What Are Analysts Really Good At?

Ohad Kadan, Leonardo Madureira, Rong Wang, and Tzachi Zach*

July 2014

Abstract

Sell-side analysts employ different benchmarks when defining their recommendations. A 'buy' for some brokers means the stock is expected to outperform its industry, while for other brokers it means the stock is expected to outperform the market, or some absolute target return. We use these benchmarks to analyze the role of stock picking, industry picking, and market timing in the investment value of stock recommendations. Analysis of the relation between analysts' recommendations and their forecasts suggests that analysts abide by their benchmarks. We find strong evidence that the investment value of stock recommendations stems from stock picking within a particular industry. We find no evidence of either industry picking or market timing skills.

JEL Classifications: G10, G24

Keywords: Analysts; Benchmarks; Stock Picking; Industry Picking, Market Timing

*Ohad Kadan (kadan@wustl.edu) is at the Olin Business School, Washington University in St. Louis; Leonardo Madureira (leonardo.madureira@case.edu) is at the Weatherhead School of Management, Case Western Reserve University; Rong Wang (rongwang@smu.edu.sg) is at the Lee Kong Chian School of Business, Singapore Management University; and Tzachi Zach (zach_7@fisher.osu.edu) is at the Fisher College of Business, Ohio State University. We thank David Aboody, Roby Lehavy, Joshua Livnat, Roni Michaely, Brian Rountree, Kent Womack, Pei-Fang Hsieh, Roger Loh, seminar participants at Case Western Reserve University, New York University, Washington University in St. Louis, Ohio State University, and conference audience at the 2012 IDC Summer Conference, the 2012 FARS Conference, the 2012 FIRS Conference, the 2012 Asian Finance Conference, the 2012 FMA Conference, the 2013 AFA Conference, and the 2013 Financial Economics and Accounting Conference for helpful comment and suggestions. An earlier version of this paper was circulated under the title "Stock Picking, Industry Picking and Market Timing in Sell-Side Research."

1 Introduction

It is well known that sell-side research analysts publish investment advice on stocks in the forms of recommendations such as 'buys,' 'holds,' and 'sells.' However, not all buys/holds/sells are created equal. An inspection of the disclosures in which analysts describe the meaning of their recommendations reveals that different brokers assign different meanings to their recommendations. For example, in one broker a 'buy' might mean that the stock is expected to outperform its industry peers (we call this broker an "industry benchmarker"); in another a 'buy' might mean that the stock is expected to outperform the market ("market benchmarker"); and in yet another, a 'buy' might mean that the stock is expected to earn a return that exceeds some pre-determined threshold such as 10% ("total benchmarker"). Thus, 'buy' recommendations from different brokers carry with them very different literal meanings and investment advice.[1]

We rely on these different benchmarks to explore analysts' different abilities as they are reflected in stock recommendations. It is standard in the literature that market professionals (analysts, money managers, etc.) can potentially provide three types of insights about future stock performance: stock picking, industry picking, and market timing. Stock picking is the ability to rank stocks within a small group of similar stocks such as an industry. Industry picking is the ability to identify hot and cold industries. Market timing is the ability to predict the future performance of the entire market. There is, however, a big debate as to whether market professionals can actually deliver these three different insights—particularly market timing— to their clients.[2]

In this paper we shed light on this debate by investigating how these abilities are manifested in the investment advice from sell-side analysts. Partitioning the sample of recommendations based on the different benchmarks provides a unique opportunity to better isolate the three abilities and directly test for their presence. Because recommendations from industry benchmarkers aim at

[1] Unless otherwise noted, we use the term 'buy' to refer to optimistic recommendations, thus including both 'strong buy' and 'buy' recommendation levels, while 'sell' refers to recommendations with a pessimistic tone, thus including both 'sell' and 'strong sell' recommendations levels.

[2] There is evidence that analysts demonstrate stock picking in firm recommendations (Boni and Womack, 2006) and industry picking in industry recommendations (Kadan et al., 2012). Market timing has been more elusive: The ability is not demonstrated by investment newsletters (Graham and Harvey, 1994, 1996, 1997), hedge fund managers (Fung, Xu, and Yao, 2002) and pension fund managers (Goggin, Fabozzi, and Rahman, 1993), while for mutual fund managers the evidence is mixed (Treynor and Mazuy, 1966; Henriksson, 1984; Grinblatt and Titman, 1994; Ferson and Schadt, 1996; and Becker, Ferson, Myers and Schill, 1999, do not find evidence of market timing, while Bollen and Busse, 2001; and Jiang, Yao, and Yu, 2007, show evidence in favor of it).

beating industry peers, they are expected to reflect only stock picking. Recommendations from market benchmarkers, whose objective is to outperform a market index, are expected to incorporate both stock picking and industry picking. Finally, recommendations from total benchmarkers are compared to an absolute return threshold, and are thus expected to reflect all three types of abilities.

Our main research question asks whether analysts possess any one of the three abilities. To address this question we proceed as follows. First, given that the data on benchmarks have not been extensively studied previously, we begin our exploration by providing some descriptive analysis of the nature of these benchmarks. Second, and more importantly, we verify whether analysts abide by their benchmarks. To do so, we examine the extent to which these benchmarks affect the way analysts incorporate fundamental information into their stock recommendations. Third, we examine the overall performance of each recommendation, taking into considerations the benchmark that is being used. Finally, we decompose the stock returns following each recommendation, allowing us to analyze the three aforementioned abilities.

Beginning in September of 2002, and following Rule NASD 2711, Rule NYSE 472, and the Global Settlement, brokers are required to define in each report the literal meaning of their recommendations, including the benchmark to be used when interpreting the recommendation advice. To examine our research questions we hand-collect, mostly from full-text analyst reports, the meaning of recommendations for 173 brokers accounting for over 94% of all recommendations issued during our sample period (September 2002-December 2009). We find that the most prevalent benchmarks are industry benchmarks (21% of brokers), market benchmarks (20% of brokers), and total benchmarks (25% of brokers). Other brokers typically use either combinations or risk-adjusted versions of these three benchmarks. Given their popularity, the simplicity of their meaning, and because they provide a more intuitive mapping to analysts' abilities, we focus our empirical analysis on brokers employing these three benchmarks exclusively.

It is possible that the benchmarks are a pure formality, and that they are ignored by analysts when they issue recommendations. We examine this conjecture by asking whether brokers indeed abide by their benchmarks. To answer this question we relate stock recommendations to analysts' outputs regarding firms' fundamentals. We expect that industry benchmarkers would practice stock picking by primarily using within-industry information about those fundamentals, while market and total benchmarkers—who profess to use both stock picking and industry picking—would also rely on across-industry information. We test this conjecture by examining how analysts' recommendations are related to other types of forecasts issued by analysts. To this end, we break down analysts' earnings and long-term growth (LTG) forecasts into within- and across-industry components. Our analysis shows that, as expected, market and total benchmarkers place more weight on across-industry expectations than industry benchmarkers when forming their recommendations. We also find evidence that total benchmarkers attempt to incorporate market timing in their recommendations. In particular, compared to market benchmarkers, total benchmarkers incorporate into their recommendations more negative news about the economy during the 2007-2009 recession. These results are consistent with analysts indeed abiding by their benchmarks.

Next, we examine whether recommendations based on a particular benchmark are successful in meeting (or beating) their performance objectives. To this end, we collect for each broker the target return associated with its benchmark. For example, a target return for a 'buy' recommendation issued by an industry (market) benchmarker specifies by how much the recommended firm is expected to beat the industry (market). Similarly, a target return for a 'buy' recommendation issued by a total benchmarker specifies an absolute return such as 10%. We then examine whether and by how much the return of a recommended firm meets or beats its stated objective—which considers both the benchmark (industry or market) and the target returns—within a year or until the recommendation is changed.

About 50% (58%) of 'buy' ('sell') recommendations issued by industry and market benchmarkers meet or beat their objective, compared to 39% (36%) of 'buy' ('sell') recommendations issued by total benchmarkers. The higher success rates of industry and market benchmarkers compared to total benchmarkers is also apparent when we examine the difference between the actual returns and the stated objective. For example, 'buy' recommendations issued by industry (market) benchmarkers beat their objective by an average of 3.12% (5.52%), while the average return following 'buy' recommendations issued by total benchmarkers is 4.81% lower than their target return. These results seem plausible, as meeting the objective for total benchmarkers is quite a heroic task. Indeed, total benchmarkers are expected to predict firm-specific returns, industry returns, and market returns.

When comparing a recommendation return with its stated objective, we are evaluating the analyst performance based on the literal meaning of her recommendation advice. While this evaluation method is relevant, it might be inadequate to capture the *incremental* insights offered by the analyst. One concern, as mentioned above, is that the stated objective might simply be too tough. Another concern is that this method does not control for the risk profile of the recommended stocks, thus *crediting* to the analyst any performance that is in fact coming from loadings on risk factors. In other words, one needs to establish a baseline against which to evaluate the analyst.

We define the baseline for the performance of a recommendation as the performance of a firm that did not receive a recommendation, but has similar risk to that of the recommended firm. To implement this, we use a propensity score methodology to match each actual recommendation (i.e., a firm receiving a recommendation at some point in time) to a control unit (some other firm and another point in time) with a similar risk profile. We compare the returns in excess of the stated objective between the actual recommendations and their associated control units. We find that for all types of benchmarks, firms for which analysts issue 'buy' ('sell') recommendations perform better (worse) than firms with similar risk characteristics that did not receive such recommendations. In particular, while the 39% success rate in meeting or beating its stated objective for 'buys' from total benchmarks seems at first to denote a poor performance, it is in fact a significant improvement over a baseline success rate of 32%.

Having attested that recommendations perform better than what their risk characteristics would imply, in our final analysis we explore the sources of this superior performance. We ask whether analysts possess any one of the three abilities: stock picking, industry picking and market timing. To evaluate these three abilities, we decompose the returns in excess of the recommendations' stated objective into components that measure each such ability. For example, for market benchmarkers the excess return following a recommendation (the difference between the firm return and the market return) is split into two components: (i) the difference between the firm return and its industry's return captures stock picking; (ii) and the difference between the industry return and the market return captures industry picking. Similar to the previous analysis, we compare each return component of an actual recommendation to that of its control unit.

We document strong evidence of stock picking ability across all types of analysts. For example, for market benchmarkers the returns associated with 'buy' recommendations exceed industry returns by 521 basis points compared to 135 basis points for the control units. This is consistent with the evidence in Boni and Womack (2006), who find that analysts are good at ranking firms within industries. On the other hand, our results do not indicate any industry picking ability for market or total benchmarkers. This contrasts with the evidence in Kadan et al. (2012), who demonstrate that *industry* recommendations issued primarily by *strategy* analysts do reflect industry picking.[3] Finally, we do not find evidence of market timing among total benchmarkers. Thus, our evidence suggests that analysts' skills are limited to stock picking: The performance of their stock recommendations is driven by the ability to pick winners and losers within an industry, even for analysts who profess, and try, to incorporate industry picking and/or market timing into their recommendations.

We contribute to the literature in several ways. First, we provide a comprehensive analysis of sources of the performance of sell-side analysts. In particular, we study how stock picking, industry picking, and market timing play a role in shaping analysts' stock recommendations. We are able to do so by relying on partitioning the sample of stock recommendations based on the benchmarks used by different brokers. This partitioning enables us to better analyze each ability because different analysts profess to use different sets of abilities. In particular, only total benchmarkers claim to incorporate market timing. As a result, we increase the power of the test that evaluates the presence of market timing by restricting it to the sample of total benchmarkers. To the best of our knowledge, this is the first study that evaluates market timing in stock recommendations. In addition, we contribute to the literature on stock picking (Boni and Womack, 2006), and on industry picking (Kadan et al., 2012).

Second, in Kadan, Madureira, Wang, and Zach (2012) we study different aspects of analysts' industry expertise. In one of the analyses we point out the existence of sell-side benchmarks, and use a small sample of disclosures from 20 brokers to study the relation between firm and industry recommendations. In contrast, in this paper we focus exclusively on these sell-side benchmarks, for which we provide the first large scale and comprehensive analysis. Thus, we contribute to the literature by documenting the attributes of these benchmarks, exploring the way in

[3] Strategy analysts (or strategists) are analysts who work in the economics and strategy group of brokerage houses. Unlike the security analysts studied in this paper, strategists typically do not cover individual firms, but rather research the equity market as a whole. As part of their research, strategies often issue recommendations for entire industries. These recommendations are the subject of investigation in Kadan et al. (2012). They are fundamentally different from the firm recommendations issued by firm-level analysts, which are the subject of this paper.

which they are reflected in analysts' recommendations, and by studying their implications for investment value. Bradshaw (2012) emphasizes the importance of these benchmarks for the study of sell-side research.

Third, our paper also relates to a long strand of literature examining the relation between stock recommendations and other outputs produced by analysts such as earnings forecasts, price-targets, and long-term forecasts (e.g., Bradshaw, 2004; Ertimur, Sunder and Sunder, 2007; Chen and Chen, 2009; Barniv et al., 2009; Brown and Huang, 2010; Kecskes, Michaely and Womack, 2010). Our analysis emphasizes that the usual method to assess the relation between recommendations and other analysts' outputs can be improved upon. When regressing recommendations on expectations of earnings and LTG, for example, we observe an inconsistency in that recommendations can be industry-adjusted statements (in the case of industry benchmarkers), while expectations of earnings and LTG are not.

Finally, in analyzing whether recommendations perform as predicted, we depart from the usual approach taken in the literature. For the most part, the literature has assessed the value of analysts' recommendations through the investment value obtained from following a set of recommendations, for example by looking at risk-adjusted returns relative to CAPM or a multifactor model, obtained from portfolios formed based on recommendations (e.g., Womack 1996; Barber, Lehavy, McNichols and Trueman, 2001 and 2006; Jegadeesh, Kim, Krische, and Lee, 2005). While this approach is useful from the perspective of an investor that diversifies her investment over many recommendations, we argue that this is at best an imperfect measure of whether each recommendation performs according to its objective. Nothing in the disclosed meaning of a recommendation suggests that it should be seen as a prediction about risk-adjusted performance (other than benchmark-adjusted performance), nor that it should be assessed after it is combined with *other* recommendations. Instead, the literal meaning of a recommendation provides a very clear predictive rule about how its advice should be taken. Our assessment of the recommendation value follows this rule directly.

We proceed as follows. Section 2 describes the data. Section 3 provides some preliminary analysis of the benchmarks used by different brokers. In Section 4 we examine whether analysts abide by their benchmarks. In Section 5 we explore whether analysts are successful in meeting their benchmark-specific targets, and evaluate whether stock recommendations reflect any one of the three abilities: stock picking, industry picking, and market timing. Section 6 concludes.

