

# iQUANT: Interactive Quantitative Investment Using Sparse Regression Factors

Submission ID: 1033

---

## Abstract

*The model-based investing using financial factors is evolving as a principal method for quantitative investment. The main challenge lies in the selection of effective factors towards excess market returns. Existing approaches, either hand-picking factors or applying feature selection algorithms, do not orchestrate both human knowledge and computational power. This paper presents iQUANT, an interactive quantitative investment system that assists equity traders to quickly spot promising financial factors from initial recommendations suggested by algorithmic models, and conduct a joint refinement of factors and stocks for investment portfolio composition. We work closely with professional traders to assemble empirical characteristics of “good” factors and propose effective visualization designs to illustrate the collective performance of financial factors, stock portfolios, and their interactions. We evaluate iQUANT through a formal user study, two case studies, and expert interviews, using a real stock market dataset consisting of 3000 stocks  $\times$  6000 days  $\times$  56 factors.*

## CCS Concepts

• **Human-centered computing**  $\rightarrow$  *Visual analytics; Visualization design and evaluation methods; Information visualization;*

---

## 1. Introduction

Quantitative investment using algorithmic trading has accounted for over 85% of transactions in US stock markets since 2012 [GK13]. There is a growing demand for machine learning models that can describe and explain the price change of an individual or a portfolio of stocks in the market. Factor investing is one of the most widely used strategies in quantitative investment. In a factor investing paradigm [Tsi18, FF92, FF04], a selective collection of financial variables called *factors* (e.g., a company’s fundamental data and its growth measures) are used as input predictors and stock returns are set as target outcomes to build predictive models. With these models, traders forecast future stock returns regularly in order to construct time-sensitive portfolios composed of a small set of high-return securities. The returns of these securities are expected to beat the average market return, which is known as the excess return. Recent global surveys show that the factor investing paradigm is employed by 90% of institutional investors to manage at least part of their portfolio [MAHG16]. The percentage of assets traded with factor investing is over 30% in size and still fast-growing.

Despite the prevalence of factor investing, selecting an effective set of factors for market investment is a difficult task. There can be hundreds of candidate factors or more available for predictive models. Currently, most quantitative traders manually select factors by comparing their corresponding portfolio returns in a standard backtesting. Such a trial-and-error approach is indirect, less intuitive, and often requires an exhaustive search in the vast combinatorial factor space. To compute returns from a set of factors, a predictive model and stock portfolio need to be constructed, whose interac-

tions with the selected factors are not fully understood and utilized in the current approach. Traders must also speculate an initial factor selection for evaluation. On the other hand, feature selection methods in machine learning research [GE03] have been applied to this domain, but not as an end-to-end approach like many other fields. Algorithm-selected factors are mostly used as a reference or starting point of the manual selection process, as these factors (especially in black-box models) are not trusted by traders until they can be explained and evaluated. There is currently few tools that can combine human intelligence (e.g., prior knowledge of a factor’s utility) and computational power (e.g., assessing stability and sensitivity of a factor) for factor selection in stock markets.

This work is committed to three essential requirements of interactive factor investing. First, traders need an intuitive and comprehensive way to assess the performance of numerous factors with respect to excess returns. Second, high-performance factors need to be selected and refined together with high-return stocks in the construction of an investment portfolio. Third, portfolios need to be evaluated and compared in the industry-standard backtesting for their returns in both past and future outlook. To meet these requirements, we present iQUANT (interactive Quantitative investment Using sparse regression facTors), a visual analytics system that seamlessly integrates human and algorithmic factor selection for effective factor investing in stock markets. iQUANT enjoys the best of both worlds from human intelligence by a heuristic search in the combinatorial factor space and from computational power by summarizing and presenting key performance metrics of factors and stocks in close interactions. The contributions of this work can be summarized as follows.

- We explore the use of sparse regression models as the initial feature selection method for factor investing. Sparse regression models are preferred as they optimize the interaction between factor selection and stock prediction in the same process. Performance metrics of factors are computed, including not only their feature weights, but also the factor contribution that describes its importance to the prediction.
- We introduce elaborate visualization designs that comprehensively illustrate the performance of factors in the regression model to predict the prices of candidate stocks. Our designs are based on expert studies about empirical characteristics of valuable factors, including stability, sensitivity, and complementarity.
- On top of the visualization interface, we propose an interaction framework for factor investing that allows traders to jointly select high-performance factors and high-return stocks in an integrated model-building and portfolio-construction process. By our framework, traders can iteratively refine the factor selection and stock portfolio based on visual evidence shown in the interface. The industry-standard backtesting is also supported.
- We evaluate the effectiveness of iQUANT in a formal user study comparing with baseline alternatives. Case studies with domain experts on targeted user tasks are conducted using a real stock market dataset. Expert feedback about iQUANT is also reported and discussed.

## 2. Related Work

### 2.1. Financial Data Visualization

Financial data, such as stock/fund prices and economic indicators, has been widely used by traders and investors to guide their investment decisions. Many visualization approaches have been proposed to support interactive analysis of financial data in the past decades [KCA\*16, DML14, Teg99]. Since the target users are often not visualization experts, classical designs such as line charts and bar charts are popular in financial visualization for better usability [CGK\*07, ZJGK10, SB13, YSZ\*19]. For example, Chang et al. [CGK\*07] applied multiple coordinated views with enhanced line charts and matrix designs to illustrate the time-varying wire transactions in banking. Ziegler et al. [ZJGK10] designed color-coded bars to display relative changes in asset prices over time. Sorenson and Brath [SB13] combined continuous line charts with icon-based event representations to inform users about the important context of financial data. With the ever-increasing data size, financial visualization methods are often integrated with automatic analytics to support visual analysis of financial data [Šim03, STKF07, SBTK08, VV99, ZJGK10]. Meanwhile, versatile visualizations beyond classical charts have also been developed to cope with the complexity of modern financial data, e.g., Growth Matrix [KNS\*06, ZNK07], wedge charts [DBBS08], and 3D visualizations [TS08, NB04].

Currently, few works have been done in integrating visualization into the factor investing process. Some existing studies [CHL\*03, RSE09, SLFE11] apply visualizations to help manage investment portfolios. FundExplorer [CHL\*03] introduces a distorted treemap to visualize both the composition of a fund's portfolio and the remaining stocks. Their design facilitates the convenient retrieval of complementary funds to achieve portfolio diversification. Fin-

Vis [RSE09] employs visualization to help non-experts understand the expected returns, risks and the changing context of multiple portfolios. PortfolioCompare [SLFE11] displays the variability and correlation of a portfolio's risks and returns by scatterplots and distribution charts. Their designs can assist users in making better investments according to their risk preferences. None of these visualization tools is integrated with predictive and feature selection models for factor investing.

### 2.2. Interactive Feature Selection

Feature selection is a key step in machine learning. Prior studies have focused on algorithmic feature selections [LCW\*17], whose mechanism and reasoning are often opaque to users. Inspired by visualizing predictive model works [LXL\*18, ZWLC19, MQB19, EASD\*19], interactive feature selections have been widely studied. These studies allow users to visually explore the feature space by novel representations and interactive interfaces. Seo and Shneiderman [JS04] proposed a rank-by-feature framework for the exploration of multidimensional data. Data features are ranked and visualized by scatter plots, bar charts, etc. Johansson and Johansson [JJ09] utilized parallel coordinates to visualize data features, with the order determined by user-defined metrics.

On the other hand, there are some works considering the relationship between features and predictive models. For example, SmartStripes [MBD\*11] combines interactive feature selection by allowing users to explore the dependencies between features and algorithms. INFUSE [KPB14] integrates feature selection with prediction models and a glyph-based design to inform users about the feature performance with respect to the prediction performance. iFEED [BS16] and FeatureMiner [CLL16] visualize the feature performance by line charts, bar charts and scatter plots. Prospector [KPN16] applies partial dependence diagnostics to analyze the correlation among features, data values, and prediction results. RegressionExplorer [DvVH\*19] combines feature space visualization of regression models with a dynamic subgroup of data analysis.

Existing studies summarize the feature dependency or performance by hybrid quality measures (e.g., feature ranking), but the temporal evolution of features are rarely considered. In factor investing, the temporal dynamics of financial factors (features) and their influence on the prediction model are crucial, which makes existing work not suitable for our scenario.

