Tversky (1992) study does not provide specific guidelines for such stratification. With this in mind, we adopt a set of subjective thresholds to test our hypotheses. Our first categorization is based on the strength proxies: i) low strength (a rating change of only one level); ii) medium strength (a rating change of two levels); and iii) high strength (a rating change of three or four levels).

The second classification is based on the level of experience of the analyst, measured in terms of the time for which the analyst is in the IBES database and the reputation of the brokerage house to which the analyst belongs. For the experience proxy, the following rule is adopted: i) low weight (less than one year of experience); ii) high weight (more than one year of experience), while the proxy based on the reputation of the investment bank to which the analyst belongs is classified as follows: i) low weight (investment bank ranking less than 6); ii) high weight (investment bank ranking equal to or greater than 6). In Section V, we report results using alternative cutoffs.

For each portfolio, we measure the *short-term* abnormal returns with daily data for the period $[t-2, t+2]$ days (t = day of recommendation change), and, alternatively, for the period $[t-2, t+19]$ days. Abnormal returns are estimated using the Brown and Warner (1985) market-adjusted methodology with standard errors computed in the cross section.[13]

We measure long-term abnormal returns using the calendar time portfolio approach (used, e.g., by Mitchell and Stafford (2000)), for different investment horizons ranging from six to 60 months. Because our goal is to measure the long-term abnormal returns beginning immediately after the end of the recommendation window, we construct our calendar time portfolios on a daily basis in contrast to the more commonly used monthly approach.

To illustrate the overall valuation effects of strength and weight, we report our main results using hedge portfolios that take long positions in upgraded stocks and short positions in downgrades, using the methodology proposed by Mitchell and Stafford (2000).[14] Note that due to short-selling restrictions (viz., Jones and Lamont (2002), Boehme, Danielsen, and Sorescu (2006), and Baker and Stein (2004)), we would expect our results to be stronger for upgrades than downgrades. Indeed, separate analyses for upgrades and downgrades in the penultimate section support this premise.

[13]We also compute abnormal returns using Brown and Warner's (1985) market model, and the results are almost identical.

[14]It may appear that the return of a hedge portfolio, which in theory requires zero net investment, is a strange construct. In reality, margin requirements do not permit the establishment of a hedge portfolio with zero investment. For example, margins of 50% each for going long and short lead us to the conclusion that the investment required up front for forming a hedge portfolio that is long one share and short one share of two stocks, each priced at $30, is in actuality precisely $30.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

A. Short-Term Return Reactions

We estimate the cumulative abnormal returns (CARs) for each firm for two different short-term windows: $[t-2, t+2]$, and $[t-2, t+19]$, measured in days ($t =$ recommendation change day). CARs are computed by adding up, beginning with day $t-2$, the firm's daily raw returns, and subtracting the sum of the corresponding daily market returns. On any day ($t*$), the cumulative abnormal return for each firm is computed as

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

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

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

We next provide a cross-sectional analysis of the short-term returns as a function of the analysts' weight proxies and the strength of the analyst recommendation change signal. The analysis is performed alternatively using a portfolio approach and a multivariate OLS model.

1. Portfolio Analysis

We first group stocks into portfolios according to their various weight and strength characteristics and take an equally-weighted average of firm-specific CAR measures within the same portfolio grouping. Hedge returns that take long positions in upgrades and short positions in downgrades are computed as the differences in the mean short-term CARs of each portfolio.[15]

Table 2 presents the short-term hedge abnormal returns corresponding to various types of recommendation changes. Results are reported for the short-term horizons of $[-2, 2]$ and $[-2, 19]$ days. Reactions for two-class changes are generally larger than for one-class changes, which conforms with intuition. More counter-intuitive is the fact that three- or four-class reactions are smaller in magnitude, especially in the case of low weight analysts. This could be an early indication that investors view such changes with skepticism, perhaps because they sense possible overreaction by analysts to an extreme information signal (Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998)).[16] We also observe that recommendations by low weight analysts lead to smaller short-term reactions, and the difference between the low and high weight hedge CARs is significant both in a statistical and economic sense. In addition, this difference is most pronounced in the case of three- or four-class changes, where it ranges from 4.4% to 6.9%. Despite the apparent skepticism with which investors view large changes made by low weight analysts, the hedge portfolio

[15] p-values for the significance of the hedge results are computed using a test of differences in means based on an equal variance assumption.

[16] Note that while the analysis of Griffin and Tversky (1992) provides sharp predictions for price reactions to high strength, low weight events relative to low strength, high weight ones, it does not appear to have unambiguous predictions for whether investor reactions should be stronger for high strength signals relative to low strength ones.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

abnormal returns are nevertheless significantly positive in the short term, indicating that prices initially move in the direction of the recommendation.[17] Later in this section, we document a price reversal for this low weight, high strength portfolio over the long term. In order to more completely understand the short-term return reactions, however, we next perform a regression analysis that considers weight and strength measures as determinants of short-term return reactions to recommendation revisions.

TABLE 2

Short-Term Abnormal Returns Portfolios

Zero-investment (hedge) cumulative abnormal returns (CARs) are shown for two different daily windows, $[t-2, t+2]$ and $[t-2, t+19]$, where t is the day of the recommendation change. CARs are classified according to weight and strength variables. The portfolios take long positions on stocks being upgraded and short positions on stocks being downgraded. *p*-values are shown in parentheses below each coefficient.

	Measurement Window					
	CAR [−2, +2]			CAR [−2, +19]		
	Strength Variables					
Weight Variables	One-Class (small)	Two-Class (medium)	Three- or Four-Class (large)	One-Class (small)	Two-Class (medium)	Three- or Four-Class (large)
Experience						
High (≥ 1yr)	7.67% (< 0.001)***	9.13% (< 0.001)***	7.45% (< 0.001)***	9.72% (< 0.001)***	11.53% (< 0.001)***	9.20% (< 0.001)***
Low (< 1yr)	6.14% (< 0.001)***	5.70% (< 0.001)***	2.84% (0.016)**	8.14% (< 0.001)***	6.44% (< 0.001)***	4.76% (0.007)***
Difference: High−Low	1.53% (< 0.001)***	3.43% (< 0.001)***	4.61% (0.014)**	1.57% (< 0.001)***	5.10% (< 0.001)***	4.44% (0.088)*
Reputation						
High (6–9)	7.82% (< 0.001)***	9.41% (< 0.001)***	9.56% (< 0.001)***	9.89% (< 0.001)***	11.70% (< 0.001)***	11.45% (< 0.001)***
Low (0–5)	5.87% (< 0.001)***	6.24% (< 0.001)***	2.70% (0.005)***	7.74% (< 0.001)***	8.14% (< 0.001)***	4.63% (0.001)***
Difference: High−Low	1.95% (< 0.001)***	3.17% (< 0.001)***	6.89% (< 0.001)***	2.15% (< 0.001)***	3.53% (< 0.001)***	6.82% (< 0.001)***

***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

2. Regression Analysis

We regress the short-term return reactions, in turn, on analyst experience, the reputation proxy, the strength of the revision, and firm size, as well as variables that capture the interaction of strength with analyst proxies for weight. These regression results appear in Table 3.[18] In considering the results for upgrades, while it is the case that large upgrades have more muted reactions than small ones,

[17]This result is consistent with that of Green (2006) in this issue who also finds that analysts' revisions have significant short-term profit potential.

[18]To minimize cross-sectional dependence, we use the return only over the shorter $[t-2, t+2]$ window as the dependent variable in the regressions. Even with this narrower window, the standard errors may be slightly overstated due to the possibility that revisions may coincide in time, but the large number of observations used in the regressions should allay this concern. For robustness, we repeat the analysis with the longer $[t+2, t-19]$ event window and find that our results are qualitatively unaltered.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

we also find that large revisions by either experienced analysts or those belonging to more reputed investment banks have greater return reactions. This suggests that the market attaches more meaning to upgrades by high weight analysts.

TABLE 3

Short-Term Abnormal Returns—Cross-Sectional Regressions

The dependent variable is the short-term CAR measured over the $[t-2, t+2]$ event window, where t is the day of the recommendation change. Independent variables are REPUTATION, a dummy variable equal to one if the analysts' brokerage house reputation index equals to 6 or higher, EXPERIENCE, a dummy variable equal to one if the analyst has been listed on the IBES database for one year or more, LARGE, a dummy variable equal to one if the strength of the change in the analyst recommendation is three levels or higher, and MEDIUM, a dummy variable equal to one if the strength of the change in the analyst recommendation is two levels. Interaction terms are also included. Separate OLS models are estimated for upgrades (columns (1) and (2)) and downgrades (columns (3) and (4)).

