Economics Letters 264 (2026) 112965

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

On the inaccuracy of relative accuracy measures

Tristan Roger

ICN Business School, Université de Lorraine, CEREFIGE, F-54000 Nancy, France

ARTICLE INFO

Keywords:
Financial analysts
Relative accuracy measures
Absolute forecast error
Earnings predictability

ABSTRACT

This article examines widely used measures of analysts' earnings forecast accuracy. Simulation results show that relative accuracy measures, such as those of Clement (1999), Hong et al. (2000), and Clement and Tse (2003), provide only an imperfect assessment of financial analysts' performance. Designed to adjust for variation in earnings predictability, these measures can also add noise to cross-analyst comparisons. The results suggest that these metrics remain informative but should be applied with caution in empirical research, particularly when analyst coverage is limited or heterogeneous.

1. Introduction

Among the numerous papers on sell-side analysts published over the last decades, an important focus has been placed on earnings forecasts, and more specifically, on earnings forecast accuracy. Early research primarily investigated forecasts in the context of market efficiency, examining whether analysts incorporate publicly available information efficiently into their forecasts. More recent work has turned to the nature of analyst expertise, asking whether some analysts are systematically better forecasters (Ramnath et al., 2008; Bradshaw, 2011). This line of inquiry raises two central questions: Do some analysts consistently produce superior forecasts, and what explains such differences? However, answering these questions presupposes a reliable way to measure forecasting ability.

The standard approach to measuring forecast accuracy is to compute the absolute forecast error, that is, the absolute value of the deviation between the forecast and the realized earnings. The complexity arises when aggregating forecast accuracy at the analyst level. Jacob et al. (1999) note that "forecasting difficulty is [...] like to differ cross-sectionally". Similarly, Hong et al. (2000) state that "some firms are more difficult than others to predict accurately". Since the difficulty of forecasting earnings is not the same for all firms, one cannot simply average the absolute forecast errors over all stocks covered by an analyst to assess their forecasting ability.

To address this issue, the literature has proposed relative measures of forecast accuracy that account for firm-level differences in predictability. Hong et al. (2000) use rank transformation. The cross-sectional differences in earnings predictability are dealt with by applying a ranking function to the absolute forecast errors for each firm. In Clement (1999), earnings predictability is taken into account by deflating the absolute forecast errors by the average absolute forecast errors for a given firm (this mean absolute forecast error aims at estimating the expected value of the absolute forecast error). Clement and Tse (2003) use a similar approach and standardize forecast accuracy by considering for a given period and a given firm the minimum and the maximum absolute forecast error. This approach is often referred to as Proportion of Maximum Scoring (POMS) or min–max transformation. Each method is designed to reduce biases arising from differences in earnings predictability, but each may introduce its own distortions in the cross-sectional comparison of analysts.

The present study evaluates whether these relative measures provide an accurate assessment of analysts' forecasting performance. I employ a simulation-based methodology in which absolute forecast errors are generated from known distributions. I then compare the rankings obtained using relative measures of accuracy to the theoretical rankings derived from the simulated data. My analysis proceeds in two complementary settings: a simple, idealized setting in which each analyst covers all firms, and a realistic setting in which coverage is heterogeneous across analysts and firms.

In the first setting, I simulate m analysts covering n firms, with each analyst issuing a forecast for every firm. Absolute forecast errors are modeled using three distributions (gamma, lognormal, and Weibull), which were chosen from a wider set of possible distributions on the basis of empirical distribution-fitting tests (see Appendix). Knowing the true distribution allows the computation of theoretical performance rankings from standardized forecast errors aggregated at the analyst level. The second setting introduces heterogeneous coverage, reflecting the real-world pattern in which some analysts follow many firms while others follow few, and some firms are heavily covered while others receive minimal attention. Using I/B/E/S data for 2019–2023, I simulate a forecast error for each analyst-firm observation and compute both

E-mail address: tristan.roger@icn-artem.com.

https://doi.org/10.1016/j.econlet.2026.112965
Received 16 October 2025; Received in revised form 4 February 2026; Accepted 29 March 2026
Available online 30 March 2026
0165-1765/

theoretical rankings and rankings computed from relative measures of accuracy. To assess the differences between theoretical and estimated rankings, I employ three metrics: Kendall rank correlation, average absolute rank deviation, and the proportion of analysts assigned to incorrect performance deciles.