## 3. System Overview

In this section, the technical background of interactive factor investing is introduced, on which key user requirements are summarized. To meet these requirements, four user tasks are defined which are achieved by a pipeline design in our system.

### 3.1. Background

Historically, factor investing originated from the Capital Asset Pricing Model (CAPM) proposed by Fama and French [FF04], which uses a single factor of a stock's sensitivity to market returns to explain stock returns. Then, researchers and practitioners continued to discover new factors related to stock returns. Notably, Fama and French [FF92] proposed the famous three-factor model which

considers a size factor (large v.s. small capitalization stocks) and a value factor (high v.s. low book-to-market ratios) in addition to the market sensitivity factor. Later, many factors explaining stock returns were discovered, including a company's recent fundamental data [Asn97], the long-term income growth of companies [PLdSSV98, DS97], etc.

In a typical quantitative investment scenario for stock markets, traders hold a portfolio composed of a collection of individual stocks appropriately mixed in value. The portfolio is adjusted in regular time intervals, which is also known as the trading cycle. Depending on the style of traders and funds, the trading cycle can range from one day/week (e.g., private equity) to several months (e.g., public funds). By the factor investing paradigm, at the beginning of each trading cycle, traders pick a few factors based on their experience, domain knowledge, market status and factor returns. Quantitative models are then built up over historical market data, which try to establish linkages between factors and stock returns. Both machine learning and statistical models have been applied in this stage. From model-predicted stock returns, traders further apply a trading strategy to determine how to adjust their investment assets, i.e., when and which stocks to buy/sell. Because it is extremely hard to forecast how the portfolio adjustment decision performs in the upcoming trading cycle, a standard backtesting helps to evaluate its returns on historical data. The model consistently earns excess returns to the market, or at least in recent cycles, is considered good in the context of quantitative investment.

### 3.2. Task Characterization

The factor investing paradigm introduced above seems straightforward. However, due to the closedness of the quantitative investment community, we only came to this understanding after pilot studies with expert traders in several mutual funds, both public and private ones. In the interviews, almost all of them mentioned a key difficulty in the factor investing practice, the selection of valuable factors. While during their long-term jobs, a number of factor combinations proven in the past have been accumulated, they never stop finding and evaluating new factors because stock fluctuations in the future can hardly be entirely the same as the past. In a sense, the ability to discover new factors is the central competence of quantitative traders and accounts for most of their performance difference.

As we learned from experts in the interview, the current factor selection process is mostly done manually. A major user pain point lies in that the evaluation of selected factors are indirect, less intuitive, and often requires an exhaustive search in the vast combinatorial factor space to identify the best factor collection. In the standard backtesting of quantitative models, to be able to compute model returns, the learning/statistical algorithm and the focused stock need to be determined in addition to the factor selection. Importantly, these model returns are often the result of interactions among factor, model and stock selections. Users often require a tedious trial-and-error comparison process to determine factors. As such a process is done in every trading cycle, efficiency is also a critical consideration not to be neglected. Meanwhile, automatic factor selections by machine learning algorithms (e.g., feature selection) have been incorporated in a trader's job. Due to explainability concerns, these algorithm-picked factors are only used as a starting point or

reference for manual factor selection until traders can understand their working mechanism. To this end, interactive factor investing systems that can integrate feature selection, model building, and human intelligence in heuristic factor search, as well as visually explain selected factors are of great demand. Below we summarize key user requirements collected from our expert studies.

**R1. Direct and intuitive factor evaluation.** As there are a large number of candidate factors, for a given pool of stocks, the system should recommend an initial factor selection for users to start with (R1.1). Importance measures of these factors should be computed in the context of quantitative models to account for the interaction between factors and models. These factor importance and the performance of underlying stock/model need to be presented comprehensively to allow intuitive factor evaluation (R1.2).

**R2. Interaction support for factor and portfolio selection.** Based on evaluations of factor and stock performance, the system should support appropriate interaction to help construct stock portfolios to invest. This requires interactions to jointly select and refine factors and stocks, as well as a mechanism to help users understand.

**R3. Integrated model building and backtesting.** In the factor selection and evaluation process, the system should automatically train quantitative models to predict stock returns based on selected factors and stocks, algorithmically or jointly (R3.1). In factor investing are, backtesting of these models should be computed to compare stock returns of different portfolios and guide trader's investment decisions (R3.2).

To meet the requirements of expert traders, we target to build a visual analytics system that can support the following key tasks for successfully conducting interactive factor investing.

**T1. Selection of an initial pool of stocks and factors.** Traders can start from all the stocks in the market for their investment analysis. More frequently, s/he focuses on a few sectors (per funder's requirements) and picked several stocks from each sector to construct a balanced initial stock portfolio. Our system provides an interface to allow users to manually determine the initial stock portfolio from the list of stocks organized by sectors. Initial factor selections are automatically recommended based on the stock portfolio. (R1.1)

**T2. Evaluation of factors through visualization.** Our system illustrates the performance of individual factors in a comprehensive visualization interface. Users can visually identify valuable factors by examining key dimensions such as stability, sensitivity, and complementarity. The stock returns of quantitative models are also displayed as context. (R1.2)

**T3. Interactive joint factor refinement and portfolio adjustment.** Traders select factors and stocks to construct investment portfolios in an interactive and iterative manner. Upon each refinement of factors or stock portfolio, the stock return of the resulting model is displayed immediately, which provides visual evidence for user's heuristic factor search. (R2)

**T4. Evaluation and comparison of portfolio returns.** To compute portfolio returns, quantitative models are automatically built up over selected factors (R3.1). Traders compare the temporal dynamics of returns from alternative portfolios and select the best portfolio in a standard backtesting (R3.2).

### 3.3. System Design

We propose iQUANT, a visual analytics system that supports all the four user tasks defined above. iQUANT features a pipelined design as shown in Figure 1. In the first stage, the system takes three types of stock market data. The first type is stock transaction data, notably the daily opening/closing prices of individual stocks. The second type is market information, including but not limited to financial statements (total assets, debts, cash flows, etc.), market news, and financial reports. The third and most critical part of data are factors computed from stock transactions and market information, which is used in our system for user selection and model building. In this application, we apply data from a major stock market consisting of 3,000 stocks during a period of 30 years (1990-2018). The data is collected from both external sources (e.g., open APIs for transaction data) or by internal computations (e.g., financial factors). The underlying technique of our system can be easily extended to support most other stock markets.

In the second stage of Figure 1, we integrate a sparse regression model to carry out factor selection and return prediction simultaneously. The performance of selected factors and predictive models are visualized in iQUANT. The result of interactive factor investing is evaluated in a standard backtesting view in the final stage.

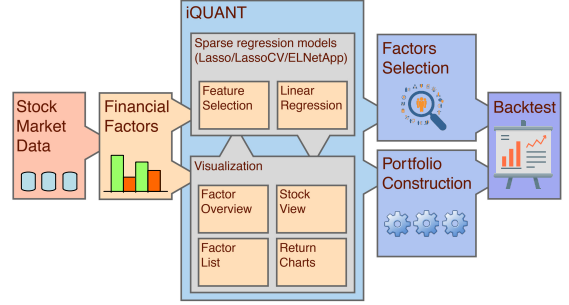
The iQUANT system is a web-based full-stack application. All data are stored in MongoDB. The front-ends are implemented by Vue.js and D3.js libraries. The sparse regression model is the most time-consuming. However, it only takes  $\sim 20$  seconds in the back-end to compute a sector (dozen) of stocks and respond to the web browser. All performance is measured in a laptop requesting a cloud server with a four-core CPU. We expect to upgrade the server to further reduce computation time.

## 4. Predictive Analytics

### 4.1. Sparse Regression Model

The interactive factor investing approach takes two steps to make final decisions on portfolio adjustment. First, the collection of financial factors are used to build models on the pool of candidate stocks for investment. Second, traders evaluate the predicted stock returns by the model and decide on the portfolio adjustment. In iQUANT, the second step is achieved by user interactions with the interface to understand the model output, and using existing trading strategies/algorithms (beyond the scope of this paper). In this part, we describe a suite of sparse regression models applied in the first step to jointly recommend factors (i.e. feature selection) and predict stock returns (i.e., model optimization).