	Direction of Change			
	Upgrade		Downgrade	
	Model			
	(1)	(2)	(3)	(4)
Intercept	0.10227 (< 0.001)***	0.09848 (< 0.001)***	−0.22033 (< 0.001)***	−0.21870 (< 0.001)***
REPUTATION	0.01150 (< 0.001)***		−0.01651 (< 0.001)***	
EXPERIENCE		0.00915 (< 0.001)***		−0.00936 (< 0.001)***
LARGE	−0.01591 (0.008)***	−0.01916 (0.030)**	0.01814 (0.004)***	0.01867 (0.103)
MEDIUM	−0.00222 (0.354)	−0.00863 (0.004)***	−0.00718 (0.009)***	−0.00270 (0.477)
LARGE∗REPUTATION	0.02167 (0.005)***		−0.02373 (0.004)***	
MEDIUM∗REPUTATION	0.00222 (0.412)		−0.00693 (0.026)**	
LARGE∗EXPERIENCE		0.01723 (0.078)*		−0.01189 (0.332)
MEDIUM∗EXPERIENCE		0.00882 (0.006)***		−0.01011 (0.012)**
Log(Size)	−0.00405 (< 0.001)***	−0.00380 (< 0.001)***	0.00890 (< 0.001)***	0.00856 (< 0.001)***
No. of Obs.	52,310	52,310	70,522	70,522
Adj. R^2	0.59%	0.53%	1.59%	1.39%

***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

For downgrades, the picture is somewhat similar, but a bit more mixed. We find that return reactions are greater for large or medium revisions by analysts who have high values of the reputation proxy. The point estimates for the interaction of the strength of the revision with analyst experience implies a similar directional effect, but this effect is significant only for the medium revision case. These results, by and large, appear consistent with the notion that big revisions by high weight analysts have greater short-term return reactions.

Overall, the short-term results support the notion that analysts' reputation as well as their experience are considered by agents in their stock investing decisions. In the next subsection, we consider return behavior following analyst revisions in the longer term.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

B. Long-Term Return Reactions

To analyze long-term returns, we begin by grouping firms into portfolios on the basis of the various strength and weight proxies. Specifically, for each horizon we form a hedge portfolio that takes long positions in upgrades and short positions in downgrades made by analysts that belong to the same weight and strength categories.

For each portfolio, we examine the long-term abnormal returns for various investment horizons that begin on day $t + 2$ (t is the day of the recommendation change) and end on the last trading day, h calendar months after the month of the recommendation change, where h can range from six to 60 months. We form separate calendar time portfolios for each horizon (h). For example, to form a 24-month portfolio for high weight upgrades on January 10, 1997 (Friday), we first isolate all firms that experience an upgrade by a high weight analyst during the period from January 1, 1995 to January 8, 1997 (Wednesday). For a firm to be retained in the portfolio, we require that the firm did not experience a revision by the same analyst during the same period.

We assume a one dollar investment in each recommendation. To mitigate biases related to bid-ask bounce (Canina, Michaely, Thaler, and Womack (1998)), portfolio weights are adjusted daily to reflect past returns on stocks comprising the portfolio.[19] Specifically, if R_{it} is the raw rate of return on stock i on day t, the rate of return of the calendar time portfolio on day t is given by

$$\left(\sum_{i=1}^{N} w_{it} R_{it}\right) \Bigg/ \sum_{i=1}^{N} w_{it},$$

where N is the number of firms in the portfolio and w_{it} is the cumulative return on day t of a one dollar position invested long in stock i.

To compute abnormal returns, we regress the excess returns of the zero-investment (or hedge) portfolios on the three daily Fama-French (1993) factors using OLS: (R_m-R_f), HML, and SMB. The corresponding intercept provides a daily measure of abnormal returns for each horizon. Throughout the paper, we report the annualized intercept corresponding to the various portfolios.[20] Owing to the controversy about the interpretation of the Fama-French (1993) factors as capturing systematic risk (see, e.g., Daniel and Titman (1997)), we also report results that perform return adjustments using a one-factor CAPM-type model.[21]

Some of our tests require that we compute statistical differences between abnormal returns obtained with two different hedge portfolios. In such cases, we introduce a method of double hedge portfolios that consists of taking long positions on one hedge portfolio and short positions on another.

[19]We thank the referee for this suggestion. This approach allows us to capture the portion of the recommendation change price effect that occurs during the calendar month of the recommendation announcement. The more traditional calendar time portfolio approach generally ignores the month containing the event occurrence and measures returns using monthly data beginning with the first trading day of the following calendar month.

[20]To annualize returns, we multiply the daily intercepts by 252.

[21]In the robustness section, we also consider the effects of adding Carhart's (1997) momentum factor.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

1. Ability to Forecast

Our first set of tests considers whether our stimuli for weight are justified. That is, across all types of revisions, 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 hedge portfolio that takes a long position in upgrades and a short position on downgrades. We then obtain the cumulative hedge abnormal returns using both the Fama-French (1993) three-factor model and the one-factor CAPM, applying the procedure described above. We report the results in Table 4 for horizons ranging from six to 60 months.

Panel A of Table 4 examines forecast changes made by analysts with high experience or reputation. The opposite case is presented in Panel B, which contains hedge portfolio results for changes made by analysts with low experience or reputation. Since the main focus of our analysis is on differences between the forecast ability for groups of analysts based on experience or reputation, we also examine a double hedge portfolio that takes long positions in the hedge portfolio from Panel A and short positions in the hedge portfolio from Panel B. The results of the double hedge portfolio analysis are presented in Panel C.[22]

In all cases, the point estimates in Panel A are higher than those in Panel B, indicating that the information content of high weight analysts' recommendations exceeds that of low weight ones. More importantly, the difference between the two sets of results is always statistically significant (Panel C), except for one of the four six-month horizon differences. For example, the annualized 48-month return for high experience analysts is about 4% larger than that for low experience ones. The magnitude of this difference is smaller when weight is measured with reputation rather than experience, but the difference itself is always statistically significant, as indicated in Panel C.[23]

An interesting finding from Panel B is that inexperienced analysts appear to have almost no forecasting power beyond the 12-month horizon, since the abnormal returns of the hedge portfolios are statistically indistinguishable from zero at the 24-month horizon and higher. By contrast, analysts associated with low reputation brokerage houses appear to have some forecasting power since the hedge portfolio abnormal returns are larger and significant. However, the double hedge portfolio abnormal returns in Panel C are always significant, indicating that the high reputation analysts have even better forecasting ability.

[22]The calendar time portfolio used with the experience proxy does not contain any observations during the year 1994 since at least one year of history is required to categorize an analyst's experience level.

[23]In unreported analysis, we also consider the weight proxy suggested by the work of MWW (2004). Specifically, each year we rank analysts into quintiles based on their stock picking abilities. For each analyst, we measure the stock picking ability as the average returns to the one-month hedge portfolio that is long on all upgrades and short on all downgrades by that particular analyst that year. We find results that are broadly consistent with those obtained by MWW. For example, categorizing analysts whose past performance falls in the bottom three quintiles as low weight and in the other quintiles as high weight, the double hedge portfolio analog of Table 4, Panel C yields an annualized drift of 1.60% ($p = 0.085$) at the 24-month horizon. This magnitude, however, is slightly smaller than the results obtained using our weight proxies. For reasons described in footnote 11, we desist from using past performance as a proxy for weight in the main part of the analysis.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

TABLE 4

Ability to Forecast Future Stock Values as a Function of Weight Proxies

Hedge portfolio long-term abnormal returns are estimated for different investment horizons as a function of the analyst's "weight," measured, alternatively, by experience and reputation. The measurement period for the long-term returns begins with $t + 2$ days (where t is the day of the change in analyst recommendation) and ends on the last trading day of the hth month after the calendar month during which the recommendation change takes place. The variable h represents the long-term investment horizon and takes one of the following values: 6, 12, 24, 36, 48, and 60 months. Abnormal returns are computed with daily data using the calendar time portfolio approach. Firms experiencing an upgrade are placed in a "long" portfolio, while downgrades are placed in a "short" portfolio. We compute a hedge portfolio that takes long positions in upgrades and short positions in downgrades and regress its excess returns on the three daily Fama and French (1993) factors and, alternatively, on the daily market factor (CAPM). The intercepts of the hedge portfolio represent the abnormal returns and are reported on an annualized basis (assuming 252 trading days per year). p-values are shown in parentheses next to each measure. Panel A shows the hedge portfolio returns for changes made by analysts having one year or more experience, or belonging to brokerage houses with a reputation index equal to 6 or higher. Panel B shows the hedge portfolio returns for recommendation changes made by analysts that have less than one experience or who belong to brokerage houses with a reputation index lower than 6. Panel C tests for the difference between the results presented in Panels A and B. The analysis in Panel C uses a double hedge portfolio that takes a long position in the hedge portfolio in Panel A and a short position in the hedge portfolio in Panel B.