The simulation results indicate that relative accuracy measures depart substantially from the theoretical benchmark, even in the idealized setting of complete coverage. Deviations appear to be even more pronounced under heterogeneous coverage. The Kendall rank correlation between the theoretical ranking and those obtained from the relative measures ranges from 0.58 to 0.72. The proportion of misclassified analysts (those assigned to deciles of performance different from the theoretical benchmark) varies between 59% and 71%. Overall, the measures of Clement (1999) and Hong et al. (2000) perform similarly, while the measure of Clement and Tse (2003) performs slightly worse. Comparing results across the complete and heterogeneous coverage scenarios shows that relative accuracy measures deteriorate substantially when coverage is low or when analysts cover few stocks.

This paper contributes to the literature on financial analysts by demonstrating that widely-used relative measures of forecast accuracy, despite their popularity, can distort the assessment of analyst performance. The measures proposed by Hong et al. (2000), Clement (1999), Clement and Tse (2003) have been cited hundreds to over a thousand times in the academic literature, indicating their widespread adoption.[1] Nonetheless, my findings suggest that empirical researchers should apply these measures with caution, particularly in settings characterized by low or heterogeneous coverage.

2. Measures of analyst performance

2.1. Theoretical measure

The theoretical performance of a forecast is measured by standardizing absolute forecast errors. For a forecast issued by analyst j on firm i in year t, theoretical performance is defined as:

$$Forecast\ theoretical\ performance_{i,j,t} = \frac{AFE_{i,j,t} - E[AFE_{i,t}]}{\sqrt{VAR(AFE_{i,t})}} \tag{1}$$

where $AFE_{i,j,t}$ is the absolute forecast error and $E[AFE_{i,t}]$ and $VAR(AFE_{i,t})$ are the expected value and variance, respectively, of the absolute forecast error, under the assumption that the distribution is known. The distributions considered include gamma, lognormal, and Weibull.

The theoretical performance of an analyst is calculated as the average of standardized forecast errors across all firms covered:

$$Performance_{j,t}^{theoretical} = \frac{1}{n_{j,t}} \sum_{i=1}^{n_{j,t}} \frac{AFE_{i,j,t} - E[AFE_{i,t}]}{\sqrt{VAR(AFE_{i,t})}} \tag{2}$$

The theoretical ranking is then obtained by ranking analysts according to their performance.

[1] As of February 2026, Hong et al. (2000) has 1750 citations on Google Scholar. This measure is also used in Hong and Kubik (2003), which has 2071 citations. Clement (1999) has 2066 citations. Clement and Tse (2003) has 639 citations. This last measure is also used in Clement and Tse (2005), an article with 1245 citations. Recent citations include articles published in top accounting journals. A non-exhaustive list of such papers contains articles published in *The Accounting Review* (e.g., Bhagwat and Liu, 2020; Bourveau and Law, 2021; Comprix et al., 2022), in *Accounting, Organizations and Society* (e.g., Hope et al., 2021), in *Journal of Accounting and Economics* (e.g., Huang et al., 2022; Du, 2023; Jiao, 2024; Lang et al., 2024), in *Journal of Accounting Research* (e.g., Aslan, 2022; Peng et al., 2022), in *Contemporary Accounting Research* (e.g., Li et al., 2023; Pisciotta, 2023; Han et al., 2024) and in *Review of Accounting Studies* (e.g., Pope and Wang, 2023; Cao et al., 2024; Dong et al., 2024).

2.2. Relative measures of accuracy

2.2.1. Measure of Hong et al. (2000)

Hong et al. (2000) rank analysts within each firm based on the absolute forecast error. Scores are computed as:

$$Score_{i,j,t} = 100 - \left[\frac{r(AFE_{i,j,t}) - 1}{m_{i,t} - 1}\right] \times 100 \tag{3}$$

where $r()$ is the ranking function that transforms, for firm i and time t, the absolute forecast errors into ranks, with 1 (respectively, 100) the rank (score) associated with the smallest absolute forecast error. $m_{i,t}$ is the number of analysts covering firm i in year t. Analysts' overall performance is obtained by averaging these scores across all firms covered.

