JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS

doi:10.1017/S0022109024000930

Social Media Analysts' Skill: Evidence from Text-Implied Beliefs

Chukwuma Dim
The George Washington University
cdim@gwu.edu

Abstract

This paper documents that 56% of nonprofessional social media investment analysts (SMAs) are skilled and declare beliefs that generate positive abnormal returns (ABRs), while 44% produce negative ABRs. 13% of all SMAs are high-skill type and produce a 1-week 3-factor alpha of 61 bps, while the remaining 87% generate only 6 bps. The distinctive features of high-skill SMAs are primarily firm and industry specializations. Although SMAs tend to extrapolate and herd, their expectations are not systematically wrong. For higher-skilled SMAs compared to the less-skilled ones, extrapolation fades more quickly, and herding is lower, consistent with theory.

I. Introduction

Social networks shape people's expectations and actions, as individuals rely on their networks for information. A crucial feature of investment-focused online social networks is the existence of influencers (i.e., nonprofessional *Social Media Investment Analysts* (SMAs)), who publish investment opinions that shape the views and actions of many individual investors. This paper examines the extent to which these individual SMAs express informative beliefs about stock returns and how they arrive at their beliefs.

This paper was formerly titled "Should Retail Investors Listen to Social Media Analysts? Evidence from Text-Implied Beliefs." I thank Stephan Siegel (the editor) and an anonymous referee for their guidance. I am grateful to Grigory Vilkov and Francesco Sangiorgi for their support and suggestions. I received helpful comments from Zhe An (discussant), Fahiz Baba-Yara, Fousseni Chabi-Yo, Tony Cookson (discussant), Alan Crane (discussant), Francesco Franzoni, Amit Goyal, Gustavo Grullon (DESI mentor), Michalis Haliassos, Jonas Happel, Rawley Heimer, Burton Hollifield, Emirhan Ilhan, Yigitcan Karabulut, Nils J. Krakow (discussant), Camelia Kuhnen, Matthias Lassak, Mario Milone (discussant), Marina Niessner (discussant), Altan Pazarbasi, Kim Peinjenburg, Stefan Pohl (discussant), Zacharias Sautner, Nic Schaub, Onur Sefiloglu (discussant), Kelly Shue (discussant), Fabio Trojani (discussant), Raman Uppal, Maximilian Voigt, and Wei Wu (discussant). I thank participants at the following conferences for their helpful comments: 2022 WFA, 2022 DESI/FMA, 2022 CICF, 2022 BFWG, 2022 FDU, 2022 SGF, 2021 FIRS, 2021 NFA, 2021 NOVA Finance PhD Pitch Perfect, 2019 Financial Markets and Corporate Governance conference, 2021 LTI/Bank of Italy Workshop, 2022 Frankfurt Reading Group on Household Finance, 2021 European Retail Investment Conference Doctoral Consortium, 2021 Queen Mary University of London Economics and Finance Workshop, 2021 International French Finance Association conference, 2021 Dauphine Finance PhD Workshop, and Frankfurt School of Finance & Management Brownbag Seminar.

https://doi.org/10.1017/S0022109024000930

In doing so, I advance the understanding of crowdsourced financial information by providing new evidence on the distribution of skill among individuals who disseminate investment ideas on the widely used Seeking Alpha platform. This is crucial because, although different social platforms have varying informational content (Cookson, Lu, Mullins, and Niessner (2024)) and have democratized investment research, little is known about the mechanism underlying value-relevant information production on each platform and how it compares with professional analysts. I show that over half of SMAs on Seeking Alpha are skilled and express beliefs that generate positive abnormal returns (ABRs), but only 13% are the high-type ones generating substantial positive returns. Furthermore, there are patterns of extrapolation and herding in the SMAs' beliefs. For higher-skilled SMAs, extrapolation fades more quickly, and herding is lower, consistent with theory.

Several factors render Seeking Alpha an ideal setting to study SMAs' skills and beliefs. First, Seeking Alpha has a long history, dating back to 2004, and is popular among retail investors. For instance, about 20 million people use Seeking Alpha monthly, and in my sample about 11,000 SMAs contributed views on roughly 7,200 firms between 2005 and 2019.[1] Second, Seeking Alpha's goal is to provide opinions and analyses rather than news, primarily from individual investors (Seeking Alpha (2006)). Third, views expressed on the platform are backed by an in-depth analysis checked by an editorial team for quality before publication. Fourth, Seeking Alpha incentivizes effort, accuracy, and the disclosure of true beliefs by providing a compensation scheme tied to one's reputation and followership on the platform. Overall, these features imply that SMAs (influencers) in this paper primarily refer to individual investors—relatively more sophisticated than the average retail investor—who share their investment beliefs on Seeking Alpha, potentially shaping other investors' views and actions.

A challenge in analyzing individual SMAs' skills is obtaining a sizeable sample of each SMA's expectations that readily map into buy/sell investment decisions. I overcome this challenge by applying natural language processing and supervised machine learning (ML) techniques to infer beliefs about a large cross-section of stocks over a relatively long period from views expressed by SMAs on Seeking Alpha. More specifically, since 2018, most SMA opinion articles on Seeking Alpha are explicitly tagged with the author's belief about a stock using one of the following descriptions: "Very Bullish," "Bullish," "Neutral," "Bearish," or "Very Bearish." I use the subset of articles with explicit belief statements to train a Support Vector Classifier (SVC), which enables me to extract the beliefs implied by the rest of the unlabelled articles dating back to 2005. I then analyze this large sample of stated and extracted beliefs.

The analysis begins with examining SMAs' stock coverage, which sheds light on whether perverse incentives aimed at price manipulation primarily drive opinion publications, leading to deliberate belief misstatements. SMAs predominantly cover large, growth, high-price, liquid, and low-volatility stocks, indicating that coverage decisions are not primarily driven by the aim to manipulate prices. More so, SMAs' stock-level consensus beliefs predict ABRs up to a 3-month horizon and

[1]See https://seekingalpha.com/page/about_us for usage statistics.

firms' earnings surprises, consistent with earlier work (Chen, De, Hu, and Hwang (2014), Farrell, Green, Jame, and Markov (2022)). These results suggest that SMAs' belief statements are meaningful and can be relied upon to analyze the distribution of SMAs' skill.

To rigorously examine the distribution of the individual SMAs' ability to express informative beliefs, I build on the recent literature on financial professionals' skill (Crane and Crotty (2020), Harvey and Liu (2018), Chen, Cliff, and Zhao (2017)) and model SMAs' ability as a mixture distribution of multiple skill groups.[2] Using the mixture distribution model enables me to address some fundamental issues unaddressed in earlier papers on crowdsourced financial information. First, the unit of analysis in the model is an individual instead of a stock day, which enables the quantification of each SMA's ability to make correct forecasts. Furthermore, using stock as the unit of analysis requires aggregating SMAs' beliefs into a consensus, thereby implicitly assuming that people on social forums base their decisions on such consensus. However, as documented in Cookson, Engelberg, and Mullins (2023), people instead operate in echo chambers on social finance forums, meaning that users heed the views of subsets of SMAs instead of the consensus. As a result, it is relevant to understand i) the distribution of skill at the individual SMA level and ii) whether users mostly follow unskilled SMAs, in which case the consensus could be informative and yet users will not benefit from it.

I estimate a two-component mixture model that measures SMAs' performance based on the average ABRs following their belief statements. The model indicates that 56% of SMAs generate true positive ABRs—that is, are skilled and state beliefs that align with future stock returns—while the remaining 44% produce negative ABRs. The average magnitude of SMAs' skill implied by the model is sizeable, with a 1-week 3-factor ABR of 0.13%, which amounts to 6.8% annualized. However, about 87% of SMAs belong to the lower return distribution, generating a moderate 6 bps 1-week average ABR. The remaining 13% high-type SMAs express beliefs that generate a much larger average ABR of 61 bps over the same horizon but with considerable dispersion. To provide some context, Crane and Crotty (2020) use a similar setup and estimate the fraction of high-type professional analysts to be 36%, with roughly 97% of all analysts being skilled. Hence, while many SMAs state beliefs that align with future returns, SMAs' skill is limited relative to professionals'.

SMAs' estimated skill is reasonably persistent: one's conditional expected skill and probability of being a high-skill type inferred from an earlier sample period (2005–2014) strongly predict the respective quantities in the later period (2015–2019). Analysis of calendar-time portfolios based on SMAs ex ante conditional expected skill indicates that the bullish beliefs of SMAs with high values of the skill measure (top tercile) yield an alpha of around 3% annualized, while that of SMAs with low values of the measure (bottom tercile) are indistinguishable from 0. However, the performance of the top-tercile of SMAs is quite comparable to a portfolio that pools all SMAs. Hence, while observable skill differences exist

[2]The mixture model uses information in the cross-section of SMA performance to reduce noise, thereby ameliorating the false discovery problem that arises from the low signal-to-noise feature of abnormal returns and the low test power in individual-level time series regressions.

among SMAs, pooling can generate considerable returns by exploiting both SMAs' skill and the wisdom-of-the-crowd.

The heterogeneity in SMAs' skill raises the question of what attributes differentiate the high- and low-type SMAs. Consistent with gains from specialization in information acquisition and cognitive capacity constraints (e.g., Van Nieuwerburgh and Veldkamp (2010), Hirshleifer, Lim, and Teoh (2011)), industry and firm specializations are the most distinctive traits of high-type SMAs. SMAs who specialize in a few industries (firms) are 34 (31) percentage points more likely to be high type. Furthermore, popular SMAs have a higher probability of being high type, suggesting that individuals engage more with skilled SMAs.

Next, I analyze the roles of fundamental information and two common behavioral patterns, return extrapolation and herding, in shaping SMAs' beliefs and how an SMA's skill relates to her belief formation pattern. I focus on extrapolation and herding because i) herding is pervasive in several domains, and social media can serve as a coordination device for mutual imitation; ii) recent research shows that extrapolative beliefs expressed on some forums mispredict returns (Da, Huang, and Jin (2021)). Hence, it is valuable to understand whether SMAs herd or extrapolate given that they express beliefs that are *not* systematically wrong.

The analyses show that SMAs extrapolate from past returns when forming beliefs about future returns. The influence of recent past returns on beliefs fades relatively quicker for the more skilled compared to the less-skilled SMAs, consistent with the models of De Long, Shleifer, Summers, and Waldmann (1990) and Barberis and Shleifer (2003) where skilled agents can front-run their less sophisticated peers to profit from the return momentum sustained by the latter. Combined with the documented informativeness of SMA beliefs, the evidence indicates that return extrapolation *does not* necessarily entail systematically biased beliefs. SMAs also herd in stating their beliefs. Herding is more pronounced among the less-skilled, consistent with Bikhchandani, Hirshleifer, and Welch's (1992) model where higher-ability agents deviate more from the consensus.

Jointly estimating the contributions of herding, extrapolation and fundamental information in shaping SMAs' beliefs indicates that all three components play prominent roles. While herding's contribution dominates, the sizeable role fundamental news plays in shaping beliefs potentially explains the informativeness of SMAs' expectations despite the presence of behavioral components. Overall, SMAs on Seeking Alpha are useful information intermediaries: over half of SMAs have some skill, while 13% are high skill, and 44% have negative skill. Moreover, the pooled beliefs of SMAs are informative and contribute to the wisdom-of-the-crowd effect documented in earlier research.

This paper relates to two conflicting strands of the literature on whether investors' social media opinions about individual stocks are informative or biased. Some studies argue that opinions expressed on some platforms exhibit biases or are uninformative, or even mispredict returns (Cookson et al. (2023), Da et al. (2021), Cookson, Engelberg, and Mullins (2020), Ammann and Schaub (2021), Heimer (2016), Antweiler and Frank (2004)). Others emphasize the informativeness of opinions on some social networks (e.g., Avery, Chevalier, and Zeckhauser (2016), Renault (2017), Cookson and Niessner (2020), Cookson et al. (2024)) and, in particular, those posted on Seeking Alpha (Chen et al. (2014), Campbell, DeAngelis, and Moon (2019), Gomez, Heflin, Moon, and Warren (2020), Farrell et al. (2022), Drake, Moon Jr, Twedt, and Warren (2023)). This paper relates more directly to the latter studies. While these existing papers study the informativeness of SMAs' consensus view, firms' information environment, and SMAs' impact on retail trading, I study the distribution of the individual SMAs' skills and the roles of cash flow news and heuristics in shaping SMAs' expectations conditional on skill.

My results contribute the following novel insights to Chen et al. (2014) and related papers on SMAs' consensus informativeness. As shown in Cookson et al. (2023), individuals engage in selective information exposure on social forums. Hence, the consensus may be informative and yet not benefit users if most SMAs are unskilled and users follow subsets of unskilled SMAs. My results, therefore, provide a deeper understanding of i) the sources of SMAs' consensus informativeness by highlighting the relevance of individual-level skill and ii) underscore that selective information exposure is likely less damaging on Seeking Alpha since most SMAs make correct forecasts and users engage more with skilled SMAs.

In terms of methodology, this paper relates to studies that also apply the mixture distribution model to study the skills of professional analysts, mutual funds, and hedge funds (Crane and Crotty (2020), Harvey and Liu (2018), Chen et al. (2017)). In contrast, this paper focuses on SMAs, a different and relatively new class of financial information intermediaries that have increasingly gained the attention of regulators and market participants due to the growing role of social media in financial markets.[3] I provide new insights into how SMAs' skills compare to that of the better-known professionals, deepening our understanding of the different classes of information intermediaries in today's financial market.

This paper adds to the literature studying how retail investors' direct participation in financial markets affects price discovery (e.g., Boehmer, Jones, Zhang, and Zhang (2021), Seasholes and Zhu (2010), Barber and Odean (2000), Kelley and Tetlock (2013), Kaniel, Liu, Saar, and Titman (2012)). Many SMAs are also individual investors, as they disclose their investments on Seeking Alpha. In demonstrating that SMAs' beliefs contain value-relevant information, this paper aligns with the extant literature arguing that retail investors produce information that likely improves price discovery. Unlike studies in this literature focusing on retail trading, this paper studies beliefs and the distribution of skill among a class of individual investors, which offer new insights since trades can be due to reasons unrelated to beliefs, such as liquidity needs.

