RankFIRST: Visual Analysis for Factor Investment by Ranking Stock Timeseries

Huijie Guo, Meijun Liu, Bowen Yang, Ye Sun, Huamin Qu, Lei Shi

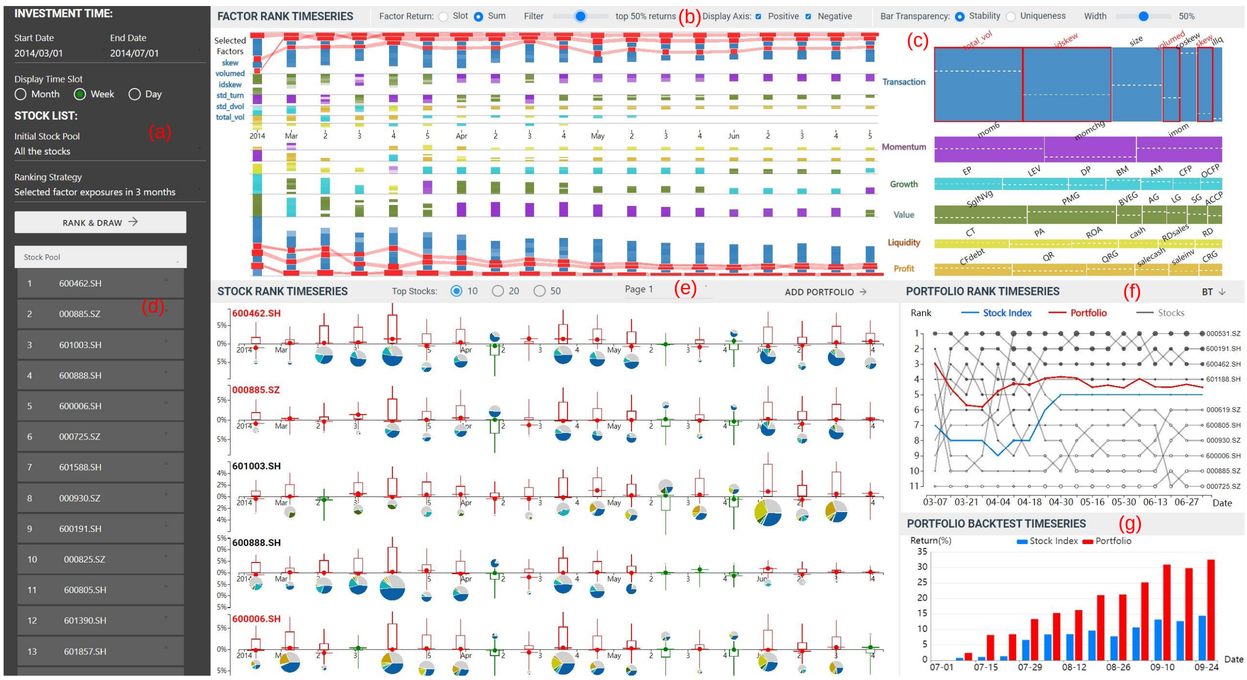

Fig. 1. The visualization interface of RankFIRST: (a) configuration panel for investment time, stock pool, and ranking strategy; (b) factor view showing the factor return time series for selection; (c) factor treemap of aggregated factor returns; (d) ranked stock list; (e) stock view showing stock time series by the firework chart for portfolio construction; (f) portfolio view depicting its return time series in the past (red curve); (g) backtest view on the actual portfolio performance in the next 3 months (red bars).

Abstract—In the era of quantitative investment, factor-based investing models are widely adopted in the construction of stock portfolios. These models explain the performance of individual stocks by a set of financial factors, e.g., market beta and company size. In industry, open investment platforms allow the online building of factor-based models, yet set a high bar on the engineering expertise of end-users. State-of-the-art visualization systems integrate the whole factor investing pipeline, but do not directly address domain users' core requests on ranking factors and stocks for portfolio construction. The current model lacks explainability, which downgrades its credibility with stock investors. To fill the gap in modeling, ranking, and visualizing stock time series for factor investment, we designed and implemented a visual analytics system, namely RankFIRST. The system offers built-in support for an established factor collection and a cross-sectional regression model viable for human interpretation. A hierarchical slope graph design is introduced according to the desired characteristics of good factors for stock investment. A novel firework chart is also invented extending the well-known candlestick chart for stock time series. We evaluated the system on the full-scale Chinese stock market data in the recent 30 years. Case studies and controlled user evaluation demonstrate the superiority of our system on factor investing, in comparison to both passive investing on stock indices and existing stock market visual analytics tools.

Index Terms—Stock market, Factor investing, Visual analysis.

1 Introduction

Quantitative investments that rely on computer-based models for trading decisions have become increasingly important in the last decade. In the American stock market, quantitative funds accounted for 36% of trading volume by institutions in 2019, up from 18% in 2010 [6]. Many leading methods in quantitative investment apply the factor investing model, which explains the risk and return of a stock by a set of factors such as the sensitivity to market risk [18], the corresponding company size, and the ratio of current stock price to its asset value [17] [19]. The model can construct an optimal stock portfolio to maximize portfolio return and control the associated risk. According to a survey by BlackRock [29], more than 50% of institutional investors used factors in their investment process, risk management, and trading strategy.

In factor investing, quantitative analysts keep a pool of useful factors which are regularly updated in every trading cycle (weeks to months). At the beginning of each trading cycle, the analysts select a set of factors

- *Huijie Guo, Meijun Liu, Bowen Yang, Ye Sun, Lei Shi are with SKLSDE and Beijing Advanced Innovation Center for Big Data and Brain Computing, School of Computer Science and Engineering, Beihang University, China. E-mail: {guo_hj, bowenyang, liumeijun, sunie,leishi}@buaa.edu.cn.*
- *Huamin Qu is with the Hong Kong University of Science and Technology. E-mail: huamin@cse.usk.hk.*
- *Lei Shi is the corresponding author.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

from the current pool to explain the recent move of stock market, and predict trends for the near future. The factor selection is traditionally based on the desired characteristics of "good" factors. Next, individual stocks are evaluated based on their risk degrees and return expectations inferred from the selected factors. A set of stocks are then selected to construct the investment portfolio, which is will be the trading target in the next cycle. Though factor/stock selection criteria vary across investors, the factor investing pipeline remains largely unchanged.

Interactive visual analytics systems are in high demand for factor investing. Analysts are now overwhelmed by the vast amount of stock market data. They find it hard to manually evaluate numerous factors and stocks one by one, even with the assistance of computer algorithms. Integrated analytics systems synthesizing the whole set of stock data, relevant computational models, and interactive user interfaces have become extremely valuable for efficient factor investing. In the industry, there have been quite a few online quantitative investment platforms (QuantConnect [2], RiceQuant [3], etc.), which bring together stock market data and cloud-based computing resources. Yet, these platforms are designed as open programming interfaces that are not favorable to ordinary analysts with limited computer science expertise. In the literature, several interactive data analysis systems have been built for stock market investment [13] [34] [35]. Their primary goals are visualization, comparison, and evaluation of existing investment portfolios [52], but not for the next decision-making via factor investing.

Recently, the iQUANT system proposed by Yue et al. [53] assembles the full pipeline of factor investing into the same visual analytics platform. Our work follows this line of research but focuses on completely different requirements raised by domain users on factor investing. First, the regularization model (i.e., sparse regression, the lasso [46]) adopted in iQUANT outputs only a small subset of high potential factors for selection. The approach prohibits domain users from selecting their most favorable factors under investment constraints. The Lasso model is also inefficient for interactive stock investment as a result is prone to change with model parameters. Second, the sophisticated Lasso model computes the importance of factors for the visual analysis process. The importance of factors is an abstract concept hardly explained by domain analysts, making the approach relatively untrusted. Last but most importantly, based on the expert study result, domain users' core requests for visual analytics for factor investing can be summarized as "ranking stock-related time series". While iQUANT faithfully displays the output of its factor models, the visualizations are not designed to highlight factor and stock ranks. With dozens of lengthy time series piled up in the same view, visual ranking by the naked eye is the key challenge in interactive factor investing.

In this work, we propose an end-to-end visual analytics system, namely RankFIRST (Factor Investment by Ranking Stock Timeseries). From the perspective of time series analysis, we adopt a model that computes the quantitative contribution of each factor to the overall stock return, so that ordinary users can rank and apply every preferable factor in their investment. In more detail, a Nobel prize-winning asset price model is introduced [20], which has been established for decades and is widely trusted in the stock market industry. From the visualization perspective, similar to iQUANT, our system supports the display of factor and stock performance for portfolio construction. Nevertheless, all the designs are customized to highlight the rank of factor and stock time series, such that the analytical capability of underlying factor investing models can be fully leveraged. The design and implementation of RankFIRST are nontrivial, in which we make the following contribution.

