# FMLens: Towards Better Scaffolding the Process of Fund Manager Selection in Fund Investments

Longfei Chen, Chen Cheng, He Wang, Xiyuan Wang, Yun Tian, Xuanwu Yue, Kam Kwai Wong, Haipeng Zhang, Suting Hong, and Quan Li

**Abstract**—The fund investment industry is heavily reliant on the expertise of fund managers, who are entrusted with the responsibility of managing portfolios on behalf of clients. Due to their investment knowledge and professional skills, fund managers are able to gain a competitive advantage over the average investor in the market. Consequently, investors have shown a preference for investing in fund managers as opposed to investing directly in funds. For this group of investors, the key issue becomes selecting a suitable fund manager. Previous studies have utilized quantitative or qualitative methods to analyze various aspects of fund managers, such as performance metrics, personal characteristics, and performance persistence. However, these studies have mostly focused on individual analyses and face challenges when dealing with a large candidate space. Furthermore, it is difficult to differentiate whether a fund manager's performance is due to skill or luck, and to meet the preferences of investors in the selection process. To address these challenges, this study characterizes the requirements of investors in selecting suitable fund managers and proposes an interactive visual analytics system called *FMLens*. This system facilitates the fund manager selection process and helps investors to efficiently evaluate and deconstruct fund managers' investment styles and capabilities from multiple dimensions. Additionally, the system enables investors to inspect and compare fund managers' performance. Two case studies and a qualitative user study demonstrate the effectiveness of the approach, and feedback from domain experts suggests that the system can analyze fund managers from multiple perspectives and improve the efficiency of fund manager evaluation and selection.

**Index Terms**—Financial Data, Fund Manager Selection, Visual Analytics.

## 1 Introduction

With the growing financial markets, funds are becoming a popular investment option. The performance of funds relies heavily on their administrators, known as fund managers, who are entrusted with managing investment portfolios on behalf of individual and institutional clients. Fund managers possess investment expertise and professional skills that give them to secure a competitive advantage over general investors, often resulting in higher returns. The growing size of the fund market has made the value of fund managers increasingly apparent, resulting in more investors tending to invest in fund managers rather than directly in funds [1]. This phenomenon can be explained in three ways: *(I) Consistent performance.* Selecting fund managers with a proven track record enables investors to expect a replication of their past successes in managing other funds, resulting in better performance over time, even if the fund is new or has a shorter track record [2], [3]. *(II) Consistency of investment philosophy.* A fund manager's investment philosophy, which typically includes their approach to risk management, research, and decision-making, can strongly influence a fund's performance [4]. Investors who share the fund manager's philosophy may prioritize the manager over the fund. *(III) Reducing the impact of short-term fluctuations.* Focusing on the fund manager rather than the fund reduces the impact of short-term fluctuations, allowing investors to rely on the fund manager's expertise to navigate volatile markets and make strategic adjustments over time.

Conventional approaches to assessing and selecting fund managers involve performance measurement [5], [6], persistence analysis [3], [7], and manager characteristics [4], [8]. Although these methods have shown preliminary effectiveness in evaluating fund managers in previous studies, they face several challenges that hinder their ability to support a comprehensive decision-making process for selecting fund managers. **(1) The large pool of candidates with various heterogeneous attributes.** The attributes of specific fund managers are diverse and may include personal background descriptions, industry investment ratios, and performance metrics, among others. Ideally, selecting fund managers from a large pool of candidates should involve integrating a variety of relevant data to provide investors with comprehensive information. However, gathering, analyzing, and comparing information about each fund manager can be time-consuming and laborious. Additionally, biased propaganda from institutions can influence investors' judgment, making it crucial to grasp available information to avoid improper selection of fund managers and jeopardizing fu-


- *Longfei Chen, Chen Cheng, He Wang, Xiyuan Wang, Haipeng Zhang, and Quan Li (corresponding author) are with the School of Information Science and Technology, ShanghaiTech University, and Shanghai Engineering Research Center of Intelligent Vision and Imaging, China. E-mail: chenlf,chengchen,wanghe1,wangxy7,zhanghp,liquan@shanghaitech.edu.cn.*
- *Yun Tian is with CFETS Information Technology (Shanghai) Co., Ltd., China. E-mail: tianyun_zh@chinamoney.com.cn.*
- *Xuanwu Yue is with Baiont Technology and Sinovation Ventures AI Institute. E-mail: xuanwu.yue@connect.ust.hk.*
- *Kam Kwai Wong is with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology. E-mail: kkwongar@connect.ust.hk.*
- *Suting Hong is with the School of Entrepreneurship and Management, ShanghaiTech University. E-mail: hongst@shanghaitech.edu.cn.*


*Manuscript received June 19, 2021; revised April 19, 2022.*

ture profits. **(2) Distinguishing luck from skill.** Due to the uncertainty of financial markets, a fund manager's performance over time is usually the result of a combination of skill and luck [9]. Extreme returns of many funds may be fortuitous, but large enough to have a significant positive impact on the performance of their managers. In other words, some fund managers may outperform over a period of time, but that does not mean they will be able to maintain the same level of performance in the future. In this sense, it is essential to differentiate between skill and luck in evaluating fund managers. **(3) Opaque actions of fund managers.** Despite extensive disclosure requirements, mutual fund investors cannot observe all actions of fund managers [10]. When a fund manager's investment strategy and decision-making process is not transparent, it is difficult for investors to assess the risk level of the fund, which affects their choice and judgment of the fund manager. **(4) Diversification of selection preferences.** Investors also have specific preferences in selecting fund managers. For example, one investor prefers to invest in high-tech industries, while another investor considers annualized returns of the fund manager more. These preferences are often highly individualized and therefore require an interactive mechanism to assist in constructing a diverse selection process. Overall, selecting fund managers requires considering multiple factors and balancing their relevance and importance to make informed decisions that align with investors' preferences and objectives.

The prevailing tools for analyzing fund managers in the current market can be broadly categorized into three types: financial data terminals such as *Wind* and *Bloomberg*, online trading platforms such as *Alipay* and *Tiantian Fund*, and rating agency websites such as *Morningstar*. These tools are specifically designed to offer investors extensive data on fund managers to facilitate comprehensive analysis. Nonetheless, they often provide information at an individual level, making it challenging to conduct cross-sectional comparisons between fund managers, thereby impeding the selection process. To address the challenges, we present an approach to better support investors in selecting fund managers. We begin by summarizing the essential requirements of investors in this selection process. Subsequently, we propose a visual analytics system, named *FMLens* that assists investors in better scaffolding the selection process. Specifically, we encode information on fund managers as feature vectors and transform the candidate space into the feature space. We then deconstruct the fund manager's investment style from multiple dimensions, including investment diversification, performance metrics, and simulated position-transfer strength. By offering different observation perspectives, investors can better differentiate between skill and luck, enabling them to assess the fund manager's real investment ability. Finally, we develop an interactive selection workflow that takes into account investors' personal preferences. The major contributions of our study can be summarized as follows.

- We integrate relevant multi-dimensional information to support the fund manager comparison and screening in various contexts.
- We develop interactive visualizations enhanced with new features to support investors to select the fund manager.
- We demonstrate the efficacy of our approach through two case studies, expert interviews, and a qualitative user study.

# 2 Related Work

## 2.1 Fund Manager Performance Analysis

In recent times, considerable attention has been devoted to examining the performance of fund managers by both scholars and practitioners. Quantitative approaches, such as the Sharpe ratio [11], Jensen's alpha [12], and the Sortino ratio [13], have been widely adopted to evaluate the performance of fund managers. In addition, researchers have investigated qualitative characteristics such as the fund manager's experience, education, gender, and investment philosophy to comprehend their impact on performance [4], [8], [14]. Moreover, William F. Sharpe's style analysis [6] has been employed to identify fund managers with consistent investment styles and strategies. Cremers et al. [15] developed a new measure, dubbed "Active Share", to assess the level of active management in mutual funds. Their study indicates that mutual funds with higher "Active Share" tend to outperform the benchmarks, implying that more active management can result in superior performance. Similarly, Kacperczyk et al. [10] arrived at a similar conclusion based on an investigation of fund managers' unobserved actions. Furthermore, Fama et al. [9] examined the impact of both luck and skill on fund manager performance and found that differentiating between the two can be a daunting task.

Numerous studies have focused on investigating the persistence of fund managers' performance. For instance, Klaas P. Baks [2] conducted cross-sectional regressions on a sample of 2086 funds from the period spanning 1992 to 1999 to determine Jensen's alpha for each period. Baks concluded that there exists persistence in fund manager performance. Similarly, Berk et al. [3] proposed a novel measure of fund performance and discovered that cross-sectional differences persisted for the subsequent ten years, thereby indicating the existence of long-term persistence in fund manager performance.

Our system design is informed by prior research and incorporates several factors that influence fund manager performance, including investment preferences, performance metrics, personal characteristics, and management motivation. Additionally, we enable investors to monitor changes in fund managers' performance metrics over time and offer a retrospective analysis of their historical fund management activities to investigate the continuity of their performance.