The transformation of absolute forecast errors into scores is motivated by the need to account for differences in forecast difficulty across firms. As Hong et al. (2000) note, some analysts are likely to receive extreme scores regardless of their actual performance. Analysts who cover few firms or firms with sparse coverage are more likely to appear at the extremes. Moreover, the transformation of absolute errors into scores discards some information, specifically the magnitude of forecast errors, so rankings based on the measure of Hong et al. (2000) can differ from the theoretical ranking. This mismatch arises from the normalization induced by the ranking function applied to transform absolute forecast errors into firm-level scores.

2.2.2. Measure of Clement (1999)

Clement (1999) proposes the proportional mean absolute forecast error (PMAFE), which scales an analyst's absolute forecast error relative to the mean absolute forecast error:

$$PMAFE_{i,j,t} = \frac{AFE_{i,j,t} - \overline{AFE}_{i,t}}{\overline{AFE}_{i,t}} \tag{4}$$

where $\overline{AFE}_{i,t}$ is the average absolute forecast error of all analysts covering firm i. Aggregating PMAFE across firms provides an overall analyst performance measure.

Clement (1999) motivates his approach by the need to control for events, such as voluntary management disclosures, mergers, or strikes, that can generate firm-year effects. His method ensures that the accuracy measure is not distorted by such unpredictable events. Beyond this consideration, the approach also helps control for differences in earnings predictability. The average absolute forecast error ($\overline{AFE}_{i,t}$) is used to estimate the expected value of the absolute forecast error ($E[AFE_{i,t}]$). However, the average absolute forecast error is a good estimator of the expected value only if the number of forecasts issued for a given firm is large enough[2] and if forecasts are unbiased. When few analysts cover a firm, the average absolute forecast error may deviate substantially from its expected value.

2.2.3. Measure of Clement and Tse (2003)

The measure of Clement and Tse (2003) scales forecast errors between the minimum and maximum observed absolute forecast errors for each firm:

$$Accuracy_{i,j,t} = \frac{AFEmax_{i,t} - AFE_{i,j,t}}{AFEmax_{i,t} - AFEmin_{i,t}} \tag{5}$$

Clement and Tse (2003) justify their scaling choice so as to facilitate comparisons in econometric analyses. All the variables in their analyses are scaled by subtracting the minimum and scaling by the difference between the maximum and the minimum so that the output lies between 0 and 1. The authors argue that "this approach controls for systematic firm-year differences".

[2] We have $\lim_{m_{i,t}\to\infty} \frac{1}{m_{i,t}} \sum_{j=1}^{m_{i,t}} AFE_{i,j,t} = E[AFE_{i,t}]$.

Table 1
Relative measures of accuracy with complete coverage (10 assets).

Panel A: AFEs generated from a gamma distribution						
	Measure of Hong et al. (2000)		Measure of Clement (1999)		Measure of Clement and Tse (2003)	
Number of analysts	Kendall rank correlation	Average rank deviation	Kendall rank correlation	Average rank deviation	Kendall rank correlation	Average rank deviation
5	0.4993	0.9020	0.6128	0.6940	0.5876	0.7348
10	0.5160	1.7820	0.7210	1.0820	0.6827	1.2100
20	0.5295	3.4140	0.7993	1.5610	0.7522	1.8950
30	0.5330	5.0410	0.8349	1.9230	0.7803	2.4980
100	0.5366	16.4000	0.9114	3.3710	0.8454	5.6930
Panel B: AFEs generated from a lognormal distribution						
	Measure of Hong et al. (2000)		Measure of Clement (1999)		Measure of Clement and Tse (2003)	
Number of analysts	Kendall rank correlation	Average rank deviation	Kendall rank correlation	Average rank deviation	Kendall rank correlation	Average rank deviation
5	0.4429	0.9874	0.6392	0.6536	0.5358	0.8264
10	0.4476	1.9860	0.7252	1.0760	0.6057	1.4740
20	0.4572	3.8800	0.7872	1.6580	0.6706	2.4660
30	0.4548	5.8210	0.8105	2.1900	0.6936	3.4130
100	0.4606	18.9700	0.8792	4.5070	0.7481	9.1000
Panel C: AFEs generated from a Weibull distribution						
	Measure of Hong et al. (2000)		Measure of Clement (1999)		Measure of Clement and Tse (2003)	
Number of analysts	Kendall rank correlation	Average rank deviation	Kendall rank correlation	Average rank deviation	Kendall rank correlation	Average rank deviation
5	0.4644	0.9528	0.6044	0.7140	0.5526	0.7940
10	0.4905	1.8550	0.7196	1.0820	0.6501	1.3170
20	0.4964	3.6190	0.7979	1.5860	0.7160	2.1570
30	0.4974	5.4020	0.8334	1.9510	0.7441	2.8800
100	0.5050	17.4800	0.9049	3.5940	0.8061	7.0830