Finally, I contribute to the literature on investor belief formation (e.g., Kuchler and Zafar (2019), Greenwood and Shleifer (2014), Choi and Robertson (2020), Giglio, Maggiori, Stroebel, and Utkus (2021)), which has largely relied on surveys of beliefs about *aggregate* outcomes. The few exceptions (e.g., Cookson and Niessner (2020), Cookson et al. (2020), Da et al. (2021)) study beliefs declared

[3]In a contemporaneous paper, Farrell, Jame, and Qiu (2020) use a similar methodology to analyze skill on Seeking Alpha, arriving at similar conclusions. This paper differs from theirs by using a precise trading signal based on SMAs' declared beliefs for analysis. Furthermore, I analyze how SMAs' ex ante skill relates to their belief formation pattern, which their paper does not. In a follow-up paper, Kakhbod, Kazempour, Livdan, and Schuerhoff (2023) studied skill among StockTwits' influencers and found a much lower incidence of skill. This indicates that platform-specific features are important, and different social forums offer different informational values.

https://doi.org/10.1017/S0022109024000930

on some other investment social media that are markedly different from Seeking Alpha both in design and incentive structure. As a result, the type of beliefs aggregated and disseminated on Seeking Alpha, the focus of this paper, is likely different and enriches our understanding of people's beliefs about the *individual stocks* they invest in.

II. Data

This section describes the data used in this study. The sample period is from January 2005 to December 2019 due to the availability of SMAs' views on Seeking Alpha. Table A1 in the Appendix describes the main variables used in the study, and Table A2 provides summary statistics.

A. Seeking Alpha Data

This paper uses SMAs' opinions and beliefs published on Seeking Alpha, a popular social finance platform aimed at promoting the exchange of investment ideas among individuals.[4] Any registered user can contribute views on Seeking Alpha by submitting a long-form opinion article detailing an investment thesis (and, more recently, accompanied by an explicit belief statement) on specific stocks. The opinion article must pass through Seeking Alpha's editorial team, which checks for quality standards without interfering with the author's viewpoint. Most Seeking Alpha users consume and comment on the views published by a smaller subset of individuals, referred to as SMAs (influencers) in this paper.

SMAs are different from professional analysts in several ways. Unlike professionals who cater to institutional investors, SMAs are primarily individual investors who share their views on the stocks in which they invest, cater to retail investors, and inform retail trades (Farrell et al. (2022), Campbell et al. (2019)). Hence, SMAs' incentives discussed in Supplementary Material Section IA differ from those of professional financial analysts.

To obtain SMAs' opinions and stated beliefs from Seeking Alpha, I use a web-scraping algorithm to download all opinion articles (and associated belief statements where available) covering a single U.S. common stock listed on the NYSE, NASDAQ, or AMEX stock exchange. I obtain the SMA's ID, investment disclosure, the stock ticker, and the publication date for each article.[5] Furthermore, I retrieved all comments posted in response to the publication by other Seeking Alpha

[4]See https://seekingalpha.com/page/about_us. Views published on Seeking Alpha can be accessed freely with some restrictions by simply registering on the platform. Alternatively, users can pay a subscription fee as low as $39 per month to remove the restrictions. Very few opinions were contributed in 2004 after Seeking Alpha's launch. As a result, the analysis in this paper uses data beginning in Jan. 2005.

[5]Most publications include a disclosure section where the author states whether they have invested in the focal stock. See Figures IF1 and IF2 in the Supplementary Material for examples of these disclosures. I manually label a randomly selected 5000 disclosures as either "Long position," "Short position," or "No position" and then use this labeled sample to train a Support Vector Classifier ML model, as described in Section III, which is used to extract the investment position stated in all other disclosures. Given the simplicity of this particular learning exercise, the trained model achieved an out-of-sample accuracy rate of 99%.

https://doi.org/10.1017/S0022109024000930

users. A nice feature of Seeking Alpha is that, unlike some other social forums, SMAs cannot delete their accounts and remove opinions previously contributed to the platform.[6] This makes survivorship bias less of a concern.

I downloaded 280,514 articles and 7.3 million comments, contributed by roughly 11,000 SMAs and 300,000 users, respectively, covering about 7,200 stocks over the period January 2005–December 2019. The belief statement that accompanies some publications is tagged "Very Bullish," "Bullish," "Neutral," "Bearish," or "Very Bearish."[7] However, most SMA opinion articles published before 2018 did not explicitly state SMAs' beliefs: only 43% of publications before 2018 include an explicit belief statement compared to roughly 75% of publications after 2018. Therefore, to obtain a large cross-sectional and time-series sample of belief statements, I train an ML model, described in Section III, to extract SMAs' underlying beliefs from the articles without explicit belief statements.

Figures IF1 and IF2 in the Supplementary Material show sample SMA opinion articles where the authors explicitly state their beliefs about a stock as "Bullish" and "Bearish," respectively. These examples indicate that individuals who contribute opinions to Seeking Alpha are generally more financially literate than the average retail investor and that Seeking Alpha differs from other investment social media in some fundamental ways. For example, Seeking Alpha differs from StockTwits, Twitter, Forcerank, and CAPS by requiring article-length posts, instead of a few 100 characters, that explain the reasoning behind investment expectations. Unlike forums where users can post anything, Seeking Alpha reviews posts to ensure they provide clear investment ideas backed by reasoned arguments. This moderation ensures that only posts related to financial investments exist on the platform, potentially attracting more financially literate individuals to contribute articles. However, Seeking Alpha also shares similarities with other forums: its user base is predominantly retail investors, and it has standard social media features such as commenting, sharing posts, following other users, and creating favorite stocks.

A valid concern with beliefs disclosed on Seeking Alpha and other social forums is whether the disclosed beliefs are the agents' *true* expectations. I address this issue in detail in Supplementary Material Section IA. To summarize, Seeking Alpha's incentive scheme induces effort to produce useful information and truthfully report one's expectations. Analysis of SMAs' stock coverage decisions and informativeness suggests that, to a large extent, value-relevant information production *not* manipulation is the SMAs' principal goal.

B. Other Data

Stock returns and prices are obtained from CRSP, and firm fundamentals are from Compustat. I compute future ABR $ABR_{k,t}(h)$ for firm k on day t for horizon h relative to 3 benchmarks: CAPM, Fama and French (1993) 3-factor model (FF3), and the Daniel, Grinblatt, Titman, and Wermers (1997) size/book-to-market/momentum characteristics-based benchmark (SBM). For the CAPM and FF3

[6]For more details, see https://feedback.seekingalpha.com/knowledge-bases/2/articles/14279-can-i-delete-a-contributor-account.

[7]Since early 2022, Seeking Alpha changed these labels to "Strong Buy," "Buy," "Neutral," "Sell," and "Strong Sell," respectively. I retain the labels observed on the platform over my sample period.

benchmarks, I estimate betas for each stock using daily data over the trading-day window $t-272$ to $t-21$, where t is the belief publication day. Merging the CRSP/Compustat data with the SMA belief data reduces the number of observations to 236,250.

Data on professional analyst stock recommendations and quarterly earnings per share forecasts are taken from the Institutional Brokers' Estimate System. I calculate earnings surprises from the unadjusted detail history of earnings forecasts. Finally, I measure the tone across a comprehensive set of cash flow relevant news events about a stock on a given day using the Event Sentiment Score from RavenPack.

III. Classifying SMAs' Beliefs from Text

Recent ML applications in finance underscore the usefulness of text data for measuring economic quantities. Inspired by such results and the recent call to construct proxies of beliefs from text (Brunnermeier, Farhi, Koijen, Krishnamurthy, Ludvigson, Lustig, Nagel, and Piazzesi (2021)), I use the subset of SMAs' opinions that includes explicit belief statements to train a supervised ML model, which is used to extract the underlying beliefs from all other articles that do not state the SMA's belief. This ML step is crucial, as the mixture model that comes next requires precise, discrete predictions that can be mapped directly to buy or sell investment decisions. There is arguably no better way to infer such precise predictions from SMAs' text than to use ML, which exploits the high dimensionality of text data to achieve high out-of-sample accuracy.

I use the linear SVC algorithm for this exercise because it performs well in very high-dimensional feature spaces (e.g., Chen, Wu, and Yang (2019), Manela and Moreira (2017)). Linear SVC can be quickly trained on high-dimensional data, as only one hyperparameter needs to be tuned. The output is interpretable, reflecting how specific word combinations relate to beliefs.[8] To train the linear SVC, I first collapse the belief labels to three classes, setting the "Very Bullish" and "Bullish" labels to "Bullish," and setting the "Very Bearish" and "Bearish" labels to "Bearish"; the third label is "Neutral." This reduces class imbalance since the "Very Bullish" and "Very Bearish" beliefs jointly account for only 3% of the labeled data. Supplementary Material Section IB provides a detailed description of the model training procedure, the SVC algorithm, and the out-of-sample model validation.

The first crucial step after model training is out-of-sample validation. Supplementary Material Figure IB1 shows that the trained model's AUC score on the out-of-sample test data is 0.94–0.97 for the different belief classes. The model's accuracy score on the test data is 90%. These validation results suggest that the trained model reliably classifies the bullish, bearish, and neutral belief classes. Supplementary Material Table IB1 provides additional validation and sanity checks. It shows the 50 terms with the largest weights used by the model for classifying the bullish and bearish beliefs, respectively, indicating that the trained model produces intuitive results. For example, articles containing terms such as "overvalue," "neutral," "short," "avoid," "take profit," and "short opportunity" are less likely to be bullish.

[8]Although penalized logistic regression has similar features, SVC performed better out of sample.

TABLE 1

Summary Statistics of SMA Beliefs, ABRs, and Size

Panel A of Table 1 reports the proportion of each belief class in the subsample of publications with explicit belief statements (Stated Beliefs), in the subsample where beliefs are extracted using ML (Extracted Beliefs), and in the combined sample of stated and extracted beliefs (All). Panel B shows the time-series average of the cross-sectional summary statistics for the stock-level aggregate beliefs (Agg. Belief), computed as the difference between the proportion of bullish and bearish beliefs. ABR(h) is ABRs over the next h trading days, starting $t+1$ following publications on day t. Column headers indicate the benchmark used to compute ABRs: CAPM or the 3-factor (FF3) model. Mkt. Cap. is market capitalization (in millions USD) on the publication day.

Panel A. Proportion of Beliefs

	Stated Beliefs	Extracted Beliefs	All
Bullish	0.82	0.79	0.81
Bearish	0.13	0.18	0.16
Neutral	0.05	0.03	0.04
Obs.	117,271	118,979	236,250

Panel B. Stock-Level Variables

			CAPM		FF3	
	Agg. Belief	Mkt. Cap.	ABR(5)	ABR(21)	ABR(5)	ABR(21)
Mean	0.6506	53,938	0.0006	0.0013	0.0006	0.0014
SD	0.6571	95,058	0.0641	0.1201	0.0631	0.1184
Min	−0.7714	1,972	−0.1511	−0.2626	−0.1503	−0.2615
P10	−0.3389	2,741	−0.0534	−0.1073	−0.0520	−0.1053
P25	0.6730	4,841	−0.0245	−0.0510	−0.0235	−0.0492
P50	0.9371	15,514	−0.0005	−0.0016	−0.0003	−0.0012
P75	0.9787	60,167	0.0235	0.0474	0.0228	0.0462
P90	0.9830	148,894	0.0540	0.1062	0.0526	0.1038
Max	0.9854	423,994	0.1775	0.3344	0.1755	0.3318

In contrast, those containing terms such as "undervalue," "upside," "buy," "opportunity," and "bullish" are more likely to be bullish. Overall, the model's impressive performance on the test data and its intuitive n-gram weights suggest that it can be used to reliably infer SMAs' beliefs from their article text.

Table 1, Panel A shows summary statistics for the stated, extracted, and all SMA beliefs. The distribution of the belief classes in the subsample of extracted beliefs is comparable to that in the subsample of stated beliefs, further revealing that the trained model produces reasonable results. Overall, bullish beliefs account for 81% of the stated and extracted beliefs (column "All"), indicating that SMA beliefs are generally bullish. Unless otherwise stated, the rest of the paper uses the stated and extracted beliefs, which provide a larger sample size that is particularly helpful for analyzing the distribution of SMAs' skill. Panel B of Table 1 shows summary statistics for the stock-level consensus SMA beliefs (Agg. Belief), firms' market capitalization, and ABRs for the 5 and 21 trading-day horizons following belief statements. Average ABRs are positive, while the average market capitalization is $53.9 billion.

IV. Distribution of SMAs' Skill

To understand the cross-sectional distribution of SMAs' ability to express informative beliefs, I build on the recent literature on financial professionals' skill (Crane and Crotty (2020), Harvey and Liu (2018), Chen et al. (2017)) and model SMAs' ability as arising from a mixture distribution of multiple skill groups. Such modeling avoids common pitfalls that arise from the low signal-to-noise feature of estimated ABRs—the standard measure of unobservable skill. Noise in estimated

ABRs could result in conventional significance tests at the SMA-level misattributing good luck to skill or bad luck to lack of skill due to low test power. Conversely, the mixture distribution model uses information from the cross-section of SMAs' performance to reduce noise and is not impeded by low test power.

The formulation of the mixture model in this paper follows Crane and Crotty (2020). Assume that there is an unknown number, J, of skill groups. For each group $j \in \{0,1,2,\ldots,J\}$, there is a fraction π_j of SMAs with true ability, captured by ABRs, centered on μ_j. The dispersion of true ABRs for SMAs in group j is driven by variations in true ability arising from investor-specific traits. Let $\alpha_i = \mu_j + \omega_i$ denote the true belief formation ability of SMA i, where ω_i captures individual-specific traits and is normally distributed with 0 mean and variance σ_j^2. On the other hand, estimated ability, $\hat{\alpha}_i$, is measured with noise, e_i, which is assumed to be independent of ω_i and normally distributed with 0 mean and variance s_i^2 (i.e., s_i is the standard error of $\hat{\alpha}_i$). Thus, the estimated abnormal performance of an SMA is $\hat{\alpha}_i = \mu_j + \omega_i + e_i$. Setting $J = 2$, the specifications boil down to a two-component distribution with the density function:

$$f(\hat{\alpha}_i) = \pi_0 \cdot \phi(\hat{\alpha}_i; \mu_0, \sigma_{i,0}) + \pi_1 \cdot \phi(\hat{\alpha}_i; \mu_1, \sigma_{i,1}), \tag{1}$$

where $\phi\left(\hat{\alpha}_i; \mu_j, \sigma_{i,j}\right)$ is the normal density function with mean μ_j and variance $\sigma_{i,j}^2 = \sigma_j^2 + s_i^2$ evaluated at $\hat{\alpha}_i$. The log-likelihood function L for a sample of N estimated SMA skill is

$$L(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_N | s_1, s_2, \ldots, s_N, \Theta) = \sum_{i=1}^{N} \log(f(\hat{\alpha}_i)), \tag{2}$$

where the parameter set $\Theta = \{\pi_0, \pi_1, \mu_0, \mu_1, \sigma_0, \sigma_1\}$ is estimated via maximum likelihood, subject to the constraints that $0 \leq \pi_0 \leq 1$, $\pi_1 = 1 - \pi_0$, and $\sigma_j \geq 0$.

