Dynamics of analyst forecasts and emergence of complexity: Role of information disparity

Chansoo Kim[1,2]☯, Daniel S. Kim[3]☯, Kwangwon Ahn[3,4]*, M. Y. Choi[5]*

1 Computational Economics Laboratory, Computational Science Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea, **2** Department of Economics, Seoul National University, Seoul 08826, Republic of Korea, **3** HSBC Business School, Peking University, University Town, Nanshan District, Shenzhen 518055, P.R. China, **4** Moon Soul Graduate School of Future Strategy, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea, **5** Department of Physics and Astronomy and Center for Theoretical Physics, Seoul National University, Seoul 08826, Republic of Korea

☯ These authors contributed equally to this work.
* k.ahn@kaist.ac.kr (KA); mychoi@snu.ac.kr (MYC)

Abstract

We report complex phenomena arising among financial analysts, who gather information and generate investment advice, and elucidate them with the help of a theoretical model. Understanding how analysts form their forecasts is important in better understanding the financial market. Carrying out big-data analysis of the analyst forecast data from I/B/E/S for nearly thirty years, we find skew distributions as evidence for emergence of complexity, and show how information asymmetry or disparity affects financial analysts' forming their forecasts. Here regulations, information dissemination throughout a fiscal year, and interactions among financial analysts are regarded as the proxy for a lower level of information disparity. It is found that financial analysts with better access to information display contrasting behaviors: a few analysts become bolder and issue forecasts independent of other forecasts while the majority of analysts issue more accurate forecasts and flock to each other. Main body of our sample of optimistic forecasts fits a log-normal distribution, with the tail displaying a power law. Based on the Yule process, we propose a model for the dynamics of issuing forecasts, incorporating interactions between analysts. Explaining nicely empirical data on analyst forecasts, this provides an appealing instance of understanding social phenomena in the perspective of complex systems.

Introduction

The twentieth century witnessed exponential growth in equity capital markets as more countries opened their securities exchanges and as the barrier to accessing capital markets became lower for the average household. In a rapidly growing capital market, one of the important players is the sell-side analyst, who provides valuable information on the issuers of securities, i.e., the firms. Sell-side analysts issue regular reports that summarize the future outlook of the business of a firm and forecast the performance, which includes forecasts on earnings per share (EPS), revenue, operating income, cash flow, and other accounting measures for both

quarterly and annual performance. Analysts use financial statements, management communication [1], and detailed voluntary disclosure from the firms [2] as their information source for producing forecasts. Analyst forecasts on the accounting performance of firms are heavily followed and monitored by both company management and market participants, because accounting performance is the basis for forming expectations for future stock prices [3]. Due to their importance to the financial markets, analyst forecasts have long been of interest to scholars in finance and accounting.

So far, the vast majority of studies on analyst forecasts have focused on analyst consensus, which is either the mean or the median of non-stale analyst forecasts, i.e., existing but not too old (usually issued within last 90 days) forecasts [4–11]. Some other studies have focused on the volatility or dispersion of analyst forecasts [12–16]. It is known that analyst forecasts tend to be optimistic in the beginning and become lower towards the year-end [17].

There has been little interest in the outliers, or the tail distribution of analyst forecasts. While the location of the distribution is undoubtedly meaningful for understanding the views of sell-side analysts, the tails of the distribution also signal important messages. Regarding the analyst forecasts that deviate from the mass, Hong and collaborators [18, 19] found that inexperienced analysts are less likely to deviate from consensus with optimistic views relative to their more reputable counterparts. Put differently, this implies that if there is a positive outlier in analyst consensus, it may come from more reputable analysts. They argue that inexperienced or less reputable analysts are more liable to face problems with their jobs when their bold (more extreme) forecasts turn out to be wrong. In other words, unknown analysts are likely to be fired if they make forecasts that are different from others and turn out to be wrong. Recently, Du, Yu, and Yu [20] studied the most extreme outlier of the forecasts during the year, to find that the outlier triggers future earnings management by the company management.

In this work, we ask if complexity emerges in the collective behavior of analyst forecasts, particularly, among outliers of optimistic analyst forecasts. *Complexity* in general stands for large variability of the macroscopic behavior displayed by a many-particle system, consisting of many elements [21]. In such a system, appropriate interactions between elements can give rise to characteristic collective properties of the whole system, which may not be reduced directly to the properties of individual constituents. The appearance of such macroscopic behavior is thus called *emergence*. Even if (microscopic) interactions between elements are rather simple, the emergent (macroscopic) behavior can be very rich and diverse [22]. In our system, consisting of analysts as its elements, herding behavior emerges, as is evident from their forecasts. While actions of analysts do not necessarily delineate complexity, the whole system nevertheless exhibits such herding behavior and associated complexity. To investigate how forecasts form complex herding, we treat the forecasts as the product of small action rules (micromotives) of individual analysts and regard complexity as macrobehavior [23, 24].

We focus on optimistic forecasts because analyst optimism is a well-documented practice, and analysts are known to walk down their forecasts towards the end of a fiscal year [17]. Note that complexity is different from "analyst herding" reported in traditional accounting and finance literature [25]: while the latter refers to the tendency of analysts to issue forecasts that are simply closer to the mean of existing forecasts, complexity may arise from the tendency to issue forecasts distributed not only in the main body but also in the tail of optimistic forecasts.

Our study attempts to answer the following questions: Does complexity emerge in forecasts by major analysts (those issuing forecasts in the main body) or extreme analysts (those issuing forecasts in the tail of the forecast distribution); how are the forecasts tied to each other; what

are the effects of the business cycle on analysts' forecasts; does the emergent complexity change over time; what are the underlying factors for the changes; and is there any relation between the behavior of all analysts and that of extreme analysts? In this study, we employ complex system analysis to address these questions. Recently, more economists attempted to understand social phenomena by applying methods of statistical physics used for describing complexity [26–28]. Further, recent studies introduced statistical methodology to better explain the tail behavior of financial markets [29]. These methods allow us to characterize economic systems and help us better understand their tail behavior.

To understand the behavior of analyst forecasts, we here consider the Yule process, described by a master equation. This turns out to allow for the emergence of skew distributions, including power-law and log-normal distributions. Power-law behavior, a prominent statistical signature for complexity [30, 31], has been recognized also in complex economic systems. The power-law exponent is found to be an important and even universal factor in characterizing financial markets, such as stock markets and foreign exchange markets [29, 32–35]. Increasing awareness of the power law has recently spurred attempts to explain emergent characteristics of various economic and social phenomena of complexity [36]. We thus probe the emergence of power-law as well as log-normal distributions, which serves as a key evidence of complexity, in analyst forecasts on EPS of publicly traded US firms.

In the following sections, details of the empirical evidence, which is based on the big-data analysis, are presented. We then provide a theoretical framework, incorporating interactions between analysts. The proposed model is shown to explain well the complexity observed in the empirical data on analyst forecasts. Finally, a summary with discussion is given.

Empirical evidence

In this section we present our empirical results based on the analyst forecast error, which measures the difference between the analyst forecast on EPS and the actual EPS outcome. We scale the difference by the stock price, to normalize the difference across sample firms.

Analyst forecast data are obtained from the Institutional Brokers' Estimate System (I/B/E/S), which reports the EPS forecast value (VALUE) z, date (ANNDATS), and time (ANNTIMS) of the forecast, along with the brokerage house (ESTIMATOR) and analyst (ANALYS). I/B/E/S also reports the actual EPS performance (ACTUAL) a of the fiscal year. Our main variable of analysis is the analyst forecast error

$$x \equiv \frac{z - a}{S},$$

(1)

which is scaled by the stock price S on the previous trading day. It turns out that a vast majority of forecast values are greater than the actual EPS performance, reflecting the optimistic views of most analysts. Accordingly, in this study, we take only observations with positive forecast errors, to focus on optimistic forecasts, and perform big-data analysis on the I/B/E/S data spanning nearly thirty years.

Our analysis manifests that the analyst forecast errors display two distinct distributional patterns: the main body and the tail. In the main body part, defined as forecast errors greater than zero but smaller than the 95th percentile value, forecast errors follow a log-normal distribution; in the tail part, defined as those larger than the 95th percentile, forecast errors follow a power-law distribution, where collective behavior among extreme forecast errors is expected [37]. For the tail, we use precisely 95.0th to 99.7th percentiles of observations to account for data errors as well as penny stock cases after the 99.8th percentile.

