



Valuing financial data: The case of analyst forecasts

Zhenghui Li^{a,b}, Yanting Xu^a, Ziqing Du^{a,b,*}

^a School of Economics and Statistics, Guangzhou University, Guangzhou, 510006, China

^b Guangzhou Institute of International Finance, Guangzhou University, Guangzhou, 510405, China

ARTICLE INFO

JEL classifications:

G10

D80

Keywords:

Valuing financial data

Analyst forecast

Investor welfare

Price impact

ABSTRACT

Investors seek financial market information to maximize returns and minimize risk. This study examines the value of financial information employing quarterly analyst data from 144 A-share listed companies in China (2008–2023). It investigates how financial data enhance investor welfare, varies by wealth levels and market competition, and differs across assets. The findings reveal three key insights: (1) Analyst forecasts enhance investor welfare, especially for wealthier investors. (2) Data value decreases in markets with price impact. (3) Financial data are most valuable for medium-sized assets with high analyst attention and transparency.

1. Introduction

Measuring the value of financial data is challenging, as it directly impacts investor welfare. Financial data shape investor expectations and decision-making in modern markets. However, investors face challenges due to a lack of expertise and information, leading to a disconnect between corporate performance and investor expectations. To bridge this gap, more analysts are providing professional forecasting (Byun and Roland, 2021). Analysts assess corporate data, including macroeconomic announcements, earnings reports, and competitor performance metrics (Lobo et al., 2017; Mangat et al., 2022). By synthesizing public and private information into earnings forecasts, analysts streamline valuation, helping investors gather and process data, shape expectations, and reduce uncertainty—key to maximizing investor utility (Call et al., 2013; Leung et al., 2023; Msomi and Kunjal, 2024; Ou and Wang, 2024; Senfi et al., 2024). This raises critical questions: To what extent do financial data analysts provide influence a stock's value? How can the value of these forecasts be accurately measured?

Data valuation has gained academic attention, but remains uncertain due to the lack of standardized units and varying use (Leonelli, 2019; Li and Hall, 2020; Veldkamp, 2023). Data valuation methods in nonfinancial corporations are categorized as follows. First, a cost-based approach estimates data value by evaluating the costs of data collection, integration, and analysis (Goodridge and Haskel, 2015). Second, an income-based approach calculates data value by discounting expected (Li et al., 2019). Third, a market-based approach determines data value by referencing the transaction prices of comparable data products on the market (Ker and Mazzini, 2020). Some scholars integrate data into economic activities, assessing its value by measuring changes in performance metrics before and after data usage, thereby quantifying its impact (Coyle and Manley, 2024; Li et al., 2024). However, financial data, e.g., transaction, market volatility, and analyst forecasts, are critical in financial product pricing, risk management, and market forecasting. Farboodi et al. (2024) assessed the value of financial data by evaluating its contribution to investors' returns; they measured changes in excess returns before and after financial data was incorporated into investor information.

* Corresponding author.

E-mail addresses: lizh@gzhu.edu.cn (Z. Li), 1112364006@e.gzhu.edu.cn (Y. Xu), duziqing@gzhu.edu.cn (Z. Du).

Existing literature that quantifies changes in performance metrics before and after data utilization offers key insights for financial data valuation. This study extends Farboodi et al. (2024) by applying the methodology to the Chinese market and underlying mechanisms. It analyzes quarterly data from A-share listed companies (2008–2023) to evaluate financial data value by assessing changes in investor welfare before and after incorporating analyst data. The study's contributions are as follows. First, it comprehensively examines how data value manifests in investor welfare while addressing the Matthew effect, where data benefits are disproportionately concentrated among certain investors. Second, it provides a deeper understanding of how market forces and data interact by examining the impact of different market structures—perfect competition and price impact—on data valuation. Third, unlike prior studies that treat data valuation as uniform, this study explores the heterogeneity of data value across underlying assets.