To take the mixture model to the data, I use the 5-day ABR relative to the 3-factor model to measure the estimated SMAs' skill.[9] ABRs ABR_i^k is computed over window $t+1$ to $t+6$ trading days for each belief statement by SMA i about stock k on day t. ABR_i^k is then signed by premultiplying it by +1 for bullish beliefs (long positions) and −1 for bearish beliefs (short positions). Neutral beliefs are excluded because they do not provide a clear investment signal. Finally, each SMA i's ABR is aggregated by averaging across all belief statements:

$$\overline{ABR}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} ABR_i^k. \tag{3}$$

[9]The use of 5-day abnormal returns allows for comparability with results in the professional analyst literature (e.g., Crane and Crotty (2020)). More so, longer horizon returns may be impacted by subsequent developments that might have prompted SMAs' to revise their beliefs. As discussed in the robustness Section VI, the results are robust to alternative benchmarks for computing abnormal returns.

TABLE 2

Summary Statistics for SMA-Specific ABRs

Table 2 summarizes the SMA-specific average 5-day ABRs $\overline{ABR}_i$ and its standard error s_i. ABRs for belief statements result from buying stocks with bullish beliefs and selling stocks with bearish beliefs. The benchmark return for each event is based on the 3-factor model. ABRs are aggregated to the SMA level by estimating the average across all of an SMA's belief statements. Standard errors, s_i, are clustered by publication date and stock. The reported "Frac. +ve" is the fraction of the SMA cross-section with a positive estimated ABR. The K-S p-value is the p-value of a Kolmogorov–Smirnov test of the null hypothesis that the demeaned cross-sectional distribution of $\overline{ABR}_i$ is normally distributed.

	Mean	SD	Min	P25	P50	P75	Max	Skew.	Kurt.	Obs.	Frac. +ve	K-S p-Value
$\overline{ABR}_i(\%)$	0.26	2.57	−18.52	−0.72	0.11	0.99	37.45	1.84	24.1	4190	0.55	0.00
$s_i(\%)$	1.36	1.61	0.01	0.54	0.94	1.61	32.78	6.20	72.4			

The main analysis uses SMAs with at least 5 belief statements over the sample period. $\overline{ABR}_i$ is an SMA's estimated ability, $\hat{\alpha}_i$. Its standard error, s_i, is calculated by clustering on the belief publication day, to account for correlation across beliefs published on the same day, and stock, to account for correlation in beliefs on the same stock.

Table 2 shows summary statistics for $\overline{ABR}_i$ and s_i. The estimated average SMA ability is 26 bps, with a median of 11 bps. The standard errors, s_i, with an average (median) of 136 (94) bps, suggest that the estimated SMA-specific ABRs are considerably noisy. The skewness (1.8) and kurtosis (24.1) of $\overline{ABR}_i$ suggest that the estimated ABR is not normally distributed, and the Kolmogorov–Smirnov test rejects normality at the 1% significance level. These statistics suggest that standard significance tests based on normality can yield incorrect inferences regarding SMAs' ability, validating the application of a mixture distribution model to isolate SMAs' true skill.

A. How Many SMAs State Informative Beliefs?

Table 3 reports the parameter estimates for the two-component mixture model, where the component $j = 0$ comprises the low-type SMAs with a lower average ABR and $j = 1$ comprises the high-type SMAs.[10] Columns 1 and 2 of Panel A show estimates corresponding to the low- and high-type SMA groups. The estimated fraction of low-type SMAs (π_0) is 87%, with the skill distribution centered on an ABR of 6 bps, with a dispersion of 0.4%. Conversely, the fraction of high-type SMAs (π_1) is 13%, with a much larger 1-week ABR of 61 bps (31% annualized) and a dispersion of 3.2%.

Panel B of Table 3 summarizes the cross-sectional distribution of SMAs' skill implied by the mixture model. Importantly, over half (56%) of SMAs have genuinely positive average ABRs following their belief statements (i.e., are skilled and express beliefs that correctly align with future stock ABRs).[11] However, a sizeable

[10]Robustness analysis in Section VI shows that the results are robust to alternative mixture model setups.

[11]Fraction positive is computed as $1-\left[\pi_0\cdot\Phi\left(\frac{0-\mu_0}{\sigma_0}\right)+\pi_1\cdot\Phi\left(\frac{0-\mu_1}{\sigma_1}\right)\right]$. For a given percentile P, the corresponding quantile q is computed numerically by solving $P=\pi_0\cdot\Phi\left(\frac{q-\mu_0}{\sigma_0}\right)+\pi_1\cdot\Phi\left(\frac{q-\mu_1}{\sigma_1}\right)$, where $\Phi(\cdot)$ is the cumulative normal distribution function.

TABLE 3

Parameter Estimates for SMA Skill Distributions

Table 3 reports the result for the two-component mixture model of SMAs' skill using data for only SMAs with at least 5 belief statements. Panel A reports the estimates of the model parameters, where π is the fraction of the low- and high-type SMAs, μ is the mean of each group's true skill, σ is its dispersion, and $\sigma_{i,j}$ is the average dispersion of the estimated skill of each group. Each SMA's estimated ABR ($\overline{ABR}_i$) is computed relative to the 3-factor model for all publications by the SMA, as in equation (3). The reported $\sigma_{i,j}$ is based on the cross-sectional average of $\overline{ABR}_i$'s standard error, s_i. Hence, $\sigma_{i,j} = \sqrt{\sigma_j^2 + \overline{s}^2}$. Estimates in Panel A are used to compute statistics for the cross-sectional distribution of skill reported in Panel B. P(10)–P(90) are percentiles of the implied cross-sectional distribution of SMA skill. Standard errors (in parentheses) are computed as the SD of the statistics from 1,000 bootstrap replications.

Panel A. Mixture Model Parameters

	Low Skill	High Skill
	1	2
Fraction of SMAs (π)	0.8719	0.1281
	(0.0419)	(0.0419)
Mean abnormal return (μ)	0.0006	0.0061
	(0.0002)	(0.0021)
Dispersion in abnormal return (σ)	0.0040	0.0322
	(0.0014)	(0.0038)
Avg. dispersion in estimated skill ($\sigma_{i,j}$)	0.0142	0.0350
	(0.0010)	(0.0034)

Panel B. Mixture Return Distribution

	Estimate	SE
Mean	0.0013	(0.0002)
SD	0.0123	(0.0007)
P10	−0.0056	(0.0007)
P25	−0.0024	(0.0004)
P50	0.0007	(0.0002)
P75	0.0038	(0.0004)
P90	0.0075	(0.0006)
Fraction positive	0.5608	(0.0164)
No. of obs.	4,190	

44% of SMAs express beliefs that yield negative ABRs, with an average value of −55 bps per week.[12] More so, there is considerable dispersion in SMAs' true ability, as the cross-sectional standard deviation of true ability is 1.2%, roughly 47% of the estimated ABR's dispersion (i.e., 2.6% reported in Table 2, which includes variations in ABR attributable to luck/noise).

The heterogeneity in SMAs' true ability and the small fraction of high-type SMAs point toward the difficulty investors might face in identifying skilled SMAs on social finance forums. To provide some context, Crane and Crotty (2020) use a similar setup to estimate the fraction of high-skill professional analysts to be 36%, with roughly 97% sufficiently skilled to generate positive ABRs. Chen et al. (2017) find that 48% of hedge funds have above-neutral skill, and Coval, Hirshleifer, and Shumway (2021) use transaction records and a different methodology to estimate that 10% to 20% of individual investors in their sample are skilled. Overall, although SMAs as a group tend to add value, SMAs' skill is relatively limited, posing a risk to social platform users that sort into echo chambers based on criteria unrelated to skill.

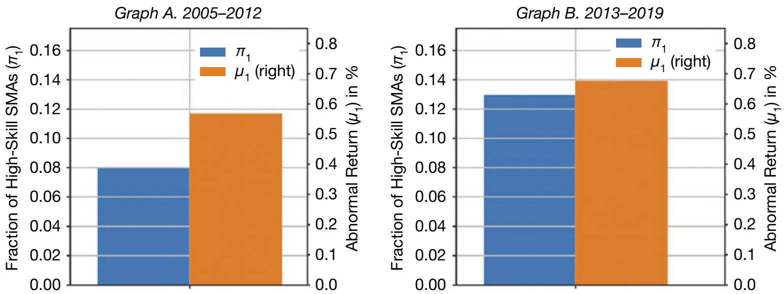
Have SMAs become better at expressing informative beliefs over time? Answering this question provides insights into the evolution of the quality of

[12]The average abnormal return for the negative skill SMAs is obtained by calculating the mean of the truncated mixture distribution using the estimated parameters. See Supplementary Material Appendix ID.1 for details.

FIGURE 1

Distribution of SMAs' Ability in Two Subsamples

Figure 1 shows the fraction of high-skill SMAs, π_1, and their true ability in terms of 5-day ABRs μ_1 in two subsamples. Graph A comprises the period from 2005 to 2012, while Graph B covers 2013–2019. The estimates are based on a two-component mixture model with only SMAs with at least 5 belief statements in each subsample. The number of SMAs in the first and second subsamples are 1,444 and 3,090, respectively.



information production/dissemination on Seeking Alpha. To proceed, I re-estimate the two-component mixture model in two subsamples. The first subsample comprises the first half of the sample, 2005–2012, while the second comprises the second half, 2013–2019. Figure 1 plots the fraction of high-type SMAs and their true ability in terms of average ABRs for the two subsamples. The fraction of high-type SMAs increased from 8% in the first subsample to 13% in the second, with the average ABR rising from 0.57% to 0.68%. This indicates that the fraction of high-type SMAs on Seeking Alpha improved over time, suggesting that the rise in Seeking Alpha's popularity among investors over the years potentially benefits investors. However, there are at least two reasons why the fraction of skilled SMAs on Seeking Alpha might improve with time: SMAs might have learned from experience, or more highly skilled individuals joined Seeking Alpha as it gained prominence. Indeed, analysis in Section IV.B points toward the second channel, as SMAs that have lasted much longer on Seeking Alpha—those that joined the platform early on—tend to be less skilled.

A natural concern is whether the foregoing results are driven by the modeling choices: the minimum number of belief statements required for computing the estimated skill ($\overline{ABR_i}$), its standard error (s_i), and the number of components in the mixture model. Robustness tests presented in the Supplementary Material and discussed in Section VI indicate that this is not the case. Across several robustness checks, I consistently find that over half of the individual SMAs are skilled and the exact fraction of skilled SMAs hovers closely around 56%.

B. SMA Characteristics and Skill

Since SMAs exhibit substantial heterogeneity in their ability, it is useful to understand which SMAs' attributes are associated with skill. I incorporate SMAs' attributes observable on Seeking Alpha in modeling the probability that an SMA is a high- or low-skill type using the logistic function:

TABLE 4

SMA Characteristics and Skill

Table 4 reports the results for the cross-sectional distribution of SMAs' skill from a two-component mixture model, where the proportion of SMAs in each component depends on an SMA characteristic, as shown in equation (4). The sample includes only SMAs with at least 5 belief statements. The table reports the estimated proportions, π_L and π_H, of the low- and high-type SMAs, respectively, for the below- ($x=0$) and above-median ($x=1$) SMAs for a given characteristic. Also shown in the table are the conditional mean of SMAs' true ability (Mean) and its standard deviation (SD), implied by the mixture model. Column headers indicate the SMA characteristics described in Table A1. SA Experience denotes experience on Seeking Alpha. ***, **, and * indicate the statistical significance of the 1-sided test of the difference between groups at the 1%, 5%, and 10% significance levels, respectively. The *p*-values are based on bootstrap distribution with 1,000 bootstrap replications.

Panel A. First Set of Characteristics

	Specialization							
	Industry		Firm		Workload		Skin-in-the-Game	
	$x=0$	$x=1$	$x=0$	$x=1$	$x=0$	$x=1$	$x=0$	$x=1$
Mean	0.0006	0.0021**	0.0008	0.0019***	0.0018	0.0010***	0.0009	0.0018***
SD	0.0139	0.0153***	0.0135	0.0136	0.0134	0.0133	0.0134	0.0133
π_L	0.9944	0.5895**	0.9566	0.7302***	0.7570	0.9267**	0.9373	0.7911***
π_H	0.0056	0.4105**	0.0434	0.2698***	0.2430	0.0733**	0.0627	0.2089***

Panel B. Second Set of Characteristics

	Effort		Disagreement		Popularity		SA Experience	
	$x=0$	$x=1$	$x=0$	$x=1$	$x=0$	$x=1$	$x=0$	$x=1$
Mean	0.0011	0.0015***	0.0016	0.0010***	0.0012	0.0015***	0.0015	0.0010**
SD	0.0133	0.0132	0.0132	0.0133	0.0133	0.0132	0.0132	0.0133
π_L	0.9136	0.8368***	0.8207	0.9244***	0.9035	0.8462**	0.7977	0.9163**
π_H	0.0864	0.1632***	0.1793	0.0756***	0.0965	0.1538**	0.2023	0.0837**

(4)
$$\pi_{i,0} = \frac{1}{1+\exp(b_0+b_1x_i)}; \quad \pi_{i,1} = 1-\pi_{i,0},$$

where x_i is a dummy variable that equals 1 if a certain SMA characteristic is above the cross-sectional median, and 0 otherwise. With this parameterization, the density function (1) has additional parameters b_0 and b_1, and the set of parameters in the maximum likelihood problem of equation (2) is now $\Theta=\{b_0, b_1, \mu_0, \mu_1, \sigma_0, \sigma_1\}$. Once these parameters are estimated, the probabilities $\pi_{i,0}$ and $\pi_{i,1}$ for SMAs with low and high values of the characteristic can be calculated using the estimates of b_0 and b_1.[13]

I consider the following SMA characteristics computed over the entire sample period: industry specialization, firm specialization, workload, effort, skin in the game, experience on Seeking Alpha, popularity on Seeking Alpha, and disagreement with other Seeking Alpha users. Table A1 describes the construction of these characteristics. Table 4 shows the results for SMAs' skills conditional on these characteristics. It reports, for each investor characteristic x, the fraction of high- and

[13] An alternative approach is to include all the SMA characteristics of interest jointly in equation (4). Doing so, however, complicates the calculation of the skill probabilities and related quantities conditional on a given characteristic since other characteristics have to be fixed as well. Nevertheless, one could focus on the estimated b_i for each characteristic to ascertain the relative contribution to the likelihood of being high skill. I conducted such analysis and summarized the estimated coefficients in the Supplementary Material Figure IF3, showing qualitatively similar characteristic relevance and direction of contribution as discussed below.

low-type SMAs, average ABR, and its standard deviation, implied by the mixture model conditional on whether x is above or below its cross-sectional median.