2 Data

We focus on analysts' stock recommendations of all U.S. firms in the period of September 2002 to December 2009. The source for the analyst recommendations, earnings forecasts and LTG projections is the IBES database. The data on firm characteristics are from COMPUSTAT. We obtain stock returns from CRSP, and equity offerings data from SDC. Industry membership is inferred through the industry classification defined by the Global Industry Classification Standard (GICS) obtained from COMPUSTAT. The GICS system is widely adopted by investment banks as an industry classification system, and has been increasingly used in academic studies—e.g., Bhojraj, Lee, and Oler (2003), Boni and Womack (2006) and Kadan et al. (2012).

We manually collect data on the benchmarks used by brokers that issued at least 100 recommendations during our sample period. There are 249,459 recommendations issued by all brokers during our sample period for U.S. firms, out of which 234,274 are issued by brokers with at least 100 recommendations. Therefore, the threshold of 100 recommendations enables us to concentrate our effort on collecting benchmark data of large brokers without significant loss of recommendation data.

We start by examining the disclosures of analysts regarding the meaning of their firm recommendations. We collect disclosures from three sources. First, we retrieve information from full-text research reports in the Investext database for brokerage houses whose reports are available. Under regulations NASD Rule 2711 and NYSE Rule 472, which were adopted in mid-2002, prior to the beginning of our sample period, analysts are required to disclose the exact meaning of their recommendations inside their reports. Analysts normally disclose the information on the ratings system, ratings distribution, and the meaning of different ratings in the last section of their reports. Secondly, for brokerage houses not appearing in Investext, we collect data from the Investars website,[4] which contains the ratings' definitions of some brokers. Finally, if necessary, we obtain data directly from brokers' websites.

< *Insert Table 1 here* >

[4] http://www2.investars.com/synopsis.asp

We rely on the analysts' disclosures to identify the benchmark they use to define their recommendations. We categorize brokers into ten different types of benchmarks. Table 1 summarizes these benchmarks and gives examples of textual descriptions from the analysts' disclosures. The three most basic benchmarks involve determining recommendations according to the expected performance of the covered stock compared to the performance of industry peers, the performance of the market, or to some return threshold. More formally, we classify brokers as industry benchmarkers if they state that their stock recommendations are benchmarked against industry performance. For example, Smith Barney's analysts rate stocks based on the "stock's performance vs. the analyst's industry coverage for the coming 12-18 months." We classify brokers as market benchmarkers if they state that their stock recommendations are benchmarked against market performance. For example, Wachovia's analysts rate a stock based on its expected performance "relative to the market over the next 12 months." Finally, we classify brokers as total return benchmarkers if they issue recommendations based on a stock's expected total return. This is the case, for example, with Deutsche Bank, where a 'buy' recommendation means that the stock's total return is "expected to appreciate 10% or more over a 12-month period."

Occasionally brokers determine their recommendations using some combination of these three basic benchmarks. We identify four such combinations. For example, Dougherty & Co combines features of market and industry benchmarks, so that its 'buy' means the corresponding stock is "expected to outperform the broader market and/or its sector." We categorize this broker as a market/industry benchmarker. Other hybrids we identify are total/market, industry/total, and market/industry/total.

Other brokers refine the basic benchmarks by adding a risk-adjustment feature. For example, Morgan Stanley establishes its recommendations based on the "stock's total return vs. analyst's coverage on a risk-adjusted basis." Notably, the nature of the adjustment for risk is often vague. In order to highlight this feature, we add a new category and classify Morgan Stanley as an industry/risk benchmarker. Similarly, we classify a broker as market/risk (total/risk) when the benchmark involves comparing the stock's expected performance to the market (a total threshold) on some type of risk-adjusted measure.

We also notice some brokers who changed their benchmarks during our sample period. For example, Merrill Lynch used a total benchmark between September 2002 and May 2008, and an industry/total benchmark since June 2008. In this case, we classify Merrill Lynch as a total benchmarker between September 2002 and May 2008, and as an industry/total benchmarker between June 2008 and December 2009. However, for some brokers, we failed to identify the exact date of the change. We classify such instances as a "Changes" category. Finally, some brokers could not be classified in any of the above categories, either because we could not find any data on their analysts' disclosures or because their disclosures did not fall into any of the above categories.

< Insert Table 2 here >

Table 2 summarizes the distribution of the different benchmarks.[5] There are 37 brokers that use the industry benchmark during our sample period, and the number of recommendations issued by such brokers accounts for about 32% of all recommendations. The number of brokers relying on a market benchmark is 34, and those brokers issued about 18% of all recommendations. There are 42 brokers that base their recommendations on a total return benchmark, and as a group they issued about 23% of all recommendations. The relevance of these basic benchmarks is apparent also when one looks at the size of each broker: Among the twenty largest brokers (according to the number of recommendations issued during our sample period), nine brokers use an industry benchmark, three brokers use market benchmark, and four brokers use total return benchmark.

Brokers using risk-adjusted benchmarks are usually big brokers, as revealed by the average number of recommendations issued by brokers in each category (Morgan Stanley is one such case), but there are relatively few of them. Therefore, as a group, these brokers account for just 11% of recommendations. Similarly, there are few brokers combining the basic benchmarks. Finally, we fail to collect data on benchmarks for 41 brokers, but these brokers are relatively small (with an average number of recommendations of 408 during the sample period), and as a group they issued about seven percent of recommendations in our sample.

In this paper we focus our attention on the three basic benchmarks. Three reasons drive our choice. First, we want to address a set of benchmarks that is representative of the universe of brokers. Industry, market, and total return benchmarkers thoroughly satisfy this requirement: Together they account for about 74% of the recommendations in our sample period, and they are adopted by 16 of the 20 largest brokers. Second, we need to address benchmarks that have a

[5] Overall, there are 173 brokers with at least 100 recommendations issued during the sample period, and 11 of them change their benchmarks during our sample period. Therefore, the total number of brokers in panel A of Table 2 is 184.

straightforward interpretation, so that clear testable hypotheses can be developed. This requirement again favors the three basic benchmarks, as they are the most precisely defined, particularly when compared to the risk-adjusted benchmarks (which do not properly document the meaning of their risk-adjustment feature) or to the benchmarks that combine more than one basic benchmark. Finally, the basic benchmarks allow for an intuitive mapping of the sets of abilities (among stock picking, industry picking and market timing) to the type of benchmarks.

3 Preliminary Analysis

3.1 Benchmark Determinants

The analysts' disclosures document that different brokerage houses rely on different benchmarks. One obvious question is why. Analysts we have interviewed hinted at a tension about which benchmark should be used. Some analysts suggest that using an industry benchmark fits well with the structure of research departments in brokerage houses, where analysts work in industry groups and are deemed industry specialists (e.g., Boni and Womack, 2006; Kadan et al., 2012). Some analysts also pointed out that ranking firms within an industry arises directly from application of techniques such as comparables.

Others expressed preference towards a total benchmark, given that a total return expectation is a direct product of applying a discounted cash flow (DCF) methodology. They also argued that an expectation about total return, as opposed to the return relative to the industry or to the market, is the most useful output from the perspective of investors. Finally, some argued that the market benchmark makes sense as well, since it is common practice to evaluate each equity asset relative to the market (or a popular index such as the S&P 500).

To add to this anecdotal evidence and provide some large sample results on the determinants of the benchmarks, we explore their possible association with brokers' characteristics. We estimate logistic models for the probability of adopting a certain benchmark. Each observation in these models is a broker-year pair, describing the benchmark used by the broker in that particular year.[6] The models presented differ in the definition of the dependent variable. As explanatory variables

[6] We also estimated similar cross-sectional regressions separately for each year during the sample period. The results are similar.

we use broker and analyst characteristics (age, size, number of industries covered, experience) as well as characteristics of the covered firms (size and book-to-market).

< Insert Table 3 here >

Table 3 presents the results. Two variables emerge as strong determinants of the choice of benchmark. The first is broker size—measured by the number of recommendations issued by a broker as a fraction of all recommendations issued during the year. Larger brokers are more likely to adopt an industry benchmark as opposed to either market or total benchmarks. It may be that large brokers that employ a large number of analysts can allow analysts to focus on a select group of firms in one particular industry, leading to more industry specialization and thereby to industry benchmarking. The second determinant is the number of industries covered. A larger number of covered industries is associated with a higher likelihood of adopting a market or total benchmark. It may be that brokers that follow many industries have a better perspective of the market, and thereby are more capable of benchmarking their recommendations to a market or total reference.

We also examined the potential linkage between the organizational structure of a broker and the benchmark it adopts. Arguably, organizing analysts by industry inside the brokerage house is less relevant for market/total benchmarkers as opposed to industry benchmarkers. We test for this possibility by comparing industry concentration of the broker's analysts across the different types of benchmarks by following the methodology suggested in Boni and Womack (2006). The results (unreported for brevity) indeed suggest that analysts employed by industry benchmarkers tend to concentrate in single industries more than their counterparts employed by market and total benchmarkers, though the differences in concentration are rather small.

3.2 Benchmark Choice and Distribution of Recommendations

Next we examine whether the choice of the benchmark is associated with the characteristics of the recommendations issued by a broker. Table 4 and Figure 1 report the distribution of recommendations broken down by the benchmark adopted by the broker. In the computation of the recommendation levels, we consider 'strong buys' and 'buys' as optimistic recommendations and assign them together a value of 1; 'holds' are assigned a value of 2; and 'sells' and 'strong sells' are considered pessimistic recommendations and are assigned a value of 3.

The table demonstrates an important and salient feature that distinguishes the behavior of industry benchmarkers from market and total benchmarkers: Industry benchmarkers tend to be less optimistic. Average recommendation levels from industry benchmarkers are significantly higher as compared to the average recommendation from market and total benchmarkers.[7] Moreover, for each year during our sample period industry benchmarkers show a smaller proportion of optimistic recommendations and a larger proportion of pessimistic recommendations compared to market or total benchmarkers. The gap between industry vs. market and total benchmarkers has diminished over the years, especially due to the industry benchmarkers reducing their share of pessimistic recommendations, but it is still significant at the end of the sample. Notably, market and total benchmarkers behave very similarly, especially with respect to the issuance of pessimistic recommendations.

< Insert Table 4 and Figure 1 here >

Table 5 further explores the relation between benchmark choice and broker optimism in a multivariate setting. We use firm fixed-effects logistic regressions including all recommendations during our sample period. The dependent variable is an indicator equal to one when the recommendation is optimistic in model (1) and pessimistic in model (2). Given the similarity in the distribution of recommendations from market and total benchmarkers, we compare these two benchmarks, as a group, with the industry benchmarkers. Our main explanatory variable is an indicator for benchmark adopted by the broker issuing the recommendation: It is equal to one if the broker is an industry benchmarker and zero otherwise.

< Insert Table 5 here >

The choice of which control variables to adopt is made easier by the firm fixed-effects specification, since it frees us from having to include firm characteristics that are not varying over time. So, instead, we focus on some broker characteristics and time-varying aspects that have been shown in prior studies to affect the optimism of brokers. There is a long literature relating conflicts of interest stemming from the relationship between investment banking and sell-side research to the optimism in analyst recommendations (e.g., Lin and McNichols,1998; Michaely and Womack, 1999). We use a broker affiliation dummy to proxy for such conflicts of interest. The affiliation

[7] In the computation of the average recommendation, 'strong buys' and 'buys' are mapped to level 1, 'holds' are mapped to level 2, and 'sells' and 'strong sells' are mapped to level 3.

dummy variable is equal to one if the broker issuing the recommendation was a lead underwriter or a co-manager in an equity offering for the firm in the 24 months before the recommendation announcement date. We also control for past market and firm performance, based on the evidence that analysts chase momentum (Jegadeesh, Kim, Krische, and Lee, 2004), and for broker and analyst characteristics. SANCT is an indicator equal to one if the recommendation is issued by an analyst who is employed by a brokerage house that was sanctioned during the Global Settlement (Barber, Lehavy, McNichols, and Trueman, 2006; Kadan, Madureira, Wang, and Zach, 2009). TIER3 is an indicator variable for whether a brokerage house uses a three-tier recommendation grid at the time a recommendation is issued (Kadan, Madureira, Wang, and Zach, 2009). Finally, we control for the experience of the individual analyst issuing the recommendation, measured as the number of days the analyst has appeared in IBES.

The results confirm the univariate inferences in Table 4, showing that the benchmarking decision is strongly associated with the bullishness of the recommendations. Industry benchmarkers are less likely to issue optimistic recommendations and more likely to issue pessimistic recommendations as compared to market and total benchmarkers.[8]

It is documented that analysts have a tendency to be overly optimistic for the subjects they cover (e.g., McNichols and O'Brien, 1997). One possible explanation for this optimism is that analysts become attached to the subjects of their coverage—be it firms or industries. Since industry benchmarkers rank firms within their industry, their firm recommendations suffer from only one source of optimism: their attachment to the firms they cover. By contrast, market and total benchmarkers incorporate both their firm and industry views into their firm recommendations. Hence, their firm recommendations might suffer from two sources of optimism. As a result, the distribution of recommendations coming from market and total benchmarkers is tilted toward optimism when compared to that of industry benchmarkers.

[8] One way to reinforce the association between a broker's benchmark and the distribution of the broker's recommendations is to look at instances where a broker changes its benchmark. We identify four events where both the old and the new benchmark are one of the three basic benchmarks analyzed here. In two of them (both changes from total to industry benchmarker), no significant change in the distribution of recommendations follows the change in benchmark. In the other two, though, there is a significant increase in the fraction of 'sell' recommendations around the event of change in benchmark: a jump from 5% to 12% in the case of a change from market to industry benchmarker, and from 3% to 17% in the case of a change from total to industry benchmarker.

4 Do Analysts Abide by their Benchmarks when Issuing Recommendations?

That an analyst asserts that her recommendation advice should be interpreted according to some specific benchmark does not imply that the benchmark is actually used when the advice is determined. In fact, the common structure of research departments along industry groups raises the possibility that all analysts determine their recommendation advice through the ranking of their coverage universe regardless of the stated benchmark. That is, recommendations could be based on stock picking ability alone. In this section, we empirically examine whether benchmarks are relevant to the way recommendations are formed. In particular, we examine whether, and to what extent, the different abilities—stock picking, industry picking and market timing—associated with each benchmark are used by the analysts when they determine their recommendations. Answering this question is important both for validating the analysts' disclosures and for better interpreting stock recommendations.

4.1 Stock Picking vs. Industry Picking

What are the implications of the proper usage of each benchmark with respect to the stock picking and industry picking abilities? Consider first analysts declaring the use of an industry benchmark. According to their disclosures, stock recommendations are statements about the analysts' expectations on how stocks will perform relative to their industry peers; that is, these analysts rely on stock picking but not on industry picking. By contrast, market and total benchmarkers would determine their recommendations by relying on their expectations of both the firm performance relative to the industry (stock picking) and the industry's overall performance relative to the market (industry picking). The challenge is that the analyst's expectations about these different components are unobservable. For example, when a market benchmarker issues a buy, stating that she expects the stock to outperform the market, we do not know her true expectation of the firm performance relative to the industry or her expectation of the industry performance relative to the market.