## 2.2 Financial Data Visualization

Financial data visualization has become a trending topic [16], [17], [18], [19], [20]. Visualization methods such as line graphs, icons, and tree diagrams are beneficial for, e.g., stock trading and quantitative investment analysis. For example, Schaefer et al. [21] proposed a line graph visualization method that allows displaying a large amount of non-overlapping data and line graphs by coloring the background according to the code. Ziegler et al. [22] used pixel bar charts to display and compare in real time stock, industry sector and country/region returns and volatilities. All of these studies use line charts to obtain additional information. In terms of visual encoding, Sorenson et al. [23] combined continuous line graphs with icon-based event representations to provide financial data to users. Treemap [24] and dendrograms [25] are used to display stock holdings. Other visualization techniques involve quantitative trading. For example, sPortfolio [26] classifies and integrates investment strategies according to their characteristics. The interactive design facilitates strategy selection and verification in large-scale investment strategies. Another work of the authors, iQUANT [27] uses an embedded glyph design to help stock stock traders select promising investment portfolios. In addition, RankFIRST [28] improves on iQUANT by introducing an interpretable factor calculation models and displaying the ranking of factors through a hierarchical slope graph design. For fund visualization, FundExplorer [25] introduces a distorted tree diagram to visualize the composition of a fund's portfolio and remaining stocks, facilitating the retrieval of additional funds for portfolio diversification. Dwyer [29] proposed a 2.5-D graph to plot the results of fund clustering based on stock holdings. Typically, extant research on visual analytics has prioritized the representation of stock data or quantitative trading, with comparatively limited attention paid to fund manager data. In light of this knowledge gap, we advance a novel approach grounded in interactive visualization techniques. This approach aims to enhance investors' ability to efficiently evaluate, compare, and ultimately select fund managers.

### 2.3 Multivariate Time-Series Data Visualization

Fund manager data are essentially time-series with multiple attributes, including but not limited to manager background data, industry investment ratios, and income related data over time. Many studies have explored time-series data [30], [31], [32]. For example, Lei et al. [33] proposed the concept of "visual signature" for financial time-series data to help professional analysts derive immediate and useful visual information from them. Sorenson et al. [23] designed and implemented a scalable visual representation for describing the many discrete timestamped events used by hundreds of thousands of financial market professionals. On the other side, other representative visualization methods focus on the analysis of time-series data with multiple attribute changes, and they are mainly dedicated to efficient visualization design for the purpose of detecting trends and patterns [34], [35], [36], [37], [38]. In this study, we combine novel visualization glyphs with a timeline design to represent multivariate attributes that illustrate how multiple characteristics of fund managers evolve. In addition, we allow target users to interactively compare fund managers across time periods or events to examine their performance.

## 3 Observational Study

### 3.1 Needfinding Interview

To gain insight into the practical needs of fund investors, we worked with eight experts with a minimum of five years of experience in the fund investment field from two sources: (1) Financial technology and investment major in a partner university, and (2) a partner financial institution. We obtained the domain experts' informed consent to disclose their detailed information, including age, gender, occupation, and years of experience (Yrs Exp), which is presented in Table 1. The experts were randomly divided into two groups of equal size and were interviewed via the online meeting platform Zoom. The entire interviews, which were divided into three parts, were audio-recorded with the experts' informed consent and lasted around 40 minutes each.

TABLE 1: The demographic characteristics of the eight domain experts (identified as E1 through E8) who participated in the needfinding interview were collected and analyzed.

| ID | Gender | Age | Occupation | Yrs Exp |
|---|---|---|---|---|
| E1 | Male | 45 | Senior Investment Manager | 17 |
| E2 | Female | 38 | Investment Major, Professor | 11 |
| E3 | Male | 33 | FinTech Major, Professor | 10 |
| E4 | Male | 36 | Financial Analyst | 9 |
| E5 | Male | 34 | Fund Risk Analyst | 7 |
| E6 | Female | 28 | Financial Analyst | 6 |
| E7 | Female | 25 | FinTech Major, PhD | 5 |
| E8 | Male | 27 | Investment Major, PhD | 5 |

During the interviews, we first asked the experts to confirm the value and importance of selecting fund managers in their actual investment process. If so, we inquired about the common methods they use to select fund managers of interest from the market. Secondly, we asked the experts to share their own analytical approaches to selecting fund managers. Thirdly, we investigated which analytical tools the experts frequently use in their daily work and explored any shortcomings or inconveniences associated with their usage. We brainstormed with the experts to identify any reasonable improvements for the identified issues. Subsequently, we organized the data and employed an iterative coding process [39] for analysis.

Based on the needfinding interviews, several findings pertaining to general and professional investors were identified.

**F.1 Fund managers play a very critical role in investment.** Most experts highlight the critical role played by fund managers in investment. E2 mentioned that "*the fund manager drives the performance, so I pay more attention to their track record and investment philosophy than the fund itself*". Furthermore, a number of experts shared instances where selecting the right fund manager led to superior investment outcomes. For example, E4 shared an experience of selecting a relatively unknown fund based on trust in the fund manager's expertise, which resulted in outstanding fund performance over time.

**F.2 Navigating a diverse pool of fund managers is challenging.** Experts noted the difficulty of efficiently accessing and consolidating information on a wide range of fund manager candidates, and pointed out the limitations of existing tools in providing a comprehensive snapshot of the candidate space. As a result, experts spent a considerable amount of time in the initial research process. E1 remarked that "*existing tools don't provide a comprehensive overview,*" while E6 suggested that "*a more streamlined approach would save time and enable us to focus on what really matters — the analysis.*"

**F.3 Investors adopt a multi-faceted approach to assess fund managers.** While different investors prioritize specific factors in the fund manager screening process, a comprehensive assessment is generally advocated. For instance, E5 prioritizes the fund manager's risk control ability, while E7 believes in a "high risk, high return" investment strategy. However, most experts consider multiple factors, including performance, investment philosophy, and risk management, when analyzing fund managers to ensure a thorough understanding of their strengths and weaknesses (as emphasized by E4).

**F.4 Existing tools lack the flexibility to tailor selection strategies to individual preferences.** A common theme among experts was the importance of alignment between an investor's investment philosophy and a fund manager's approach. Experts highlighted the need for tools that allow customization of selection strategies for selecting fund managers to prioritize factors that are most important to them. E6 remarked, *I feel more secure in my investments and achieve better returns when my investment objectives align with the fund manager's approach."* However, experts also noted that current tools provide limited support for the customization of selection strategies. E8 suggested, *"an ideal system would enable me to rank fund managers based on my investment preferences, enabling me to focus on the factors that matter most to me."*

**F.5 Evaluation of fund managers should go beyond the metrics.** The limitations of metrics in evaluating fund managers were highlighted by the experts. As some fund managers manage multiple funds simultaneously, it is essential to examine performance at the individual fund level rather than relying solely on aggregate metrics. E3 emphasized the importance of scrutinizing claims of high annualized returns, stating, *"It is crucial to investigate if such returns are achieved across all managed funds or if they are limited to only one fund, thereby concealing poor performance in other funds."*

**F.6 More information about "unobserved actions" is needed to help understand the fund manager's investment ability.** The term "unobserved actions" refers to the decisions, strategies, or behaviors that are not easily observable or transparent to investors or analysts [10]. Although these actions can have a significant impact on the fund's performance, limited data or a lack of disclosure often make them difficult to assess. For example, the disclosure of a fund's positions by a fund manager is mandated only once in each quarterly report, thereby limiting access of investors to pertinent information at other times. Experts in the study pointed out that current tools do not adequately provide insight into these unobserved actions, leaving investors with an incomplete picture. E1 noted that unobserved actions reflect the fund manager's responsiveness to market events, highlighting the need for better tools to assess these actions."

**F.7 Comparisons of fund managers are necessary.** E4 and E6, expressed the need for comparisons when evaluating fund managers. According to them, existing tools do not adequately support this type of analysis. E6 stated that current tools only integrate fund manager data and do not provide sufficient support for comparisons. This inadequacy hinders their work and makes it difficult to make informed decisions.

### 3.2 Requirement Characterization

Following several rounds of discussion and ideation, the subsequent requirements (**R.1–R.6**) were derived from the needfinding survey.

**R.1 Provide an overview of the fund manager candidate pool.** The first requirement pertains to the need for an efficient way to provide investors with an overview of the available fund manager candidate pool (**F.2**). The overview should include details about the investment style and performance of the fund managers. The aim is to minimize the cost of collecting information and enable investors to quickly and effectively assess the fund manager landscape. A clear and concise visualization of this information can help investors identify managers who align with their investment criteria.

**R.2 Evaluate a fund manager's investment skills in various dimensions.** The second requirement entails the need to evaluate a fund manager's investment skills in various dimensions, such as risk management, asset allocation, and market timing. To facilitate this evaluation, a visual analytics system should offer detailed metrics that allow investors to make informed investment decisions in accordance with their individual investment preferences and risk tolerance, as highlighted by the findings in **F.3**.

**R.3 Personalize ranking strategies based on investor preferences.** Each investor has individual goals and preferences, and our system should take these into account when ranking fund managers (**F.4**). The system should enable investors to assign weights to their specific preferences, such as risk tolerance and returns. By doing so, the system can produce personalized rankings that prioritize fund managers who meet the investor's preferences. This approach will result in a better match between the investor and the selected fund manager.

**R.4 Provide a historical review of management records.** The fourth requirement pertains to the need for a historical review of management records of fund managers, beyond just the consideration of their performance metrics (**F.5**). Specifically, when a single individual manages multiple funds, relying solely on performance metrics may not provide a complete picture of their ability. To address this, our system should provide investors with access to detailed historical management records, enabling them to undertake a more comprehensive analysis of the fund manager's abilities.

**R.5 Simulate the fund manager's position[1] adjustment actions.** As noted by E1, a fund manager's unobserved actions reflect their investment capabilities and are essentially a series of private position adjustments. However, due to their importance and difficulty in tracking (**F.6**), our system should generate daily position simulations using industrial sector indices and historical net values of the fund. By doing so, investors can gain deeper insights into the fund manager's investment style, judgment of the market, and timing, ultimately aiding them in making informed investment decisions.