Supporting theoretical results on gains from specialization in information acquisition (e.g., Van Nieuwerburgh and Veldkamp (2010)), industry and firm specializations are the most distinctive characteristics that separate high- and low-type SMAs. For instance, SMAs specializing in a few industries (above median specialization) have a 41% probability of being high type compared to 0.5% for SMAs that cover many industries. The model-implied average ABR is 15 bps lower for SMAs with less industry specialization. Similarly, SMAs with a lower workload (below median average publications per year) have a 24% probability of being high type compared to 7% for SMAs with more publications per year, consistent with models of limited attention and cognitive capacity constraints (e.g., Hirshleifer et al. (2011)). These results indicate that SMAs who specialize less and have a heavier workload are less able to effectively process information to obtain informative signals.

SMAs who have stayed longer on Seeking Alpha (SA experience) have a 12-percentage-point lower probability of being high type compared to those with below-median years of Seeking Alpha experience. This suggests that the earlier results in Section IV.A, showing a higher fraction of high-type SMAs on SA in the second half of the sample, are driven by more skilled individuals who joined SA as the platform became more popular.

SMAs who often invest in the stocks they declare beliefs about (above median skin in the game) have a 14-percentage-point higher probability of being high type, with 9 bps higher performance. To the extent that SMAs' truthfully disclose their investments, this result suggests that more skin in the game motivates more diligent information acquisition and processing, leading to superior performance. Consistent with this view, Campbell et al. (2019) show that having an investment position in a stock does not impair the informativeness of opinions expressed by nonprofessional analysts. Table 4 further shows that SMAs who invest more effort (write longer articles), are more popular (receive more comments on their publications), and whose views other investors tend to disagree less with are also more likely to be high type.

The analyses indicate that some SMA characteristics, particularly firm and industry specializations, can help investors identify skilled SMAs. Furthermore, there are differences in how nonprofessional and professional analyst characteristics relate to skill. While the literature shows that professional analysts who issue more bearish recommendations tend to be more skilled (e.g., Barber, Lehavy, McNichols, and Trueman (2006)), my findings do not support this pattern among SMAs.

C. Persistence in Skill

We have seen that SMAs' skill is substantially heterogeneous and that certain individual characteristics relate to skill. However, it is essential to understand whether SMAs' ability is reasonably persistent so investors can potentially benefit from heeding their views. I address this by comparing an SMA's conditional probability of being a high type and her conditional expected skill in two non-overlapping subsamples.

For a given subsample, an SMA's conditional expected skill in return space can be inferred from the two-component mixture model as follows:

$$ESkill_i = \hat{\pi}_{i,0}\hat{\mu}_{i,0} + \hat{\pi}_{i,1}\hat{\mu}_{i,1}, \tag{5}$$

where $\hat{\pi}_{i,j}$ denotes the conditional probability that SMA i is from skill group j and $\hat{\mu}_{i,j}$ denotes the expected value of the SMA's skill conditional on the SMA belonging to skill group j. The following expressions give both quantities:

$$\begin{aligned}\hat{\pi}_{i,j} &= \text{Prob}(\text{SMA } i \text{ belongs to group } j | \hat{\alpha}_i, s_i, \Theta) \\ &= \frac{\pi_j \cdot \phi\left(\hat{\alpha}_i; \mu_j, \sigma_{i,j}\right)}{\pi_0 \cdot \phi(\hat{\alpha}_i; \mu_0, \sigma_{i,0}) + \pi_1 \cdot \phi(\hat{\alpha}_i; \mu_1, \sigma_{i,1})},\end{aligned} \tag{6}$$

$$\hat{\mu}_{i,j} = \left(\frac{1/\sigma_j^2}{1/\sigma_j^2 + 1/s_i^2}\right)\mu_j + \left(\frac{1/s_i^2}{1/\sigma_j^2 + 1/s_i^2}\right)\hat{\alpha}_i. \tag{7}$$

Once equations (5) and (6) are computed using parameter estimates from each subsample, respectively. Persistence can be assessed using an SMA-level cross-sectional regression. Equation (8) is estimated to determine whether an SMA's conditional skill measures in the first subsample predict the respective quantities in the subsequent subsample:

$$SkillMeasure_{i,s2} = \beta_0 + \beta_1 SkillMeasure_{i,s1} + \epsilon_{i,s2}, \tag{8}$$

where i indexes the individual SMA and $s1$ and $s2$ represent the first and second subsamples, respectively. *SkillMeasure* is either the conditional expected skill from equation (5) or the conditional probability of being a high type (i.e., $\hat{\pi}_{i,1}$) from equation (6).[14] In the specification, a positive and sizeable β_1 indicates persistence in SMAs' ability.

Figure 2 depicts the estimated β_1 coefficient from equation (8) and its 95% confidence intervals, indicating positive, sizeable, and statistically significant estimates. The first bar shows that β_1 is around 35% when *SkillMeasure* is the conditional probability of high type. This implies a 35% higher likelihood of being a high type in the second subsample if an SMA's probability of being high type approximately equals 1 in the first subsample. The second and third bars are for the case where *SkillMeasure* is the conditional expected skill. The third bar shows the economic magnitude of the estimate in the second bar by dividing the left- and right-hand sides of equation (8) by their means before the estimation. An increase in

[14]Unlike in the analysis of Figure 1, where the full sample is divided into 2 equal periods, here I split subsamples in a way that gives a roughly equal number of SMAs in the two subsamples. Specifically, the first subsample covers 2005–2014, while the second covers the remaining periods. This is because SMAs rarely last very long on the platform, leading to many people present at the beginning of the sample not remaining in the more recent periods. Consequently, if the sample is split into 2 equal periods, the number of SMAs simultaneously present in both samples will be very small, leading to a relatively low number of observations for estimating equation (8). Even with the slightly different sample split implemented here, we still lose a sizeable number of SMAs in the analysis, ending with 659 observations in the regression.

FIGURE 2

Persistence of SMAs' Ability Across Subsamples

Figure 2 shows the result of analyzing persistence in SMAs' ability across two nonoverlapping subsamples. Plotted are the estimated prediction coefficients and their 95% confidence intervals from regressing an SMA-specific conditional skill measure computed in the second subsample on the same measure computed in the first subsample, as shown in equation (8). The *x*-axis indicates the skill measure used for the estimation.

an SMA's conditional expected skill to the tune of the mean value in the first subsample predicts roughly 40% higher conditional expected skill relative to the mean in the second subsample.

Overall, there is some evidence of a reasonable level of persistence in SMAs' ability to express informative beliefs.

D. Portfolio Returns Based on SMAs' Beliefs

Can investors rely on individual SMAs' beliefs to form profitable portfolios, and can they do better by heeding SMAs with an ex ante higher skill? This question is addressed by examining simple transaction-based, calendar-time portfolios, as in Jeng, Metrick, and Zeckhauser (2003) and Seasholes and Zhu (2010). I form buy-and-hold bullish and bearish portfolios by putting a unit of stock k in the bullish (bearish) portfolio on day t whenever an SMA publishes a bullish (bearish) belief about stock k on $t-1$. The position is held for 1 month. Daily returns for each portfolio are value-weighted based on the number of units of each stock in the portfolio on day t and the closing stock prices on day $t-1$.[15]

I repeat the portfolio formation process conditional on SMAs' conditional expected skill based on equation (5) estimated using data up to the beginning of the last 2 calendar weeks before day t (i.e., $ESkill_{i,t-\tau}$). The 2-week gap between portfolio formation and the estimation of $ESkill_{i,t-\tau}$ ensures that all information used to estimate $ESkill_{i,t-\tau}$ is available before portfolio formation. The two-component mixture model is estimated using an expanding window at the end of

[15]Compared to daily rebalancing, using buy-and-hold portfolios for the analysis allows for realistic portfolio strategies, implementable at moderate trading costs in practice. Furthermore, buy-and-hold portfolios correct for noisy prices that can bias portfolio-based tests (see Blume and Stambaugh (1983)). Adding stocks to portfolios 1 day after the belief statement ensures investors have sufficient time to observe SMAs' beliefs and trade on them. It further avoids complications surrounding the time of day when beliefs are expressed.

https://doi.org/10.1017/S0022109024000930

each week, with the first window starting at the end of 2010 to allow for enough estimation data. Then, SMAs are ranked into three groups based on $ESkill_{i,t-\tau}$, with those in the top tercile categorized as high type and those in the bottom-tercile low type. I then examine the performance of these groups alongside that of all SMAs.

Table 5 summarizes the bullish and bearish portfolios' performance for All SMAs over the full sample (Panel A) and All SMAs over the period the conditional expected skill is estimated (Panel B). These "All SMAs" portfolios use all belief declarations by any SMA, including those SMAs excluded from the mixture model and expected skill estimation due to insufficient data. Panels C and D show portfolio results for the top-tercile and bottom-tercile SMAs, respectively. Panel A shows that the bullish (bearish) portfolio has an average excess return of 1.2% (0.67%) per month over the full sample. While the bullish portfolio has a positive and significant alpha between 0.32% and 0.44% per month, the bearish portfolio's alphas are statistically indistinguishable from 0. Across board, the bearish portfolios have fewer stocks than the bullish portfolios, as SMAs' beliefs are more bullish.

TABLE 5

Belief-Based Portfolio Performance

Table 5 reports average monthly excess returns (Exc. Ret.) and alphas for portfolios based on bullish and bearish SMA beliefs. For each SMA publication about stock k on day $t-1$, a unit of the stock is added to the bullish (bearish) portfolio on t if the belief is bullish (bearish). The position is then held for 1 month (21 trading days). Panel A shows results for portfolios based on all SMAs over the full sample. Panel B shows results based on all SMAs starting from 2011. Panel C (D) shows results for SMAs in the top (bottom) tercile of expected conditional skill ($ESkill_{i,t-\tau}$) from equation (5) estimated using data up to the beginning of the last 2 calendar weeks before t. Due to data limitation, $ESkill_{i,t-\tau}$ is estimated using expanding window. The first estimation at the end of 2010 uses data starting from 2005. The daily portfolio returns are cumulated to monthly frequency to compute excess returns and alphas. The column "Avg. No. Stocks" indicates the average number of unique stocks in each portfolio daily. Alphas are relative to the CAPM, 3-factor (FF3), 4-factor (Carhart), and 5-factor plus momentum (FF5 + MOM) benchmarks indicated in the column headers. Reported in parentheses are standard errors based on the Newey and West (1987) method.

			Alpha (%)			
	Avg. No. Stocks	Exc. Ret. (%)	CAPM	FF3	Cahart	FF5 + MOM
Panel A. All SMAs (full sample)						
Bullish	460	1.205***	0.444***	0.320**	0.321**	0.321**
		(0.256)	(0.149)	(0.152)	(0.152)	(0.147)
Bearish	117	0.669	−0.152	−0.279	−0.218	−0.131
		(0.410)	(0.247)	(0.202)	(0.199)	(0.240)
Panel B. All SMAs (from 2011)						
Bullish	620	1.373***	0.341**	0.197	0.180	0.207*
		(0.146)	(0.151)	(0.146)	(0.139)	(0.118)
Bearish	158	1.103***	−0.035	−0.126	−0.003	0.122
		(0.345)	(0.348)	(0.308)	(0.297)	(0.278)
Panel C. Top-Tercile SMAs (from 2011)						
Bullish	249	1.398***	0.376***	0.219*	0.200	0.257**
		(0.160)	(0.140)	(0.131)	(0.126)	(0.123)
Bearish	62	1.234***	0.122	−0.013	0.198	0.390
		(0.428)	(0.436)	(0.368)	(0.327)	(0.290)
Panel D. Bottom-Tercile SMAs (from 2011)						
Bullish	252	1.204***	0.159	0.051	0.060	0.084
		(0.126)	(0.127)	(0.124)	(0.114)	(0.082)
Bearish	53	1.067***	0.127	0.045	0.181	0.317
		(0.403)	(0.375)	(0.381)	(0.383)	(0.355)

https://doi.org/10.1017/S0022109024000930

Panel B of Table 5 shows qualitatively similar results for all SMAs in the more recent sample period. However, only alphas relative to the CAPM and 5-factor plus momentum benchmarks are significant. There are at least two plausible reasons for the decline in the significance in the more recent periods: i) As SA gained prestige, market participants likely responded quickly to SMAs' publications, leading to their views being reflected quicker in stock prices. ii) The decline in significance may also arise from low test power since the sample only starts from 2011. Overall, results in Panels A and B are consistent with Chen et al. (2014) who also observe significant alphas based on article tone.

Turning to the portfolios of the top-tercile (Panel C) and bottom-tercile (Panel D) SMAs, we observe noticeable differences in the 2 groups' performance. The alphas for both groups' bearish portfolios are once again insignificant. While the bottom-tercile's bullish portfolio alphas are also insignificant and relatively small, those of the top-tercile SMAs are mostly significant and sizeable. Top-tercile SMAs' alpha relative to CAPM (5-factor + momentum) is 0.38% (0.26%) per month, which corresponds to a 3.1%–4.5% annualized alpha.

Examining alphas for the difference portfolios, "All SMAs minus Top-tercile SMAs," "All SMAs minus Bottom-tercile SMAs," and "Top-tercile minus Bottom-tercile SMAs," shown in Supplementary Material Table IF12, we do not observe significant alphas for the bearish beliefs' difference portfolios. In contrast, while alphas for the bullish beliefs' "Top-tercile minus Bottom-tercile SMAs" difference portfolio are positive, significant, and sizeable, those of "All SMAs minus Top-tercile SMAs" are mostly insignificant.