However, some measures of analysts' expectations are observable. Besides issuing recommendations, analysts also consistently release forecasts about the firm's upcoming earnings and about the firm's long-term growth (LTG). Our strategy is thus to rely on the analysts' revealed expectations in order to assess whether benchmarks are in fact used when recommendations are formed. In considering the relation between analysts' recommendations and analysts' other outputs such as earnings and LTG forecasts, we are following a long literature (e.g., Bradshaw, 2004; Ertimur, Sunder and Sunder, 2007; Chen and Chen, 2009; Barniv et al., 2009; Brown and Huang, 2010; Kecskes, Michaely and Womack, 2010). One way to analyze this relation is to regress recommendations on measures of analysts' earnings and LTG forecasts. A typical model looks like

$$\text{Rec} = \beta_0 + \beta_{LTG}\text{LTG} + \beta_{E/P}\text{E/P} + \varepsilon, \tag{1}$$

where Rec is an integer mapping the recommendation levels—for example, 'optimistic' recommendations are mapped to 1, 'neutral' to 2, and 'pessimistic' to '3'. The independent variables are obtained from the analysts' expectations about LTG and earnings. Given that the earnings number is mechanically linked to the number of outstanding shares (and the prevalence of the use of comparables techniques by sell-side analysts when analyzing companies), the earnings-price ratio is used instead of the raw measure of earnings per share estimates. To avoid extreme values in the independent variables, researchers use rankings of the LTG and E/P measures, where values are scaled to range evenly between 0 and 1. The results in the literature show that the coefficients β_{LTG} and $\beta_{E/P}$ are negative: Higher expectations about LTG and forward earnings-price ratios are associated with lower levels of—that is, more optimistic—recommendations.[9]

The model above needs to be revamped if brokers rely on different benchmarks when determining their recommendations. To see this point, consider industry benchmarkers. For these brokers, while recommendations are just a ranking relative to industry peers, expectations about earnings-price ratios and LTG are by nature absolute, and do not immediately translate into an industry ranking. There is, thus, an inconsistency between the left-hand side (LHS) and right-hand side (RHS) variables: The LHS variable, the recommendation, *is* industry-adjusted while the RHS variables are *not*.

We aim at extending model (1) in a way that will capture both within- and across-industry relative expectations. To see the idea, suppose we have the analyst's expectations about (i) how the firm's LTG compares with the LTG of its peers in the industry ("within-industry" LTG expectation); *and* (ii) how the LTG of its industry compares to the LTG of the other industries

[9] LTG and price-earnings ratios are just two examples of "valuation" proxies based on analysts' estimates that can be used in a regression model to explain recommendations. Other proxies have been explored in the literature, such as the residual income valuation model analyzed by Bradshaw (2004). We focus on the LTG and price-earnings proxies in this study for two reasons. They are the simplest and most parsimonious proxies (other proxies such as the residual income depend on further assumptions for their estimation) and their associations with recommendation levels are the most robust across the studies relating recommendations and other outputs from sell-side analysts.

("across-industry" LTG expectation). Within-industry expectation is relevant for stock picking and across-industry expectation matters for industry picking. Thus a market or total benchmarker will rely on both expectations when determining her recommendation advice, while an industry benchmarker will mostly (or totally) rely on the first component. In other words, all brokers (industry, market, or total benchmarkers) would "load" on their within-industry expectations, but industry benchmarkers would not load (or at least load less) on the across-industry expectations when compared to market and total benchmarkers.

We do not observe the within-industry and across-industry expectations directly, but we can infer them from the raw forecasts issued by the analysts. More specifically, we decompose analysts' expectations of LTG and earnings into an across-industry (AI) and within-industry (WI) components as follows. Starting with the LTG forecasts, each month we first compute for each firm the consensus LTG as the average LTG forecast amongst the outstanding forecasts available for that firm. In the next step, we define for each industry an industry LTG forecast as the average LTG consensus across all firms in that industry. Then, for each firm in that month we compute the firm's industry-adjusted LTG forecast as the firm's LTG forecast minus its industry LTG forecast. We assign each firm a score between 0 and 1 based on the ranking of industry-adjusted LTG forecasts in each industry. We denote this score by WI_LTG. For each firm we also calculate an across-industry LTG score based on the ranking of its industry LTG forecasts among all industries. The latter is denoted AI_LTG. Similarly, we calculate a within- and across-industry earnings estimate rankings denoted by WI_E/P and AI_E/P respectively, based on the analyst earnings forecast scaled by the stock price prevailing when the earnings data are collected.[10]

We then estimate the following model:

$$Rec = \beta_0 + \beta_{AI\_LTG} AI\_LTG + \beta_{WI\_LTG} WI\_LTG + \beta_{AI\_E/P} AI\_E/P + \beta_{WI\_E/P} WI\_E/P + \varepsilon, \qquad (2)$$

where Rec takes the value of 1, 2, or 3 for "optimistic," "neutral," and "pessimistic," respectively.[11] In line with the prior literature we expect all the coefficients to be negative. More relevant for our

[10] We use unadjusted measures of forecasts of 1-year ahead earnings. Forecasts that are older than 12 months are deleted. Results are robust to using 2-year ahead projections, and to relaxing the 12-months limit on the outstanding measures.

[11] Optimistic refer to 'strong buy' and 'buy' recommendations; neutral refer to 'hold' recommendations; and pessimistic refer to 'sell' and 'strong sell' recommendations. This 3-tier mapping differs from the usual 5-tier mapping adopted by the literature. The change is motivated by the sample period of our study. After 2002 (the period of our study), most of the brokers have adopted a three-tier rating system. The qualitative inferences reported here are robust to mapping the recommendations into a range of 1 through 5 (from 'strong buy' to 'strong sell', respectively).

focus, we run these models separately for industry and market or total benchmarkers. We then expect β_{AI_LTG} and $\beta_{AI\_E/P}$ for market and total benchmarkers to be *more* negative than the corresponding coefficients for industry benchmarkers.

< *Insert Table 6 here* >

We estimate models (1) and (2) using monthly regressions. The results are reported in Table 6. The table shows the Fama-MacBeth's (1973) style coefficients from averaging the monthly regressions from September 2002 through December 2009, where the standard errors for the mean coefficients are adjusted for autocorrelation (see, for example, Loughran and Schultz, 2005; Fama and French, 2002). Specifications (i) and (ii) in Table 6 show estimates of model (1), the one traditionally pursued in the literature, by which LTG and E/P are not broken into within- and across-industry components. As expected, the coefficients are significantly negative for both industry and non-industry (market or total) benchmarkers, reflecting that better views on earnings and LTG prospects of the company do translate on average into a more favorable recommendation.

In specifications (iii) and (iv) we estimate model (2) separately for industry and for market and total benchmarkers. We also estimate a model on a pooled sample that allows us to compare the coefficients related to different benchmarks (using appropriate dummy variables). The results show that both within- and across-industry expectations are incorporated into the recommendations of both analyst types as all the coefficients are negative. Take the effect of analysts' expectations of long-term growth, for example: The coefficients on both across-industry (AI_LTG) and within-industry (WI_LTG) expectations are significantly negative for all types of benchmarkers.[12] Notice, however, that the loadings on across-industry expectations are significantly higher in absolute value for market and total benchmarkers compared to industry benchmarkers (0.271 vs. 0.176 for LTG and 0.092 vs. 0.038 for E/P, both different at the 1% level). This suggests that market and total benchmarkers put more weight on across-industry expectations when issuing recommendations compared to industry benchmarkers. By contrast, we do not find a significant difference in

[12] In order to assess the economic magnitudes of these results, let's examine the model for industry benchmarkers. A change from the 25th percentile to the 75th percentile in AI_LTG is associated with a shift of 0.088 (0.5*0.176), which translates to about 18% of a standard deviation, towards a more optimistic recommendation. In comparison, a similar change in WI_LTG is associated with a shift of 0.17 (0.5*0.341), which translates to about 34% of a standard deviation, towards a more optimistic recommendation.

coefficients of the within-industry measures of expectations for LTG and E/P, suggesting that all brokers take this information into account to a similar degree when issuing recommendations.[13]

These results support the hypothesis that market and total benchmarkers do behave differently from industry benchmarkers in how they use expectations about the firms' fundamentals when determining their recommendations. Industry benchmarkers mostly rely on the ranking of a firm's fundamentals within its industry (though they also use the across-industry expectation of LTG). Market and total benchmarkers, while also ranking firms within industry, use their expectations about the firm's industry performance as compared to the other industries to a larger degree than industry benchmarkers. In other words, stock picking is practiced by all types of benchmarkers, and industry picking matters *more* for market and total benchmarkers when compared to industry benchmarkers. This behavior is consistent with the stated benchmark in the analysts' disclosures.[14]

4.2 Market Timing

A recommendation from a market benchmarker—a measure of the expected return of a firm relative to the market—can be seen as a statement about how the firm will perform relative to its industry (stock picking) plus how its industry will perform relative to the market (industry picking). A recommendation from a total benchmarker—a measure of the firm's expected absolute return—in turn can be interpreted as a statement about how the firm will perform relative to its industry (stock picking) plus how the industry will perform relative to the market (industry picking) plus how the market will perform (market timing). Thus, what distinguishes a total from a market benchmarker is the reliance on market timing.

[13] A natural concern is that the firms covered by industry and market benchmarkers are fundamentally different, and hence the results we uncover are driven by differences in the characteristics of the covered firms, rather than by the adopted benchmark. To address this issue we repeat the analysis in Table 6 for a subsample of firms that are covered by both industry and market/total benchmarkers. The results of this analysis are very similar to those reported in Table 6 (and available upon request).

[14] This methodology also sheds some light on why market and total benchmarkers are in general more optimistic than industry benchmarkers. If it was only for the within-industry expectation of the firm's fundamentals, brokers with different benchmarks would be similar in the optimism presented in their recommendations. It is the extra loading on the analysts' expectations about how the fundamentals of the firm's industry compare to the fundamentals of the other industries that distinguishes market and total benchmarkers from the industry benchmarkers. If you take two analysts having the same relative expectations about the firms and their industries fundamentals, the analyst that works based on a market or total benchmark becomes more optimistic compared to an industry benchmarker because she puts extra weights on the across-industry dimensions of her expectations. (This interpretation is made easier given that RHS variables are normalized between 0 and 1.)

We can then examine whether total benchmarkers abide by their benchmark and aim at market timing by comparing their recommendations with the recommendations from market benchmarkers. A starting point in testing for market timing in recommendations is to compare the recommendations' optimism with a proxy for expectations about market performance. Successful market timing would entail being more optimistic (pessimistic) when the market is expected to perform well (poorly), for example during expansions (recessions). Figure 1 shows, for example, that all types of benchmarkers decrease their overall optimism (measured by either a decreasing proportion of buys or an increasing proportion of sells) as the 2007-2009 recession develops.

However, overall optimism cannot be necessarily linked to market timing. For both market and total benchmarkers, optimism can also originate from the other skills—stock picking and industry picking—employed by these analysts. Therefore, we need to isolate optimism that is linked to market timing. For that, we extract the degree of optimism in recommendations after netting out the effects of stock picking and industry picking. Recall that the regression model (2) above explicitly incorporates the effects of stock picking and industry picking in shaping recommendations. In that model, we can interpret the intercept β_0 as capturing the baseline level of optimism before the effects of stock picking and industry picking abilities are incorporated.[15] In fact, it is more appropriate to interpret the intercept as capturing the pessimism in recommendations; given that the LHS variable in model (2) takes values from 1 (optimistic recommendation) to 3 (pessimistic recommendation), higher values of the intercept are associated with more pessimistic recommendations.

We estimate model (2) separately for total and for market benchmarkers. The difference between their corresponding intercepts, ($\beta_{0,Total}$ - $\beta_{0,Market}$), is the estimate of the difference in baseline pessimism between total benchmarkers and market benchmarkers.[16] We refer to this difference as 'excess pessimism' throughout the discussion. Model (2) is estimated monthly, yielding a time-series of monthly estimates of 'excess pessimism'. To test for market timing, we then compare this 'excess pessimism' with a proxy for expectations about market performance. For this proxy, we adopt the Chicago Fed National Activity Index (CFNAI). According to the Chicago Federal Reserve, the CFNAI "is a monthly index designed to gauge overall economic activity and

[15] Given the normalization of the independent variables, β_0 captures the recommendation of the least favorable firm.
[16] We combine the two estimations in one single regression, by pooling data from both market and total benchmarkers and interacting each coefficient with a "Total" dummy for the recommendations coming from total benchmarkers. The interaction of the intercept with the "Total" dummy is the estimate for the ($\beta_{0,Total}$ - $\beta_{0,Market}$).

related inflationary pressure."[17] The idea is that analysts would have a direct way to assess the overall state of the economy and, to the extent that market performance correlates with economic activity, analysts could rely on CFNAI to adjust their expectations about market performance.

< Insert Figure 2 here >

Figure 2 shows monthly estimates of ($\beta_{0,Total}$ - $\beta_{0,Market}$) and the CFNAI. While for roughly the first half of our sample period no clear pattern emerges regarding comovements between 'excess pessimism' and CFNAI, a strong negative correlation between these measures emerges during the later part of the sample, particularly during the 2007-2009 recession. There, we see CFNAI collapsing and the 'excess pessimism' booming. Formal statistical tests confirm the visual pattern. The overall correlation between 'excess pessimism' and CFNAI is -0.11 (t-stats=-2.90), though the bigger effect is in the 2nd half of the sample: The correlation for the first half of the sample is not significantly different from zero, while for the second half it stands at -0.60 (t-stats=-5.14). This is evidence consistent with the idea that total benchmarkers rely more on market timing than market benchmarkers. They become significantly more pessimistic than market benchmarkers during a recession, and this is not because of stock picking or industry picking.[18]

5 The Performance of Recommendations and How They Reflect Analysts' Abilities

5.1 General

The results in the previous section suggest that analysts do indeed take the different benchmarks into account when issuing their recommendations. Different benchmarks imply different objectives for recommendations. For industry benchmarkers the objective is to beat the industry peers; for market benchmarkers it means beating the market; and for total benchmarkers it

[17] The monthly index is a weighted-average of 85 monthly indicators published by the Chicago Fed. The index is designed to have an average value of zero and standard deviation of one. Values above (below) zero indicate economic growth above (below) trend. See http://www.chicagofed.org/webpages/publications/cfnai/index.cfm for more information.