Hedge portfolio abnormal returns computed using daily data during the window [$t + 2$ days, $t + h$ months] where t is the date of the recommendation change and h is the long-term measurement horizon shown below (in months). The hedge portfolio takes long positions on upgrades and short positions on downgrades.		Model Estimated (The number of trading days in the calendar equals 2,266 for reputation-based proxies and 2,015 for experience-based proxies)			
		Calendar Time Fama-French Three-Factor Model		Calendar Time CAPM (one-factor model)	
Weight Proxy	Horizon	Annual Abnormal Returns (%)	p-Value	Annual Abnormal Returns (%)	p-Value
Panel A. Ability to Forecast for Analysts with High Experience/Reputation					
Experience	6 months	11.61%	(< 0.001)***	10.47%	(< 0.001)***
Reputation		11.41%	(< 0.001)***	10.47%	(< 0.001)***
Experience	12 months	9.18%	(< 0.001)***	7.69%	(< 0.001)***
Reputation		9.19%	(< 0.001)***	7.94%	(< 0.001)***
Experience	24 months	6.37%	(< 0.001)***	4.78%	(< 0.001)***
Reputation		6.39%	(< 0.001)***	5.03%	(< 0.001)***
Experience	36 months	5.05%	(< 0.001)***	3.71%	(< 0.001)***
Reputation		5.12%	(0.001)***	3.98%	(< 0.001)***
Experience	48 months	4.52%	(< 0.001)***	3.28%	(0.001)***
Reputation		4.71%	(< 0.001)***	3.65%	(< 0.001)***
Experience	60 months	4.32%	(< 0.001)***	3.17%	(0.001)***
Reputation		4.61%	(< 0.001)***	3.61%	(< 0.001)***
Panel B. Ability to Forecast for Analysts with Low Experience/Reputation					
Experience	6 months	9.89%	(0.002)***	5.93%	(0.045)**
Reputation		9.19%	(< 0.001)***	7.71%	(< 0.001)***
Experience	12 months	5.96%	(0.016)**	2.13%	(0.230)
Reputation		6.78%	(< 0.001)***	5.04%	(0.001)***
Experience	24 months	1.24%	(0.285)	−1.36%	(0.276)
Reputation		4.41%	(0.001)***	2.95%	(0.019)**
Experience	36 months	−0.08%	(0.482)	−2.04%	(0.147)
Reputation		3.54%	(0.002)***	2.33%	(0.034)**
Experience	48 months	0.35%	(0.416)	−1.54%	(0.193)
Reputation		2.70%	(0.001)***	1.58%	(0.095)*
Experience	60 months	1.23%	(0.215)	−0.75%	(0.328)
Reputation		2.56%	(0.011)**	1.59%	(0.084)*

(continued on next page)

TABLE 4 (continued)

Ability to Forecast Future Stock Values as a Function of Weight Proxies

Panel C. Double Hedge Portfolio: Differences in Abilities to Forecast According to the Experience Level and Reputation of the Analysts' Brokerage House

Hedge portfolio abnormal returns computed using daily data during the window [t + 2 days, t + h months] where t is the date of the recommendation change and h is the long-term measurement horizon shown below (in months).

The hedge portfolio takes long positions in the hedge portfolio from Panel A and short positions in the hedge portfolio from Panel B.

		Model Estimated (The number of trading days in the calendar equals 2,266 for reputation-based proxies and 2,015 for experience-based proxies)			
		Calendar Time Fama-French Three-Factor Model		Calendar Time CAPM (one-factor model)	
Weight Proxy	Horizon	Annual Abnormal Returns (%)	*p*-Value	Annual Abnormal Returns (%)	*p*-Value
Experience	6 months	1.72%	(0.273)	4.54%	(0.058)*
Reputation		2.12%	(0.084)*	2.76%	(0.043)**
Experience	12 months	3.22%	(0.079)*	5.56%	(0.008)***
Reputation		2.41%	(0.036)**	2.90%	(0.016)**
Experience	24 months	5.12%	(0.003)***	6.14%	(< 0.001)***
Reputation		1.98%	(0.044)**	2.08%	(0.038)**
Experience	36 months	5.13%	(< 0.001)***	5.75%	(< 0.001)***
Reputation		1.58%	(0.069)*	1.65%	(0.063)*
Experience	48 months	4.17%	(0.002)***	4.82%	(< 0.001)***
Reputation		2.01%	(0.023)**	2.07%	(0.022)**
Experience	60 months	3.09%	(0.009)***	3.92%	(0.001)***
Reputation		2.05%	(0.019)**	2.02%	(0.022)**

***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Overall, the results in Table 4 present a compelling picture that both experience and reputation count in the analyst industry: the information content of upgrades and downgrades issued by experienced and reputable analysts is significantly greater than those issued by analysts that are relatively inexperienced or who are employed by lower-ranked brokers.[24]

2. Return Reactions as Functions of Strength and Weight

We now turn to tests of the Griffin and Tversky (1992) hypothesis. Table 5 presents statistical tests for the holding period abnormal return on hedge portfolios that represent the high weight, low strength, and low weight, high strength cases.

[24]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 rebalance 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 rebalancing, reasonable estimates of transaction costs drastically reduce potential abnormal profits. By contrast, we follow changes in recommendation by the same analyst, resulting in less frequent rebalancing. Our results also contrast with Desai and Jain (1997) and Metrick (1999), who show that stocks recommended by Super Star money managers at the Barron's Annual Roundtable and in newsletters tracked by the Hulbert Financial Digest (HFD) do not exhibit abnormal performance. In these cases, the differences in results could be due to the different agents considered. For example, the Barron's roundtable consists of money managers, not analysts, and the criteria used for selecting Super Stars are not clear. Further, newsletters oftentimes do not provide specific investment advice (as Metrick (1999) suggests), and HFD has to translate this advice using proprietary but subjective procedures which involve, for example, considering only the most highly recommended category.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

In particular, we document the long-run cumulative abnormal returns for various strength and weight combinations, using windows of different sizes for up to 60 months following the recommendation shift.

Panel A provides the evidence for high weight, low strength changes in recommendations. Abnormal returns measured *after* the end of the short-term event window (i.e., beginning with $t+2$ days) are all positive and statistically significant at better than the 1% conventional level. These results hold for both proxies of weight as well as for the three-factor and CAPM methodologies. The magnitude of the annualized drift is approximately 6%–9% per year for the 12- to 24-month post-event period and about 2%–4% per year for the longer 48-month horizon.

TABLE 5

Strength, Weight, and Long-Term Returns

Hedge portfolio long-term abnormal returns are estimated for different investment horizons. Two separate cases are considered. Panel A includes stocks with minor changes in recommendation (one level) made by analysts whose experience level is one year or higher, or who belong to brokerage firms with a reputation index equal to 6 or higher. Panel B includes stocks with major changes in recommendation (three levels or higher) made by analysts that have less than one year of experience or belong to brokerage firms with a reputation index less than 6. Panel C tests for the difference between the results presented in Panels A and B. The analysis in Panel C uses a double hedge portfolio that takes a long position in the hedge portfolio in Panel A and a short position in the hedge portfolio in Panel B. The measurement period for the long-term returns begins with $t + 2$ days (where t is the day of the change in analyst recommendation) and ends on the last trading day of the hth month after the calendar month during which the recommendation change takes place. The variable h represents the long-term investment horizon and takes one of the following values: 6, 12, 24, 36, 48, and 60 months. Abnormal returns are computed with daily data using the calendar time portfolio approach. Firms experiencing an upgrade are placed in a "long" portfolio, while downgrades are placed in a "short" portfolio. We compute a hedge portfolio that takes long positions in upgrades and short positions in downgrades and regress its excess returns on the three daily Fama and French (1993) factors and, alternatively, on the daily market factor (CAPM). The intercepts of the hedge portfolio represent the abnormal returns and are reported on an annualized basis (assuming 252 trading days per year). *p*-values are shown in parentheses under each measure.