2. Study design

2.1. Theoretical analysis and research hypothesis

The value of analyst data lies in its ability to reduce investor uncertainty and enhance welfare. Financial markets often operate under the semi-strong form of the efficient market hypothesis, where information is not fully transparent. Analysts bridge this gap by offering in-depth analysis and forecasts. The framework for valuing financial data is as follows.

Assets. Let us consider N distinct risky market assets indexed by j . Net supply is denoted as \bar{x} . Each asset's dividend stream is $\{d_{jt}\}_{t=0}^{\infty}$, where vector d_t follows an autoregressive process given by Eq. (1). Here, the exogenous dividend shock $y_{t+1} \sim N(0, \Sigma_d)$ is independently and identically distributed across time periods. t indicates that the variable is known at the end of period t .

$$d_{t+1} = \mu + G(d_t - \mu) + y_{t+1} \quad (1)$$

Investors. Each investor i born at time t has an initial wealth \bar{w}_{it} while engaging in lifetime consumption c_{it+1} . At time t , investor i selects a portfolio of risky assets. Each asset's share is denoted by q_{it} . They also select a risk-free asset with a return rate of r . Their budget constraint is formulated as Eq. (2), where the set of investable assets is $\mathcal{C}_i = Re^N$. Matrix θ_i is $|\mathcal{C}_i| \times N$, a binary matrix. $|\mathcal{C}_i|$ represents the number of investable assets for investor i . Each row of θ_i contains a single 1 and zero otherwise.

$$c_{it+1} = r(\bar{w}_{it} - q_{it}'\theta_i p_t) + q_{it}'\theta_i(p_{t+1} + d_{t+1}) \quad (2)$$

Data. Each investor has access to H distinct data sources. Each source h provides information on dividend volatility y_{t+1} , modeled as:

$$\eta_{iht} = \psi_h y_{t+1} + \Gamma_h e_{it} \quad (e_{it} \sim \mathcal{N}(0, I)) \quad (3)$$

Beginning each period t , investors possess an information set I_t^- . Observing data signals, investors incorporate this information into priors to form posterior information sets I_{it} :

$$I_{it} = \{I_t^-, \eta_{it}, p_t\} \quad (4)$$

Upon this, they select portfolios q_{it} that maximize expected utility $E[U(c_{it+1})|I_{it}]$.

Value of data. The payoff for investor i from purchasing assets in the investable set is $R_t = \Pi_{it} \odot \theta_i p_t$. Investor i 's expected utility can be expressed in a perfectly competitive market as follows:

$$\tilde{U}(I_{it}) \approx \frac{1}{2} E[R_{it}]' E[V[R_{it}|I_{it}]^{-1}] E[R_{it}] + \frac{1}{2} \text{Tr}[V[R_{it}] V[R_{it}|I_{it}]^{-1} - I] + r p_i \bar{w}_{it} \quad (5)$$

In the presence of price impact dp/dq_j , the expected utility becomes

$$\tilde{U}(I_{it}) \approx E[R_{it}]' \hat{V}_{it}^{-1} E[R_{it}] + \text{Tr}[(V[R_{it}] - V[R_{it}|I_{it}]) \hat{V}_{it}^{-1}] + r p_i \bar{w}_{it} \quad (6)$$

where

$$\hat{V}_{it}^{-1} = \left(I - \frac{1}{2} V[R_{it}|I_{it}] \tilde{V}_{it}^{-1} \right) \tilde{V}_{it}^{-1} \quad (7)$$

and

$$\tilde{V}_{it}^{-1} = V[R_{it}|I_{it}] + \frac{1}{\theta_i p p' \theta_i} \frac{dp}{dq} \quad (8)$$

Consequently, the value of data for investor i can be quantified as follows:

$$\text{Value of Data}_i(\text{RMB}) = \frac{1}{p_i} \tilde{U}(I_{it} \cup \text{data}) - \tilde{U}(I_{it}) \quad (9)$$

Assuming a standard utility function equivalent to the standard exponential utility function, the absolute risk aversion coefficient ρ

satisfies:

$$\frac{c^{1-\sigma}}{1-\sigma} = -\exp^{-\rho c} \quad (10)$$

This formula indicates that the absolute risk aversion coefficient decreases as wealth increases. Thus, wealthier investors exhibit lower absolute risk aversion and are more willing to take on risk and assign greater value to financial data. Wealthier investors tolerate more risk, placing a higher value on the same data.