Consider the model for a stock  $Y$  during a time period of  $\Gamma = [T+1, T+L]$ , where  $L$  is the length of  $\Gamma$  and any time  $t \in \Gamma$  indicates the  $t$ th trading day.  $\Gamma$  is further partitioned into  $N$  trading cycles:  $\{[T+1, T+D], \dots, [T+(N-1)D+1, T+ND]\}$ , where  $D$  indicates the length of a trading cycle by days,  $D = L/N$ . Take the first trading cycle  $\tau_1 = [T+1, T+D]$  as an example, the market information in a previous training period  $[1, T]$  is used in the predictive modeling.  $T$  denotes the fixed length of training period by days. Similarly, for the modeling of the  $i$ th trading cycle  $\tau_i = [T+(i-1)D+1, T+iD]$ , the training period of  $[(i-1)D+1, (i-1)D+T]$  is used.



**Figure 1:** iQUANT system pipeline composed of two modules: the sparse regression model for feature selection and prediction, and the visualization of factors and stocks for factor selection and portfolio construction.

On stock  $Y$ , denote the actual daily stock returns by percentage in the time period  $[1, T+L+1]$  by  $\mathbf{y} = (y_1, \dots, y_{T+L+1})$ . Let the model take  $F$  factors as input. Their normalized factor values in  $[1, T+L+1]$  compose a matrix  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{T+L+1})$  where  $\mathbf{x}_i = (x_{i1}, \dots, x_{iF})'$ . The modeling of stock  $Y$  in  $\Gamma$  takes  $N$  steps, corresponding to the trading cycles of  $\tau_1, \dots, \tau_N$ , respectively. Consider the  $i$ th trading cycle  $\tau_i$ , a sub-vector of actual stock returns is used as outcomes, denoted by  $\mathbf{y}^{(i)} = (y_{(i-1)D+2}, \dots, y_{(i-1)D+T+1})$ . A sub-factor-matrix is used as the design matrix (predictors), denoted by  $\mathbf{X}^{(i)} = (\mathbf{x}_{(i-1)D+1}, \dots, \mathbf{x}_{(i-1)D+T})$ . We fit a sparse regression model:

$$\mathbf{y}^{(i)} = \mathbf{w}^{(i)'} \mathbf{X}^{(i)} + b^{(i)} + \varepsilon^{(i)} \quad (1)$$

Subject to the objective function:

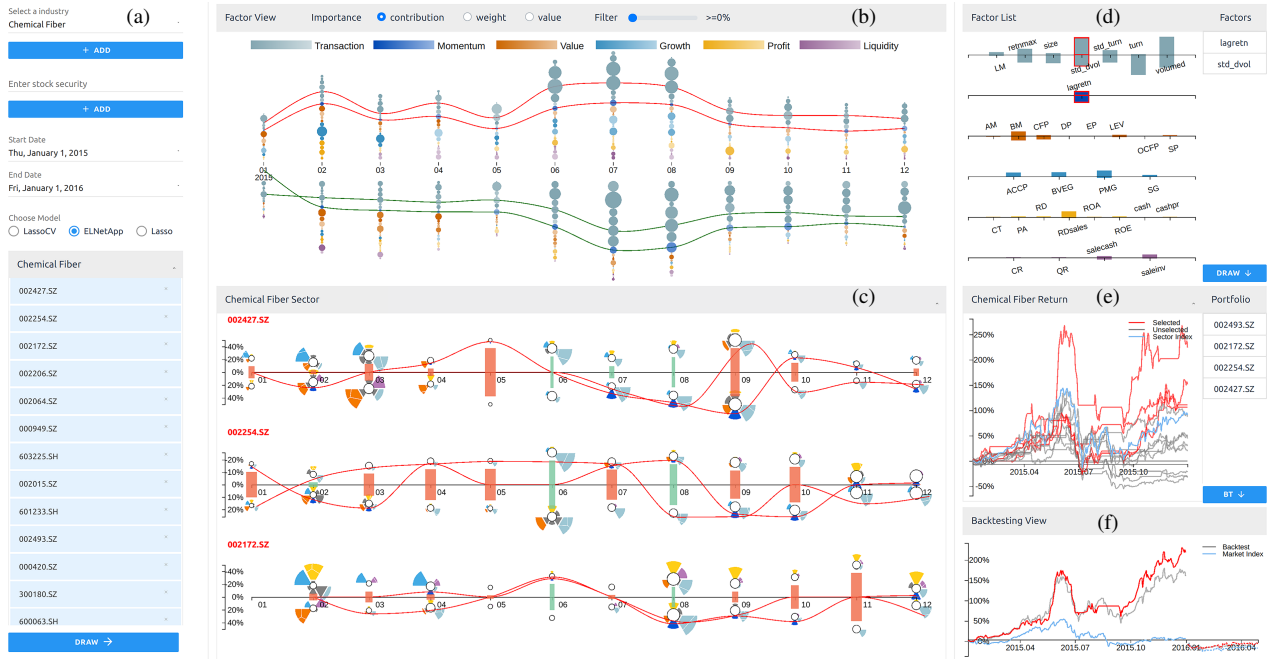
$$\text{Minimize } \text{Lasso\_NLL} = \text{NLL} + \frac{1}{2} \lambda [\alpha \|\mathbf{w}^{(n)}\|_1 + (1-\alpha) \|\mathbf{w}^{(n)}\|_2^2] \quad (2)$$

where  $\mathbf{w}^{(i)} = (w_1^{(i)}, \dots, w_F^{(i)})'$  denotes the weight vector of financial factors in the fitted model,  $b^{(i)}$  is the bias,  $\varepsilon^{(i)}$  is the error term,  $\|\cdot\|_1$  ( $\|\cdot\|_2$ ) is  $L_1$  ( $L_2$ ) norm of the vector.  $D$  is set to 21 (average number of monthly trading days). The best weight vector is computed by minimizing the negative log likelihood (NLL) of a linear regression term plus  $L_1$  and  $L_2$  penalty terms of the weight vector. The penalty terms shrink the weight of unimportant factors to zero. On each trading cycle, the model is applied to predict daily stock returns.

Note that  $\lambda$  and  $\alpha$  in Eq. (2) control the degree of model sparsity and weights between  $L_1$  and  $L_2$  terms. In this work, we provide three choices of modeling:  $\alpha = 1$  (without the  $L_2$  term), the famous *Lasso* model [Tib96] with a fixed  $\lambda$ ; *LassoCV* model using 10-fold cross-validation to determine  $\lambda$  of the *Lasso* model [UGH09]; and  $\alpha = 0.5$ , a balanced *Elastic Net* model (*ELNetApp*) [ZH05]. The trade-off of the three models in predictive performance and the number of selected features can be found in recent literature [Tib96] [UGH09] [ZH05].

The sparse regression model achieves feature selection ( $L_1$  and  $L_2$  penalty terms) and model optimization (regression term) simultaneously. This is extremely helpful to alleviate the overfitting effect in the factor investing case when the number of predictors ( $F = 56$  in our implementation) is comparable to the number of training samples ( $T = 200$ ). Though we apply a basic linear regression for stock prediction, the sparse models with  $L_1$  and  $L_2$  penal-





**Figure 2:** Interactive selection of four stocks from the Chemical sector in 2015 using a two-factor model (**lagretn+std\_dvol**): (a) control panel; (b) factor view; (c) stock view by sectors; (d) factor list for selection; (e) stock return view for portfolio construction; (f) backtesting view for evaluation of returns.

ties can be easily extended to support more sophisticated models such as nonlinear models using customized kernel functions.

#### 4.2. Financial Factors

For factor computation, we adopt the result of a latest white paper on the target stock market [Tsi18], which recommends six types of 56 factors with each factor characterizes a key aspect of a stock:

- **Transaction friction:** 17 factors delineating overall status of a company (e.g., firm size by its market value, age of a firm) and transaction statistics (e.g., volatility, risks measured by stock returns compared with the market return);
- **Momentum:** 5 factors computing daily stock returns in the recent 6/12 months, the change of momentum, and specially defined momentums;
- **Value:** 8 factors including the famous book-to-market ratio proposed by Fama and French [FF92], the asset-to-market ratio, and other company performance related ratios;
- **Growth:** 11 factors describing the growth of a company's asset, debt, market value, sales, profit, tax, etc;
- **Profitability:** 8 factors quantifying return on equity, return on asset, and other factors related to company profits;
- **Liquidity:** 7 factors characterizing the liquidity performance of a company, including current ratio, quick ratio, cash flow to debt ratio, etc.

