

Relative accuracy of analysts' earnings forecasts over time: a Markov chain analysis

Derann Hsu · Cheng-Huei Chiao

Published online: 16 October 2010
© Springer Science+Business Media, LLC 2010

Abstract The main purpose of this paper is to analyze the time patterns of individual analysts' relative accuracy ranking in earnings forecasts using a Markov chain model. Two levels of stochastic persistence are found in analysts' relative accuracy over time. Factors underlying analysts' performance persistence are identified and they include analyst's length of experience, workload, and the size and growth rate of firms followed by the analyst. The strength and the composition of these factors are found to vary markedly in different industries. The findings support the general notion that analysts are heterogeneous in their accuracy in earnings forecasts and that their superior/inferior performance tends to persist over time. An analysis based on a refined measure of analysts' forecast accuracy ranking that strips off firm-specific factors further enhances the empirical validity of the findings. These findings provide a concrete basis for researchers to further explore why and how analysts perform differently in the competitive market of investment information services.

Keywords Earnings forecasts · Persistence of performance · Information efficiency · Markov chain models

JEL Classification G14 · C21 · C22 · C53 · D84

D. Hsu (✉)
Sheldon B. Lubar School of Business, University of Wisconsin, Milwaukee,
P.O. Box 742, Milwaukee, WI 53201-0742, USA
e-mail: hsu@uwm.edu

C.-H. Chiao
Department of Business, Missouri Western State University,
4525 Downs Drive, St. Joseph, MO 64507, USA
e-mail: cchiao@missouriwestern.edu

1 Introduction

Financial information conveyed through analysts' earnings forecasts has been long considered an important element in the valuation of a firm, as it can impact strongly the firm's stock price returns.¹ Research findings documenting this understanding have been reported by Abarbanell and Bernard (1992), Bao et al. (1997), Beneish and Harvey (1998), Lim (2001), Diether et al. (2002), Copeland et al. (2004), Loh and Mian (2006), and numerous other authors whose work has been reviewed in a survey article by Ramnath et al. (2008, Sect. 3.3).

It is hence important to understand the inner working of analysts' information processing mechanism, its efficiency, and underlying contributing factors. To this end, researchers have long studied the relative accuracy of analysts' earnings forecasts in hopes of acquiring greater insight into these subjects. Such research work has been reported by Richards (1976), Brown and Rozeff (1980), O'Brien (1987, 1990), Butler and Lang (1991), Stickel (1992), Sinha et al. (1997), and other authors whose empirical findings have been summarized in Ramnath et al. (2008, Sect. 3.2).

Earlier research papers on analysts' relative earnings forecast accuracy, approximately prior to year 1991, reported results that showed little evidence of persisting differential performance among analysts. The empirical results were viewed as evidence in support of the hypothesis that superior (inferior) analysts do not exist.² Contrary to this view, Stickel (1992) and Sinha et al. (1997) provide empirical findings that suggest the existence of superior analysts. These two studies have motivated and provided a guidepost for our research presented in this paper. To help the reader capture the main idea, the questions raised in their work, and reasons behind our present work, we highlight in the next two paragraphs the research findings reported therein.

Stickel (1992) analyzed earnings forecast accuracy of the members of the All-American Research Team for the years 1981–1985. He showed evidence suggesting that team members exhibited superior accuracy in earnings forecasts during their tenure on the team. Two observations are of special interest. First, the membership of the team was elected by users of analysts' services, predominantly institutional investors, via a voting scheme sponsored by the publisher of *Institutional Investor*. Second, Sinha et al. (1997, p. 38, endnote #2, based on the data of 1985–1990) later pointed out that the length of tenure of membership on the team was on average less than 2 years. The first observation raises the question: In what other ways can we sort out the “superior” analysts from the “inferior” analysts without a time consuming and obviously subjective voting scheme? The second observation leads us to another question: How long can a superior (inferior) analyst stay superior (inferior) over multiple years?

The second paper by Sinha et al. (1997) presents an *ex ante* analysis that investigates whether an analyst's past superior (inferior) performance in a formation sample period can carry forward to a holdout sample period (of 1 year) on a firm-specific, and industry-specific, basis. They show that based on the data for the period 1984–1993, there was a significant carryover effect for superior performance, but not for inferior performance.

¹ A discussion that relates analysts' accuracy in earnings forecasts to their ability in recommending stocks that generate superior returns is given in Loh and Mian (2006). It is seen to suggest the significance of information contents in earnings forecasts that apply to equity valuation.

² There are several possible explanations for such findings. They include: (1) lack of extensive database, (2) primitive nature of statistical analysis as constrained by the computer facilities available at that time, and (3) the wide spread belief at that time in the implausibility of the existence of superior analysts in an efficient market.

The question left unanswered was: How prevalent and strong this phenomenon is over different industries and time periods?

Besides these open questions, we noted that to date, research on this subject has been carried out on an ad hoc basis. Relative accuracy of analysts' earnings forecasts was analyzed without an organized statistical model structure. A case in point is Sinha et al. (1997). On the other hand, characteristics of analysts with superior/inferior forecast accuracy normalized by average absolute forecast errors in a firm were analyzed, but without first having their relative ranking identified, or persistence checked (Clement 1999; Ghosh and Whitecotton 1997; Jacob et al. 1999; Mikhail et al. 1997, 2003). In light of these observations, in this paper we conduct a comprehensive empirical analysis employing newer database and more efficient statistical modeling concepts and techniques³ derived from the Markov chain model.⁴ This is with the aim to systematically explore the nature, stochastic properties, and determinants of analysts' relative forecast accuracy.

More specifically, we conduct our analysis to address three linked questions: first, does an analyst's superior (inferior) forecasting accuracy statistically persist over time; second, if so, how long will the superior (inferior) performance last; and third, what are the determinants of analysts' relative forecasting accuracy?

A brief preview of our analytic approach and empirical findings follows.

We adopted a measure for analysts' forecast accuracy called Absolute Proportional Forecast Error (AFE). We then compiled historical transition probability matrices of the quintile rank of analysts' average annual AFE from a formation year to five subsequent observation years. This was done for the overall dataset and each of the thirteen major industries separately. Our results suggest that analysts are highly heterogeneous in their forecast accuracies, and their relative performance persists much more strongly than as suggested by a pure chance model. This persistence is then classified into a "first order" and, when appropriate, a "second order" Markov chain model. We call such persistence "first degree Markov persistence" and "second degree Markov persistence",^{5,6} respectively.

We then identified a two-component Markov chain model that partitions the hidden influencing factors into two separate components. One is named "transient" and the other "long-lasting". We then estimated the proportional share and the strength of each component. We have found that, for the overall dataset, the transient and the long-lasting component each contributes to analysts' time pattern in relative forecast accuracy in a ratio of about 58–42, while the staying power, in terms of the half-life span, of the latter is about ten times that of the former. However, the share and strength of either component varies considerably among different industries. We also observe that, while analysts' relative forecast accuracy exhibits the first degree Markov persistence in each of the thirteen industries, it shows evidence of the second degree Markov persistence in just five industries.

To search for long-lasting influencing factors, we linked analysts' relative performance to their professional characteristics and salient features of the firms that they follow. The results show that, based on data from the years 2004 through 2006 and for the overall

³ Use of ranking statistics in the study of performance persistence arises naturally. For instance, Carhart (1997) uses it in the study of mutual fund performance.

⁴ References for Markov chains can be found in Grimmet and Stirzaker (2001), Chap. 6.

⁵ It is similar to the actuarial table for life expectancy conditional on a person's current age. The conditional probability of survival shifts as the person's current age increases, but not in a monotonic fashion.

⁶ This finding applies to both superior and inferior performers. Our results are somewhat different from those reported in Sinha et al. (1997). In the latter work, it was found that while superior performance persists, inferior performance does not.

dataset, with other things being held constant, an analyst with longer experience performs better, while an analyst with heavier workload performs worse. In a similar manner, an analyst's forecast accuracy is better for firms with larger capitalization sizes. Further, the size of the sponsoring brokerage house and the average capitalization-size growth rate of the followed firms do not significantly affect an analyst's performance in the sample data that we study. But these findings for the overall dataset do not apply uniformly to individual industries. This observation appears to suggest that analysts' performance in different industries hinges importantly on different types of information gathering skills, analytical capacities, and other factors peculiar to individual industries. A discriminant function analysis on the predictive power of the set of explanatory variables confirms the validity of our findings.

As a check on the robustness of our findings, we used an alternative relative accuracy measure designed to control firm-specific factors. The empirical results confirm the existence of the "first degree Markov persistence" for most industries.

We hasten to point out that the analysis reported in this study can help improve investors' utilization of analysts' earnings forecasts. For instance, this analysis can be used to sort out "superior" analysts from "inferior" analysts. A composite earnings forecast can hence be constructed by placing more weights on forecasts made by superior analysts and less on forecasts by inferior analysts. Such a forecast will likely perform better than the now-prevalent simple consensus forecast, which places equal weight on forecasts made by all analysts following a firm. It is also hoped that a modified composite earnings forecast can provide early signs of shifts (innovations) in earnings trend, and help characterize intrinsic investment risks in a real-time investment analysis.

The organization of the remainder of this paper is as follows: Section 2 describes the data structure and constraints. Section 3 defines the basic variables and explains the methodology. Section 4 provides descriptive statistics of the analyst data and reports preliminary findings on Markov persistence in various sub-samples. Section 5 reports formal test results for the *first and second degree Markov persistence* for thirteen industries. Section 6 presents estimation results for a two-component Markov chain model. Section 7 reports results that link analysts' performance quintile ranking to characteristics of the analysts and the followed firms. Section 8 provides additional results based on an alternative measure for accuracy ranking. Section 9 summarizes the findings of the study and provides concluding remarks.

2 Data structure and constraints

The data for our study come from two sources and consists of all firms that appear in both (a) the Institutional Brokers Estimate System (I/B/E/S) files, and (b) the Center for Research in Security Prices (CRSP) files. We obtain earnings data from I/B/E/S and price and capitalization data from CRSP. Forecasts of quarterly earnings reported by analysts from over 300 brokerage firms are extracted from I/B/E/S Detail History files. Each record of forecast made by an analyst for a firm and a fiscal quarter contains the company ticker, brokerage firm identifier, analyst identifier, ending date of the forecasted quarter, earnings estimate, and the date on which the forecast was made. I/B/E/S retains, in its best efforts, the analyst identifier code as an analyst moves from a brokerage firm to another. The sample covers the period from the first quarter of 1984 through the fourth quarter of 2006, consisting of 92 consecutive quarters. However, in light of insufficient number of analysts at the industry level in earlier years, we conduct our main analysis using data starting from 1988.

To alleviate undesirable influences from irregular data points and in consideration of compatibility of quality and timing of earnings forecasts made by different analysts, we impose several requirements on the data. They are stated and explained below.

First, we focus on quarterly predictions made up to 90 days but no less than 30 days prior to the end of the fiscal quarter being forecasted.⁷ If an analyst makes more than one forecast for that quarter within that time window, only the *first* forecast is considered. This is to avoid the data complexity arising from continual revisions of earnings estimates by individual analysts. This is also to alleviate potential distortions from herding among analysts during the latest part of the quarter. We feel that such restriction of the timing window for earnings forecasts is appropriate and necessary.^{8,9} When a firm's fiscal quarter is different from a regular calendar quarter, but ends in a particular calendar quarter, the earnings estimates of that fiscal quarter are identified with that calendar quarter.

Second, we require that each firm in the sample must have earnings forecasts from *at least four different analysts* in the quarter immediately prior to the quarter being studied. This consideration follows the examples set by Hilary and Menzly (2006) and Abarbanell and Lehavy (2003). The purpose is to reduce the number of occurrences of unduly large proportional forecasting errors often associated with little known firms followed by only one or two analysts. These firms usually lack reliable financial information.