The distribution of forecast errors in the main body is found to be skewed, and turns out to fit the log-normal distribution, specified by the probability density function (PDF):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]. \tag{2}$$

In fitting our sample, we first construct histograms with optimal binwidth for the forecasts errors in the main body. We then use the maximum likelihood estimation (MLE) method, to estimate the mean and deviation of the log-normal distribution within the 95% confidence interval. Fig 1 illustrates the resulting log-normal fit of our sample for calendar year 1993 with mean $\mu = -5.46$ and deviation $\sigma = 1.94$. The forecast errors for other years in our sample, left out for brevity, also fit log-normal distribution very closely.

To disclose the detailed behavior of the tail, we consider the cumulative distribution function (CDF) $F(x)$, which is given by the integral of the PDF from $-\infty$ to x. The power-law behavior described by the CDF [29, 37]

$$1 - F(x) \sim x^{-\alpha}, \tag{3}$$

corresponds to the PDF $f(x) = dF(x)/dx \sim x^{-(1+\alpha)}$.
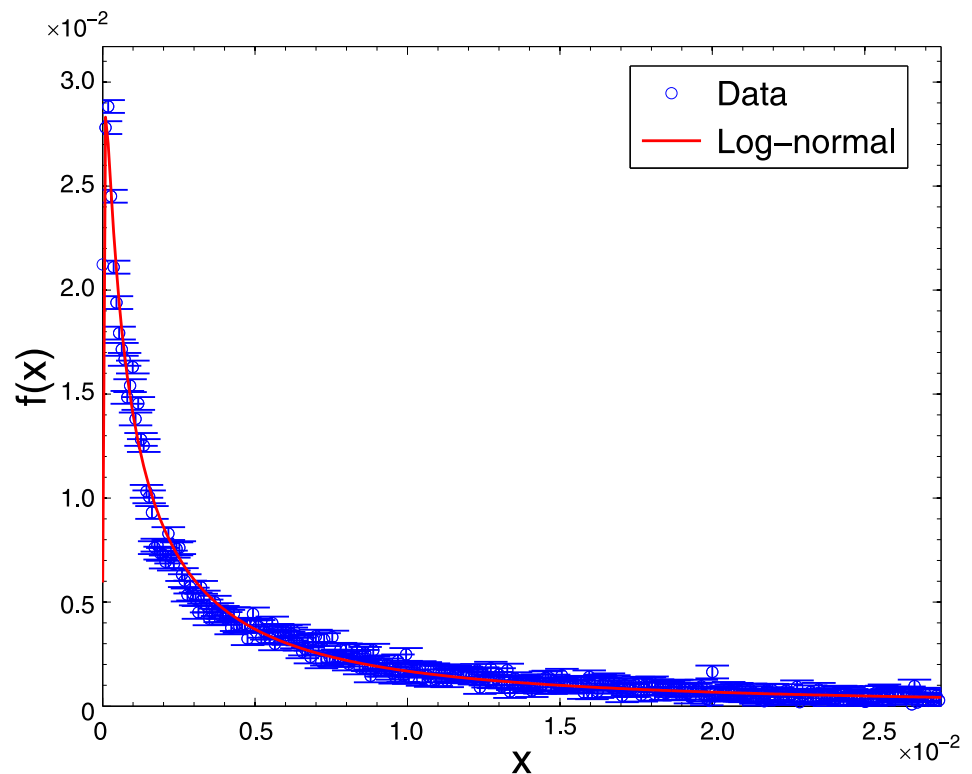
Fig 1. Log-normal fit of the main body part of forecast errors. We fit the main body part of analyst forecast errors to the log-normal distribution. (The figure is based on non-stale analyst forecast errors on the last trading day of 1993; main body parts from other calendar years fit log-normal distributions as well.) We define the main body part as the bottom 95 percent of analyst forecast errors (129202 observations in 1993). The red line is the fitted log-normal density while blue circles represent the data points. Error bars have been estimated from the deviation of the data. The fitted parameter values are: location parameter $\mu = -5.46$ and scale parameter $\sigma = 1.94$.
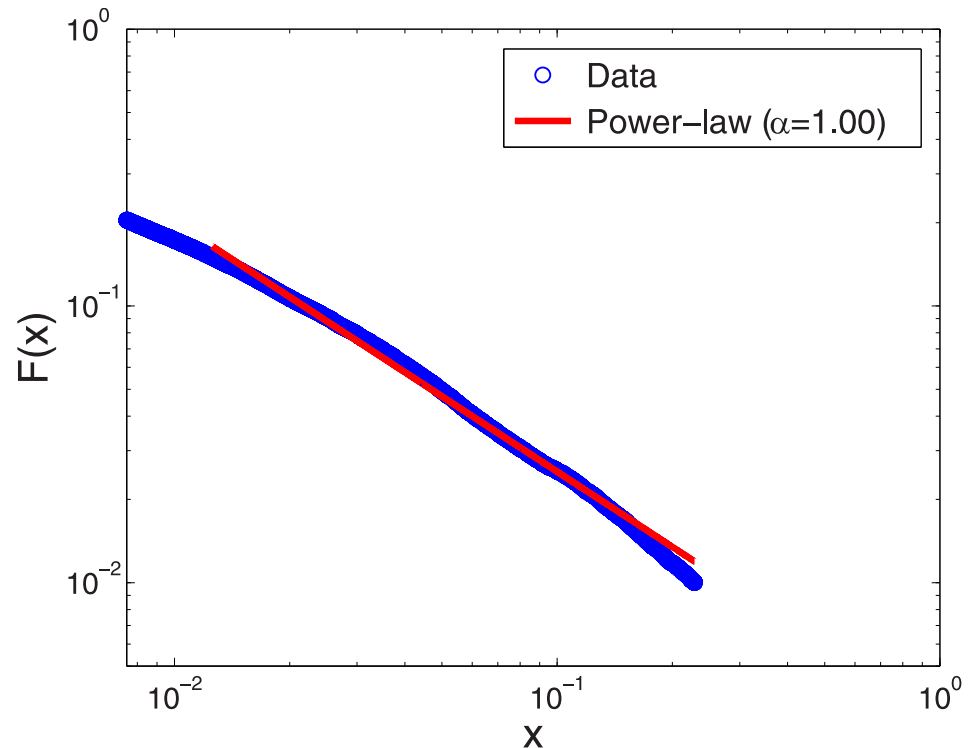
Fig 2. Power-law fit of the tail of forecast errors. We fit the tail part of analyst forecast errors to the power-law distribution. (The figure is based on non-stale analyst forecast errors on the last trading day of 1993; tail parts from other calendar years fit power-law distributions as well). Since the power-law distribution in Eq (3) implies a linear relationship in the logarithmic scale, we use ordinary least squares (OLS) to fit the empirical cumulative distribution. The red line depicts the fitted power-law distribution while blue circles represent the data points. In 1993, the power-law exponent approximates unity ($\alpha \approx 1$), which indicates that analyst forecast errors follow Zipf's law.

https://doi.org/10.1371/journal.pone.0177071.g002

Since Eq (3) exhibits a linear relationship in the logarithmic scale, we use the ordinary least squares (OLS) method to fit the power-law PDF $f(x)$. Constructing an empirical cumulative distribution from the data on forecast errors in the tail, we use OLS regression to estimate the exponent α of the power-law distribution. It minimizes the errors between $F(x)$ and the empirical distribution. From Fig 2, we find that the tail of analyst forecast errors indeed fits the power-law distribution, implying collective behavior among forecast errors. In year 1993, with 16508 forecast errors in the tail, the estimated power-law exponent is $\alpha = 1.00 \pm 0.01$.

The power-law behavior of forecast errors suggests that analyst forecasts are based on the "principle of least effort" [38, 39]. Specifically, the power-law exponent unity, following Zipf's law, implies that information seeking analysts tend to use the most convenient search method and such extremely optimistic analysts behave collectively. In Theoretical Framework section, emergence of such skew distributions as log-normal and power-law distributions is explained by means of controlled growth processes [40].