**R.6 Compare the performance of fund managers.** According to experts, it is essential to provide comparisons

1. Fund manager's position, which refers to the amount of securities, assets, or property held or sold short by an individual or entity

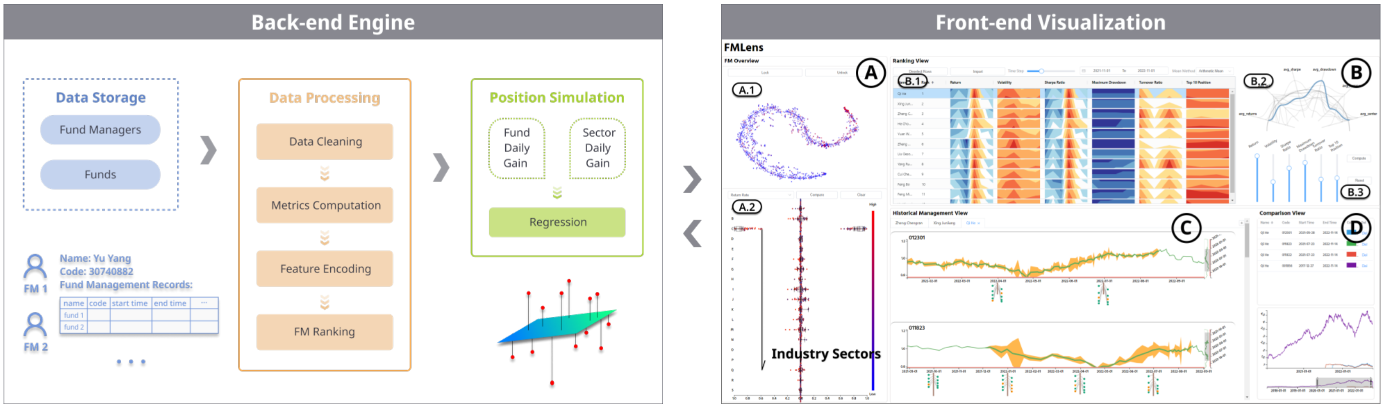



Fig. 1: The architecture of *FMLens* comprises a back-end engine and a front-end visualization. In the back-end engine, the data processing module quantifies the fund manager's characteristics into performance metrics. The simulation module then uses regression methods to simulate the fund manager's position adjustment actions. In the front-end visualization, there are four views: (A) The *FM Overview* serves to provide a summary of the fund manager candidate space, while (B) the *Ranking View* facilitates the examination of fund managers' performance evolution and supports interactive ranking. In addition, (C) the *Historical Management View* offers the ability to review the historical experience of fund managers, and (D) the *Comparison View* provides additional comparative analysis between funds.

between fund managers (**F.7**) after a multi-level summary of their performance. For instance, experts suggested comparing the performance of products managed by fund managers over the same time period. As noted by E2, "*it is not uncommon for fund managers to have the same high returns, yet for different reasons.*" Therefore, identifying the differences between fund managers can assist investors in making more informed decisions.

## 4 Approach Overview

The architecture of *FMLens*, as depicted in Figure 1, is comprised of two main components: a back-end engine and a front-end visualization. In the back-end engine, relevant features pertaining to fund managers' investment capabilities are initially quantified, and high-dimensional performance metrics are subsequently computed from raw data using the data processing module. The simulation module then employs regression methods based on the fund's net value to simulate the fund manager's positions. In the front-end visualization, we provide four well-designed views for analysis. The *FM Overview* serves to provide a summary of the fund manager candidate space. The *Ranking View* facilitates the examination of fund managers' performance evolution and supports interactive ranking. In addition, the *Historical Management View* provides to review the historical experience of fund managers, and the *Comparison View* provides additional comparative analysis between funds.

## 5 Back-end Engine

### 5.1 Data Description and Processing

TABLE 2: Two datasets used in our study.

| Dataset | Capacity | Characteristics Dimensions |
|---|---|---|
| Fund Manager | 2047 | *code, name, fund management records* |
| Fund | 5439 | *code, name, size, daily net value, turnover ratio, top 10 position* |

This study utilizes data sourced from a publicly available investment website[2]. As shown in Table 2, equity fund managers were filtered out from this dataset, yielding a total of 2047 fund managers. Each entry in this dataset corresponds to a distinct fund manager and includes pertinent information, such as the fund manager's *code*, *name*, *employment duration*, and *fund management records* - the latter of which encompass the manager's historical fund management experience and include details such as the fund's *code*, *name*, *management start time*, and *end time*. In order to accurately calculate fund manager performance metrics, it was necessary to obtain additional information about the fund through the fund code and a publicly available data API[3]. As a fund may be managed by multiple fund managers at different points in time and subsequently appear in the management records of different managers, the issue of duplicate queries was addressed by first aggregating all funds that appeared in the management records and subsequently querying and storing data from all 5439 funds. Each entry in this resultant dataset relates to a unique fund and comprises various informative variables, including *daily net value*, *size*, *industry allocation*, *turnover ratio*, and *top 10 position*. Notably, the *daily net value* data is time-series in nature and is updated at the conclusion of each trading day, while the other data points are discrete and disclosed quarterly in the fund's financial reports.

### 5.2 Performance Metrics Computation

Drawing from both relevant literature [3], [40] and expert recommendations, the following 6 performance metrics have been identified for the purpose of evaluating the investment style and capabilities of fund managers:

- ***Return***: This metric, also referred to as financial return, denotes the net profit or loss generated by investment within a specified time frame.

2. http://fund.eastmoney.com
3. https://www.ricequant.com/doc/rqdata/python/fund-mod.html

- ***Volatility***: This statistical measure reflects the extent of variation in returns of a given security or market index.
- ***Sharpe Ratio***: This metric compares the return of an investment with its risk level, using a mathematical expression that accounts for the potential impact of excess returns over time in signaling greater volatility and risk than investment skill.
- ***Maximum Drawdown***: This metric is the maximum observed loss from the peak to the trough of a portfolio before a new peak is reached and serves as an indicator of downside risk over a given period.
- ***Turnover Ratio***: This metric represents the proportion of holdings within a mutual fund or other portfolios that are replaced within a given time period.
- ***Top 10 Position***: This metric pertains to the top ten stocks held within the fund's position, expressed as a percentage of net value.

It is noteworthy that a fund manager's performance evaluation is solely based on the funds managed, which implies that the aforementioned metrics are calculated on a per-fund basis [3], [12], [40]. For instance, suppose a fund manager is currently managing two funds, A and B. In the past three months, fund A has yielded a return of $20\%$, while fund B has yielded a return of $40\%$. Then the fund manager's return is the average ($30\%$) of the returns of both funds, regardless of their size. As per expert recommendations, the fund manager's performance metrics are presented from a time-series perspective. Based on the time-series intervals, these metrics can be categorized into "arbitrary interval" and "fixed interval" performance metrics. The "arbitrary interval" performance metrics consist of *Return*, *Volatility*, and *Maximum Drawdown*.

$$\mathit{Return} = \sum_{k=1}^{n} w_k \times \frac{P_k^{(j)} - P_k^{(i)}}{P_k^{(i)}}, \tag{1}$$

where $n$ represents the number of funds managed by the fund manager in the interval, $P_k^{(i)}$ denotes the net value of the $k^{th}$ fund at the beginning of the interval, $P_k^{(j)}$ denotes the net value of the $k^{th}$ fund at the end of the interval, and $w_k$ represents the weighting, which is dependent on the averaging method;

$$\mathit{Volatility} = \sum_{k=1}^{n} w_k \times \sigma_k, \tag{2}$$

where $\sigma_k$ represents the standard deviation of the daily net value of the $k^{th}$ fund within the interval; and

$$\mathit{Sharpe\ Ratio} = \sum_{k=1}^{n} w_k \times \frac{R_p^{(k)} - R_f^{(k)}}{\sigma_p^{(k)}}, \tag{3}$$

where $R_p^{(k)}$ denotes the return of the $k^{th}$ fund within the interval, $R_f^{(k)}$ denotes the risk-free rate of return of the $k^{th}$ fund, and $\sigma_p^{(k)}$ denotes the standard deviation of the excess return of the $k^{th}$ fund. Since the net value of the fund is updated on a daily basis, the values of the "arbitrary interval" metrics can be computed for any time interval. On the other hand, the "fixed interval" performance metrics, namely *Maximum Drawdown*, *Turnover Ratio*, and *T op 10 Position*, cannot be calculated using publicly available data, as these data are only published quarterly, which restricts their time intervals to three months.

### 5.3 Fund Manager Ranking

Once the performance metrics have been calculated, a multi-dimensional ranking approach is developed. This approach considers the above-mentioned six performance metrics as optional dimensions based on expert recommendations. As these dimensions differ in their scales and magnitudes, a Z-score normalization [41] is employed to transform the score of each fund manager in a specific dimension into a standard score. These standardized scores are then weighted based on user-defined weights to obtain a total score for each fund manager, which is then used to rank them.

### 5.4 Position Simulation

In order to assess the degree of abrupt changes in fund positions, machine learning-based techniques have been proposed to evaluate fund portfolios. These methods can be classified into two groups, namely *regression-based* [42], [43], [44] and *deep learning-based* methods [45], [46]. In view of computational efficiency, we have opted for the linear regression method since real-time calculations of positions are required on a daily basis. The common approach currently used to measure fund position is based on index simulation [42], [47], [48]. This involves carrying out regression calculations by using fund net worth data and index point data. Following this conventional approach, we have devised the following regression equation:

$$N_{f,t} = \sum_{i=1}^{n} \gamma_{i,t} I_{i,t} + \varepsilon_t \tag{4}$$

Here, $N_{f,t}$ represents the net worth of fund $f$ on the $t_{th}$ day, $I_{i,t}$ represents the index of the $i_{th}$ sector on the $t_{th}$ day, $\gamma_{i,t}$ denotes the respective share of stocks in sector $i$ on the $t_{th}$ day, which serves as the regression coefficient, and $\varepsilon_t$ represents the residual term. The total position $\sum_i \gamma_{i,t}$ is publicly announced in the quarterly report. The degree of change intensity is gauged by $\sum_i (\gamma_{i,t+1} - \gamma_{i,t})^2$. We compared three regression techniques (principle component regression (PCR) [49], ridge regression [50], and lasso regression [51]) for 90 common stock funds between the end of the third quarter of 2020 and the end of the third quarter of 2021, for a total of five quarter-end cross-sections. We used fund return data from $T - 14$ to $T + 15$ trading days and 28 primary sector index return data to determine the fund's position weights at the close of the $T_{th}$ trading day. We then compared the results by calculating the average of total fitted positions with the true values released in the quarterly reports, and measured the variances. According to government regulations, the lower limit of holding position for common stock funds is 80%; therefore, we have set the range of position forecasts to $[0.8, 1]$. If the predicted value calculated by the regression method is outside this range, then the predicted value is treated as being similar to the boundary value. The results are presented in Table 3. Our findings suggest that Lasso regression has greater accuracy, principle component regression is marginally weaker, while ridge regression displays systematic overestimation.