H1. Data value is indicated by investor welfare.

The impact of market competition on the value of analyst financial data varies significantly between perfect competition and price-impact markets. In perfect competition, prices are determined by public information, with no single investor able to influence them, as outlined by the efficient market hypothesis (Chen et al., 2020). Analyst data provide valuable insights, giving investors a decision-making advantage. The value of analyst data decreases in price-impact markets as high-wealth investors directly influence prices through trade actions (Beyer and Guttman, 2011; Lobão et al., 2024). Investors leverage market power to drive price movements, exploiting market sentiment and irrational behavior to achieve excess returns. Smaller investors tend to be more reactive, adjusting to market changes rather than shaping them.

H2. Financial data value depends on market competition.

The value of financial data varies by asset size, analyst attention, and transparency. Information asymmetry affects market efficiency and asset allocation, with analysts helping reduce these gaps (Attilio, 2024; Li et al., 2022; Liao et al., 2021; Ou and Wang, 2024; So, 2013). Analyst data are most valuable for medium-sized assets, where balanced asymmetry and market attention maximize impact. Value declines due to information saturation for large assets, while small assets benefit the least due to low visibility and limited resources. Low analyst attention limits access to reliable information, reducing data utility. High attention enables markets to absorb data quickly, leading to more accurate pricing. In highly transparent markets, prices reflect information almost immediately, leaving little room for additional data to add value. Less transparent markets increase information asymmetry, where analyst data are crucial for guiding investment decisions.

H3. Financial data value exhibits heterogeneity across differing underlying assets.

2.2. Valuation model

For each asset in period t , we calculated the excess return R_{jt} :

$$R_{jt} = \frac{p_{jt+1} - d_{jt+1} - p_{jt}}{p_{jt}} - r_t \quad (11)$$

and the analysts' forecast revision ratio, i.e., the current forecast divided by the previous.¹ These metrics were weighted by the asset's market capitalization to obtain unconditional expected return $E[R_t]$ and variance $V[R_t]$. We projected R_t onto the available information set (valuation data X_t and control variable Z_t already in investors' possession) and solely on Z_t :

$$R_t = \beta_1 X_t + \beta_2 Z_t + \varepsilon_t^{XZ} \quad (12)$$

$$R_t = \gamma_2 Z_t + \varepsilon_t^Z \quad (13)$$

We estimated the conditional variances $V[R_t|X_t, Z_t]$ and $V[R_t|Z_t]$. We quantified the informational gains to determine the data value.