Yearly pattern in extreme forecast errors

For each year in our sample period, 1984 to 2012, we estimate the power-law exponent α for analyst forecast errors and obtain the mean value 1.00 with standard deviation 0.15. From $\alpha \approx 1$, we recognize that the tails of the analyst forecast errors follow Zipf's law [38] in the whole (sample) period, as presented in Fig 3.
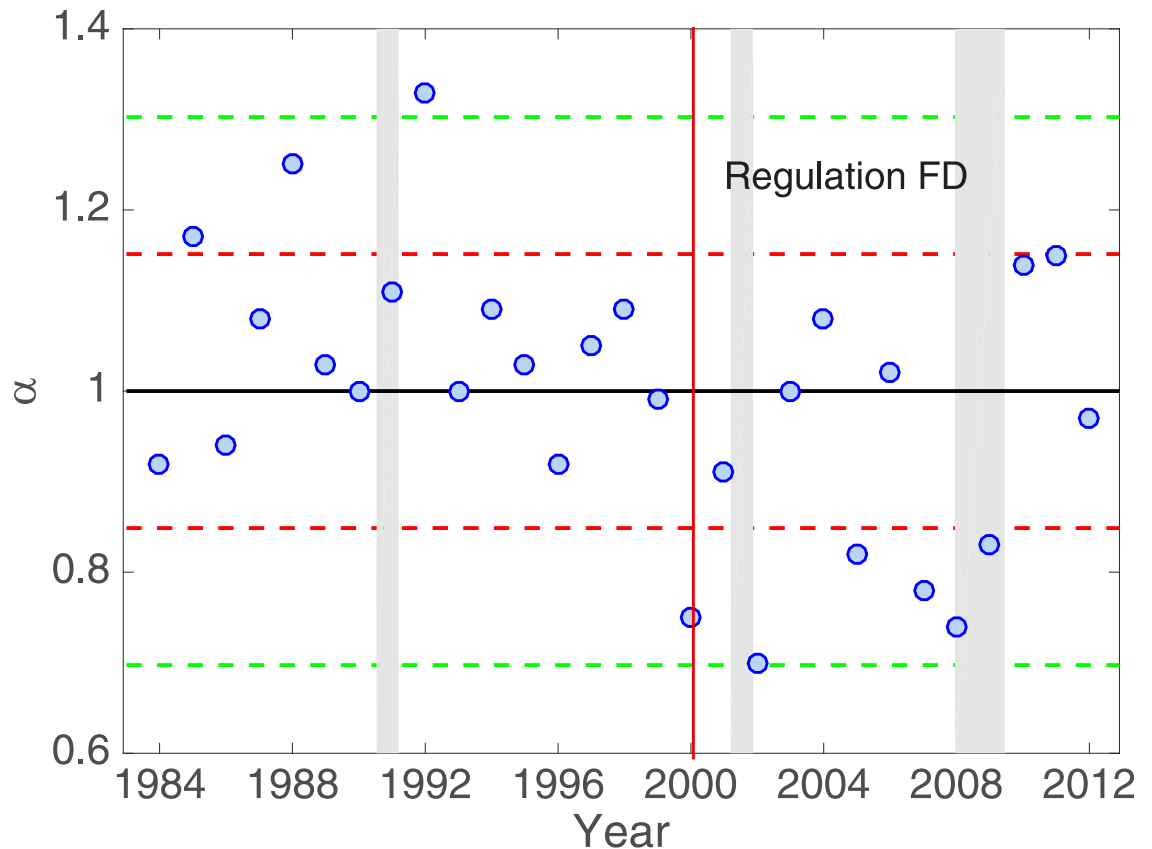
Fig 3. Annual values of the power-law exponent. Blue dots represent annual estimates of the power-law exponent α from year 1984 to 2012. The annual value of the exponent is estimated by fitting all forecast errors in each calendar year. The black horizontal line represents the power-law exponent of unity ($\alpha = 1$), which corresponds to Zipf's law, while two red/green dashed lines indicate one/two standard deviations away from Zipf's law. The vertical line designates year 2000, the year of "Regulation Fair Disclosure (FD)", while the shaded periods correspond to NBER recessions.

It is of interest that while most years have α close to unity, in certain years α deviates from unity by more than one standard deviation. It takes values lower by more than one standard deviation in years 2000, 2002, 2005, 2007, 2008, and 2009. The lower the value of α is, the thicker the tail distribution is. This implies that collective behavior is weaker in the above mentioned years. In contrast, collective behavior is stronger in years with α higher than unity by more than one standard deviation: 1985, 1988, and 1992.

Comparing Fig 3 with recession data from National Bureau of Economic Research (NBER), we notice that the years when α deviates by more than one standard deviation overlap largely with the recession period in the U.S. This indicates that analysts, facing economic recessions, tend to either follow extreme opinions (higher α) or reject them (lower α).

It is also broadly accepted that information asymmetry or disparity is more severe during a recession as the prospects of the economy are more uncertain, and as a consequence, analyst forecasts are more dispersed than in other periods [41, 42]. Information disparity in the market is higher just before an economic recession, as it is in general more difficult to forecast the future before huge changes take place. Thus the exponent α deviating from unity by more than one standard deviation is consistent with our intuition that extreme analyst forecasts should behave differently during a recession compared with a normal period. The relationship

between the recession and the S&P500 index (https://research.stlouisfed.org/fred2/series/SP500) indeed confirms this intuition [43, 44].

Yet, it is not straightforward to explain the discrepancy between higher and lower values of α in different recession periods. The year separating the two cases provides us with an important clue. Those values deviating largely from unity display sharp contrast in the general trend before and after 2000: They are higher than the normal values before 2000, but are lower during and after 2000. We propose that this is the evidence of a structural break in information disparity in the U.S. markets. Specifically, in October 2000, the Securities Exchange Commission (SEC) passed "Regulation Fair Disclosure (FD)", which requires that all information be fairly disseminated to the market [45]. This regulation is intended to reduce information disparity in the financial market by allowing equal access to firm information for all market participants. Prior to Regulation FD, companies could disclose firm information selectively to only some of the analysts and participants in the market; the regulation prohibits such selective disclosure practice. While there are arguments about the actual outcome of the regulation, some researchers do find improvements in the information environment after the regulation [46, 47].

When there is higher information disparity, namely, when only some analysts can access information, it is plausible to expect that those without the information will follow the opinions of those with the information. This mimicking behavior will cause some analysts to issue forecasts that are closer to extreme forecasts of other sell-side analysts, usually more reputable ones, giving rise to collective behavior among the outliers of analyst forecasts. Such collective behavior resulted in higher values of α during recession periods in the years before Regulation FD. However, when this information disparity weakens in the market, all financial analysts share the same set of information and there is no need for the sell-side analysts to follow extreme analyst forecasts. This results in lower values in the years of post-Regulation FD. It was also claimed that the positive effects of Regulation FD are more pronounced for firms that suffer from more information disparity (illiquid firms) [46], which is consistent with our findings. We thus conclude that Regulation FD made the information environment more effective during recessions.

We now test formally our conjecture from Fig 3. We choose the analyst forecast dispersion, defined as the standard deviation of analyst forecasts, as the proxy for information disparity. Unlike existing studies adopting the analyst forecast dispersion as the proxy for information asymmetry at the firm level, we consider the information disparity of the whole market. As such, we develop a market-wide analyst forecast dispersion for each sample year. We first scale the forecast dispersion by analyst consensus (mean of analyst forecasts) for each firm in our sample. The market-wide forecast dispersion is then given by the means of the scaled forecast dispersions of all companies in our sample. We also adopt the forecaster bias [48] as the proxy for economic uncertainty, which is in literature found high during the periods of economic downturn. We use the economic uncertainty as the continuous alternative to the NBER recession indicator; however, results using the NBER recession indicator are qualitatively similar. Table 1 presents regression results of information disparity and the power-law exponent.