For each firm-analyst observation, I simulate an absolute forecast error from a known distribution. In Panel A, AFEs follow a gamma distribution. In Panel B, they follow a lognormal distribution. In Panel C, they follow a Weibull distribution. The parameters of the distributions are selected so that the expected value (variance) of AFE is equal to the estimated average AFE (variance) over the total population. Column 1 indicates the number of analysts considered in the simulations (the number of assets is equal to 10). Column 2 reports the Kendall rank correlation coefficient between the analysts' performance as calculated in Hong et al. (2000) and the theoretical analysts' performance. Column 3 reports the average rank deviation for each analyst between the ranking established using analysts' performance as calculated in Hong et al. (2000) and the theoretical ranking (*i.e.* how far from their theoretical ranking an analyst is located on average). Columns 4 and 5 (columns 6 and 7) report the same statistics as in columns 2 and 3 except that analysts' performance is measured using the relative measure of accuracy of Clement (1999), Clement and Tse (2003).

3. Simulations

3.1. Design of the simulations

I simulate an economy in which m analysts issue forecasts for n firms, with forecast errors drawn from specified distributions. The simulation proceeds as follows:

1. For each firm and analyst, I generate a simulated absolute forecast error from a chosen distribution (gamma, lognormal, or Weibull) with parameters estimated from I/B/E/S data (2019–2023).[3]
2. I compute theoretical performance/rankings, and compute relative measures of accuracy according to Hong et al. (2000), Clement (1999), Clement and Tse (2003).
3. I assess the similarity between theoretical and estimated rankings using Kendall rank correlation, average absolute rank deviation (a modified Spearman's footrule),[4] and the proportion of analysts assigned to incorrect deciles of performance.

These steps are repeated 1000 times to obtain stable estimates of the performance of each relative measure.

[3] The parameters of the distribution are estimated using the *fitdist* function in Matlab.

[4] Formally, it is defined as

$$Average\,rank\,deviation_t = \frac{1}{m}\sum_{j=1}^{m}\left|Ranking_{j,t}^{estimated} - Ranking_{j,t}^{theoretical}\right| \quad (6)$$

4. Empirical analysis

4.1. Complete coverage

In the complete coverage setting, each analyst issues forecasts for all firms. Table 1 presents three main results. First, there are large deviations between the theoretical ranking and those obtained with relative measures of accuracy. Second, relative measures improve as the number of analysts increases. Third, the measure proposed by Clement (1999) performs best, whereas the one proposed by Hong et al. (2000) yields substantially different rankings, with Kendall correlations as low as 0.4429.

4.2. Heterogeneous coverage

When analyst coverage is heterogeneous, the results reported in Table 2 indicate that all relative accuracy measures perform poorly and exhibit little differentiation. For gamma-distributed AFEs, for instance, Kendall rank correlations range from 0.6816 to 0.7226, with up to 62% of analysts assigned to incorrect performance deciles. Similar results are obtained when AFEs are generated from a Weibull distribution. When AFEs follow a lognormal distribution, Kendall rank correlations are lower, ranging from 0.5833 to 0.6162, and misclassification rates are higher, between 67.44% and 70.91%. The measures proposed by Clement (1999) and Hong et al. (2000) perform similarly, whereas the measure of Clement and Tse (2003) performs slightly worse. Nevertheless, across all three distributions, performance differences remain small.

Table 2
Relative measures of accuracy with heterogeneous coverage.