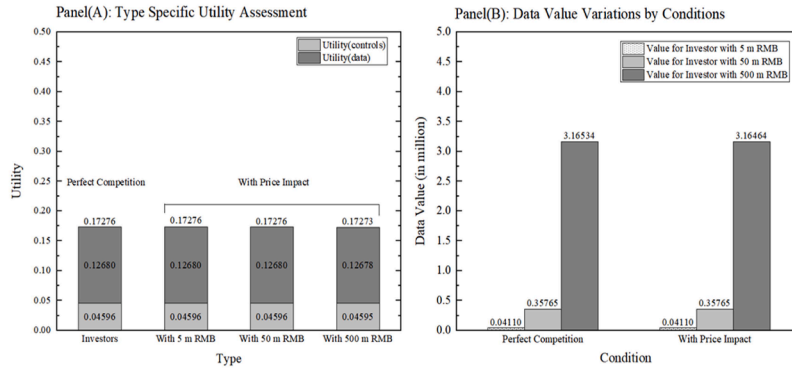
3. Valuing financial data

3.1. Sample selection and data

This study employed data from the National Bureau of Statistics and China Stock Market & Accounting Research (CSMAR) database to construct a panel of 144 A-share listed firms (2006Q2–2023Q4; 71 quarters). We averaged their predictions for each firm–quarter to mitigate the inherent noise and subjectivity in multiple concurrent analyst forecasts. Price represents the closing stock price at quarter-end; dividends encompass total pretax per-share dividends distributed quarterly. The risk-free rate was based on the average three-month China bond Government Securities Yield for each quarter. Market capitalization indicates corporate value at quarter-end; the CSI300 dividend yield was a relevant investor consideration. To eliminate the influence of price factors, we deflated the cash flow data of listed companies (total market capitalization, dividends, and price). We employed the GDP deflator (with 2010 as the base year) to deflate the financial data.

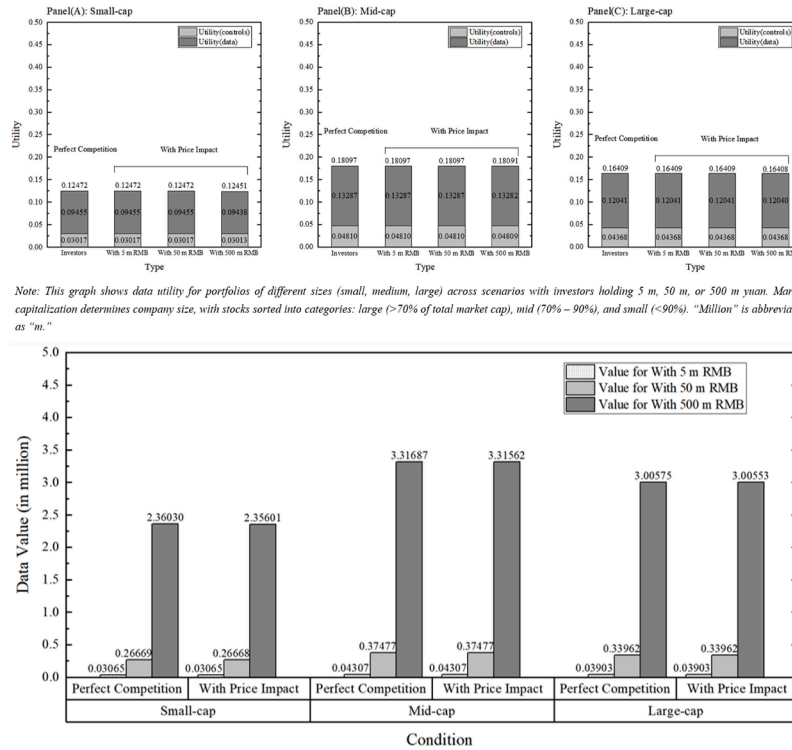
Following Frazzini et al. (2018), we posit that trading 2.5 % of a stock's daily volume leads to a 15-basis points price impact.

¹ The analyst forecast revision ratio serves as a proxy for financial data, reflecting market adjustments to changes in a company's financial condition and the broader economic environment.



Note: Panel (A) shows data utility: the first bar represents investors in a perfect competition market, and the next three represent investors with 5 m, 50 m, and 500 m yuan in a market with price impact. Panel (B) shows data values for investors under different conditions: left for perfect competition, right for price impact. Data values are in millions. "Million" is abbreviated as "m."

Fig. 1. Financial data value and investor wealth across market types.