Third, we require that an analyst make forecasts for *at least ten different quarters for a given firm* over our sample period in order to be included in our sample for that firm. This is to assure that an analyst's performance persistence is examined on firms being followed by the analyst on a reasonably steady basis.

Fourth, we cap (winsorize) at 1.0 the absolute error of a quarterly earnings forecast after it is deflated by the absolute value of the corresponding actual earnings number (i.e., quarterly AFE as defined below). This is to reduce the undue impacts of an actual earnings number that is near zero.

Finally, fifth, in the event that forecasts from an analyst for a firm are missing for some quarters in a year, the average of the remaining available quarterly AFEs is used as the analyst's yearly AFE.

⁷ Only 1% of analyst-firm combinations in the I/B/E/S database have forecasts recorded at more than 90 days prior to the end of the fiscal quarter.

⁸ We recognize that a forecast made closer to the end of the quarter for which an earnings forecast is made may be based on more newly available information and thus result in better accuracy. For that reason, in our measuring system, an analyst can gain an edge by postponing the forecast as long as possible. However, a forecast made later in the designated quarter is less useful to the analyst's clients and can be detrimental to the analyst's job evaluation. An analyst is expected to provide the forecast within a reasonable time window, usually within the first 2 months of the fiscal quarter. The fact that a definite majority of analysts in our sample (over 75%) made their first forecast for the quarter during the first 2 months of the quarter confirms this contention.

⁹ An in-depth investigation, conducted by the present authors, of analysts' timing of quarterly earnings forecasts reveals that only a small percentage of analysts (about 3%) have more than 50% of their quarterly earnings forecasts made later than 30 days prior to the end of the fiscal quarter. These analysts often carry a heavy forecasting workload because they follow a large number of firms. We also note that few analysts have a preset narrow time range in which they made all, or a majority of, their forecasts. Instead, they made forecasts at varying times in the quarter over the sample years and for different firms. In that sense, the 60-day window that we use to rank analysts' forecast accuracy does not run a significant risk of generating biased results in favor of chronically late forecasters. To narrow or divide the time window further will unduly reduce the available data points and make analysis difficult and less reliable. It will take another special study to devise a mechanism to adjust the effects of timing differences across a spectrum of analysts and firms.

In analysis reported in later sections, we divide our sample firms into thirteen industry groups based on the first two digits of their SIC numbers, as suggested in Breeden et al. (1989, pp. 243–245, Table 2). These thirteen industries are: Petro (petroleum), Consum (consumer durables), Basic (basic industries), Food (food and tobacco), Constr (construction), Financ (finance and real estates), Capita (capital goods), Transpo (transportation), Utilit (utilities), Textile (textiles and trade), Servic (services), Leisur (leisure), and Other (others which are not included in the preceding twelve groups).

3 Methodology

Suppose that an analyst, j , makes, at time $t - 1$, a forecast of earnings per share (EPS) of firm k for a quarter ending at time t . We shall denote that forecast as ${}_{t-1}EPS_{j,t}^k$ and the actual EPS of firm k for a quarter ending at time t as EPS_t^k . We then measure an analyst's forecast accuracy for the quarter ending at time t by the absolute value of the ratio of the forecast error to the actual earnings (abbreviated as AFE_t) [see Sinha et al. (1997), and Cooper et al. (2001)].¹⁰ It is formally defined below.¹¹

$${}_{t-1}AFE_{j,t}^k = \left| \frac{{}_{t-1}EPS_{j,t}^k - EPS_t^k}{EPS_t^k} \right|,$$

The formula is an adaptation of the standard statistical measure for forecasting precision, termed “absolute percentage error” (APE), which is commonly employed in statistical modeling and forecasting literature (Hanke and Wichern 2009).

For a firm and a specific quarter, an analyst's quarterly AFE is computed as defined above. The analyst's annual AFE for a firm in a year is the average of the four quarterly AFEs of the year. The analyst's “average annual AFE” in a year is then the average of annual AFEs of all of the firms that the analyst follows in that year. Such a composite measure is needed to assess an analyst's general intrinsic ability in information acquisition and analysis. It is distinct from unique firm-specific talents that may have limited value to investors in general because of susceptibility to herding or imitations among analysts who follow the same firm. Further, a firm-by-firm assessment of analysts' relative forecast accuracy is difficult to carry out reliably because of analysts' frequent turnovers. A relative accuracy measure that is specifically designed to neutralize firm-specific effects is presented and discussed in a later section.

In a particular year, all analysts selected under our data selection criteria as specified in Sect. 2 are ranked and separated into five equal-sized groups, that is, by their quintiles, based on their average annual AFEs. The quintile group with smallest average annual AFEs is called “quintile 1”, or the “top quintile”, while the quintile group with the largest average annual AFEs is called “quintile 5”, or the “bottom quintile”. The other three quintile groups are defined similarly. After analysts are identified by these five initial

¹⁰ The time point $t - 1$ in our context here is actually a fraction of a calendar quarter prior to the quarter end point t . This is because we take the first forecast of an analyst within 3 months, but more than 30 days prior to the time point t .

¹¹ Our definition of the AFE here is just one of the several possible measures used to rank analysts' accuracy in earnings forecasts. We have examined advantages and disadvantages of several measures that have been employed in the literature. In the end, we chose to use the definition provided above, in conjunction with a carefully chosen winsorization scheme, for the main analysis in this paper. In a later section, we also conduct a robustness study based on another measure of relative forecast accuracy.

quintile ranks in a formation year, we track their quintile ranks in the subsequent five observation years. We started the first tracking cycle by placing the formation year in 1988 and moved the formation year and the trailing “five-year observation window” forward one year at a time till it reached year 2001. Fourteen tracking cycles resulted.

Development of the Markov chain model for our analysis is explained below. In a pure chance model, where accuracy among analysts has no intrinsic difference, an analyst’s quintile rank in the formation year should come from a random draw from integers 1 through 5. This same phenomenon should then repeat itself in the subsequent observation years. This means that the quintile rank of an analyst in subsequent years should have equal probability to fall within each of the five possible quintile ranks. As such, an analyst’s average rank in each of the subsequent observation years, regardless of the initial quintile group membership, should be equal to 3.0 (the mathematical center point). Evidence of a departure from this pure chance model in favor of positive persistence suggests that there are intrinsic differences in analysts’ forecast accuracy and that truly superior/inferior performers exist.

To prepare our discussions in Sect. 5, we describe below the Markov transition probability matrix and its use in obtaining the frequency distribution under the one-step Markov chain model.

Define the transition probability matrix consisting of 6×6 elements, i.e., $P_{i,j}$, for a transition from quintile i (i th row) in one year to quintile j (j th column) in the next year as follows.

$$\mathbf{P}^{(i)} = \begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} \\ P_{21} & P_{22} & P_{23} & P_{24} & P_{25} & P_{26} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} & P_{36} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} & P_{46} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} & P_{56} \\ P_{61}^{(i)} & P_{62}^{(i)} & P_{63}^{(i)} & P_{64}^{(i)} & P_{65}^{(i)} & P_{66}^{(i)} \end{pmatrix} = \begin{pmatrix} P'_1 \\ P'_2 \\ P'_3 \\ P'_4 \\ P'_5 \\ P_{6.}^{(i)'} \end{pmatrix} \tag{1}$$

where $P_{6j}^{(i)} = (0.05) \left[\left(P_{ij} / \sum_{l=1}^5 P_{il} \right) \right]$, and $P_{66}^{(i)} = 0.95$, $i, j = 1, 2, 3, 4, 5$, $\sum_{j=1}^6 P_{ij} = 1$, for $i = 1, 2, 3, 4, 5$, and the index “ i ” represents the initial quintile rank. Entries on the last (6th) row and the last (6th) column of the matrix represent the probabilities of missing observations (due to temporary or permanent exit of analysts from the sample). Due to the programming complexity in calculating precise frequencies for the entries on the 6th row, we estimate $P_{66}^{(i)}$ to be 0.95 (probability to stay missing in the next year) and maintain a proportionally identical distribution on the remaining elements of the 6th row (as seen on the transition probability array), going from the corresponding initial quintile group to all quintile groups, summed to 0.05. We also note that in a steady one-step Markov chain process the values of the elements in the first five rows of the transition probability matrix are fixed when an analyst’s quintile number moves from the time origin to the next year, and then to the year after, and so on. These elements are unaffected by the membership of an analyst’s initial or subsequent quintile rank. Further, all of the values of the 6th row are set equal to zero for the transition from the formation year to the first observation year.

Let the expected probability arrays for t subsequent years computed based on the one-step Markov chain model which imposes a constant one-step transition probability matrix beyond the first year, be expressed as

$$M^{(i)} = \left(M_1^{(i)}, M_2^{(i)}, M_3^{(i)}, \dots, M_t^{(i)} \right), \tag{2}$$

where $M_1^{(i)} = P_{i.}$, and $M_t^{(i)} = \mathbf{P}^{(i)} \cdot M_{t-1}^{(i)}$, $i = 1, 2, 3, 4, 5$, and $t = 2, 3, 4, 5$. We define the term ‘‘first degree Markov persistence’’ as a state in which a violation of the equal probability hypothesis takes place in favor of positive persistence of initial quintile rank, where the null hypothesis is

$$H_0^{(1st)} : P_{ij} = P_{ik}, \quad \text{for } j \neq k, \quad \text{and } i, j, k = 1, 2, 3, 4, 5, \quad \text{and} \\ P_{i6} = P_{j6}, \quad \text{for } i \neq j, \quad \text{and } i, j = 1, 2, 3, 4, 5.$$

In other words, under the null hypothesis $H_0^{(1st)}$, $P_{ij} = (1 - \bar{P}_6)/5$, for $i, j = 1, 2, 3, 4, 5$, and $P_{i6} = \bar{P}_6$, for $i = 1, 2, 3, 4, 5$, where \bar{P}_6 is the estimated average probability of missing observations in the first observation year across all five initial quintile groups.

Further, to test the conformity of the observed frequency distributions of quintile ranks in five subsequent years to the one-step Markov chain model, we define the state of ‘‘second degree Markov persistence’’ as a rejection, in favor of a higher degree of positive persistence, of the null hypothesis listed below.

$$H_0^{(2nd)} : M_t^{(i)} = M_t^{(i)*}, \quad i = 1, 2, 3, 4, 5 \quad \text{and} \quad t = 2, 3, 4, 5,$$

where $M_t^{(i)}$ is the array of observed relative frequencies, while $M_t^{(i)*}$ is the array derived from the one-step Markov chain model, based on the one-step transition probability matrix $\mathbf{P}^{(i)}$ as described in Eq. 2. In summary, a state of second degree Markov persistence can occur only after the first degree Markov persistence is established, for the former is at a higher level of persistence.

In testing the first hypothesis, we use a χ^2 test statistic on a two-way contingency table. In the second test, we use a one-way χ^2 test for the equality of an observed discrete frequency distribution to its theoretical counterpart. The latter procedure applies to the individual frequency distribution of each initial quintile group and each of the five observation years.

Further, we define $\text{rank}(k, t)$ as the average quintile rank value at time lag t (the t -th observation year) for the group of analysts initially ranked in the k -th quintile, that is,

$$x(k, t) = |\text{rank}(k, t) - 3.0|, \quad t = 0, 1, 2, 3, 4, 5, \tag{3}$$

where $t = 0$ refers to the formation year, and $x(k, 0) = |\text{rank}(k, 0) - 3| = |k - 3|$. Further, let

$$y(k, t) = x(k, t) / x(k, t - 1), \quad t = 1, 2, 3, 4, 5.$$

We define

$$\beta_1 = y(k, 1), \tag{4}$$

and

$$\beta_2 = \text{geometric average } [y(k, 2), y(k, 3), y(k, 4), y(k, 5)]. \tag{5}$$

The term β_1 is the convergence rate of the average quintile rank to the median rank (i.e., 3.0) from $t = 0$ to $t = 1$, while β_2 is the geometric average of the convergence rates of the average quintile rank from $t = 1$ to $t = 5$, both for the analysts initially in the k -th quintile group. Under $H_0^{(2nd)}$, the value of β_1 is approximately equal to β_2 . On the other hand, a violation of $H_0^{(2nd)}$ in favor of a stronger positive persistence beyond the first observation

year (i.e., exceeding the first degree Markov persistence) implies $\beta_2 > \beta_1$. Testing the equality of the two parameters can shed light on whether the second degree Markov persistence exists.