For the above factors, the ranges of their values differ significantly from each other. Also, their values can change dramatically along with the stock market. The use of a sparse regression model will normalize the relative importance of each factor, which can facilitate the factor analysis. The sparse regression model in iQUANT outputs three measures to delineate the importance of each factor

in the prediction. Take the  $j$ th factor in the  $i$ th trading cycle (e.g., daily, monthly or quarterly) as an example, the three measures are as follows:

- Factor **weight** learned in the model, denoted by  $w_j^{(i)}$ ;
- Factor **value** changing in a daily basis, denoted by  $\mathbf{X}^{(i)}$ ;
- **Contribution** of factors to the prediction of daily stock returns, denoted by  $\sum_{k=T+(i-1)D+1}^{T+iD} w_j^{(i)} x_{kj}$ .

By collecting the feedback from traders in our pilot study, we further extracted three metrics that are widely-used in industry to delineate the utilization of each factor. These metrics are visualized in iQUANT to aid the interactive factor selection and stock portfolio construction process.

- **Sensitivity:** a metric that indicates whether the utilization of a factor can be replaced by other factors. The sensitivity is computed by  $(\xi_{-j}^i - \xi^i)^+$  where  $\xi_{-j}^i$  denotes the prediction error by removing the  $j$ th factor from sparse regression models,  $\xi^i$  denotes the prediction error of the original model;
- **Stability:** a good factor for the predictive model should have stable contributions in a long period of time before confidently applied. Quantitatively, it is measured by the number of times the factor's contribution to a stock's prediction flips from negative to positive or from positive to negative.
- **Complementarity:** a set of factors will perform better when they are complementary to each other, i.e., contribute to a stock's prediction in separate time periods.

#### 5. Visualization

The iQUANT interface is composed of six coordinated views.

### 5.1. Control Panel

The *control panel* in Figure 2(a) allows users to select an initial pool of stocks for analysis ( $T1$ ). S/he can input the stock code to select an individual stock or choose an entire sector of stocks from a sector list. All the selected stocks will appear in the bottom part of the control panel, aggregated by sectors. Next, users need to specify a time period and a model for factor investing. Finally, users click the “Draw” button to visualize factors related to the selected stocks.

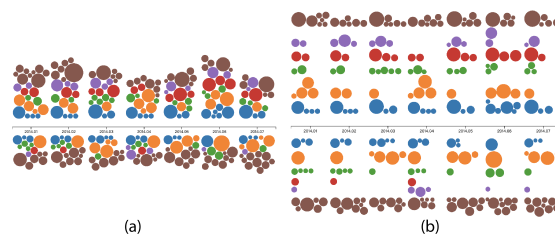
### 5.2. Factor View

*Factor view* (Figure 2(b)) is designed to provide traders with a quick overview of financial factors relevant to the selected stocks and time period. These factors are initially recommended by sparse models in iQUANT ( $T1$ ). The model measures the importance of each factor by its weight, value, or contribution to the prediction of stock returns. By default, traders first look at the *contribution* measure which covers both its current value and the overall weight. The contribution can be either *positive or negative* depending on the relationship between the factor and stock returns. According to our expert study, they also consider several other metrics in the factor evaluation and selection process, including the temporal *stability* of factors and their *sensitivity* in the model. ( $T2$ )

To meet the above design requirements, we propose a customized visual design to integrate time-varying factor importance metrics in a single factor view. As shown in Figure 2(b), the x-axis of the view indicates a timeline divided into multiple time units (months by default). Each factor is visualized as a series of filled circles in these time units. On each circle, the color hue encodes factor type, the color opacity encodes the factor sensitivity to the model, and the circle radius represents the aggregated factor importance to all the selected stocks. Three importance metrics can be displayed by the checkboxes in the top row of the factor view: weight, value, and contribution (default). All the circles on the same time unit are vertically aligned by a fixed order of their corresponding factors to facilitate factor tracking and comparison along the timeline. As each factor can have a positive or negative contribution in different time units, we depict positive and negative factors as circles above and below the x-axis respectively.

User interactions are designed in the factor view to help identify useful factors for modeling. When a user hovers a factor circle, all the circles representing the same factor are linked by a curved line to reveal its stability over time. (Figure 2(b)). The interface could interactively filter factors to reduce visual complexity.

**Alternative designs** of the factor view have been considered. The line chart is the most intuitive form to visualize factor importance over time. When we have as many as 56 factors, line charts suffer from severe visual clutter. Stacked flow chart [BW08] is another possible visual design. A quick overview of factor composition can be provided, but it is not good at displaying the temporal dynamics of individual factors (especially for unimportant factors). For the circle layout, we also experimented with circle packing in which positive and negative factors are placed separately. The circle packing leads to more compact visualization (Figure 3(a)). However, our domain experts complained that the design made it difficult to track individual factors across the timeline. The relative position of the same factor circle varied a lot across different time



**Figure 3:** Alternative designs of the circle layout in the factor view: (a) circle packing; (b) layered circle packing.

units. In an improved layered circle packing (Figure 3(b)), the same type of factors is aligned horizontally. It facilitates tracking of the same factor, but the space usage is inefficient and the connecting lines on factors lead to heavy crossings.

### 5.3. Stock View

*Stock view* (Figure 2(c)) splits the aggregated factor time series in the factor view according to their importance to individual stocks. This helps to evaluate the factor performance among different stocks and sectors ( $T2$ ). The visual design is challenging, as the space for each stock is limited and there could be many factors relevant to a stock. On each factor, there are also multiple attributes to be visualized, e.g., the magnitude and polarity of importance (contribution), the model bias, and the trend of the stock price.

In our design, the visualization of each stock follows a similar layout to the display in the factor view. As shown in Figure 2(c), the x-axis is a timeline divided into units. On each unit, a stock bar is drawn, whose height represents the ratio of stock price change in this time unit and width represents the rate of prediction error on the price change by models. A red (green) bar indicates a rise (drop) of the stock price. On top of and below each stock bar, there are two *factor glyphs* which summary the factors having a positive and negative contribution to the price change in the predictive model.

The factor glyph (Figure 6(d)) has a circular shape composed of one inner circle and six angular sectors in the outside. The radius of the inner circle encodes the model bias. Each outer sector represents one type of factors, with the size of each sector indicating the sum of all factor’s importance in this type. On each factor glyph, the top-5 factors with the highest importance in all factor types are displayed. When more than one factors from the same type belong to the top-5, a Voronoi diagram is drawn to pack these factors in the same angular sector. Each region in the Voronoi diagram corresponds to one factor, with the region size indicating its importance. The sum of all the other factor’s importance out of the top-5 category is represented by the grey region close to the glyph center. When users hover a Voronoi region of one factor, a line will be drawn to connect all Voronoi regions of this factor over time. When the factor does not belong to the top-5 category, the line will be linked to zero position on the x-axis in the time unit.

**Alternative designs** such as line chart and streamgraph [HHN00] are commonly used for visualizing temporal evolution of variables. In our scenario, the top-5 factors can be quite different in separate time units, making line charts or streamgraph discontinuous over time and hard to track by users. Commodity designs such as stacked bar charts and pie charts are also considered. Nevertheless, the stacked bar chart is not scalable

to support dozens of factors in the same view. The pie chart can clearly show the percentage of each factor in importance, but again the relative position of the same factor can vary a lot in different units, making it hard to compare the evolution of the same factor.

#### 5.4. Factor List

*Factor list* (Figure 2(d)) aggregates factor data together to depict the overall factor performance (T2). The six types of factors are listed in rows, with factors in one type uniformly distributed as factor bars in the same row. Each factor bar represents the positive and negative contribution of the factor in the entire time period, using the height of the bar above and below the x-axis respectively. Users can manually select factors by clicking factor bars, then the stock prediction models are re-built using the selected factors after triggering the “Draw” button. The factor performance in the stock view is also updated for iterative factor/stock selection. (T3)

#### 5.5. Stock Return View

*stock return* view (Figure 2(e)) is organized by sectors. In each expandable sector box, the investment returns of all stocks in that sector are calculated by backtesting and displayed in multiple grey line charts. An additional blue line indicates the market return. Users could select promising stocks in the stock view and examine their returns by the lines highlighted in red. The stocks could be selected in the return view and added into a *portfolio* table. A “Backtest” button will trigger the backtesting computation by the selected factors (factor list) and selected stocks (portfolio table).