- We investigate a real-life collection of 6 classes of 56 factors for the Chinese stock market (Fig. 1(c)). The main challenge lies in selecting an appropriate factor model from numerous candidates in the literature. The cross-sectional regression model [20] is finally applied because of its high interpretability via visualization and full coverage of all factors. Several performance measures for good factors are characterized and displayed based on the study with factor investing experts.
- We introduce novel designs for the time series visualization of factors and stock performance. The challenges come from the visual complexity caused by the big financial data (3k stocks$\times$7k days$\times$56 factors) and the design requirement to visually rank tens of lengthy time series. For the visualization of factors, a hierarchical slope graph design is proposed to accommodate all factor returns in a single view while still highlighting important performance measures for factors (Fig. 1(b)). For the visualization of stock returns, a new firework chart design is invented by extending the well-known candlestick chart (Fig. 1(e)).
- We build an interactive factor investing framework with RankFIRST. The framework aligns well with the best practice of domain experts by supporting ranking-based visual factor selection, factor-based stock ranking (Fig. 1(d)), interactive stock picking, portfolio management and evaluation (Fig. 1(f)(g)). Case and user studies demonstrate the usability and advantages of our system over existing stock market analysis and factor investing tools.

2 Related Work

2.1 Financial Market Visualization

Visualizing financial market data such as stock/fund prices and economic indicators have been studied for decades. Techniques were proposed to help users better understand financial information in the history, explore these data to uncover useful patterns, and detect fraudulent behaviors that are dangerous to financial institutions. Recent surveys provide more details of existing work [27] [16] [43].

The financial market data is mostly time series. Hence temporal data visualization techniques predominate in the literature. Chang et al. presented WireVis [11, 12], a visual analytics system that uses multiple coordinated views to illustrate the complex and time-varying behavior of wire transactions in large US banks. Ziegler et al. designed color-coded bars to visualize relative percentage changes of asset prices over time [54]. Bloomberg Inc. introduced a scalable visualization that combines continuous line charts with icon-based discrete event representations to inform users of the important context in the financial time series data [42]. Financial time series visualizations are often combined with automatic analysis methods. For example, clustering algorithms are proposed to aggregate similarly-shaped time series together [36, 37, 41, 48, 54]. The clusters are displayed by self-organizing map [36,37,41], hierarchical view [24], or heat maps [54]. Beyond standard charts, novel designs are also developed to cope with the complexity of modern financial data, e.g., the Growth Matrix [26] [55], wedge charts [14], animated transition [44], and 3D visualizations [31, 45].

On quantitative investment, known methods in the literature mostly focus on the visualization and comparison of existing investment portfolio [13] [34] [35] [52]. FundExplorer [13] introduced a distorted treemap to show the composition of the mutual fund's portfolio. Investors could visually query complementary mutual funds to achieve portfolio diversification. The FinVis tool [34] helps non-experts to understand the expected returns, aggregated risk, and changing context of investment portfolios. PortfolioCompare [35] applied scatterplots and distribution charts to display the variability and correlation of the portfolio's risk and return measures. Personal investment decisions can be aligned with the user's risk preference. sPortfolio [52] explores the characteristics of an investment portfolio from the perspective of style and industry factors. It helps to compare different portfolios and examine the management detail of each individual portfolio.

The above portfolio visualization techniques are not specially designed for portfolio construction by factor investing methods. Recently, the iQUANT tool [53] was proposed, which implemented the standard pipeline of factor investing with visual analytics techniques. They focused on factor and stock selection for portfolio construction. As mentioned earlier, our work improves from the iQUANT proposal by introducing explainable factor computation models, well-motivated (ranking stock-related time series) new visualization designs, and the support of human flexibility in portfolio construction.

2.2 Ranking Data Visualization

Ranking data is ubiquitous to visualization solutions, such as visual sports analysis and interactive recommendation. These techniques can be classified into two categories: universal designs for generic ranking data and specific visualizations for a certain application or data genre.

In the generic category, ranking data is generally considered as a sequence or time series of items associated with multiple attributes (ranks, values, etc.). Hence, classical visualization techniques for multivariate/temporal data can be applied, including spreadsheets, Table Lens [33], stacked bar charts, and parallel coordinates. Especially, a well-known method called bump charts (slope graphs [47]) is proposed to highlight the ranking information, which replaces Y axis of the original line chart with the rank attribute. On advanced ranking visualization techniques, LineUp [22] focused displaying synthetic ranking from a combination of attributes through a nice orchestration of slope graphs and bar charts. It also supports the online adjustment of attribute weights and parameters to track, explain, and compare ranking changes. Also, on multi-attribute ranking, Podium [49] chose an opposite approach that allows users to specify a preferred ranking of data items. Through a machine learning model, Podium supports the visual analysis of attribute contribution to the ranking.

On temporal ranking data, RankExplorer [39] aggregated a large data collection into several ranking categories for visualization. The original ThemeRiver design and trend charts are modified to show the evolution of aggregated ranking categories over time. RankBrushers [23] introduced parallel histograms to characterize the evolution, distribution, and uncertainty of a large collection of temporal ranking data. On multidimensional data, Seo and Shneiderman [38] implemented the Hierarchical Clustering Explorer according to their rank-by-feature framework. The technique enhances the conventional statistical charts with additional ranking support. Behrisch et al. [8] studied the visual comparison of a set of ranking results. They proposed a radial node-link graph design to display the difference between rankings. Mylavarapu et al. [30] conducted a user study to compare 6 visualization designs for ranked list data. On 3 tasks involving ranking, comparison, and average, their results suggested that wrapped bars offered the best trade-off between accuracy and interaction cost.

In the application category, RankViz [32] is designed to visually interpret learning to rank (LtR) algorithms and results. Customized glyph designs were proposed to illustrate the difference between ranked lists and the influence of features/rules on ranking. uRank [15] is a visual analytics tool for the adaptive recommendation. The interface mixes stacked bars and tag clouds to display both document ranking and content. SRVis [50] proposed an optimized visual analysis framework to rank a large set of spatial items. Both matrix visualizations and stacked bar charts are refined and integrated into their design. Xia et al. [51] designed a calendar heat map visualization with sparklines overlays to illustrate the ranked time series of Wikipedia top page views. TrajRank [28] combines conventional spatial map visualization and sankey diagrams to analyze the ranking, distribution, and variation among trajectories passing the same route. Compared with our work, no existing literature has studied the user requirement and customized visualization design for ranking stock-related time series in the context of factor investing.

3 Factor Investing Problem

3.1 Background and Expert Study

In the quantitative investment of the stock market, factor investing is a leading method, which explains the performance of individual stocks by a set of characteristics shared among a class of stocks, known as *factors*. For example, in the first factor investing model called CAPM [18], the non-risk-free return of a stock is shown to change linearly with the stock's market beta (the single factor), which quantifies its sensitivity to the market risk. In all factor models, the stock return earned from each factor comes at a substantial cost, known as the *factor risk exposure*. The stock return rewarded by the risk exposure is thus called *factor risk premium* to optimize the investment. In the stock market, quantitative analysts manage an appropriate allocation of stocks, known as *portfolio*. The ultimate goal is to increase overall return, reduce investment volatility, and diversify factor risk exposures.

To learn the best practice of factor investing, we conducted longitudinal studies with two domain experts. These experts worked at public funds for more than five years. In the last two years, we have arranged bimonthly discussions with them to understand their factor investing workflow. Routinely, quantitative analysts maintain a large pool of candidate factors, mostly from known factor models in the domain. The exposure value of these factors for individual stocks is updated regularly. At the beginning of each trading cycle (e.g., a month) for factor investing, analysts first select a few factors from their pool as the anchor of the following investment. Later, we will elaborate on the factor selection process (Sect. 4.2). Secondly, the exposure values of the selected factors are aggregated by a scoring function to rank individual stocks. Highly ranked stocks are picked and mixed in appropriate weights to form the portfolio for investment. The factor-based stock ranking and portfolio weighting algorithm is also known as the factor investment strategy (Sect. 4.3, Sect. 4.4). Finally, after the decision of portfolio construction/adjustment is made, relevant stocks need to be purchased or sold in the best market time. Our work focuses on factor investing in the first and second steps, and avoids sophisticated market timing. All stock transactions are assumed to be executed at the average price of the first day within each trading cycle.