We first test whether information disparity in the stock market measured by the dispersion in analyst forecasts is statistically associated with economic uncertainty and whether the association has changed significantly since the Regulation FD. Our empirical models describe how information disparity I is related to economic uncertainty E:

$$I = c_1 + \beta_1 E + \varepsilon_1 \tag{4}$$

Table 1. Regressions of information disparity and the power-law exponent. Models (1) and (2) test the association of information disparity and economic uncertainty; Models (3) and (4) test how the power-law exponent α deviates from unity due to information disparity. Models (5) and (6) test how collective behavior changed after the Regulation FD was introduced. R takes the value unity for years after Regulation FD was introduced and zero otherwise, I is the proxy for information disparity measured by the scaled forecast dispersion of the market, E is the economic uncertainty measure from Ref. [48] and c is constant. The numbers in parentheses are t-statistics calculated from heteroskedasticity and autocorrelation consistent standard errors according to Newey and West [49]. ** and *** represent significance at 5% and 1% levels, respectively.

Model	(1)	(2)	(3)	(4)	(5)	(6)
	I	I	$\lvert\alpha-1\rvert$	$(\alpha-1)^2$	α	α
R		−0.148*** (−8.62)			−0.147*** (−5.40)	0.00601 (0.09)
I			0.841*** (3.29)	0.196*** (2.45)		0.832*** (2.64)
E	0.0438** (2.26)	0.0544*** (4.62)				
$E \cdot R$		0.402*** (9.73)				
$R \cdot I$						−1.499*** (−3.00)
c	0.0800*** (6.25)	0.0653*** (9.28)	0.0294 (1.30)	0.00205 (0.28)	1.062*** (89.50)	0.981*** (30.83)
Observations	29	29	29	29	29	29
Adjusted r^2	0.077	0.394	0.025	0.002	0.215	0.173

https://doi.org/10.1371/journal.pone.0177071.t001

and

$$I = c_2 + \beta_2 R + \beta_3 E + \beta_4 R \cdot E + \varepsilon_2 \qquad (5)$$

with appropriate constants c_i and β_i, where R takes into account Regulation FD ($R = 0$ and 1 for calendar years before and after Regulation FD is introduced, respectively), $R \cdot E$ represents the interaction of R and E, and ε_i denotes the error term.

In Model (1), the analyst forecast dispersion is significantly positively associated with economic uncertainty, supporting our argument that a higher level of economic uncertainty leads to a higher level of information disparity in the market. Further, the significantly negative loading on Regulation FD in Model (2) indicates that information disparity in the stock market has generally decreased since Regulation FD put in place. Economic uncertainty still seems to positively influence information disparity in general. Interestingly, the loading on the interaction term of economic uncertainty and regulation is significant and positive. This indicates that economic uncertainty is more closely linked to information disparity in the stock market after the introduction of Regulation FD. This is rather intuitive when we understand the potential sources of information disparity. Specifically, information disparity is caused by at least two factors: 1) imperfect information disclosure and 2) uncertainty in the market. The Regulation FD was introduced to reduce information disparity in the stock market by forcing companies to share information equally among market participants. Thus the Regulation FD effectively eliminates the first source of information disparity, making uncertainty more influential in the stock market.

Models (3) to (6) test how information disparity affects estimates of the power-law exponent α. In Fig 3, we observe that the power-law exponent deviates from unity, corresponding to Zipf's law, during recessions. Models (3) and (4) employ the absolute deviation from unity, $\lvert\alpha-1\rvert$, and the squared deviation from unity, $(\alpha-1)^2$, respectively, as the dependent variables and are thus specified by

$$\lvert\alpha-1\rvert = c_3 + \beta_5 I + \varepsilon_3 \qquad (6)$$

and

$$(\alpha - 1)^2 = c_4 + \beta_6 I + \varepsilon_4, \tag{7}$$

respectively. As evident from Table 1, in years with higher levels of information disparity, the power-law exponent deviates more from unity, either upward or downward.

In Models (5) and (6) the level of the power-law exponent is adopted as the dependent variable in the following way:

$$\alpha = c_5 + \beta_7 R + \varepsilon_5 \tag{8}$$

and

$$\alpha = c_6 + \beta_8 R + \beta_9 I + \beta_{10} R \cdot I + \varepsilon_6. \tag{9}$$

In Model (5), where only the Regulation FD is used as the independent variable, it appears that the power-law exponent generally becomes smaller after the introduction of Regulation FD. However, in Model (6), one can clearly see that this is not the case. The Regulation FD itself does not affect the power-law exponent. Model (6) suggests that the power-law exponent is unconditionally close to unity, indicated by the constant term 0.98. Furthermore, the power-law exponent is significantly positively and negatively affected by information disparity before and after the Regulation FD, respectively (specifically, a unit increase in information disparity will lead to the increase and decrease in the power-law exponent by 0.83 and 0.67, respectively). These results are consistent with our observation in Fig 3 that analyst forecast errors follow Zipf's Law in ordinary years. They are also consistent with our observation during recessions that analysts tend to herd more in the tail before Regulation FD and more to the consensus after Regulation FD.

Monthly pattern in extreme forecast errors

Observing that α can take different values over time, we next investigate whether there is any time-series pattern in α within a given year. Since we use analyst forecasts on annual EPS, which will be realized after the fiscal year end, analysts forecasts issued earlier in the year suffer from higher uncertainty than those issued later in the year. Thus the change of the power-law exponent within a year reflects the change in collective behavior of analysts due to the change in the informational environment. We first plot in Fig 4 monthly changes of α in eight sample years, which are selected randomly. It is observed that the majority shows a clear descending pattern as it approaches the end of a year. (Data for all other years also exhibit qualitatively the same results, albeit not shown for the sake of readability.) Referring to our previous explanation, the decreasing value of α implies that the collective behavior becomes weaker.

This leads to a counterfactual conclusion: extreme analysts tend to flock less as the year-end approaches, i.e., as information disparity reduces closer to the year-end, there are fewer uncertainties regarding the outcome of the fiscal year. Analysts have less reason to follow blindly extreme forecasts by other analysts because all analysts are provided with richer information about the firms. This result is consistent with our finding in the annual analysis in the previous section. On the other hand, it is conceivable that better information should lead analysts to issue similar forecasts closer to the actual outcome.

A key to resolve this discrepancy is to check the cut-off point (95th percentile) of the tail as our findings for the collective behavior of analyst forecast errors come from the tail. Fig 5 shows the monthly cut-off points in each year. We clearly observe that the cut-off point decreases as it gets closer to the year-end. This indicates that the distribution of analyst forecast errors clusters more around zero towards the year-end.
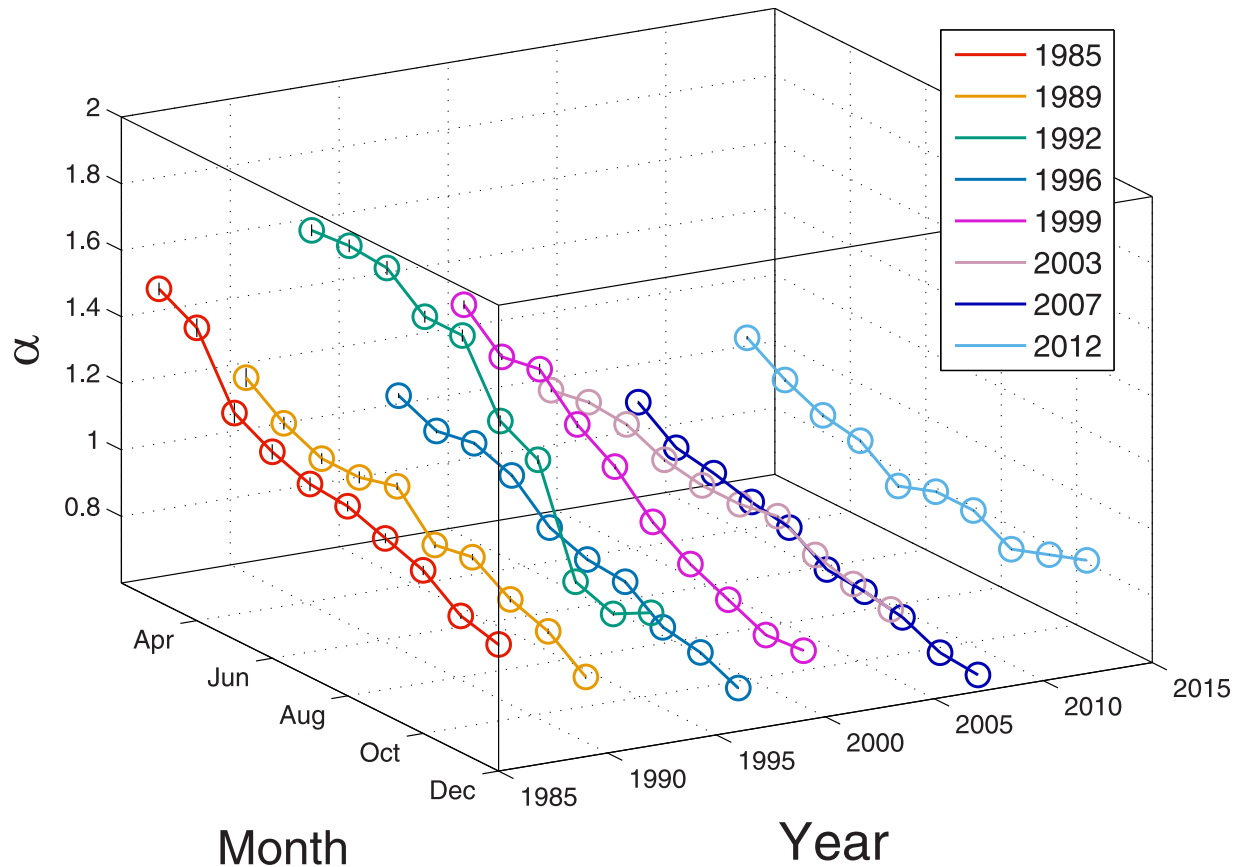
Fig 4. Monthly variations of the power-law exponent by year. Shown are the data in calendar years 1985, 1989, 1992, 1996, 1999, 2003, 2007, and 2012; these sample years have been chosen for the sake of brevity, but the trends in other calendar years are largely similar to those shown in this figure. The colored circles represent the monthly estimates of the power-law exponent from March to December in each given year, with error bars estimated from the power-law fit. Lines connecting circles are merely guides to the eye. There appears a clear descending pattern of the power-law exponent within a year. Namely, the power-law exponent tends to decrease towards the year-end.

