Application of the Algorithm for Analyzing Financial Instruments Based on Correlation Coefficient

Laura Karimova[1], and Sabina Rakhmetulayeva[1]

[1] International Information Technology University, Manas St. 34/1, Almaty, 050000, Kazakhstan

Abstract

Stock prices depend not exclusively on the demand of investors or the business of the company itself. Individual stocks frequently follow the prices of different stocks or stock indices. In this paper, a model has been developed that determines the connection between the stock prices of various assets. Stock costs of two assets from a similar area as a rule have a strong positive correlativity, whereas there are times when stocks have a negative influence on another stock or have absolutely no linear relationship with another stock. To show this dependence, the principle of correlation is used. The article presumes that correlation shows the exact relationship within assets and can become the basis for price forecasting depending on the change in the price of another asset.

Keywords

Cross market analysis, algorithms, stock prices, correlation coefficient, financial instruments

1. Introduction

Since the stock market is characterized not only by high returns but also by high risks, investors can choose decentralized portfolios to reduce non-systematic risks. Portfolio theory is that a diversified portfolio with low relativity between assets can effectively mitigate unsystematic risks.

Many scientists today prefer to analyze data on investor behavior, corporate governance, and shareholder information. However, a large amount of stock research is currently being done in terms of stock price correlations.

Thus, the relevance of the topic of this work is due to the problem of finding a more accurate trend based on the behavior of asset prices from one or adjacent sectors. Traditionally, we have relied on manual analysis, which is becoming impractical as data volumes grow exponentially. New big data analytics technologies help analyze huge amounts of data and extract useful patterns and relationships. From the point of view of various factors influencing the relativity of stocks, this paper examines the relationship between the similarity of different assets and different degrees of correlation in stock prices and then analyzes the similarities in the structures of correlations under the strongest correlation scenarios. Thus, the purpose of the study is to provide investors with broad information to reduce the complexity of developing their portfolio strategies.

2. Literature review

Research on determining the relationship between financial instruments, indices and other factors has been ongoing over the past few decades. During this time, scientists and researchers from all over the world have written a huge number of works. This article will review the most recent work and their results.

Proceedings of the 7th International Conference on Digital Technologies in Education, Science and Industry (DTESI 2022), October 20–21, 2022, Almaty, Kazakhstan
EMAIL: laurakarimova20@gmail.com (Laura Karimova); ssrakhmetulayeva@gmail.com (Sabina Rakhmetulayeva)
ORCID: 0000-0002-8144-3975 (Laura Karimova); 0000-0003-4678-7964 (Sabina Rakhmetulayeva)

The work of Ben-Salha and Mokni (2022) is a consideration of the relationship within oil price-stock market using the detrended cross-correlation analysis (DCCA) with a generalization to the quantile structure. Thus, four different scenarios depending on market conditions were considered. The results of this work are extremely useful for investors, as they indicate which assets are best combined in a portfolio under certain market conditions.

Zhao and Wang (2022), in turn, determined the relationship between different asset classes and stocks using the DCC-GARCH model, which is an excellent fixer for correlation between assets in order to reveal the importance of economic policy uncertainty (EPU) and monetary policy uncertainty (MPU) USA and China. The article can also serve investors as well as governments.

The impact of oil price hikes on stock market returns was explored by Maghyereh and Abdoh (2022). The authors used vector autoregression (VAR) and quantile cross-spectrum (QS) methods to avoid non-linearity and discontinuities. It was found that different jumps affect the economies of countries in different ways.

A study on how the coronavirus affected the stock markets was presented by Dong et al. in 2021 using the dynamic model averaging approach. The authors come to the conclusion that the market situation has changed a lot after the coronavirus outbreak and propose measures to prevent such problems for countries in the future.

Other researchers have been working to identify the connection between China's natural resources and the stock market using the novel ARDL model (Zhang, 2022). As a result, it was confirmed that the impact of natural resources on the market is significant, which cannot be said to be the opposite.

The following researcher in his work developed a model characterizing the correlation between exchange rates and stocks (Ding, 2021). In the article, the author comes to the conclusion that the relationship between these elements is determined by the close dependence of share prices on ordinary shares.

Wang et al. (2019) analyzed various factors affecting the state of the stock market, as well as their relationship with each other. The result of the study was to determine the threshold at which the correlations are the strongest.