3.2 Requirement Analysis

This work targets to popularize factor investing practice to broader financial users without factor modeling or engineering background. Two groups of users were interviewed to collect requirements in adopting factor investing. The first group was quantitative investment *practitioners* from private funds, who applied other types of stock investment methods such as mean-variance analysis and temporal analysis rather than factor investing models. The other group was independent *investors* (the same group invited for user study in Sect. 6). Because of the barrier to implementing fully functional systems, few of these ordinary investors had ever adopted quantitative investment in the stock market. After an initial prototype of RankFIRST was released, we presented the system to the two groups of users as a pilot study to collect their feedback. Key user requirements are summarized as below.

R1. Rank factors by their explainable performance and select top factors. Ordinary users have little knowledge on how to compare financial factors out of computational models. It is important to provide an initial rank of these factors according to the expert view of factor investing. Moreover, the factor ranking criterion should be better based on well-known performance metrics that are explainable to domain users. The factor investing method gains the trust of these users through explainability. The selection of top-ranked factors works as the first step of factor investing and is key to investment success.

R2. Rank stocks with the selected factors, understand stock performance over time, and optimally manage portfolios. In factor investing, users also need a ranking of stocks given the selected set of factors. As stock investment is sophisticated with many subjective constraints, an automated stock ranking algorithm may not be sufficient. An explanation of the stock ranking by factors is necessary to allow the incorporation of human knowledge and intervention. Users can then customize their portfolios based on both algorithmic ranking and manual tuning to achieve optimal management. It is also useful to include stock performance measures that are irrelevant to factors.

R3. Evaluate investment portfolio and understand portfolio performance. Given the portfolio, domain users demand an evaluation of its performance. The positive outcome of the portfolio will indicate the success of factor selection and stock picking. Note that according to the industry standard, the portfolio's performance should not be measured only by the overall historical return. More importantly, the evaluation should also consider the possible returns in the future. The backtest strategy for stock investment commonly achieves it.

R4. Semi-automate factor investing process with the support of user input. Ordinary investors do not have the resource to conduct factor investing, as many processes are quite expensive, e.g., factor modeling and portfolio evaluation. A system with built-in computa-

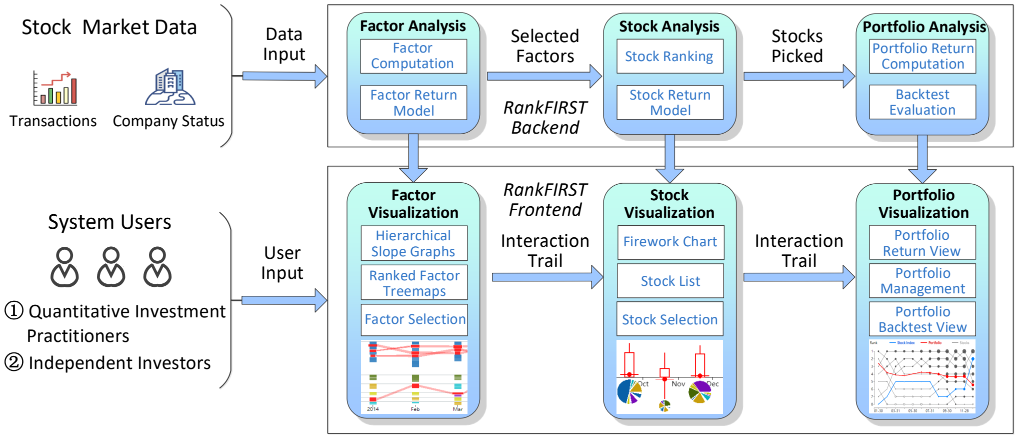

Fig. 2. RankFIRST system pipeline composed of analysis backend and visualization frontend on factors, stocks, and portfolios.

tional components is required to popularize factor investing. In addition, the system should also support incorporating user input, including experience and preference on good factors, stocks, and portfolios. The joint force of computational factor models and human investment knowledge enjoys the best of both worlds for factor investing.

3.3 Visual Analysis Tasks and Challenges

We propose to apply visual analytics techniques to satisfy these user requirements. The RankFIRST system is designed to achieves the following tasks.

T1. Visually analyze the rank of factor return time series that are multivariate on each time slot (R1, R4). The factor performance is inherently time-varying, with the most important feature being its rank of return within all the factors. In addition, there are other key factor measures over the static return, including the dynamic return volatility over time and the interaction with other factors. The number of factors in the analysis could be hundreds (56 in the model applied here), which brings challenges with high visual complexity.

T2. Visually analyze the rank of stock time series from the new perspective of factor investing models (R2, R4). Most existing time series visualizations of individual stocks focus on the dynamic of stock prices (returns), e.g., the standard line charts and candlestick charts. In comparison, the RankFIRST design should focus on the stock performance modeled by factor investing, working as a visual explanation of the model. Visualizing stock return volatility and risk diversification will help recommend an appropriate stock ranking to construct the portfolio.

T3. Visualization of portfolio time series in the history and in the future for backtest evaluation (R3, R4). The performance of portfolios should be illustrated in contrast with baseline portfolios such as stock indices. What-if analysis in portfolio construction will also be helpful in optimally managing the investment.

4 RankFIRST Techniques

The system pipeline of RankFIRST is illustrated in Fig. 2. It takes stock market data as input (Sect. 4.1) and is composed of analytics and visualization components on factors (Sect. 4.2), stocks (Sect. 4.3), and portfolios (Sect. 4.4). User inputs are also incorporated in the visual analytics process. The system as a whole fulfills the T4 task.

4.1 Data and Factor Collections

Throughout this work, we showcase the use of RankFIRST in the China A-Shares stock market. Two data sets are collected as the system input: transaction data of all the 3,000 stocks in China A-shares from 1990 to 2018, including daily opening/closing prices, trading volume, etc.; market information of these companies, including but not limited to financial statements, market news, and financial analysis reports.

From the raw data, 56 factors recommended in a latest white paper for the Chinese stock market [5] are computed for each stock, which can be categorized into 6 classes: *Transaction factors* [17,21], *Momentum factors* [10], *Value factors* [17], *Growth factors*, *Profitability factors* [19], *Liquidity factors*. The detailed definitions of these 56 factors are given in Appendix B.

4.2 Visual Analysis of Factors

4.2.1 Factor Investing Model

Factor investing models explain the return of an individual stock by the risk exposure of the stock on a set of financial factors. Here stock return indicates the money earned by investing the stock, and is denoted by $\boldsymbol{r}$. The factor risk exposure of a stock indicates its risk degree on that factor relative to the market average, and is denoted by $\boldsymbol{x}$. In this work, we adopt the cross-sectional regression model originally proposed by Fama and MacBeth [20]. The cross-sectional model has long been proved effective in the stock industry and has good explainability even for ordinary users. Many large quantitative investment platforms such as RiceQuant [3] and Uqer [4] also adopt this fundamental model. Mathematically, the model is described as below.

$$r_i^{(t)} = \sum_{j=1}^{F} x_{ij}^{(t)} \cdot f_j^{(t)} + b^{(t)} + \varepsilon_i^{(t)}, \qquad \forall i \in [1, N] \tag{1}$$

where i, j, t denote the i-th stock (N stocks in total), the j-th factor (F factors in total), and the t-th trading day, respectively. r is the return of the stock, x is the risk exposure, f is the return of the factor, b is the bias and averaged to zero in the long term, ε_i is the residual of the model on the i-th stock known as the idiosyncratic return of the stock. The risk exposure of all stocks on the same factor are standardized to a mean of zero and a standard deviation of 1. The regression model is solved on each trading day by ordinary least square methods.

From pilot studies with domain experts, we summarize several metrics for good factors over this time series data. These metrics should be highlighted in the visual analytics system.

- **Rank** of absolute factor returns. The selection of highly ranked factors, either positive or negative, can bring large overall return to corresponding stocks. The polarity of stock returns can be controlled by picking the factor exposure polarity on stocks;
- **Volatility** of factor returns. The investment risk can be reduced by selecting factors with low instability over time;
- **Versatility** of factors (classes). Selecting more factors (classes) can help to diversify risk by increasing investment variety;
- **Uniqueness** of factors indicating their low-degree interaction with other factors. This is measured by removing the factor from Eq. (1) and calculating the increase of regression residual.