https://doi.org/10.1371/journal.pone.0177071.g004

In other words, we see that analysts lose their collective behavior in the tail, and at the same time they herd toward the center as time goes by. Overall, Figs 4 and 5 suggest that analyst forecasts clearly become less biased closer to the year-end and analysts are less likely to follow blindly extreme forecasts.

Monthly pattern in non-extreme forecast errors

As shown in Fig 1, forecast errors in the main body form a log-normal distribution. To probe the monthly pattern in non-extreme forecast errors, we fit the monthly data to the log-normal distribution for each month in calendar year 1993 and examine how the distribution evolves in time. Investigation of such monthly dynamics of the main body may also help substantiate the monthly pattern in extreme forecast errors as the pattern of the main body should relate with that of the tail.

Note that a log-normal distribution has two parameters: location parameter μ corresponding to the mean and scale parameter σ measuring the deviation. If the location parameter is smaller, the bulk of the distribution locates closer to zero whereas the scale parameter affects the skewness as well as thickness of the tail. The larger the scale parameter, the more positively
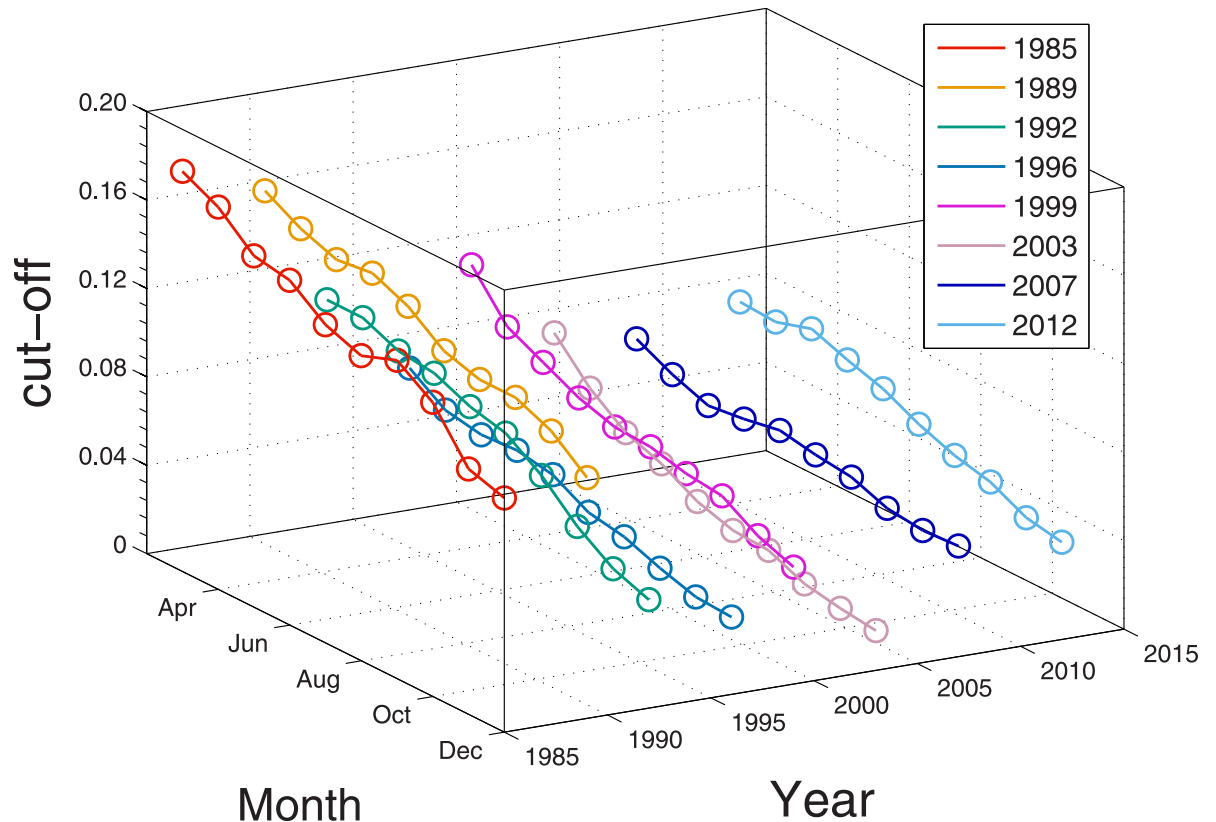
Fig 5. Monthly cut-off points by year. For comparison, the same sample years as those in Fig 4 have been selected. We show the cut-off points for the tail part of forecast errors (95th percentile of the observations) from March to December in each given year. The descending trend of cut-off points manifests that the forecast errors converge to zero towards the year-end, which implies that analysts are becoming more precise in their forecasts.

https://doi.org/10.1371/journal.pone.0177071.g005

skewed and fatter the tail. Namely, the tail is pulled more in the positive direction and becomes thicker [50] in the main body.

Fig 6 exhibits the monthly change of the two parameters. It is observed that μ reduces linearly with time t and that σ increases in proportion to the square root of time: $\mu = At + \mu_0$ and $\sigma = B\sqrt{t} + \sigma_0$ with the slopes $A = -(3.2 \pm 0.5) \times 10^{-3}$ and $B = (1.5 \pm 0.6) \times 10^{-2}$ from March to December (with t in units of day). This indicates that the distribution of the main body moves toward smaller values in each month while becoming fatter in the tail.

Theoretical framework

The fitted distributions can be understood conveniently by the master equation description of the Yule process, originally proposed to explain growth of particle size [40, 51]. Making use of the similarities between the particle growth and analysts' behavior, this approach explains resulting skew distributions of analyst forecasts, including log-normal as well as power-law distributions.