Panel A. High Weight, Low Strength Rating Changes Using Hedge Portfolios

	Proxy for Low Strength: Hedge portfolios with long positions in stocks with one-class upgrades and short positions in stocks with one-class downgrades			
	Proxy for High Weight			
	Experience ≥ 1 year		Reputation ≥ 6	
	Model			
	Fama-French 3-Factor	CAPM	Fama-French 3-Factor	CAPM
	No. of Trading Days			
Horizon	2,015	2,015	2,266	2,266
6 months	9.63% (< 0.001)***	8.60% (< 0.001)***	8.62% (< 0.001)***	8.45% (< 0.001)***
12 months	6.92% (< 0.001)***	6.30% (< 0.001)***	6.85% (< 0.001)***	6.45% (< 0.001)***
24 months	5.16% (< 0.001)***	4.05% (0.001)***	4.63% (< 0.001)***	4.15% (< 0.001)***
36 months	4.24% (< 0.001)***	3.13% (0.002)***	3.63% (< 0.001)***	3.25% (0.001)***
48 months	3.82% (< 0.001)***	2.74% (0.002)***	2.20% (0.005)***	1.79% (0.022)**
60 months	3.78% (< 0.001)***	2.87% (0.001)***	2.37% (0.001)***	2.04% (0.007)***

(continued on next page)

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

TABLE 5 (continued)

Strength, Weight, and Long-Term Returns

Panel B. Low Weight, High Strength Rating Changes Using Hedge Portfolios

	Proxy for High Strength: Hedge portfolios with long positions in stocks with three- or four-class upgrades and short positions in stocks with three- or four-class downgrades			
	Proxy for Low Weight			
	Experience < 1 year		Reputation < 6	
	Model			
	Fama-French 3-Factor	CAPM	Fama-French 3-Factor	CAPM
	No. of Trading Days			
Horizon	2,015	2,015	2,266	2,266
6 months	20.2% (0.253)	17.9% (0.277)	3.18% (0.347)	3.29% (0.343)
12 months	2.40% (0.461)	1.94% (0.468)	2.04% (0.380)	2.39% (0.361)
24 months	−21.0% (0.190)	−21.1% (0.189)	−0.92% (0.429)	−0.50% (0.461)
36 months	−27.8% (0.063)*	−25.5% (0.080)*	−3.83% (0.208)	−3.20% (0.253)
48 months	−19.2% (0.052)*	−15.8% (0.092)*	−6.53% (0.078)*	−5.67% (0.114)
60 months	−3.31% (0.357)	−2.27% (0.401)	−5.89% (0.113)	−4.93% (0.161)

Panel C. Double Hedge Portfolio: Differences in Long-Term Returns between the High Weight, Low Strength Case (Panel A) and the Low Weight, High Strength Case (Panel B)

	Hedge Portfolio Construction: Long positions in the hedge portfolio in Panel A and short positions in the hedge portfolio in Panel B			
	Proxy for Weight			
	Experience		Reputation	
	Model			
	Fama-French 3-Factor	CAPM	Fama-French 3-Factor	CAPM
	No. of Trading Days			
Horizon	2,015	2,015	2,266	2,266
6 months	−10.6% (0.364)	−9.33% (0.379)	5.43% (0.255)	5.17% (0.266)
12 months	4.53% (0.426)	4.36% (0.429)	4.80% (0.238)	4.06% (0.273)
24 months	26.2% (0.139)	25.2% (0.148)	5.56% (0.144)	4.65% (0.188)
36 months	32.1% (0.040)**	28.7% (0.059)*	7.45% (0.060)*	6.46% (0.091)*
48 months	23.0% (0.028)**	18.5% (0.063)*	8.72% (0.032)**	7.45% (0.060)*
60 months	7.10% (0.219)	5.14% (0.287)	8.25% (0.048)**	6.97% (0.083)*

***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

In Panel B, we document the abnormal returns for the high strength, low weight case. Every point estimate is negative in this case for the experience proxy for the 24-month and longer horizons, which is the exact opposite of Panel A. For the 36- and 48-month horizons, results using the experience proxy attain statistical significance at the 10% level. Turning now to the case of analysts belonging to brokerage houses with low reputation, we also observe an initial continuation of returns for up to 24 months followed by a reversal: abnormal returns for both models are negative at the 24-month horizon and longer, but only attain significance in the case of the Fama-French model at the 48-month horizon. The lower significance level in the high strength, low weight case is likely due to the low power of the calendar time portfolio, given the relatively low number of observations in that case (viz. footnote 9).

For a more complete test of overreaction to low weight, high strength events by financial market agents, it is of interest to consider results that include the announcement period return in the long-term return analysis. Thus, we repeat the test by including the announcement window ($t - 2$ days, $t + 2$ days) in the calendar time portfolio. Since high strength, low weight hedge returns during the announcement window are significantly positive (Table 2), post-event reversal would cause these returns to become insignificant and close to zero over the long term. The results (not reported in the tables) conform with this conjecture. Returns including the announcement period are small in magnitude and always statistically insignificant. For example, hedge portfolio abnormal returns corresponding to large recommendation changes made by inexperienced analysts amount to -22.5% per year (p-value $= 0.128$), while hedge abnormal returns for large changes made by low reputation analysts for the same horizon equal -1.54% per year (p-value $= 0.372$).

In Panel C, we make use of the double hedge portfolio methodology previously developed to examine differences in returns between the low strength, high weight case (Panel A) and the high strength, low weight case examined in Panel B. The double hedge results are positive at the 12-month horizon and longer, and generally attain statistical significance at the 36-month horizon and beyond. This further supports our previous conclusion of return continuation in Panel A and reversals in Panel B.

Our results suggest substantial cross-sectional variation in the response of stock prices to analyst recommendations. The results support the notion that more experienced or reputable analysts render more useful recommendations. In addition, stock prices appear to drift following low strength recommendations issued by experienced or reputable analysts and, to a lesser extent, experience reversals following high strength recommendations issued by analysts that are either inexperienced or employed by lower reputation brokerage firms.

At this point, a word about the economic magnitude of our effects is in order. The relevant magnitudes of the abnormal return are mostly less than 1% (0.05%) at the monthly (daily) level. While our portfolios do not require daily rebalancing (since revisions do not occur on a day to day basis), the magnitude of the abnormal return suggests that our results are more likely to be exploitable by institutions as opposed to retail investors. Our results do have normative implications for retail investors. Specifically, our analysis indicates that individuals should avoid stocks

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

that have experienced dramatic upward revisions by inexperienced analysts and that more attention should be paid to revisions by relatively experienced analysts. Overall, while our analysis supports the Griffin and Tversky (1992) theory of how agents react to information, the results are not inconsistent with the notion that the market is efficient within the transaction cost bounds of individual investors.

V. Robustness Checks

In this section, we report the results of various robustness checks we conduct and present the results in Table 6. We examine different subsamples to account for market capitalization and the "dot-com" sector, and also experiment with different cutoffs for the weight variable and with the inclusion of Carhart's (1997) momentum factor.[25]

One potential concern with our results is the extent to which our analysis is influenced by the Internet bubble. Specifically, a buoyant Internet sector and enthusiastic analyst recommendations could potentially combine to influence our results in spurious ways. For example, Internet stocks recommended for purchase by high weight analysts could have realized strong positive returns merely because the Internet sector experienced a gradual run-up in stock prices (viz. Ofek and Richardson (2003)). To rule out this possibility, we rerun the tests of Table 5 after excluding all "new economy" stocks (defined as in Anderson and Reeb (2003)) from the analysis.[26] The results from that subsample are presented in Panel A of Table 6, and continue to support our hypotheses. However, the experience proxy for weight is more strongly supported by the data in this case. Moreover, the double hedge portfolios returns are always positive and are often significant. Overall, the results suggest that our results are not an artifact of the dot-com phenomenon.

We next verify the extent to which our phenomena are influenced by market capitalization. More specifically, to ensure that our results are not driven by the relatively small and "neglected" stocks, we report results for the subsample that consists of stocks in CRSP deciles six through 10, and present the results in Panel B of Table 6. Again, the returns are all of a sign consistent with Griffin and Tversky's (1992) arguments. Further, the results using the experience proxy for weight are generally stronger than those that use the reputation proxy. This suggests that our phenomenon is not confined to the small-cap sector.

We now report results that use thresholds for the weight variable that are different from those used in earlier tables. In particular, we experiment with a cutoff of two years for experience and seven for reputation. The results are reported in Panel C of Table 6. We again find that the directions of all return effects are consistent with Griffin and Tversky (1992) at 36-month and higher horizons. However, the statistical significance of the results is diminished considerably, es-

[25] For conciseness, results are presented for the three-factor asset pricing model, except in the case of Panel D, which examines the effect of the momentum factor. Results obtained with the CAPM are not qualitatively different from those obtained with the three-factor model.

[26] These excluded stocks consist of two-digit SIC codes of 35 (industrial machinery and equipment), 36 (electronic and other electrical equipment), 38 (instruments and related products), and 73 (business services).