	Measure of Hong et al. (2000)		Measure of Clement (1999)		Measure of Clement and Tse (2003)	
	Kendall rank correlation	Proportion of analysts in incorrect deciles of performance	Kendall rank correlation	Proportion of analysts in incorrect deciles of performance	Kendall rank correlation	Proportion of analysts in incorrect deciles of performance
Panel A: Heterogeneous coverage with AFEs drawn from a gamma distribution						
2019	0.6936	0.5960	0.7198	0.5992	0.6845	0.6166
2020	0.6938	0.5963	0.7176	0.6052	0.6828	0.6178
2021	0.6935	0.5967	0.7220	0.6014	0.6829	0.6196
2022	0.6897	0.5994	0.7169	0.6043	0.6816	0.6194
2023	0.6906	0.5997	0.7226	0.5993	0.6832	0.6205
Panel B: Heterogeneous coverage with AFEs drawn from a lognormal distribution						
2019	0.6099	0.6753	0.6162	0.7041	0.5891	0.7048
2020	0.6099	0.6759	0.6128	0.7085	0.5884	0.7061
2021	0.6069	0.6775	0.6125	0.7058	0.5833	0.7079
2022	0.6078	0.6744	0.6153	0.7039	0.5877	0.7028
2023	0.6042	0.6806	0.6152	0.7063	0.5838	0.7091
Panel C: Heterogeneous coverage with AFEs drawn from a Weibull distribution						
2019	0.7017	0.5864	0.7120	0.6088	0.6764	0.6261
2020	0.7027	0.5861	0.7102	0.6156	0.6738	0.6291
2021	0.7002	0.5892	0.7136	0.6126	0.6723	0.6330
2022	0.6962	0.5920	0.7067	0.6169	0.6725	0.6296
2023	0.6938	0.5962	0.7093	0.6173	0.6700	0.6372

Each year (column 1), I select the structure of coverage as obtained in I/B/E/S (the firms for which analysts issue earnings forecasts). For each firm-analyst pair, I simulate an absolute forecast error from a known distribution. In Panel A, absolute forecast errors follow a gamma distribution. In Panel B, they follow a lognormal distribution. In Panel C, they follow a Weibull distribution. For each firm-year, the parameters of the distributions are obtained by fitting the distribution to the actual AFEs (using maximum likelihood estimation). Column 2 reports the Kendall rank correlation coefficient between the analysts' performance as calculated in Hong et al. (2000) and the theoretical analysts' performance. Column 3 reports the proportion of analysts who are not assigned to the same decile of performance when using analysts' performance as calculated in Hong et al. (2000) and when using the theoretical performance (*i.e.* analysts whose assignment to a decile of performance is incorrect when using the measure of Hong et al. (2000)). Columns 4 and 5 (columns 6 and 7) report the same statistics as in columns 2 and 3 except that analysts' performance is measured using the relative measure of accuracy of Clement (1999) (respectively, Clement and Tse, 2003).

Table A.1
Distribution fitting of AFEs.

Panel A: Kolmogorov–Smirnov test									
Significance level	Exponential	Gamma	Half Normal	Inverse Gaussian	Logistic	Lognormal	Normal	Rayleigh	Weibull
5%	0.7091	0.8901	0.5573	0.4687	0.7544	0.9008	0.6208	0.2984	0.9310
10%	0.6362	0.8413	0.4947	0.3952	0.6793	0.8322	0.5313	0.2415	0.8921
Panel B: Anderson–Darling test									
Significance level	Exponential	Gamma	Half Normal	Inverse Gaussian	Logistic	Lognormal	Normal	Rayleigh	Weibull
5%	0.7230	0.9235	0.5577	0.4871	0.7173	0.9083	0.6176	0.2087	0.9442
10%	0.6476	0.8798	0.4908	0.4166	0.6303	0.8470	0.5261	0.1700	0.9051

For each firm, I take the vector of AFE observations over the 2019–2023 period. Conditional on having more than 10 observations, I then conduct a Kolmogorov–Smirnov test (Panel A) and an Anderson–Darling test (Panel B) to investigate whether the sample is drawn from a given probability distribution. I consider several distributions: (1) the exponential distribution; (2) the gamma distribution; (3) the half normal distribution; (4) the inverse Gaussian distribution; (5) the logistic distribution; (6) the lognormal distribution; (7) the normal distribution; (8) the Rayleigh distribution; and, (9) the Weibull distribution. For each distribution, I report the percentage of firms for which the null hypothesis that the sample is indeed drawn from said distribution is not rejected. I consider both a 5% significance level and a 10% significance level. The initial sample is composed of 6595 firms for which only 4393 have more than 10 AFEs associated.

5. Discussion

The simulations yield three main insights. First, relative measures of accuracy are informative but depart substantially from theoretical analyst performance. The extent of misclassification depends on the distribution of forecast errors, the number of analysts, and the coverage structure. Second, coverage heterogeneity substantially reduces the accuracy of relative measures, underscoring the importance of accounting for coverage structure in empirical applications. Third, in the more realistic heterogeneous coverage setting, the three relative measures of accuracy perform similarly, with slightly weaker performance for Clement and Tse (2003).