In the master equation, the transition rate is constructed from analyst's forecast updates and we obtain the evolution equation for the distribution of analyst forecasts. Brown and collaborators [52] suggested important information sources or considerations for analysts' earnings forecasts, which are, in the order of importance, industry knowledge, communication with management, earnings conference calls, management earnings guidance, quality or
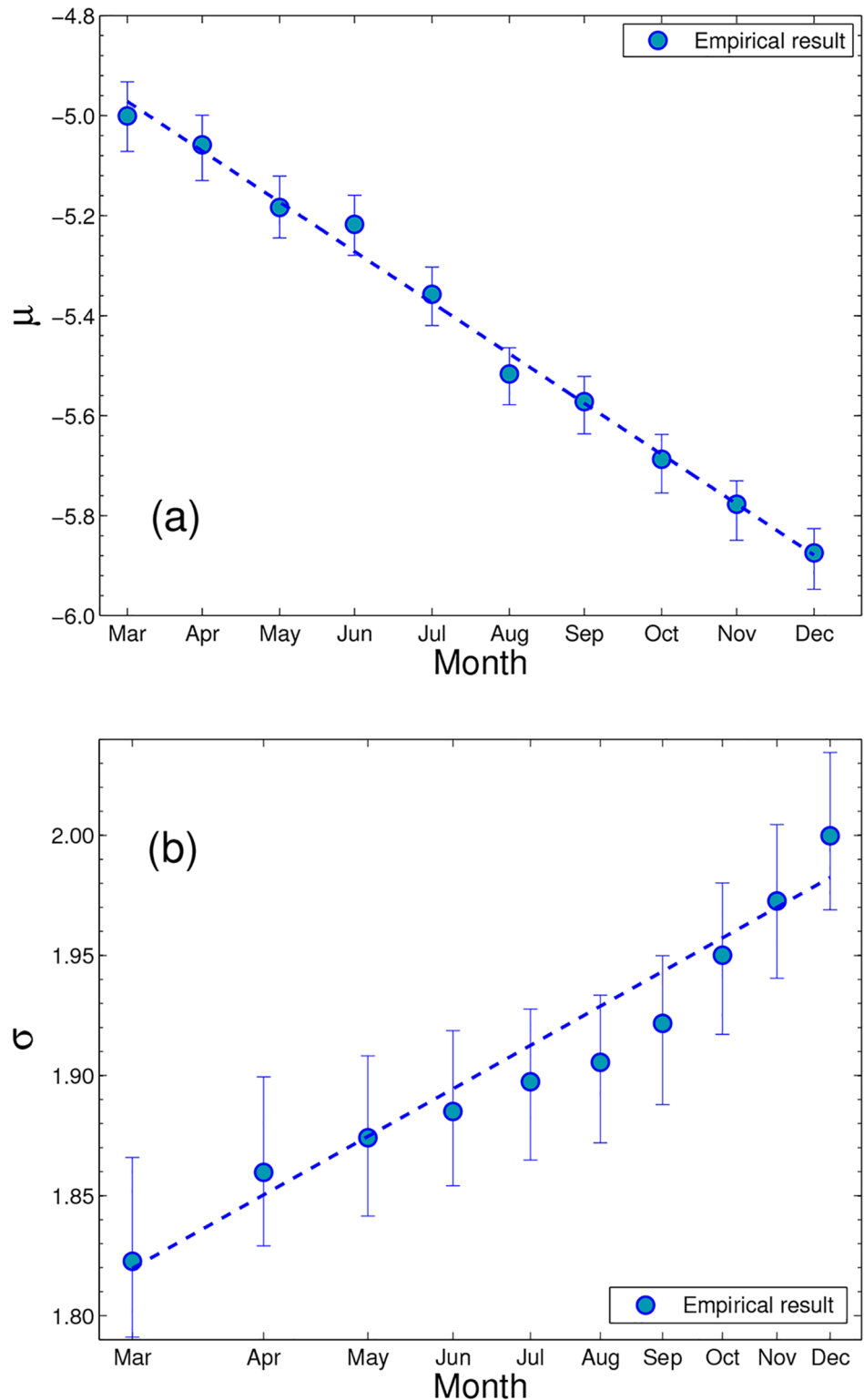
Fig 6. Monthly evolution of (a) the location parameter *μ* and (b) the scale parameter *σ*. This figure presents the monthly changes in μ and σ of the log-normal fit of the main body part of forecast errors in 1993. As predicted by the model, μ is linear in time and σ is linear with respect to the square root of time (note that the scale of the horizontal axis is given in the square root of time). Blue dots in (a) and (b) are the estimates of the location and the scale parameters, respectively; dashed lines in (a) and (b) have the slopes −0.0032 and 0.0148, respectively, per day.

reputation of management, recent earnings performance, recent Form 10-K (annual report of a company's financial performance required by the SEC) or Form 10-Q (quarterly report), primary research, other analysts' earnings forecasts, own stock recommendations, and recent stock price performance.

While earnings conference calls, management earnings guidance, quality or reputation of management, recent earnings performance, recent Form 10-K or Form 10-Q, and own stock recommendations are available only sporadically or quarterly over time, other factors can be updated more often. In this view, if an analyst does not update the earnings forecast on a particular day, that analyst is counted issuing the same earnings forecast as the previous day. Namely, according to the Bayesian inference, we assume that an analyst, when no new information is available, would not issue a new forecast.

We thereby consider the forecasts to be evolving elements and the forecast errors to form a distribution evolving according to forecast updates, which reveals analysts' behavior. Specifically, an analyst's forecast error is viewed as the "size" of the element. The size x grows or diminishes, namely, the forecast error x changes to x', when an analyst updates a forecast upon the arrival of new information; we call this *an event*.

It is expected that the amount of growth, $\Delta x \equiv x' - x$, depends on the present value x. Based on the Yule process, we define that Δx is proportional to x: $\Delta x = bx$, and write the transition rate associated with an event in the form

$$\omega(x \to x') = \lambda\delta[x' - (1 + b)x], \tag{10}$$

where λ is the occurrence rate of events. The growth factor b may be estimated from the average of the changes of the forecast values in the data, i.e., $b = (x' - x)/x = (z' - z)/z$.

Suppose that there are a total of N forecasts at given time t. The error in each forecast can change according to the transition rate in Eq 10. Then the time evolution of the probability $P(x_1, \ldots, x_N; t)$ of the N forecast errors is conveniently described by a master equation [51]. In terms of the distribution function, relating with the probability $P(x_1, \ldots, x_N; t)$ via

$$f(x, t) = \frac{1}{N} \int dx_1 \cdots dx_N \sum_i \delta(x_i - x)P(x_1, \ldots, x_N; t), \tag{11}$$

the master equation reduces to the evolution equation

$$\frac{\partial}{\partial t}f(x, t) = -(r + \lambda)f(x, t) + \frac{\lambda}{1 + b}f\left(\frac{x}{1 + b}, t\right) + rg(x, t), \tag{12}$$

where $g(x, t)$ is the distribution function of newly produced forecast errors with the production rate r. Here production of a forecast is referred to as the issuance of the first forecast on a firm by an analyst in a given (fiscal) year.

We first consider the case that new forecast errors are produced according to the current distribution. In this case of *self-size production*, where the distribution of updated forecast errors is identical to that of existing forecast errors [$g(x, t) = f(x, t)$], the solution of the master equation is given by the log-normal distribution [40]:

$$f(x, t) = \frac{1}{\sqrt{2\pi}\sigma_t x} \exp\left[-\frac{(\ln x - \mu_t)^2}{2\sigma_t^2}\right], \tag{13}$$

where the mean and the deviation evolving in time are given by

$$\begin{aligned} \mu_t &= \lambda t \ln(1+b) + \mu_0 \\ \sigma_t &= \sqrt{\lambda t}|\ln(1+b)| + \sigma_0 \end{aligned}$$

(14)

with the initial mean μ_0 and deviation σ_0.

It is remarkable that the empirical fit in Fig 1 exhibits the same form of the log-normal distribution in Eq 13. In particular, the time evolution is the same, with the identification $A = \lambda \ln(1+b)$ and $B = \sqrt{\lambda}|\ln(1+b)|$; this manifests that the analyst forecast indeed grows in a similar manner to the particle size growth.