[18] Two caveats are in order. The first is power. Given that we rely on monthly measures of ($\beta_{0,Total}$ - $\beta_{0,Market}$) and CFNAI, our inferences are based on only 88 data points. Second, we are assuming that the analyst's expectation of market performance is captured by the CFNAI measure. CFNAI measures current economic activity, but what the analyst incorporates in the stock recommendation advice is her expectation of market return over the next year. It is possible that the analyst gets her expectations from other sources. Nevertheless, it is unlikely that the analyst—or any type of financial expert—would ignore the economic indicators when attempting to assess market performance. (During our sample period, for example, the CFNAI and the market return have a significant positive correlation of 0.40.) This is particularly true during what has been named the "great recession." It is hard to make a case that analysts would be bullish about the market during the harsh economic times (as indicated by the CFNAI) between 2007 and 2009.

means beating some absolute threshold. In this section, we analyze the performance of analysts based on whether the recommended stocks behave "as promised" in the analysts' disclosures, meeting or beating their declared objective. We then explore the sources of performance in terms of stock picking, industry picking, or market timing.

< *Insert Table 7 here* >

In order to ascertain whether the recommendation's objective is achieved, we first take a closer look at how analysts state their objectives. Besides the benchmark, the recommendation's objective (or, its literal meaning) carries a target threshold as well, and this threshold varies across brokers. For example, in the case of a 'buy', some analysts may expect the recommended stock return to surpass the benchmark return by 10%, while others may require a 5% outperformance.[19] Table 7 presents summary statistics of the target thresholds used by the brokers in our sample. Panel A shows the thresholds used by market benchmarkers. The most frequent target is zero, saying that a typical 'buy' recommendation issued by a market benchmarker means that the recommended stock's return will exceed the market return over the forecast horizon. This threshold is used by 20 out of the 34 market benchmarkers in our sample. Panel B shows that for industry benchmarkers the most common threshold is also zero, which corresponds to the expectation that the stock's return of a buy recommendation exceeds the industry return over the forecast horizon. Finally, Panel C presents the threshold distribution for total benchmarkers. Here, the most prevalent threshold is 15%, which corresponds to the expectation that the total return of a stock with a buy recommendation over the forecasted horizon should be at least 15%. Though, notably, in this case targets of 10% or 20% are also quite popular. Target thresholds for 'sell' recommendations are typically symmetric, and are not reported for brevity.

5.2 Methodology

We evaluate whether the recommendation's objective has been achieved in two ways. In the first approach, we simply compare the performance of the recommendation to its stated objective, as follows: (i) If the recommendation has not been changed for a year, we compare the cumulative stock return during the year to the stated objective; (ii) If the recommendation advice has been

[19] The literal meaning of the recommendation also includes the forecast horizon: how long should it take for the recommendation prediction to materialize. In this case, though, a very common trend emerges, with the vast majority of the brokers working on a 12-month horizon. In a few cases, the broker adopts a range for its forecast horizon (for example, saying that the recommendation is based on the "stock's performance vs. the analyst's industry coverage for the coming 12-18 months"), though in these situations the 12-months period tends to be part of the declared range.

changed within 12 months after it was issued (e.g., through a cancelation or an upgrade/downgrade), we compare the cumulative stock return until the end of the day when the recommendation was changed to the stated objective. We thus refer to the target date of a recommendation as the earlier of 12 months and the date in which the recommendation advice has been revoked.[20] Under this approach, we follow the literal meaning of the recommendation's stated objective, without accounting for risk. This is consistent with how the analysts' employers and the institutional investors most often judge recommendations' performance.[21]

In the second approach, we also consider the risk profile of stock recommendations. We want to isolate any performance that is associated with loadings on risk factors, and only measure performance that is due to some key insights offered by the analysts. To do so, we match each recommendation (a firm i that receives a 'buy' or 'sell' at time t) to a control unit (another firm i_c and another time period t_c) such that firm i at time t and firm i_c at time t_c have a similar risk profile based on the four Fama-French factors: beta, size, book-to-market and momentum. The matching procedure is based on the nearest neighbor matching of propensity scores (Rosembaum and Rubin, 1983). The propensity score matching procedure has the appealing feature of solving the problem of the "curse of dimensionality" that appears when matches over multiple dimensions are required, and has been used in many different corporate finance settings (e.g., Bharath et al., forthcoming; Drucker and Puri, 2005; Villalonga, 2004; Colak and Whited, 2007; Hellman, Lindsey and Puri, 2008). We provide a detailed discussion of the matching procedure in Appendix A.

5.3 Results

Panel A of Table 8 presents the fraction of 'buy'/'sell' recommendations that meet their stated objective. We show this success rate broken down by the three different benchmarks, and we

---

[20] In other words, a recommendation is evaluated throughout its stated life span as long as its advice is still outstanding. This definition of the life span of a recommendation is similar to the approach used in the literature when examining the investment value of recommendations. When forming portfolio based on recommendations, stocks are included in a portfolio when a new recommendation appears, and the stock is kept in the portfolio until the earlier date between (1) the end of the stated life span of the recommendation and (2) the date when the recommendation advice is revoked. See, for example, Barber, Lehavy, McNichols, and Trueman (2006) and Barber, Lehavy, and Trueman (2007).

[21] Conversations with sell-side analysts indicated that the benchmarks are in fact used internally by the brokers when assessing the performance of their analysts. A recent article in the press reinforces the view that analysts do want their recommendations to be interpreted relative to the adopted benchmarks. The article discusses Credit Suisse decision to switch to an industry benchmark, an event that was accompanied by some reshuffling of their outstanding recommendation. In explaining why Hess Corp. was downgraded, its analyst wrote that Hess "could still outperform the broader market. However, Hess spent more on energy exploration and development than expected this year, so that could prove a drag on its results relative to its peers." See "Credit Suisse: These Downgrades Aren't Personal," *The Wall Street Journal*, October 2nd, 2012.

report the results separately for the actual recommendations and for their control units. The results indicate that about 50% of buy recommendations issued by industry and market benchmarkers meet or beat their objective. By contrast, less than 40% of buy recommendations issued by total benchmarkers do so. These results seem plausible, as meeting the objective for total benchmarkers is arguably a harder task. Indeed, total benchmarkers need to base their advice on predictions related to firm-specific returns, industry returns, and market returns. Note also, that the most target used by total benchmarkers is 15% (see Table 7). This quite high threshold could also contribute to total benchmarkers' lower success rate in hitting their targets.

Next, we consider whether the success rate for 'buy' recommendations is related to the recommendation itself or is driven by either risk characteristics or more stringent objectives. [22] To do that, we compare the success rates between the actual recommendations and the control units obtained from the propensity score matching procedure. The comparison suggests that for all types of benchmarks, firms for which analysts issue 'buy' recommendations perform better than firms with similar risk characteristics that did not receive such recommendations.[23] For example, 49.8% of buy recommendations issued by industry benchmarkers hit their targets, compared with 43.7% of control units. Also, it is worth noting that while recommendations issued by total benchmarkers underperform those that were issued by market or industry benchmarkers, they perform much better than their control units (38.9% vs. 31.7%). The results for 'sell' recommendations are similar. More than 58% of 'sells' coming from market and industry benchmarkers meet their objective but only about 36% of 'sells' coming from total benchmarkers do so. In addition, compared to the control units, the actual recommendations perform significantly better for all types of benchmarks.

[22] To illustrate the first possibility, take two analysts, *A* and *B*, working for the same broker (let's say a market benchmarker). Analyst *A* covers large company stocks and *B* covers small company stocks. Assume each analyst randomly issues 'buys' for the stocks within their coverage set, without trying to add any insight. If size is indeed a risk factor—such that large (small) companies on average underperform (underperform) the market—then the 'buys' from *A* will underperform her stated objective while 'buys' from *B* will overperform it, but the performance difference is not due any special insight being offered by neither analyst (other than different loadings on risk factors). For the second possibility, now take analysts *A* and *B* working for different brokers, but issuing exactly the same 'buys' to the same firms. Assume that analyst *A*'s employer adopts a higher target return, and thus a higher stated objective, compared to *B*'s employer. In the measure of recommendation performance—the return of recommended firm minus the stated objective—the first term is the same for analysts *A* and *B*, while the second term is higher for analyst *A*. Thus, analyst *B* performs better than *A* but, again, not due to any specific insight (other than having chosen to work for a less demanding broker!).

[23] Recall that we measure the performance of a recommendation over the period during which it was active, or over one year, whichever is shorter. For the control unit, we measure performance over the same length period that we used for its corresponding actual recommendation. In addition, for the control unit, we adopt the same stated objective as the one used by the corresponding recommendation.

Moreover, the difference in success rates between actual recommendations and their control units for 'sell' recommendations is larger than for 'buy' recommendations.

Panel B of Table 8 considers the magnitudes by which analysts beat (or miss) their stated objectives. The table reports the average, as well as the median, difference between the realized return and the stated objective for each recommendation in our sample as well as for the control units. The results are consistent with those in Panel A. Indeed, industry and market benchmarkers significantly beat their stated objective for both 'buy' and 'sell' recommendations. For example, a 'buy' recommendation from an industry benchmarker yields a return that exceeds the stated objective (the industry return plus the target) by 312 basis points.[24] By contrast, total benchmarkers on average miss their stated objective. For example, a 'sell' recommendation issued by a total benchmarker misses the target by 1,148 basis points, on average. To evaluate the performance of recommendations relative to the performance of stocks with similar risk characteristics and facing the same stated objective, we consider the control units. We find that for all types of benchmarkers, the excess returns over the stated objectives for 'buy' ('sell') recommendations are better (worse) than those of the control units.

< Insert Table 8 here >

In Panel C we report the raw returns associated with the different stock recommendations broken by benchmark type. This analysis focuses on the performance of recommendation abstracted from the recommendation's stated objective. As before, the time period we use is the earliest of 12 months or until the recommendation has been changed. Notice that the raw returns following 'buy' recommendations issued by market, industry, and total benchmarkers are not very different from each other (10.4%, 10.9%, and 9.94%). More importantly, we see that the better performance of 'buys' and 'sells' compared to their control units is also observed in raw returns. Thus, the results do not seem to depend on differences in the stated objectives across different benchmarkers.

In an alternative analysis (not reported and available upon request) we perform a multivariate analysis of the relation between raw returns and the benchmark type. We regress raw returns following the recommendations on benchmark indicators and a set of control variables including past firm and market performance to account for momentum, analysts' experience, broker

[24] We report medians to ensure that our inferences are not affected by extreme observations that might have an undue influence on the means. This is an important concern when dealing with long term returns. We discuss the results on medians whenever they might yield a different inference than that of the means.

size, firm size, and book-to-market. We also control for the general tendency of a broker to issue each type of recommendation. If a broker is in general more stringent with respect to issuing 'buys' it is likely that its 'buys' are more meaningful.[25] We follow Barber, Lehavy, McNichols and Trueman (2006) and include dummies for the broker's favorableness quintiles. These quintiles are determined each quarter by ranking brokers in ascending order according to the percentage of each type of recommendation at the end of the previous quarter.[26] We include in the regression dummies for quintiles 1 (least favorable) through quintile 4 (that is, quintile 5, the most favorable, is the baseline to which the other dummies should be compared). Consistent with the results in Panel C of Table 8, we do not observe a difference between the three groups of benchmarkers. This reinforces our interpretation that the difference in abnormal performance of recommendations across different benchmarkers comes from the stated objective and not from the returns.

In sum, the analysis in Table 8 reveals that for all types of benchmarks 'buy' ('sell') recommendations outperform (underperform) stocks with similar risk profiles and subject to the same investment objective. Also, it is important to emphasize that the seemingly weak performance of recommendations issued by total benchmarkers, relative to those issued by industry and market benchmarkers, is a result of a more stringent stated objective.

In our next analysis, we are interested in identifying the source of value in stock recommendations. There are three possible contributors to the performance of stock recommendations. First, stock recommendations can reflect analysts' ability to identify winners and losers within a particular industry (Boni and Womack, 2006). We refer to this dimension as stock picking. Second, it is possible that stock recommendations also reflect analysts' opinions about the industry prospects of the firms they cover (Kadan et al, 2012). We refer to this dimension as

---

[25] Barber, Lehavy, McNichols and Trueman (2006) report that indeed the investment value of recommendations depends on the overall 'favorableness' (or proclivity to issue 'buys') of each broker. Given the results in Section 3 that the distribution of recommendations differs across different benchmarkers, we also need to control for this favorableness here.

[26] Barber et al (2006) considered favorableness based on fraction of 'buys' only, while we separately look at favorableness towards 'buys' for the regression examining 'buys' and favorableness towards 'sells' for the regression examining 'sells'. The difference is explained by the sample period of the two studies. For Barber et al (2006), most of the data comes from the period before September 2002, when sells were rare, so the vast majority of the recommendations were in practice spread between 'buys' and 'hold', and therefore the favorableness towards 'buys' would be a good summary of the overall distribution of recommendations for the broker. Our sample period starts in September 2002, when recommendations become more balanced between 'buys' and 'sells,' so a broker's favorableness towards 'buys' does not denote necessarily its lack of favorableness towards 'sells.'

industry picking. Third, stock recommendations could be influenced by the general sentiment of analysts towards the market as a whole (market timing).

The disclosure of recommendations' benchmarks allows us to better evaluate the three dimensions of analysts' abilities, because each dimension is manifested differently in each benchmark type. Industry benchmarkers, who state that their recommendations aim at beating an industry threshold, are expected to rely on stock picking ability alone. Market benchmarkers state that their recommendations will beat a market threshold. Thus, their recommendations are expected to incorporate both stock picking and industry picking abilities. Finally, total benchmarkers present an absolute threshold that is influenced by the performance of firms, industries and the market as a whole. Thus, we expect recommendations issued by total benchmarkers to reflect all three dimensions of analysts' abilities. Our objective is to examine whether and how the performance of recommendations demonstrate the presence of these three abilities. Our setting provides us with a cleaner and more powerful test of such abilities, compared with a setting that does not differentiate between the different types of analysts, as we can focus on the subsets of analysts that claim to exploit a particular ability. For example, we can study the presence of (the yet unexplored) market-timing ability of analysts by focusing on total-benchmarkers, who claim to provide market-timing advice. The results in Section 4 add credibility to this approach, since analysts indeed appear to abide by their benchmarks.

To evaluate the different abilities of analysts, we decompose the returns in excess of the recommendations' stated objective into components that measure stock picking, industry picking and market timing. For industry benchmarkers, excess returns only reflect analysts' stock picking and are measured as

$$R - (R_{\text{industry}} + \text{Target}), \tag{3}$$

where the "Target" is the one given in Table 7.

For market benchmarkers, we decompose the difference between the actual returns and the stated objective into two components,

$$R - (R_{\text{market}} + \text{Target}) = (R\text{-}R_{\text{industry}})+(R_{\text{industry}} - (R_{\text{market}} + \text{Target})), \tag{4}$$

where the first term on the RHS reflects stock picking and the second term reflects industry picking.