Given that individual investors primarily follow SMAs on social media, it is interesting that a long-only portfolio with a decent holding period and, hence, modest portfolio turnover and transaction cost yields significant ABRs.[16] Therefore, investors could trade on SMAs' belief statements with modest gains. However, because the performance of the top-tercile SMAs is quite comparable to that of all SMAs, it follows that there is value in pooling all SMAs and relying on the wisdom-of-the-crowd. The wisdom-of-the-crowd can yield informative signals on average when views are heterogeneous, even though individuals do not know more than their peers. In this case, however, the wisdom-of-the-crowd benefits from SMAs' skill besides just averaging out noise in the absence of skill.

V. Heuristics and SMAs' Beliefs

A natural question arises from the preceding results: How do low- versus high-type SMAs form their beliefs? Do they rely on heuristics, and does it depend on one's type, as economic theory suggests? I examine two common behavioral patterns that manifest in investors' beliefs: return extrapolation and herding.

[16]To further address the issue of transaction costs, I conduct robustness tests, discussed in Section VI, by excluding penny stocks and microcap stocks, which are traded infrequently and hence pose high transaction costs. The results are qualitatively similar.

https://doi.org/10.1017/S0022109024000930

A. Evidence on Return Extrapolation

Return extrapolation is the idea that people's expectation of a stock's future return is a weighted average of its past returns, with the weights on the past returns positive and higher for recent past returns. Theoretical models (De Long et al. (1990), Barberis and Shleifer (2003)) formalize how extrapolative beliefs can arise from heuristics and their implications. Recently, Da et al. (2021) show that extrapolative beliefs from another social forum predict stock returns with the wrong sign. On the contrary, the analyses in previous sections of this paper indicate that SMAs' beliefs correctly provide value-relevant information, raising the issue of whether SMAs extrapolate while still having informative beliefs.

To test this, I regress beliefs about stock k on the stock's past nonoverlapping weekly returns over the past 3 months:

$$\text{(9)} \qquad Belief_{i,k,t} = \beta_0 + \sum_{\tau=1}^{12} \beta_\tau Ret(\tau)_{k,t} + \mathbf{X}\,\boldsymbol{\Gamma} + \epsilon_{i,k,t},$$

where $Belief_{i,k,t} \in \{-1, 0, 1\}$ is SMA i's belief about stock k on day t. The variable equals -1 if beliefs are bearish, 0 if neutral, and 1 if bullish. $Ret(\tau)_{k,t}$ is stock k's past τ'th nonoverlapping 1-week (5 trading days) return, with the most recent return window ending 2 days before the belief statement day t.[17] $\mathbf{X}$ is a vector of stock-specific control variables that might influence beliefs, namely lagged average belief, cash flow news tone averaged over the past week ending $t-2$, and professional analysts' consensus forecast of quarterly earnings as of the last calendar month. Missing values for lagged average belief and cash flow news tone are replaced with the neutral value of 0 and 0.5, respectively. The controls also include characteristics that capture a stock's attractiveness to individual investors: market beta, log of market capitalization, stock price, idiosyncratic volatility, and idiosyncratic skewness. The regression includes year-month fixed effects to absorb common trends.

Figure 3 plots the coefficient estimates and 95% confidence intervals for the lagged weekly returns, $Ret(\tau)$. The most recent 1-week return has the largest influence on beliefs, with a relatively tight 95% confidence interval. More so, the effect of past returns generally declines with time, consistent with return extrapolation in SMAs' belief formation. In terms of economic magnitude, the coefficient of 0.31 for the most recent 1-week return implies that belief becomes significantly more bullish by roughly 40% of its standard deviation when a stock's price doubles over the past 1 week. Conversely, the influence of older returns is much lower, becoming indistinguishable from 0 by 2 months.[18]

The degree of return extrapolation might depend on an SMA's skill. On the one hand, high-type SMAs might bet against extrapolators to profit from the potential

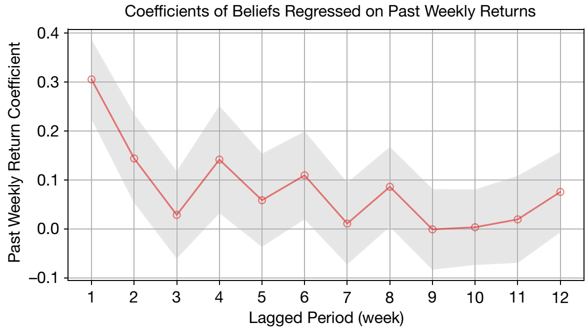
[17]Calculation of past weekly returns and control variables ends at least 2 days before the belief publication day to ensure that the variables were observable by SMAs and hence could guide belief formation.

[18]Table IF15 in the Supplementary Material shows the estimation results with and without control variables, as well as subsample analysis, all supporting the extrapolation hypothesis.

FIGURE 3

Return Extrapolation in SMAs' Belief Formation

Figure 3 shows coefficient estimates and 95% confidence intervals (grey area) from regressing SMA beliefs, $Belief_{i,k,t}$, on stocks' past nonoverlapping weekly (5 trading days) returns, as described under equation (9). Control variables included in the regression are described under equation (9). Standard errors are clustered by stock and year-month.



mispricing they generate. On the other hand, high-type SMAs could try to front-run extrapolators by i) reacting quicker to recent past returns and ii) allowing the impact of these returns on their beliefs to die out quicker, making it seem as though the high-type SMAs have a higher extrapolation intensity. SMAs' return extrapolation might also differ across stocks depending on the salience of returns. For example, extrapolation might be stronger for stocks with high media coverage, making past returns easy to observe and recall.

To explore these potential sources of heterogeneity, I use an exponential decay specification (as in Cassella and Gulen (2018), Da et al. (2021)) to summarize the degree of extrapolation conditional on SMAs' skill and stocks' news coverage as follows:

$$Belief_{i,k,t} = \lambda_0 + \lambda_{1,l} \cdot \sum_{\tau=1}^{12} 1_l \cdot w_{l,\tau} Ret(\tau)_{k,t} + \lambda_{1,h} \cdot \sum_{\tau=1}^{12} 1_h \cdot w_{h,\tau} Ret(\tau)_{k,t} + \mathbf{X\Gamma} + \epsilon_{i,k,t}, w_{l,\tau} = \frac{\lambda_{2,l}^{\tau-1}}{\sum_{j=1}^{12} \lambda_{2,l}^{j-1}}, w_{h,\tau} = \frac{\lambda_{2,h}^{\tau-1}}{\sum_{j=1}^{12} \lambda_{2,h}^{j-1}}, \quad 0 \le \lambda_{2,l} < 1 \text{ and } 0 \le \lambda_{2,h} < 1, \tag{10}$$

where $Belief_{i,k,t}$, $Ret(\tau)_{k,t}$, and $\mathbf{X}$ are as defined under equation (9). Subscripts l (h) index either low (high) SMA skill type or stock news coverage level. Hence, 1_l is an indicator function that equals 1 if a given SMA or stock characteristic is below its median. Conversely, 1_h is an indicator function that equals 1 if the characteristic is above its median. $\lambda_{2,j}$ governs the importance of recent returns relative to older returns in shaping beliefs about stocks' future returns conditional on SMAs' skill type or stocks' news coverage. $\lambda_{2,j}$ closer to 0 suggests that recent returns primarily

https://doi.org/10.1017/S0022109024000930

influence beliefs compared to older returns—that is, the impact of past returns on beliefs dies out quickly. In contrast, $\lambda_{2,j}$ close to 1 indicates that SMAs give roughly the same weight to older and recent returns. $\lambda_{1,j}$ captures the overall extent to which SMAs' beliefs respond to past returns. Da et al. (2021) note that $\lambda_{1,j}(1-\lambda_{2,j})$ is an appropriate summary for the degree of extrapolation.

I use three proxies estimated over the full sample to capture SMAs' skill: conditional expected skill from equation (5), conditional probability of being high type from equation (6), and an indicator variable that equals 1 if both variables are above their medians. I measure a stock's news coverage (salience of returns) as the number of news articles published about the stock over the past week ending $t-2$ and then compare it to the monthly median.

Table 6 shows the nonlinear least squares estimation results. Columns 1–6 show the estimated extrapolation coefficients conditional on low versus high SMA skill. Across board, more skilled SMAs have a higher extrapolation intensity, captured by a lower λ_2 and summarized in the last row by $\lambda_1(1-\lambda_2)$. For example, focusing on columns 5 and 6, we see that SMAs with above median skill level based on both the conditional probability of high type and conditional expected skill have $\lambda_2 = 0.506$ versus 0.866 for their lower-skilled peers. This implies that for the above (below) median skill SMAs, the most recent 1-week return has about 8 (2) times the influence of the 4th-week return in shaping beliefs. Hence, the impact of past returns on beliefs dies out quicker for the higher-skilled SMAs.

At face value, it may seem contradictory that high-type SMAs have higher extrapolation intensity. However, this is consistent with high-type SMAs front-running their lower-skilled peers, potentially profiting from the return momentum sustained by the latter and other market participants who have a higher λ_2 and hence correct their mispricing relatively much slower. Therefore, high-type SMAs' skills likely arise from their ability to integrate relevant cash flow information with

TABLE 6

Return Extrapolation, Skill, and Salience: Nonlinear Least Squares

Table 6 reports the results for return extrapolation conditional on SMAs' skill and the salience of stock returns based on equation (10). Columns 1 and 2 present results for the case where skill is based on the conditional probability of high skill (equation (6)). In columns 3 and 4, skill is based on the conditional expected skill (equation (5)). In Columns 1–4, an SMA is categorized in the High (Low) group if their value for the denoted skill measure is above (below) the median. Columns 5 and 6 combine both the conditional probability of high skill and the expected skill to classify an SMA in the High group if the SMA is above the median for both variables and Low otherwise. Columns 7 and 8 present results for extrapolation conditional on return salience. A stock is categorized as having High (Low) salient returns if the number of news articles published about the stock over the past week is above (below) the monthly median. Control variables included in the regressions are described under equation (9). Standard errors are shown in parentheses.

	Prob. High Skill		Exp. Skill		Prob. High Skill × Exp. Skill		Stock News Coverage	
	Low	High	Low	High	Low	High	Low	High
	1	2	3	4	5	6	7	8
λ_0	0.464		0.463		0.464		0.473	
	(0.022)		(0.022)		(0.022)		(0.021)	
λ_1	1.448	0.697	1.472	0.760	1.339	0.729	1.094	1.261
	(0.124)	(0.163)	(0.145)	(0.122)	(0.115)	(0.205)	(0.131)	(0.158)
λ_2	0.884	0.565	0.895	0.590	0.866	0.506	0.954	0.710
	(0.024)	(0.117)	(0.028)	(0.076)	(0.025)	(0.153)	(0.034)	(0.047)
Obs.	169,030	169,030	169,030	169,030	169,030	169,030	176,237	176,237
$\lambda_1(1-\lambda_2)$	0.167	0.303	0.155	0.311	0.180	0.360	0.051	0.365

identifying, broadcasting, and profiting from momentum trends earlier than their peers. This interpretation lines up with the models of De Long et al. (1990) and Barberis and Shleifer (2003), where sophisticated investors who understand the existence of extrapolators in the market can ride bubbles (i.e., exacerbate mispricing) in a manner that allows them to profit from the future buying of other naive extrapolators who correct their mispricing less quickly than the sophisticated investors.[19]

Turning to heterogeneity in extrapolation conditional on low versus high news coverage, columns 7–8 of Table 6 show that SMAs extrapolate more on the past returns of stocks that received higher news coverage. $\lambda_2 = 0.64$ (0.95) for stocks with above (below) median news coverage. Summarizing extrapolation intensity by $\lambda_1(1-\lambda_2)$, we see that the intensity is 7 times larger for stocks with above median news coverage and hence have more salient returns. This corroborates existing studies (e.g., Alok, Kumar, and Wermers (2020), Tversky and Kahneman (1973)) suggesting that salience fuels certain biases.

Overall, SMAs extrapolate from past returns when forming beliefs, and the impact of past returns on beliefs dies out quickly for the more skilled SMAs. However, as previous sections show, the extrapolation does not render SMAs' beliefs systematically wrong, as they remain informative about future stock returns and cash flow. These results differ from that of Da et al. (2021), documenting systematically wrong beliefs for extrapolators. A potential explanation for the divergent findings is that, unlike the individuals in Da et al. (2021), SMAs tend to be more sophisticated than the average retail investor and are, therefore, better equipped to combine past return trends with fundamental information when forming expectations. The texts accompanying SMAs' belief declarations provide fundamental arguments supporting their beliefs. I revisit this point in Section V.C by decomposing belief drivers into behavioral and fundamental components. There, we see that fundamental information plays a prominent role in shaping SMAs' beliefs. Overall, the results highlight the usefulness of analyzing the activity and beliefs expressed on different social platforms.

B. Evidence on Herding

Social media can serve as a coordination mechanism for mutual imitation (i.e., herding), because it quickens information transmission and enhances the ability to observe actions of others. Hence, SMAs may herd in stating their beliefs, in line with the reputational herding and information cascade models (e.g., Banerjee (1992), Bikhchandani et al. (1992), and Scharfstein and Stein (1992)) or naive (irrational) herding models (e.g., Eyster and Rabin (2010)). It is equally possible that herding is less pervasive among SMAs, given their incentive to attract followers. Ultimately, herding might intensify mispricing if it is naively based on little or no information or promote price discovery if it is caused by fundamental information.

[19]Accordingly, Cassella and Gulen (2018) show that for the aggregate market, when λ_2 is low, mispricing is corrected much quicker compared to when λ_2 is high, as investors more quickly forget the positive return that made them excited in the first place.

To test for herding, I adopt the herding test of Welch (2000), which is appropriate in settings where choices are discrete (e.g., bullish, neutral, and bearish belief statements). The test estimates a parameter θ that governs whether belief statements depend on the prevailing consensus. If $\theta = 0$, new beliefs are independent of the consensus—the null hypothesis. Conversely, $\theta > 0$ indicates a tendency for beliefs to follow the consensus, while $\theta < 0$ suggests a tendency to avoid the consensus. Supplementary Material Section ID.2 describes the test and estimation procedure in detail.

Table 7, Panel A, shows the estimated herding coefficient $\hat{\theta}$ and its χ^2 p-value for the equal-weighted and characteristic-weighted consensus (average) beliefs. The estimation uses all SMAs' belief revisions made within a year to avoid stale beliefs, and the consensus is calculated only if there are at least 2 belief statements by other SMAs over the 6-month period prior to an SMA's revision. The estimated coefficients are all significant and around 0.28, suggesting that SMAs herd toward the consensus when stating their beliefs. Furthermore, the views of the more likely skilled SMAs do not disproportionately influence belief revisions.

