Statistical Analysis of Backgrounds with Technology Stocks Graphing

Nathan Lang*
nathan.lang@marquette.edu
Marquette University
Milwaukee, Wisconsin

Raymond Adams
raymond.adams@marquette.edu
Marquette University
Milwaukee, Wisconsin

Matthew Leiser
matthew.leiser@marquette.edu
Marquette University
Milwaukee, Wisconsin

*All authors contributed equally to this research.

ABSTRACT

In this research paper, we analyzed how users interacted with our web application based upon technology stocks and charting/comparing them. We first split technology stocks into three categories: hardware, software, and big data. We also utilized two different plots, daily price change and daily percent change, as well as ranging stock datasets from April 1 2018 to April 1 2019 that could be parsed out. With these three main features, we began to test Specifically, we had Amazon mechanical Turk users take our survey. Our survey consisted of demographic data about them as well as their familiarity with stocks. Furthermore, we had them perform tasks on our web application to track the results which included number of clicks on page and time taken on a task. With the goal to relate somebody's background with their performance on our web application, we formed two two-way ANOVA's. One of the ANOVA's had the independent variables being stock familiarity, later defined in the paper, as well as chart type. While the other ANOVA had stock chart familiarity and education as its independent variables. Through the ANOVA's we were able to answer our research questions. We then finish the paper regarding our limitations and learning outcomes from this experiment.

KEYWORDS

Data analysis, Statistical Analysis, Stocks, Stock Charting, Crowd sourcing, Amazon Mechanical Turk

INTRODUCTION

Many successful businesses turn into publicly traded companies on the stock market. At time of writing the SP 500 Index is nearing an all time. The SP 500 represent the 500 largest U.S. publicly traded companies [1]. With that being said many people are interested in buying and selling stocks whether it be for an investment, retirement plan, vacation savings, etc. So, it is important to understand stock charting and the volatility of large stocks [2]. In recent years technology stocks have been some of the most rapid growing stocks. This is mainly due to the expansion of advancing technology such machine learning, artificial intelligence, and data science entering into non-technology based industries and companies [3, 4]. Therefore, the companies supplying these technologies are growing and driving more revenue to themselves leading to a higher stock price [5]. It is a must for people investing in stock markets to understand the comparisons of the main 500 stocks. It would be a very difficult and nearly impossible task to chart and compare 500 stocks manually [6], let alone 5 or 10 that somebody decided to focus on.

There are three main sectors that technology stocks are broken into: hardware, software, and big data [7]. Of course, there are many more breakdowns that can be made i.e. software broken into front-end, backend, etc. Furthermore, there are two popular methods in which stocks are charted, namely stock price over time and percent change of price over time [8]. With these main ideas our group decided to make a web application allowing users to pick a few technology stocks of the same subcategory (hardware, software, or big data), and chart them against one another. Moreover, we knew it was essential for users to see the comparisons on both types of charts, stock price and percent change. Lastly, we determined our web app should allow for varying time frames once charted, but since our web application was meant for longer term our design decision for time frame options included: 1 week, 1 month, 3 months, 6 months, 9 months, and 1 year. Our data ranged from April 1st, 2018 to April 1st, 2019. This was an arbitrary date range decided upon at the beginning of our project when we were collecting the historical data. One design decision that proved difficult was deciding which stocks to use for each subcategory of technology stocks. After reviewing a few top 10 lists across the internet as well as which stocks have a high market cap and which we feel many people heard of would be the best method. We have chose the following stocks:

(1) Hardware: Apple, IBM, Nvidia, AMD, Intel, Motorola Solutions, Microsoft, 3M, Qualcomm
(2) Software: Microsoft, Oracle, IBM, Synacor, VMware, Salesforce, Facebook
(3) Big Data: IBM, Oracle, SAP, Microsoft, Teradata Corp.