Finally, for total benchmarkers, we decompose the difference between actual returns and the target into three components reflecting stock picking, industry picking and market timing abilities.

$$R - \text{Target} = (R - R_{industry}) + (R_{industry} - R_{market}) + (R_{market} - \text{Target}). \qquad (5)$$

Similar to the analysis in Table 8, in Panel A of Table 9 we compare the returns and their components between the actual recommendations and their control units. We begin with 'buy' recommendations from industry benchmarkers, for whom we can only evaluate stock picking ability. We document a significant difference between the stock picking component associated with the actual recommendations and the one associated with the control units (3.1% vs. -0.77%), suggesting that stock picking ability exists.[27]

< Insert Table 9 here >

When examining market benchmarkers, we can evaluate both stock picking and industry picking. We confirm that stock picking is also present, as the returns exceed the industry index by 513 basis points for the actual recommendations, compared to 143 basis points for the control units.[28] On the other hand, accounting for risk, our results do not indicate any industry picking ability, as the industry picking components are not significantly different between the actual recommendations and their control units (39 basis points compared to 51 basis points).

Studying total benchmarkers allows us to examine all three possible abilities of analysts. Like before, we find stock picking ability, where the difference between actual returns and industry returns is 478 basis points for the actual recommendations compared to 155 basis points for the control units. We do not find any evidence of industry picking, as the difference between 189 basis points for the actual recommendations and 176 basis points for the control units is not significant. Finally, we do not find evidence of market timing among total benchmarkers. Market returns following 'buy' recommendations are not higher than those of the control units. In fact, the average difference between market returns and the targets following actual 'buy' recommendations is lower

[27] Note that because the stock picking component and the returns in excess of the stated objective are equal for industry benchmarkers, the numbers in the top of Table 9 are identical to those for industry benchmarkers in Table 8.

[28] Notice that the stock picking components in equations (4) and (5) are slightly different due to the way the target return is assigned. To compare the stock picking component of industry benchmarkers to that of market benchmarkers, one needs to add the weighted average of the targets among 'buys' of industry benchmarkers (153 basis points) to their average stock picking component of 312 basis points. The difference in computation, however, does not affect the inferences from comparing the recommendations and their control units because the returns for each control unit are measured in the same way as for the corresponding actual recommendation.

than that following the control units (-1,148 basis points versus -1,054 basis points), though not significantly so when comparing the medians. Results for ‘sell’ recommendations are very similar. Like in the case of ‘buy’ recommendations, we find evidence of stock picking, but not of industry picking or market timing.

Overall, Panel A of Table 9 provides evidence that analysts possess stock picking ability across all three types of benchmarks. The results are consistent with the analysts’ disclosures of their investment objectives, as each of the three benchmarks suggests analysts’ reliance on the stock picking ability. The evidence is also consistent with prior studies (starting with Boni and Womack, 2006) that argue that analysts are good in ranking firms within an industry.

We do not find evidence of industry picking among both market and total benchmarkers. This contrasts with their disclosures implying reliance on industry picking. It is worth emphasizing that our test for industry picking is a joint test of analysts abiding by their stated benchmark, as well as being successful at industry picking. It is possible, for example, that market benchmarkers are, de facto, acting like industry benchmarkers, not attempting to provide any industry picking. If that is the case, we obviously would not expect to find any evidence of industry picking. However, our results in Table 6 provide evidence that market and total benchmarkers’ recommendations rely more on across-industry information, suggesting that these analysts are attempting to abide by their stated objective. Thus, we conclude that our results are more likely consistent with analysts not demonstrating industry picking ability, as opposed to analysts not attempting to provide industry picking.

It is important to contrast this conclusion with the results of Kadan et al. (2012). In that paper, we provide evidence that strategy analysts possess industry picking ability demonstrated in their industry recommendations. In this paper we focus our attention on firm-level analysts and recommendations to individual firms rather than industries. The different results emphasize the difference in skills and scope between strategy analysts and firm-level analysts. Also, in one analysis Kadan et al. (2012) rely on firm recommendations and present some mild evidence of industry picking among market and total benchmarkers. In this paper, we rely on a different methodology of analyzing industry picking in firm recommendations.[29] The other important

[29] For example, in this paper we analyze the performance of a recommendation over its entire life span, while in Kadan et al. (2012) we only evaluate performance over a short-term window of one month. In addition, in this paper we control for risk through the use of a matched sample, while in Kadan at el. (2012) we use a four-factor alpha.

difference is that in Kadan et al. (2012) we examine a sample of the twenty largest brokers, while in this paper we examine a more comprehensive sample of brokers. In untabulated results we re-examine the analysis of Kadan et al. (2012) on the more expanded sample of brokers. The results show no evidence of industry picking among market and total benchmarkers in the larger sample. Thus, we believe that to the extent that industry picking among firm-level analyst exists, it is not robust, whereas such skill does seem to exist among strategy analysts.

As for market timing, we do not find evidence of superior performance by total benchmarkers, even as they profess and try to rely on it. As before, we emphasize the joint nature of this test. Given the earlier evidence that total benchmarkers do try to incorporate market timing in their recommendations, lack of evidence of superior market timing performance is more likely consistent with total benchmarkers not demonstrating market timing ability, rather than not attempting to do so. Such lack of results might be consistent with the task's difficulty. The absence of market timing ability among sell-side analysts mirrors the inability of other market professionals to successfully time the market. These include investment newsletters (see Graham and Harvey, 1994, 1996, 1997), hedge fund managers (Fung, Xu and Yao, 2002), and pension fund managers (Coggin, Fabozzi and Rahman, 1993).[30]

Panel B of Table 9 reinforces these results by exploring whether all three types of brokers possess all three types of skills, regardless of their benchmark. For example, in this panel we test whether industry benchmarkers possess industry picking and market-timing skills, even though they do not commit to such abilities. To facilitate this analysis we no longer use the broker's specified target as in Panel A. Instead, we set each broker's target (be it an industry, market or absolute target) to 0. The results show that no such skills exist. In particular, industry benchmarkers do not show any industry picking or market timing abilities compared to the control units, and market benchmarkers do not show any market timing abilities.

5.4 Robustness Analyses

Same-Industry Bias. There is a concern on whether our matching procedure stacks the analysis against finding evidence of industry picking. If firms in the same industry are similar according to

[30] The ability to time the market has also been extensively tested in the context of mutual funds, with mixed results. While most of the literature has failed to identify such ability in mutual funds (e.g., Treynor and Mazuy, 1966; Henriksson, 1984; Grinblatt and Titman, 1994; Ferson and Schadt, 1996; Becker, Ferson, Myers and Schill, 1999), more recent developments on how market timing is tested do ascribe some positive timing to mutual fund managers (Bollen and Busse, 2001; Jiang, Yao, and Yu, 2007).

the risk dimensions adopted in the matching procedure, forcing the control unit to look like the actual recommended stock along these risk dimensions can bias the control unit to belong to the industry of the recommended stock. To address this concern, we search for evidence of a "same industry" bias in the matching procedure. We compare the actual fraction of control units belonging to the same industry of the recommended stock to what such fraction would be in a random match—for which no bias can exist. We bootstrap the empirical distribution of the fraction of same-industry matches from samples of randomly chosen control units. We find no evidence of a "same industry" bias. For example, our propensity score procedure has 3.29% of the sells matched to a control unit in the same industry, which is within the 95% confidence interval for the bootstrapped average fraction (Similar results apply to 'buys'.)

Same-Time Bias. Uncovering evidence of market timing would also be impaired if the matching procedure biased the control unit to be picked at the same calendar period as the recommendation. We test for and find no evidence of a timing bias from the matching procedure. For example, the actual fraction of sells (or buys) matched to control units in the same month of the recommendation is not significantly different from what would result in a random allocation of the time period of each control unit.

Stock-picking vs. Market-timing. Another concern is that our measure of stock picking ability may actually be capturing market timing. If an analyst bets on beta—say, issuing 'buy' for a high-beta stock within her industry coverage prior to an upward market move, the analyst is relying on market timing skill. In this situation, the excess return that we use to diagnose stock picking—equation (3)—would in fact derive from market timing. To examine this possibility, we repeat the performance analysis of the recommendations after forcing each control unit to belong to the same industry *and* to be picked at exactly the same time as the actual recommendation. Given that the match is done based on beta, it follows that each control unit carries roughly the same beta as the actual recommendation.[31] If the analyst successfully used market timing in recommending a stock based on its beta, the control unit would also enjoy market timing, and thus would perform similarly to the actual recommendation. In particular, we would find evidence of stock picking for the control units as well. We do not. Regenerating Table 9 under this alternative sampling yields results

[31] Betas can still differ given that the matching procedure is based on four dimensions. Alternately, we can run the matching procedure based on beta alone—guaranteeing that betas from the control unit and from the recommended stocks are indistinguishable. Inferences are robust to this alternative.

(available upon request) that are qualitatively the same as the results discussed in the paper: The recommendations, but not the control units, reveal stock picking. In summary, market timing on stock betas does not seem to be responsible for stock picking.

Industry Beta. We also attempt to expand the risk dimensions used to define the control units. In particular, we include industry as a risk factor. This can be important in order to properly measure risk-adjusted performance—if industry beta is relevant. There are problems with this approach, though. First, the industry beta does not show up at all significantly in the propensity score regressions. Second, when including industry beta, we create a bias in that a firm in the same industry is more likely to be chosen as the control unit. Nevertheless, defining control units based on the expanded set of risk dimensions does not change any of the qualitative inferences discussed above. [32]

6 Conclusion

In this paper we examine the literal meaning of sell-side analysts' stock recommendations. We document that different brokers rely on different benchmarks with respect to which the investment advice in each recommendation should be interpreted. For example, a 'buy' from a market benchmarker is a prediction that the recommended stock is expected to outperform the market; a 'buy' from an industry benchmarker denotes the analyst's expectation that the stock will outperform its peers in the same industry; finally, a buy from a total benchmarker suggests the stock will beat some absolute return threshold.

We show that these benchmarks are not an irrelevant detail in the analyst's disclosure about how recommendations should be viewed. Instead, such benchmarks are in fact used when analysts form their recommendation advice. For example, industry benchmarkers, who profess to basically rank firms within each industry, do rely less on across-industry expectations about fundamentals—such as earnings and LTG projections—when compared to market and total benchmarkers. Also,

[32] Adding an industry beta to the matching procedure also allows us to examine whether analysts try market timing on industry beta—say, issuing 'buys' for stocks of high-beta industries prior to an upward market move. This would show up as excess industry performance in the second component in equation (4). We thus would diagnose as industry picking an ability that in fact originates in market timing. This possibility is less relevant here, given that we cannot find industry picking anyway. Nevertheless, we test for the possibility as well. We force the control unit to be part of a different industry and to be picked at the same time as the actual recommendation (and to have similar industry beta, given that this risk factor is now included in the matching procedure). No evidence of market timing on industry beta surfaces.

.

consistent with the assertion in the analysts' disclosures that total, but not market benchmarkers, rely on market timing, we observe that total benchmarkers do become more pessimistic relative to market benchmarkers during the recession in our sample period. This suggests that the *use* of each recommendation—by investors or by academics—should take into consideration the benchmark under which it is formed.

We exploit the different benchmarks to better understand the sources of value that are reflected in stock recommendations. Each benchmark implies the use of a different set of skills, which could include stock picking, industry picking and market timing. We show that stock recommendations from all benchmarkers perform better than stocks with similar risk profiles that were not issued the same type of recommendation. The improved performance of stock recommendations comes solely from stock picking. We find no evidence of industry picking or market timing, even for the benchmarkers that imply the use of these abilities.

Our study suggests that both academics and investors should pay more attention to the declared objective of each recommendation. In particular, the fact that different recommendations carry different meanings can be used to shed new light on a range of empirical questions. Ramnath, Rock, and Shane (2008), for example, advocate the need for a better understanding of how analysts operate. The different benchmarks employed by brokers suggest that information shocks would affect recommendations differently depending on the broker's benchmark—e.g., with industry shocks affecting more the recommendations from market and total benchmarkers when compared to recommendations from industry benchmarkers. Another potential area worth of a second look is the long literature on how incentives affect bias and performance of recommendations (e.g., Lin and McNichols, 1998; Michaely and Womack, 1999). This comes naturally once one recognizes that performance is a comparison between the return path of the recommended stock and its stated objective, and thus should take into consideration the benchmark adopted by the broker. In fact, determining superiority among analysts in terms of their stock picking abilities (e.g., Mikhail, Walther, and Willis, 2004) might need adjustment as well, given that different analysts arguably pick stocks according to different objectives. These are left as avenues for future research.

References

Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman, 2001, Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns, *Journal of Finance* 56, 531-563.

Barber, Brad, Reuven Lehavy, Maureen McNichols, and Brett Trueman, 2006, Buys, Holds, and Sells: The Distribution of Investment Banks' Stock Ratings and the Implications for the Profitability of Analysts' Recommendations, *Journal of Accounting & Economics* 41, 87-117.

Barber, Brad, Reuven Lehavy, and Brett Trueman, 2007, Comparing the Stock Recommendation Performance of Investment Banks and Independent Research Firms, *Journal of Financial Economics* 85, 490-517.

Barniv, Ran, Ole-Kristian Hope, Mark Myring, and Wayne Thomas, 2009, Do Analysts Practice What They Preach and Should Investors Listen? Effects of Recent Regulations, *The Accounting Review* 84, 1015-1039.

Becker, C., W. Ferson, D. Myers and M. Schill, 1999. Conditional Market Timing With Benchmark Investors. *Journal of Financial Economics* 52: 119-148.

Bharath, Sreedhar., Sandeep Dahiya, Anthony Saunders, and Anand Srinavasan, forthcoming. Lending Relationships and Loan Contract Terms. *Review of Financial Studies*.

Bhojraj Sanjeev, Charles M. C. Lee, and Derek K. Oler, 2003. What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research, *Journal of Accounting Research* 41, 745-774.

Bollen, Nicolas and Jeffrey Busse, 2001. On the Timing Ability of Mutual Fund Managers. *Journal of Finance* 56: 1075-1094.

Boni Leslie, and Kent L. Womack, 2006, Analysts, Industries, and Price Momentum, *Journal of Financial and Quantitative Analysis* 41, 85-109.

Bradshaw, Mark, 2004, How Do Analysts Use Their Earnings Forecasts in Generating Stock Recommendations?, *The Accounting Review* 79, 25-50.

Bradshaw, Mark, 2009, Analyst Information Processing, Financial Regulation, and Academic Research, *The Accounting Review* 84, 1073-1083.

Bradshaw, Mark, 2012, Discussion of "Analysts' Industry Expertise, *Journal of Accounting and Economics* 54, 95-120.

Brown, Lawrence, and Kelly Huang, 2010, Forecast-Recommendation Consistency and Earnings Forecast Quality, Working Paper, Georgia State University.

Chance, Don. And Michael Hemler, 2001. The Performance of Professional Market Timers: Daily Evidence from Executed Strategies. *Journal of Financial Economics* 62: 377-411.

Chen, Chih-Ying and Peter Chen, 2009, NASD Rule 2711 and Changes in Analysts' Independence in Making Stock Recommendations, *The Accounting Review* 84, 1041-1071.