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

TABLE 6

Robustness Tests

Hedge portfolio long-term abnormal returns as a function of strength and weight are estimated for two selected subsamples (Panels A and B), and for the full sample using alternative definitions of weight (Panel C). Panels A, B, and C use the daily Fama and French (1993) three-factor asset pricing model. In Panel A, we examine the long-term abnormal returns for stocks that *do not* belong to the high-tech sector (as in Anderson and Reeb (2003), we use two-digit SIC codes of 35, 36, 38, and 73 to identify high-tech firms). In Panel B, we examine the long-term abnormal returns for stocks belonging to CRSP size deciles 6 through 10. Panel C examines the long-term abnormal returns for the full sample, using different definitions of weight: a cutoff of two years experience instead of one, and a reputation index cutoff of 7 instead of 6. Panel D examines the long-term abnormal returns for the full sample, using a daily four-factor asset pricing model that includes Carhart's (1997) momentum factor in addition to the three Fama-French factors. In all cases, firms experiencing an upgrade are placed in a "long" portfolio, while downgrades are placed in a "short" portfolio. We compute a hedge portfolio that takes long positions in upgrades and short positions in downgrades and regress its excess returns on the three daily Fama and French (1993) factors (Panels A, B, and C), and on the four daily Fama-French-Carhart factors (including momentum, Panel D). The left half of each panel presents results using analyst experience as a proxy for weight; the right half uses reputation as a proxy for weight. Within each panel, columns (a1) and (a2) examine stocks with minor changes in recommendation (one level) made by analysts whose experience level is one year or higher, or who belong to brokerage firms with a reputation index equal to 6 or higher. Columns (b1) and (b2) include stocks with major changes in recommendation (three levels or higher) made by analysts that have less than one year of experience or belong to brokerage firms with a reputation index less than 6. Columns labeled (a1)−(b1) and (a2)−(b2) test for the difference between the results presented in columns (a1) or (a2) and those presented in columns (b1) or (b2), using double hedge portfolios that take long positions in the hedge portfolios in columns (a1)/(a2) and short position in the hedge portfolios from columns (b1)/(b2). The measurement period for the long-term returns begins with $t + 2$ days (where t is the day of the change in analyst recommendation) and ends on the last trading day, h months after the end of the calendar month during which the recommendation change takes place. The variable h represents the long-term investment horizon and takes one of the following values: 6, 12, 24, 36, 48, and 60 months. Abnormal returns are computed with daily data using the calendar time portfolio approach. The intercepts from each regression represent the abnormal returns and are reported on an annualized basis, assuming 252 trading days per year. p-values are show in parentheses under each measure.

	Proxy for Strength					
	High Strength: hedge portfolios with long positions in stocks with three- or four-class upgrades and short positions in stocks with three- or four-class downgrades					
	Low Strength: Hedge portfolios with long positions in stocks with one-class upgrades and short positions in stocks with one-class downgrades					
	Proxies for Weight					
	High Weight: Experience $\geq$ 1 year Low Weight: Experience $<$ 1 year (no. trading days in the portfolio = 2,015)			High Weight: Reputation $\geq$ 6 Low Weight: Reputation $<$ 6 (no. trading days in the portfolio = 2,266)		
	Portfolio					
Horizon	(a1) Low Strength, High Weight	(b1) High Strength, Low Weight	(a1)−(b1) Double Hedge Portfolio	(a2) Low Strength, High Weight	(b2) High Strength, Low Weight	(a2)−(b2) Double Hedge Portfolio
Panel A. Old Economy Stocks (all stocks except two-digit SIC codes 35, 36, 38, and 73)						
6 months	8.22% (< 0.001)***	−16.1% (0.334)	23.3% (0.235)	8.82% (< 0.001)***	−3.67% (0.335)	12.5% (0.074)*
12 months	6.79% (< 0.001)***	−47.9% (0.116)	54.2% (0.086)*	7.49% (< 0.001)***	−2.34% (0.374)	9.82% (0.089)*
24 months	4.77% (< 0.001)***	−32.4% (0.129)	37.2% (0.098)*	5.18% (< 0.001)***	−0.73% (0.451)	5.90% (0.157)
36 months	3.93% (< 0.001)***	−40.2% (0.043)**	44.2% (0.030)**	4.24% (< 0.001)***	−2.48% (0.322)	6.72% (0.106)
48 months	3.62% (< 0.001)***	−28.5% (0.047)**	32.1% (0.030)**	4.05% (< 0.001)***	−2.66% (0.303)	6.70% (0.098)*
60 months	3.48% (< 0.001)***	−18.7% (0.113)	22.3% (0.076)*	4.02% (< 0.001)***	−4.98% (0.160)	9.00% (0.037)**
Panel B. Large Stocks (size deciles 6 to 10)						
6 months	9.34% (< 0.001)***	−11.1% (0.387)	20.6% (0.278)	9.33% (< 0.001)***	5.20% (0.261)	4.13% (0.309)
12 months	7.51% (< 0.001)***	−44.9% (0.902)*	52.4% (0.060)*	7.79% (< 0.001)***	1.44% (0.416)	6.36% (0.172)
24 months	5.52% (< 0.001)***	−35.1% (0.056)*	40.6% (0.034)**	5.76% (< 0.001)***	−1.75% (0.368)	7.50% (0.077)*
36 months	4.44% (< 0.001)***	−52.5% (0.002)***	56.9% (0.001)***	4.64% (< 0.001)***	−2.17% (0.321)	6.81% (0.077)*
48 months	4.02% (< 0.001)***	−30.1% (0.021)**	34.2% (0.012)**	4.26% (< 0.001)***	−2.58% (0.283)	6.84% (0.068)*
60 months	3.91% (< 0.001)***	−20.8% (0.067)*	24.7% (0.039)**	4.25% (< 0.001)***	−2.29% (0.301)	6.54% (0.072)*

(continued on next page)

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

TABLE 6 (continued)

Robustness Tests

Panel C. Alternative Weight Definitions

	Proxy for Strength: High Strength: Hedge portfolios with long positions in stocks with three- or four-class upgrades and short positions in stocks with three- or four-class downgrades; Low Strength: Hedge portfolios with long positions in stocks with one-class upgrades and short positions in stocks with one-class downgrades					
	Proxies for Weight					
	High Weight: Experience $\geq$ 2 years; Low Weight: Experience $<$ 2 years; (no. of trading days in the portfolio = 1,763)			High Weight: Reputation $\geq$ 7; Low Weight: Reputation $<$ 7; (no. of trading days in the portfolio = 2,266)		
	Portfolio					
Horizon	(a1) Low Strength, High Weight	(b1) High Strength, Low Weight	(a1)−(b1) Double Hedge Portfolio	(a2) Low Strength, High Weight	(b2) High Strength, Low Weight	(a2)−(b2) Double Hedge Portfolio
6 months	10.0% ($<$ 0.001)***	10.7% (0.344)	−0.65% (0.490)	9.66% ($<$ 0.001)***	7.18% (0.168)	2.48% (0.372)
12 months	7.91% ($<$ 0.001)***	3.41% (0.428)	4.50% (0.404)	7.81% ($<$ 0.001)***	5.20% (0.203)	2.61% (0.339)
24 months	5.35% ($<$ 0.001)***	−8.11% (0.281)	13.5% (0.168)	5.43% ($<$ 0.001)***	1.67% (0.368)	3.76% (0.226)
36 months	4.23% ($<$ 0.001)***	−15.1% (0.092)*	19.3% (0.045)**	4.41% ($<$ 0.001)***	−1.65% (0.356)	6.06% (0.090)*
48 months	3.83% ($<$ 0.001)***	−11.3% (0.116)	15.1% (0.055)*	4.04% ($<$ 0.001)***	−2.49% (0.279)	6.53% (0.065)*
60 months	3.72% ($<$ 0.001)***	−3.77% (0.327)	7.49% (0.187)	4.01% ($<$ 0.001)***	−1.33% (0.380)	5.34% (0.113)

Panel D. Effect of the Momentum Factor (abnormal returns computed using the four-factor model)

	Proxy for Strength: High Strength: Hedge portfolios with long positions in stocks with three- or four-class upgrades and short positions in stocks with three- or four-class downgrades; Low Strength: Hedge portfolios with long positions in stocks with one-class upgrades and short positions in stocks with one-class downgrades					
	Proxies for Weight					
	High Weight: Experience $\geq$ 1 year; Low Weight: Experience $<$ 1 year; (no. trading days in the portfolio = 2,015)			High Weight: Reputation $\geq$ 6; Low Weight: Reputation $<$ 6; (no. trading days in the portfolio = 2,266)		
	Portfolio					
Horizon	(a1) Low Strength, High Weight	(b1) High Strength, Low Weight	(a1)−(b1) Double Hedge Portfolio	(a2) Low Strength, High Weight	(b2) High Strength, Low Weight	(a2)−(b2) Double Hedge Portfolio
6 months	6.63% ($<$ 0.001)***	25.8% (0.203)	−19.1% (0.268)	4.59% ($<$ 0.001)***	3.49% (0.334)	1.10% (0.446)
12 months	4.42% ($<$ 0.001)***	0.44% (0.493)	3.98% (0.436)	3.70% (0.001)***	0.83% (0.451)	2.87% (0.335)
24 months	3.09% (0.001)***	−10.6% (0.328)	13.7% (0.283)	2.31% (0.006)***	−0.86% (0.434)	3.17% (0.271)
36 months	2.76% (0.001)***	−16.7% (0.176)	19.4% (0.140)	1.86% (0.015)**	−3.68% (0.218)	5.53% (0.123)
48 months	2.75% ($<$ 0.001)***	−11.1% (0.167)	13.85% (0.115)	2.10% (0.007)***	−6.41% (0.082)*	8.51% (0.036)**
60 months	2.64% ($<$ 0.001)***	1.59% (0.430)	1.04% (0.454)	2.37% (0.001)***	−6.20% (0.101)	8.57% (0.043)**

***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

pecially for the high strength, low weight cases.[27] While this analysis suggests the results are somewhat sensitive to the definitions used for the strength and weight proxies, we note that Griffin and Tversky's (1992) arguments do not offer precise guidelines for high weight events. Overall, given the necessarily subjective interpretation of the weight/strength stimuli, we view the results as broadly supportive of their arguments.