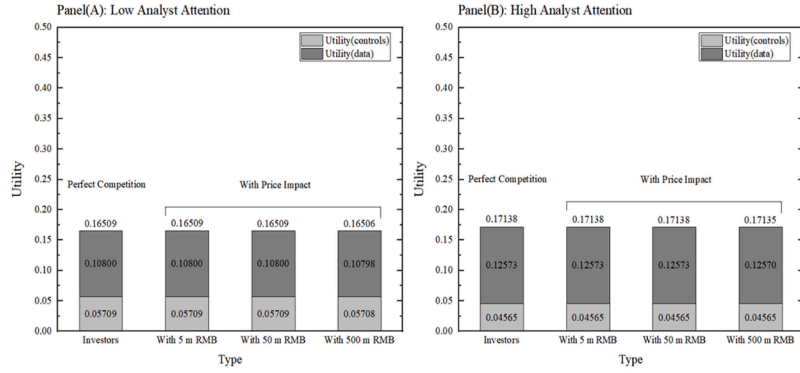
Note: This graph shows data values for portfolios of varying sizes (small, medium, large) across scenarios with investors holding 5 m, 50 m, or 500 m yuan. Market capitalization determines company size, with stocks sorted into categories: large (>70% of total market cap), mid (70%–90%), and small (<90%). Data values are in millions. "Million" is abbreviated as "m."

Fig. 2. Financial data value and underlying asset size.

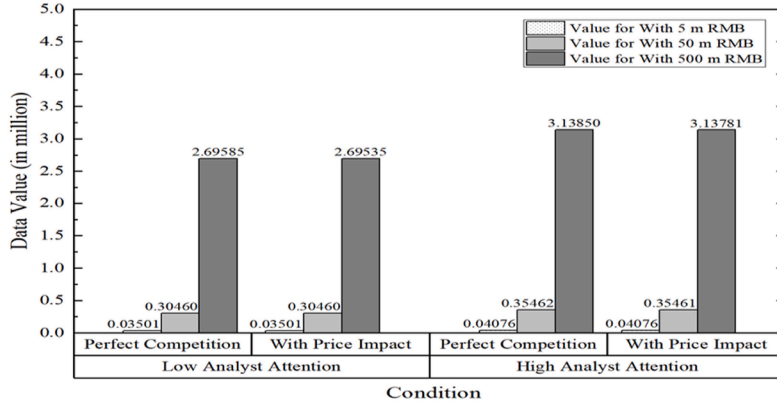
Assuming an annual turnover of 100 % over 250 trading days, we calculated an elasticity:

$$\frac{dp}{dq} = 1.5 \quad (14)$$

A market capitalization of 1 billion RMB translates to:



Note: This figure shows data utility for portfolios with varying levels of analyst attention (low, high). The criterion for analyst attention is based on the average number of analysts who have tracked and analyzed the company. High attention means more analysts than average; low attention means fewer. "Million" is abbreviated as "m."



Note: This figure shows data for portfolios with varying levels of analyst attention (low, high), based on the average number of analysts tracking the company. High attention means more analysts; low attention means fewer. Data values are in millions. "Million" is abbreviated as "m."

Fig. 3. Financial data value and analyst attention.

$$\lambda = \frac{\frac{dp}{dq}}{\theta_i p p' \theta_i} = 1.5 \times 10^{-8} \quad (15)$$