In Figure 1 you can see our web application with check lists that are shown on selection of the three categories, as previously mentioned. Figure 2 shows what a percent change chart would look like for the selected stocks. Our web application is hosted on github pages at https://nathanlang14.github.io/TechStocksGraphing/. As well as our github source code at https://github.com/Nathanlang14/

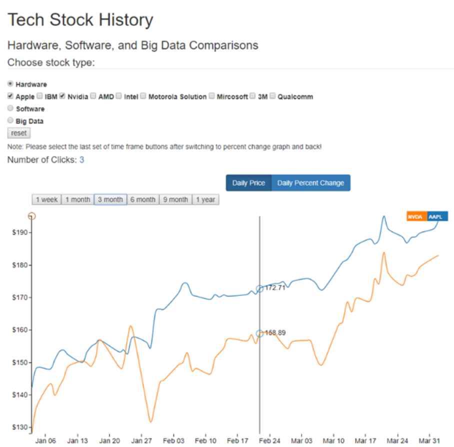
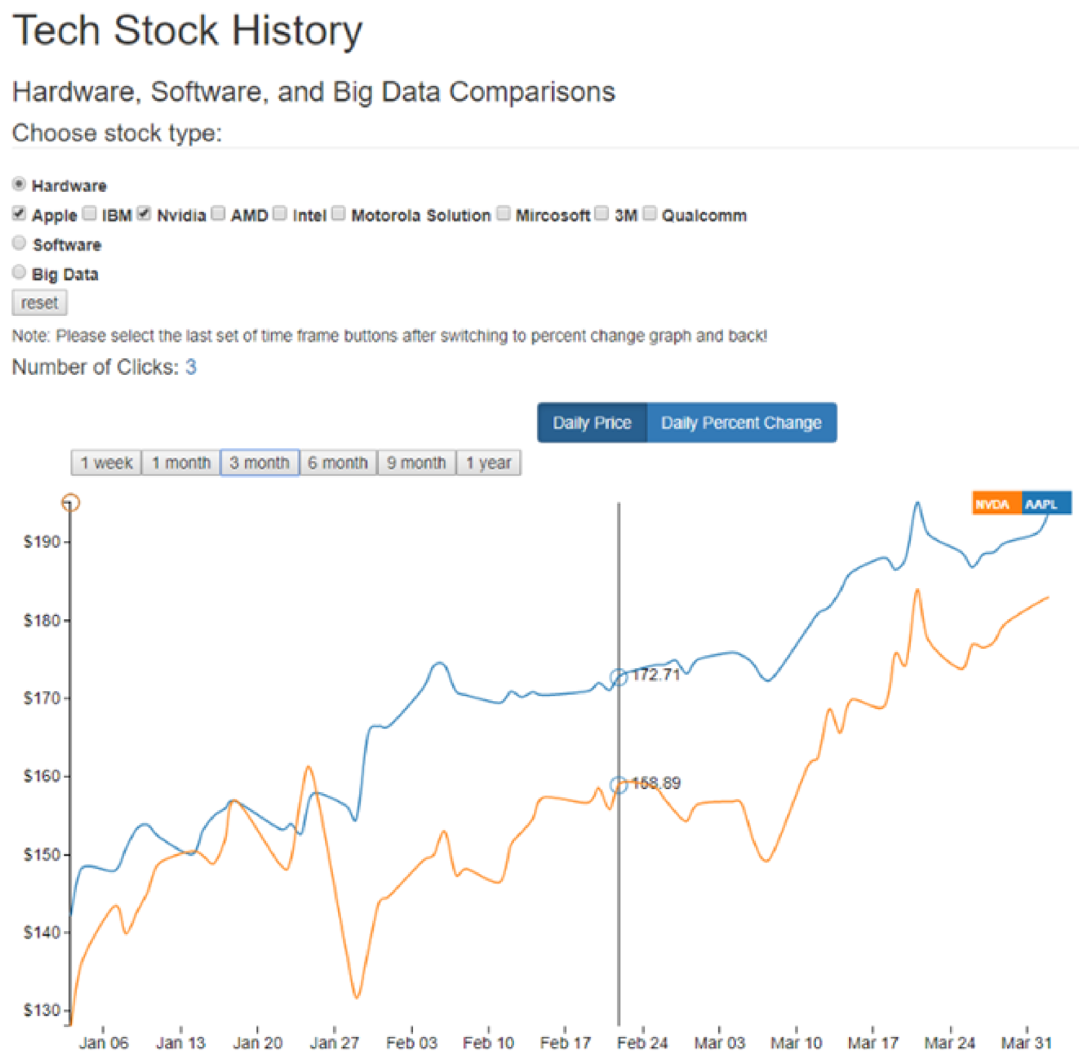
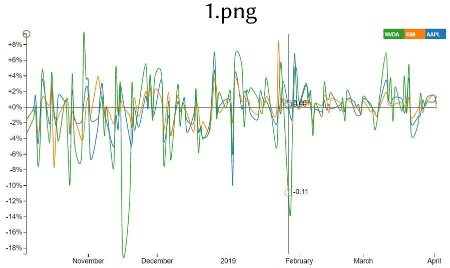