In our final robustness check, we consider how the analysis is affected by consideration of momentum, an exercise that allows us to consider the extent to which our results capture this well-known feature of the cross section of expected stock returns. Specifically, we repeat the analysis in Table 5 using the Fama-French-Carhart four-factor model instead of the three-factor model. We do not include the momentum factor at the outset in our analysis because the interpretation of momentum is controversial. For example, some authors, e.g., Chordia and Shivakumar (2002), suggest that it may represent compensation for risk, and others such as Grundy and McNichols (2001) and Fama and French (1996) argue that it is a genuine anomaly. Nonetheless, explicitly incorporating the momentum factor allows us to shed additional light on the sources of the abnormal returns documented earlier. Results incorporating this factor are presented in Panel D of Table 6.

The drift results for low strength, high weight revisions are still significant, but are lower in magnitude in Panel D of Table 6 than they are in Panel A of Table 5. This is consistent with the idea that these analysts upgrade (downgrade) stocks following good (poor) performance, so part of the post-recommendation change drift (about half) appears to be explained by momentum. The high strength, low weight returns remain negative for the longer horizons, but only attain significance in the case of the reputation proxy at the 48-month horizon. It appears that young analysts are more contrarian in that they dramatically upgrade (downgrade) stocks following poor (good) performance, and accounting for momentum attenuates the continuation of the performance post-recommendation. The results for the less reputed analysts also are consistent with contrarian behavior by such analysts. Overall, controlling for momentum allows for more insight, but is consistent with our basic conclusion that return reactions differ for high weight, low strength revisions relative to low weight, high strength ones in a manner consistent with the psychological literature. The analysis reiterates the normative implication that investors should pay more attention to revisions by high weight analysts than by low weight ones. Specifically, the point estimates of the hedge portfolio returns and the significance of the double hedge portfolio returns for the reputation proxy in Panel D of Table 6 indicate that investors may earn greater returns on average by investing on the basis of revisions by high weight analysts.

VI. Separate Price Effects for Upgrades and Downgrades

Thus far, we have analyzed abnormal returns surrounding analyst revisions by taking the difference between a portfolio with positive returns and one with

[27] In unreported results, we also experimented with a cutoff of two or more changes in categories for the strength variable. Similar to the results in Panel B, while the high weight, low strength results are preserved, the high strength, low weight results are no longer significant.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

negative returns. This helps keep our sample size large and partly allays concerns about the power of the Fama-French calendar time portfolio technique to detect abnormal returns (e.g., Loughran and Ritter (2000)). For completeness, however, we now consider how the results change when we consider upgrades and downgrades separately.

Panel A of Table 7 presents long-term abnormal returns for one-class upgrades made by experienced or reputable analysts. As expected, all but one of the long-term abnormal returns are positive and more than half are statistically significant, consistent with the conjecture that returns continue to drift upward following an upgrade made by experienced or reputable analysts. Panel B examines the long-term abnormal returns for one-class downgrades made by experienced or reputable analysts. The abnormal returns are negative and generally significant only for the six-, 12-, and 24-month horizons. Given that downgrades may require short positions, short-sales constraints could explain the result that drift is more significant on upgrades than downgrades.

TABLE 7

Separate Effects for Upgrades and Downgrades

Calendar time long-term abnormal returns are estimated for different investment horizons. Four separate cases are considered. Panels A and B include stocks with minor changes in recommendation (one level) made by analysts whose experience level is one year or higher, or who belong to brokerage firms with a reputation index equal to 6 or higher. Panel A examines upgrades and Panel B looks at downgrades. Panels C and D include stocks with major changes in recommendation (three levels or higher) made by analysts that have less than one year of experience or belong to brokerage firms with a reputation index less than 6. Panel C examines upgrades and Panel D looks at downgrades. The measurement period for the long-term returns begins with $t + 2$ days (where t is the day of the change in analyst recommendation) and ends on the last trading day h months after the end of the calendar month during which the recommendation change takes place. The variable h represents the long-term investment horizon and takes one of the following values: 6, 12, 24, 36, 48, and 60 months. Abnormal returns are computed with daily data using the calendar time portfolio approach. In each panel, we compute a calendar time portfolio that takes long positions in the selected securities. We then regress the portfolio's excess returns on the three daily Fama and French (1993) factors and, alternatively, on the daily market factor (CAPM). The intercepts of the regression represent the abnormal returns and are reported on an annualized basis (assuming 252 trading days in a year). *p*-values are show in parentheses under each measure.

Panel A. Upgrades for High Weight, Low Strength Rating Changes

	Proxy for Low Strength Long positions in stocks with one-class upgrades			
	Proxy for High Weight			
	Experience ≥ 1 year		Reputation ≥ 6	
	Model			
	Fama-French 3-Factor	CAPM	Fama-French 3-Factor	CAPM
	No. of Trading Days			
Horizon	2,015	2,015	2,266	2,266
6 months	1.34% (0.181)	−0.14% (0.477)	2.84% (0.026)**	4.06% (0.024)**
12 months	1.69% (0.101)	0.83% (0.358)	1.88% (0.103)	3.14% (0.059)*
24 months	3.39% (0.005)***	3.91% (0.041)**	1.40% (0.167)	2.84% (0.072)*
36 months	3.21% (0.006)***	3.45% (0.056)*	1.63% (0.127)	3.17% (0.048)**
48 months	3.01% (0.008)***	3.16% (0.063)*	1.79% (0.103)	3.39% (0.035)**
60 months	2.66% (0.013)**	2.90% (0.072)*	1.91% (0.086)*	3.59% (0.026)**

(continued on next page)

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

In Panel C of Table 7, we show the long-term abnormal returns for three- or four-class upgrades made by inexperienced analysts, or analysts associated with low reputation brokerage houses (the high strength, low weight case). These

TABLE 7 (continued)

Separate Effects for Upgrades and Downgrades

Panel B. Downgrades for High Weight, Low Strength Rating Changes

	Proxy for Low Strength Short positions in stocks with one-class downgrades			
	Proxy for High Weight			
	Experience ≥ 1 year		Reputation ≥ 6	
	Model			
	Fama-French 3-Factor	CAPM	Fama-French 3-Factor	CAPM
	No. of Trading Days			
Horizon	2,015	2,015	2,266	2,266
6 months	−8.29% (< 0.001)***	−8.74% (0.001)***	−5.79% (0.008)***	−4.42% (0.064)*
12 months	−5.24% (< 0.001)***	−5.47% (0.019)**	−5.01% (0.012)**	−3.35% (0.112)
24 months	−1.77% (0.079)*	−0.15% (0.475)	−3.29% (0.050)**	−1.37% (0.296)
36 months	−1.03% (0.181)	0.32% (0.443)	−2.20% (0.116)	−0.24% (0.461)
48 months	−0.81% (0.227)	0.42% (0.421)	−1.81% (0.152)	0.20% (0.467)
60 months	−1.12% (0.149)	0.03% (0.495)	−1.68% (0.166)	0.37% (0.436)

Panel C. Upgrades for Low Weight, High Strength Rating Changes

	Proxy for High Strength Long positions in stocks with three- or four-class upgrades			
	Proxy for Low Weight			
	Experience < 1 year		Reputation < 6	
	Model			
	Fama-French 3-Factor	CAPM	Fama-French 3-Factor	CAPM
	No. of Trading Days			
Horizon	2,015	2,015	2,266	2,266
6 months	−21.3% (0.125)	−26.3% (0.080)*	−0.94% (0.440)	0.49% (0.469)
12 months	−14.38% (0.177)	−16.0% (0.157)	0.54% (0.460)	2.47% (0.326)
24 months	−31.4% (0.037)**	−29.0% (0.049)**	−4.67% (0.134)	−2.19% (0.311)
36 months	−31.4% (0.012)**	−27.8% (0.024)**	−6.18% (0.058)*	−3.51% (0.200)
48 months	−18.1% (0.030)**	−12.4% (0.103)	−6.19% (0.054)*	−3.64% (0.185)
60 months	−6.64% (0.175)	−3.49% (0.315)	−4.88% (0.103)	−2.33% (0.283)

(continued on next page)

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

TABLE 7 (continued)
Separate Effects for Upgrades and Downgrades

Panel D. Downgrades for Low Weight, High Strength Rating Changes

	Proxy for High Strength: Short positions in stocks with three- or four-class downgrades			
	Proxy for Low Weight			
	Experience < 1 year		Reputation < 6	
	Model			
	Fama-French 3-Factor	CAPM	Fama-French 3-Factor	CAPM
	No. of Trading Days			
Horizon	2,015	2,015	2,266	2,266
6 months	−41.6% (0.049)**	−44.3% (0.039)**	−3.55% (0.275)	−2.23% (0.362)
12 months	−16.8% (0.189)	−17.9% (0.175)	−0.94% (0.424)	0.65% (0.450)
24 months	−10.3% (0.263)	−7.86% (0.316)	−3.27% (0.199)	−1.21% (0.392)
36 months	−3.54% (0.382)	−2.07% (0.431)	−1.86% (0.294)	0.20% (0.480)
48 months	1.14% (0.436)	3.33% (0.323)	−1.06% (0.368)	0.87% (0.408)
60 months	−3.33% (0.288)	−1.22% (0.421)	−0.67% (0.411)	1.26% (0.361)

***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

stocks experience a remarkable reversal in value after the upgrade, especially in the case of inexperienced analysts. For low reputation analysts, the results are weaker although statistical significance is attained for the 36- and 48-month horizons using the three-factor model. These results are very supportive of the reversal hypothesis. Panel D examines the case of large (three- or four-class) downgrades made by inexperienced analysts or those employed by low reputation brokers. The Griffin and Tversky hypothesis predicts a reversal, but we do not find any evidence of positive abnormal returns. A possible explanation for this finding is that the initial overreaction to the downgrade may be muted by short-selling constraints.