#### 5.6. Backtesting View

*Backtesting view* (Figure 2(f)) evaluates the investment performance of the selected portfolio (T4). Backtesting calculates the portfolio returns using historical stock data. The blue line indicates the average return in the market and the grey one indicates the portfolio return. A dotted line after the selected time period represents the predicted return in future. Comparison among multiple portfolios is enabled by adding more portfolio from the stock return view.

#### 5.7. Multiple Views Interaction

iQUANT supports interactions among multiple views for comprehensive evaluation of factor performance (T2) and interactive selection of stocks and factors (T3).

**Selection of factors:** Users can select/de-select factors by clicking factor circles in the factor view, Voronoi regions in the stock view, and factor bars in the factor list. Each change to the factor selection will trigger updates in all the views simultaneously.

**Selection of stocks:** An initial collection of stocks can be selected in the control panel. Users can further compose an investment portfolio by clicking the stock name in the stock view or the return curve in the stock return view. This helps to interactively and iteratively refine stock selection according to their returns or the performance of relevant factors.

## 6. User Experiment

### 6.1. Design

We conducted a formal user study to compare the effectiveness of our iQUANT visualization design with a baseline factor visualization using commodity multiple line charts (See the supplementary

material for more details of the baseline factor visualization used in the study). Since there is no such an efficient tool that can fulfill the tasks mentioned in *Task Characterization*, we chose to compare iQUANT with line charts, the widely-used visualization in the stock market. In both designs, the time series of financial factors and stock prices were displayed. Users were required to complete multiple tasks involving the selection of factors and stocks for investment (see below for details), which corresponded well with T1~T3 in *Task Characterization*. In total, 14 expert quantitative traders were recruited as subjects, who were paid for their valuable time. We applied a between-subject design that subjects were randomly assigned to two groups of equal size. The seven subjects in each group will complete all the tasks with only one visualization, iQUANT or baseline. The experiment was composed of two sessions. In the first training session, subjects were presented with a tutorial and allowed a trial usage of the visualization interface to be tested. In the subsequent formal test session, each subject took seven tasks on the tested visualization. To alleviate the influence of financial data on the result and increase statistical power, we presented two stock data. Each subject was required to complete all the tasks with both data using the tested visualization. The full documents of the experiment design are provided in Appendix A. The user tasks can be categorized into four groups.

*Q1/Q2 (factor selection by contribution):* Among all the factors, which factors have the highest positive/negative contribution to the return of a particular stock during a given time period?

*Q3 (factor selection by stability):* Among all the factors, which factors have the highest unstable contribution to the return of a particular stock during a given time period?

*Q4/Q5 (factor selection by contribution to a sector of stocks):* Among all the factors, which factors have the highest positive/negative contribution to the return of all stocks in a particular sector during a given time period?

*Q6/Q7 (stock selection by factor contribution):* Among all the stocks in a given sector, whose return receive the highest positive/negative contribution from a selected factor during a given time period?

For each question, the subject was required to provide 1 ~ 2 answers. Because of the between-subject design, subjects were not asked for their subjective rating on each visualization (no within-subject comparison). Instead, we collected their verbal feedback on the pros and cons of each design.

### 6.2. Result and Analysis

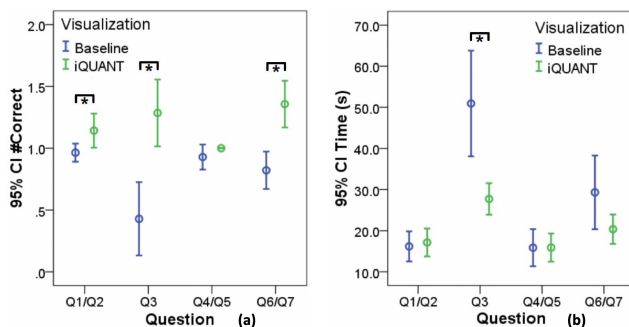
We compute the accuracy of each user’s response to a question by comparing them with top-2 answers. The overall accuracy is measured by the number of correct answers to each question. In Figure 4, we depict the accuracy and completion time of four groups of seven questions. Because both accuracy and time do not follow normal distributions, we apply the Mann-Whitney test to analyze.

On task accuracy (Figure 4(a)), the result reveals that, for factor selection by contribution (Q1/Q2), by stability (Q3), and stock selection by factor contribution (Q6/Q7), the accuracy of iQUANT is significantly higher than the baseline design:  $U = 324.0$ ,  $p = .024$

for Q1/Q2;  $U = 30.0$ ,  $p = .001$  for Q3;  $U = 207.0$ ,  $p < .001$  for Q6/Q7. The average task accuracies are  $1.14 \pm 0.14$  (iQUANT) and  $0.96 \pm 0.07$  (baseline) for Q1/Q2;  $1.29 \pm 0.27$  (iQUANT) and  $0.43 \pm 0.3$  (baseline) for Q3;  $1.36 \pm 0.19$  (iQUANT) and  $0.82 \pm 0.15$  (baseline) for Q6/Q7. For factor selection in a sector (Q4/Q5), the accuracy difference between the iQUANT and baseline is not significant ( $U = 364.0$ ,  $p = .15$ ), though the average accuracy of iQUANT ( $1.0 \pm 0.0$ ) is higher than the baseline ( $0.93 \pm 0.1$ ).

On task completion time (Figure 4(b)), the result reveals that, for factor selection by stability (Q3), the completion time using iQUANT is significantly shorter than the baseline design ( $U = 30.5$ ,  $p = .002$ ). The average completion times are  $27.2s \pm 3.8s$  (iQUANT) and  $50.9s \pm 12.8s$  (baseline). For factor selection by contribution (Q1/Q2), in a sector of stocks (Q4/Q5), and stock selection by factor contribution (Q6/Q7), the completion time using iQUANT is not significantly shorter than the baseline design:  $U = 359.0$ ,  $p = .58$  for Q1/Q2;  $U = 367.0$ ,  $p = .68$  for Q4/Q5;  $U = 342.0$ ,  $p = .41$  for Q6/Q7. The average completion times are  $17.1s \pm 3.4s$  (iQUANT) and  $16.2 \pm 3.6s$  (baseline) for Q1/Q2;  $15.9s \pm 3.4s$  (iQUANT) and  $15.9s \pm 4.5s$  (baseline) for Q4/Q5;  $20.4s \pm 3.6s$  (iQUANT) and  $29.3s \pm 9.0s$  (baseline) for Q6/Q7.

The study result indicates that, on most tasks, iQUANT leads to higher task accuracy than the baseline design. The only exception happens on the factor selection task for a sector of stocks (Q4/Q5). It is found that in the stock data used for these two questions, the factor having the highest positive/negative contribution is quite separated from all the other factors. Hence, only the top-1 factor is used as the correct answer, and in both designs, subjects could easily identify this factor (a 100% accuracy for iQUANT). For similarly easy tasks (Q1/Q2 and Q6/Q7), though the baseline design in most case helps to identify one (but only one) top factor, the accuracy of iQUANT is significantly higher because subjects could quickly discover more top factors in comparison to the baseline design. For the hard task of selecting factors by stability (Q3), the advantage of iQUANT in accuracy becomes the highest. Meanwhile, on completion time, iQUANT is favored only in the hard task of Q3, for which the iQUANT design can better support temporal comparison and analytics of factor contribution. For the other tasks, the completion time is similar in both iQUANT and baseline design.



**Figure 4:** User experiment results: (a) the number of correct answers provided by users in two visualization designs; (b) the task completion time.

### 6.3. Expert Feedback

In the user experiment, experts also provided valuable feedback about several aspects of iQUANT.

On positive aspects, some experts stressed that iQUANT presented an effective way to visually compare salient factors among different sectors of the market. In practice, the factor constituent and dynamics are truly different across sectors, which is of great importance to factor investing. As supported by the study result, experts also found iQUANT to be useful in analyzing the temporal dimension of factors, as well as the relationship between factors and stocks, which are exactly our design goal. Overall, subjects reported the iQUANT interface to be user-friendly, as our design was similar to their familiar and popular tools. Some experts even commented that our design could be part of standard stock market analysis tools (e.g., the work from Bloomberg [SB13]).