Figure 1: Tech Stock Web app

TechStocksGraphing. In our research, we utilized Amazon mechanical turk and Qualtrics to test our webapp and collect data to be able to answer our research questions. We also hosted our web application on github pages. Qualtrics was the site we used to record our survey data. First we asked mturk users a series of demographic questions regarding their demographics like gender, education, country, and job type. Then, we also asked the users their familiarity with stocks, if they trade stocks, follow stocks, and their familiarity with reading charts. After we got some background information about them, we then decided to utilize tasks to determine how users are operating our web application. Our six tasks were split into two types, three tasks regarding the price chart and the other three tasks regarding the percent change chart.


Tech Stock History
Hardware, Software, and Big Data Comparisons
Choose stock type:
Hardware
Apple IBM Nvidia AMD Intel Motorola Solution Mircosoft 3M Qualcomm
Software
Big Data
reset
Note: Please select the last set of time frame buttons after switching to percent change graph and back!
Number of Clicks: 3
Daily Price
Daily Percent Change
1 week
1 month
3 month
6 month
9 month
1 year
NVDA
AAPL
$190
$180
$170
$160
$150
$140
$130
172.71
158.89
Jan 06
Jan 13
Jan 20
Jan 27
Feb 03
Feb 10
Feb 17
Feb 24
Mar 03
Mar 10
Mar 17
Mar 24
Mar 31

1.png


November
December
2019
February
March
April

RESEARCH QUESTIONS

(1) (RQ1) How does somebody's familiarity determine which chart they spent more time on?
(2) (RQ2) How does the number of stocks chosen to plot while on the web application vary with somebody's familiarity with stocks and their education?

For RQ1, we utilized a qualtrics mechanism that recorded the first click on a page and the amount of seconds it took for somebody to submit a page. By taking the difference if these values, specifically submit time - first click time, we are able to determine how long it took for somebody to complete a task. Because we put each of our six tasks on separate page, we are able to determine task time. Furthermore, since we split the tasks for each type of chart, we are able to use that and the user's stock familiarity as our independent variables.

For RQ2, on our web application we kept track of how many total clicks a user did on our web application. This calculation was done by simply counting up every time a user changed the graph in someway whether that be. We then had a users stock chart familiarity and education as our independent variables. Both of our research questions were analyzed using a two-way anova. For our first research question, we found that there was in fact a statistical significance between the time it took for mturk participants to answer a task . Our second research question, we found that there was statistical significance with the total number of clicks on our web application and the interaction of our independent variables education and stock chart familiarity.

DATA METHODS

Research Question 1

The results were obtained through the use of two-way analysis of variance (ANOVA) statistical tests. A two-way ANOVA differs from a one-way ANOVA by a one-way ANOVA comparing the group mean variance of one independent variable to one dependent variable, while a two-way ANOVA compares the group mean variances of two independent variables to one dependent variable [9]. Two-way ANOVA is done to test the mean differences between groups that have been split into two independent variables [10]. The primary purpose is to test for an interaction between the independent variables and the dependent variable. Finding an interaction to be significant shows that the effect of one of the independent variables on the dependent variable is the same for all data point of the other independent variable.

To answer our research questions, we created a survey through Qualtrics. The survey asked participants to answer basic background questions and questions to determine their base familiarity with stock charts. Participants were then asked to use our stock tracker application to answer six questions based around manipulating the data shown on the chart. The questions were divided based on if the question referred to using the price change-based chart model or the percent change-based chart model. The topics were split evenly with three of the six questions asking about the price change chart and the other three asked about the data in the percent change chart.

The survey participants were found using Amazon Mechanical Turk (MTurk) system. Our batch asked for 129 participants to take our survey. Of those that responded, we kept 61 in total. Our data cleaning methods looked for participants who did not answer to the best of their abilities such as answering 10,000 when the maximum possible correct answer was 200. We also removed answers from participants that answered each question with the same answer as it provided poor data. By working with MTurk, we encountered some biases that could have ended up influencing the data. One such bias was that there was not an even distribution across the nationalities of the participants. The participant's nationality breakdown ended up being 56% from India, 41% from the United States, and 3.2 from Canada or Indonesia. Another bias came from MTurk's mechanics itself with most of the participants who took that survey seemingly rushed through the survey in an effort to finish as fast as possible possibly providing us with lower quality data.