Panel B of Table 7 shows the economic implications of the estimated herding coefficient. The column with $\hat{\theta} = 0$ captures the unconditional probabilities of an SMA stating a bearish, neutral, or bullish revision. Comparing the column $\hat{\theta} = 0$ with the column $\hat{\theta} = 0.25$, which is close to the estimated herding coefficients, we can infer that herding increases the probability of declaring a belief that matches a hypothetical bullish (bearish) belief consensus by about 5 (6) percentage points. That is, with a herding coefficient of 0.25, the probability of stating a bearish belief when the consensus is bearish increases from 0.159 (when $\hat{\theta} = 0$) to 0.218 (when $\hat{\theta} = 0.25$). This indicates a moderate level of herding, which is interesting since SMAs can instead deviate completely from the consensus in an attempt to attract attention and readership.

TABLE 7

SMAs' Herding

Table 7 shows the herding test results in Panel A and the economic significance of herding in Panel B. Panel A shows the estimated herding coefficient, $\hat{\theta}$, and χ^2 p-value for different consensus estimates (targets). Panel B shows the probability of a belief revision hitting a hypothetical bearish, neutral, or bullish target for different $\hat{\theta}$ values. If $\hat{\theta} = -\infty$, the target will always be avoided. If $\hat{\theta} = 0$, the probability of hitting the target is equal to the unconditional probability of hitting the target. If $\hat{\theta} = \infty$, the target will always be hit. Values in Panel B are based on the unconditional belief transition probability matrix and hypothetical values for $\hat{\theta}$ shown on the column headers.

Panel A. Estimated Herding Coefficient

Consensus is	$\hat{\theta}$	$\chi^2 p-value$
Equal-weighted	0.280	0.000
Specialization-weighted	0.273	0.000
Effort-weighted	0.283	0.000
Popularity-weighted	0.280	0.000

Panel B. Probability of Hitting Target

	Herding Coefficient $\hat{\theta}$							
Target	–10	–1	0	0.15	0.25	0.5	1	10
1 (Bearish)	0.000	0.037	0.159	0.193	0.218	0.292	0.469	1.000
2 (Neutral)	0.000	0.019	0.037	0.041	0.044	0.052	0.072	0.975
3 (Bullish)	0.000	0.481	0.804	0.835	0.854	0.892	0.941	1.000

TABLE 8

Conditional Herding

Table 8 reports results for the herding test, conditional on the realization of some variable, y, such that the estimated herding coefficient $\hat{\theta}=\hat{\theta}_0+\hat{\theta}_1 y$. Panel A shows results where y is an indicator variable of high-skilled SMA. Panel B reports results where y is CC, measured as $CO \times Ret(h)$. CO stands for consensus optimism, measured as the equal-weighted consensus minus 2, and $Ret(h)$ is the future stock return over horizon $h \in \{5, 63, 126\}$ trading days starting $t+1$.

Panel A. SMA Skill

	Prob. High Skill		Exp. Skill		Prob. High Skill × Exp. Skill	
	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$
Estimate	0.286	−0.050	0.286	−0.016	0.300	−0.301
χ^2 *p* − *value*	0.000	0.069	0.000	0.384	0.000	0.000

Panel B. Cons. Correctness

	$CO \times Ret(h=5)$		$CO \times Ret(h=63)$		$CO \times Ret(h=126)$	
	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_0$	$\hat{\theta}_1$
Estimate	0.280	0.476	0.279	0.112	0.283	−0.014
χ^2 *p* − *value*	0.000	0.005	0.000	0.028	0.000	0.689

When is Herding More Pronounced?

To further understand SMAs' herding, let θ be a function of some variable y (i.e., $\theta(y)=\theta_0+\theta_1 y$). Then, if $\theta_1=0$, herding does not depend on y. Conversely, $\theta_1>0$ indicates that herding increases with y, while $\theta_1<0$ indicates that herding decreases with y.

Table 8 shows results for the estimated $\hat{\theta}_0$ and $\hat{\theta}_1$ and the associated χ^2 *p*-values. Panel A shows results where y is an indicator variable that captures whether the SMA revising her belief has above-median skill measure based on the three measures used in Table 6. Panel B shows results where y is a measure of consensus correctness (CC), quantified as consensus optimism (CO) times future stock return over horizon h starting $t+1$: $CC=CO \times Ret(h)$. CO is measured as the equal-weighted consensus minus 2, which is positive (negative) if the consensus is optimistic (pessimistic). CC is positive if the consensus is correct ex post, that is, an optimistic (pessimistic) consensus is followed by a positive (negative) future stock return.

Table 8, Panel A, shows that the incremental herding coefficient, $\hat{\theta}_1$, is negative, suggesting that higher-skilled SMAs tend to herd less. For example, focusing on the last 2 columns, we see that $\hat{\theta}_1=-0.30$ for SMAs simultaneously above the median in the conditional probability of high skill and the conditional expected skill measures, rendering their total herding coefficient close to 0. This is consistent with Bikhchandani et al.'s (1992) model, which implies more deviation from the consensus for high-ability agents.

Panel B shows that except for one future return horizon, $\hat{\theta}_1$ for CC is consistently positive and significant. This implies stronger herding when the consensus is correct. In terms of economic magnitude, the standard deviation of $CO \times Ret(h)$ for the 5-day and 63-day horizons are 5% and 17%, which implies an increase in herding by 7%–9% of its unconditional value for a standard deviation increase in CC. The evidence that the incremental herding when the consensus is correct does not significantly reverse sign even for the 6-month return horizon suggests that SMAs tend to herd on fundamental information incorporated in the consensus rather than just naive herding, as in Eyster and Rabin (2010). Therefore, SMAs likely learn fundamental information from their peers' publications to improve their forecasts, consistent with information-based herding models (e.g., Banerjee (1992)).[20]

C. Heuristics Versus Fundamental Information in Beliefs

Having documented that return extrapolation and herding shape SMAs' beliefs about future stock returns, I now jointly quantify the respective contributions of extrapolation, herding, and cash flow information in shaping SMAs' beliefs in the spirit of Fuster, Laibson, and Mendel's (2010) model, where beliefs are a weighted average of rational forecasts and heuristics-based forecasts. Building on the preceding results, I use the following specification:

$$
\begin{aligned}
(11)\qquad \mathit{Belief}_{i,k,t} = {} & \beta_0 + \beta_1 \mathit{WeightedPastReturn}_{k,t} + \beta_2 \mathit{ConsensusBelief}_{k,t} \\
& + \beta_3 \mathit{EarningsEstimate}_{k,t} + \beta_4 \mathit{CashFlowNewsTone}_{k,t} \\
& + \mathbf{X\Gamma} + \epsilon_{i,k,t}.
\end{aligned}
$$

$\mathit{Belief}_{i,k,t}$ and $\mathbf{X}$ are as defined under equation (9). $\mathit{WeightedPastReturn}_{k,t}$ is a weighted sum of the nonoverlapping weekly returns used in equation (10), with weights based on exponential decay parameters $\lambda_1 = 1.12$ and $\lambda_2 = 0.80$ estimated from a version of equation (10) that does not condition on any characteristic. $\mathit{ConsensusBelief}_{k,t}$ is the equal-weighted consensus SMAs' belief used in Table 7, which captures herding tendencies. $\mathit{EarningsEstimate}_{k,t}$ and $\mathit{CashFlowNewsTone}_{k,t}$ are proxies for fundamental news. The former is the average of professional analysts' earnings per share estimate. $\mathit{CashFlowNewsTone}_{k,t}$ is the average tone of fundamental-relevant news published in the media about stock k over the past week, which increases in tone positivity. The dependent variable and regressors are normalized to unit variance to make the estimated coefficients comparable.

The coefficients of interest are β_1, which captures the contribution of extrapolation in shaping beliefs, β_2, which captures the contribution of herding tendencies, and $\beta_3 + \beta_4$, which captures the combined contribution of fundamental news. equation (11) is estimated at the end of each month using data over the past year, obtaining a time series of the coefficients. Figure 4, Graph A, shows the evolution of the estimated coefficients over time, indicating that extrapolation, herding, and fundamental news contribute incrementally and positively in shaping SMAs' belief bullishness. There is a noticeable time-series variation in how much each component matters. Return extrapolation's contribution declined over 2008–2009, while the contribution of fundamental news increased sharply during this period. On the other hand, herding's contribution remained relatively stable. Overall, herding contributed the most in shaping SMAs' beliefs, followed by fundamental news and then extrapolation.

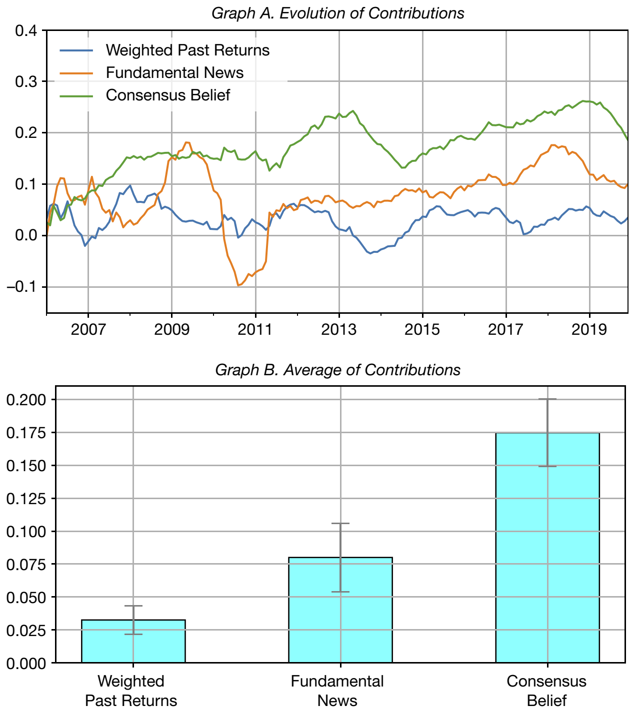
Graph B of Figure 4 indicates that, on average, a standard deviation increase in weighted past returns—the proxy for return extrapolation—is associated with

[20]An alternative explanation, which cannot be fully differentiated, is that SMAs independently follow the same fundamental information. There is also the possibility that the consensus moves prices. I, however, favor the fundamental information story because, as shown in the Supplementary Material Section IC, SMAs' aggregate beliefs are informative about firms' future cash flows.

FIGURE 4

Fundamental Versus Non-Fundamental Belief Drivers

Figure 4 Graph A depicts the time series of coefficients from estimating equation (11) at the end of each month using data over the past year. Graph B shows the average of the coefficients' time series and the 95% confidence intervals based on the Newey and West (1987) standard errors. Fundamental News is the sum of coefficients of *EarningsEstimate* and *CashFlowNewsTone* (i.e., $\beta_3 + \beta_4$).



SMAs' belief becoming significantly more bullish by roughly 3% of its standard deviation (SD). Similarly, beliefs become more bullish by 8% of the SD per one SD improvement in fundamental news versus 18% relative increase in bullishness per one SD increase in consensus bullishness.

In sum, herding, which may be naive but more likely information-based, heuristics such as representativeness and the law of small numbers, manifesting in return extrapolation, and fundamental news all play crucial roles in shaping SMAs' beliefs. While herding appears to dominate the latter two, the sizeable importance of fundamental information sources in shaping beliefs likely explains why SMAs' expectations are informative despite the presence of behavioral components in them.

VI. Additional Results and Robustness

This section discusses additional results and several robustness analyses collected in the Supplementary Material.

https://doi.org/10.1017/S0022109024000930

https://doi.org/10.1017/S0022109024000930

SMAs' Skill and Trading Volume. The heterogeneity in skill across SMAs raises the question of whether, compared to their peers, the more skilled SMAs generate more volume and volatility following their publications. More trading volume and volatility following higher-skilled SMAs' publications could suggest that i) skilled SMAs' publications tend to attract more investors to the focal stock, leading to more trading and price movements, and ii) skilled SMAs use their superior skills to uncover fundamental news before the information is made public, leading to higher trading and volatility after the subsequent announcement. While I do not pursue which of these potential channels is predominant due to space limitation, the results reported in Table IF2 indicate that, after controlling for other potential drivers, abnormal trading volume and volatility are significantly higher over the next week following publications by SMAs with above median skill level. In terms of economic magnitude, abnormal volume (abnormal volatility) is higher by roughly 5% (7%) relative to its SD when a publication is authored by an SMA with an above-median conditional probability of being high type.

Excluding Earnings Announcement Days. Earnings announcement days are well-known information events associated with significant market activity. A potential concern with this paper's analysis is that SMAs could cluster their publications around earnings announcements making it challenging to differentiate their identified skill from the market moves associated with earnings announcements. Although the analysis of Table IC1 tried to account for this by directly controlling for cash flow news tone, and Table IC2 shows that SMAs' beliefs even predict earnings announcement surprises, some doubts may remain.

Therefore, I conduct additional tests by excluding all SMAs' publications within n days centered around an earnings announcement day. Interestingly, only a small fraction of SMAs' views are published around earnings announcement days: roughly 4% on earnings announcement days and 17% within 7 days centered on an earnings announcement day. Next, the two-component mixture model is re-estimated, excluding all articles published within $n \in \{3, 7, 11\}$ days window around an earnings announcement. Table IF3 tabulates the results, indicating a consistent picture as in the main analysis. Across board, over half of SMAs are skilled and express beliefs that align with future ABRs. In addition, as in the main analysis, most SMAs belong to the low-type group, with the high-type fraction ranging from 9% to 11%. Overall, publications around earnings announcements do not materially drive the main results.

Alternative Mixture Model Setup and Data Requirement. The analysis of Section IV uses a two-component mixture model and only SMAs that have at least 5 belief statements over the entire sample. To ensure that this specific setup does not drive the results, Table IF4 considers an alternative setup with three components in the mixture model. Again, most SMAs (66%) belong to Component 0 (lowest skill group), while Component 1 (medium) and Component 2 (highest skill group) comprise 29% and 5% of SMAs, respectively. Overall, the three-component mixture model indicates that 60% of SMAs are skilled, comparable to the 56% in the main analysis using the two-component mixture model.