Further, the power-law type distribution can also be explained with the same model. Specifically, analysts in the tail are expected to produce new (extreme) forecasts, whose value may vary within a limited range and are thus more or less uniform. In such a case of *uniform-size production*, i.e., $g(x, t) = \delta(x - x_0)$, the stationary-state solution of Eq 12 is given by the power-law distribution [40]:

$$f(x) = \frac{r}{\lambda b x_0} \left(\frac{x}{x_0}\right)^{-(\alpha+1)} \theta(x - x_0)$$

(15)

with the Heaviside step function $\theta(\cdot)$, where the power-law exponent depends on the parameters as follows:

$$\alpha = \frac{\ln(1 + r/\lambda)}{\ln(1+b)}.$$

(16)

Note that the distribution shape of the newly produced forecasts works as a discriminating factor between the main body and the tail. A log-normal distribution can be well explained if the distribution of newly produced forecasts are almost identical to that of the existing forecasts. On the other hand, if newly produced forecasts are almost uniform, there emerges a power-law distribution. The proposed model, when appropriate parameters associated with analyst forecast behaviors are recognized, thus yields distributions for both cases.

For quantitative comparison, we use $b = -0.12 \pm 0.04$, $\lambda = (1.11 \pm 0.48) \times 10^{-2}$, and $r = (5.5 \pm 4.4) \times 10^{-3}$, estimated from empirical data for the main body, to obtain the slopes of μ_t and of σ_t: $\lambda \ln(1+b) = -0.0015 \pm 0.0008$ and $\sqrt{\lambda}|\ln(1+b)| = 0.014 \pm 0.006$, respectively. Within error bars, these values are reasonably consistent with the slopes A and B obtained from the empirical analysis, which confirms the relevance of our model. As for the tail part, which exhibits power-law behavior, we obtain the growth factor $b = 0.30 \pm 0.05$, update rate $\lambda = (9.3 \pm 2.6) \times 10^{-3}$, and production rate $r = (5.4 \pm 2.0) \times 10^{-3}$, estimated from empirical data for the extreme forecasts. Using these parameters in Eq 16 leads to $\alpha = 1.8 \pm 0.7$, which, albeit demonstrating the emergence of complexity, is appreciably larger than the empirical value of unity.

To resolve this discrepancy in α, we extend our model to incorporating the effects of sensing other analysts' forecasts. Such *interactions* among constituents have never been considered in the growth problem. We assume that there is a tendency toward reducing the difference from other forecasts, which leads to the contribution $c_{ij}(x_j - x_i)$ of the jth forecast to the amount of growth of the ith forecast error. In case that detailed information on c_{ij} is not available, it is rational to avoid any bias. We thus take the uniform interaction strength $c_{ij} = c/N$, which is least biased and corresponds to the mean-field theory in physics [53], and sum over j to obtain $\Delta x_i = bx_i + c\bar{x}$, where c/N has been absorbed into b and $\bar{x} \equiv N^{-1}\sum_{j(\neq i)} x_j$ is

essentially the average forecast error. In consequence, Eq 10 changes into

$$\omega(x_i \rightarrow x_i') = \lambda \delta[x' - (1+b)x_i - c\bar{x}], \tag{17}$$

and the ith argument x_i of the probability $P(x_1, \ldots, x_N; t)$ is therefore replaced by $x_i - c\bar{x}$ in the master equation. In terms of $\tilde{x}_i \equiv x_i - c\bar{x}$, we thus have the same expression, which is to be integrated over all x_i. This can be performed by noting $dx_1 \ldots dx_N = J_N^{-1} d\tilde{x}_1 \ldots d\tilde{x}_N$, where J_N is the Jacobian determinant of the mapping between x and $\tilde{x}$. Accordingly, the presence of interactions introduces the Jacobian determinant in the result. Although it is formidable to calculate exactly the $N \times N$ determinant, the lowest-order correction may be obtained for weak interactions ($c < 1$).

In the case of the stationary (power-law) distribution, which is our concern, Eq 16 is modified as follows:

$$\alpha^{\text{int}} = \frac{\ln(1 - c^2/2) + \ln(1 + r/\lambda)}{\ln(1+b)}. \tag{18}$$

Empirical data indeed show that analysts in the tail interact with each other. With the mean growth factor b, the interaction strength is defined according to $z_{i+1} - z_i = bz_i + c_i(\bar{z} - z_i)$, which leads to $c = [z_{i+1} - (1+b)z_i]/(\bar{z} - z_i)$ with the precalculated value of b. Taking the average, we thus estimate $c \approx 0.58$, although large uncertainty due to strong fluctuations among analysts apparently restricts the relevance of this mean value. Putting this value into Eq 18, we obtain $\alpha^{\text{int}} \approx 1.1$. This value is substantially closer to the empirical value $\alpha \approx 1$, which is remarkable in view of the simple consideration of the interactions.

Discussion

Financial analysts play a substantial role in financial markets as they collect, process, and disseminate valuable information to market participants, especially to retail investors who generally lack the time and resources to perform research on companies individually. Hence, whether analyst forecasts offer unbiased information in a timely manner is important. Our study focuses on the influence of information disparity on the analyst behavior: effects of 1) Regulation FD, 2) information dissemination throughout a fiscal year, and 3) interactions among financial analysts.

By dividing optimistic analyst forecast errors into extreme forecasts above the 95th percentile and the remaining non-extreme ones below, we document different analyst forecast behaviors in the two groups. The main body of forecast errors fits a log-normal distribution. On the other hand, the tail of forecast errors follows a power-law distribution with exponent $\alpha \approx 1$ in normal years. During recessions, there are more intensive collective behaviors among extreme analyst forecasts before Regulation FD, while we observe less collective behavior after Regulation FD.

Carrying out big-data analysis on the time-series of annual power-law exponents, we find that the power-law exponents of forecast errors are mostly close to unity, exhibiting Zipf's law. During the years of economic recessions, the exponents deviate from unity, implying that analyst forecasts behave differently during the period of high information disparity. Within a year, the monthly power-law exponent decreases with time, indicating that less collective behavior comes into play among the extreme forecasts. Further, the cut-off point of the 95th percentile moves closer to the center towards the year-end, suggesting that more forecasts herd with the majority. On the other hand, probing the monthly pattern in non-extreme forecasts, we observe that the corresponding log-normal distribution evolves in accordance with the mean

and the variation growing linearly in time and proportionally to the square root of time, respectively.

Making use of the analogy with the growth of particles, we have proposed a theoretical model for the dynamics of forecasts. It provides a good description, not only qualitatively but also quantitatively, of such features as the power-law distribution of the tail as well as the log-normal distribution of the main body, evolving through the mean and the deviation growing in time. In particular, the model, incorporating interactions between analysts, gives the power-law exponent in excellent agreement with the empirical result. We thus believe that our theoretical model, despite its simplicity, captures the key ingredients relevant to the actual dynamics of analyst forecasts.

These findings shed light on our understanding of how analysts form their forecasts, based not only on material public information but also on existing forecasts of other analysts; this influences substantially analyst forecasts. This is more distinct when the level of information disparity in the market is higher.

Acknowledgments

Author Contributions

Conceptualization: KA MYC.

Data curation: DSK.

Formal analysis: CK MYC.

Funding acquisition: CK DSK KA MYC.

Investigation: CK MYC.

Methodology: CK DSK MYC.

Project administration: KA.

Supervision: MYC.

Visualization: CK.

Writing – original draft: CK DSK KA MYC.

Writing – review & editing: CK DSK KA MYC.