The article by Liu et al. (2022) describes the dynamic evolution of stock correlations, the factors influencing it, and in your article a model is built with an emphasis on new energy. According to the results, it can be concluded that such analyzes can help investors optimize their portfolios.

China stock market forecasting was done in Zhang et al. (2020) to determine the effectiveness of exchange rate information for forecasting. The authors determined that the effectiveness of this kind of data is comparable to other popular sources.

In the following study, the changes in four known indicators of market liquidity and their correlation were analyzed in the market where HFT was involved and not. As a result of the study, it was concluded that market liquidity is provided by HFT transactions, and liquidity itself can be measured by the speed of execution (Yagi et al., 2020).

Chavan and Patil (2013) in their work worked on finding input parameters to ensure the highest accuracy of forecasts. The authors concluded that basically all machine learning methods are used to predict a certain set stock price based on basic variables.

Enke and Thawornwong (2005), using machine learning, or rather information retrieval techniques, developed a model to identify relationships in financial forecasts. The results of the authors' work show that, based on rating models, such strategies have the ability to bring more profit.

Chong et al. (2017) used deep learning networks as they do not depend on prior knowledge of predictors. In their work, the authors cited a number of advantages and disadvantages of using this method for analyzing and predicting the stock market.

The next group of researchers developed a new system based on neural networks and an autoregressive moving average (ARMA) algorithm that predicts future stock data based on current stocks. Based on this, the output data of the system is the closest to real (Navale et al., 2016).

Jasic and Wood in 2004 developed a neural network to predict daily stock market returns. Performance was evaluated with a linear regression reference model, resulting in an improvement in prediction.

The work of Jiang (2021) contains relevant information about the latest progress in this area, especially for the last three years. The author analyzed research works where for stock market prediction deep learning models were used. One of the most important points is that the author also analyzed and compared the implementation and effectiveness of different models. It was found that in almost all works, supervised learning forecasting models were used.

Kumar et al. (2021) in their work summarized research on stock prediction based on machine learning models using narrative synthesis and vote-counting methods. It was revealed which statistical and machine learning methods were used mostly in the studied. Besides, a small number of researchers have used hybrid ML approaches to increase the accuracy of their predictions.

In the study of Liu et al. (2021) the specific data from two Chinese stock companies was taken, and for prediction a basic deep neural networks method was chosen. In addition, stock price charts were treated as images. After testing the deep learning model and the single-layer model for predicting movement, the authors conclude that the first method works much more accurately in the case of the stock market.

In a study of Chhajer et al. (2022) an overview of machine learning and artificial intelligence in the stock market forecasting problem is provided. Such ML technologies as support vector machines, artificial neural networks and long-term memory were considered and critically analyzed. One of the key points of this article is that the authors raised the question of how machine learning can change the dynamics of investing.

In addition to the support vector machines and classification methodologies mentioned above, Akhtar et al. (2022) also used the R - forest classifier in their work. The Kaggle dataset was used in this study given its size and security. Through testing, the authors measured the accuracy of various algorithms, and concluded that according to their results, the most suitable algorithm for predicting the random wooded area algorithm with 80.8% accuracy.

3. Analysis of financial instruments and indices

As a rule, for designing an effective strategy of investment, as well as choosing means of stocks for an investment portfolio, a correlation measure value can greatly help an investor. This rate can be operated to find stocks from different areas that tend to move in tandem, or vice versa, hedging stocks, in order to ensure against the dangers of price changes so that if one stock fails, the other is likely to make a profit. Also, there's a third behavior where stocks have a measure value of 0. This case doesn't mean that stocks never behave the identical way; rather, it'll mean that the stocks under study can move both in a different way and together, making them changeable. In spite of that, it could also mean that numerous uncorrelated stocks are less likely to decline at the same time.

Diversification of an investment portfolio is a technique of allocating funds in a portfolio between different groups of assets, taking into account projected returns, in order to reduce threat. In other words, diversification helps to reduce the threat of losing earnings when prices of some stocks fall, and gain benefits from different other instruments. Therefore, opting various stocks with different degrees and trends of correlation values is one of the most effective and considerably used diversification strategies. The result is a portfolio with an overall upward trend because at any given time at least one security should be in good shape, indeed if the others fail.