Another way to look at the second-degree Markov persistence from the perspective of a multiple-component Markov process is through a process in which multiple components (categories) of hidden factors, embedded among analysts or within each analyst, are mixed in a random fashion to steer the observed transition probability matrix. For simplicity, we use a two-component model for analysis. A straightforward way to estimate the structure of the model is to examine the pattern of convergence of the average quintile rank toward three over the five observation years for the extreme (the top and the bottom) initial quintile groups. The two-component Markov model implies the following pattern of the convergence of the average quintile rank toward three over the five subsequent observation years.

$$f(p, a, b) = x(k, t)/x(k, 0) = pe^{-at} + (1 - p)e^{-bt}, \quad k = 1, 5; \quad t = 0, 1, \dots, 5, \quad (6)$$

where $a > b$ and the quantity $x(k, t)$ has been defined in Eq. 3. In the above equation the quantity p is the probability assigned to the first component, which has an exponential convergence parameter “ $-a$ ” associated with some “transient” factors. The remainder of the probability is assigned to the second component which has a convergence parameter “ $-b$ ” associated with some “long-lasting” factors. The values e^{-a} and e^{-b} represent the convergence rate toward the median rank, three, for the “transient” and the “long-lasting” factors respectively.¹²

Details of the computational procedures for the defined terms and relevant test statistics under the two hypotheses and for the two-component Markov chain model outlined above are provided in later sections.

4 Descriptive statistics and distributions of quintile ranks

We first tabulated the frequency distribution, before the data reductions as explained in Sect. 2, of the number of analysts who have ever covered a given number of firms/industries for the period 1984–2006. A total of 415,654 data points (quarter-analyst-firm triplets) are involved. We observed that 40% of analysts cover only one industry, while 17% of analysts cover more than three industries, with the average number of industries covered by an analyst equal to 2.25. Furthermore, we note that 13% of analysts cover only one firm, while about 30% cover up to three firms. On the other hand, about 28% of analysts cover more than ten firms. The average number of firms ever covered, at one time or another, by an analyst is 8.8. The number of distinct analysts involved in our qualified sample is 2,655.

We then constructed summary statistics of analysts’ average annual AFEs and the number of analyst-firm-year triplets in our data sample, cross-tabulated by industry and by year. Some noticeable characteristics of the data are summarized below. *First*, the value of the average AFE, aggregated over the 23-year sample period, varies substantially in different industries, striding a range between 0.15 (for finance) and 0.32 (for petroleum). Two industries (petroleum and transportation) post average AFE values above 0.25, while the

¹² The “transient” factors may include pure luck, big earnings surprises, occasional information advantages, irregularity in firms’ accounting treatments, etc. The “long-lasting” factors may involve analysts’ information gathering skills, analytic capacity, workload, in-depth understanding of the industry and firm being followed, etc.

remaining eleven industries below 0.25. *Second*, the AFE value, averaged over thirteen industries, for each of the 23 years, is close to 0.22, except for the periods 1994–1997, year 2000, 2005 and 2006, during which the average AFE value is at around 0.19. It exhibits the time-varying nature of the AFE. *Third*, basic, finance and consumer durable have the highest overall level of analyst participation, as is reflected by the number of analyst-firm-year triplets recorded in the sample, while other, construction, and transportation have the lowest level of analyst participation.¹³ The three characteristics outlined above suggest the need to separate analysts by industry. Further, a quintile ranking system that focuses on the relative performance among analysts in a year is a natural way to normalize the impacts of time varying level of the average AFE over different years.¹⁴

Table 1 displays, for the overall dataset as an illustration for our analytic procedure, the frequency distributions of quintile ranks in five successive observation years averaged over the fourteen tracking cycles from 1988 through 2001.¹⁵ The entries in this table show that the probability of staying in the initial quintile rank is about 40% for the two extreme (the first and the fifth) initial quintile groups in the first observation year. This probability declines to about 17% (32% when excluding missing observations) in the fifth year. The probabilities of missing observations climb steadily from about 15 to about 45% in 5 years for the two extreme quintile groups. On the other hand, the probabilities of missing observations move up from about 8% to about 41% in 5 years for the middle three initial quintile groups.¹⁶

As previously noted, the percentage frequency distributions exhibited in Table 1 are the average of fourteen sets of such distributions with the formation year starting in 1988 and ending in 2001. In each replication of the tabulation, the average rank value was computed and the fourteen average rank values were used to conduct a test against the hypothesis of a true mean equal to 3.0 under the pure chance model. The *t*-values so computed are listed in Table 1 on the last row of each panel. All *t*-test results, except for the third quintile group, are statistically significant at less than 1%, two-sided, level. This indicates a strong persistence in analysts' performance for the overall dataset.

As evidenced by our discussion earlier, the average accuracy of earnings forecasts in terms of AFEs varies depending on the firm's industry. This then necessitates a separate analysis for each industry. To begin, we compiled the number of analysts involved in a quintile group for each of the thirteen industries, and in each formation year, from 1988 through 2001.¹⁷ We observed that a quintile size (i.e., the number of analysts in a quintile) ranges from 6 to 110. The average number of analysts in a quintile group, averaged over

¹³ We also note that the value of standard deviation of the AFEs (not shown in the table) is strikingly parallel to the corresponding mean value in both industry and year dimensions, due to the positive correlation between them.

¹⁴ Several studies mentioned in the introductory section appear to overlook the significance of the industry and year factors in their analysis (Clement 1999; Ghosh and Whitecotton 1997; Jacob et al. 1999; Mikhail et al. 1997, 2003).

¹⁵ Recognizing that the overall dataset pools together analysts who follow different industries, and hence involves rankings of AFEs from heterogeneous groups, we present the results in Table 1 mainly to illustrate our analytical procedure. Results from analysis on individual industries are to follow.

¹⁶ Missing observations can happen for various reasons, such as switching to other roles as research director, money manager or investment officer in asset management firms, firm executive, or exiting the profession in pursuit of other interests. It can also result, in a small number of cases, from our data trimming schemes outlined in Sect. 2.

¹⁷ When an analyst covers firms in more than one industry, the analyst is counted as a distinct analyst in each of the industries. For that reason, the analyst's performance in each of the different industries is ranked separately competing against analysts in that particular industry.

Table 1 Frequency distributions of analysts' accuracy quintile ranks in five subsequent years after the initial rank in a formation year (overall AFE data—up to 2006—forecasting day within -90 to -30 of the end of a fiscal quarter)

Successive year	1	2	3	4	5
Panel A: Percentage frequency of quintile rank based on average AFEs in five test years with initial quintile 1 group					
Quintile 1 (%)	40.27	31.27	25.16	21.18	17.15
Quintile 2 (%)	20.20	18.45	17.27	15.09	14.24
Quintile 3 (%)	11.36	10.85	11.16	11.16	10.72
Quintile 4 (%)	6.60	7.69	7.50	6.87	6.90
Quintile 5 (%)	5.37	6.40	6.27	6.22	5.24
Missing (%)	16.19	25.33	32.64	39.48	45.74
Survived (%)	83.81	74.67	67.36	60.52	54.26
Average rank	2.006	2.194	2.292	2.359	2.412
<i>t</i> -value	-29.95	-23.43	-23.00	-14.32	-13.99
Panel B: Percentage frequency of quintile rank based on average AFEs in five test years with initial quintile 2 group					
Quintile 1 (%)	19.32	16.89	14.57	12.74	11.85
Quintile 2 (%)	28.07	22.19	17.75	15.76	13.41
Quintile 3 (%)	20.94	18.86	18.07	15.97	13.91
Quintile 4 (%)	12.70	13.08	13.11	13.04	11.47
Quintile 5 (%)	7.62	9.94	8.73	7.20	7.02
Missing (%)	11.35	19.05	27.78	35.30	42.34
Survived (%)	88.65	80.96	72.22	64.70	57.66
Average rank	2.564	2.717	2.769	2.779	2.789
<i>t</i> -value	-16.50	-9.29	-8.19	-5.74	-5.93
Panel C: Percentage frequency of quintile rank based on average AFEs in five test years with initial quintile 3 group					
Quintile 1 (%)	9.67	10.70	9.25	8.23	7.93
Quintile 2 (%)	22.41	17.99	17.43	15.55	12.29
Quintile 3 (%)	26.10	21.85	18.90	17.22	15.49
Quintile 4 (%)	21.36	20.10	17.58	15.29	14.01
Quintile 5 (%)	12.18	12.53	11.81	10.29	8.97
Missing (%)	8.28	16.83	25.04	33.41	41.32
Survived (%)	91.72	83.17	74.96	66.59	58.68
Average rank	3.046	3.070	3.066	3.053	3.057
<i>t</i> -value	1.82	2.55	2.43	1.35	1.81
Panel D: Percentage frequency of quintile rank based on average AFEs in five test years with initial quintile 4 group					
Quintile 1 (%)	5.47	5.42	5.92	5.73	4.93
Quintile 2 (%)	10.62	11.62	11.63	10.14	9.93
Quintile 3 (%)	22.36	20.84	18.09	15.40	13.77
Quintile 4 (%)	29.40	25.03	21.80	19.30	15.86
Quintile 5 (%)	23.57	19.44	17.32	15.21	12.74
Missing (%)	8.57	17.66	25.24	34.23	42.77
Survived (%)	91.43	82.34	74.76	65.77	57.23

Table 1 continued

Successive year	1	2	3	4	5
Average rank	3.600	3.502	3.443	3.429	3.377
<i>t</i> -value	23.77	20.34	17.98	21.53	11.78
Panel E: Percentage frequency of quintile rank based on average AFEs in five test years with initial quintile 5 group					
Quintile 1 (%)	4.38	4.85	4.76	5.10	4.32
Quintile 2 (%)	6.92	7.62	7.57	7.24	7.18
Quintile 3 (%)	12.60	13.42	12.70	12.52	10.76
Quintile 4 (%)	24.50	21.60	19.31	16.89	15.95
Quintile 5 (%)	37.96	28.69	24.94	20.39	16.37
Missing (%)	13.65	23.81	30.72	37.85	45.41
Survived (%)	86.35	76.19	69.28	62.15	54.59
Average rank	3.980	3.808	3.757	3.660	3.609
<i>t</i> -value	26.06	28.86	20.53	14.89	16.85

The term $|_{t-1}AFE_{j,t}^k|$ is the absolute value of proportional forecast error in earnings, deflated by the actual earnings, made by analyst j at time $t - 1$ for the earnings of firm k at time t . The quintile ranking is based on average quarterly AFEs in a year across all firms covered by an analyst. The subsequent 5-year tracking window after the initial quintile formation year is moved 1 year at a time with the formation year shifting from 1988 through 2001, with the final test year reaching 2006. The *t*-value is based on the 14 sets of mean values of the yearly quintile ranks of the analysts identified in an initial quintile group in the formation year, moving from 1988 through 2001. The null hypothesis tested here is H_0 : average rank = 3, the expected rank under the pure chance hypothesis. The two-sided critical values of the *t*-statistic at **5 and 1%** significance level for 13 degrees of freedom are **2.160 and 3.012**, respectively

fourteen formation years, ranges among the industries from 12.1 (for transportation industry) to 77.2 (for capital goods industry). Moreover, the number of analysts in a quintile group in the overall dataset rises from 124 in 1988 to 375 in 2001, a 202% increase.