TABLE 3: The performance of three regression methods is evaluated based on the average fitted values for a sample of 90 equity funds.

| Quarter | PCR | | Ridge | | Lasso | |
|---|---|---|---|---|---|---|
| | Mean | Variance | Mean | Variance | Mean | Variance |
| 2020/10/01 – 2020/12/31 | 4.48% | 1.24% | 3.54% | 1.33% | 2.00% | 1.32% |
| 2021/01/01 – 2021/03/31 | 4.63% | 0.89% | 4.15% | 1.41% | 0.93% | 1.07% |
| 2021/04/01 – 2021/06/30 | 5.69% | 0.97% | 9.94% | 1.07% | 1.22% | 1.13% |
| 2021/07/01 – 2021/09/30 | 7.65% | 0.75% | 12.98% | 0.88% | 2.46% | 0.98% |
| 2021/10/01 – 2021/12/31 | 0.47% | 0.62% | 13.39% | 0.60% | 3.12% | 0.69% |

**Time window length sensitivity.** In addition, we assess the sensitivity of the three regression methods with varying time window lengths by selecting three cross-sections at the end of the fourth quarter of 2020, the end of the first quarter of 2021, and the end of the second quarter of 2021 and traversing window lengths from 15 to 59 days (Figure 2). Our findings indicate that in most cases, the mean prediction error of each method stabilizes at a value after the window length exceeds 40 days, indicating convergence of the solutions; no clear pattern is observed for window lengths less than 40 days. It is not recommended to use excessively long window lengths (generally not more than one quarter, about 60 trading days) since the regression coefficients represent the average fund position in the past window period, and this value is used to predict the fund position at the current time. If the window length is too long, the prediction may lag. Based on our assessment, we select Lasso regression with a 30-day time window as our simulation module to fit the intensity and frequency of changes in fund positions.

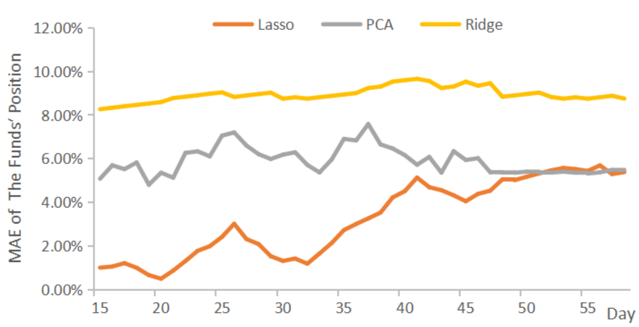



Fig. 2: The present chart displays the variation in the average mean of the length of the time window across three quarters, spanning a range of 15 to 59 days. Notably, the performance of Lasso regression stands out as the best, reaching its nadir at approximately 30 days.

## 6 Front-End Visualization

The basic design principle behind *FMLens* is to leverage and enhance familiar visual metaphors to enable investors to concentrate on analysis. As depicted in Figure 1, we have devised four principal visualizations to assist investors in efficiently evaluating and selecting prospective fund managers: the *FM Overview* that summarizes the fund manager candidate space (**R.1**); the *Ranking View* that demonstrates the evolution of various performance metrics over time and facilitates fund manager ranking based on investor preferences (**R.2 – R.3**); the *Historical Management View* that integrates the past fund management records and simulates their positions to provide investors with insights into the investment styles and abilities of fund managers (**R.4 – R.5**), and the *Comparison View* that enables investors to conduct a detailed comparison of the performance of different fund managers (**R.6**).

### 6.1 FM Overview

The *FM Overview* summarizes the candidate space and helps to identify potential similarities and outliers among fund managers. Dimensionality reduction techniques such as *t-SNE*, *PCA*, and *MDS* have been widely adopted to generate low-dimensional representations that preserve local similarities to convey neighborhood structures [52], [53]. Following the conventional practice, we project all fund managers into a 2D space to view potential clusters and outliers. After discussing with domain experts, we use the amount of money invested by fund managers in each industry sector as the feature dimensions to assess their investment style at the industry level. Specifically, we count each fund manager's cumulative investment amount in 19 primary industry sectors defined by the China Securities Regulatory Commission (CSRS) as the feature vector. We then use *t-SNE* to project these high-dimensional vectors into a 2D space because *"it reveals meaningful insights about data and shows superiority in generating 2D projections"* [54]. As shown in Figure 1(A.1), each fund manager in the projection space is represented by a point, and the color indicates the cumulative returns. Users can lasso any of the points and then the summary chart below (Figure 1(A.2)) will show the specific industry allocation of the lassoed points. In the summary chart, each industry sector corresponds to one of the horizontal axes. Referring to [], the industry codes (A–S) on the left side of the axes and the industry sectors are one-to-one. The lassoed fund manager has a corresponding point on each horizontal axis, and the position of the point depends on the percentage of the fund manager's investment in the corresponding sector. Users can brush entities on any axis and a box plot is displayed to show the distribution of the brushed fund manager's investments in each industry.

### 6.2 Ranking View

The *Ranking View* facilitates the acquisition of fund managers' performance metrics' evolution and overall distribution over time, while also enabling the weighting of the metrics to rank fund managers. This view is comprised of three subviews: the metrics table (Figure 1(B.1)), the semicircular radar chart (Figure 1(B.2)), and the ranking control panel (Figure 1(B.3)).

**Metrics Table.** In the metrics table, six performance metrics, including *Return*, *Volatility*, *Sharpe Ratio*, *Maximum Drawdown*, *Turnover Ratio*, and *Top 10 Position*, are displayed for each fund manager. Horizon charts [55], a type of chart that shows data values over time using multiple time units, are used to represent these metrics. These metrics can be classified into two categories: "arbitrary interval" and "fixed interval", with the former having variable time units and the latter having fixed time units. The x-axis of each chart corresponds to a timeline divided into multiple time units. Users can customize the time range, time interval, and averaging method through the control panel above the metrics table. The time range determines the x-axis range for all metrics, while the time interval adjusts the time unit of the "arbitrary interval" metrics. The averaging method adjusts the weights ($w_k$) used in calculating the composite score of each fund manager, as described in subsection 5.2.

**Semicircular Radar Chart and Ranking Control Panel.** The semicircular radar chart in the *Ranking View* is composed of six concentric circular axes, which represent the average performance metrics of a fund manager over a given time horizon. The points representing the average of each metric for a fund manager are connected by lines to form a polygon, with the distance from the center indicating the magnitude of the metric value. Users can interact with this chart by clicking on a row in the metrics table to highlight the corresponding fund manager's polygon. Moreover, the Ranking Control Panel enables users to assign weights to each metric, which in turn alters the fund managers' rankings in the metrics table. The ranking order can also be modified, either in ascending or descending order, by adjusting the weights of the metrics. For instance, returns are commonly ranked in ascending order, implying that higher returns correspond to higher rankings.

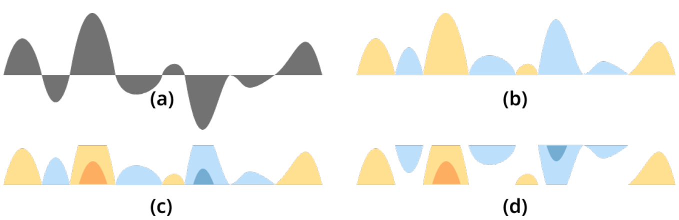


Fig. 3: Design alternatives for the performance metrics.

**Design Alternatives.** Several alternative designs were considered for representing performance metrics over time, as shown in Figure 3. The first design used a simple area chart, depicted in Figure 3(a), with positive and negative values represented by different directions, as the charts were assigned to cells in a table. To save space, a mirrored area chart was proposed in Figure 3(b). However, when displaying performance metrics of multiple fund managers, aligning the vertical axis scales between different charts was necessary to avoid visual misinterpretation. Otherwise, changes in one fund manager's performance could be visually misleading when comparing with other managers' performance on the same scale. To address this issue, a mirrored horizon graph was proposed in Figure 3(c), which splits the mirrored area chart into non-overlapping bands of the same size along the vertical axis and stacks all other bands on top of the band closest to the horizontal axis. This design solved the problem associated with uniform vertical axis scales but resulted in a flipped slope of the negative part, which was not preferred by experts during discussions. Therefore, an offset horizon graph was finally chosen, as shown in Figure 3(d), which uses an offset rather than a mirroring method.