In most cases, the correlativity between stocks can be evident. For example, when comparing two stocks belonging to the same area, it is most likely that stock prices will have one change trend and respond in the same way to market conditions. On the contrary, the correlation may not be as easy to describe in a portfolio if the investor owns shares of joint or exchange finance.

Holding stocks with negative correlation rates is one strategy investors should look into; in other words, this strategy is called" hedging." Hedging balances stocks with positive correlations in your portfolio to more manage dangers.

A sample can be given in real estate and stocks. grounded on historical data, they had a very low correlation with each other. numerous investors use bonds to balance their investment portfolios and manage threats, as bonds also tend to have a weak relationship with the stock market. However, portfolio hedging also has the disadvantage that it can potentially impact investors' return on investment over market cycles. So, when a single stock or investment makes a solid profit, the negatively correlated stocks you bought as insurance can cut your gains down significantly.

Finding the relationship between the activity of two stocks or their indices is an important step in modeling an investment strategy. However, since the ratio is just a number indicating the degree to which two stocks behave in an identical manner, the correlation ratio alone cannot affect the stock market in any way. Analysts typically use the value of correlation coefficient to make calculations of future values about how a stock will perform based on the historical data of another security with which the relationship has been noted.

4. Correlation coefficient

The correlation coefficient is widely used in data analysis to select variables in analytical models and identify the most significant features in terms of the problem being solved. In mathematical statistics, the correlativity is an indicator that characterizes the degree of a statistical connection among several components. The value of this indicator can vary from -1 to 1 (Figure 1). In this case, any two stocks with a measurement of 1 are called "perfectly" correlated. This connection means that when one stock rises three points, another stock does the same thing at the same time. A correlation with a value of -1 is a "strong" negative correlation, meaning that when one stock rises three points, the other loses three points. Such cases are rare enough in the stock market that ideal correlations are almost entirely theoretical.

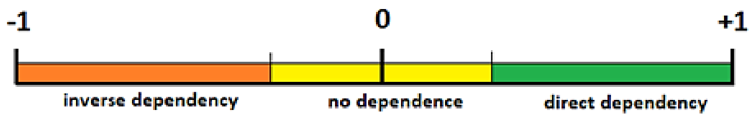


Figure 1: Correlation coefficient value limits

Correlation values can be divided into 3 groups:

1. Negative linear correlation (Figure 2) has a value among -1 and 0. With a negative correlation, an increase (or decrease) in the prices of one asset contributes to a regular decrease (or increase) in another asset.