The frequency table in Table 1 for the distributions of quintile rank in five observation years was produced for each of the thirteen industries. For economy of space, we present in Table 2 only the key results from the test of the pure chance hypothesis for the initial top and bottom quintile groups. The main entries on the table are the values of the average quintile rank, one for each industry-observation-year combination, with its *t* statistic (against the null hypothesis that the average rank is equal to 3.0) listed in the parentheses immediately beneath. Results exhibited in Table 2 suggest that quintile rank persistence among analysts is strong in nearly all industries, with the strongest cases in basic industries, finance, capital goods, textile, and services; transportation is the only marginal case.¹⁸

¹⁸ In an earlier version of this paper, we produced results in a format similar to Table 2 for nine Morningstar Investment Style groups. The reason for doing so is to determine whether, after separating firms into more homogeneous groups by their capitalization size and P/B ratio, the average quintile ranking scores deviate in a significant way from the median value, 3.0, in successive observation years. The results, not shown here for economy of space, behave with even stronger degree of departure from the pure chance model compared with the case with the 13 industries as shown in Table 2.

Table 2 Average quintile ranks based on AFEs in each of the 5 years subsequent to the formation year for the two extreme initial quintile groups for each of the thirteen industries

Observation year	Petro	Consum	Basic	Food	Constr	Financ	Capita	Transpo	Utilit	Textile	Service	Leisure	Other	All
A. Average quintile ranks at each successive year following the formation year and the <i>t</i> -ratios for testing their deviations from 3.0 (initial top-quintile group)														
1	2.190 (-5.74)	2.253 (-10.16)	1.896 (-16.25)	2.354 (-7.06)	2.244 (-6.14)	2.070 (-13.61)	2.063 (-15.99)	2.614 (-2.48)	2.082 (-8.46)	2.019 (-11.26)	2.240 (-10.87)	2.321 (-7.33)	2.376 (-3.94)	2.195 (-25.55)
2	2.450 (-4.42)	2.611 (-5.18)	2.042 (-11.85)	2.693 (-2.07)	2.278 (-5.95)	2.232 (-11.70)	2.184 (-18.23)	2.628 (-2.56)	2.344 (-5.62)	2.175 (-8.59)	2.441 (-8.37)	2.535 (-5.50)	2.685 (-2.17)	2.368 (-18.50)
3	2.492 (-3.23)	2.789 (-3.87)	2.150 (-12.96)	2.921 (-0.51)	2.261 (-5.01)	2.289 (-8.35)	2.351 (-11.67)	3.010 (0.07)	2.589 (-3.10)	2.228 (-8.17)	2.486 (-6.17)	2.629 (-4.20)	2.646 (-2.97)	2.440 (-19.83)
4	2.713 (-1.25)	2.827 (-2.41)	2.172 (-12.86)	2.892 (-0.54)	2.449 (-3.09)	2.350 (-8.39)	2.411 (-8.73)	2.804 (-1.53)	2.515 (-3.31)	2.217 (-9.38)	2.653 (-3.57)	2.464 (-5.94)	2.480 (-2.37)	2.570 (-15.21)
5	2.593 (-2.12)	2.791 (-1.88)	2.201 (-8.98)	2.875 (-1.39)	2.432 (-3.75)	2.583 (-5.54)	2.323 (-8.07)	2.811 (-1.71)	2.718 (-1.81)	2.192 (-7.75)	2.676 (-4.82)	2.540 (-3.97)	2.610 (-2.25)	2.576 (-18.72)
B. Average quintile ranks at each successive year following the formation year and the <i>t</i> -ratios for testing their deviations from 3.0 (initial bottom-quintile group)														
1	3.644 (7.09)	3.718 (9.86)	4.147 (23.43)	3.932 (7.74)	3.635 (4.27)	4.100 (13.77)	3.963 (18.44)	3.403 (2.67)	3.977 (11.70)	3.845 (9.20)	3.948 (14.71)	3.874 (10.36)	3.772 (5.03)	3.810 (21.81)
2	3.403 (2.87)	3.541 (6.68)	3.961 (16.44)	3.623 (4.47)	3.616 (3.79)	3.857 (7.87)	3.746 (15.08)	2.962 (-0.24)	3.713 (7.46)	3.563 (4.84)	3.582 (7.80)	3.758 (9.15)	3.678 (5.45)	3.689 (20.10)
3	3.477 (2.83)	3.472 (9.61)	3.926 (13.82)	3.713 (5.20)	3.549 (3.48)	3.808 (7.69)	3.571 (10.22)	3.233 (1.63)	3.671 (6.16)	3.512 (4.95)	3.642 (10.05)	3.526 (3.70)	3.041 (0.16)	3.597 (19.42)
4	3.420 (2.34)	3.351 (4.30)	3.883 (10.77)	3.689 (3.95)	3.360 (2.77)	3.727 (6.25)	3.465 (6.37)	3.299 (2.01)	3.400 (4.22)	3.498 (5.30)	3.535 (8.60)	3.552 (4.79)	3.540 (3.02)	3.524 (18.40)
5	3.280 (1.28)	3.345 (4.26)	3.793 (11.21)	3.567 (2.84)	3.331 (1.59)	3.683 (5.08)	3.488 (10.18)	2.954 (-0.14)	3.166 (0.95)	3.554 (4.90)	3.426 (5.18)	3.696 (5.06)	3.099 (0.36)	3.460 (11.70)

Listed above are average quintile ranks of analysts in the specified initial quintile group (the first and the fifth) for each of the subsequent five observation years. The *t*-ratio is to test the difference between the observed average quintile ranks and the theoretical average quintile rank under the pure chance hypothesis, that is, 3. The fourteen sets of the mean quintile ranks (each set contains five numbers, one for each of the subsequent 5 years) obtained from each of the fourteen formation years from 1988 through 2001, are assumed to be random draws from a steady distribution and are used to test the hypothesis that the overall mean estimate is compatible to 3. These tests are carried out for analysts participating in each of the thirteen industries and for analysts ever involved in the dataset. The reported *t*-values are in the parentheses beneath the coefficients. The two-sided critical values of the ***t*-statistic at 5 and 1%** significance level for 13 degrees of freedom are **2.160 and 3.012**, respectively. The basic term $_{t-1}AFE_{jt}^k$ is the absolute value of proportional forecast error made by analyst *j* at time *t* - 1 for the earnings of firm *k* at time *t*, deflated by the actual reported earnings. The overall dataset in this table contains analysts' records of performance ranking grouped by industry. An analyst who covered firms in multiple industries appears in our overall dataset multiple times, one for each industry involved. The analysis displayed in the last column of the table above is based on the sub-divided data counts. The industry grouping method used here is from Breeden et al. (1989)

5 Markov persistence for thirteen industries

In search for further insights, we examined in more depth the stochastic properties of analysts' performance persistence. The two sets of tests against $H_0^{(1st)}$ and $H_0^{(2nd)}$ outlined in Sect. 3 were used to conduct the analysis. Table 3 presents the results of such analysis for the *overall dataset*. Panel A of Table 3 provides the estimated one-year Markov transition probability matrix using the elements contained in Table 1. The numbers on or near the diagonal positions in the matrix reveal that the probability of an analyst to remain in or close to his/her initial quintile group is uniformly substantially above that predicted by the equal probability model, that is, $(1 - P_{i6})/5 = (1 - \bar{P}_6)/5$, or approximately 0.177.

We also observe, as previously noted, that the probabilities of missing observations for the middle three quintiles (between 8 and 11% as displayed on the last column) are measurably lower than those for the first and the fifth quintile (16 and 14%, respectively). This pattern of frequency distribution appears to link analysts' career mobility (moving up or out from an analyst position) to their relative performance in forecast accuracy.

A two-way contingency table test is used to examine whether the computed transition probability matrix conforms to the equal probability hypothesis $H_0^{(1st)}$. The χ^2 test statistic thus computed carries an exceedingly large value of 1,913.2, giving virtually a zero p -value.¹⁹ This result shows a strong first-degree Markov persistence in analysts' performance in earnings forecasts for the overall dataset.

Panels B and C of Table 3 provide the expected (under $H_0^{(2nd)}$) and the observed frequency distributions of quintile rank, one for each of the five observation years, for the two extreme initial quintile groups. (Details for the middle three initial quintile groups are omitted for economy of space.) The expected frequency distribution is obtained through the use of Eq. 2 explained in Sect. 3. Results from a series of χ^2 tests are listed in Panel D of Table 3. For each initial quintile group and a particular observation year, we computed a χ^2 statistic to test the conformity of the observed relative frequencies to the expected relative frequencies (one-way contingency).²⁰ The p -values of the χ^2 test statistics are provided for the second through the fifth observation year for each of the five initial

¹⁹ We first constructed the expected frequency table under $H_0^{(1st)}$ as described in Sect. 3. We then calculated the quantity $\chi^2 = \sum_{i=1}^5 \sum_{j=1}^6 \left[\frac{(f_{ij}^{(o)} - f_{ij}^{(e)})^2}{f_{ij}^{(e)}} \right]$, where $f_{ij}^{(e)}$ is the expected number of observations in the cell at the i -th row and j -th column, and $f_{ij}^{(o)}$ is its observed counterpart. The degrees of freedom for the χ^2 statistic is $(5 - 1) \times (6 - 1) = 20$. (Since the frequencies on the 6th row are estimated values as explained in Sect. 3, we exempt them from the test.) The number of observations in each cell used in the calculation of the χ^2 statistic here is set equal to one half of the observations we actually count. This is to allow for the overlapping nature of our moving window. In this particular circumstance, the (first) observation year of this cycle overlaps with the formation year in the next cycle. The true number of independent observations is not exactly known, but it should be between 50 and 100% of the number of observations that we actually count. To make the matter simple and to be on the conservative side, we use the lower bound, that is, one half. Specifically, the number of observations used in the test statistic is equal to $n \times 5 \times (14/2) = 35n$, where n is the number of analysts in a quintile, averaged over the fourteen formation years (1988–2001), and the number "5" is because there are five quintiles in a dataset. The number "14" in the expression above is because of the fourteen formation cycles in our sample—altogether, it is $267.6 \times 35 = 9,366$, the number of estimated independent observations for the overall dataset. In Table 4, we apply the same calculation procedure to each of the thirteen industries, with the value of n replaced by the average quintile group size for that particular industry.

²⁰ The degrees of freedom in the test is $(6 - 1) = 5$. The numbers of (independent) observations are estimated to be $4n$, $3n$, $3n$, and $3n$ for year 2, 3, 4, and 5 respectively, where n is the average number of analysts in a quintile over the 14-year cycle. Again, these estimates are on the conservative side.