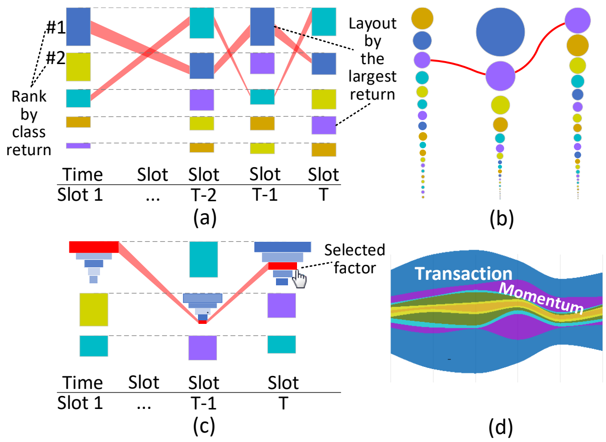

Fig. 3. Factor rank time series visualization: (a) high-level slope graph for factor classes; (b) alternative design with single-level compact layout; (c) low-level slope graph with reversed pyramid chart to display factors; (d) alternative design with stacked stream graph.

4.2.2 Factor Visualization

Design rationale: For the visual analysis task T1, our design focuses on ranking factor returns over time since the primary goal of an investment is to increase the overall return. Other factor metrics, such as uniqueness, and volatility, can be displayed by contextual visual attributes. The factor visualization design also tries to trade off between high visual abundance (a large number of factors) and low visual complexity (to reduce the cognitive and interaction burden).

Hierarchical slope graph: We propose a novel hierarchical slope graph design to visualize the factor return time series associated with multiple classes, as shown in Fig. 3. At the high level, the visualization is composed of multiple stacks of factor bars in the chosen investment period (Fig. 3(a)). Each stack corresponds to the factor returns within one time slot of the period (a month by default) and is partitioned into two sides by the timeline axis (Fig. 1(b)). The factors having the positive (negative) sum of returns within the time slot are placed above (below) the timeline. The stack on each side is aggregated into 6 bars by the factor class. These bars are ranked by the absolute value of their factor return sum in the class (also used as the bar height). The filled color hue of each bar indicates its factor class (see legends on the right).

The factor bar stacks are stacked vertically to emphasize the rank of overall return for each factor class. According to a default compact layout algorithm, which places factors next to one another, as shown in Fig. 3(b), a factor class with a higher rank in one time slot may have a different layout from the same factor class in a time slot with a lower rank. We propose a margined layout algorithm that vertically aligns factor bars of the same rank in all time slots. In Fig. 3(a), take upper-side stacks as an example. The vertical position of thw bar is computed from the top-ranked row to the bottom-ranked row. The vertical position of each row is determined by the bar with the largest height. By the margined layout, the connecting line of the same factor class can help the user identify the ranking change of factor class returns over time.

The high-level graph design abstracts 56 types of factors into class-based factor time series to control the overall visual complexity. To illustrate the details of individual factors, we propose an interaction-based low-level design. Upon mouse hovering over a factor class bar, the bar switches to a reversed pyramid chart composed of multiple equal-height horizontal strips (Fig. 3(c)). Each strip represents a factor in the class, with the strip width indicating the ratio of factor return to the average factor return in the class. By this design, the actual factor return will be linearly mapped to the area size of the factor strip and can be compared across time slots. Upon mouse hovering, a segmented line will connect the strips belonging to the same factor across time slots. This line will reveal the relative ranking change of the factor in the same class by the vertical line positions in all the pyramid charts. Also, the connecting line is drawn in a tapered style with line width indicating the temporal dynamics of factor returns. On each factor strip, an additional factor metric (i.e., return volatility by default) is displayed by fill color opacity, the more opaque the color, the lower the volatility.

The ranked factor graph design is augmented with versatile interactions to fulfill relevant user tasks. The focused factor strip can be mouse-clicked, working as a factor selection in our investing paradigm. All the selected factors are also listed in the leftmost column of the slope graph. In the configuration panel (Fig. 1(a)), users can choose the investment period and set the granularity of time slots (month/week/day). The factor return value is shown in the factor class bar, and the factor strip can also be switched to the sum of factor returns in a recent time window (6 months by default) instead of the per-timeslot display. A filtering mechanism is also introduced where small-return factors can be filtered out to highlight high-return factors. To serve the same purpose of reducing visual complexity, the hierarchical slope graph can only show positive or negative factors via checkbox options. On each factor strip, the fill color saturation can be switched from the factor volatility to its uniqueness to display more info. The default width of the factor bar is also adjustable in case many time slots are analyzed together.

Ranked factor treemap: In complement to the time series visualization of factor returns, the factor list on the top right (Fig. 1(c)) displays aggregated factor returns in the chosen investment period. A ranked treemap design is adopted, in which each row indicates a factor class, ordered by the overall return (also used as the row height). Each treemap row's columns are divided and ranked by the individual factor return (also used as the rectangle area size). To visualize the polarity of factor returns, a dashed line is drawn over each treemap rectangle, with the area size above (below) the dashed line indicating the sum of positive (negative) returns on this factor within the chosen period. The factor rectangle can also be selected or de-selected, having the same functionality as clicking on factor strips in the factor graph. The names of the selected factors above the rectangle are highlighted in red, which is used as the second selected factor list.

Alternative design: We have considered several alternatives for visualizing factor time series at the prototype stage. For instance, a line chart with symbols is the most natural way to visualize multiple factor returns over time, as shown by the iQUANT layout in Fig. 3(b). However, it suffers from serious visual clutter, as there are 56 factors. Also, comparing the factor rank in the iQUANT layout is not straightforward, as factor symbols are packed together vertically. The stream graph design is also tested (Fig. 3(d)), where stacked line charts are applied to illustrate the temporal dynamics of factor returns better. As the number of factors is too big to display in the same view by stacked line charts, the stream graph is limited only to show the factor return time series aggregated by 6-factor classes.

4.3 Visual Analysis of Stocks

4.3.1 Analysis of Individual Stock Returns

After user-preferred factors are selected, the second step of factor investing is to pick stocks to form an investment portfolio. Our method employs a scoring function to rank stocks by their performance on selected factors. The default scoring aggregates the risk exposure of selected factors in a recent time window on each stock. In Fig. 1(a), the stock ranking strategy and initial stock pool can be configured. When there are few good factors, the stocks are ranked by all factors. In China A-shares, users can choose the top-sized 300 stocks to start.

The candidate portfolio composed by top-ranked stocks can be appropriate for professional factor investors. They would like the portfolio to contain stocks as many as possible (e.g., hundreds) to diversify factor risks. Yet, independent investors normally can not purchase hundreds of stocks due to the constraints on capital and management cost. The shortlisted top stocks also may not be the best choice for a portfolio as no single stock-picking algorithm considers the preferences of all the investors. To popularize factor investing, we propose to allow users refine the portfolio of top-ranked stocks through a deeper understanding of individual stock's return performance. Following the cross-sectional regression model on each stock return (Eq. (1)), we introduce an AR+GARCH model to explain the idiosyncratic return of

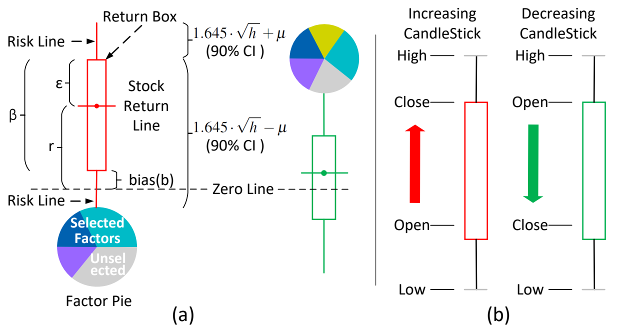

Fig. 4. Stock time series visualization: (a) firework chart showing components of individual stock return; (b) the original candlestick chart.

the stock ($\varepsilon_i^{(t)}$, denoted as $\varepsilon^{(t)}$ for convenience).

$$\varepsilon^{(t)} = \mu^{(t)} + \sqrt{h^{(t)}} \cdot e^{(t)} \tag{2}$$

where $\varepsilon^{(t)}$ forms the input time series of the model. $\mu^{(t)}$ is the mean process by an autoregressive (AR) model with a order of 5 (optimal parameter). The model order is determined by empirical analysis on Chinese stock market. $\sqrt{h^{(t)}} \cdot e^{(t)}$ is the error term modeled by a GARCH(1,1) model [9]. GARCH is a commonly employed model for financial time series that exhibit time-varying volatility and volatility clustering. Here $e^{(t)} \sim N(0,1)$ is the standard normal variable. $h^{(t)}$ is the time-dependent variance of the time series and represents the risk/volatility of the idiosyncratic stock return.