### 6.3 Historical Management View

The *Historical Management View* provides a comprehensive review of a fund manager's management records and helps users understand the performance metrics by fund data. As illustrated in Figure 1(C), all funds managed by a fund manager are displayed from top to bottom in this view. Tabs at the top of the view correspond to different fund managers, allowing users to view the funds they manage separately. For each fund, as depicted in Figure 4, a unique fund code (Figure 4(b)) is displayed in the top left corner to aid identification. An abbreviated timeline (Figure 4(c)) on the right shows the fund's operational history, with the red masked portion indicating the duration that the fund manager has managed this fund. Users can control the time range displayed on the x-axis of the main chart by brushing on the abbreviated timeline. In the main chart (Figure 4(d)), the green line above the x-axis represents the daily net value of the fund, while the yellow ripple area around the line indicates the change in the day's simulated position relative to the previous day, with larger areas indicating more significant position changes. Each glyph (Figure 4(e)) below the x-axis corresponds to a series of disclosures in the quarterly report, with the height of the brown bar representing the fund's size, and each circle attached to the bar corresponding to one of the fund's top 10 stock holdings. The relative position of the connected circles on the bars is proportional to the percentage of stocks to net value. When a stock first becomes a fund's top 10 holding, the corresponding circle is marked yellow; if it reappears on a subsequent reporting date, it is marked green.

### 6.4 Comparison View

The *Comparison View* is designed to enable users to compare the fund performance of one or more fund managers. As illustrated in Figure 1(D), this view comprises a table and a line chart. Users can add funds to the *Comparison View* by selecting them in the *Historical Management View*. The table displays basic information about the selected fund, including the fund code, manager, start management time, and end management time. The line chart shows the daily growth rate of each fund using different colored curves. The user can hover over the funds in the table to highlight the corresponding curves. Since the start time and end time of different funds may not be the same, the time range of the line chart starts from $min(all\ start\ time)$ and ends at $max(all\ end\ time)$ by default, so that all selected funds can be displayed. Users can also control the time range of the line chart by brushing the abbreviated timeline below. Note that if a fund's $[start\ time, end\ time]$ does not overlap with the brushed time range, the fund will not be displayed.

## 7 Evaluation

We present the efficacy of *FMLens* through various means in the context of the visualization field [56]. To begin with, we illustrate two case studies identified by our collaborative experts in subsection 3.1 during their exploration of the system. Furthermore, we conduct a user study with 12 participants to assess their experience with our system.

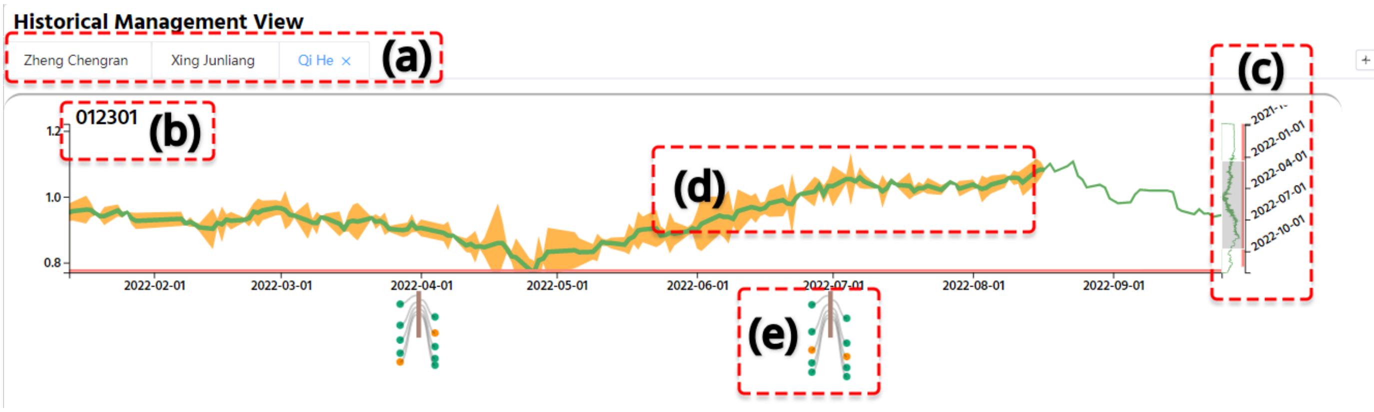



Fig. 4: The *Historical Management View*: (a) The tabs are divided into columns to show the history of different fund managers' management records. (b) The fund code is the unique identification of the fund. (c) The abbreviated timeline shows how long the fund has been in operation, and the red mask indicates the period in which the fund manager was involved in the management. (d) The green line represents the daily net value of the fund, and the yellow ripple area indicates the change in the fund's position on that day compared to the previous day. (e) The glyph corresponds to a series of disclosures in the quarterly report.

## 7.1 Case Study

### *7.1.1 Case I: A Leader in Manufacturing Investment*

The first case study describes the process of using *FMLens* to find a satisfactory fund manager for the financial analyst E6. In the *FM Overview*, E6 first selected two regions with a higher concentration of red points for analysis (Figure 5(1)). In the summary plot, she then focused on those fund managers with a high percentage of manufacturing investments (Figure 5(2)). She explained, "*since manufacturing has the largest market volume of any industry sector, I prefer to invest in it.*" E6 then found that the group of fund managers with the highest percentage of investments in manufacturing on the left-hand side of the region have significantly higher cumulative returns than the right-hand side. "*It looks like the left-hand side is more trustworthy.*" As she did so, E6 brushed a few fund managers in the manufacturing header on the left (Figure 5(3)). After these fund managers were added to the ranking view, E6 chose to view their performance metrics for the most recent year (Figure 5(4)) and adjusted the weights in the control panel (Figure 5(5)) to begin ranking them according to her investment preferences. After the ranking was generated, E6 checked the performance metrics of the top fund managers (Figure 5(6)). "*The results are in line with my expectations, but the performance of the top three looked very close,*" said E6. So she decided to look at the management records of the top three (Figure 5(7)&(8)) to find where the differences lie in the *Historical Management View*. E6 first looked at the third, *Zheng Chengran*, and found that the manager currently manages a large number of funds, but only one (code: 004997) is highly profitable. E6 discovered the size of this fund is much larger than any other fund managed by *Zheng Chengran* during the same period. She commented, "*Zheng Chengran's outstanding performance on metrics was heavily influenced by this fund, masking his poor performance on other funds.*" Moving on to *Xing Junliang* in the second place, E6 noted that this person has only managed five funds to date, with a track record of first management since 2021. E6 also found that *Xing Junliang* has very few position adjustment actions, and said, "*Xing Junliang's investment style is conservative, which is why he is in second place in the rankings. But his management experience is lacking, and we need to watch his future performance further.*" E6 finally looked at the history of the first place *Qi He*, a relatively experienced fund manager with six years of investment experience. Upon further exploration, E6 was surprised to find that whenever *Qi He* took over a new fund, its daily net value would magically rise, and the trend was very similar. To verify the findings, E6 added the four funds managed by *Qi He* to the *Comparison View* (Figure 5(9)). Three of the funds managed by *Qi He* since 2021 have very similar good development curves as well as being close to the early trend of the excellent fund (code: 001856) he once managed. Ultimately, E6 believed that *Qi He*'s comprehensive skills were stronger and held greater investment value.

### *7.1.2 Case II: The Present and Past Life of a Star Manager*

The second case study shows how E4 used *FMLens* to identify the investment style of the fund manager of interest. Starting with the *FM Overview*, E4 selected a few points with the darkest red color (highest return) (Figure 6(1)) and found that manufacturing was favored by most fund managers. Then E4 became interested in those that were not in the mainstream, i.e. those with a low percentage of manufacturing investments (Figure 6(2)). E4 brushed these fund managers to see their overall performance over the past eight years. Surprisingly, *Ge Lan*, a star fund manager rooted in healthcare, appeared (Figure 6(3)). Hearing some negative news about *Ge Lan*'s Waterloo, E4 clicked and recalled *Ge Lan*'s management records. E4 thoroughly browsed the net value of all the funds and found that *Ge Lan* had performed poorly when she was a freshman. E4 found that her investment style was very different from what it is now. *Ge Lan* had bought *Leeco* as her third-heaviest position, then bought *Storm* into her top 10 positions in the third quarter of 2015 and added her heaviest position at the end of the year (Figure 6(4.1)). These stocks were equally sought after by the market capital at the time and valuations were