Table 3 Tests of conformity of observed frequency distributions of quintile ranks to the pure chance transition model and the one-step Markov chain model (overall data—based on average AFEs)

Initial group	Relative frequency distribution of quintile ranks in the first observation year									
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Missing				
Panel A: One-step Markov transition probability matrix										
Quintile 1	0.403	0.202	0.114	0.066	0.054	0.162				
Quintile 2	0.193	0.281	0.209	0.127	0.076	0.114				
Quintile 3	0.097	0.224	0.261	0.214	0.122	0.083				
Quintile 4	0.055	0.106	0.224	0.294	0.236	0.086				
Quintile 5	0.044	0.069	0.126	0.245	0.380	0.137				
Missing	X1	X2	X3	X4	X5	0.95				
Chi-square contingency-table test statistic = 1,913.2 (<i>p</i> -value = 0.0000)										
Quintile rank	Expected probabilities in five subsequent observation years					Observed relative frequencies in five subsequent observation years				
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 1	Year 2	Year 3	Year 4	Year 5
Panel B: Probability distributions through multi-year transitions: initial top-quintile group										
Quintile 1	0.403	0.222	0.153	0.122	0.107	0.403	0.313	0.252	0.212	0.172
Quintile 2	0.202	0.176	0.147	0.127	0.113	0.202	0.185	0.173	0.151	0.142
Quintile 3	0.114	0.140	0.136	0.126	0.116	0.114	0.109	0.112	0.112	0.107
Quintile 4	0.066	0.110	0.122	0.120	0.113	0.066	0.077	0.075	0.069	0.069
Quintile 5	0.054	0.087	0.102	0.104	0.101	0.054	0.064	0.063	0.062	0.052
Missing	0.162	0.264	0.340	0.400	0.449	0.162	0.253	0.326	0.395	0.457
Panel C: Probability distributions through multi-year transition: initial bottom-quintile group										
Quintile 1	0.044	0.074	0.088	0.090	0.087	0.044	0.049	0.048	0.051	0.043
Quintile 2	0.069	0.109	0.119	0.116	0.108	0.069	0.076	0.076	0.072	0.072
Quintile 3	0.126	0.156	0.148	0.136	0.124	0.126	0.134	0.127	0.125	0.108
Quintile 4	0.245	0.206	0.172	0.149	0.134	0.245	0.216	0.193	0.169	0.160

Table 3 continued

Quintile rank	Expected probabilities in five subsequent observation years					Observed relative frequencies in five subsequent observation years				
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 1	Year 2	Year 3	Year 4	Year 5
Quintile 5	0.380	0.228	0.171	0.144	0.128	0.380	0.287	0.249	0.204	0.164
Missing	0.137	0.228	0.302	0.365	0.418	0.137	0.238	0.307	0.379	0.454
Initial quintile rank	<i>p</i> -value of chi-square test result									
	Year 2		Year 3		Year 4		Year 5			

Panel D: Results of chi-square tests for conformity to one-step Markov chains	
Quintile 1	0.000
Quintile 2	0.007
Quintile 3	0.062
Quintile 4	0.000
Quintile 5	0.000

Based on AFEs as previously defined, the entries above provide the observed (averaged across fourteen formation cycles) one-step transition probability matrix for quintile ranks and the expected frequency distributions of quintile ranks in successive years derived from the one-step Markov chain model for analysts' forecast accuracy in the first and the fifth initial quintile group. The average number of analysts for each quintile in this table is 267.6. The missing probabilities (Xs) are profiled according to the transition probabilities from the initial quintile onto quintiles 1 through 5 and are summed to a total of 0.05. The formation year moves from 1988 to 2001 one year at a time, with a total of fourteen formation cycles

quintile groups. The test results are exceedingly significant for the two extreme initial quintile groups where they really count. This result suggests a strong second-degree Markov persistence in analysts' accuracy in earnings forecasts for the overall dataset.

Because the nature and degree of difficulty in earnings forecasting may vary from industry to industry, we carried out the same analysis as presented in Table 3 for each of the individual thirteen industries. These results are displayed in Table 4. Two main findings emerge and are outlined below.

First, results of analysis for all thirteen industries exhibit strong first-degree Markov persistence. The pure chance hypothesis for analysts' relative forecast accuracy is rejected at exceedingly low significance levels for each of the thirteen industries.

Second, the second-degree Markov persistence is confirmed for the following five industries: basic industries, finance, capital goods, textiles, and services, occurring mainly with the initial top quintile group.

6 The two-component Markov chain model and the behaviors of β_1 and β_2

Analysis using β_1 and β_2 , as explained in Sect. 3, has several benefits beyond the analysis presented in Sect. 5. First, it condenses the various sets of test statistics contained in Tables 3 and 4 to two parameters. Second, it captures some important characteristics of the Markov persistence: the *rate of convergence* of the average quintile rank toward the median rank (i.e., 3) in successive observation years, from the initial quintile rank. Third, the values of the two parameters can be displayed in the form of time series to exhibit the strength of persistence over time. And fourth, a simple test on the equality between β_1 and β_2 (as explained in Sect. 3) can shed light on the existence of the second-degree Markov persistence.

We calculated the values of β_1 and β_2 for each combination of formation year (1988–2001, 14 formation years) and industry (13, plus the overall dataset), and for each of the two extreme quintile groups. Because of the large volume of numerical results, we summarize below important observations distilled from the results.

First, for the overall dataset, the values of both β_1 and β_2 for the two extreme initial quintile groups exhibit a stable and horizontal time pattern, with the average values equal to approximately 0.54 and 0.85, respectively. The observed values of β_1 and β_2 over the fourteen tracking cycles are within $\pm 15\%$ range of the average values. An analysis of variance test confirms the hypothesis of constant value for both β_1 and β_2 .

Second, a similar horizontal pattern, but at different average levels and with larger proportional variations over time, is observed for both β_1 and β_2 with individual industries. A special analysis for the stationarity of the two parameters for the two extreme initial quintile groups and for each industry group was carried out. The general finding is that there is little evidence to suggest a significant degree of non-stationarity, or a discernable non-horizontal time pattern, on the two parameters for any of the thirteen industries. The proportional variations over time around the average parameter values are largely within a $\pm 25\%$ range.

Third, tests of the equality of the values of the two parameters β_1 and β_2 for the overall dataset and for each of the thirteen industries were carried out. The tests were based on the estimates over the fourteen tracking cycles. The test results confirm the second-degree Markov persistence for the overall dataset and the following seven industries: consumer durables, basic industries, construction, finance, capital goods, textiles, and leisure. Four out of the seven industries identified here are also identified by the tests against $H_0^{(2nd)}$ presented in Sect. 5.

Table 4 Tests of conformity of observed frequency distributions of quintile ranks to the pure chance transition model and the one-step Markov chain model for thirteen industries (based on AFEs)

Industry	Petro	Consum	Basic	Food	Constr	Financ	Capita	Transpo	Utilit	Textile	Servic	Leisur	Other	Overall
A. Chi-square two-way contingency-table test values & the corresponding p -values for the test against $H_0^{(1st)}$														
Chi-square	109.8	289.4	547.9	87.5	58.8	371.5	503.2	23.3	161.9	271.4	289.0	119.5	53.9	1,913.2
(p -value)	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
B. p -values for the test against $H_0^{(2nd)}$ for the initial top-quintile group														
<i>Test year</i>														
2	0.824	0.272	0.009	0.765	0.349	0.006	0.000	0.540	0.652	0.126	0.035	0.669	0.488	0.000
3	0.646	0.520	0.002	0.936	0.196	0.016	0.000	0.889	0.799	0.032	0.008	0.625	0.622	0.000
4	0.611	0.689	0.000	0.967	0.579	0.014	0.000	0.941	0.420	0.012	0.112	0.304	0.522	0.000
5	0.452	0.550	0.001	0.980	0.462	0.368	0.000	0.993	0.992	0.005	0.374	0.405	0.901	0.000
C. p -value for the test against $H_0^{(2nd)}$ for the initial bottom-quintile group														
2	0.470	0.085	0.085	0.862	0.680	0.278	0.211	0.989	0.407	0.299	0.877	0.422	0.745	0.000
3	0.360	0.091	0.007	0.449	0.865	0.031	0.361	0.967	0.061	0.626	0.176	0.815	0.917	0.000
4	0.663	0.519	0.003	0.711	0.770	0.034	0.610	0.871	0.800	0.458	0.231	0.694	0.790	0.000
5	0.977	0.488	0.020	0.601	0.685	0.352	0.311	0.786	0.832	0.460	0.340	0.332	0.797	0.000

Based on AFEs as previously defined, the entries above provide chi-square test results from testing the null hypothesis $H_0^{(1st)}$ and $H_0^{(2nd)}$, that is, the pure chance transition model and the one-step Markov chain model, respectively. The intermediate steps of the test procedure have been illustrated in Table 3. The industry grouping method used here is from Breeden et al. (1989)

The findings reported in the first two points above are important, and in fact remarkable, in the sense that the two characterizing measures of the accuracy ranks persistence pattern appear to be stable over the sample period 1988–2006. This is valid not only for the overall dataset, but also for the data of individual industries. This may imply that the stochastic dynamic behaviors of analysts' forecasting activities within the investment community stay remarkably stable throughout the sample years. From an analytical point of view, this stationarity enables us to drastically simplify our analysis by aggregating data and measured variables across the time dimension. In that case, we can focus our attention on the cross-sectional characteristics of individual industries and explore the driving factors that are peculiar to each particular industry.

As explained in the immediately preceding paragraph, we can now conduct more in-depth analysis of individual industries based on data aggregated over the period 1988–2006. In this case, we conduct a factors partition analysis employing the two-component Markov chain model. We estimated the proportional weight placed on the transient and the long-lasting component of factors and then computed the respective strength of each of the two components in steering the convergence rate of the average quintile rank for analysts who participate in an identified dataset. In such estimations, we fitted the function $f(p, a, b)$ from Eq. 6 to the values of the average quintile rank reported in Table 2 using a hybrid adaptive splines algorithm (Lou and Wahba 1997). The results of estimation are displayed in Table 5.

From the entries on the last row in Table 5, we observe that, for the overall dataset, the weight taken by the transient component, p , averaged over the two extreme initial quintile groups, is about 58%, while the long-lasting factors capture the remaining 42% weight. Further, the value of the parameter “ a ” is 2.38 for quintile 1 and 4.26 for quintile 5, meaning a steep exponential convergence ratio of $e^{-a} = 0.093$ and 0.014, for the transient factors, for the two extreme groups, respectively. On the other hand, the value of the parameter “ b ” is about 0.12, averaged over the two extreme quintile groups, implying $e^{-b} = 0.887$, a slow exponential declining rate for the long-lasting factors. In other words, the long-lasting component lasts more than ten times in duration that of the transient component, i.e., $(e^{-b})^{10} = e^{-a}$. Furthermore, the values of β_1 and β_2 calculated based on the estimated function $f(p, a, b)$ are displayed on the last line, the right compartment of each of the two panels, of Table 5. They are about 0.45 and 0.85, again averaged over the two extreme quintile groups, respectively.

The same analysis was individually carried out for each of the thirteen industries and the results are separately tabulated in the main body of Table 5 for each of the top and the bottom quintile group. It is seen that the estimated values of the parameters a , b , and p for different industries vary markedly. This implies that the nature and the strength of factors included in the transient and long-lasting components that affect analysts' forecast accuracy vary substantially among different industries. Several significant observations on the values of these parameters, as well as on the derived values of β_1 and β_2 , are highlighted below.

First, among different industries, a relatively low value of “ p ” means weaker influences of the transient factors, or equivalently, stronger impacts of the long-lasting factors. Industries with relatively low p values are basic industries, capital goods, and utility.

Second, a smaller value of e^{-a} (larger value of a) coupled with a larger value of e^{-b} (smaller value of b) result in stronger—that is, second-degree—persistence. Again the three industries mentioned in the first point above serve as examples of this condition.