4.3.2 Stock Visualization

Design rationale: For the visual analysis task T2, we decided that our design should visually explain the critical components of stock returns as defined in the cross-sectional regression model and the AR+GARCH model. The system supports several useful stock sorting strategies. Users can refine the initial ranked stock list to optimize their portfolios. For example, factor investing for long sale advocates picking stocks that: a) earn high returns recently (large $r^{(t)}$); b) stock returns can be explained by selected factors (large $\beta^{(t)} = \sum_{j=1}^{F_S} x_{ij}^{(t)} \cdot f_j^{(t)}$); and c) have low risks (small $h^{(t)}$). Besides, our design can be used to study changes in stock performance over time. Meanwhile, to flatten the learning curve, we would like the visual representation to mimic standard stock time series visualizations such as the candlestick chart [1].

Stock firework chart: As shown in Fig. 1(e), the stock view is composed of multiple rows, with each row corresponding to one stock in the ranked list (Fig. 1(d)). For stocks beyond the first page, users can explore through a page navigator.

We propose a firework chart design based on the famous candlestick chart to visualize the return time series of each stock. The original candlestick chart, as shown in Fig. 4(b), displays the open/close price by a box and the polarity of the stock return by the color of the box. The highest/lowest prices of the stock are shown by the lines extending above/below the box. Inheriting this design, the firework chart, as shown in Fig. 4(a), is made up of a vertical box to display factor-related stock returns (return box), the lines extending from the box to indicate risk (risk line), a horizontal line either inside or outside the box to indicate the overall stock return (return line), and a pie chart connected to the risk line indicating factor diversification of returns (factor pie). We named our design the firework chart as its shape looks like a firework stick (return box) fueled by a ball of powder (factor pie). This visual metaphor resembles the factor investing model in that the stock return is fueled by its factor exposures.

In detail, the height of the return box indicates the sum of individual stock's return within the time slot explained by all factors ($\beta^{(t)} = \sum_{j=1}^{F} x_{ij}^{(t)} \cdot f_j^{(t)}$). Positive (negative) factor returns are drawn by upward red (downward green) boxes. The vertical position of the starting edge of the return box indicates the bias of the regression model in the current time slot ($b^{(t)}$). The vertical position of the return line represents the overall stock return in the time slot ($r^{(t)} = \beta^{(t)} + b^{(t)} + \varepsilon^{(t)}$). As a result, the height from the ending edge of the return box to the return line indicates the idiosyncratic return of the stock ($\varepsilon^{(t)}$). A dot tag is drawn at the center of the return line to disambiguate it from the edges of the return box. Extending outside the return box, the length from the ending edge of the box to the top (bottom) of the risk line indicates the maximal (minimal) idiosyncratic return with a confidence interval (CI) of 90% ($1.645 \cdot \sqrt{h^{(t)}} \pm \mu^{(t)}$). As a result, the range of risk line represents the probabilistic range of the stock's daily return under the regression and AR+GARCH model.

At another endpoint of the risk line, a pie chart is attached to represent the diversification of factor-driven returns. Each slice of the pie indicates a factor whose return has the same sign (β^{+}) with the sum of all factor returns ($\beta^{(t)}$). The slice size is proportional to its factor return. Unselected factors integrated together are designated grey. There are several issues with directly using each factor return value as the area size of the corresponding factor slice. First, we would like the size of the factor pie to be comparable to the return box, but it is hard to compare the size of different shapes visually. Second, we would like a larger factor pie to represent more preferred stock in the time slot. However, larger same-sign factor returns with the same overall return (return box size) suggest larger opposite-sign factor returns ($|\beta^{-}|$), an indicator of more controversy in all factors. Third, for aesthetic consideration, we would like the size of the factor pie to be contained, while the sum of same-sign factor returns is larger than the overall factor return and can be unlimited. To resolve these issues, we propose a weighted factor return measure $\beta^{*(t)}$, which is visually encoded by the height of the factor pie.

$$\beta^{(t)} = \beta^{+} + \beta^{-}, \quad \beta^{*(t)} = \beta^{(t)} \times \frac{\beta^{(t)}}{\beta^{+}} = \beta^{(t)} \times (1 - |\frac{\beta^{-}}{\beta^{+}}|) \tag{3}$$

By this design, the height of return box ($\beta^{(t)}$) and factor pie ($\beta^{*(t)}$) can be compared to understand the inconsistency ratio in factor returns ($|\frac{\beta^{-}}{\beta^{+}}|$). A larger factor pie indicates a more consistent factor return with the ratio close to 0. The factor pie can not be taller than the return box.

The user can add the corresponding stock to the portfolio or remove it by clicking on the label in the upper-left corner of the corresponding row. A ranked list of stocks is also displayed at the bottom-left of the interface (Fig. 1(d)). Clicking the "RANK&DRAW" button after changing the strategy will rearrange the list.

4.4 Visual Analysis of Portfolios

In the portfolio view (Fig. 1(f)), we introduce the weighted bump chart to depict the cumulative returns of selected stock time series. The bump chart highlights the rank of individual stocks in the portfolio for possible adjustment. It is challenging to interpret individual stock returns on conventional multiple-line charts because they may be similar or grouped. Since the actual stock return is also important for investment, we revise the original bump chart to visually encode the cumulative return of each stock at a time point by the circle size at the corresponding line plot. A red segmented line depicts the overall cumulative return in the chart for the entire portfolio. For comparison, another blue line displays the average market returns during the same time frame.

While the portfolio view illustrates the portfolio returns in the past, the visualization does not reveal its predictive performance in the future. To meet the visual analytics task of T3, we introduce backtest visualization based on the best practice of quantitative investment experts. As shown in Fig. 1(g), the cumulative returns of the selected portfolio are inferred, starting from the investment period's end time. These return outlooks, as well as the average market return, are depicted by clustered bar charts. The default lookahead window is set to 3 months as the portfolio will probably be adjusted after this period.

5 Case Study

5.1 Factor Investing on China A-Shares

The RankFIRST system was applied to the Chinese stock market to fulfill domain users' requirements in Sect. 3.2. We first chose the time

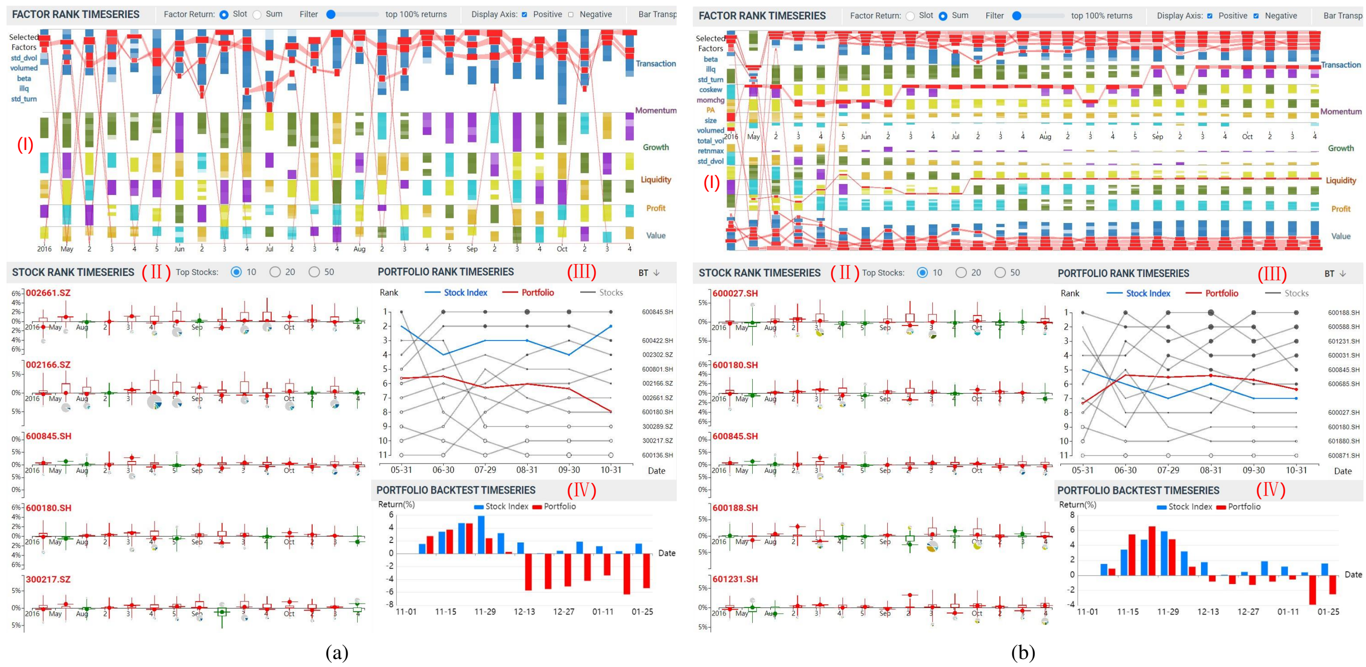

Fig. 5. Factor investing with RankFIRST on all China A-Shares at the time of Nov. 1st, 2016. 10 top-ranked stocks by their selected factor exposures are used to construct the investment portfolio: (a) selecting 5 factors in the "slot" mode; (b) selecting 11 factors in the "sum" mode.