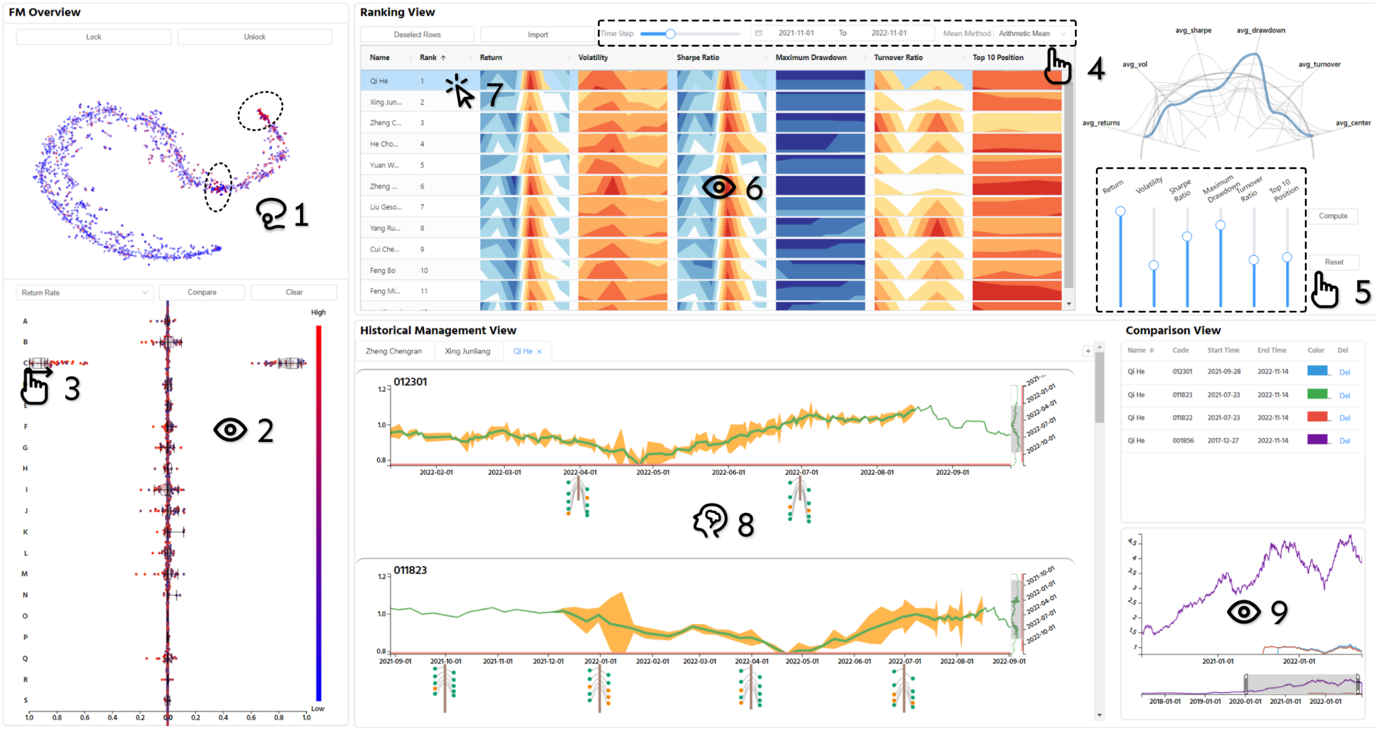



Fig. 5: Case I. (1) Lasso regions with a higher concentration of red points. (2) Observe fund managers with a high percentage of manufacturing investments. (3) Brush some managers for further analysis. (4) Set the time range and interval to calculate metrics. (5) Adjust metric weights for ranking. (6) View the performance metrics of the top rankers. (7) View historical management records. (8) Generate findings by management records. (9) Comparison shows the funds with a similar pattern of development after the "*Qi He*" takeover.

pushed all the way up. However, five years later, *Storm* and *Leeco* were delisted from the exchange. E4 said that it was easy to see that *Ge Lan* liked to invest in hot star companies at the time and enjoyed the gains from valuation expansion. But because of liquidity issues, these highly valued stocks shuddered back down during the year. In addition, E4 found that *Ge Lan* liked to change positions frequently at the beginning of her career (Figure 6(4.1)&(4.2)), i.e., *Ge Lan* did not develop her own investment style. In addition, E4 said that the fund documented all *Ge Lan*'s style transitions in (Figure 6(4.2). The above conclusion about her style was also confirmed in the early part of this chart. Looking at the color variation of the leaves in (Figure 6(4.2), E4 found that *Ge Lan* started to focus on the biopharmaceutical sector, a very good long-term growth track, and developed a "long-term holdings and concentrated positions" style that matches the characteristics of the biopharmaceutical sector. E4 also mentioned speculation about *Ge Lan*'s recent failures. The growth in management scale requires fund managers to diversify their positions due to the limited capacity of a single track. From the tree trunk length, E4 found that *Ge Lan* did not change her strategy quickly in 2021 as the size under management increases, and most of the leaf nodes are green and the top 10 positions implied by the tree's forked position remain high in 2021.

### 7.2 User Study

In order to perform a quantitative evaluation of the efficacy of *FMLens*, a user study was conducted to compare its performance with that of a baseline visualization. The *Wind*[4] financial data terminal, an established system, was selected as the baseline for this study.

**Participants.** A total of 12 participants, with at least two years of experience in fund investment, were recruited for the study. None of them had previously utilized our system. Of these participants, five participants (four male, one female) were recruited from financial institutions where E1, E4-6 work, and the remaining seven participants (five male, two female) were recruited from universities where E2-3, E7-8 are located. The participants' ages were recorded, with a mean of 29.6 years (SD = 2.15) for the first group and a mean of 24.3 years (SD = 2.80) for the second group.

**Tasks and Procedure.** At the start of the study, participants were given a brief explanation of the workflow of *FMLens* and the *Wind* terminal (used as a benchmark). The explanation lasted approximately 20 minutes, during which the performance metrics appearing in *FMLens* were explained in detail. Following this, two links were provided to the participants, allowing them to access the *FMLens* and *Wind* terminal online, respectively. Participants were then given 20 minutes of free exploration time to familiarize themselves with the two systems. Subsequently, participants were given four tasks to complete. **T.1** required them to find a fund manager with a diversified investment sector and a fund manager with a concentrated investment sector (**R.1**). **T.2** required them to select among multiple fund managers with consistently better monthly returns over the period

4. https://www.wind.com.cn/

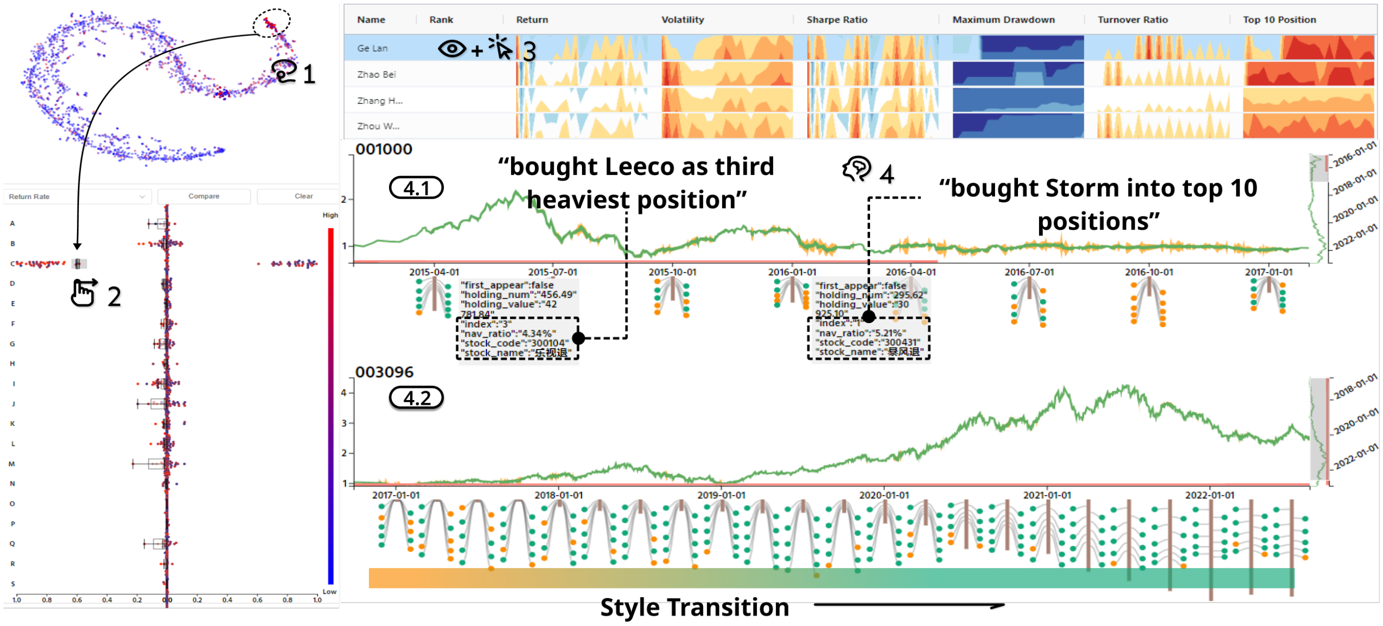



Fig. 6: Case II: (1) Selected the points with high returns. (2) Brushed fund managers with low manufacturing investment ratios. (3) Found a star manager. (4.1) The fund manager's early immature investment style, preferring to pursue star companies such as Leeco and Storm. (4.2) Record the fund manager's style shift.

2021-2022 (**R.2**, **R.3**). **T.3** involved investigating the number of funds managed by the managers selected in **T.2** in the last year and how their positions had changed (**R.4**, **R.5**). Finally, **T.4** required participants to compare multiple fund managers to find differences in their investment styles (**R.5**). Participants were asked to complete these tasks first using *FMLens*, and then again using the *Wind* terminal. Following completion of the tasks using each system, participants were asked to anonymously fill out a 7-point Likert questionnaire, with responses ranging from 1 (strongly disagree) to 7 (strongly agree). Feedback and comments were collected from both questionnaires. Finally, participants were provided with a 15 voucher as a reward for their participation in the study.

**Results.** Our questionnaire aimed to compare the performance of the *FMLens* system with that of a baseline system in terms of system effectiveness, the convenience of visual designs, and system usability. The results of the questionnaire are presented in Figure 7. Overall, the participants expressed that the *Wind* system provides a substantial amount of financial data (Q1), but its interface is not focused on providing a flexible presentation and comparison of fund manager information (Q2–Q5), making the selection process inconvenient. Participants felt that *Wind*'s interface was more complex (Q9) and its design of presenting fund manager performance metrics in spreadsheets made participants feel that it was not easy to capture dynamic information (Q10). Additionally, some participants indicated that they had to spend more time searching and filtering data in *Wind* because it does not present information about fund managers in a targeted manner (Q12-Q13). In contrast, the participants found *FMLens* to be more convenient and efficient, providing a more concise overview of fund managers (Q1) and enabling quick examination of performance metrics for fund managers of interest (Q2). Participants found it easier to adjust the time horizon and interval in *FMLens* to obtain the appropriate performance metrics (Q3), view the distribution of each performance metric (Q4), rank fund managers based on their investment preferences (Q5), track a fund manager's historical assets under management (Q6), learn more about a fund manager's investment style (Q7), and compare the daily returns of funds managed by different fund managers on a daily basis (Q8). Regarding visual design, participants found *FMLens* to be more suitable for novice users (Q9), providing enough information to vet fund managers (Q10), and having a visual design and interactions that helped participants explore and filter fund managers (Q11). Finally, in terms of system usability, participants found *FMLens* easier to find fund managers of interest (Q12), understand fund managers' investment styles and capabilities (Q13), and have workflows that fit their operational logic (Q14).