Our first research question asked, how does somebody's familiarity determine which chart they spent more time on? This question was tested using a two-way ANOVA. For the test, the ANOVA was created to test three hypotheses:

(1) H0: There is no difference of the time spent on a page with a price-based stock chart vs. a page with a percent change-based stock chart. H1: There is a difference of the time spent on a page with a price-based stock chart vs a page with a percent change-based stock chart.
(2) H0: There is no difference in time spent on a page based on someone's familiarity with reading stock charts. H1: There is a difference in time spent on a page based on someone's familiarity with reading stock charts.
(3) H0: There is no interaction between a person's familiarity with reading stock charts and the use of a price-based chart vs. a percent change-based chart. H1: There is an interaction between a person's familiarity with reading stock charts and the use of a price-based chart vs. a percent change-based chart.

Should the first null hypothesis be rejected, then the data would show that there is a difference in the time it takes someone to read a price change-based stock chart vs. the time taken to read a percent change-based stock chart. The second null hypothesis being rejected would show that there is a difference between the time it takes a person to read a stock chart and their baseline familiarity with reading and understanding stock charts. The rejection of the third null hypothesis would mean that there is an interaction between a person's baseline familiarity of reading stock charts and whether the chart being read is price change-based or percent change-based in regards to time spent on a page. Should an interaction be found, then it would mean that the chart type and the familiarity of a person affect one another to influence the continuous variable (time). The data was displayed using a box plot graph.

Research Question 2

Participants in this study included 129 Amazon Mechanical Turk users. However, many users did not properly complete the tasks so we were forced to remove those participants. After elimination of unwanted data there were 61 suitable Amazon Mechanical Turk participants. All members of this study were volunteers.

Participants took the survey on Qualtrics. The first and second page of the survey displayed an informed consent form and instructions stating information about the procedure, purpose and benefits of participating. The third page displayed 10 demographic questions: age, gender, education, country, state, job type, "How familiar are you with the stock market?", "Have you traded stocks before?", "Do you follow stocks?", and "How familiar are you with reading stock charts?".

The majority of our participants were located in India or the United States. However, we also had participants from Canada and Indonesia. In this survey 56% of the participants were located in India, 41% were located in the United States, 1.6% were located in Canada, and 1.6% were located in Indonesia. The fact that the data was not evenly distributed across the countries poses a potential bias.

Participants were asked to state their highest level of education. This questions was asked to determine whether the level of education had an effect on the participants ability to read the stock chart. After analyzing the data it was determined that 18% of the participants had some other level of education, 54% had received their bachelor's degree, and 28% had received their master's degree.

Participants were asked to state how familiar they were with stock charts. This question ranged from "Not at all" to "Extremely" familiar. After evaluating the data it was determined that 5% answered "Not at all", 10% answered "Slightly", 34% answered "Moderately", 31% answered "Very", and 20% answered "Extremely".

The following pages contained a total of six tasks. Participants were asked to use a total of two charts, percent change and price. Three questions involved participants interacting with the percent change chart. The other three questions involved participants interacting with the price chart. For example, a participant was asked "What was the maximum price of Microsoft, Oracle, IBM, and Facebook over the last year".

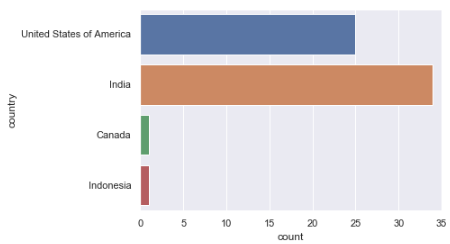


Figure 3: Count plot of the countries that the participants are from

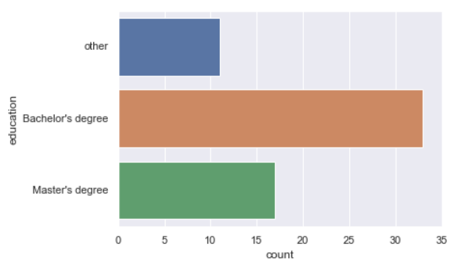
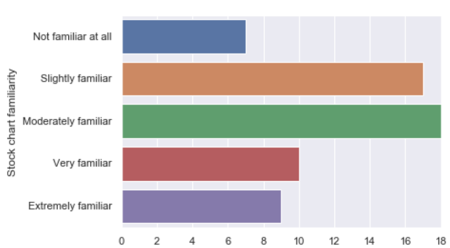