of Nov 1st, 2016, when China A-Shares had been stable for a while. Suppose a factor investor would like to construct a stock portfolio at this time and s/he used our system to assist the investing process. The investor chose the recent 6 months as the investment time period (May to Oct, 2016) and picked the "slot" setting to visualize factor returns aggregated by weekly slots. As shown in Fig. 5(a).I, the ranked slope graph in the factor view reveals several factor classes having significant returns in this period. Notably, the factor class of "Transaction" ranks at the top in generating positive and negative returns. The temporal dynamics of these factor returns can be visually analyzed by hovering over each factor bar on the graph. The investor selected 5 factors with desired characteristics according to the criteria discussed in Sect. 4.2.1. Fig. 5(a).I only gives 3 factors on the positive side: volumed, illq, beta. The two factors with negative returns are std_turn and std_dvol. All the 5 factors have relatively stable return time series in the recent 3 months and are ranked high in their factor class. The other unselected factors are low in either stability over time or factor return ranking as placed close to the timeline of the slope graph.

After the factor selection, the investor ranked all the stocks in China A-Shares by the exposure of these factors in the chosen period. This ranking visualizes the stocks in firework charts, as shown in Fig. 5(a).II. The investor chose the default top-10 stocks to form a portfolio. Fig. 5(a).III illustrates the time series of individual stock returns in this period, and the return of the entire portfolio. It was found that the portfolio return (red curve) was slightly below the market return (blue curve) during this period. The investor further performed backtest evaluation and obtained future returns of the portfolio as shown in Fig. 5(a).IV. The backtest also resulted in a low portfolio return. The highest investment loss widened to -7% within 3 months, whereas the market return is always positive.

We examined the reason for this low performance and hypothesized that it was because of the small number of selected factors and factor classes. All 5 factors are from the "Transaction" class. The gray slices dominating the factor pie in Fig. 5(a).III indicates that unselected factors mostly drive the stock returns. Due to the factor selection priority on the temporal stability of factor return, very few factors are eligible in the current graph aggregated by time slots. In another trial, the investor switched to the "sum" mode when the slope graph showed the sum of factor returns in a recent 6-month window instead of the weekly average return (Fig. 5(b).I). In this mode, the factor selection focuses more on accumulated factor returns and less on temporal stability. Finally, the investor selected 11 factors with top accumulated returns. Notably, the largest factor in the "Momentum" and "Profit" classes were also selected (momchg, PA). Some factors, e.g., mom6 in the Momentum class, were excluded due to their extremely high volatility measure, as displayed in high transparency. The investor applied the same stock ranking and picking strategy to construct the portfolio. The return performance of the new portfolio was given in Fig. 5(b).III (past) and Fig. 5(b).IV (future), respectively. For most times, the portfolio return (red curve) was higher than or close to the market return (blue curve). Note that it is generally believed to be hard for independent investors to constantly outperform the market return. In the backtest, the portfolio gain was shown to increase to 7% within the first month of investment (5% in the "slot" mode) and the possible loss was reduced to -4% within three months (-7% in the "slot" mode). In Fig. 5(b).II, it was shown that the selected factors indeed represent a majority of factor returns, as illustrated by the dominating colored slices in the factor pie.

The investor went on to refine the portfolio of top-10 stocks by the "sum" mode factor selection. With the same selected factors, s/he examined the firework chart associated with each stock in Fig. 6(a). According to the criteria of good stocks in factor investing, stocks having large upward factor bars (high overall factor returns), large factor pies (consistent factor returns), and colored factor pies (stock returns dominated by selected factors) will be preferred. The investor finally picked 10 stocks to construct the new portfolio. Two stocks picked are listed as the red-labeled ones in Fig. 6(a): 000039.SZ, 000768.SZ. As shown in Fig. 6(c), the return of the new portfolio in the next 3 months increases by about 4% from the top-10 stocks' portfolio (Fig. 5(b).IV). This shows an improvement of factor investing performance through the visual analysis of stock returns decomposed by factors.

5.2 Investing via Stock Return Time Series

In the second case, we chose another time on July 1st, 2014, when the market was also stable in China (Fig. 1). To confirm the effectiveness of RankFIRST, the investor repeated the visual analysis process for factor investing as in Sect. 5.1. Because of the significant number of factors with large returns, a filter was applied, and only the factors with top 50% returns were left in the display, as shown in Fig. 1(b). On the "sum" mode factor visualization, 6 factors were selected. It was found that the factor treemap in Fig. 1(c) was also helpful in selecting top factors. The factor with a bar biased toward a positive or negative side could be selected due to its high positive/negative returns accumulated

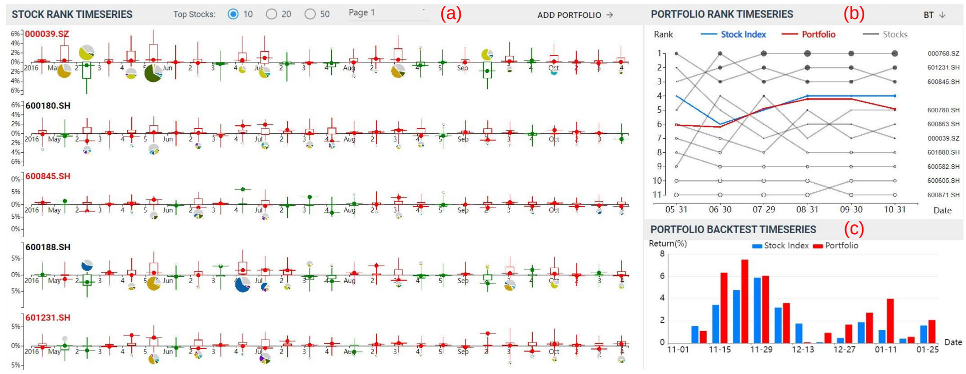

Fig. 6. The portfolio refined according to the stock firework chart: (a) stock view; (b) portfolio view; (c) backtest view.

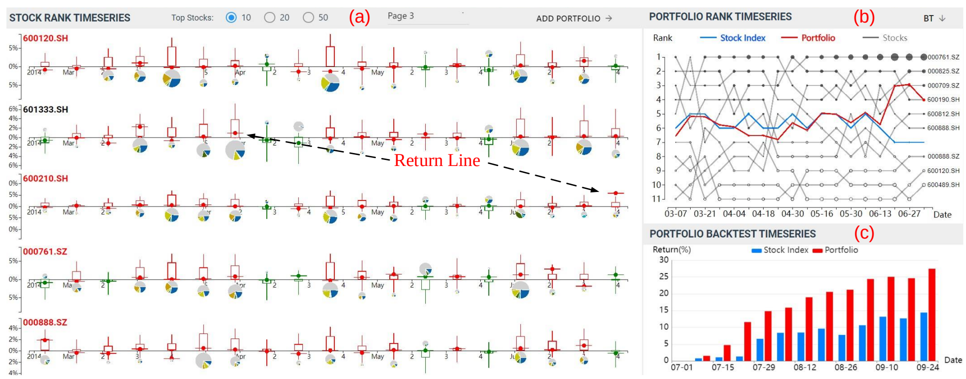

Fig. 7. The portfolio by stocks with positive returns and low risks: (a) stock view; (b) portfolio view; (c) backtest view.

over time. The fill color opacity was also switched to show the factor's uniqueness. A few factors with low uniqueness (less opaque in color) were excluded, such as illq and SgINVg. Finally, 10 stocks were picked according to the ranking strategy and the stock performance shown in the firework chart. Three of such stocks picked were highlighted in red in Fig. 1(e) (600462.SH, 000885.SZ, 600006.SH), all with large upward factor bars and large colored factor pies, especially in the most recent months. In the backtest (Fig. 1(g)), it was shown that a portfolio gain up to 33% can be achieved, 18% higher than the market index (15%). The investor also tried a manually refined portfolio with 20 stocks. A similar excess return of 18% can be achieved over the index.