Table 5 Estimated parameter values for a two-component Markov chain model (based on AFEs)

Industry	Initial top-quintile group				Initial bottom-quintile group					
	a	b	p	β_1	β_2	a	b	p	β_1	β_2
Petroleum	2.477 (0.084)	0.287 (0.751)	0.518	0.405	0.730	3.344 (0.035)	0.177 (0.838)	0.643	0.322	0.823
Consumer	2.946 (0.053)	0.607 (0.545)	0.348	0.374	0.539	2.388 (0.092)	0.118 (0.889)	0.665	0.359	0.848
Basic	3.402 (0.033)	0.119 (0.888)	0.393	0.552	0.883	1.851 (0.157)	0.065 (0.938)	0.466	0.574	0.906
Food	1.777 (0.169)	0.442 (0.643)	0.675	0.323	0.576	4.372 (0.013)	0.243 (0.784)	0.413	0.466	0.782
Construction	3.201(0.041)	0.120 (0.887)	0.542	0.428	0.876	3.551 (0.029)	0.130 (0.878)	0.601	0.368	0.867
Finance	2.180 (0.113)	0.060 (0.942)	0.575	0.465	0.907	1.899 (0.149)	0.145 (0.865)	0.440	0.550	0.838
Capital	2.461 (0.085)	0.186 (0.830)	0.485	0.469	0.811	1.731 (0.177)	0.240 (0.787)	0.419	0.532	0.758
Transportation	5.138 (0.006)	0.327 (0.721)	0.738	0.193	0.717	1.277 (0.279)	0.159 (0.853)	0.786	0.402	0.702
Utility	3.150 (0.043)	0.300 (0.741)	0.404	0.459	0.734	2.051 (0.129)	0.203 (0.816)	0.477	0.489	0.789
Textiles	2.143 (0.117)	0.050 (0.951)	0.552	0.491	0.918	2.285 (0.102)	0.296 (0.744)	0.500	0.423	0.720
Services	1.667 (0.189)	0.192 (0.825)	0.700	0.380	0.742	1.229 (0.292)	0.175 (0.839)	0.668	0.474	0.737
Leisure	2.372 (0.093)	0.204 (0.816)	0.659	0.340	0.776	2.315 (0.099)	0.238 (0.788)	0.437	0.487	0.770
Others	2.093 (0.123)	0.110 (0.896)	0.756	0.312	0.820	1.660(0.190)	0.182 (0.834)	0.540	0.486	0.785
Overall	2.378 (0.093)	0.106 (0.900)	0.616	0.527	0.855	4.258 (0.014)	0.143 (0.867)	0.542	0.405	0.862

Based on the two-component Markov chain model specified, the three parameters for the model, a , b , and p are estimated from average quintile rank statistics in five observation years after a formation year, averaged over the fourteen formation cycles, 1988–2001, using a spline smoothing technique. Each of the number in the parentheses, which is e^{-x} , where x is the number to its left, provides an indication of the decaying rate toward the central level, 3, of the quintile rank of the component in question. The variables β_1 and β_2 are the estimated decaying rate from the initial quintile rank to the quintile rank of the first observation year and the average decaying rate for the remaining four observation years, respectively

Third, a large value of β_1 coupled with a large value of β_2 , the latter usually being close to 1.0, implies a strong first-degree Markov persistence. In addition, a large differential between the values of β_1 and β_2 enhances the second-degree Markov persistence.

Fourth, the estimated values of the parameters for the initial top quintile group (left panel of Table 5) and those for the initial bottom quintile group (right panel) are in fairly close agreement for the vast majority of the industries. This implies that the persistence is fairly symmetric on either side, that is, for superior or inferior performance.

7 Factors influencing analysts' performance

To understand the factors that may explain accuracy differentials among analysts, we chose to conduct a special study that links analysts' quintile rank scores for the last 3 years in our sample period (2004–2006) to characteristics of analysts and the firms they follow.²¹ The quintile ranks score of an analyst in the three sample years is calculated by summing the quintile rank number of the analyst in each of the 3 years. For instance, an analyst whose average AFE is ranked in the first quintile for each of the 3 years has a total quintile ranks score of 3. For that reason, the range of an analyst's quintile rank score for the 3 years is between 3 and 15, inclusively. We call this score SCORE-A (quintile ranks score based on average AFE).²²

In our attempt to identify factors that may explain analysts' relative forecasting accuracy, we noted in the extant literature several variables employed in related work. In the research papers by Clement (1999), Jacob et al. (1999), Mikhail et al. (1997, 2003), Ghosh and Whitecotton (1997), Abarbanell and Lehavy (2003), and Stickel (1992), three variables were used. They are: (1) analyst's workload, (2) analyst's length of experience, and (3) the size of the brokerage house that employed the analyst.

Reasoning for the selection of these three variables is fairly intuitive. To achieve higher accuracy in earnings forecasts for a firm, an analyst will need to devote more time and efforts to analyze the firm. But a heavy workload in terms of the number of firms covered by the analyst will constrain the analyst's ability to carry out the task for each firm thoroughly. Further, the longer an analyst's experience in forecasting earnings for a firm, just like in most professional careers, the better the analyst is supposed to perform. Finally, the more resources and supports an analyst has the greater forecasting accuracy the analyst is likely to achieve. And it is often the case that larger brokerage houses are more resourceful.

In addition, Brown (1997) and Ho (2004) pointed out that an analyst's forecast accuracy is inversely related to the capitalization size of the firm being followed. That is, analysts' forecasts for larger firms tend to be more accurate than for smaller firms, holding other things constant. This is often because of the better quality and accessibility of financial information afforded by larger firms.

²¹ The motivation for using the last three sample years, 2004–2006, was mainly that they are relatively recent and are more relevant to current conditions. We also noted that Regulation FD (fair disclosure) became effective on 10/23/2000 and may have altered analysts' relative performance in earnings forecasts compared with the period prior to it. For that reason, our study reported below is seen to have incorporated the impact of the new regulation and may provide insights into the factors that affect analysts' performance in the new regulatory environment.

²² From our analysis presented in Sect. 6, we recognize that the influences of the transient factors have a half-life span in most cases much less than 0.35 years. The variable SCORE-A as defined above is predominantly driven by the long-lasting factors.

We also noticed that high growth firms tend to attract more attention from major investors and prominent analysts. Closer scrutiny and wider dissemination of the firm's financial information in the market place may help analysts achieve better forecasting accuracy.

To prepare for the empirical investigation, the variables discussed above that characterize analysts and the firms that they follow are defined and explained in details below.

FN: an analyst's workload, proxied by the average number of firms followed by that analyst in the 3 year period. Specifically, it is the ratio of the total number of quarter-firm counts to the number of quarters for which the analyst provided earnings forecasts in the years 2004–2006.

ASIZE: the natural logarithm of the overall average of the capitalization sizes of the firms which were included in our data sample and followed by the analyst, computed at the beginning of each quarter, observed in the 3 year period.

SB: the size of the brokerage house which employed the analyst, proxied by the average number of analysts affiliated with the firm in the 3 year period, calculated at the beginning of each year.

AE: an analyst's length of experience prior to the beginning of year 2004, proxied by the number of quarters from the first recorded participating quarter to the last in our sample data over the period from the second quarter of 1984 through the fourth quarter of 2003, ignoring discontinuity.²³

FGR: the average annual compound growth rate prior to the end of year 2003 in the capitalization size of the firms that were included in our sample and were followed by the analyst in the 3 year period.

Among the five variables described above, FN, SB, and AE are analyst-related, while ASIZE and FGR are firm-related. By involving both sets of influencing factors in our analysis, we in effect postulate an expansion of the analytic framework in the extant literature. The analysis provides a broader perspective to factors that influence analysts' forecasting prowess.

We further note that, in order to be included in our sample for the calculation of the variables defined above, an analyst must fully participate in the forecasting activities. By "full participation" we mean that an analyst must provide at least one quarterly forecast for any firm in our sample in each of the 3 years under study. Analysts who do not meet the requirement are excluded from the sample.

Summary statistics of the variables listed above for our sample of analysts and related firms are provided in Table 6. There are 1,582 analysts meeting the full participation requirement and each carries a set of values for the variables SCORE-A,²⁴ FN, ASIZE, and SB. However, some of the analysts (171 in number) among the 1,582 do not have experience records prior to year 2004, in part due to our data trimming criteria as described in Sect. 2. Further, four analysts are involved with firms which do not have necessary capitalization numbers for the calculation of the variable FGR in our sample timeframe. This leads to the usable sample of analysts of $1,582 - 171 - 4 = 1,407$.

Results of regression analysis for the overall dataset and each of the thirteen industries for the sample period identified above are displayed on the left panel of Table 7. In this part of the table, we display each estimated regression coefficient with its corresponding

²³ We skipped the first quarter of 1984 because of the absence of some necessary records in the I/B/E/S database in that very first quarter.

²⁴ The other score variable SCORE-R is for a different ranking score system to be explained in the next section.

Table 6 Descriptive statistics of analysts' characteristics

Variable	<i>N</i>	Mean	SD	Min	Q1	Median	Q3	Max
SCORE-A	1582	9.005	3.495	3.000	6.000	9.000	12.000	15.000
SCORE-R	1582	9.012	2.683	3.000	7.000	9.000	11.000	15.000
FN	1582	19.910	13.172	1.500	9.077	18.056	28.167	89.390
ASIZE	1582	8.278	1.124	5.145	7.503	8.230	9.012	12.846
SB	1582	32.187	20.401	1.000	15.676	31.561	51.000	78.200
AE	1411	18.996	18.394	1.000	5.000	13.000	26.000	77.000
FGR	1578	0.011	0.034	-0.480	-0.003	0.010	0.027	0.142

The variable SCORE-A is the total AFE quintile rank score of analyst j and the variable SCORE-R is the total within-firm-rank quintile rank score of analyst j , for the 3 years, 2004–2006. The other variables are: (1) FN, the average number of firms followed by analyst j in the 3 year period, which is equal to the total number of firm-quarter counts divided by analyst j 's number of participating quarters in the 3 years; (2) ASIZE, natural logarithm of the grand average of capitalization sizes of all firms followed by analyst j observed in the 3 year period; (3) AE, analyst's length of experience as proxied by the number of quarters from the first recorded participating quarter to the last of analyst j , over the period from the second quarter of 1984 through the fourth quarter of 2003, ignoring discontinuity; (4) FGR, the average annual compound growth rate in the capitalization size of all of the firms followed by the analyst in question over the 3 year period; and (5) SB, the size of the brokerage house which employed analyst j , as proxied by the average number of analysts affiliated with the brokerage house in the 3 year period. In the table, N = number of analysts with valid data for the variable in question; analysts with zero AE or following a firm with only 1 year of record, which makes FGR not computable, are excluded from the sample

t -statistic embraced by the pair of parentheses immediately to the right. From the results for the overall dataset placed at the bottom row of the panel, we observe several significant phenomena. First, the heavier an analyst's workload, in terms of the number of firms followed by the analyst, the worse the analyst's performance, as suggested by the positive sign of the regression coefficient on FN. However, the corresponding t -ratio, 1.41, is not statistically significant at the 5% two-sided level.²⁵ Second, the longer an analyst's experience is as recorded in our sample data, the better the analyst performs as indicated by the negative sign of the regression coefficient on the variable AE. But its t -ratio value, -1.76 , is again not statistically significant at the 5% two-sided level. Third, the larger the size of the firm being followed, the more accurately an analyst can (on average) forecast its earnings, as is reflected by a very significant negative regression coefficient, with a t -ratio equal to -11.05 , on ASIZE. Fourth, the size of the brokerage firm that employed the analyst does not materially affect the analyst's relative performance as evidenced by the near zero t -ratio, 0.26, on the SB variable. And fifth, a firm with a higher average capitalization growth rate has a fairly significant favorable effect on analysts' relative forecast accuracy as suggested by a t -ratio of -2.52 on the variable FGR. In summary, we note that the first three observations are consistent (two in correct signs, though insignificantly) with the bulk of findings reported in the literature, while the fourth is not, and the fifth is new.²⁶

Observations of greater relevance are found for analysts within each industry, for each industry provides a more homogeneous forecasting environment. As shown by the thirteen

²⁵ Heretofore, our analysis is based on the 5% two-sided significance level on the t -statistic which is near the normal distribution given the large sample size as indicated on the column with the caption "N".

²⁶ The reader is referred to Clement (1999), Jacob et al. (1999), Mikhail et al. (1997, 2003), Ghosh and Whitecotton (1997), Abarbanell and Lehavy (2003), and Stickel (1992) for the findings related to FN, SB and AE. A lone finding about the effect of ASIZE is reported in Brown (1997). A similar finding in a foreign country is also documented by Ho (2004). The variable FGR is new in this study.