Coggin, Daniel, Frank Fabozzi and Shafiqur Rahman, 1993. The Investment Performance of US Pension Funds: An Empirical Investigation. *Journal of Finance* 48: 1039-1056.

Colak, Gonu and Toni Whited, 2007. Spin-offs, Divestitures, and Conglomerate Investments. *Review of Financial Studies* 20: 557-595.

Drucker, Steven, and Manju Puri, 2005 On the Benefits of Concurrent Lending and Underwriting. *Journal of Finance* 60:2763–99.

Ertimur, Yonca, Jayanthi Sunder, and Shyam Sunder, 2007, Measure for Measure: The Relation Between Forecast Accuracy and Recommendation Profitability of Analysts, *Journal of Accounting Research* 45, 567-605.

Fama, Eugene and Kenneth French, 2002, Testing Trade-Off and Pecking Order Predictions About Dividends and Debt, *Review of Financial Studies* 15, 1-33.

Ferson, W. and R. Schadt, 1996. Measuring Fund Strategy and Performance in Changing Economic Conditions. *Journal of Finance* 51: 425-462.

Fung, Hung, Xiaoqing Xu, and Jot Yau, 2002. Global Hedge Funds: Risk, Return and Market Timing. *Financial Analysts Journal* 58: 19-30.

Graham, John and Campbell Harvey, 1994. Market Timing Ability and Volatility Implied in Investment Newsletters' Asset Allocation Recommendations. Working Paper, NBER.

Graham, John and Campbell Harvey, 1996. Market Timing Ability and Volatility Implied in Investment Newsletters' Asset Allocation Recommendations. *Journal of Financial Economics* 42: 397-422.

Graham, John and Campbell Harvey, 1996. Grading the Performance of Market-Timing Newsletters. *Financial Analysts Journal* 53 (November/December): 54-66.

Grinblatt, Mark and Sheridan Titman, 1994. A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques.

Hellman, Thomas, Laura Lindsey, and Manju Puri, 2009. Build Relationships Early: Banks in Venture Capital. *Review of Financial Studies* 21: 513-541.

Henriksson, Roy, 1984. Market Timing and Mutual Fund Performance: An Empirical Investigation. *Journal of Business* 57: 73-96.

Jegadeesh, Narasimhan., Joonghyuk Kim, Susan D. Krische, and Charles M. Lee, 2004, Analyzing the Analysts: When Do Recommendations Add Value? *Journal of Finance* 59, 1083-1124.

Jiang, George, Tom Yao and Tong Yu, 2007. Do Mutual Funds Time the Market? Evidence from Portfolio Holdings. *Journal of Financial Economics* 86: 724-758.

Kadan Ohad, Leonardo Madureira, Rong Wang, and Tzachi Zach, 2009, Conflicts of Interest and Stock Recommendations: The Effect of the Global Settlement and Related Regulations, *Review of Financial Studies* 22, 4189-4217.

Kadan Ohad, Leonardo Madureira, Rong Wang, and Tzachi Zach, 2012, Analysts' Industry Expertise, *Journal of Accounting and Economics* 54, 95-120.

Kesckes, Ambrus, Roni Michaely, and Kent Womack, 2010, What Drives the Value of Analysts' Recommendations: Earnings Estimates or Discount Rate Estimates? , Working Paper, Tuck School of Business.

Lin, Hsiou-wei and Maureen F. McNichols, 1998, Underwriting Relationships, Analysts' Earnings Forecasts and Investment Recommendations, *Journal of Accounting and Economics* 25, 101-127.

Loughran, T., and P. Schultz, 2005, Liquidity: Urban Versus Rural Firms, *Journal of Financial Economics* 78, 341-374.

McNichols, Maureen and Patricia C. O'Brien, 1997, Self-Selection and Analyst Coverage, *Journal of Accounting Research* 35, 167-199.

Michaely, Roni and Kent Womack, 1999, Conflict of Interest and the Credibility of Underwriter Analyst Recommendations, *Review of Financial Studies* 12, 653-686.

Mikhail, Michael, Beverly Walther, and Richard Willis, 2004, Do Security Analysts Exhibit Persistent Differences in Stock Picking Ability? *Journal of Financial Economics* 74, 67-91..

Ramnath Sundaresh, Steve Rock and Philip B. Shane, 2006, Financial Analysts' Forecasts and Stock Recommendations: A Review of the Research, *Foundations and Trends in Finance* 2, 311-421.

Rosenbaum, Paul, and Donald Rubin, 1983. The Central Role of the Propensity Score in Obsevational Studies for Causal Effects. *Biometrika* 70: 373-401.

Stickel, Scott E., 1995, The Anatomy and Performance of Buy and Sell Recommendations, *Financial Analysts Journal* 51, 25-39.

Treynor, Jack and Kay Mazuy, 1966. Can Mutual Funds Outguess the Market? *Harvard Business Review* 44: 131-136.

Womack Kent, 1996, Do Brokerage Analysts' Recommendations have Investment Value?, *Journal of Finance* 54, 137-157.

Appendix A. Propensity Score Matching Procedure

The procedure for matching 'buys' is as follows. We estimate a probit model of the likelihood that a firm receives a buy recommendation in a particular month. We use a pooled sample of the monthly cross-sections of firms trading on NYSE, Nasdaq and AMEX between September 2002 and December 2009. We rely on the predicted probabilities (the propensity scores) from the probit model to find a nearest-neighbor match with replacement for each 'buy'. A 'buy' is defined by the firm receiving the recommendation and the recommendation announcement day. In the probit model, this 'buy' maps to the data point (i,t), corresponding to firm i that received the recommendation and to month t during which the recommendation is issued. We define the matched unit for (i,t) as the data point (i_c,t_c) —corresponding to firm i_c and month t_c—with the propensity score that is closest to the propensity score from (i,t). We also require that firm i_c has not received a 'buy' in month t_c. Given that the probit model is estimated at the monthly level, the matching procedure does not define the day of the month for the matched observation. We assume it to be the same day of the month as the original 'buy' (or the last day of t_c if t_c does not have the day of the month of the original 'buy').[33,34]

For explanatory variables in the probit model, we use the risk characteristics associated with the Fama-French 4-factor model: beta, size, book-to-market and momentum. There are two concerns with using the raw measures of the risk characteristics in the pooled probit model. First, it is possible that a time trend in the raw measures would result in weaker matches. For example, if we use raw measure of size and the average market capitalization increases over time, a 'buy' for a relatively large firm in the early part of the sample could be matched to a relatively small firm in the late part of the sample. Second, measures like size and book-to-market can be prone to skewness and the presence of outliers. We address both concerns by adopting a normalized version of each risk measure.[35] We rank all firms in each month according to that measure, and then define a score

[33] For example, assume a 'buy' for firm i is announced on March 31st, 2002. If this 'buy' is matched to firm i_c in October 2006, we define the matched recommendation day as October 31st, 2006; if it is matched to June 2003, which does not have 31 days, we assume the matched recommendation day is June 30th, 2006.

[34] Notice that all 'buys' for the same firm and the same month are mapped to one single data point in the probit model, and thus have the same propensity score. In a matching procedure with replacement, they are all matched to the same pair (i_c,t_c), though the resulting recommendation day for the control unit differs if the 'buys' are announced on different days of the month. We can force dispersion by requiring that each of these buys is matched to a different control unit (that is, without replacement). Results are not sensitive to this choice.

[35] The raw measures, on which we base the scores, are computed as follows. For a firm i and month t, we define the firm beta as the coefficient from a regression of the firm daily return on the market return over the preceding year; firm

variable that goes from 0 (for the firms with the smallest measure in that month) to 1 (for the firm with largest measure in that month).[36] Our probit model for 'buys' thus becomes (yearly dummies are also included):

$$\text{Prob(BUY)}=\alpha+\beta_1\text{*score(beta)}+\beta_2\text{*score(size)}+\beta_3\text{*score(beme)}+\beta_0\text{*score(mom)}+\varepsilon$$

We estimate this model for 'buys' in a sample of 372,163 firm-month observations, an average of 4,229 firms per month. The results, reported under the "Pre-Match" column in Panel A of Table A1, confirm that risk measures are important determinants of 'buy' recommendations. Analysts are more likely to issue these recommendations for firms with higher betas, higher market values, lower values of book-to-market (growth firms) and better performance in the recent past. The pseudo-R^2 of the model is 12.4%.

< Insert Table A1 here >

Our matching process, discussed above, defines one control unit for each 'buy',[37] with the goal that the sample of 'buys' and the control sample would be very similar with respect to the risk measures. We, next, evaluate the matching process in terms of reaching that goal. First, Panel B of Table A1 shows that each 'buy' and its respective control unit are indeed very close in terms of their propensity scores—with the maximum difference between them across all pairs being a mere 0.1%. Panel C compares 'buys' with either randomly matched units or with their nearest-neighbor matches, across the four dimensions of risk used in the matching process. The pre-match analysis reinforces the inferences from the regression: firms receiving a 'buy' are much bigger, have lower book-to-market, higher betas and better performance, when compared to the average firm, and all differences are statistically significant. After the match, when compared to the nearest-neighbor, 'buys' and control units do not differ significantly with respect to size, book-to market and momentum; there are still differences with respect to beta, though these differences are economically very small.

---

size is the market value of its equity 7 months prior to month t; book-to-market is the ratio of the book value of equity to the market value of equity, for the fiscal year preceding *t*; and momentum is defined as the average monthly return over the 6 month-period preceding *t*. We restrict the analysis to firms with share codes 10 or 11 and remove penny stocks (average trading price during the month below $5). We also require at least 60 days of past returns for an estimated beta to be used in the regressions.

[36] Take firm size, for example. The normalization works as follows. Each month we sort all firms according to firm size and define a variable $rank_{i,t}$ that is equal to 1 for the smallest firm, equal to 2 for the next firm, and equal to n for the biggest firm, where n is the number of firms in that month. The score measure is defined as $score_{i,t}=100-100*(rank_{i,t}-1)/(n-1)$.

[37] Results are qualitatively the same if we define more than one control unit—let's say, 3 or 5—per recommendation.

We can also analyze the accuracy of the matching process by restricting the probit regression to the original sample of 'buys' plus their control units. Results are shown in the column labeled "Post-Match" in Panel A of Table A1. The magnitude of the coefficients on the risk measures decline substantially, and all coefficients become insignificant. Moreover, the pseudo-R^2 drops from 12.4% to 3%. In summary, the results suggest that the matching process ensures 'buys' and their control units are similar with respect to beta, size, book-to-market and momentum.

We then repeat the propensity score method to construct a control sample for 'sells'. We start with a probit modeling the likelihood that a firm receives a 'sell' recommendation. As with 'buys', 'sells' are more likely to be issued for firms with higher betas and for bigger firms. Contrary to 'buys', though, 'sells' are more commonly issued for value and low-performing firms. The different loadings on the measures of risk for the probits modeling 'sells' vs. 'buys' reinforce the need of different matching procedures for each type of recommendation. The matching procedure also does a good job with 'sells'. The "Post-Match" probit leaves only one coefficient (on momentum) significant at the 5% level. 'sells' are also very similar to their control units with respect to the risk measures—with the exception of the score of beta, for which 'sells' and control units differ at the 5% level.

Table A1. Propensity Score Diagnostics

This table presents diagnostics on the propensity score methodology used to create matched samples to the samples of 'buys' and sells. Panel A contains parameter estimates of the probit models generating the propensity scores used to match 'buys'/sells to control units. The sample includes monthly cross-sections of firms trading in NYSE, Nasdaq and Amex, from September 2002 to December 2009. Only firms with shares codes equal to 10 or 11 are included, and stocks with monthly average price below \$5 are excluded. When modeling 'buys' (sells), the dependent variable of the probit regression is a dummy equal to 1 if the firm was issued a recommendation with a buy (sell) signal in that month. The independent variables are normalized measures of beta, size, book-to-market and momentum. The normalized measure of *X*, *score(X)*, is defined as follows. Each month we sort all firms according to *X* and define a variable $rank_{i,t}$ that is equal to 1 for the firm with smallest *X*, equal to 2 for the next firm, and equal to *n* for the firm with biggest value of *X*, where *n* is the number of firms in that month; we then define $score(X)_{i,t}=100-100*(rank_{i,t} - 1)/(n - 1)$. For a firm *i* and month *t*, we define the firm beta as the coefficient from a regression of the firm daily return on the market return over the preceding year; firm size is the market value of its equity 7 months prior to month *t*; book-to-market is the ratio of the book value of equity to the market value of equity, for the fiscal year preceding *t*; and momentum is defined as the average monthly return over the 6 month-period preceding *t*. The *Pre-Match* column contains the parameter estimates for entire sample, prior to matching. The *Pre-Match* probits are used to generate the propensity scores for matching 'buys'/sells. The *Post-Match* column contains the parameter estimates of the probit estimated on the subsample of original recommendations (buys/sells) and the corresponding control observations, after matching. The matching procedure is the nearest-neighbor match of treatment and control firms with replacement. Panel B presents pairwise comparisons, across the dimensions used to match the original recommendations to the matched sample, of the recommendation (buys/sells) and control samples. Panel C shows the distribution of propensity scores for the treatments, controls, and the difference in estimated propensity scores. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Probit Regression Results

	Modeling Buys		modeling Sells	
	Pre-Match	Post-Match	Pre-Match	Post-Match
constant	-2.3664***	0.00467	-3.2134***	0.0768
	(0.0198)	(0.0323)	(0.0334)	(0.0711)
score_beta	0.4291***	-0.027	0.5264***	-0.0789*
	(0.0127)	(0.0204)	(0.0216)	(0.0446)
score_size	1.2794***	0.0351*	1.4449***	0.00224
	(0.0132)	(0.022)	(0.0235)	(0.0516)
score_beme	-0.3998***	0.00357	0.1616***	0.0269
	(0.0121)	(0.0191)	(0.0197)	(0.0382)
score_mom	0.2352***	-0.0378**	-0.156***	-0.0664*
	(0.0112)	(0.0172)	(0.0185)	(0.0357)
Year fixed effects	Yes	Yes	Yes	Yes
# observations	372,163	69,508	372,163	16,002
Pseudo R^2	12.38%	2.98%	10.74%	2.27%

Table A1. (*Continued*)

Panel B: Estimated Propensity Score Distributions

Matching Buys

	# obs	Mean	SD	Min	P5	Median	P95	Max
difference		0.000	0.000	-0.001	0.000	0.000	0.000	0.001
treatment	34,754	0.134	0.064	0.004	0.032	0.132	0.243	0.326
control	34,754	0.134	0.064	0.004	0.032	0.132	0.243	0.326

Matching Sells

	# obs	Mean	SD	Min	P5	Median	P95	Max
difference		0.000	0.000	0.000	0.000	0.000	0.000	0.000
treatment	8,001	0.040	0.023	0.001	0.007	0.038	0.083	0.127
control	8,001	0.040	0.023	0.001	0.007	0.038	0.083	0.128

Panel C: Pairwise Comparisons

	Buys						Sells					
	Pre-Match			Post-Match			Pre-Match			Post-Match		
	Buy	Control	t-test	Buy	Control	t-test	Sell	Control	t-test	Sell	Control	t-test
score_beta	0.639	0.497	81.25	0.644	0.649	-2.91	0.668	0.506	47.28	0.671	0.679	-2.29
score_size	0.697	0.478	130.31	0.707	0.703	1.98	0.745	0.491	75.93	0.749	0.750	-0.17
score_beme	0.386	0.506	-70.02	0.381	0.381	0.31	0.443	0.497	-16.05	0.442	0.439	0.81
score_mom	0.516	0.492	13.62	0.517	0.521	-1.75	0.452	0.495	-12.58	0.445	0.451	-1.49

Table 1. Description of Benchmarks

This table summarizes the different types of benchmarks brokers use in our sample. For each type of benchmark, the description of the benchmark and one example of the textual description of recommendations are provided.