The RankFIRST system also supports stock picking strategies other than factor-based investing. In our case, the investor turned to examine the time series of stock returns in Fig. 7(a). S/he would like to pick the stocks having positive overall returns (indicated by the return line) and low risks (indicated by the risk line) recently. These desirable features lead to positive portfolio returns in the chosen investment time period [7], which will probably persist in the near future. The investor finally picked 10 stocks as the portfolio. Two of these stocks were listed as red-labeled ones in Fig. 7(a): 600190.SH and 000761.SZ. The actual portfolio return depicted in Fig. 7(c) achieved 27% in the next 3 months, which was 12% higher than the maximal market return (15%).

5.3 Contrarian Investment Strategy

In the third case, we chose the time of Oct 1st, 2015. Unlike previous cases, the market was undergoing heavy fluctuations, and a "hill" curve was drawn in the market index during this time (the market advanced from Jan to June and crashed from July to Sept). Because of this rapid change in the market, there were hardly any stable factors that could be selected for factor investing. The investor decided to focus on large-capital stocks in the CSI-300 index and adopted a different contrarian investment strategy. The contrarian strategy looked at the alpha component of stock return, which was the idiosyncratic stock return (ε in Eq. (1)). It was believed that the idiosyncratic return would average to zero in the long term for every stock. Hence, stock having negative idiosyncratic returns will be more likely to bring positive returns soon, a generic stock market effect called 'return reversals' [25].

Fig. 8. The portfolio by the contrarian investment strategy: (a) stock view with a few stocks picked; (b) portfolio view; (c) backtest view.

The investor examined the stock view of Fig. 8(a) and picked 10 stocks having negative idiosyncratic returns recently (indicated by the height from the ending edge of the return box to the return line). Two such stocks were labeled red: 000999.SZ, 600132.SH. The actual portfolio return depicted in Fig. 8(c) achieved as high as 28% in the next 3 months, 9% higher than the maximal market return (19%).

6 User Evaluation

To evaluate the effectiveness of our system for factor investing, we also carried out quantitative studies with independent investors (see Sect. 3.2 for more details on the target user). The RankFIRST (RF) visual analytics system was compared with the baseline iQUANT (IQ) visualization tool [53]. Users were required to construct their portfolios with the assistance of the two systems. System performance was assessed by the returns earned in the next 3 months with the corresponding portfolio.

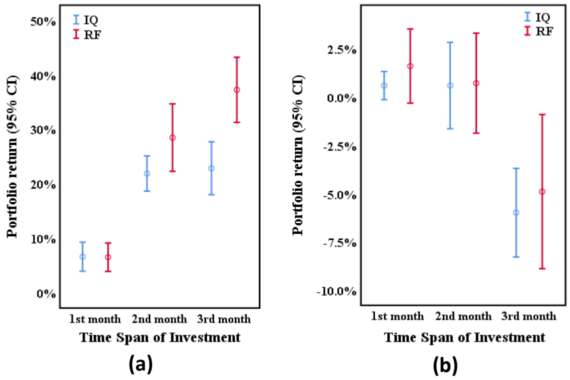

Fig. 9. The user study result on alternative stock visualization methods measured by portfolio returns: (a) time period I; (b) time period II.

The screenshots of the two systems are given in Appendix A.

Design. We recruited 12 graduate students as subjects of the study. All subjects were required to be experienced investors in the Chinese stock market, so that they could understand and execute the user task well. The experiment was designed to have two sessions. The first session was to train subjects in order to familiarize them with the visualization system and the task. The training session included one user task on a different investment time from the test session (March to Aug. 2016). The organizer checked the output of the training task and explained any ambiguity immediately.

In the formal test session, subjects were required to complete the same task again. To eliminate learning and ordering effect, we adopted a shuffled between-subject design. Each subject performed the same task with the two systems in turn, but in two different time periods: Feb. to July, 2014 (I); Mar. to Aug., 2017 (II). In detail, 12 subjects were partitioned into four groups by the test order of visualization systems and investment times. Because of the long time overhead for users to pick stocks one by one with IQ charts, we fixed the initial stock pool to 100 stocks sampled from China A-shares. The detailed user study document is described in Appendix A.

User tasks. Each subject was asked to select 10 stocks from the pool to construct their portfolio. The backtest functionality was disabled for fairness consideration. After the portfolio was built, subjects were asked to write down the selected factors, all the stocks in the portfolio, and unstructured feedback on the strategy to build the portfolio.

Result and analysis. The portfolio selected by each subject in the test session was used to compute the return curve in the next 3 months. The maximal returns of each portfolio in the next 1st/2nd/3rd months were extracted. Fig. 9 illustrates the comparison result of the two systems on portfolio returns with a 95% confidence interval. For time period I (Fig. 9(a)), RF achieves higher or close maximal portfolio return than IQ in average for all the next three months: 6.55%±2.5% ≈6.64%±2.5% (1st month); 28.5%±5.9%>21.9%±3.1% (2nd month); 37.3%±5.7%>22.9%±4.6% (3rd month). By the independent t-test, the differences are significant in the 2nd month ($t(10) = 2.43$, p=0.036) and 3rd month ($t(10)$= 4.80, p=0.001), but not in the 1st month ($t(10)$=0.065, p=0.949). The significance level is set to 0.05. The assumptions of normality/variance-homogeneity hold for all t-tests.

For time period II (Fig. 9(b)), the performance on maximal portfolio returns of the two systems is closer (RF vs. IQ): 1.6%±1.8%>0.6%±0.7% (1st month); 0.7%±2.5%≈0.6%±2.1% (2nd month); -4.9%±3.8%>-6.0%±2.2% (3rd month). By the independent t-test, the differences are not significant for all the next 3 months: $t(10) = 0.11$, p=0.19 (1st month); $t(10)$=1.85, p=0.86 (2nd month); $t(10)$=1.9, p=0.29 (3rd month).

The study result is divided. On the one hand, among all the 6 return measures reported, our system significantly outperforms the alternative (IQ) in the average statistics of 4 measures and is closer in the other 2 measures. On the other hand, in only 2 out of 6 measures, the differences are significant. It is found that subjects achieve better comparative performance with our system in the period I than period II. The primary index of A-shares was relatively stable in period I, with a maximal fluctuation of 8% (comparing the lowest index to the highest in this period). In the period II, the maximal fluctuation increased to 15%. We hypothesize that advanced factor investing with explainable models and rank-based visualization might be better than the baseline method in more stable stock markets. As described in the case studies, few good factors can be found in the market with rapid and significant fluctuations (Sect. 5.3). Yet, more research is required to validate this hypothesis.

Threats to validity. We caution that the analysis result only holds for the controlled study setting in this experiment. First, the performance is reported on ordinary investors to demonstrate the system effectiveness for the mass. For experts on factor investing, their customized factor selection and stock picking may lead to better comparative performance for RankFIRST. Second, we focus on the Chinese stock market and the factors designed for it. The result may not extend to other stock markets. Third, no time limit is set for the tasks. We assume portfolio construction to be the most critical in stock investment which worths significant cost. For amateur investors, this assumption may not hold.

Discussion. Despite the effectiveness of RankFIRST in factor investing, we observed a relatively high learning curve for first-time users. The user study mostly followed the recommended analysis trail advised in the training session. For example, the top factors and stocks in the corresponding view are often selected without further customization. Currently, we are preparing the deployment of RankFIRST for field study among ordinary users. It is hoped that the long-term user studies [40] can answer the question of whether ordinary investors can improve their expertise with the assistance of our factor investing tool.

Some users commented on our system as an explanation tool for the stock market's change in the past. They stated that a predictive system would be more demanding. Indeed, factor investing models are not entirely predictive, and the removal of backtesting in the user study further levels its predictive features. Yet, a fundamental question remains: would interactive factor investing accepting human input be better than the black box prediction model on the stock market?

7 Conclusion and Future Work

This paper presents RankFIRST, a visual analytics system built for ordinary stock investors in a typical factor investing scenario. On the analysis side, our system integrates a cross-sectional regression model and an AR+GARCH model. These models are studied and adopted due to their excellent visual explainability for factor risks and stock returns. On the visualization side, we introduce novel designs on the time series visualization of factor and stock ranks. A hierarchical slope graph design is introduced to highlight key characteristics of good factors from a large collection. To adapt to the current practice of stock time-series visualization, a new firework chart is invented by extending the well-known candlestick chart. Finally, the system stitches together the visual analytics techniques on factors, stocks, and portfolios to cover the entire cycle of factor investing. Case study results demonstrate that the portfolio return by factor investing with our system often outperforms the market index. In a complex market environment where factor investing is ineffective, our system offers versatile investing options, including stock picking by overall return and risk and the contrarian strategy. The user study evaluation also reveals our system's superiority over a baseline visual analysis tool.