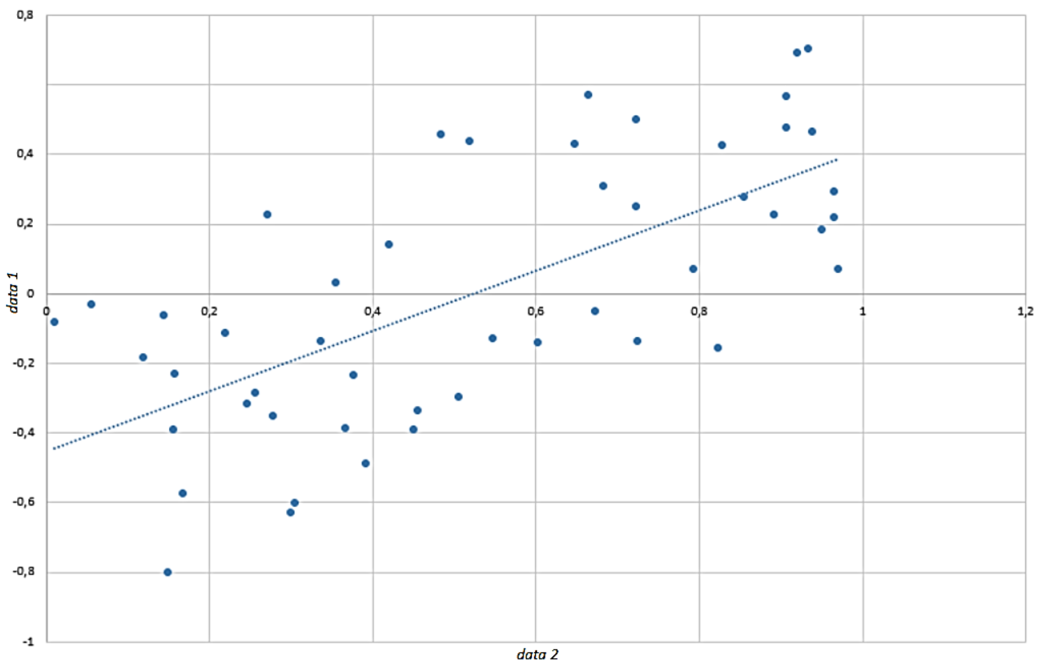


Figure 2: Example of a negative straight-line correlation

2. Positive linear correlation (Figure 3) has a value among 0 and 1. With a positive correlation, a reduction (or expansion) in one asset leads to a regular expansion (or reduction) in another asset.

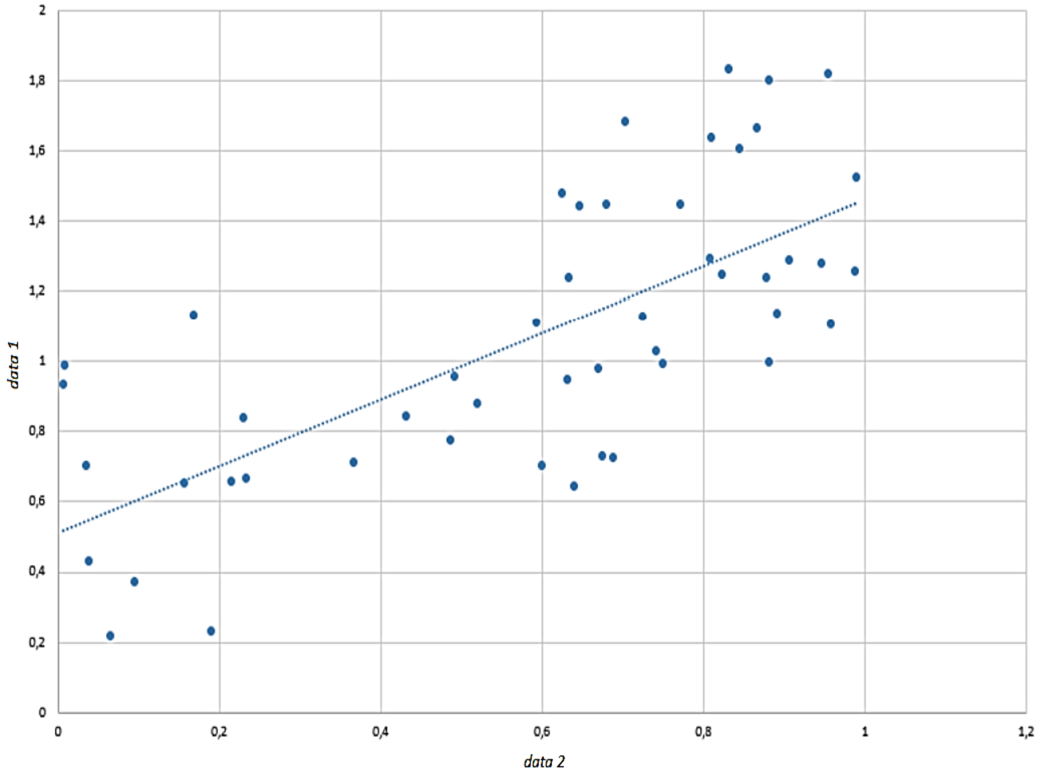


Figure 3: Example of a positive linear correlation

3. The absence of correlation (Figure 4) is applicable only if the value of the correlation coefficient is 0. In this case, the measurement indicates a non-linear relationship between the compared assets.