3.2. Value of financial data results

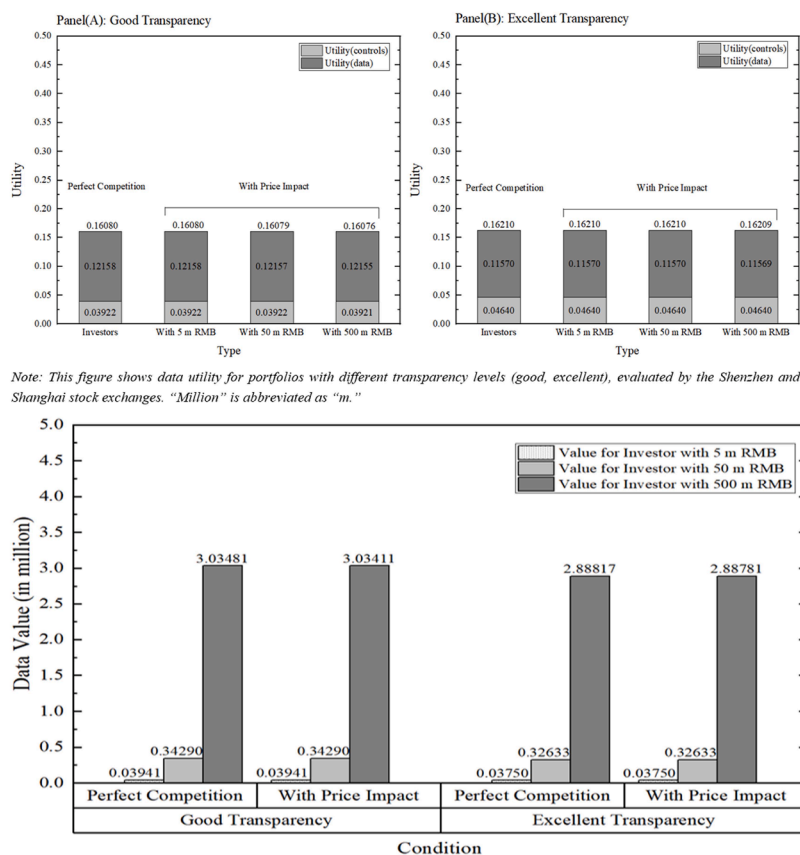
Analyst forecast data enhanced investor utility, with a Matthew effect observed. High-wealth investors, benefiting from well-diversified portfolios, gained even more value from analyst data (Fig. 1). This disparity stemmed from differences in asset allocation and risk tolerance. Their stronger risk tolerance and market position enabled them to leverage data effectively for investment decisions and risk management, exploiting information gaps and adopting bolder strategies to maximize returns.

Moreover, the price impact diminished the value of analyst data for high-wealth investors from a market competition perspective. In a market with a price impact, high-wealth investors gain relatively less utility from analyst data (Fig. 1), reducing the value of financial data. Compared to a perfectly competitive market, the trading behavior of high-wealth investors directly drove market price changes, making the incremental utility of analyst data lower for them than for middle- and low-wealth investors in the market with price impact (H2).

4. Further analysis

4.1. Financial data value for differing underlying asset sizes

The value of financial data varied across asset sizes, particularly in price-impact markets. Analyst data were most valuable for medium-sized assets, followed by large ones, with the least value for small assets (Fig. 2). As asset size increased, the diminishing effect of price impact on financial data value for high-wealth investors decreased. This was most significant for small-sized underlying assets,



Note: This figure shows data utility for portfolios with different transparency levels (good, excellent), evaluated by the Shenzhen and Shanghai stock exchanges. "Million" is abbreviated as "m."

Note: This figure shows data values for portfolios with different transparency levels (good, excellent), evaluated by the Shenzhen and Shanghai stock exchanges. Data values are in millions. "Million" is abbreviated as "m."

Fig. 4. Financial data value and differing transparency.

where high-wealth investors more easily influenced prices through trading (H3). Large-sized underlying assets experience had minimal price impact due to high liquidity. Medium-sized underlying assets attracted more market attention than small but offered less comprehensive disclosure than large underlying assets, making analyst data most valuable.

4.2. Financial data value for differing analyst attention

The value of analyst data increased as analyst attention on the underlying asset portfolio rose. The high-attention underlying asset provided greater utility by reducing market noise and addressing information asymmetry for investors (Fig. 3). When analyst coverage was low, all investors, regardless of wealth, benefited less due to the lack of available information. For high-wealth investors, the price impact further diminished analyst data value for high-attention underlying assets. Despite additional market information, large trades can directly influence prices, making the incremental value of analyst data relatively lower. In well-covered markets, abundant information makes prices more responsive, reducing reliance on analyst insights (H3).