## 8 Discussion and Limitation

We conducted one-hour semi-structured interviews with E1–E8 to collect their opinions and suggestions about *FMLens*.

**Contributions Over the Previous Work.** The current study introduces several significant contributions compared to prior research. First, *FMLens* is centered on investing in fund managers rather than investing directly in funds, representing a novel idea. As stated by E1, this approach optimizes the selection process of fund managers, which is not adequately supported by existing tools. Second, *FMLens* provides a comprehensive evaluation of fund managers by employing a multidimensional assessment framework. As mentioned by E4, this approach integrates performance metrics and position adjustment simulation to provide insight into fund managers' unobservable actions. Finally, the user interface of *FMLens* is designed to be simpler and more

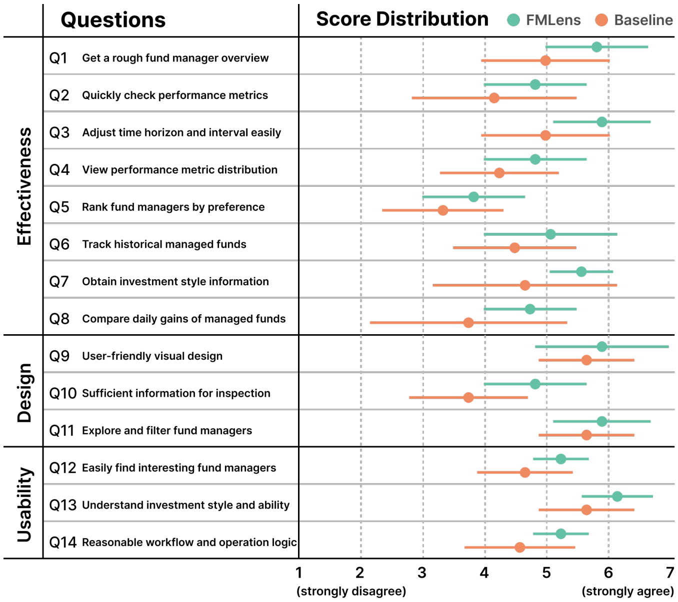



Fig. 7: The questionnaire results of 12 participants in terms of system effectiveness, the convenience of visual design, and system usability.

intuitive compared to conventional financial data analytics platforms such as *Wind* and *Bloomberg*. As noted by E7, this feature makes the system user-friendly and can be utilized quickly by users with investment backgrounds and experiences.

**System Performance.** All experts appreciated that *FMLens* considered multidimensional features to help investors select fund managers. E1 commented that "*the system is effective in specific operations, such as analyzing time-series changes in fund manager performance, which nicely compensates for the lack of existing tools.*" E2 and E3 spoke of the system's ability to assist in tracking shifts in fund managers' investment styles. "*It helps us make sense of the data in the analysis,*" said E3. E8 liked the design of the simulated position adjustment actions. He indicated that it helps to have a deeper understanding of the fund manager's investment philosophy and skills. E7 described our system as a "*good retrospective analysis tool*". She said, "*FMLens does a good job of integrating data about fund managers and presenting it in a novel way.*" In addition, experts mentioned that the system runs smoothly and the interactive logic is well organized.

**Learning Curve and Target Users.** When employing *FMLens*, all experts reached a consensus that the system was user-friendly. Despite the fact that certain visual designs necessitated a certain degree of training, comprehension of the visual encoding enabled swift initiation of exploration. Experts remarked that the workflow was conventional yet innovative, thus permitting the system to facilitate intricate analytical duties without imposing excessive barriers to use. Nonetheless, E1 voiced apprehensions regarding specific performance metrics in the system that novice investors may find challenging to comprehend. E1 emphasized that although familiarity with financial matters could enhance the system's performance, even fledgling users could easily evaluate fund managers based on rudimentary metrics. For instance, users with minimal investment experience could simply scrutinize the metric of "returns" and, therefore, appraise fund managers. The experts suggested that the system's target audience should comprise individuals with some investment experience rather than complete novices. Novices may find it more advantageous to choose funds based on "return" rankings rather than fund managers through extant commercial tools or apps. Nevertheless, E5 also voiced similar reservations regarding the system's metrics and proposed that it encompass further details concerning the metrics, including their calculation methods, significance, and evaluation criteria.

**Generalizablility and Scalability.** During the interviews, we engaged the experts in a discussion of which components of *FMLens* could be applied to other scenarios and which one(s) would require customization. They noted that the system was highly versatile within the financial sector and could be used in any financial product management organization, provided that the data were preprocessed. E2 also remarked that there was research value in substituting fund companies for fund managers in the system. Moreover, relevant financial data, such as stock data, could be easily integrated into *FMLens*. With regard to scalability, *FMLens* presents multidimensional data on fund managers interactively. Since there are currently no more than $4,000$ fund managers and no more than $10,000$ funds in the entire market, and our workflow follows the visualization mantra of "overview first, zoom and filter, then details-on-demand" [57], there are no significant scalability issues for the front end. The bottleneck lies in the real-time calculation of performance metrics in the back end. When the user selects many fund managers at once or sets a small time interval, the increase in the number of calculations can cause slower response times.

**Limitation.** This work has several limitations. First, market rules impose a time lag in updating fund information. Although the system utilizes much of the data disclosed in quarterly reports, the data changes on a daily basis, and live access to this data is not available. Second, there is a dearth of analysis on factors that influence performance. Fund managers' performance is typically impacted by several factors, including social factors (market prosperity, public opinion, social events) and personal factors (education, age, gender). Relevant attribution analysis is not yet supported by *FMLens*. Third, to enhance the effectiveness of clustering after dimensionality reduction, more attributes such as education level and years of employment can be considered for summarizing the candidate space.

## 9 Conclusion and Future Work

This study proposes a visual analytics approach for facilitating the fund manager selection process and enabling investors to evaluate candidate fund managers effectively. With *FMLens*, investors can compare investment styles, capabilities, and performance metrics of different fund managers and dynamically rank them according to their preferences. The efficacy of *FMLens* is demonstrated through two case studies and a qualitative user study. To further enhance the analysis, future work includes integrating additional fund manager characteristics, such as age, education, and the number of awards.

## Acknowledgments

We would like to thank our industry collaborator for offering valuable resources. We also thank our domain experts and the anonymous reviewers for their insightful comments. This work is partially supported by the Shanghai Frontiers Science Center of Human-centered Artificial Intelligence (ShangHAI) and Key Laboratory of Intelligent Perception and Human-Machine Collaboration (ShanghaiTech University), Ministry of Education.

## References

[1] R. C. Jones and R. Wermers, "Active management in mostly efficient markets," *Financial Analysts Journal*, vol. 67, no. 6, pp. 29–45, 2011.

[2] K. P. Baks, *On the performance of mutual fund managers*. University of Pennsylvania, 2002.

[3] J. B. Berk and J. H. Van Binsbergen, "Measuring skill in the mutual fund industry," *Journal of financial economics*, vol. 118, no. 1, pp. 1–20, 2015.

[4] A. A. Gottesman and M. R. Morey, "Manager education and mutual fund performance," *Journal of empirical finance*, vol. 13, no. 2, pp. 145–182, 2006.

[5] M. Grinblatt and S. Titman, "Performance measurement without benchmarks: An examination of mutual fund returns," *Journal of business*, pp. 47–68, 1993.

[6] W. F. Sharpe, "Asset allocation: Management style and performance measurement," *Journal of portfolio Management*, vol. 18, no. 2, pp. 7–19, 1992.

[7] J. A. Christopherson, W. E. Ferson, and D. A. Glassman, "Conditioning manager alphas on economic information: Another look at the persistence of performance," *The Review of Financial Studies*, vol. 11, no. 1, pp. 111–142, 1998.

[8] J. Chevalier and G. Ellison, "Are some mutual fund managers better than others? cross-sectional patterns in behavior and performance," *The journal of finance*, vol. 54, no. 3, pp. 875–899, 1999.

[9] E. F. Fama and K. R. French, "Luck versus skill in the cross-section of mutual fund returns," *The journal of finance*, vol. 65, no. 5, pp. 1915–1947, 2010.

[10] M. Kacperczyk, C. Sialm, and L. Zheng, "Unobserved actions of mutual funds," *The Review of Financial Studies*, vol. 21, no. 6, pp. 2379–2416, 2008.

[11] W. F. Sharpe, "Mutual fund performance," *The Journal of business*, vol. 39, no. 1, pp. 119–138, 1966.

[12] M. C. Jensen, "The performance of mutual funds in the period 1945-1964," *The Journal of finance*, vol. 23, no. 2, pp. 389–416, 1968.

[13] F. A. Sortino and L. N. Price, "Performance measurement in a downside risk framework," *the Journal of Investing*, vol. 3, no. 3, pp. 59–64, 1994.

[14] A. Niessen-Ruenzi and S. Ruenzi, "Sex matters: Gender bias in the mutual fund industry," *Management Science*, vol. 65, no. 7, pp. 3001–3025, 2019.

[15] K. M. Cremers and A. Petajisto, "How active is your fund manager? a new measure that predicts performance," *The review of financial studies*, vol. 22, no. 9, pp. 3329–3365, 2009.

[16] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto, "Wirevis: Visualization of categorical, time-varying data from financial transactions," in *2007 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 2007, pp. 155–162.