Figure 4: Count plot of highest education



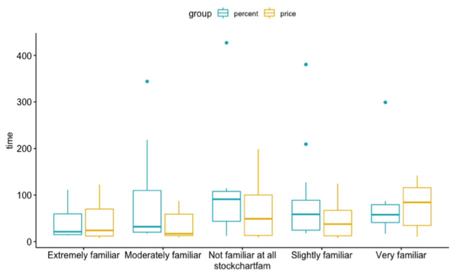


Figure 7: pending

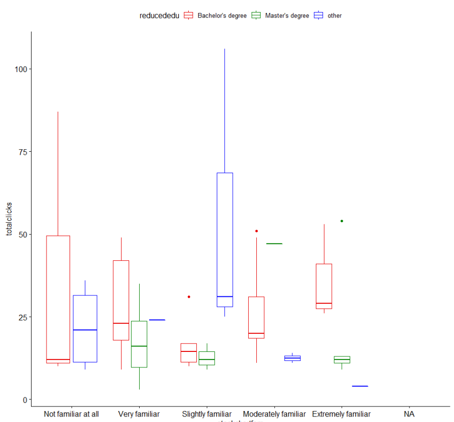


Figure 8: Box plot: x-axis is stockchartfam, y-axis is totalcliks

RESULTS

Research Question 1

Using R Studio, the data from the survey and the Qualtrics built in timing and click records, the data was plotted on a box plot as a representation of the data collected. The x-axis of the graph represents the familiarity level of the participant with labels, âĂIJExtremely familiarâĂİ, âĂIJModerately familiarâĂİ, âĂIJNot familiar at allâĂİ, âĂIJSlightly familiarâĂİ, and âĂIJVery familiarâĂİ. The y-axis represents the time spent on the survey questions in total which ranges from 0-400 seconds. The blue colored boxes represent the percent change stock chart data while the yellow boxes represent the price change stock chart data.

We ran a two-way ANOVA on the data using R Studio. The independent variables were the group (price chart vs. percent chart) and the participant's familiarity with reading stock charts. The dependent variable for this experiment was the total time a participant spent answering the survey questions from the first click to the last click. The p-value when the group was compared to the time taken was 0.0258. When comparing the participant's familiarity with reading stock charts against the time taken, the p-value was found to be 0.2400. When looking at the interaction between the group and the participant's familiarity, the p-value was 0.6439. The interpretations of the values found are discussed below.

Two-way ANOVA					
	df	Sum Sq	Mean Sq	F value	Pr(>F)
group	1	25751	25751	5.106	0.0258
stockchartfam	4	28160	7040	1.396	0.2400
group:stockchartfam	4	12658	3165	0.627	0.6439
Residuals	112	564869	5043		

Research Question 2

The data was gathered and manipulated in R Studio to create a box plot where the x-axis represents how familiar the participant is with stock charts and the y-axis represents the total number of clicks. The x-axis ranges from "Not familiar at all" to "Extremely familiar". The box plots were then grouped by education. Therefore, there were three categories: "Other", "Bachelor's degree", and "Master's degree".

The chart shows that a higher level of education usually lead to less clicks.

A two-way ANOVA was then ran on the data. The dependent variable was the total clicks and the significance level was set at 0.05. The p-value for education was 0.2284 which implies that the levels of education are not associated with total clicks. The p-value for stock chart familiarity was 0.9962 which implies that the levels of stock chart familiarity are not associated with total clicks. Lastly, the p-value for the interaction between education*stock chart familiarity was 0.0309. This indicates that

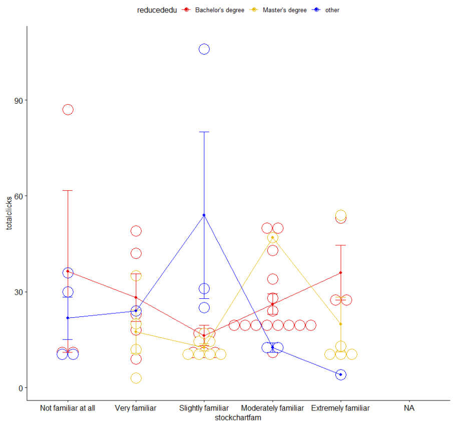


Figure 9: Interaction plot between stockchartfam education

the interaction between stock chart familiarity and education is significant for representing total clicks.