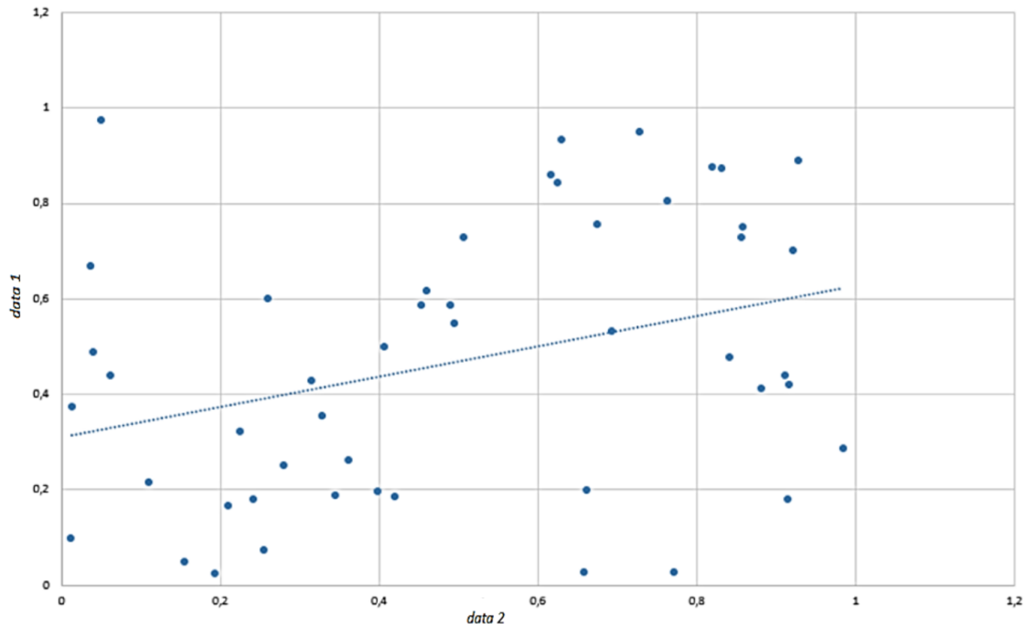


Figure 4: Example of the lack of correlation

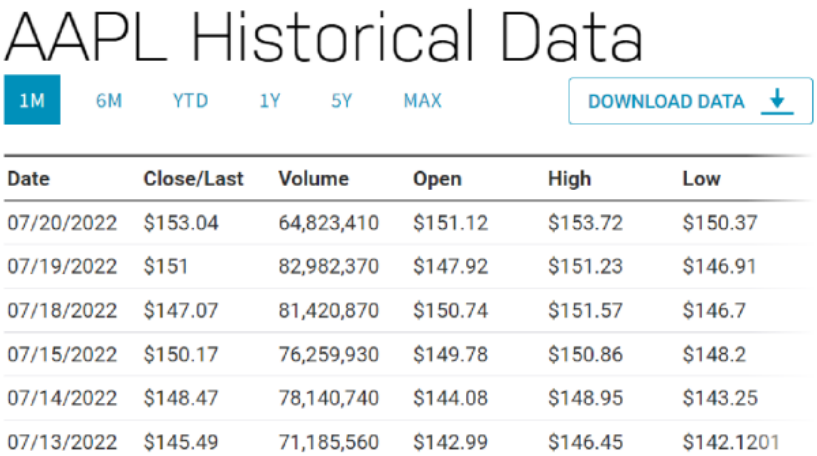
Also, the table below (Table 1) shows the detailed ranges of values and the degree of relationship.

Table 1
Interpretation of the values of the correlation coefficient

Coefficient value	Relationship between variables
[-1; -0,9)	Very strongly reversed
[-0,9; -0,7)	Strong reverse
[-0,7; -0,5)	Middle reverse
[-0,5; -0,2)	Weak inverse
[-0,2; 0]	Very weak reverse
0	No connection
(0; 0,2]	Very weak straight
(0,2; 0,5]	Weak straight
(0,5; 0,7]	Middle straight
(0,7; 0,9]	Strong straight
(0,9; 1]	Very strong straight

To develop the model, historical data on stocks were taken from the National Association of Securities Dealers Automated Quotation (NASDAQ) — the Quotation Service of the National Association of Securities Dealers. The exchange specializes in shares of high-tech companies (manufacturing of electronics, software, etc.). The NASDAQ is one of the major US stock exchanges, along with the NYSE.

The site provides information on historical stock price data for more than 3,200 companies. There are options for uploading data for a different period of time: 1 month, 6 months, 1 year, 5 years, and the maximum period is 8 years. In the table, you can find data on the date, price at the close of the day, volume, price at the time of the opening of the day, the highest and lowest prices for the day (Figure 5).