Benchmark	Description	Examples of textual description of recommendations
Industry	Recommendation is benchmarked against performance of peers in the same industry	"Our ratings reflect expected stock price performance relative to each analyst's coverage universe."
Market	Recommendation is benchmarked against market performance	Performance "relative to the market index over the next 12 months."
Total Return	Recommendation is based on a stock's total return.	"The rating system is based on a stock's forward -12-month expected total return (price appreciation plus dividend yield)."
Market/Industry	Recommendation is benchmarked against market and/or industry performance.	Buy: Expected to outperform the broader market and/or its sector over the next six to twelve months.
Total/Market	Recommendations is based on a stock's total return and/or benchmarked against market performance.	Buy means the stock is expected to appreciate and produce a total return of at least 10% and outperform the S&P 500 over the next 12-18 months;
Industry/Total	Recommendations is based on a stock's total return and/or benchmarked against industry performance.	STRONG BUY–The company has strong fundamentals and/or positive near-term catalysts. The stock's total return is expected to exceed the peer group's return in the industry and/or appreciate 15% or more over the next 12 months;
Market/Industry/Total	Recommendations is based on a stock's total return and/or benchmarked against market and/or industry performance.	Buy - anticipates appreciation of 10% or more within the next 12 months, and/or a total return of 10% including dividend payments,and/or the ability of the shares to perform better than the leading stock market averages or stocks within its particular industry sector.
Market/risk	Recommendation is based on a stock's risk-adjusted return relative to the market performance.	"Underperform (U) Expected to underperform on a total return, risk-adjusted basis the broader U.S. equity market over the next 12 months."
Industry/risk	Recommendation is based on a stock's risk-adjusted return relative to industry performance.	"Stock's total return vs. analyst's coverage on a risk-adjusted basis, for the next 12-18 months."
Total/risk	Recommendation is based on a stock's risk-adjusted return.	"Based on the stock's total return for the next 12-18 months on a risk-adjusted basis"
Not sure	Cannot identify which benchmark a broker uses.	"Buy/Add – Buy if you do not own or Add to existing positions. We believe that the shares offer an attractive reward versus risk profile over the next 12-18 months given current information and defined objectives. Shares seem undervalued based on current valuation measures and expectations."
Changes	A broker changes the benchmark during our sample period and we cannot identify when the broker made the change.	Janney Montgonery Scott LLC used a total return benchmark in 2004, and used an industry benchmark by the end of 2009.
No data	Cannot find data on the definition of ratings.	

Table 2. Summary Statistics

This table presents the summary statistics on the different types of benchmarks. Only brokerage houses which issued at least 100 recommendations to U.S. firms during our sample period (9/2002 – 12/2009) are included in the analysis. For each type of benchmark, we report the number of brokers using this type of benchmark, the distribution of the number of recommendations issued by each broker, the total number of recommendations issued by all brokers and the percentage to the total number of recommendations, and the number of brokers which is amongst the biggest 20 brokers in IBES according to the total number of recommendations issued.

		# of recommendations per broker						
Benchmark	No. of Brokers	Mean # rec	25 percentile	median	75 percentile	Total # rec	% of all	No. of brokers amongst biggest 20
Industry	37	2021	332	737	2668	74788	31.92%	9
Market	34	1230	306	627	1506	41822	17.85%	3
Total	42	1274	267	742	1467	53518	22.84%	4
No Data	41	408	164	211	391	16745	7.15%	0
Industry/Risk	4	2453	694	1081	4212	9810	4.19%	1
Total/Risk	8	1094	346	1159	1466	8753	3.74%	0
Market/Risk	2	3307	3103	3307	3511	6614	2.82%	1
Total/Market	4	1622	249	983	2995	6487	2.77%	1
Changes	2	2376	1347	2376	3405	4752	2.03%	0
Industry/Total	3	2056	359	2080	3730	6169	2.63%	1
Market/Industry	4	495	392	463	599	1981	0.85%	0
Not Sure	2	772	685	772	859	1544	0.66%	0
Market/Industry/Total	1	1291	1291	1291	1291	1291	0.55%	0
All	184					234274		

Table 3. Determinants of Benchmarks

This table reports the results of estimating logistic models of the probability of adopting a certain benchmark. The models are estimated for all brokers which use either industry or market or total benchmark and with at least 100 recommendations issued during our sample period (9/2002 – 12/2009). The dependent variables are as follows: **Broker Age** is the number of years a broker has appeared in IBES, **Broker Size** is defined as the ratio of the number of recommendations issued by a broker to the total number of recommendations by all brokers in the last year, **Number of Industries** is the number of industries covered by a broker in last year, **Average Experience** is the average analyst experience across all analysts employed by the brokerage house at the beginning of the year, where analyst experience is measured as the number of days the analyst has appeared in IBES. **Firm Size** is the average market value of equity of all firms covered by a broker by the end of last year, **BE/ME** is the average ratio of book equity to market equity of all firms covered by a broker in last year. Robust standard errors (in parentheses) are calculated after clustering at the broker level. ***, **, * denote statistical significance at the 1%, 5%, 10% levels, respectively.

	(1)	(2)	(3)	(4)
	Industry vs. Market or Total	Industry vs. Market	Industry vs. Total	Market vs. Total
VARIABLES	Prob(Benchmark=Industry)	Prob(Benchmark=Industry)	Prob(Benchmark=Industry)	Prob(Benchmark=Market)
Log(1+Broker Age)	0.961**	0.649	1.178**	0.378
	(0.479)	(0.570)	(0.546)	(0.475)
Broker Size	101.5***	133.4**	86.47**	-51.73
	(34.19)	(52.10)	(39.64)	(46.30)
Log(Number of Industries)	-1.025***	-1.168***	-1.012***	0.189
	(0.283)	(0.403)	(0.332)	(0.331)
Log(1+Average Experience)	-0.215	-0.249	-0.274	0.171
	(0.366)	(0.488)	(0.465)	(0.420)
Log(Firm Size)	0.0308	-0.208	0.197	0.374**
	(0.184)	(0.197)	(0.214)	(0.187)
Log(1+BE/ME)	-0.199	-0.255	-0.110	0.388
	(0.288)	(0.306)	(0.356)	(0.312)
Constant	0.198	5.770	-1.649	-8.007*
	(3.877)	(4.376)	(4.744)	(4.275)
Observations	702	438	490	476

Table 4. Distribution of Recommendations

This table presents the summary statistics on the distribution of recommendations according to the types of benchmarks. Only brokerage houses which issued at least 100 recommendations to U.S. firms during our sample period (9/2002 – 12/2009) are included in the analysis. Summary statistics are obtained for each year of the sample. Each observation in a yearly sample is a pair of firm and broker such that the broker has an outstanding recommendation for the firm at the end of the year, where an outstanding recommendation is the most recent recommendation issued by the broker to the firm during the year and that has not been cancelled by the broker. The table presents for each year of the sample and each type of broker, the distribution of the outstanding recommendations at the end of the year, the average recommendation level, and the standard deviation of the recommendation level. In the computation of the recommendation levels, 'strong buys' and 'buys' are considered optimistic recommendations and are mapped to level 1; 'holds' are mapped to level 2; and 'sells' and 'strong sells' are considered pessimistic recommendations and are mapped to level 3.

		Dec-02	Dec-03	Dec-04	Dec-05	Dec-06	Dec-07	Dec-08	Dec-09
Market benchmarkers	%buy	50%	46%	45%	47%	47%	52%	46%	46%
	%hold	45%	48%	49%	47%	47%	45%	49%	49%
	%sell	5%	7%	6%	5%	6%	4%	5%	5%
	Avg rec	1.53	1.62	1.60	1.60	1.57	1.52	1.58	1.59
	Std dev rec	0.51	0.52	0.51	0.53	0.51	0.48	0.49	0.47
Total benchmarkers	%buy	49%	48%	46%	49%	49%	53%	50%	52%
	%hold	45%	48%	48%	47%	46%	43%	45%	43%
	%sell	6%	5%	5%	4%	5%	4%	5%	6%
	Avg rec	1.53	1.56	1.58	1.55	1.56	1.50	1.55	1.54
	Std dev rec	0.51	0.49	0.49	0.48	0.48	0.47	0.48	0.46
Industry benchmarkers	%buy	37%	38%	39%	42%	40%	42%	40%	42%
	%hold	45%	47%	48%	49%	50%	49%	50%	49%
	%sell	18%	15%	13%	10%	10%	9%	10%	9%
	Avg rec	1.81	1.79	1.73	1.68	1.70	1.67	1.71	1.69
	Std dev rec	0.59	0.56	0.53	0.51	0.50	0.49	0.48	0.47

Table 5. Logistic Regressions Relating Optimistic/Pessimistic to Different Benchmarks

The table presents results of logistic regressions whose dependent variable equals 1 when a recommendation is either optimistic or pessimistic. Our sample period is between 9/2002 and 12/2009. All models use firm fixed effects. Optimistic recommendations are 'strong buy' and 'buy,' and pessimistic recommendations are 'sell' and 'strong sell.' **Industry** takes value of 1 if a broker uses an industry benchmark and 0 if a broker uses market or total return benchmarks. **AFF** is an indicator variable equal to 1 if the broker issuing the recommendation was a lead underwriter or a co-manager in an equity offering for the firm in the 24 months before the recommendation announcement date. **SANCT** is an indicator variable equal to 1 if the recommendation is issued by an analyst who is employed by a sanctioned brokerage house. **PASTFIRMPERF** is the average daily stock return over [-180, -2]. **PASTMKPERF** is the average daily market return over [-180, -2]. **ANALYST EXPERIENCE** is defined as the number of days the analyst has appeared in IBES. **TIER3** is an indicator variable for whether a brokerage house uses a three-tier recommendation grid at the time a recommendation is issued. Robust standard errors (in parentheses) are clustered at the firm level. ***, **, * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1) Prob(Rec=OPT)	(2) Prob(Rec=PESS)
Industry	-0.200***	0.475***
	(0.0122)	(0.0229)
AFF	0.314***	-0.652***
	(0.0240)	(0.0489)
PASTMKTPERF	-3.967	-9.011
	(6.798)	(10.39)
PASTFIRMPERF	43.03***	-53.34***
	(2.986)	(4.405)
SANCT	-0.248***	0.310***
	(0.0143)	(0.0243)
LOG(1+ANALYST EXPERIENCE)	-0.0182***	0.0260***
	(0.00346)	(0.00644)
TIER3	-0.276***	0.0390*
	(0.0126)	(0.0229)
Observations	149,673	129,290

Table 6. The Relation Between Recommendations, Earnings Forecasts and LTG Projections

This table presents average parameter values from running monthly Fama and MacBeth (1973) cross-sectional regressions—models (1) and (2)—of recommendation levels on measures of analysts' forecasts regarding earnings and long-term growth (LTG). The observations are monthly firms for each month between September 2002 and December 2009. A firm is included in the regression for month *t* only if the firm has outstanding recommendations and outstanding forecasts regarding next annual earnings and forecasts of LTG available at the end of that month. An outstanding recommendation (forecast) issued by a broker to a firm at time *t* is the most recent recommendation (forecast) issued by the broker to that firm that is not older than 12 months and that has not been cancelled by the broker. Models (i) and (iii) [(ii) and (iv)] is based on recommendations and forecasts issued by industry (market or total) benchmarkers only. The dependent variable is the average recommendation level among the outstanding recommendations available for the firm at the end of the month. **E/P** is a score based on the average earnings-price ratio forecasts for the firms in the sample, where earnings forecasts are average 1-year ahead annual earnings forecasts and price is the observed stock price when earnings data are collected. **AI_LTG** and **WI_LTG** (**AI_E/P** and **WI_E/P**) refer respectively to measures of across-industry and within-industry expectations of LTG (earnings-price ratio), and are computed as follows. Starting with the LTG forecasts, each month we first compute for each firm the consensus LTG as the average LTG forecast amongst the outstanding forecasts available for that firm. We then define for each industry an industry LTG forecast as the average LTG consensus across all firms in that industry. Then, for each firm in that month we compute the firm's industry-adjusted LTG forecast as the firm's LTG forecast minus its industry LTG forecast. We compute **WI_LTG** as a score between 0 and 1 based on the ranking of industry-adjusted LTG forecasts in each industry. For each firm we also calculate an across-industry LTG score, denoted as **AI_LTG,** based on the ranking of the industry LTG forecasts across all industries. Similarly, we calculate the within- and across-industry earnings estimate rankings denoted by WI_E/P and AI_E/P respectively, based on the analyst earnings forecast scaled by the stock price prevailing when the earnings data are collected. Robust standard errors (in parentheses) are calculated using the Fama-MacBeth (1973) autocorrelation-adjusted t-statistics. ***, **, * denote statistical significance at the 1%, 5%, 10% levels, respectively. The reported R^2s and number of observations are the time-series averages of the monthly cross-sectional regression measures.

Table 6. (*Continued*)

	(i) Industry Benchmarkers	(ii) Market or Total Benchmarkers	p-value (i)=(ii)	(iii) Industry Benchmarkers	(iv) Market or Total Benchmarkers	p-value (iii)=(iv)
Intercept	2.978***	2.853***	<0.0001	3.056***	2.991***	0.0488
	(0.052)	(0.044)		(0.036)	(0.048)	
LTG	-0.391***	-0.420***	0.2891			
	(0.031)	(0.023)				
AI_LTG				-0.176***	-0.271***	0.0180
				(0.036)	(0.027)	
WI_LTG				-0.341***	-0.334***	0.7489
				(0.037)	(0.018)	
E/P	-0.140**	-0.163***	0.2670			
	(0.058)	(0.046)				
AI_E/P				-0.038	-0.092***	0.0005
				(0.024)	(0.023)	
WI_E/P				-0.124***	-0.132***	0.4965
				(0.027)	(0.025)	
Observations	973	1,338		973	1,338	
R-square	5.20%	6.80%		6.10%	7.70%	

Table 7. Distribution of Recommendation Targets

This table summarizes the distribution of buy recommendation targets for market, industry and total return benchmarkers in our sample. For market (or industry) benchmarkers, a buy recommendation target is defined as the 'x' percent return a stock is expected to outperform the market (or industry) performance. For total return benchmarkers, a buy recommendation target is defined as the 'x' percent total return a stock is expected to achieve.