On the other hand, it was suggested that the interface could incorporate more statistical information according to the industry standard (e.g., the change of stock holdings in the backtesting view). The Voronoi map for top-5 factors could be reduced to only show the top-3 or less, in order to better reveal the most salient signal for the stock market. More detailed visual encoding could be applied to distinguish the factors of the same type, which are now depicted in the same color. Notably, we thought the complexity of the interface could raise learning issues. During the study, most subjects could be well prepared after a 10-minute training session, as indicated by the high task accuracy in applying iQUANT.

## 7. Case Study

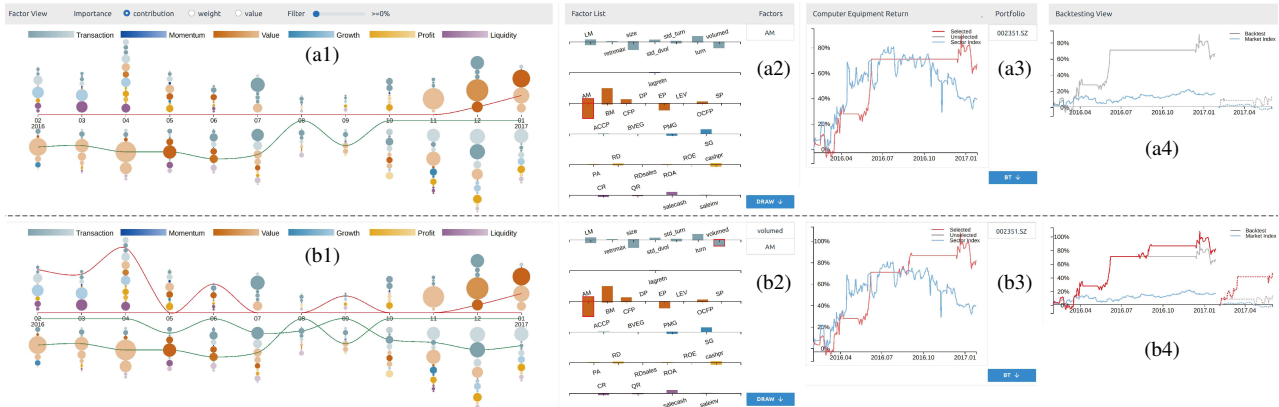
This section showcases how iQUANT helps users in two key scenarios of quantitative investment.

### 7.1. Selection of Investment Time

We first invited a fund manager to select the best investment time for individual stock with iQUANT. He looked at the year of 2016.2-2017.1 when the market index rose by 15.7%. From a list of candidates, the manager examined each stock until a feasible investment time for excess return is identified.

The manager reported stock in the Computer sector (code: 002351.SZ) for which the composite index of the whole sector rose by 2.5%. In his analysis, the Lasso model was selected. As shown in Figure 5(a1), the initial factor view displays a summary of important factors recommended by the Lasso model. In the factor list view (Figure 5(a2)), the manager identified several useful factors having large contribution to the stock return. He hovered these factors one by one to examine their performance in the given period. The AM factor (asset-to-market ratio) was chosen because its contributions are stable in the first six months: all negative to the rise of stock prices. There are two months when the factor sensitivity is high, as indicated by opaque colors. By this one-factor model, the investment return (red curve in Figure 5(a3)) surpasses the stock price change (blue curve) by 26.5% at the end of the period. A maximal return of 11% can also be forecast for the next three months, as shown by the grey dotted curve in the backtesting view of Figure 5(a4). Notably, the manager suggested applying the model from June when the AM factor has been stable for four





**Figure 5:** The selection of investment time during 2016.2-2017.1 on 002351.SZ by: (a) a single-factor model (AM); (b) a two-factor model (AM+volumed). The manager switched from a single-factor model to a two-factor model to optimize his investment strategy.

months and enjoys a high sensitivity in the latest month (May). In case this strategy was adopted, an excess return of 39.1% could be earned in June when the overall market only rose by 1.6%.

To optimize the investment for the whole year, the manager explored other factors that can characterize the price change in the second half of the year. He found that the **volumed** factor was complementary to **AM**, as the **volumed** factor contributed to the model mainly from July (Figure 5(b1)). By applying a **AM+volumed** two-factor model, a return of 81.8% can be obtained by the end of the period, which is 14.5% more than the one-factor model (Figure 5(b3)). Notably, the predicted return for the next three months increases to 40.9% with the new model (Figure 5(b4)).

## 7.2. Selection of Stock Portfolio

In the second scenario, we invited the fund manager to select a portfolio of stocks to invest from an initial stock pool in a given time period. The sector-based division was adopted. The manager analyzed every candidate sector, and combined the selected stocks in each sector into a final portfolio for investment. The year of 2015 was considered when the market experienced a spike with the maximal index climbing as high as 60% in the first six months. The yearly growth was 9.4% after the downturn in the second half of the year. It is thus an appropriate scenario to test whether our tool can improve factor investing even in rapidly changing markets.

The manager first studied the Chemical sector which had a similar trend to the market in 2015. The factor view in Figure 2(b) visualizes an initial collection of factors using the ELNetApp model. It can be seen in the factor view that many factors are quite stable over time. Different from the study on individual stocks, for portfolio construction, the manager would like to select both valuable factors and a set of stocks whose selected factors behave in a similarly stable and effective way. Our tool provides support with an interactive selection process. Users can click on each important factor in the factor list (Figure 2(d)) and scroll in the stock view (Figure 2(c)) to identify candidate stocks. By this process, the manager found factor **lagretn**, which had a moderate overall contribution to stock returns of the sector (Figure 2(d)) and the contribution was stable on several individual stocks. As shown in Figure 6(a), the selected stocks share similar patterns on **lagretn**. When the stock price started to

rise again from September (notice red/green bars in Figure 2(c)), the contribution of **lagretn** reversed to negative and stayed for few months, as shown by the connecting lines in Figure 6(a). It is found in Figure 6(b) that five stocks (red lines) sharing this pattern on **lagretn** receive excess returns above the sector average (blue line). The prediction performance in Figure 6(c) (grey dotted line) also surpasses the market average (blue dotted line).

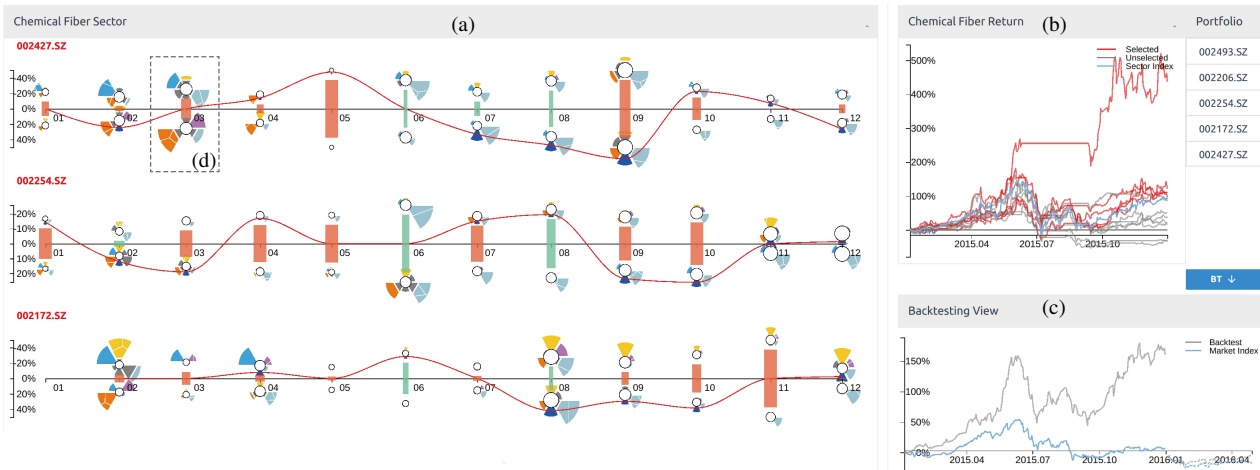
The manager proceeded to interactively examine all the five selected stocks. In the factor list view, he checked again all the candidate factors and found the factor **std.dvol**, which had a favorable pattern as the factor **lagretn** on all the selected stocks. After applying the two factors together (Figure 2(d)), the stock returns were re-computed and updated in Figure 2(e). One stock appeared to receive less return after adding the second factor, while all the other four stocks remained superior to the sector average. In the backtesting view (Figure 2(f)), the final portfolio of two factors and four stocks (red lines) turned out to be better than the single-factor portfolio (grey lines) in both the backtest performance in 2015 (solid lines) and an outlook in the first three months of 2016 (dotted lines).