References

1. Lang M, Lundholm R. Corporate disclosure policy and analyst behavior. Accounting Review. 1996; 71:467–492.

2. Healy P, Hutton A, Palepu K. Stock performance and intermediation changes surrounding sustained increases in disclosure. Contemporary Accounting Research. 1999; 16:485–520. https://doi.org/10.1111/j.1911-3846.1999.tb00592.x

3. Graham J, Harvey C, Rajgopal S. The economic implications of corporate financial reporting. Journal of Accounting and Economics. 2005; 40:3–73. https://doi.org/10.1016/j.jacceco.2005.01.002

4. Abarbanell J, Lehavy R. Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts. Journal of Accounting and Economics. 2003; 36:105–146. https://doi.org/10.1016/j.jacceco.2003.11.001

5. Bartov E, Givoly D, Hayn C. The rewards to meeting or beating earnings expectations. Journal of Accounting and Economics. 2002; 33:173–204. https://doi.org/10.1016/S0165-4101(02)00045-9

6. Bhojraj S, Hribar P, Picconi M, McInnis J. Making sense of cents: an examination of firms that marginally miss or beat analyst forecasts. Journal of Finance. 2009; 64:2361–2388. https://doi.org/10.1111/j.1540-6261.2009.01503.x

7. Bradshaw M. How do analysts use their earnings forecasts in generating stock recommendations? Accounting Review. 2004; 79:25–50. https://doi.org/10.2308/accr.2004.79.1.25

8. Brown LD. A Temporal Analysis of Earnings Surprises: Profits versus Losses. Journal of Accounting Research. 2001; 39:221–241. https://doi.org/10.1111/1475-679X.00010

9. Gleason C, Lee C. Analyst forecast revisions and market price discovery. Accounting Review. 2003; 78:193–225. https://doi.org/10.2308/accr.2003.78.1.193

10. O'Brien PC. Analysts' forecasts as earnings expectations. Journal of Accounting and Economics. 1988; 10:53–83. https://doi.org/10.1016/0165-4101(88)90023-7

11. Welch I. Herding among security analysts. Journal of Financial Economics. 2000; 58:369–386. https://doi.org/10.1016/S0304-405X(00)00076-3

12. Abarbanell J, Lanen W, Verrecchia R. Analysts' forecasts as proxies for investor beliefs in empirical research. Journal of Accounting and Economics. 1995; 20:31–60. https://doi.org/10.1016/0165-4101(94)00392-I

13. Diether K, Malloy C, Scherbina A. Differences of opinion and the cross section of stock returns. Journal of Finance. 2002; 57:2113–2141. https://doi.org/10.1111/0022-1082.00490

14. Johnson T. Forecast dispersion and the cross section of expected returns. Journal of Finance. 2004; 59:1957–1978. https://doi.org/10.1111/j.1540-6261.2004.00688.x

15. Park C. Stock return predictability and the dispersion in earnings forecasts. Journal of Business. 2005; 78:2351–2375. https://doi.org/10.1086/497047

16. Garfinkel J, Sokobin J. Volume, opinion divergence, and returns: A study of post-earnings announcement drift. Journal of Accounting Research. 2006; 44:85–112. https://doi.org/10.1111/j.1475-679X.2006.00193.x

17. Chopra V. Why so much error in analysts' earnings forecasts? Financial Analysts Journal. 1998; 54(6): 35–42. https://doi.org/10.2469/faj.v54.n6.2223

18. Hong H, Kubik J, Solomon A. Security analysts' career concerns and herding of earnings forecasts. RAND Journal of Economics. 2000; 31:121–144.

19. Hong H, Kubik J. Analyzing the analysts: Career concerns and biased earnings forecasts. Journal of Finance. 2003; 58:313–351. https://doi.org/10.1111/1540-6261.00526

20. Du Q, Yu F, Yu X. Standing out from the crowd: The outlier's effect on corporate governance. Unpublished working paper. 2014; 1–58.

21. Bak P. How Nature Works. Copernicus; 1999.

22. Johnson S. Emergence: The Connected Lives of Ants, Brains, Cities and Software. Scribner; 2002.

23. Schelling TC. Micromotives and Macrobehavior. Norton; 2006.

24. Johnson NF. Simply Complexity: A Clear Guide to Complexity Theory. Oneworld; 2007.

25. Clement M, Tse S. Financial analyst characteristics and herding behavior in forecasting. Journal of Finance. 2005; 60:307–341. https://doi.org/10.1111/j.1540-6261.2005.00731.x

26. Gabaix X. Zipf's law for cities: an explanation. Quarterly Journal of Economics. 1999; 114:739–767. https://doi.org/10.1162/003355399556133

27. Lux T. The stable paretian hypothesis and the frequency of large returns: an examination of major german stocks. Applied Financial Economics. 1996; 6:463–475. https://doi.org/10.1080/096031096333917

28. Acemoglu D, Carvalho V, Ozdaglar A, Tahbaz-Salehi A. The network origins of aggregate fluctuations. Econometrica. 2012; 80:1977–2016. https://doi.org/10.3982/ECTA9623

29. Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. A theory of power-law distributions in financial market fluctuations. Nature. 2003; 423:267–270. https://doi.org/10.1038/nature01624 PMID: 12748636

30. Mainzer K. Thinking in Complexity. Springer; 2007.

31. Nicolis G, Nicolis C. Foundations of Complex Systems. World Scientific; 2012.

32. Gabaix X. Power-laws in economics and finance. Annual Review of Economics. 2009; 1:255–294. https://doi.org/10.1146/annurev.economics.050708.142940

33. Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. A unified econophysics explanation for the power-law exponents of stock market activity. Physica A. 2007; 382:81–88. https://doi.org/10.1016/j.physa.2007.02.030

34. Plerou V, Gopikrishnan P, Gabaix X, Stanley HE. Quantifying Stock Price Response to Demand Fluctuations. Physical Review E. 2002; 66(027104):1–4.

35. Gabaix X, Gopikrishnan P, Plerou V, Stanley HE. Institutional investors and stock market volatility. Quarterly Journal of Economics. 2006; 121:461–504. https://doi.org/10.1162/qjec.2006.121.2.461

36. Krause A. Herding behavior of financial analysts: a model of self-organized criticality. In: Galegatti M, Kirman A, Marsili M, editors. The Complex Dynamics of Economic Interaction, Lecture Notes in Economics and Mathematical Systems. vol. 531/Part IV; 2004. p. 257–267.

37. Hull J. Risk Management and Financial Institutions, 3rd ed. Wiley & Sons, Hoboken, NJ; 2012.

38. Zipf GK. Human Behavior and the Principle of Least Effort. Addison-Wesley, Reading, MA; 1949.

39. Cancho RF, Sole RV. Least effort and the origins of scaling in human language. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100:788–791. https://doi.org/10.1073/pnas.0335980100

40. Goh S, Kwon HW, Choi MY, Fortin JY. Emergence of skew distributions in controlled growth processes. Physical Review E. 2010; 82:061115–061118. https://doi.org/10.1103/PhysRevE.82.061115

41. Bloom N. The impact of uncertainty shocks. Econometrica. 2009; 77:623–685. https://doi.org/10.3982/ECTA6248

42. Loh R, Stulz R. Is sell-side research more valuable in bad times? NBER Working Paper. 2014; 19778:1–36.

43. Baran LC, King THD. S&P500 Index reconstitutions and information asymmetry. Applied Financial Economics. 2014; 24. https://doi.org/10.1080/09603107.2014.904489

44. Chen HL, Bondt WD. Style momentum within the S&P-500 index. Journal of Empirical Finance. 2004; 11:483–507. https://doi.org/10.1016/j.jempfin.2004.04.005

45. Securities and Exchange Commission. Selective Disclosure and Insider Trading. vol. Release 33; 2000.

46. Eleswarapu VR, Thompson R, Venkataraman K. The Impact of Regulation Fair Disclosure: Trading Costs and Information Asymmetry. Journal of Financial and Quantitative Analysis. 2004; 39:209–225. https://doi.org/10.1017/S0022109000003045

47. Heflin F, Subramanyam KR, Zhang Y. Regulation FD and the Financial Information Environment: Early Evidence. Accounting Review. 2003; 78:1–37. https://doi.org/10.2308/accr.2003.78.1.1

48. Bloom N. Fluctuations in uncertainty. Journal of Economic Perspectives. 2014; 28: 153–175. https://doi.org/10.1257/jep.28.2.153

49. Newey WK, West KD. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica. 1987; 55: 703–708. https://doi.org/10.2307/1913610

50. Chaudhuri A, Ghosh SK. Quantitative Modeling of Operational Risk in Finance and Banking Using Possibility Theory. Springer Int'l Pub.; 2015.

51. Choi MY, Choi H, Fortin JY, Choi J. How skew distributions emerge in evolving systems. Europhysics Letters. 2009; 85(30006):1–5.

52. Brown LD, Call AC, Clement MB, Sharp NY. Inside the "Black Box" of sell-side financial analysts. Journal of Accounting Research. 2015; 53:1–47. https://doi.org/10.1111/1475-679X.12067

53. Goldenfeld N. Lectures on Phase Transitions and the Renormalization Group. Addison-Wesley; 1992.