Panel A - Market Benchmarkers	Target	No. of Brokers
	0	20
	5%	5
	10%	1
	15%	4
	20%	1
	N.A.	3
	All	34

Panel B - Industry Benchmarkers	Target	No. of Brokers
	0	31
	5%	1
	10%	3
	20%	1
	N.A.	1
	All	37

Panel C - Total Benchmarkers	Target	No. of Brokers
	7%	1
	10%	10
	15%	14
	20%	7
	25%	1
	30%	1
	N.A.	8
	All	42

Table 8. Performance of Recommendations and Control Units

This table analyzes the performance of 'buy' and 'sell' recommendations issued by market/industry/total benchmarkers. Our sample period is between 9/2002 and 12/2009. Each recommendation is paired with a propensity score matched (control) unit according to the procedure described in Table A1. The table reports performance measures for the sample of recommendations and the corresponding sample of control units. In Panel A, the performance variable for each recommendation (control unit) is a dummy equal to 1 if the recommendation (control unit) achieved its stated objective. For a 'buy' recommendation, the stated objective from an industry (market) [total] benchmarker is $R_{industry}+target$ ($R_{market}+target$)[$target$], so achieving the objective means $R-R_{industry}-target>0$ ($R-R_{market}-target>0$)[$R-target>0$]. For a 'sell' recommendation, the stated objective from an industry (market) [total] benchmarker is $R_{industry}-target$ ($R_{market}-target$)[$target$], so achieving the objective means $R-R_{industry}+target<0$ ($R-R_{market}+target<0$)[$R-target<0$]. For a control unit, the stated objective is the same as in its corresponding recommendation. In Panel B, the performance variable is the difference between the cumulative stock return and the stated objective. In Panel C, the performance variable is the raw return. Returns associated with a recommendation (the stock return R, the industry return $R_{industry}$ and the market return R_{market}) are computed during the stated life span of a recommendation—the period in which the recommendation advice is kept alive. This is the period between the recommendation issuance and the earliest of (i) 12 months following the recommendation issuance and (ii) the date when the recommendation advice is changed (e.g., though a cancelation or an upgrade/downgrade by the same analyst). Returns associated with a control unit are computed for the period starting with the control unit issuance date (as defined in Table A1) and with the same number of days as the stated life span of its corresponding recommendation. P-values for test of difference of proportions is reported under the column *Diff (p-value)*.

Panel A: Proportion of Recommendations Achieving the Stated Objective

	Buys				Sells			
		% achieving the objective				% achieving the objective		
	# obs	Buy	Control	Diff p-value	# obs	Sell	Control	Diff p-value
Industry	11,108	49.8%	43.7%	0.0000	4,166	58.8%	48.0%	0.0000
Market	8,121	52.3%	45.6%	0.0000	1,553	58.3%	43.1%	0.0000
Total	11,870	38.9%	31.7%	0.0000	1,605	36.3%	17.4%	0.0000

Table 8. (*Continued)*

Panel B: Return in Excess of the Stated Objective

	Buys						Sells					
	Recommendation		Control		Diff (p-value)		Recommendation		Control		Diff (p-value)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Industry	0.0312	-0.0005	-0.0077	-0.0252	0.0000	0.0000	-0.0327	-0.0389	0.0357	0.0077	0.0000	0.0000
Market	0.0552	0.0128	0.0194	-0.0198	0.0000	0.0000	-0.0320	-0.0417	0.0483	0.0240	0.0000	0.0000
Total	-0.0481	-0.0843	-0.0723	-0.1094	0.0001	0.0000	0.1148	0.0891	0.1977	0.1649	0.0000	0.0000

Panel C: Raw Return

	Buys						Sells					
	Recommendation		Control		Diff (p-value)		Recommendation		Control		Diff (p-value)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Industry	0.1040	0.0729	0.0836	0.0473	0.0003	0.0000	0.0112	-0.0190	0.0871	0.0415	0.0000	0.0000
Market	0.1090	0.0747	0.0852	0.0379	0.0008	0.0000	-0.0290	-0.0465	0.0637	0.0317	0.0000	0.0000
Total	0.0994	0.0616	0.0752	0.0324	0.0001	0.0000	-0.0233	-0.0455	0.0596	0.0266	0.0000	0.0000

Table 9. Decomposition of Returns

This table analyzes the decomposition of returns in excess of the stated objective for recommendations issued by market/industry/total benchmarkers. The sample of recommendations and control units is described in Table 9. *R* ($R_{industry}$) [R_{market}] refer to firm (industry) [market] returns. Such returns for a recommendation are computed during the stated life span of a recommendation—the period in which the recommendation advice is kept alive. This is the period between the recommendation issuance and the earliest of (i) 12 months following the recommendation issuance and (ii) the date when the recommendation advice is changed (e.g., though a cancelation or an upgrade/downgrade by the same analyst). The returns associated with a control unit are computed for the period starting with the control unit issuance date (as defined in Table A1) and with the same number of days as the stated life span of its corresponding recommendation. Panel A relies on the actual benchmark target associated with each recommendation, while Panel B repeats the computation after ignoring the benchmarks (that is, setting each benchmark to 0). P-values for test of difference of mean (t-test) and median (Wilcoxon) are reported under the columns *Diff (p-value)*.

Panel A: Decomposition of Returns Using Brokers' Targets

Buys

	# obs	Recommendation Mean	Recommendation Median	Control Mean	Control Median	Diff (p-value) Mean	Diff (p-value) Median
Industry benchmarkers							
R - ($R_{industry}$+target)	11,108	0.0312	-0.0005	-0.0077	-0.0252	0.0000	0.0000
Market Benchmarkers							
R - $R_{industry}$	8121	0.0513	0.0122	0.0143	-0.0133	0.0000	0.0000
$R_{industry}$ - (R_{market}+target)	8121	0.0039	-0.0037	0.0051	-0.0021	0.5830	0.8480
Total Benchmarkers							
R - $R_{industry}$	11,870	0.0478	0.0086	0.0155	-0.0112	0.0000	0.0000
$R_{industry}$ - R_{market}	11,870	0.0189	0.0072	0.0176	0.0069	0.4414	0.6185
R_{market} - target	11,870	-0.1148	-0.0892	-0.1054	-0.0902	0.0000	0.1168

Sells

	# obs	Recommendation Mean	Recommendation Median	Control Mean	Control Median	Diff (p-value) Mean	Diff (p-value) Median
Industry benchmarkers							
R - ($R_{industry}$ - target)	4,166	-0.0327	-0.0389	0.0357	0.0077	0.0000	0.0000
Market Benchmarkers							
R - $R_{industry}$	1,553	-0.0701	-0.0645	0.0139	-0.0042	0.0000	0.0000
$R_{industry}$ - (R_{market} - target)	1,553	0.0381	0.0175	0.0344	0.0212	0.4211	0.4647
Total Benchmarkers							
R - $R_{industry}$	1,605	-0.0680	-0.0590	0.0168	-0.0032	0.0000	0.0000
$R_{industry}$ - R_{market}	1,605	0.0162	0.0024	0.0144	0.0032	0.6612	0.7799
R_{market}+target	1,605	0.1667	0.1656	0.1665	0.1692	0.9508	0.4773

Table 9. (*Continued*)

Panel B: Decomposition of Returns Settting Target to 0

Buys

	# obs	Recommendation Mean	Recommendation Median	Control Mean	Control Median	Diff (p-value) Mean	Diff (p-value) Median
Industry benchmarkers							
$R - R_{industry}$	11,108	0.0465	0.0137	0.0076	-0.0128	0.0000	0.0000
$R_{industry} - R_{market}$	11,108	0.0152	0.0034	0.0218	0.0080	0.0002	0.0000
R_{market}	11,108	0.0423	0.0670	0.0541	0.0703	0.0000	0.0000
Market Benchmarkers							
$R - R_{industry}$	8121	0.0513	0.0122	0.0143	-0.0133	0.0000	0.0000
$R_{industry} - R_{market}$	8121	0.0209	0.0089	0.0221	0.0075	0.5830	0.8480
R_{market}	8121	0.0368	0.0595	0.0488	0.0634	0.0000	0.0003
Total Benchmarkers							
$R - R_{industry}$	11,870	0.0478	0.0086	0.0155	-0.0112	0.0000	0.0000
$R_{industry} - R_{market}$	11,870	0.0189	0.0072	0.0176	0.0069	0.4414	0.6185
R_{market}	11,870	0.0327	0.0556	0.0421	0.0562	0.0000	0.1168

Sells

	# obs	Recommendation Mean	Recommendation Median	Control Mean	Control Median	Diff (p-value) Mean	Diff (p-value) Median
Industry benchmarkers							
$R - R_{industry}$	4,166	-0.0470	-0.0530	0.0214	-0.0041	0.0000	0.0000
$R_{industry} - R_{market}$	4,166	0.0200	0.0030	0.0231	0.0067	0.3109	0.1388
R_{market}	4,166	0.0381	0.0564	0.0426	0.0584	0.1629	0.3332
Market Benchmarkers							
$R - R_{industry}$	1,553	-0.0701	-0.0645	0.0139	-0.0042	0.0000	0.0000
$R_{industry} - R_{market}$	1,553	0.0171	-0.0003	0.0134	0.0060	0.4211	0.4647
R_{market}	1,553	0.0240	0.0335	0.0364	0.0371	0.0084	0.0372
Total Benchmarkers							
$R - R_{industry}$	1,605	-0.0680	-0.0590	0.0168	-0.0032	0.0000	0.0000
$R_{industry} - R_{market}$	1,605	0.0162	0.0024	0.0144	0.0032	0.6612	0.7799
R_{market}	1,605	0.0286	0.0311	0.0284	0.0305	0.9508	0.4773

Figure 1. End-of-Month Distribution of Outstanding Recommendations

This figure presents, for each month between September 2002 and December 2009, the fraction of 'buys' and fraction sells among the outstanding recommendations issued by market, total, and industry benchmarkers. Only brokerage houses which issued at least 100 recommendations to U.S. firms during our sample period (9/2002 – 12/2009) are included in the analysis.

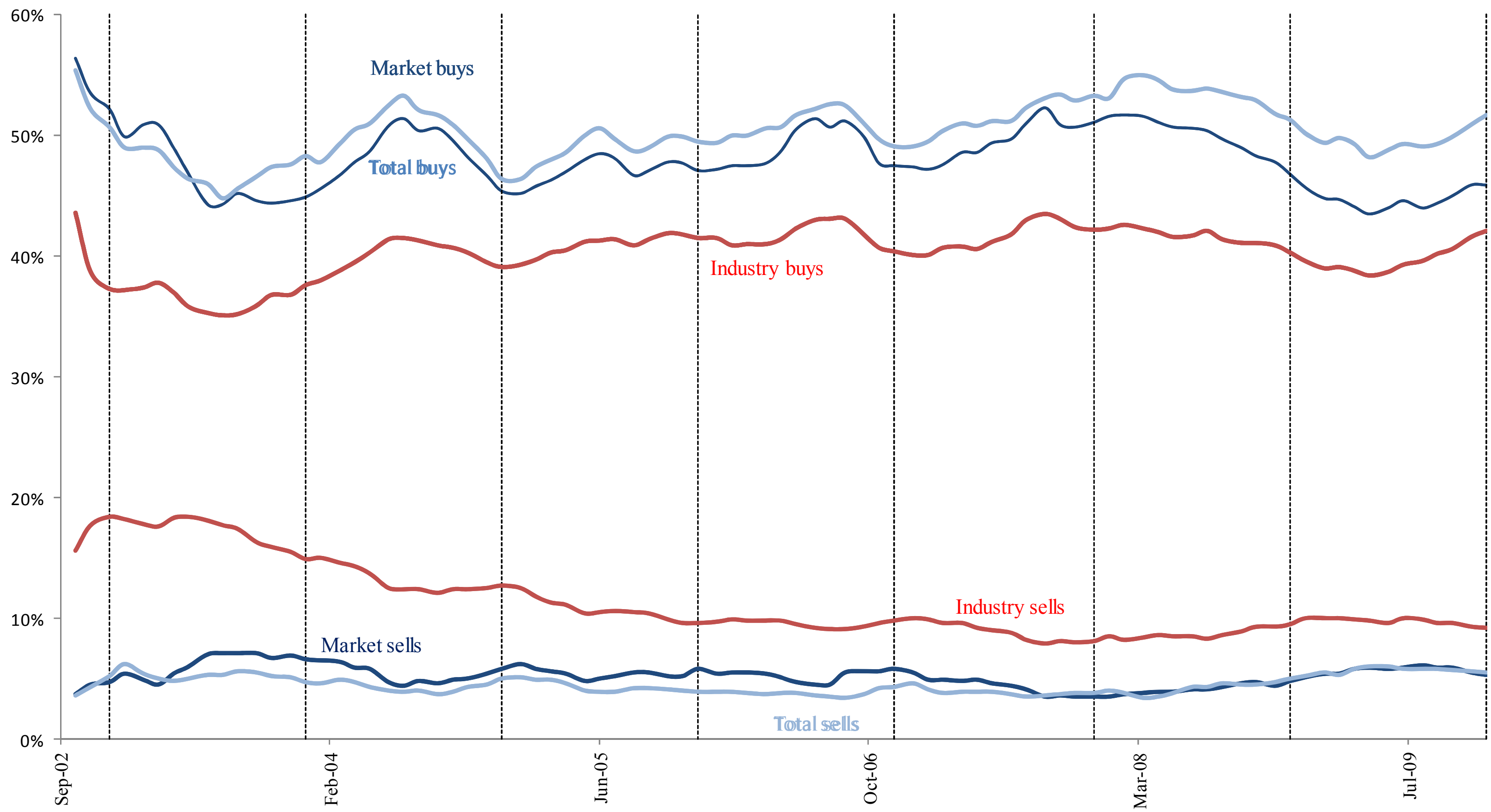

Figure 2. Monthly Net Pessimism of Total Against Market Benchmarkers ($\beta_{0,Total}$- $\beta_{0,Market}$) vs. CFNAI

This figure presents, for each month between September 2002 and December 2009, the measures of ($\beta_{0,Total}$ - $\beta_{0,Market}$) and of the Chicago Fed National Activity Index (CFNAI). The estimate of $\beta_{0,Total}$ ($\beta_{0,Market}$) for a specific month is the intercept from running model (2) for the sample of recommendations from total (market) benchmarkers for that specific month. The solid vertical line represents the peak (December 2007) and the dashed vertical line represents the trough (June 2009) of business cycles within our sample period, according to the NBER's Business Cycle Dating Committee.