In another study, we asked the manager to work in the Household Appliances sector in the same year of 2015. By repeating a similar process, he was able to build a two-factor model of **size+retmax**. Three stocks having stable contributions were selected. The portfolio returns in both 2015 and 2016 (outlook) are better than the market average. We omit screenshots due to the space limit.

By comparing the visual analysis result in the above two sectors, it is noticed that the Appliance sector is almost exclusively influenced by the transaction factors, while the Chemical sector is influenced by transaction factors and other types of factors such as momentum, value, and growth (Figure 2(d)). We hypothesize that because of the relationship of the Chemical industry and oil refinement, the Chemical sector is not only affected by transactions in the stock market but also linked to the price of crude oil, which is independently traded.

## 8. Discussion

We discuss some uncovered issues in iQUANT design by the order of the system pipeline.



**Figure 6:** Portfolio construction from the Chemical sector using the single factor (**lagretn**) in 2015: (a) **lagretn** factor contributed similarly to these three stocks; (b) selected high return stocks for portfolio; (c) backtesting results from the portfolio; (d) an example of Voronoi-based glyph.

**The sparse regression model and its flexibility.** In our system, users routinely start from an initial set of factors recommended by the build-in sparse regression model, which potentially limits the flexibility of factor investing by iQUANT. This issue is alleviated from three directions: in the factor list view (Figure 2(d)), users can manually select any set of factors as the starting point of analysis; the sparsity parameter of the regression model ( $\lambda$ ) can be tuned in the backend to offer more or less initial factors; finally, the iQUANT system is fully compatible with any other sparse models or a combination of feature selection and prediction models.

**The joint human-model approach and its advantage.** We have mentioned in previous sections, the model-only approach in factor selection suffers from the interpretability issue. Traders need to understand the factor before applied in practice. Also, a human is better in a heuristic search of the best combination within the vast factor space. More importantly, in the financial domain, there is no model that works in the long term. It will be critical to have human stand by models, and would be better to early adjust accordingly.

**The visual design and its learning curve.** Notice that stock traders may not have a strong background in visualization. Thus, we tried to strike a balance between the intuitiveness and expressiveness of visualization. On one hand, classical designs such as bar charts and line charts are employed in iQUANT, due to their intuitiveness and users' familiarity. On the other hand, our designs are tailored to the target users. For example in the factor view, green/red are used to indicate the decline/rise of stock prices, according to the convention of our target market. The effectiveness of iQUANT design is confirmed by the user study result and expert feedback. Most subjects found iQUANT easy to learn and understand.

**The user study and its design choices.** We evaluated iQUANT through a formal user study focusing on the performance of factor/stock selection. The study would have accounted for more stages of iQUANT by looking at the final stock returns, or the accuracy in predicting. We stayed in the scope of evaluating factors for two reasons. First, the stock return by an interactive factor investing system is determined in complex interactions among factor/stock selection, predictive model, and trading strategy. A higher

stock return in the study may or may not be directly correlated with a better factor selection design. Second, there is currently no commercial tool that supports end-to-end interactive factor investing. We compare iQUANT with a baseline design, which is not natively integrated with prediction models and trading strategies.

**The deployment of iQUANT and its extensibility.** The investment industry is highly closed as the winning secret can be too valuable to disclose. This closedness prohibits a more practical approach to directly optimize deployed tools. Instead, we developed a homegrown interactive factor investing pipeline and an end-to-end system. Because of the complexity of quantitative investment, we mainly focus on improving the factor selection stages in the pipeline, which is well demonstrated by the evaluation. As a next step, we plan to customize and integrate the factor selection technique in this work into the production system of our collaborators in the domain, in the hope of making real impacts on the industry.

## 9. Conclusion

This paper presents iQUANT, a visual analytics system that seamlessly integrates algorithmic and user-steered feature selection methods for interactive factor investing. In the analytics side, we explore the use of sparse regression models that have the merit of jointly optimizing factor selection and stock prediction processes. On the visualization side, we propose multiple coordinated visual designs to comprehensively illustrate importance metrics of each factor, including its positive/negative contribution to models, the factor stability, sensitivity, and complementarity. On top of the prediction model and visualization design, iQUANT develops an interactive framework that encloses factor refinement and portfolio construction in the same visual analytics loop. With iQUANT, traders can optimize their investment strategy and evaluate them in industry-standard backtesting. We demonstrate the value of iQUANT through one formal user study and two case studies, using a real stock market dataset recording the stock data in the past 30 years. On the selection of investment time and stock portfolio, our domain experts achieved excess returns from both individual stock and sector-based portfolios.