Two-way ANOVA					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Education	2	964	482.1	1.525	0.2284
Stock chart familiarity	4	56	13.9	0.044	0.9962
Education*stock chart familiarity	7	5459	779.9	2.467	0.0309
Residuals	46	14541	316.1		

These results show that education alone has no effect on the total number of clicks taken on the web application. Also, being familiar with stock charts has no impact on the number of clicks neither. However, having a higher level of education and being familiar with stock charts typically does lead to less clicks. Therefore, it should be noted that obtaining a higher level of education and being familiar with a certain task may lead to a better understanding of that task.

CONCLUSION

To summarize, technology has been continuously advancing which has lead to an increase in interest of technology stocks. The main technology stocks can be group into three subcategories: big data, hardware, and software. Through our design on our web application we were able to understand how a background of somebody has an effect on the way they utilized our application and went through our tasks. By employing a survey on Qualtrics and using Amazon mturk employees we were able to gather the desired data. By having users complete three tasks with the stock price chart and three tasks with the percent change chart, we were able to track how long users took on tasks related to each chart. We used a Qualtrics feature to calculate the time spent on each page, representing total time on task. By summing the three tasks for each chart, we developed our two-way anova null hypotheses.

For RQ1, we were able to determine that there was a statistical significance between the means for the type of graph to impact the time. This lead to the conclusion that there is in fact an increased difficulty in reading percent change charts. We believe this is because traditionally graphs are represented as total price over time, inherently this meant that percent change charts are more complex and foreign to users.

As For RQ2, we again were interested in how many times a user clicked on our web application as we correlated that to understanding the task at hand. The p-value for the interaction between Education and Stock chart familiarity is 0.0309. This indicates that the interaction between stock chart familiarity and education is significant for representing total clicks. Furthermore, we were able to conclude that if user had some background in a topic, then they are likely to be better at understanding the task at hand. Of course, this makes sense. Now what is interesting is that it seemed if somebody had a background in the topic and a higher level education: bachelors and then masters they were more also more likely to understand the task better. Paring these together is what we found.

Some limitations we faced include many of our Amazon mturk users were scams or the Qualtrics answers were unbelievable and were deleted. We learned that Amazon mturk can have biases, and also be plain out wrong. Furthermore, this was all members in our groups first time using d3 and due to lack of time and knowledge we had to resort to a simpler web application. Again, our group has never used Amazon mturk before or done a survey so our settings on Amazon mturk could have been optimized better to adjust for better results. Our survey tasks could have been more clear, as we learned through feedback that some tasks could be interpreted differently that intended. Lastly, since we had minimal funding we were only able to run one small batch and many resulted in lack of quality data. Overall, this project was successful in learning about the front-end web development stack and A/B testing.

REFERENCES

[1] Will Kenton and Chris B Murphy. (2019, April 19). Understanding SP 500 Index âĂŞ Standard Poor's 500 Index. Retrieved from https://www.investopedia.com/terms/s/sp500.asp
[2] Hamilton, J. D., Lin, G. (1996). Stock market volatility and the business cycle. Journal of applied econometrics, 11(5), 573-593.
[3] Su, J. S., Zhang, B. F., Xu, X. (2006). Advances in machine learning based text categorization. Ruan Jian Xue Bao(Journal of Software), 17(9), 1848-1859.
[4] Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.
[5] Akhigbe, A., Larson, S. J., Madura, J. (2002). Market underreaction and overreaction of technology stocks. The Journal of Psychology and Financial Markets, 3(3), 141-151.
[6] Chavarnakul, T., Enke, D. (2008). Intelligent technical analysis based equivolume charting for stock trading using neural networks. Expert Systems with Applications, 34(2), 1004-1017.
[7] Hobijn, B., Jovanovic, B. (2001). The information-technology revolution and the stock market: Evidence. American Economic Review, 91(5), 1203-1220.
[8] Shiller, R. J. (1990). Speculative prices and popular models. Journal of Economic perspectives, 4(2), 55-65.
[9] Wilcox, R. R. (1989). Adjusting for unequal variances when comparing means in one-way and two-way fixed effects ANOVA models. Journal of Educational Statistics, 14(3), 269-278.
[10] Kremer, E. (1982). IBNR-claims and the two-way model of ANOVA. Scandinavian Actuarial Journal, 1982(1), 47-55.