These findings carry implications for empirical research. Studies often rely on relative accuracy measures to analyze analyst skill, career outcomes, or market reactions. The results suggest that reported differences in analyst ability may be partly shaped, or obscured, by the measurement approach rather than by actual variation in skill.

It should be acknowledged that the choice of the theoretical benchmark is not entirely neutral and may influence the results. While centering and standardizing appears to be a natural choice,[5] as it allows performance to be compared across firms by accounting for both scale (the magnitude of earnings) and difficulty (the variability of earnings), such a benchmark may be less appropriate once analysts' career concerns are considered. Several studies argue that the labor market for analysts resembles a tournament (Leone and Wu, 2007; Yin and Zhang, 2014), in which performance is evaluated relative to peers rather than in absolute terms. Under these incentives, analyst performance may be better captured by ranking-based measures or by performance functions that incorporate absolute forecast errors nonlinearly. However, evidence on how analysts' compensation and career progression are determined remains limited.[6] This lack of information likely explains why empirical studies nearly always rely on linear functions of absolute forecast errors, in the absence of better alternatives.[7]

6. Conclusion

While this paper does not offer a definitive solution for measuring analyst forecast accuracy, and while the results necessarily depend on the choice of the theoretical benchmark, it shows that commonly used relative measures can introduce distortions, particularly when coverage is sparse or heterogeneous.

Appendix A

Little research exists on the distribution of forecast errors (FE) and absolute forecast errors (AFE). A couple of papers (Abarbanell and Lehavy, 2003; Cohen and Lys, 2003) show that FE are not normally distributed. Abarbanell and Lehavy (2003) note "two relatively small but statistically influential asymmetries in the tail and the middle of distributions of analysts' forecast errors". Cohen and Lys (2003) confirm this finding but do not attempt to identify the underlying distribution ("To derive such a test, one would have to first define [...] the distribution [...]. The derivation of what the distribution should be [..] is beyond the scope of our discussion".).

The goal of this article is not to determine the true distribution of AFEs with certainty, but to consider several candidate distributions for representing AFEs. For each firm, I collect the vector of AFEs over the 2019–2023 period and apply both Kolmogorov–Smirnov and Anderson–Darling tests to compare the empirical observations with reference probability distributions. I consider nine distributions: (1) exponential, (2) gamma, (3) half normal, (4) inverse Gaussian, (5) logistic, (6) lognormal, (7) normal, (8) Rayleigh, and (9) Weibull. For each distribution, I report the percentage of firms for which the null hypothesis that AFEs follow the distribution cannot be rejected.

Table A.1 presents the results. Overall, the Weibull distribution fits the data best: for over 90% of firms, the Kolmogorov–Smirnov test at the 5% significance level does not reject the null hypothesis that AFEs follow a Weibull distribution. Fit is also strong for the gamma and lognormal distributions, at roughly 90%, but substantially weaker for the other distributions considered.

Data availability

The authors do not have permission to share data.

References

Abarbanell, J., Lehavy, R., 2003. Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts. J. Account. Econ. 36 (1–3), 105–146.

Aslan, H., 2022. Personal financial distress, limited attention. J. Account. Res. 60 (1), 97–128.

Bhagwat, V., Liu, X., 2020. The role of trust in information processing: Evidence from security analysts. Account. Rev. 95 (3), 59–83.

Bourveau, T., Law, K.K., 2021. Do disruptive life events affect how analysts assess risk? Evidence from deadly hurricanes. Account. Rev. 96 (3), 121–140.

Bradshaw, M.T., 2011. Analysts' Forecasts: What do We Know After Decades of Work? Working Paper, Boston College.

Cao, Y., Hao, R., Yang, Y.G., 2024. National culture and analysts' forecasting. Rev. Account. Stud. 29 (2), 1147–1191.

[5] Accounting for scale is common practice in the forecasting literature (Hyndman and Koehler, 2006), although the approach used in this literature differs somewhat, as it deals with time series forecasts. In addition, z-score approaches are standard in finance (e.g., the Sharpe ratio).

[6] Only a handful of studies examine the effect of forecast accuracy on analysts' career incentives (Mikhail et al., 1999; Hong et al., 2000; Hong and Kubik, 2003; Emery and Li, 2009; Roger, 2018), and career outcomes are measured indirectly using proxies.