Table 7 Regression and discriminant analysis of factors influencing analysts' persistence in accuracy in thirteen industry groups and the overall dataset

	INT	FN	ASIZE	SB	AE	FGR	Pr > F	N	No. of best	No. of worst	Wilks' λ	<i>p</i> -value
<i>Industries</i>												
Petro	6.119 (0.34)	-0.017 (-0.79)	-0.140 (-0.38)	0.019 (0.11)	-0.021 (-1.10)	4.928 (0.30)	0.6583	97	7	18	0.8945	
Consum	27.364 (6.16)	0.103 (6.29)	-1.380 (-11.40)	-0.052 (-2.02)	0.007 (0.84)	-8.022 (-1.85)	<0.001	407	76	72	<0.001	
Basic	23.142 (3.87)	0.124 (6.82)	-1.177 (-7.44)	-0.027 (-0.70)	-0.021 (-2.18)	-4.823 (-0.88)	<0.001	263	40	51	<0.001	
Food	-6.678 (-0.19)	0.045 (0.94)	-0.749 (-1.52)	0.567 (1.16)	0.021 (0.94)	19.340 (0.58)	0.2745	49	10	6	0.2333	
Constr	24.973 (1.20)	0.058 (1.25)	-1.401 (-5.53)	0.293 (1.24)	-0.010 (-0.48)	-5.376 (-0.27)	<0.001	70	15	15	0.0002	
Financ	-2.752 (-0.22)	0.014 (0.87)	0.324 (1.79)	-0.090 (-2.00)	-0.001 (-0.08)	9.273 (0.77)	0.1830	235	43	43	0.4147	
Capita	15.189 (2.30)	0.032 (1.26)	-1.171 (-6.30)	0.045 (1.22)	-0.005 (-0.47)	2.833 (0.42)	<0.001	308	61	70	<0.001	
Transpo	40.011 (5.64)	0.051 (1.66)	-1.835 (-3.55)	0.327 (0.69)	-0.006 (-0.19)	-18.523 (-2.93)	0.0002	49	12	11	0.0332	
Utilit	17.434 (1.40)	0.075 (1.87)	-0.886 (-3.41)	-0.066 (-0.56)	-0.017 (-0.69)	-0.820 (-0.07)	0.0051	116	26	22	0.0420	
Textile	30.441 (5.51)	0.146 (7.96)	-2.092 (-11.23)	-0.184 (-2.21)	-0.008 (-0.73)	-5.818 (-1.04)	<0.001	164	42	30	<0.001	
Servic	23.734 (4.21)	0.088 (3.76)	-1.010 (-6.71)	-0.035 (-0.91)	-0.021 (-1.83)	-6.938 (-1.29)	<0.001	343	61	58	<0.001	
Leisur	15.726 (1.22)	-0.056 (-2.05)	-0.459 (-1.57)	0.017 (0.12)	0.002 (0.13)	-2.631 (-0.20)	0.0688	127	18	17	0.0213	
Other	59.718 (7.67)	0.277 (3.03)	-1.167 (-4.11)	0.178 (1.84)	0.010 (0.42)	-44.410 (-5.82)	<0.001	138	25	27	<0.001	
Overall	23.224 (8.81)	0.011 (1.41)	-0.933 (-11.05)	0.001 (0.26)	-0.009 (-1.76)	-6.503 (-2.52)	<0.001	1407	262	273	<0.001	

The dependent variable SCORE-A is the total AFE quintile rank score of analyst *j* for the 3 years, 2004–2006. The independent variables are: (1) FN, the average number of firms followed by analyst *j* in the 3 year period, which is equal to the total number of firm-quarter count divided by analyst *j*'s number of participating quarters in the 3 years; (2) ASIZE, natural logarithm of the grand average of capitalization sizes of all firms followed by analyst *j* observed in the 3 year period; (3) AE, analyst's length of experience as proxied by the number of quarters from the first recorded participating quarter to the last of analyst *j*, over the period from the second quarter of 1984 through the fourth quarter of 2003, ignoring discontinuity; (4) FGR, the average annual compound growth rate in the capitalization size of all of the firms followed by the analyst in question over the 3 year period; (5) SB, the size of the brokerage house which employed analyst *j*, as proxied by the average number of analysts affiliated with the brokerage house in the 3 year period. In the table, INT represents the estimated intercept term of the linear regression model. The value of the *t*-statistic for each estimated regression coefficient as displayed is placed in the parentheses to its right. The caption "Pr > F" represents the *p*-value of the *F* test statistic for the regression model in question. The caption *N* represents the number of observations (analysts) involved in the regression analysis in each of the cases indicated. The caption "N of Best (Worst)" represents the number of analysts in the best (worst) group of analysts, where the best group carries Score-A value equal to 3, or 4, or 5 and the worst group has Score-A value equal to 13, or 14, or 15. The caption "Wilks' λ (*p*-value)" represents the significance level of the discriminant function based on the five explanatory variables defined above. The industry grouping method employed here is from the paper by Breeden et al. (1989)

sets of regression coefficients displayed on the left panel of Table 7, the strength of each of the five explanatory variables does not apply evenly to all industries. The adverse effect of an analyst's workload, FN, is most pronounced for firms in consumer durables, basic industries, textiles, services, and other. The firm size, i.e., ASIZE, effect is most significant for firms in consumer, basic industries, construction, capital goods, transportation, utility, textiles, services and other. Further, the capitalization growth rate, FGR, shows significant effects on analysts' relative accuracy in earnings forecasts for firms in transportation and other. Interestingly, analyst's length of experience, AE, as in the case for the overall dataset, does not show significant effect on analyst's relative forecast accuracy in any of the thirteen industries.²⁷ The lack of beyond-border-line significant effect of the size of brokerage house, SB, is observed for each of the thirteen industries, as it is the case for the overall dataset. In summary, some of the markedly diverse observations on individual industries as we have noted above in this paragraph caution us against applying the findings from the overall dataset to each individual industry. Instead, they suggest that the effect of each of the analyst- or firm-related characteristics on analysts' relative accuracy in earnings forecasts varies considerably in different industries.

For completeness of our discussion, we provide in Table 7 a column with the caption "Pr > F" that lists the p -value from the F test for the effectiveness of the entire regression equation in each case under study. The entries in this column show that the five-variable regression equation possesses significant explanatory power, at the 5% significance level, in nine out of the thirteen industries. The four industries that defy explanation by the regression equation are petroleum, food, finance, and leisure. Incidentally, the column with caption "N" lists the number of observations (analysts) involved in the particular industry under study. We remind the reader that an analyst can be involved in more than one industry. The analyst's performance in each of the different industries is analyzed separately.

We further conducted a discriminant analysis to evaluate the ability of our analysis to identify superior or inferior analysts, measured by their relative average AFE, employing the five explanatory variables discussed in the immediately preceding three paragraphs.²⁸ We first identify two extreme groups of performers, in which the best performing group consists of analysts with the 3-year total quintile rank scores equal to 3, 4, or 5, while the worst group with scores 13, 14, or 15. The values of the five explanatory variables of the two groups of analysts were then used to construct and estimate a linear discriminant function. The calculated functional value of each analyst was then used as the marker of that analyst's ability in accurately forecasting earnings. An optimal dividing point of the functional value was decided and used to cast each analyst in the sample into either the "best" or the "worst" group. The effectiveness of the discriminant function in correctly identifying an analyst's group membership is captured by a test statistic, called "Wilks' λ (lambda)". A statistically significant Wilks' λ value means that the chance of correct classification of an analyst's group membership using the functional value exceeds that of random guessing supplemented by the knowledge of the relative size of the two groups.

²⁷ We note that the enactment of the Regulation FD starting in October 2000 may have reduced the importance of analyst's experience, which may in part proxy for an analyst's connection with corporate executives whose firms the analyst follows. The uncertain economy in the early 2000s in the aftermath of the burst of the high tech bubble in the stock market may also have affected the usefulness of analysts' past experience.

²⁸ The reader is referred to Johnson and Wichern (2002), Chap. 11, for detailed methodological explanations of this analysis.

On the right panel (the rightmost three columns) of Table 7, we display relevant analyst counts and the results from the discriminant analysis for each of the thirteen industries and for the overall dataset. The captions “*N* of Best” and “*N* of Worst” represent the number of observations (analysts) in the best and the worst groups, respectively, as previously defined. For instance, the results displayed on the bottom row of the table for the overall dataset show that there are 262 analysts in the “best” group and 273 analysts in the “worst” group. The value of the Wilks’ λ computed based on the covariance matrix of the five explanatory variables for the two groups in question has a *p*-value of less than 0.0001, indicating that the discriminant function works effectively in separating superior analysts from inferior analysts, much better than random guessing. Our analysis also reveals that the five explanatory variables provides a 72% correct classification for analysts who are in the ex post “best” group, and 55% correct classification for those in the ex post “worst” group (details not shown in Table 7). As the numbers suggest, the discriminant function does a better job in picking up good analysts than weeding out bad analysts.

The key test result of the discriminant analysis for each of the thirteen industries, the *p*-value of the Wilks’ λ , is displayed on the rightmost column in Table 7. The numbers reported therein show that the effectiveness of the discriminant function is statistically significant at the 5% level for ten out of the thirteen industries, closely matching the industries that reveal significant *F*-test statistics for the five-variable regression equation.

In summary, the findings reported in this section provide useful information about which analyst- and firm-related characteristics are significantly related to analysts’ forecast accuracy, in what fashion, to what extent, and in which industries.

8 An alternative scoring method and the empirical results

Our analysis thus far relies on the measure of annual AFE averaged over all of the firms followed by an analyst in a particular year. A potential shortcoming of that measure is its reliance on environmental factors that put analysts’ forecasting accuracy on uneven footing, such as the size and the complexity of the business, or businesses, of the firms being followed. As such, in order to assess the robustness of our results reported in the previous sections and, more importantly, to gain further insights into analysts’ persistence in relative accuracy in earnings forecasts, we employed an alternative measure that neutralizes all firm-related factors in the performance evaluation. The measure is to rank all qualified analyst’s AFEs *for a firm in a particular year* and convert the ranking into a sequence of fractional numbers equally spaced from zero to one.^{29,30} Specifically, we define

²⁹ A same measure is employed by Clarke and Subramanian (2006, p. 92, Eq. 11), in a study of herding behaviors among analysts in earnings forecasting.

³⁰ A different standardized measure of forecasting accuracy is proposed by Clement (1999) and used by Jacob et al. (1999), in which the absolute values of forecast errors of individual analysts for firm *k* in year *t* are normalized (i.e., divided) by the average absolute forecast error from all participating analysts in the same firm and year. This measure has the tendency to magnify the accuracy differences among analysts because a large proportion (about 50%) of quarterly earnings forecast errors in our sample are within a small range of ± 3 cents per common stock share. Further, more than 10% of all forecast errors in our sample are exactly zero, which renders the normalization infeasible in many cases. In summary, such a normalization process could produce serious distortions. Because of these concerns, we did not pursue a measure of accuracy along that line.

$${}_{t-1}R_{j,t}^k = \left({}_{t-1}I_{j,t}^k - 1 \right) / \left(N_t^k - 1 \right), \quad j = 1, 2, \dots, N_t^k$$

where N_t^k : number of analysts providing forecasts for the earnings of firm k in year t ; ${}_{t-1}I_{j,t}^k$: the rank of analyst j 's AFE, i.e., ${}_{t-1}AFE_{j,t}^k$, among all participating analysts' AFEs for firm k in year t , usually from 1 to N_t^k (except for ties, which will be treated in the way as explained below), with 1 being the best, and N_t^k being the worst.

In the calculation for ${}_{t-1}I_{j,t}^k$, in case of a tie, we use the average of the ranks which otherwise would be assigned to the observations. For instance, when there are three analysts providing identical forecasts for a firm for each of the four quarters in a year, the ${}_{t-1}I_{j,t}^k$ value assigned to each of the three analysts is 2, the average of the three rank numbers, 1, 2, and 3, which would otherwise be assigned to the analysts should their forecasts be all distinct. In the special case where only one qualified analyst provides forecasts for the firm in a year, the value of the only analyst's ${}_{t-1}R_{j,t}^k$ is set equal to 0.5. In the analysis of an analyst's performance in a year, we average the R -scores across all of the firms that the analyst follows in that year. In short, in the new analysis we substitute the average AFE value in the analysis presented in the previous sections by the average R -score.