4.3. Financial data value for different transparency

There was an inverse relationship between the value of analyst forecast data and underlying asset transparency. For the underlying asset with excellent transparency, investor utility and data value are lower than those with good transparency. In highly transparent markets, the impact of price effect on reducing data value for high-wealth investors was minimal (Fig. 4). When an asset has excellent transparency, enough information is already available, limiting the additional value that analyst data can provide. As prices reflected most information, the utility and value of analyst forecasts decreased, reducing the need for new data. For transparent assets, analyst data filled information gaps, reduced asymmetry, and provided greater value to investors. Reliance on analyst data increased, boosting its value. In highly transparent markets, timely disclosure of large trades ensured fairer price adjustments, preventing high-wealth investors from significantly influencing price movements, as they might in less transparent markets (H3).

5. Conclusion

Financial data value is crucial for understanding how information asymmetry impacts investor decision-making and welfare in semi-strong, efficient markets. This study has established a theoretical foundation for valuing financial data. It examined valuation across investor wealth levels and market competition. We discussed heterogeneity across underlying assets. The main findings are as follows. First, financial data value is indicated by investor welfare; we observed the Matthew effect across investor wealth levels. Second, the trading behavior of high-wealth investors influences asset prices more than perfect competition markets, diminishing the value of financial data in markets with price impact. Third, financial data value across underlying assets was heterogeneous; market size showed a left-skewed distribution and higher value in underlying assets with high analyst attention and excellent transparency.

Future research should explore financial data valuation dynamics under varying market conditions (e.g., price effects across markets and the role of macroeconomic factors). Since market sentiment is pivotal in determining short-term market trends, its inclusion in the information set may significantly affect investment decisions. While this study draws on Frazzini et al. (2018) regarding price effects, further research should examine how effects vary across markets, particularly in the context of the Chinese stock market. Macroeconomic factors should be incorporated to enhance the understanding of data valuation in diverse economic environments.

Funding

This project was supported by Major Program of National Fund of Philosophy and Social Science of China (Grant No. 22&ZD163).

Statement on generative AI and AI-assisted technologies

During the preparation of this work no generative AI and AI-assisted technologies were used in the writing process.

CRediT authorship contribution statement

Zhenghui Li: Writing – review & editing, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Yanting Xu:** Visualization, Software, Methodology, Data curation. **Ziqing Du:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis.

Declaration of competing interest

The authors declare no competing interests.

Data availability

The authors do not have permission to share data.