Table IF5 further reports the results for the two-component mixture model using data for only SMAs with at least 10 (instead of 5) belief statements over the sample period. Although this more stringent data requirement reduces the number of SMAs in the cross-section by about half, we still find a qualitatively similar result as in the main analysis.

Finally, Tables IF6 and IF7 show the counterparts of Table 3 in the main text and Table IF4, respectively, but now using CAPM instead of the 3-factor model as the benchmark for SMAs' estimated ABRs. The fraction of high-type SMAs and the proportion of SMAs with positive ABRs are very similar in both settings, indicating that the results are robust to alternative ABR benchmarks. Overall, the robustness analysis shows that alternative modeling and estimation choices do not significantly influence earlier results on the cross-sectional distribution of SMAs' skill.

Informativeness of Beliefs Using Only Explicitly Stated Beliefs. Although several model validation exercises demonstrate that the trained ML model classifies beliefs with a high degree of accuracy out-of-sample and has intuitive feature weights, some doubts may remain regarding whether the ML model somehow drives the results in this paper. To address this concern, the ability of SMA beliefs to predict future ABRs is estimated using only explicitly stated beliefs (i.e., ignoring beliefs extracted using ML). Although this leads to a smaller sample size, Table IF8 shows that aggregate SMA beliefs' predictability of ABRs holds. In particular, the coefficient of $AggBelief_{k,t}$ is significant for all horizons and is comparable to the results in Table IC1, which uses both the stated and ML-inferred beliefs.

I conduct additional robustness tests using only the stated beliefs but now averaging the beliefs over the past 2 weeks and 1 month, respectively. Precisely, for each firm-date observation of $AggBelief_{k,t}$, $\overline{AggBelief}_{k,t}$ is calculated by averaging $AggBelief_{k,t}$ over the window $t-h$ to t. Panels A and B of Table IF9 show the results for $h = 14$ and 30 days, respectively, indicating that averaging SMAs' beliefs over the past reduces the significance of the ABR predictability for the CAPM benchmark and future 63-day horizon. Nonetheless, the results are qualitatively comparable to Table IC1. Conducting similar exercises with the stated and extracted beliefs combined produces similar insights, as Table IF10 shows.

Transaction-Based Calendar-Time Portfolios Excluding Small Stocks. To demonstrate that small stocks do not drive the results of portfolios formed on SMAs' beliefs, the portfolio analysis is conducted excluding penny stocks (price less than \$5) and microcap stocks (market capitalization less than the 2nd NYSE decile), respectively. Precisely, a stock is not included in a portfolio on day t if on day $t-1$, when beliefs are published, the stock's price is less than \$5, or its market capitalization is less than the 2nd NYSE decile. Once a unit of a stock is in a portfolio, it is held until the end of the holding period, regardless of whether the stock failed to meet the inclusion criteria on some dates during the holding period. Table IF13 shows the results for the exclusion of penny stocks, while Table IF14 shows the results for the exclusion of microcap stocks. In both cases, the bullish portfolio returns are qualitatively similar to those in the main analysis across different benchmarks, with lower statistical significance in some instances. Conversely, alphas for the bearish portfolio are statistically indistinguishable from 0, corroborating the results in the main analysis.

Extrapolation in Subsamples. Table IF15 shows that the analysis of SMAs' return extrapolation holds across subsamples. Columns 3 and 4 show the results for the first and second half of the sample, respectively. In both cases, the 2 most recent 1-week returns have the strongest and most statistically significant influence on SMAs' beliefs. The influence of older 1-week returns is mostly statistically insignificant.

Herding Test with Consensus Based on Alternative Window. Finally, the main analysis on SMAs' herding is based on consensus computed over the past 6-month period $t-180$ to $t-2$. Tables IF16 and IF17 show the robustness results, where the consensus is instead computed over the past 3-month period $t-90$ to $t-2$, indicating that the window used in computing the consensus does not significantly influence the results. In particular, the estimated herding coefficient is close to the 0.28 obtained in the main analysis, with the results on conditional herding also qualitatively similar to the main analysis.

VII. Conclusion

This paper uses natural language processing techniques to infer nonprofessional SMA beliefs about a large cross-section of stocks from opinions expressed on the popular social finance forum, Seeking Alpha (SA). The paper studies the distribution of SMAs' skill and how two common behavioral patterns—return extrapolation and herding—interact with SMAs' skill in shaping their beliefs.

On average, SMAs' belief statements contain value-relevant information. However, substantial heterogeneity exists in SMAs' ability: while over half of SMAs are skilled and express beliefs that generate modest positive abnormal stock returns, only 13% of SMAs belong to the high-type group that declare beliefs that produce a much larger 1-week 3-factor ABR of 61 bps. The analysis of portfolios formed on SMAs' beliefs and ex ante skills indicates that investors could profitably trade on SMAs' forecasts. However, the performance of the top-tercile SMAs' portfolio is quite comparable to the portfolio that follows all SMAs, indicating value in the wisdom-of-the-crowd. Since a sizeable fraction of SMAs are skilled, it follows that the wisdom-of-the-crowd effect here does, in fact, benefit from SMAs' skill in addition to averaging out noise.

There are behavioral patterns in SMAs' belief formation. On the one hand, SMAs tend to extrapolate from past stock returns and herd on the views of their peers. On the other hand, higher-skilled SMAs appear to i) front-run their peers whose extrapolative beliefs depend more on older returns and ii) herd less on the consensus, both in line with existing theories. Crucially, the extrapolation and herding tendencies do not result in systematically wrong beliefs, likely due to the distinct and sizeable role fundamental information plays in shaping SMAs' expectations. Therefore, individuals can form informative expectations even though the belief formation process is not entirely consistent with rational models.

In light of the concerns surrounding social media's growing influence over financial markets, this paper suggests that SMAs on SA are useful information intermediaries. Most SMAs are skilled, and investors seem to benefit from their expectations since users engage more with the higher-skilled SMAs' views. Although encouraging, this conclusion may not readily extend to other platforms with different contributor incentives and design features that impact contributor quality and the effort devoted to information production.

Appendix

TABLE A1

Variable Definitions

Variable	Definition
Stock-Level Variables	
$ABR_{k,t+1\rightarrow t+1+h}$	Stock k abnormal return relative to either the CAPM, Fama and French (1993) 3-factor model (FF3), or the Daniel et al. (1997) size/book-to-market/momentum characteristics-based benchmark (SBM), where t is the SMA belief publication day. For the CAPM and 3-factor benchmarks, betas for each stock are estimated using daily data over the trading-day window $t-272$ to $t-21$.
AggBelief	Stock-level aggregate SMA belief about stock k on day t computed as the number of bullish beliefs ($NBullish_{k,t}$) minus number of bearish beliefs ($NBearish_{k,t}$) divided by the total number of beliefs ($NBelief_{k,t}$): $(NBullish_{k,t} - NBearish_{k,t})/NBelief_{k,t}$.
Abnormal Dollar Volume	Computed for stock k following a publication on day t as the log of the average daily dollar trading volume over the 5 trading days starting $t+1$ minus the log of the average daily dollar trading volume over the prior 60 trading days ending $t-21$. Dollar volume is day-end stock price times trading volume.
Abnormal Volume	Computed for stock k following a publication on day t as the log of the average daily trading volume over the 5 trading days starting $t+1$ minus the log of the average daily trading volume over the prior 60 trading days ending $t-21$.
Abnormal Volatility	Computed for stock k following a publication on day t as the log of the volatility of daily returns over the 5 trading days starting $t+1$ minus the log of the volatility of daily returns over the prior 60 trading days ending $t-21$.
Book-to-Market	The log of stock k's book-to-market ratio as of the most recent fiscal year-end.
CashFlowNewsTone	The tone of fundamental-relevant news (i.e., $NewsSentiment_{k,t}$, defined below) published in the media about stock k averaged over the past week ending $t-2$. Missing values are replaced with the neutral value of 0.5.
Characteristic-weighted Consensus	For an SMA's belief revision about a stock on day t, the characteristic-weighted consensus is computed as the weighted average of other SMAs' beliefs on the same stock over the past 6-month period ending $t-2$, where the weights correspond to each SMA's value for a given characteristic x computed as of last calendar month using data over the past 1 year. Missing values for each characteristic are replaced with the median value, and the weights are normalized to sum to 1. The consensus is estimated if there are at least two belief statements by other SMAs over the past 6-month period. Bearish, neutral, and bullish beliefs are assigned the values 1, 2, and 3, respectively.
Consensus Optimism (CO)	A stock's equal-weighted consensus (C) on day t minus 2 (i.e., $CO = C - 2$). Since beliefs are labeled as 3 (bullish), 2 (neutral), and 1 (bearish), $CO > 0$ implies an optimistic consensus, while $CO < 0$ implies a pessimistic consensus.
Consensus Correctness (CC)	A stock's consensus optimism (CO) on day t times its future return (i.e., $CC = CO \times Ret(h)$), where $Ret(h)$ is the future horizon h return of the stock starting $t+1$ for belief revision on day t. CC is positive when the consensus is correct ex post, that is, an optimistic (pessimistic) consensus is followed by a positive (negative) future stock return.
Downgrade	Number of professional stock analysts that downgraded stock k on day t. If there are no downgrades, the value is set to 0.
EarningsEstimate	The average of professional analysts' earnings per share estimate over the last calendar month.
Equal-weighted Consensus	For an SMA's belief revision about a stock on day t, the consensus is computed as the average of other SMAs' beliefs on the same stock over the past 6-month period ending $t-2$. The consensus is estimated if there are at least two belief statements by other SMAs over the 6-month period. Bearish, neutral, and bullish beliefs are assigned the values 1, 2, and 3, respectively.
Forecast Dispersion	Dispersion of professional analyst forecasts of firm k's earning per share computed as the standard deviation of the analyst forecasts over the 30-day window ending 1 day before earnings announcement day t scaled by the stock price at the end of the previous quarter.
Idio. Skewness	Stock k's idiosyncratic skewness as of the end of the last calendar month prior to day t. It is computed as in Kumar (2009) based on the normalized third central moment of return residuals obtained by fitting a 2-factor model using 6 months of daily returns data, where the two factors are the excess market returns and the squared excess market returns. A minimum of 3 months of daily data is required.
Idio. Volatility	Stock k's idiosyncratic volatility as of the end of the last calendar month prior to day t. It is computed as in Kumar (2009) based on the standard deviation of return residuals relative to the Carhart 4-factor model using 6 months of daily returns data. A minimum of 3 months of daily data is required.
Market Beta	Stock k's CAPM market beta computed as of the end of the last calendar month prior to day t using six months of daily returns data. A minimum of 3 months of daily data is required.
NewsSentiment	Average news tone across a comprehensive set of cash flow relevant news events about stock k on day t (unless otherwise stated) computed as the Event Sentiment Score (ESS) from RavenPack News Analytics divided by 100. ESS ranges between 0 and 100, where 50 indicates neutral sentiment, values above 50 indicate positive sentiment and values below 50 indicate negative sentiment. I use only news events with a relevance score of at least 75 to focus on news events that mostly relate to a specific firm. If there are no news events for a stock on a given day, $NewsSentiment_{k,t}$ is set to its neutral value of 0.5. Supplementary Material Table IF1 lists the cash flow relevant news categories.

(continued on next page)

TABLE A1 (continued)

Variable Definitions

Variable	Definition
Price	Stock k's price per share as of the end of the last calendar month prior to day t.
Ret$[m, n]$	Stock k's return computed over window m to n trading days relative to day t, where t is the belief publication day.
Size	The log of stock k's market capitalization as of June of the previous calendar year.
Stock News	
Coverage	Number of news articles (No. of News) published about a stock over the past week ending $t-2$ (unless otherwise stated) based on RavenPack news data.
SUE	Earnings surprise for firm k computed as the difference between actual earnings per share and the average forecasts across professional analysts (consensus estimate) divided by the stock price at the end of the last quarter. To avoid stale forecasts, only forecasts published within the 30 days ending 1 day before the earnings announcement day t are used.
Upgrade	Number of professional stock analysts that upgraded stock k on day t. If there are no upgrades, the value is set to 0.
Volatility	The sum of squared daily returns in the calendar month before day t.
WeightedPastReturn	Weighted sum of the nonoverlapping weekly returns $Ret[-m, -n]$ used in equation (10), with weights based on exponential decay parameters $\lambda_1 = 1.1.2$ and $\lambda_2 = 0.80$ estimated from a version of equation (10) that does not condition on characteristic.
SMA-Level Variables	
$\overline{\mathrm{ABR}}_i$	SMA i's estimated ABR computed as the average of signed ABRs across all belief statements by SMA i. To sign the ABRs, $ABR_{k,t+1\rightarrow t+1+h}$ is premultiplied by +1 for a bullish belief about stock k and by -1 for a bearish belief.
Belief$_{i,k,t}$	SMA i's belief about stock k published on day t. The variable equals -1 if beliefs are bearish, 0 if neutral, and 1 if bullish
Exp. Skill	SMA i's conditional expected skill inferred from the two-component mixture model using equation (5).
Disagreement	The average of the absolute difference between the fraction of negative words in an SMA's opinion article and the average fraction of negative words in the comments posted within 2 days of article publication. The negative word list is from the Loughran and McDonald (2011) dictionary.
Effort	Proxied by the average number of words in opinion articles corresponding to each belief statement.
Firm Specialization	Proxied by 1 divided by the average number of unique firms an SMA expressed beliefs about in a given year.
Industry Specialization	Proxied by 1 divided by the number of unique SIC industries across which an SMA expressed beliefs over the sample period.
Popularity	Proxied by the average number of comments on each belief statement by an SMA within 2 days of publication.
Prob. High Skill	SMA i's conditional probability of being high-skilled inferred from the two-component mixture model using equation (6).
Prob. Low Skill	SMA i's conditional probability of being low-skilled inferred from the two-component mixture model using equation (6).
s_i	The standard error of SMA i's estimated ABR ($\overline{ABR}_i$). It is calculated by clustering on the belief statement day, to account for correlation across belief statements on the same day, and stock level, to account for correlation in belief statements on the same stock.
SA Experience	Captures an SMA's experience on the Seeking Alpha (SA) platform, computed as the number of years between an SMA's first and last publication on Seeking Alpha.
Skin-in-the-game	Proxied by the fraction of time an SMA discloses an investment position in the stock about which they express belief. For each SMA's publication about a stock, I create a dummy variable that equals 1 if the SMA discloses a "long" or "short" position in the stock and 0 if there is no disclosure or "no position" is disclosed. Finally, the indicator variable is averaged over the sample period for each SMA.
Workload	Proxied by the average number of belief statements published per year.