Factor investing is a complex and worthwhile task to work on. Our proposed system, RankFIRST, still needs improvement. The system only adapts the cross-sectional regression model, and more factor investment models will be added in the future. The system's layout meets users' needs, but some designs could be optimized. For example, the firework design is novel, but it takes time to learn. The pie charts could be re-considered with suitable alternatives. In addition, long-term user and expert research is our important ongoing work.

Acknowledgments

This work was supported by National Key R&D Program of China (2021YFB3500700), NSFC Grant 62172026, the Fundamental Research Funds for the Central Universities, and SKLSDE. We would like to thank Dr. Xuanwu Yue for providing the iQUANT code.

References

[1] Candlestick chart. https://en.wikipedia.org/wiki/Candlestick_chart.

[2] Quantconnect. https://www.quantconnect.com.

[3] Ricequant. https://www.ricequant.com.

[4] Uqer. https://uqer.datayes.com.

[5] White paper for quantitative factor investing in chinese a-shares. http://mscf.pbcsf.tsinghua.edu.cn/2018/0316/20180316025702935.pdf, 2018.

[6] The stockmarket is now run by computers, algorithms and passive managers. https://www.economist.com/briefing/2019/10/05/the-stockmarket-is-now-run-by-computers-algorithms-and-passive-managers/, 2019.

[7] A. Ang, R. J. Hodrick, Y. Xing, and X. Zhang. High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, 91(1):1–23, 2009.

[8] M. Behrisch, J. Davey, S. Simon, T. Schreck, D. Keim, and J. Kohlhammer. Visual comparison of orderings and rankings. In *EuroVis*, 2013.

[9] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[10] M. M. Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82, 1997.

[11] 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 *VAST'07*, pp. 155–162, 2007.

[12] R. Chang, A. Lee, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto. Scalable and interactive visual analysis of financial wire transactions for fraud detection. *Information visualization*, 7(1):63–76, 2008.

[13] C. Csallner, M. Handte, O. Lehmann, and J. Stasko. Fundexplorer: Supporting the diversification of mutual fund portfolios using context treemaps. In *InfoVis'03*, pp. 203–208, 2003.

[14] H. T. Dao, A. L. Bazinet, R. Berthier, and B. Shneiderman. Nasdaq velocity and forces: An interactive visualization of activity and change. *Journal of Universal Computer Science*, 14(9):1391–1410, 2008.

[15] C. Di Sciascio, V. Sabol, and E. E. Veas. Rank as you go: User-driven exploration of search results. In *IUI'16*, pp. 118–129, 2016.

[16] M. Dumas, M. J. McGuffin, and V. L. Lemieux. Financevis.net-a visual survey of financial data visualizations. In *Poster Abstracts of IEEE Conference on Visualization*, vol. 2, 2014.

[17] E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

[18] E. F. Fama and K. R. French. The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, 18(3):25–46, 2004.

[19] E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

[20] E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

[21] A. Frazzini and L. H. Pedersen. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, 2014.

[22] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013.

[23] D. Han, J. Pan, F. Guo, X. Luo, Y. Wu, W. Zheng, and W. Chen. Rankbrushers: interactive analysis of temporal ranking ensembles. *Journal of Visualization*, 22(6):1241–1255, 2019.

[24] M. C. Hao, U. Dayal, D. A. Keim, and T. Schreck. Importance-driven visualization layouts for large time series data. In *InfoVis'05*, pp. 203–210, 2005.

[25] J. Kang, M.-H. Liu, and S. X. Ni. Contrarian and momentum strategies in the china stock market: 1993–2000. *Pacific-Basin Finance Journal*, 10(3):243–265, 2002.

[26] D. A. Keim, T. Nietzschmann, N. Schelwies, J. Schneidewind, T. Schreck, and H. Ziegler. A spectral visualization system for analyzing financial time series data. In *Eurographics/IEEE TCVG Symposium on Visualization*, pp. 195–202, 2006.

[27] 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, pp. 599–617, 2016.

[28] M. Lu, Z. Wang, and X. Yuan. Trajrank: Exploring travel behaviour on a route by trajectory ranking. In *PacificVis'15*, pp. 311–318, 2015.

[29] M. McCombe, A. Ang, K. Hogan, and Y. Gelfand. The rise of factor investing. Technical report, BlackRock, 2016.

[30] P. Mylavarapu, A. Yalcin, X. Gregg, and N. Elmqvist. Ranked-list visualization: A graphical perception study. In *CHI'19*, pp. 1–12, 2019.

[31] K. V. Nesbitt and S. Barrass. Finding trading patterns in stock market data. *IEEE Computer Graphics and Applications*, 24(5):45–55, 2004.

[32] M. M. Pereira and F. V. Paulovich. Rankviz: A visualization framework to assist interpretation of learning to rank algorithms. *Computers & Graphics*, 93:25–38, 2020.

[33] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *CHI'94*, pp. 318–322, 1994.

[34] S. Rudolph, A. Savikhin, and D. S. Ebert. Finvis: Applied visual analytics for personal financial planning. In *VAST'09*, pp. 195–202, 2009.

[35] A. Savikhin, H. C. Lam, B. Fisher, and D. S. Ebert. An experimental study of financial portfolio selection with visual analytics for decision support. In *HICSS'11*, pp. 1–10, 2011.

[36] T. Schreck, J. Bernard, T. Von Landesberger, and J. Kohlhammer. Visual cluster analysis of trajectory data with interactive kohonen maps. *Information Visualization*, 8(1):14–29, 2009.

[37] T. Schreck, T. Tekušová, J. Kohlhammer, and D. Fellner. Trajectory-based visual analysis of large financial time series data. *ACM SIGKDD Explorations Newsletter*, 9(2):30–37, 2007.

[38] J. Seo and B. Shneiderman. A rank-by-feature framework for interactive exploration of multidimensional data. *Information visualization*, 4(2):96–113, 2005.

[39] C. Shi, W. Cui, S. Liu, P. Xu, W. Chen, and H. Qu. Rankexplorer: Visualization of ranking changes in large time series data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2669–2678, 2012.

[40] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *BELIV'06*, pp. 1–7, 2006.

[41] K. Šimunić. Visualization of stock market charts. In *WSCG*, 2003.

[42] E. Sorenson and R. Brath. Financial visualization case study: Correlating financial timeseries and discrete events to support investment decisions. In *IV'13*, pp. 232–238, 2013.

[43] D. P. Tegarden. Business information visualization. *Communications of the Association for Information Systems*, 1(4):1–38, 1999.

[44] T. Tekusova and J. Kohlhammer. Applying animation to the visual analysis of financial time-dependent data. In *IV'07*, pp. 101–108, 2007.

[45] T. Tekusova and T. Schreck. Visualizing time-dependent data in multivariate hierarchic plots-design and evaluation of an economic application. In *IV'08*, pp. 143–150, 2008.

[46] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[47] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 2001.

[48] J. J. Van Wijk and E. R. Van Selow. Cluster and calendar based visualization of time series data. In *InfoVis'99*, pp. 4–9, 1999.

[49] E. Wall, S. Das, R. Chawla, B. Kalidindi, E. T. Brown, and A. Endert. Podium: Ranking data using mixed-initiative visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):288–297, 2017.

[50] D. Weng, R. Chen, Z. Deng, F. Wu, J. Chen, and Y. Wu. Srvis: Towards better spatial integration in ranking visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):459–469, 2018.

[51] J. Xia, Y. Hou, Y. V. Chen, Z. C. Qian, D. S. Ebert, and W. Chen. Visualizing rank time series of wikipedia top-viewed pages. *IEEE computer graphics and applications*, 37(2):42–53, 2017.

[52] 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*, 26(1):601–610, 2020.

[53] X. Yue, Q. Gu, D. Wang, H. Qu, and Y. Wang. iquant: Interactive quantitative investment using sparse regression factors. In *Computer Graphics Forum*, vol. 40, pp. 189–200, 2021.

[54] H. Ziegler, M. Jenny, T. Gruse, and D. A. Keim. Visual market sector analysis for financial time series data. In *VAST'10*, pp. 83–90. IEEE, 2010.

[55] H. Ziegler, T. Nietzschmann, and D. A. Keim. Relevance driven visualization of financial performance measures. In *Eurographics/IEEE TCVG Symposium on Visualization*, pp. 19–26, 2007.