An advantage of this new measure of average R -score is that it neutralizes all firm-specific factors, which tend to vary from firm to firm, and instead focuses on analysts' relative intrinsic abilities within the setting of a single firm. Potential significant drawbacks in this new measure exist too. Firstly, the dispersion of the R -scores for a firm depends on the number of participating analysts and is somewhat oblivious to the practical significance of the magnitudes of earnings forecast errors. For instance, when there are only two participating analysts in a firm, the spread of the R -scores between the two analysts is automatically 1, regardless of how trivial the discrepancy between the two forecast numbers is. On the other hand, a far-off-the-mark forecast by an analyst in a group of, say, 20 analysts will not result in severe penalty against the analyst, for the worst R -score the analyst will get is 1. Secondly, the measure based on the R -score tends to oversimplify and artificially space the difference in relative accuracy by a multiple of $1/(N_t^k - 1)$.

However, we recognize that at present time a completely satisfactory measure for analysts' relative forecast accuracy is yet to be developed. Keeping this in mind, we present below results based on the R -score using the same dataset and methodology as reported in previous sections.

The frequency distribution of the number of qualified analysts who ever participate in a firm in our sample was collected. (For economy of space, it is not shown here.) From the data, we observed that about 30% of firms in our sample have three or fewer qualified participating analysts.³¹ On the other hand, about 10% of sample firms have 27 or more distinct participating analysts over the sample years. We also note that the average number of analysts per firm is 10.7, over a total of 2,934 firms.

We replicated the analysis as presented in Table 3 for the overall dataset using the new measure and observed that the newly compiled one-step (1-year) transition probability matrix carries patterns remarkably similar to that in Table 3, albeit much weaker in the

³¹ In Sect. 2 we state that only firms with four or more participating analysts in a quarter immediately prior to the quarter in question, as listed in the original I/B/E/S database, are included in our sample for that quarter. But other trimming criteria, such as that an analyst will have to provide quarterly forecasts for at least ten quarters for the firm, further reduced the number of qualified analysts in our sample. This explains why we have a substantial number of firms with three or fewer qualified analysts.

Table 8 Tests of conformity of observed frequency distributions of quintile ranks to the pure chance transition model (based on the R -score)

Industry	Petro	Consum	Basic	Food	Constr	Financ	Capita
Chi-square two-way contingency-table test values and the corresponding p -values for the test against $H_0^{(1st)}$							
Chi-square	32.2	115.6	74.1	19.2	20.8	60.6	66.6
(p -value)	0.04119	0.0000	0.0000	0.51101	0.40812	0.00001	0.0000
Industry	Transpo	Utilit	Textile	Servic	Leisur	Other	All
Chi-square	26.4	38.2	48.2	80.3	32.7	30.0	258.3
(p -value)	0.15352	0.00845	0.00040	0.0000	0.03656	0.07032	0.0000

Based on the R -score for within-firm-year-rank as defined in the text, the entries above provide chi-square test results from testing the null hypothesis $H_0^{(1st)}$, that is, the pure chance transition model. The intermediate steps of the test procedure have been illustrated in Table 3, Panel A. The industry grouping method used here is from Breeden et al. (1989)

strength of persistence. The chi-square test statistic against the pure chance hypothesis, equal to 258.3, is still exceedingly significant. However, the new results on testing $H_0^{(2nd)}$, that is, the one-step Markov chain model, comparable to those in Panel D of Table 3, show few signs that suggest the overall dataset to carry the second degree Markov persistence in analysts' relative forecast accuracy, as it is measured by the R -score.

We then redid Table 4 based on the R -score for the analysis of individual industries and present the results in Table 8. The entries therein (equivalent to Panel A in Table 4) reveal that for nine out of the thirteen industries at the 5% significance level, analysts' relative forecasting accuracy maintains the property of "first degree Markov persistence" in the petroleum, consumer durables, basic, finance, capital goods, utilities, textiles, services, and leisure industries. However, the χ^2 test results comparable to those in Panels B and C of Table 4 reveal that the "second degree Markov persistence" is no longer present in any one of the thirteen industries. This may not be surprising, because the new measure, as we have pointed out earlier, tends to be overly simplistic and carries certain distortive effects. Thanks to these drawbacks, detections of analysts' performance differentials based on the average R -score have been made more difficult.

In short, in our view, the fact that analysts' relative accuracy performance sustains the first degree Markov persistence in nine out of the thirteen industries based on the R -score measure is rather remarkable.

9 Summary and concluding remarks

In this paper we present a systematic analysis using a Markov chain model that casts new light on the nature, stochastic properties, and the level of persistence in analysts' relative accuracy in earnings forecasts. The concepts of "first degree Markov persistence" and "second degree Markov persistence" are formulated to distinguish between two levels of performance persistence. The two levels of persistence are then linked to the two-component Markov chain model, which identifies the proportion and the strength of each of the two classes of "transient" and "long-lasting" factors that influence the superiority/inferiority of analysts' performance. We also define and use the parameters β_1 and β_2 to characterize the convergence rate of an analyst's average quintile rank—from an initial

quintile rank toward the median rank—in successive observation years. These two characterization parameters are then used to study the stationarity of the transitional probabilities of quintile ranks. They are also used to examine the existence of the second degree Markov persistence in analysts' relative forecast accuracy.

We find from our empirical results that analysts' workload and the size and growth rate of the firms followed are among the long-lasting influencing factors. We also find that the strength of each of these long-lasting factors varies markedly from one industry to another. Overall, the results provided in this paper demonstrate that the spread and persistence of performance differentials among analysts is remarkably strong and deep-rooted in various industries. The results are also fairly robust to a more stringent (perhaps overly-rigid) measure, termed *R*-score, formulated to neutralize all firm-related effects.

In the context of the analysis presented in this paper, several issues warrant further attention and research.

First, in practice, there is no unified timing to submit earnings forecasts. It varies widely across analysts, quarters, and firms. A strict equal-based comparison of analysts' accuracy is infeasible. A mechanism is needed to properly adjust accuracy for the timing of forecasting. Such a mechanism might hopefully improve the reliability of the analysis of relative forecasting accuracy.

Second, there are considerable turnovers of analysts in our data sample. The number of missing data points for the fifth observation year is highest for the initial top and bottom quintile groups. There are needs to track the paths of departure (promotions, dismisses, etc.) of the missing analysts. The information may cast more light on the ultimate intrinsic ability of an analyst and the consequences of being a superior or inferior analyst.

Third, there is a need to understand analysts' incentives, or lack thereof, in achieving higher accuracy in earnings forecasts. Many other factors (besides the obvious ones explored in this paper) may explain why analysts' performance tends to persist in a competitive market of investment information services. In an ideal market, competition should quickly weed out inferior analysts, if the sole criterion of job retention is accuracy in forecasting. In the same environment, superior analysts would also lose their edge quickly. But of course the reality is much more complex. What other factors are at work and how do they work in reality? An outline of discussions on this subject is available in Ramnath et al. (2008).

Despite the potential issues inherent in our analysis, this paper presents for the first time in the literature a broad-based analysis demonstrating that analysts' abilities in earnings forecasts are indeed heterogeneous and that differentials persist over time. We also show that the strength of persistence and the potential influencing factors appear to vary from industry to industry. This provides a concrete base for future research work in the following directions.

First, accuracy of composite earnings forecasts can be improved by placing more weights on forecasts made by identified superior analysts, in contrast to the conventional equal-weighted consensus forecasts. Optimization algorithms can be designed for the tasks at hand.

Second, information gathering and processing in different industries conceivably requires different sets of skills and analytical capacities. The degree of difficulty and the intensity of work in earnings forecasting may thus vary across industries. How one would characterize and analyze these phenomena and how the related factors could affect information efficiency in the overall financial system are subjects of future research interest.

Acknowledgments The authors gratefully acknowledge receipt of research supports from Lubar School of Business at the University of Wisconsin-Milwaukee and Department of Business at Missouri Western State University. These supports include access to the databases used in this paper.

References

- Abarbanell JS, Bernard VL (1992) Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *J Finance* 47:1181–1207
- Abarbanell JS, Lehavy R (2003) Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts. *J Acc Econ* 36:105–146
- Bao D, Chien C, Lee C (1997) Characteristics of earnings-leading versus price-leading firms. *Rev Quant Financ Acc* 8:229–244
- Beneish M, Harvey C (1998) Measurement error and nonlinearity in the earnings-returns relation. *Rev Quant Financ Acc* 11:219–247
- Breeden DT, Gibbons MR, Litzenberger RH (1989) Empirical tests of the consumption-oriented CAPM. *J Finance* 44:231–262
- Brown LD (1997) Analyst forecasting errors: additional evidence. *Financ Anal J* 53:81–88
- Brown LD, Rozeff MS (1980) Analysts can forecast accurately! *J Portfolio Manag* 6(3):31–34
- Butler KC, Lang LHP (1991) The forecast accuracy of individual analysts: evidence of systematic optimism and pessimism. *J Acc Res* 29:150–156
- Carhart M (1997) On persistence in mutual fund performance. *J Finance* 52:57–82
- Clarke J, Subramanian A (2006) Dynamic forecasting behavior by analysts: theory and evidence. *J Financ Econ* 80:81–113
- Clement MB (1999) Analyst forecast accuracy: do ability, resources, and portfolio complexity matter? *J Acc Econ* 27:285–303
- Cooper RA, Day TE, Lewis CM (2001) Following the leader: a study of individual analysts' earnings forecasts. *J Financ Econ* 61:383–416
- Copeland T, Dolgoff A, Moel A (2004) The role of expectations in explaining the cross-section of stock returns. *Rev Acc Stud* 9:149–188
- Diether KB, Malloy CJ, Scherbina A (2002) Differences of opinion and the cross section of stock returns. *J Finance* 57:2113–2141
- Ghosh D, Whitecotton SM (1997) Some determinants of analysts' forecast accuracy. *Behav Res Acc* 9:50–68
- Grimmett GR, Stirzaker DR (2001) Probability and random processes, 3rd edn. Oxford University Press, Oxford
- Hanke JE, Wichern DW (2009) Business forecasting, 9th edn. Prentice Hall, Upper Saddle River, NJ
- Hilary G, Menzly L (2006) Does past success lead analysts to become overconfident? *Manag Sci* 52:489–500
- Ho L (2004) Analysts' forecasts of Taiwanese firms' earnings: some empirical evidence. *Rev Pacific Basin Financ Mark Policies (RPBFMP)* 7:571–597
- Jacob J, Lys TZ, Neale MA (1999) Expertise in forecasting performance of security analysts. *J Acc Econ* 28:51–82
- Johnson RA, Wichern DW (2002) Applied multivariate statistical analysis, 5th edn. Prentice Hall, Upper Saddle River, NJ
- Lim T (2001) Rationality and analysts' forecast bias. *J Finance* 56:369–385
- Loh R, Mian G (2006) Do accurate earnings forecasts facilitate superior investment recommendations? *J Financ Econ* 80:455–483
- Lou Z, Wahba G (1997) Hybrid adaptive splines. *J Am Stat Asso* 92:107–116
- Mikhail MB, Walther BR, Willis RH (1997) Do security analysts improve their performance with experience? *J Acc Res* 35:131–166
- Mikhail MB, Walther BR, Willis RH (2003) The effect of experience on security analyst underreaction. *J Acc Econ* 35:101–116
- O'Brien P (1987) Individual forecasting ability. *Manag Finance* 13:3–9
- O'Brien P (1990) Forecast accuracy of individual analysts in nine industries. *J Acc Res* 28:286–304
- Ramnath S, Rock S, Shane P (2008) The financial analyst forecasting literature: a taxonomy with suggestions for further research. *Int J Forecast* 24(1):34–75

-
- Richards RM (1976) Analysts' performance and the accuracy of corporate earnings forecasts. *J Bus* 49:350–357
- Sinha P, Brown LD, Das S (1997) A re-examination of financial analysts' differential earnings forecast accuracy. *Contemp Acc Res* 14:1–42
- Stickel SE (1992) Reputation and performance among security analysts. *J Finance* 47:1811–1836