[7] Although Hong et al. (2000) propose a relative accuracy measure motivated by analysts' career concerns, the symmetric and linear nature of the ranking function is unlikely to fully capture real-world performance evaluations.

Clement, M.B., 1999. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? J. Account. Econ. 27 (3), 285–303.

Clement, M.B., Tse, S.Y., 2003. Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? Account. Rev. 78 (1), 227–249.

Clement, M.B., Tse, S.Y., 2005. Financial analyst characteristics and herding behavior in forecasting. J. Financ. 60 (1), 307–341.

Cohen, D.A., Lys, T.Z., 2003. A note on analysts' earnings forecast errors distribution. J. Account. Econ. 36 (1–3), 147–164.

Comprix, J., Lopatta, K., Tideman, S.A., 2022. The role of gender in the aggressive questioning of CEOs during earnings conference calls. Account. Rev. 97 (7), 79–107.

Dong, Y., Liu, X., Lobo, G.J., Ni, C., 2024. Overprecise forecasts. Rev. Account. Stud. 29 (1), 276–326.

Du, M., 2023. Locked-in at home: The gender difference in analyst forecasts after the COVID-19 school closures. J. Account. Econ. 76 (1), 101603.

Emery, D.R., Li, X., 2009. Are the Wall Street analyst rankings popularity contests? J. Financ. Quant. Anal. 44 (2), 411–437.

Han, Y., Mao, C.X., Tan, H., Zhang, C., 2024. Climatic disasters and distracted analysts. Contemp. Account. Res. 41 (2), 1120–1150.

Hong, H., Kubik, J.D., 2003. Analyzing the analysts: Career concerns and biased earnings forecasts. J. Financ. 58 (1), 313–351.

Hong, H., Kubik, J.D., Solomon, A., 2000. Security analysts' career concerns and herding of earnings forecasts. RAND J. Econ. 31 (1), 121–144.

Hope, O.K., Li, C., Lin, A.P., Rabier, M., 2021. Happy analysts. Account. Organ. Soc. 90, 101199.

Huang, A.H., Lin, A.-P., Zang, A.Y., 2022. Cross-industry information sharing among colleagues and analyst research. J. Account. Econ. 74 (1), 101496.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688.

Jacob, J., Lys, T.Z., Neale, M.A., 1999. Expertise in forecasting performance of security analysts. J. Account. Econ. 28 (1), 51–82.

Jiao, Y., 2024. Managing decision fatigue: Evidence from analysts' earnings forecasts. J. Account. Econ. 77 (1), 101615.

Lang, M., Pinto, J., Sul, E., 2024. MiFID II unbundling and sell-side analyst research. J. Account. Econ. 77 (1), 101617.

Leone, A.J., Wu, J.S., 2007. What does it take to become a superstar? Evidence from institutional investor rankings of financial analysts. In: Evidence from Institutional Investor Rankings of Financial Analysts (May 23, 2007). pp. 02–12, Simon School of Business Working Paper No. FR.

Li, C., Lin, A.P., Lu, H., 2023. The effect of social skills on analyst performance. Contemp. Account. Res. 40 (2), 1418–1447.

Mikhail, M.B., Walther, B.R., Willis, R.H., 1999. Does forecast accuracy matter to security analysts? Account. Rev. 74 (2), 185–200.

Peng, L., Teoh, S.H., Wang, Y., Yan, J., 2022. Face value: Trait impressions, performance characteristics, and market outcomes for financial analysts. J. Account. Res. 60 (2), 653–705.

Pisciotta, K., 2023. Analyst workload and information production: Evidence from IPO assignments. Contemp. Account. Res. 40 (3), 1605–1640.

Pope, P.F., Wang, T., 2023. Analyst ability and research effort: non-EPS forecast provision as a research quality signal. Rev. Account. Stud. 28 (3), 1263–1315.

Ramnath, S., Rock, S., Shane, P.B., 2008. The financial analyst forecasting literature: A taxonomy with suggestions for further research. Int. J. Forecast. 24 (1), 34–75.

Roger, T., 2018. The coverage assignments of financial analysts. Account. Bus. Res. 48 (6), 651–673.

Yin, H., Zhang, H., 2014. Tournaments of financial analysts. Rev. Account. Stud. 19 (2), 573–605.