Overall, the results are generally consistent with the hypothesis of Griffin and Tversky (1992) within the constraints imposed by market frictions such as short-selling costs.

VII. Conclusion

In this paper, we consider how heterogeneity of analyst attributes in the cross section influences market reaction to upgrades and downgrades issued by analysts. We formulate our hypotheses based on recent research on how agents process information. In particular, Griffin and Tversky (1992) indicate that agents place emphasis on the strength of a signal (e.g., the warmth of a recommendation letter) and de-emphasize its weight (the credibility of the letter writer). In turn, this yields differential return implications for high strength (vivid), low weight signals and high weight, low strength signals. In our study, the proxy for strength is the

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

number of categories covered by an upgrade or downgrade. We classify three or more category changes in recommendation as high strength events. Low strength events are those where the recommendation changes by only one category. The number of years of experience (as measured by the amount of time the analyst is present in the IBES 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 financial research on investor behavior is often conducted by obtaining ex post explanations for return phenomena.

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. We also present evidence consistent with the idea that competition in the labor market attracts the best analysts to the most reputable investment banks.

Our analysis indicates that market reactions to analyst recommendations occur in a manner consistent with that suggested by Griffin and Tversky (1992). Specifically, the abnormal returns are negative after large recommendation changes made by analysts who are either relatively inexperienced or who are employed by less reputable investment banks. By contrast, there is return drift after small recommendations by experienced or reputable analysts. The reversals for inexperienced analysts are very strongly evident for upgrades (in contrast to downgrades), suggesting that short-selling constraints possibly mute the initial reaction to the analysts' revisions.

While the phenomena we document are by and large statistically significant, the magnitude of the abnormal return is usually below 100 basis points on a monthly basis. Thus, the results accord with the notion that the phenomena we document are more likely to be exploitable by institutions than individuals. From a normative standpoint, our results suggest that investors should avoid paying undue attention to upward revisions by high strength, low weight revisions and should focus attention on those stocks that experience revisions by high weight analysts.

The study demonstrates that it is possible to adapt the literature on human decision making to the analysis of financial market agents in a manner that yields directly testable implications. One potential limitation of our study is that we have only considered return reactions, and not trading activity. Partitioning trades by individuals and institutions would shed light on who causes the return patterns that we do find. Our study also suggests extensions to other events that may be characterized by their weight and strength. For example, pronouncements by the chairman of the Federal Reserve Board are likely high weight events but may often be of low strength. In contrast, opinions expressed by young correspondents of financial networks such as CNBC may have the opposite characteristics. Exploring return patterns and trading activity following such events would be an interesting area for future research.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

Appendix

This appendix formally analyzes the implications of the work of Griffin and Tversky (1992) for security prices. 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 take the form $P = (1-k)\bar{\theta} + k(\theta + \epsilon)$, where k is the coefficient of the private signal in the formation of the posterior. The rational Bayesian value of k, denoted by k_b, is given by

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

We propose that k differs from k_b in a manner consistent with research on individual decision making. As in Odean (1998b), we assume that investors' calculated noise variance, v_r, differs from its true counterpart, v_ϵ. Further, we assume that investors use $k = k'$ in the formation of their posterior, where k' is given by

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

For the moment, we take the variable s to be non-stochastic. Later in this section, we will model s as a function of the information signal and precisely link k' with available experimental evidence.

The price P is given by $P = (1 - k')\bar{\theta} + k'(\theta + \epsilon)$. Substituting for k', straightforward calculations indicate that the sign of the covariance of price changes with the signal, $\text{cov}(\theta - P, \theta + \epsilon)$, is given by

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

It can be seen that if $v_r = v_\epsilon$ and $s = 0$, then prices are set by perfect Bayesian agents so that the covariance is zero. If the noise variance is assessed to be higher than v_ϵ, prices tend to drift. However, a positive s contributes to a negative covariance. The net effect depends on the magnitude of each type of effect.

To more explicitly connect strength and weight to investor reaction and the equilibrium price, and to model the phenomena documented 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 depend on the strength (extremeness) of the signal. In this case, the price is nonlinear 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 high strength, low weight, and low strength, high weight. 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. The high (low) strength case is represented by $|\theta + \epsilon| > 1$ (< 1). For the high weight (low weight) case, v_ϵ is 0.4 (1.0). Thus, the notion that investors do not respond to signal weight in a manner appropriate to a perfect Bayesian is captured by noting that their calculated noise variance v_r is invariant to whether v_ϵ (which is inversely related to signal accuracy) is high or low.

The correlations and conditional price changes yielded by the simulation for the low strength, high weight and high strength, low weight cases described above, are given in Panel A of Table 8.[28] Conforming to intuition, conditional price change patterns match the

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

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. Since Table 3 shows that the investor reaction is stronger for more extreme (high strength) signals, the market overly reacts to such signals. At the same time, the market reacts less than sufficiently 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.[29]

TABLE 8

Simulations of Price Reactions to Information Signals

Table 8 shows the result of a simulation of stock price responses 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 Positive Signal	Expected Future Price Change Conditional on Negative Signal
Panel A. Base Case			
LSHW	0.183	0.086	−0.086
HSLW	−0.543	−0.429	0.429
Panel B. Smaller Investor Reaction to a High Strength Signal			
LSHW	0.026	0.013	−0.013
HSLW	−0.346	−0.243	0.243

To briefly illustrate the dependence of the preceding effects on parameter values, Panel B of Table 8 describes the correlations and return reactions for the two cases of Panel A, with the difference that s_1 is set to be 0.3 rather than 0.4. In this case, the reaction to high strength signals is more muted than in Panel A and, consequently, the magnitudes of continuations and reversals are smaller.[30] This indicates that from an empirical standpoint the return reactions to signals of different stimuli will depend on the (unobserved) extent to which the decision making documented by Griffin and Tversky (1992) is present in financial market agents. The design and discussion of our tests take into account the subjective issues involved in the translation of experimental evidence to the design of proxies and hypotheses using non-experimental data.

In simulation results not reported for brevity, the correlation between price changes and the signal increases in v_ϵ^{-1}, the inverse of the variance of signal noise (and our measure of the signal's weight), which is a very intuitive result. This indicates that high weight signals will exhibit stronger price reactions in the direction of the signal, irrespective of signal strength, which justifies the hypothesis that there will be more drift following revisions by analysts with high values of the weight proxy.

References

Abreu, D., and M. Brunnermeier. "Synchronization Risk and Delayed Arbitrage." *Journal of Financial Economics*, 66 (2002), 341–360.

Amihud, Y., and H. Mendelson. "Asset Pricing and the Bid-Ask Spread." *Journal of Financial Economics*, 17 (1986), 223–249.

[29]Our analysis could also be further extended to consider implications for the high strength, high weight and low strength, low weight cases. However, we do not incorporate these extensions since the experimental work of Griffin and Tversky (1992), on which we rely to formulate our hypotheses, does not shed direct light on these cases.