[17] A. Samak, H. C. Lam, B. D. Fisher, and D. S. Ebert, "An experimental study of financial portfolio selection with visual analytics for decision support," *2011 44th Hawaii International Conference on System Sciences*, pp. 1–10, 2011.

[18] S. Ko, I. Cho, S. Afzal, C. Yau, J. Chae, A. Malik, K. Beck, Y. Jang, W. Ribarsky, and D. S. Ebert, "A survey on visual analysis approaches for financial data," in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 599–617.

[19] E. V. F. Lapura, J. K. J. Fernandez, M. J. K. Pagatpat, and D. D. Dinawanao, "Development of a university financial data warehouse and its visualization tool," *Procedia Computer Science*, vol. 135, pp. 587–595, 2018.

[20] A. Arleo, C. Tsigkanos, C. Jia, R. A. Leite, I. Murturi, M. Klaffenböck, S. Dustdar, M. Wimmer, S. Miksch, and J. Sorger, "Sabrina: Modeling and visualization of financial data over time with incremental domain knowledge," in *2019 IEEE Visualization Conference (VIS)*. IEEE, 2019, pp. 51–55.

[21] M. Schaefer, F. Wanner, R. Kahl, L. Zhang, T. Schreck, and D. Keim, "A novel explorative visualization tool for financial time series data analysis," in *VAW2 011: The Third International UKVAC Workshop on Visual Analytics*, 2011.

[22] H. Ziegler, M. Jenny, T. Gruse, and D. A. Keim, "Visual market sector analysis for financial time series data," in *2010 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 2010, pp. 83–90.

[23] E. Sorenson and R. Brath, "Financial visualization case study: Correlating financial timeseries and discrete events to support investment decisions," in *2013 17th International Conference on Information Visualisation*. IEEE, 2013, pp. 232–238.

[24] W.-A. Jungmeister and D. Turo, "Adapting treemaps to stock portfolio visualization," Tech. Rep., 1992.

[25] C. Csallner, M. Handte, O. Lehmann, and J. Stasko, "Fundexplorer: Supporting the diversification of mutual fund portfolios using context treemaps," in *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No. 03TH8714)*. IEEE, 2003, pp. 203–208.

[26] X. Yue, J. Bai, Q. Liu, Y. Tang, A. Puri, K. Li, and H. Qu, "sportfolio: Stratified visual analysis of stock portfolios," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 1, pp. 601–610, 2019.

[27] X. Yue, Q. Gu, D. Wang, H. Qu, and Y. Wang, "iquant: Interactive quantitative investment using sparse regression factors," *arXiv preprint arXiv:2104.11485*, 2021.

[28] H. Guo, M. Liu, B. Yang, Y. Sun, H. Qu, and L. Shi, "Rankfirst: Visual analysis for factor investment by ranking stock timeseries," *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[29] T. Dwyer, "A scalable method for visualising changes in portfolio data," in *Proceedings of the Asia-Pacific symposium on Information visualisation-Volume 24*, 2003, pp. 17–25.

[30] J. Li, H. Izakian, W. Pedrycz, and I. Jamal, "Clustering-based anomaly detection in multivariate time series data," *Applied Soft Computing*, vol. 100, p. 106919, 2021.

[31] A. Abdella, J. K. Brecht, and I. Uysal, "Statistical and temporal analysis of a novel multivariate time series data for food engineering," *Journal of Food Engineering*, vol. 298, p. 110477, 2021.

[32] B. D. Nguyen, R. Hewett, and T. Dang, "Netscatter: Visual analytics of multivariate time series with a hybrid of dynamic and static variable relationships," in *2021 IEEE 14th Pacific Visualization Symposium (PacificVis)*. IEEE, 2021, pp. 52–60.

[33] S. T. Lei and K. Zhang, "Visual signatures for financial time series," in *Proceedings of the 2011 Visual Information Communication-International Symposium*, 2011, pp. 1–10.

[34] H. Hochheiser and B. Shneiderman, "Dynamic query tools for time series data sets: timebox widgets for interactive exploration," *Information Visualization*, vol. 3, no. 1, pp. 1–18, 2004.

[35] Y. Wu, N. Pitipornvivat, J. Zhao, S. Yang, G. Huang, and H. Qu, "egoslider: Visual analysis of egocentric network evolution," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 1, pp. 260–269, 2015.

[36] R. D. Peng, "A method for visualizing multivariate time series data," 2008.

[37] Y. Chen, Q. Chen, M. Zhao, S. Boyer, K. Veeramachaneni, and H. Qu, "Dropoutseer: Visualizing learning patterns in massive open online courses for dropout reasoning and prediction," in *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2016, pp. 111–120.

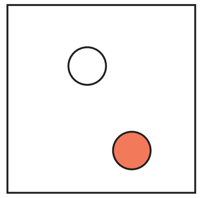
[38] Q. Li, X. Wei, H. Lin, Y. Liu, T. Chen, and X. Ma, "Inspecting the running process of horizontal federated learning via visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[39] D. J. Hruschka, D. Schwartz, D. C. St. John, E. Picone-Decaro, R. A. Jenkins, and J. W. Carey, "Reliability in coding open-ended data: Lessons learned from hiv behavioral research," *Field methods*, vol. 16, no. 3, pp. 307–331, 2004.

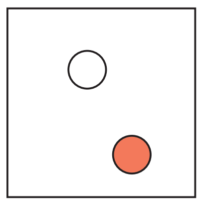
[40] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *The journal of finance*, vol. 19, no. 3, pp. 425–442, 1964.

[41] H. Abdi, "Z-scores," *Encyclopedia of measurement and statistics*, vol. 3, pp. 1055–1058, 2007.

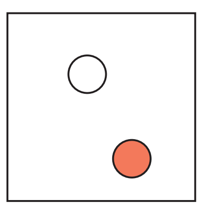
[42] R. D. Harris and M. Mazibas, "Dynamic hedge fund portfolio construction: A semi-parametric approach," *Journal of Banking & Finance*, vol. 37, no. 1, pp. 139–149, 2013.
[43] B. Murthi, Y. K. Choi, and P. Desai, "Efficiency of mutual funds and portfolio performance measurement: A non-parametric approach," *European Journal of Operational Research*, vol. 98, no. 2, pp. 408–418, 1997.
[44] M. M. L. de Prado, *Machine learning for asset managers*. Cambridge University Press, 2020.
[45] J. Huang, J. Chai, and S. Cho, "Deep learning in finance and banking: A literature review and classification," *Frontiers of Business Research in China*, vol. 14, no. 1, pp. 1–24, 2020.
[46] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios," *Applied Stochastic Models in Business and Industry*, vol. 33, no. 1, pp. 3–12, 2017.
[47] M. Arouri, H. Ben Ameur, N. Jawadi, F. Jawadi, and W. Louhichi, "Are islamic finance innovations enough for investors to escape from a financial downturn? further evidence from portfolio simulations," *Applied Economics*, vol. 45, no. 24, pp. 3412–3420, 2013.
[48] R. Fernholz, R. Garvy, and J. Hannon, "Diversity-weighted indexing," *Journal of Portfolio Management*, vol. 24, no. 2, p. 74, 1998.
[49] R. Liu, J. Kuang, Q. Gong, and X. Hou, "Principal component regression analysis with spss," *Computer methods and programs in biomedicine*, vol. 71, no. 2, pp. 141–147, 2003.
[50] G. C. McDonald, "Ridge regression," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 1, pp. 93–100, 2009.
[51] J. Ranstam and J. Cook, "Lasso regression," *Journal of British Surgery*, vol. 105, no. 10, pp. 1348–1348, 2018.
[52] F. Heimerl and M. Gleicher, "Interactive analysis of word vector embeddings," in *Computer Graphics Forum*, vol. 37, no. 3. Wiley Online Library, 2018, pp. 253–265.
[53] P. E. Rauber, S. G. Fadel, A. X. Falcao, and A. C. Telea, "Visualizing the hidden activity of artificial neural networks," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 101–110, 2016.
[54] Q. Li, K. S. Njotoprawiro, H. Haleem, Q. Chen, C. Yi, and X. Ma, "Embeddingvis: A visual analytics approach to comparative network embedding inspection," in *2018 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2018, pp. 48–59.
[55] H. Reijner *et al.*, "The development of the horizon graph."
[56] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller, "A systematic review on the practice of evaluating visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2818–2827, 2013.
[57] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations - sciencedirect," *The Craft of Information Visualization*, pp. 364–371, 2003.



**Chenyang Zhang** is currently a Master's student in the Department of Computer Science at the University of Illinois Urbana-Champaign (UIUC). He received his Bachelor's degree from the School of Information Science and Technology (SIST) at ShanghaiTech University. Chenyang's research interests include Human-Computer Interaction, Creative Process, Data Visualization, Interaction Design, and Education. He has published his research results in top-tier conferences and journals, such as VIS and TMI. Chenyang is working hard to innovate and contribute to the intersection fields of Computer Science, Design, and Humanities. For more information about his work and achievements, please visit his personal website: https://chenyang.me.



**He Wang** is currently pursuing his M.S. degree at the School of Information Science and Technology, ShanghaiTech University. He received his B.S. degree from the School of Computer Science and Information Engineering at Hefei University of Technology. His research interests include data visualization, data-centric AI, and eXplainable Artificial Intelligence (XAI). His research goal is to promote human-AI collaboration by developing scalable and interpretable visual analytics tools.



**Quan Li** is currently a tenure-track assistant professor at School of Information Science and Technology, ShanghaiTech University. He received his Ph.D. from the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology. His research interests lie in artificial intelligence, big data visualization, and visual analysis, interpretable machine learning, data science, data storytelling, human-computer interaction (HCI), and computational social sciences. For more details, please refer to https://faculty.sist.shanghaitech.edu.cn/liquan/.