References

- Attilio, L.A., 2024. Impact of Chinese financial shocks: a GVAR approach. *Natl. Account. Rev.* 6, 27–49. <https://doi.org/10.3934/NAR.2024002>.
- Beyer, A., Guttman, I., 2011. The effect of trading volume on analysts' forecast bias. *Account. Rev.* 86, 451–481. <https://doi.org/10.2308/accr.00000030>.
- Byun, S., Roland, K., 2021. Analyst bias and forecast consistency. *Account. Financ.* 61, 5403–5437. <https://doi.org/10.1111/acfi.12763>.
- Call, A.C., Chen, S., Tong, Y.H., 2013. Are analysts' cash flow forecasts naive extensions of their own earnings forecasts? *Contemp. Account. Res.* 30, 438–465. <https://doi.org/10.1111/j.1911-3846.2012.01184.x>.
- Chen, L., Du, Z., Hu, Z., 2020. Impact of economic policy uncertainty on exchange rate volatility of China. *Financ. Res. Lett.* 32, 101266. <https://doi.org/10.1016/j.frl.2019.08.014>.
- Coyle, D., Manley, A., 2024. What is the value of data? A review of empirical methods. *J. Econ. Surv.* 38, 1317–1337. <https://doi.org/10.1111/joes.12585>.
- Farboodi, M., Singal, D., Veldkamp, L., Venkateswaran, V., 2024. Valuing financial data. *Rev. Financ. Stud.* <https://doi.org/10.1093/rfs/hhae034>.
- Frazzini, A., Israel, R., Moskowitz, T.J., 2018. Trading costs. Available at SSRN 3229719. <https://doi.org/10.2139/ssrn.3229719>.
- Goodridge, P., Haskel, J., 2015. How Much is UK Business Investing in Big data? Imperial College, London. No. 25159. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.926.243&rep=rep1&type=pdf>.
- Ker, D., Mazzini, E., 2020. Perspectives on the Value of Data and Data Flows. OECD Publishing, Paris. <https://doi.org/10.1787/a2216bc1-en>. No. 299.
- Leonelli, S., 2019. Data—from objects to assets. *Nature* 574, 317–320. <https://doi.org/10.1038/d41586-019-03062-w>.
- Leung, M.F., Jawaid, A., Ip, S.W., Kwok, C.H., Yan, S., 2023. A portfolio recommendation system based on machine learning and big data analytics. *Data Sci. Financ. Econ.* 3, 152–165. <https://doi.org/10.3934/DSFE.2023009>.
- Li, W.C.Y., Hall, B.H., 2020. Depreciation of business R&D capital. *Rev. Income Wealth.* 66, 161–180. <https://doi.org/10.1111/roiw.12380>.
- Li, W.C.Y., Makoto, N., Kazufumi, Y., 2019. Value of data: there's no such thing as a free lunch in the digital economy. RIETI discussion paper. <http://www.rieti.go.jp/jp/publications/dp/19e022.pdf>.
- Li, Z., Lai, Q., He, J., 2024. Does digital technology enhance the global value chain position? *Borsa Istanbul. Rev.* 24, 856–868. <https://doi.org/10.1016/j.bir.2024.04.016>.
- Li, Z., Zou, F., Mo, B., 2022. Does mandatory CSR disclosure affect enterprise total factor productivity? *Ekon. Istraz.* 35, 4902–4921. <https://doi.org/10.1080/1331677X.2021.2019596>.
- Liao, G., Hou, P., Shen, X., Albitar, K., 2021. The impact of economic policy uncertainty on stock returns: the role of corporate environmental responsibility engagement. *Int. J. Financ. Econ.* 26, 4386–4392. <https://doi.org/10.1002/ijfe.2020>.
- Lobão, J., Pacheco, L., Naia, M.B., 2024. Navigating the herd: the dynamics of investor behavior in the Brazilian stock market. *Quant. Financ. Econ.* 8, 635–657. <https://doi.org/10.3934/QFE.2024024>.

- Lobo, G.J., Song, M., Stanford, M.H., 2017. The effect of analyst forecasts during earnings announcements on investor responses to reported earnings. *Account. Rev.* 92, 239–263. <https://doi.org/10.2308/accr-51556>.
- Mangat, M.K., Reschenhofer, E., Stark, T., Zwatz, C., 2022. High-frequency trading with machine learning algorithms and limit order book data. *Data Sci. Financ. Econ.* 2, 437–463. <https://doi.org/10.3934/DSFE.2022022>.
- Msomi, S., Kunjal, D., 2024. Industry-specific effects of economic policy uncertainty on stock market volatility: a GARCH-MIDAS approach. *Quant. Financ. Econ.* 8, 532–545. <https://doi.org/10.3934/QFE.2024020>.
- Ou, R., Wang, Q., 2024. Are analysts' forecasts reliable? A machine learning-based analysis of the target price accuracy. *J. Behav. Financ.* 1–17. <https://doi.org/10.1080/15427560.2024.2368151>.
- Senfi, S., Sheikh, R., Sana, S.S., 2024. A portfolio selection using the intuitionistic fuzzy analytic hierarchy process: a case study of the Tehran Stock Exchange. *Green Financ.* 6, 219–248. <https://doi.org/10.3934/GF.2024009>.
- So, E.C., 2013. A new approach to predicting analyst forecast errors: do investors overweight analyst forecasts? *J. Financ. Econ.* 108, 615–640. <https://doi.org/10.1016/j.jfineco.2013.02.002>.
- Veldkamp, L., 2023. Valuing data as an asset. *Rev. Financ.* 27, 1545–1562. <https://doi.org/10.1093/rof/rfac073>.