## References

- [Asn97] ASNESS C. S.: The interaction of value and momentum strategies. *Financial Analysts Journal* 53, 2 (1997), 29–36. doi:10.2469/faj.v53.n2.2069. 3
- [BS16] BANG H., SELVA D.: ifeed: Interactive feature extraction for engineering design. In *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (2016), American Society of Mechanical Engineers, pp. V007T06A037–V007T06A037. doi:10.1115/DETC2016-60077. 2
- [BW08] BYRON L., WATTENBERG M.: Stacked graphs - geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1245–1252. doi:10.1109/TVCG.2008.166. 6
- [CGK\*07] CHANG R., GHONIEM M., KOSARA R., RIBARSKY W., YANG J., SUMA E., ZIEMKIEWICZ C., KERN D., SUDJANTO A.: Wirevis: Visualization of categorical, time-varying data from financial transactions. In *2007 IEEE Symposium on Visual Analytics Science and Technology* (Oct 2007), pp. 155–162. doi:10.1109/VAST.2007.4389009. 2
- [CHL\*03] CSALLNER C., HANDTE M., LEHMANN O., STASKO J.: Fundexplorer: Supporting the diversification of mutual fund portfolios using context treemaps. In *IEEE Symposium on Information Visualization 2003* (Oct 2003), pp. 203–208. doi:10.1109/INFVIS.2003.1249027. 2
- [CLL16] CHENG K., LI J., LIU H.: Featureminer: A tool for interactive feature selection. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management* (New York, NY, USA, 2016), CIKM '16, ACM, pp. 2445–2448. URL: <http://doi.acm.org/10.1145/2983323.2983329>, doi:10.1145/2983323.2983329. 2
- [DBBS08] DAO H. T., BAZINET A. L., BERTHIER R., SHNEIDERMAN B.: Nasdaq velocity and forces: An interactive visualization of activity and change. *Journal of Universal Computer Science* 14, 9 (2008), 1391–1410. doi:10.3217/jucs-014-09-1391. 2
- [DML14] DUMAS M., MCGUFFIN M. J., LEMIEUX V. L.: Financevis.net—a visual survey of financial data visualizations. In *IEEE Conference on Information Visualization* (2014), vol. 2. 2
- [DS97] DECHOW P. M., SLOAN R. G.: Returns to contrarian investment strategies: Tests of naive expectations hypotheses. *Journal of Financial Economics* 43, 1 (1997), 3–27. 3
- [DvVH\*19] DINGEN D., VAN'T VEER M., HOUTHUIZEN P., MESTROM E. H. J., KORSTEN E. H., BOUWMAN A. R., VAN WIJK J.: Regressionexplorer: Interactive exploration of logistic regression models with subgroup analysis. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 246–255. doi:10.1109/TVCG.2018.2865043. 2
- [EASD\*19] EL-ASSADY M., SPERRLE F., DEUSSEN O., KEIM D., COLLINS C.: Visual analytics for topic model optimization based on user-steerable speculative execution. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 374–384. doi:10.1109/TVCG.2018.2864769. 2
- [FF92] FAMA E. F., FRENCH K. R.: The cross-section of expected stock returns. *The Journal of Finance* 47, 2 (1992), 427–465. doi:10.1111/j.1540-6261.1992.tb04398.x. 1, 2, 5
- [FF04] FAMA E. F., FRENCH K. R.: The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives* 18, 3 (2004), 25–46. doi:10.1257/0895330042162430. 1, 2
- [GE03] GUYON I., ELISSIEFF A.: An introduction to variable and feature selection. *The Journal of Machine Learning Research* 3 (2003), 1157–1182. URL: <http://dl.acm.org/citation.cfm?id=944919.944968>, doi:10.5555/944919.944968. 1
- [GK13] GLANTZ M., KISSELL R.: *Multi-Asset Risk Modeling: Techniques for a Global Economy in an Electronic and Algorithmic Trading Era*. Academic Press, 2013. 1
- [HHN00] HAVRE S., HETZLER B., NOWELL L.: Themeriver: visualizing theme changes over time. In *Proceedings of IEEE Symposium on Information Visualization 2000*. (Oct 2000), pp. 115–123. doi:10.1109/INFVIS.2000.885098. 6
- [JJ09] JOHANSSON S., JOHANSSON J.: Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 993–1000. doi:10.1109/TVCG.2009.153. 2
- [JS04] JINWOOK SEO, SHNEIDERMAN B.: A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections. In *IEEE Symposium on Information Visualization* (Oct 2004), pp. 65–72. doi:10.1109/INFVIS.2004.3. 2
- [KCA\*16] KO S., CHO I., AFZAL S., YAU C., CHAE J., MALIK A., BECK K., JANG Y., RIBARSKY W., EBERT D. S.: A survey on visual analysis approaches for financial data. *Computer Graphics Forum* 35, 3 (2016), 599–617. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12931>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12931>, doi:10.1111/cgf.12931. 2
- [KNS\*06] KEIM D. A., NIETZSCHMANN T., SCHELWIES N., SCHNEIDWIND J., SCHRECK T., ZIEGLER H.: A spectral visualization system for analyzing financial time series data. In *Proceedings of the Eighth Joint Eurographics / IEEE VGTC Conference on Visualization* (Aire-la-Ville, Switzerland, Switzerland, 2006), Eurographics Association, pp. 195–202. URL: <http://dx.doi.org/10.2312/VisSym/EuroVis06/195-202>, doi:10.2312/VisSym/EuroVis06/195-202. 2
- [KPB14] KRAUSE J., PERER A., BERTINI E.: Infuse: Interactive feature selection for predictive modeling of high dimensional data. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1614–1623. doi:10.1109/TVCG.2014.2346482. 2
- [KPN16] KRAUSE J., PERER A., NG K.: Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), ACM, pp. 5686–5697. URL: <http://doi.acm.org/10.1145/2858036.2858529>, doi:10.1145/2858036.2858529. 2
- [LCW\*17] LI J., CHENG K., WANG S., MORSTATTER F., TREVINO R. P., TANG J., LIU H.: Feature selection: A data perspective. *ACM Computing Surveys* 50, 6 (2017), 94:1–94:45. URL: <http://doi.acm.org/10.1145/3136625>, doi:10.1145/3136625. 2
- [LXL\*18] LIU S., XIAO J., LIU J., WANG X., WU J., ZHU J.: Visual diagnosis of tree boosting methods. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 163–173. doi:10.1109/TVCG.2017.2744378. 2
- [MAHG16] MCCOMBE M., ANG A., HOGAN K., GELFAND Y.: *The rise of factor investing*. BlackRock, 2016. 1
- [MBD\*11] MAY T., BANNACH A., DAVEY J., RUPPERT T., KOHLHAMMER J.: Guiding feature subset selection with an interactive visualization. In *Proceedings of 2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Oct 2011), pp. 111–120. doi:10.1109/VAST.2011.6102448. 2
- [MQB19] MING Y., QU H., BERTINI E.: Rulematrix: visualizing and understanding classifiers with rules. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 342–352. doi:10.1109/TVCG.2018.2864812. 2
- [NB04] NESBITT K. V., BARRASS S.: Finding trading patterns in stock market data. *IEEE Computer Graphics and Applications* 24, 5 (2004), 45–55. doi:10.1109/MCG.2004.28. 2
- [PLDSSV98] PORTA R. L., LOPEZ-DE SILANES F., SHLEIFER A., VISHNY R. W.: Law and finance. *Journal of Political Economy* 106, 6 (1998), 1113–1155. doi:10.1086/250042. 3
- [RSE09] RUDOLPH S., SAVIKHIN A., EBERT D. S.: Finvis: Applied visual analytics for personal financial planning. In *2009 IEEE Symposium*

- on *Visual Analytics Science and Technology* (Oct 2009), pp. 195–202. doi:10.1109/VAST.2009.5333920. 2
- [SB13] SORENSON E., BRATH R.: Financial visualization case study: Correlating financial timeseries and discrete events to support investment decisions. In *17th International Conference on Information Visualisation* (2013), pp. 232–238. doi:10.1109/IV.2013.31. 2, 8
- [SBTK08] SCHRECK T., BERNARD J., TEKUSOVA T., KOHLHAMMER J.: Visual cluster analysis of trajectory data with interactive kohonen maps. In *2008 IEEE Symposium on Visual Analytics Science and Technology* (Oct 2008), pp. 3–10. doi:10.1109/VAST.2008.4677350. 2
- [Šim03] ŠIMUNIĆ K.: Visualization of stock market charts. In *The 11th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision* (2003), vol. 3, pp. 129–132. 2
- [SLFE11] SAVIKHIN A., LAM H. C., FISHER B., EBERT D. S.: An experimental study of financial portfolio selection with visual analytics for decision support. In *44th Hawaii International Conference on System Sciences* (2011), pp. 1–10. doi:10.1109/HICSS.2011.54. 2
- [STKF07] SCHRECK T., TEKUSOVÁ T., KOHLHAMMER J., FELLNER D.: Trajectory-based visual analysis of large financial time series data. *ACM SIGKDD Explorations Newsletter* 9, 2 (2007), 30–37. doi:10.1145/1345448.1345454. 2
- [Teg99] TEGARDEN D. P.: Business information visualization. *Communications of the Association for Information Systems* 1, 4 (1999), 1–38. doi:10.17705/1CAIS.00104. 2
- [Tib96] TIBSHIRANI R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 1 (1996), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x. 4
- [TS08] TEKUSOVA T., SCHRECK T.: Visualizing time-dependent data in multivariate hierarchic plots - design and evaluation of an economic application. In *2008 12th International Conference Information Visualization* (July 2008), pp. 143–150. doi:10.1109/IV.2008.51. 2
- [Tsi18] White paper for the list of quantitative factors in a major stock market, 2018. URL: <http://mscf.pbcscf.tsinghua.edu.cn/uploadfile/2018/0316/20180316025702935.pdf>. 1, 5
- [UGH09] USAI M. G., GODDARD M. E., HAYES B. J.: Lasso with cross-validation for genomic selection. *Genetics Research* 91, 6 (2009), 427–436. doi:10.1017/S0016672309990334. 4
- [VV99] VAN WIJK J. J., VAN SELOW E. R.: Cluster and calendar based visualization of time series data. In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99)* (Oct 1999), pp. 4–9. doi:10.1109/INFVIS.1999.801851. 2
- [YSZ\*19] YUE X., SHU X., ZHU X., DU X., YU Z., PAPADOPOULOS D., LIU S.: Bitextract: Interactive visualization for extracting bitcoin exchange intelligence. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 162–171. doi:10.1109/TVCG.2018.2864814. 2
- [ZH05] ZOU H., HASTIE T.: Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 2 (2005), 301–320. doi:10.1111/j.1467-9868.2005.00503.x. 4
- [ZJGK10] ZIEGLER H., JENNY M., GRUSE T., KEIM D. A.: Visual market sector analysis for financial time series data. In *2010 IEEE Symposium on Visual Analytics Science and Technology* (Oct 2010), pp. 83–90. doi:10.1109/VAST.2010.5652530. 2
- [ZNK07] ZIEGLER H., NIETZSCHMANN T., KEIM D. A.: Relevance driven visualization of financial performance measures. In *Eurographics/IEEE TCVG Symposium on Visualization* (2007), pp. 19–26. doi:10.2312/VisSym/EuroVis07/019-026. 2
- [ZWLC19] ZHAO X., WU Y., LEE D. L., CUI W.: iforest: Interpreting random forests via visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 407–416. doi:10.1109/TVCG.2018.2864475. 2