TABLE A2

Summary Statistics

Table A2 reports summary statistics for the variables used in the paper. The construction of the variables is described in Table A1. Panel A summarizes SMA-level variables computed over the full sample. Panel B summarizes variables associated with a given SMA belief publication, and Panel C summarizes the characteristics of the stocks the publications relate to. The numbers in Panels B and C are obtained by first averaging each variable across publications on a given day and then summarizing the resulting time series.

	Mean	SD	Min	P10	P25	P50	P75	P90	Max
Panel A. SMA Variables									
$\overline{\text{ABR}}$	0.003	0.026	−0.185	−0.018	−0.007	0.001	0.010	0.025	0.375
Std. Error of $\overline{\text{ABR}}$ (s)	0.014	0.016	0.000	0.003	0.005	0.009	0.016	0.028	0.328
Prob. Low Skill	0.872	0.194	0.000	0.761	0.884	0.932	0.959	0.972	0.982
Prob. High Skill	0.128	0.194	0.018	0.028	0.041	0.068	0.116	0.239	1.000
Exp. Skill	0.001	0.010	−0.125	−0.003	−0.001	0.001	0.003	0.005	0.147
Industry Specialization	0.172	0.185	0.002	0.029	0.059	0.125	0.200	0.333	1.000
Firm Specialization	0.307	0.471	0.002	0.024	0.060	0.152	0.360	0.723	6.258
Workload	28.528	50.120	0.463	2.510	4.795	11.186	28.884	70.889	645.971
Skin-in-the-game	0.444	0.369	0.000	0.000	0.038	0.412	0.800	1.000	1.000
Effort	695.925	446.919	62.051	272.033	419.351	593.099	852.318	1197.576	5154.385
Disagreement	0.029	0.012	0.001	0.017	0.022	0.028	0.034	0.043	0.167
Popularity	16.236	25.658	0.000	2.800	5.061	9.357	17.707	33.741	531.200
SA Experience	2.811	2.652	0.008	0.296	0.784	1.956	4.059	6.573	14.395
Panel B. Publication Variables									
Belief	0.647	0.249	−1.000	0.389	0.545	0.667	0.778	1.000	1.000
Equal-weighted Consensus	2.635	0.166	1.000	2.429	2.562	2.660	2.737	2.803	3.000
Specialization-weighted Consensus	2.633	0.180	1.000	2.413	2.556	2.659	2.742	2.820	3.000
Effort-weighted Consensus	2.632	0.168	1.000	2.423	2.558	2.658	2.736	2.805	3.000
Popularity-weighted Consensus	2.617	0.177	1.018	2.398	2.531	2.637	2.727	2.819	3.000
Consensus Correctness (h=5)	0.002	0.024	−0.159	−0.023	−0.009	0.002	0.013	0.025	0.231
Consensus Correctness (h=63)	0.020	0.083	−0.546	−0.068	−0.018	0.022	0.062	0.105	0.589
Consensus Correctness (h=126)	0.042	0.121	−0.714	−0.084	−0.016	0.042	0.097	0.168	0.894
Ret[−6, −2]	0.003	0.052	−0.780	−0.042	−0.017	0.005	0.023	0.044	0.885
Ret[−11, −7]	0.003	0.040	−0.355	−0.038	−0.015	0.004	0.021	0.038	0.557
Ret[−16, −12]	0.002	0.040	−0.247	−0.038	−0.015	0.003	0.021	0.039	0.637
Ret[−21, −17]	0.002	0.039	−0.585	−0.037	−0.015	0.003	0.020	0.038	0.419
Ret[−26, −22]	0.002	0.039	−0.426	−0.037	−0.015	0.003	0.020	0.037	0.437
Ret[−31, −27]	0.002	0.038	−0.320	−0.037	−0.015	0.003	0.020	0.038	0.421
Ret[−36, −32]	0.002	0.039	−0.425	−0.036	−0.014	0.004	0.020	0.037	0.340
Ret[−41, −37]	0.002	0.038	−0.600	−0.036	−0.015	0.003	0.020	0.038	0.369
Ret[−46, −42]	0.003	0.038	−0.254	−0.037	−0.014	0.003	0.020	0.039	0.368
Ret[−51, −47]	0.003	0.038	−0.378	−0.035	−0.014	0.004	0.020	0.038	0.512
Ret[−56, −52]	0.003	0.036	−0.228	−0.035	−0.014	0.004	0.020	0.037	0.374
Ret[−61, −57]	0.002	0.038	−0.444	−0.037	−0.014	0.004	0.020	0.037	0.482
Earnings Estimate	0.662	0.428	−3.130	0.216	0.450	0.661	0.859	1.080	5.600
Weighted Past Return	0.003	0.018	−0.254	−0.015	−0.004	0.004	0.011	0.018	0.225
Cash Flow News Tone	0.530	0.038	0.220	0.493	0.511	0.529	0.548	0.570	0.850
Stock News Coverage	48.831	38.416	1.000	18.146	27.238	40.460	59.547	84.986	708.000
Panel C. Stock Characteristics									
Stock Price	85.149	156.547	1.160	32.247	52.308	73.134	97.897	126.322	6412.191
Log(Mkt. Cap.)	16.116	0.882	9.916	15.214	15.752	16.185	16.583	17.004	19.722
Market Beta	1.024	0.184	−1.704	0.869	0.958	1.025	1.094	1.176	3.967
Idio. Volatility	0.022	0.008	0.006	0.015	0.017	0.020	0.023	0.029	0.147
Idio. Skewness	0.198	0.564	−7.001	−0.284	−0.042	0.188	0.445	0.707	5.041

Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109024000930.

References

Alok, S.; N. Kumar; and R. Wermers. "Do Fund Managers Misestimate Climatic Disaster Risk." *Review of Financial Studies*, 33 (2020), 1146–1183.

Ammann, M., and N. Schaub. "Do Individual Investors Trade on Investment-Related Internet Postings?" *Management Science*, 67 (2021), 5679–5702.

Antweiler, W., and M. Z. Frank. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *Journal of Finance*, 59 (2004), 1259–1294.

Avery, C. N.; J. A. Chevalier; and R. J. Zeckhauser. "The "CAPS" Prediction System and Stock Market Returns." *Review of Finance*, 20 (2016), 1363–1381.

Banerjee, A. V. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, 107 (1992), 797–817.

Barber, B. M.; R. Lehavy; M. McNichols; and B. Trueman. "Buys, Holds, and Sells: The Distribution of Investment Banks' Stock Ratings and the Implications for the Profitability of Analysts' Recommendations." *Journal of Accounting and Economics*, 41 (2006), 87–117.

Barber, B. M., and T. Odean. "Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance*, 55 (2000), 773–806.

Barberis, N., and A. Shleifer. "Style Investing." *Journal of Financial Economics*, 68 (2003), 161–199.

Bikhchandani, S.; D. Hirshleifer; and I. Welch. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*, 100 (1992), 992–1026.

Blume, M. E., and R. F. Stambaugh. "Biases in Computed Returns: An Application to the Size Effect." *Journal of Financial Economics*, 12 (1983), 387–404.

Boehmer, E.; C. M. Jones; X. Zhang; and X. Zhang. "Tracking Retail Investor Activity." *Journal of Finance*, 76 (2021), 2249–2305.

Brunnermeier, M.; E. Farhi; R. S. J. Koijen; A. Krishnamurthy; S. C. Ludvigson; H. Lustig; S. Nagel; and M. Piazzesi. "Review Article: Perspectives on the Future of Asset Pricing." *Review of Financial Studies*, 34 (2021), 2126–2160.

Campbell, J. L.; M. D. DeAngelis; and J. R. Moon. "Skin in the Game: Personal Stock Holdings and Investors' Response to Stock Analysis on Social Media." *Review of Accounting Studies*, 24 (2019), 731–779.

Cassella, S., and H. Gulen. "Extrapolation Bias and the Predictability of Stock Returns by Price-Scaled Variables." *Review of Financial Studies*, 31 (2018), 4345–4397.

Chen, H.; P. De; Y. J. Hu; and B.-H. Hwang. "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media." *Review of Financial Studies*, 27 (2014), 1367–1403.

Chen, M. A.; Q. Wu; and B. Yang. "How Valuable is FinTech Innovation?" *Review of Financial Studies*, 32 (2019), 2062–2106.

Chen, Y.; M. T. Cliff; and H. Zhao. "Hedge Funds: The Good, the Bad, and the Lucky." *Journal of Financial and Quantitative Analysis*, 52 (2017), 1081–1109.

Choi, J. J., and A. Z. Robertson. "What Matters to Individual Investors? Evidence from the Horse's Mouth." *Journal of Finance*, 75 (2020), 1965–2020.

Cookson, J. A.; J. E. Engelberg; and W. Mullins. "Does Partisanship Shape Investor Beliefs? Evidence from the COVID-19 Pandemic." *Review of Asset Pricing Studies*, 10 (2020), 863–893.

Cookson, J. A.; J. E. Engelberg; and W. Mullins. "Echo Chambers." *Review of Financial Studies*, 36 (2023), 450–500.

Cookson, J. A.; R. Lu; W. Mullins; and M. Niessner. "The Social Signal." *Journal of Financial Economics*, 158 (2024), 103870.

Cookson, J. A., and M. Niessner. "Why Don't We Agree? Evidence from a Social Network of Investors." *Journal of Finance*, 75 (2020), 173–228.

Coval, J. D.; D. Hirshleifer; and T. Shumway. "Can Individual Investors Beat the Market?" *Review of Asset Pricing Studies*, 11 (2021), 552–579.

Crane, A., and K. Crotty. "How Skilled Are Security Analysts?" *Journal of Finance*, 75 (2020), 1629–1675.

Da, Z.; X. Huang; and L. J. Jin. "Extrapolative Beliefs in the Cross-Section: What Can We Learn from the Crowds?" *Journal of Financial Economics*, 140 (2021), 175–196.

Daniel, K.; M. Grinblatt; S. Titman; and R. Wermers. "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks." *Journal of Finance*, 52 (1997), 1035–1058.

De Long, J. B.; A. Shleifer; L. H. Summers; and R. J. Waldmann. "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." *Journal of Finance*, 45 (1990), 379–395.

Drake, M. S.; J. R. Moon Jr; B. J. Twedt; and J. D. Warren. "Social Media Analysts and Sell-Side Analyst Research." *Review of Accounting Studies*, 28 (2023), 385–420.

Eyster, E., and M. Rabin. "Naive Herding in Rich-Information Settings." *American Economic Journal: Microeconomics*, 2 (2010), 221–43.

Fama, E. F., and K. R. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

Farrell, M.; T. C. Green; R. Jame; and S. Markov. "The Democratization of Investment Research and the Informativeness of Retail Investor Trading." *Journal of Financial Economics*, 145 (2022), 616–641.

Farrell, M.; R. Jame; and T. Qiu. "The Cross-Section of Non-Professional Analyst Skill." Working Paper, available at SSRN 3682490 (2020).

Fuster, A.; D. Laibson; and B. Mendel. "Natural Expectations and Macroeconomic Fluctuations." *Journal of Economic Perspectives*, 24 (2010), 67–84.

Giglio, S.; M. Maggiori; J. Stroebel; and S. Utkus. "Five Facts About Beliefs and Portfolios." *American Economic Review*, 111 (2021), 1481–1522.

Gomez, E.; F. Heflin; J. Moon; and J. Warren. "Can Financial Analysis on Social Media Help Level the Playing Field Among Investors? Evidence from Seeking Alpha." *Georgia Tech Research Paper* 18–45 (2018).

Greenwood, R., and A. Shleifer. "Expectations of Returns and Expected Returns." *Review of Financial Studies*, 27 (2014), 714–746.

Harvey, C. R., and Y. Liu. "Detecting Repeatable Performance." *Review of Financial Studies*, 31 (2018), 2499–2552.

Heimer, R. Z. "Peer Pressure: Social Interaction and the Disposition Effect." *Review of Financial Studies*, 29 (2016), 3177–3209.

Hirshleifer, D.; S. S. Lim; and S. H. Teoh. "Limited Investor Attention and Stock Market Misreactions to Accounting Information." *Review of Asset Pricing Studies*, 1 (2011), 35–73.

Jeng, L. A.; A. Metrick; and R. Zeckhauser. "Estimating the Returns to Insider Trading: A Performance-Evaluation Perspective." *Review of Economics and Statistics*, 85 (2003), 453–471.

Kakhbod, A.; S. M. Kazempour; D. Livdan; and N. Schuerhoff. "Finfluencers." Working Paper, available at SSRN 4428232 (2023).

Kaniel, R.; S. Liu; G. Saar; and S. Titman. "Individual Investor Trading and Return Patterns Around Earnings Announcements." *Journal of Finance*, 67 (2012), 639–680.

Kelley, E. K., and P. C. Tetlock. "How Wise Are Crowds? Insights from Retail Orders and Stock Returns." *Journal of Finance*, 68 (2013), 1229–1265.

Kuchler, T., and B. Zafar. "Personal Experiences and Expectations About Aggregate Outcomes." *Journal of Finance*, 74 (2019), 2491–2542.

Kumar, A. "Who Gambles in the Stock Market?" *Journal of Finance*, 64 (2009), 1889–1933.

Loughran, T., and B. McDonald. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance*, 66 (2011), 35–65.

Manela, A., and A. Moreira. "News Implied Volatility and Disaster Concerns." *Journal of Financial Economics*, 123 (2017), 137–162.

Newey, W. K., and K. D. West. "A Simple, Positive-Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (1987), 703–708.

Renault, T. "Market Manipulation and Suspicious Stock Recommendations on Social Media." Working Paper, available at SSRN 3010850 (2017).

Scharfstein, D. S., and J. C. Stein. "Herd Behavior and Investment." *American Economic Review*, (1992) 465–479.

Seasholes, M. S., and N. Zhu. "Individual Investors and Local Bias." *Journal of Finance*, 65 (2010), 1987–2010.

Seeking Alpha. "About Seeking Alpha." https://about.seekingalpha.com/ (2006). Accessed Dec. 10, 2020.

Tversky, A., and D. Kahneman. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology*, 5 (1973), 207–232.

Van Nieuwerburgh, S., and L. Veldkamp. "Information Acquisition and Under-Diversification." *Review of Economic Studies*, 77 (2010), 779–805.

Welch, I. "Herding Among Security Analysts." *Journal of Financial Economics*, 58 (2000), 369–396.