[30]Other cases with varying values of C, s_1, and s_2 yield varying values of the numbers in Table 8, but do not add additional insight while preserving the signs of the various quantities. Hence, we omit these other cases from the discussion.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

Anderson, R., and R. Reeb. "Founding Family Ownership and Firm Performance: Evidence from the S&P 500." *Journal of Finance*, 58 (2003), 1301–1328.
Baker, M., and J. Stein. "Market Liquidity as a Sentiment Indicator." *Journal of Financial Markets*, 7 (2004), 271–299.
Barber, B.; R. Lehavy; M. McNichols; and B. Trueman. "Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns." *Journal of Finance*, 56 (2001), 531–563.
Barber, B.; R. Lehavy; and B. Trueman. "Comparing the Stock Recommendations of Investment Banks and Independent Research Firms." Working Paper, Univ. of California at Davis (2005).
Barber, B., and D. Loeffler. "The 'DartBoard' Column: Second-Hand Information and Price Pressure." *Journal of Financial and Quantitative Analysis*, 28 (1993), 273–284.
Barberis, N.; A. Shleifer; and R. Vishny. "A Model of Investor Sentiment." *Journal of Financial Economics*, 49 (1998), 307–343.
Benartzi, S., and R. Thaler. "Naïve Diversification Strategies in Defined Contribution Saving Plans." *American Economic Review*, 91 (2001), 79–98.
Boehme, R.; B. Danielsen; and S. Sorescu. "Short Sale Constraints, Differences of Opinion, and Overvaluation." *Journal of Financial and Quantitative Analysis*, forthcoming (2006).
Boni, L., and K. Womack. "Analysts, Industries, and Price Momentum." *Journal of Financial and Quantitative Analysis*, 41 (2006), 85–109.
Brav, A., and R. Lehavy. "An Empirical Analysis of Analysts' Target Prices: Short-Term Informativeness and Long-Term Dynamics." *Journal of Finance*, 58 (2003), 1933–1967.
Brennan, M., and A. Subrahmanyam. "Investment Analysis and Price Formation in Securities Markets." *Journal of Financial Economics*, 38 (1995), 361–381.
Brown, S., and J. Warner. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics*, 14 (1985), 3–31.
Canina, L.; R. Michaely; R. Thaler; and K. Womack. "Caveat Compounder: A Warning About the Daily CRSP Equal-Weighted Index to Compute Long-Run Excess Returns." *Journal of Finance*, 53 (1998), 403–416.
Carhart, M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.
Carter R.; F. Dark; and A. Singh. "Underwriter Reputation, Initial Returns, and the Long-Run Performance of IPO Stocks." *Journal of Finance*, 53 (1998), 285–311.
Carter, R., and S. Manaster. "Initial Public Offerings and Underwriter Reputation." *Journal of Finance*, 45 (1990), 1045–1067.
Chevalier, J., and G. Ellison. "Are Some Mutual Fund Managers Better than Others? Cross-Sectional Patterns in Behavior and Performance." *Journal of Finance*, 54 (1999), 875–899.
Chordia, T., and L. Shivakumar. "Momentum, Business Cycle, and Time-Varying Expected Returns." *Journal of Finance*, 57 (2002), 985–1019.
Clarke, J.; A. Khorana; A. Patel; and R. Rau. "Analyst Turnover, Stock Coverage, and Investment Banking Deal Flow." Working Paper, Purdue Univ. (2004).
Clement, M., and S. Tse. "Financial Analyst Characteristics and Herding Behavior in Forecasting." *Journal of Finance*, 60 (2005), 307–341.
Cliff, M. "Do Independent Analysts Provide Superior Stock Recommendations?" Working Paper, Virginia Polytechnic Institute and State Univ. (2005).
Coval, J., and T. Shumway. "Do Behavioral Biases Affect Prices?" *Journal of Finance*, 60 (2005), 1–34.
Daniel, K.; D. Hirshleifer; and A. Subrahmanyam. "Investor Psychology and Security Market Under- and Overreactions." *Journal of Finance*, 53 (1998), 1839–1885.
Daniel K., and S. Titman. "Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns." *Journal of Finance*, 52 (1997), 1–33.
Desai, H., and P. Jain. "An Analysis of the Recommendations of the 'Superstar' Money Managers at Barron's Annual Roundtable." *Journal of Finance*, 50 (1997), 1257–1273.
Dimson, E., and P. Marsh. "An Analysis of Brokers' and Analysts' Unpublished Forecasts of U.K. Stock Returns." *Journal of Finance*, 39 (1984), 1257–1292.
Eames, M.; S. Glover; and J. Kennedy. "The Association between Trading Recommendations and Broker-Analysts' Earnings Forecasts." *Journal of Accounting Research*, 40 (2002), 85–104.
Easterwood, J., and S. Nutt. "Inefficiency in Analysts' Earnings Forecasts: Systematic Misreaction or Systematic Optimism?" *Journal of Finance*, 54 (1999), 1777–1797.
Edwards, W. "Conservatism in Human Information Processing." In *Formal Representation of Human Judgment*, B. Klienmuntz, ed. New York, NY: John Wiley and Sons (1968).
Fama, E. "Market Efficiency, Long-Term Returns, and Behavioral Finance." *Journal of Financial Economics*, 49 (1998), 283–306.
Fama, E., and K. French. "Common Risk Factors in Returns on Stocks and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press

Fama E., and K. French. "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance*, 51 (1996), 55–84.
Fox, J. "Employer-Size Wage Gaps: Testing Worker Preference and Span of Control Explanations." Working Paper, Univ. of Chicago (2004).
Green, C. "The Value of Client Access to Analyst Recommendations." *Journal of Financial and Quantitative Analysis*, 41 (2006), 1–24.
Griffin, D., and A. Tversky. "The Weighing of Evidence and the Determinants of Confidence." *Cognitive Psychology*, 24 (1992), 411–435.
Grundy, B., and S. Martin. "Understanding the Nature of the Risks and the Source of the Rewards to Momentum Investing." *Review of Financial Studies*, 14 (2001), 29–78.
Hong, H., and J. Kubik. "Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts." *Journal of Finance*, 58 (2003), 313–351.
Hong, H.; J. Kubik; and A. Solomon. "Security Analysts' Career Concerns and Herding of Earnings Forecasts." *Rand Journal of Economics*, 321 (2000), 121–144.
Hong, H.; J. Kubik; and J. Stein. "Social Interaction and Stock-Market Participation." *Journal of Finance*, 59 (2004), 137–163.
Irvine, P. "The Incremental Impact of Analyst Initiation of Coverage." *Journal of Corporate Finance*, 9 (2003), 431–451.
Jegadeesh, N.; J. Kim; S. Krische; and C. Lee. "Analyzing the Analysts: When Do Recommendations Add Value?" *Journal of Finance*, 59 (2004), 1083–1124.
Jegadeesh, N., and W. Kim. "Value of Analyst Recommendations: International Evidence." Working Paper, Emory Univ. (2003).
Jones, C., and O. Lamont. "Short-Sale Constraints and Stock Returns." *Journal of Financial Economics*, 66 (2002), 207–239.
Lichtenstein, S.; B. Fischoff; and L. Phillips. "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, MA: Cambridge Univ. Press (1982).
Lim, T. "Rationality and Analysts' Forecast Bias." *Journal of Finance*, 56 (2001), 369–385.
Loughran, T., and J. Ritter. "Uniformly Least Powerful Tests of Market Efficiency." *Journal of Financial Economics*, 55 (2000), 361–389.
Malloy, K. "The Geography of Equity Analysis." *Journal of Finance*, 60 (2005), 719–755.
Metrick, A. "Performance Evaluation with Transactions Data: The Stock Selection of Investment Newsletters." *Journal of Finance*, 54 (1999), 1743–1775.
Michaely, R., and K. Womack. "Conflict of Interest and the Credibility of Underwriter Analyst Recommendations." *Review of Financial Studies*, 12 (1999), 653–686.
Mikhail, M.; B. Walther; and R. Willis. "Do Security Analysts Improve Their Performance with Experience?" *Journal of Accounting Research*, 35 (1997), 131–157.
Mikhail, M.; B. Walther; and R. Willis. "The Effect of Experience on Security Analyst Underreaction." *Journal of Accounting and Economics*, 35 (2003), 101–116.
Mikhail, M.; B. Walther; and R. Willis. "Do Security Analysts Exhibit Persistent Differences in Stock Picking Ability?" *Journal of Financial Economics*, 74 (2004), 67–91.
Mitchell, M.; T. Pulvino; and E. Stafford. "Limited Arbitrage in Equity Markets." *Journal of Finance*, 57 (2002), 551–584.
Mitchell M., and E. Stafford. "Managerial Decisions and Long-Term Stock Price Performance." *Journal of Business*, 73 (2000), 287–329.
Odean, T. "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53 (1998a), 1775–1798.
Odean, T. "Volume, Volatility, Price, and Profit when all Traders are Above Average." *Journal of Finance*, 53 (1998b), 1887–1934.
Ofek, E., and M. Richardson. "DotCom Mania: The Rise and Fall of Internet Stock Prices." *Journal of Finance*, 58 (2003), 1113–1137.
Palia, D. "The Impact of Regulation on Manager Labor Markets." *Rand Journal of Economics*, 31 (2000), 165–179.
Shleifer, A., and R. Vishny. "The Limits of Arbitrage." *Journal of Finance*, 52 (1997), 35–55.
Stickel, S. "The Anatomy of the Performance of Buy and Sell Recommendations." *Financial Analysts Journal*, 51 (1997), 25–39.
Womack, K. "Do Brokerage Analysts' Recommendations Have Investment Value?" *Journal of Finance*, 51 (1996), 137–167.

https://doi.org/10.1017/S0022109000002453 Published online by Cambridge University Press