AAPL Historical Data

1M 6M YTD 1Y 5Y MAX DOWNLOAD DATA

Date	Close/Last	Volume	Open	High	Low
07/20/2022	$153.04	64,823,410	$151.12	$153.72	$150.37
07/19/2022	$151	82,982,370	$147.92	$151.23	$146.91
07/18/2022	$147.07	81,420,870	$150.74	$151.57	$146.7
07/15/2022	$150.17	76,259,930	$149.78	$150.86	$148.2
07/14/2022	$148.47	78,140,740	$144.08	$148.95	$143.25
07/13/2022	$145.49	71,185,560	$142.99	$146.45	$142.1201

Figure 5: Calculation of the coefficient of correlations

To perform the calculations, historical data for the period 15.07.2012 - 15.07.2022 for AAPL shares (Apple shares) and MSFT (Microsoft Corporation shares) were taken. Both companies are known to be in the information technology sector.

To write the coefficient calculation algorithm, the necessary libraries were installed: pandas, numpy, csv, scipy and matplotlib.

As a result of calculating the coefficient, we get the result r = 0.98338641 (Figure 6).

```
In [78]:  r = np.corrcoef(array_AAPL_new, array_MSFT_new)
          r

Out[78]: array([[1.        , 0.98338641],
                [0.98338641, 1.        ]])
```

Figure 6: Calculation of the coefficient of correlations

To build a visualization of the scatterplot, the matplotlib library and its functions were used (Figure 7).

```
# Visualization of Correlation
```

```
In [88]:  import matplotlib.pyplot as plt


          plt.style.use('correlation_plot')

          slope, intercept, r, p, stderr = scipy.stats.linregress(array_AAPL_new, array_MSFT_new)

          line = f'Regression line: y={intercept:.2f}+{slope:.2f}x, r={r:.2f}'
          line

Out[88]: 'Regression line: y=4.79+1.89x, r=0.98'
```

```
In [90]:  fig, ax = plt.subplots()
          ax.plot(array_AAPL_new, array_MSFT_new, linewidth=0, marker='s', label='Data points')
          ax.plot(array_AAPL_new, intercept + slope * array_AAPL_new, label=line)
          ax.set_xlabel('x')
          ax.set_ylabel('y')
          ax.legend(facecolor='white')
          plt.show()
```

Figure 7: Building a visualization

As a result, a diagram was built, according to the results of which one can observe a strong positive straight-line correlation of assets (Figure 8).

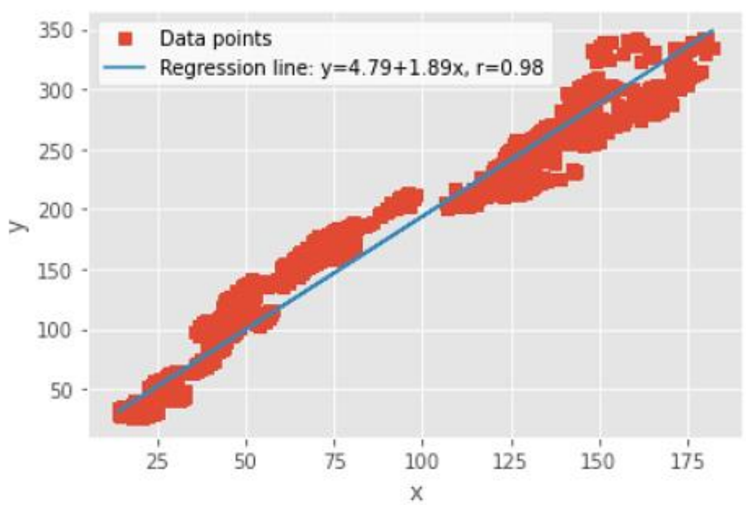


Figure 8: Scatterplot diagram

For the second test, AAPL (Apple stock) and AAL (American Airlines stock) were taken. As a result of finding the coefficient, we get the value (-0.76478884), which shows us a negative straight-line relationship among the two assets (Figure 9).

```
In [102]:   1 r = np.corrcoef(array_AAPL_new, array_AAL_new)
            2 r

Out[102]: array([[ 1.        , -0.76478884],
                 [-0.76478884,  1.        ]])
```

Figure 9: Calculation of the coefficient of correlations (2)

We build a scatterplot for assets with a negative correlation (Figure 10).

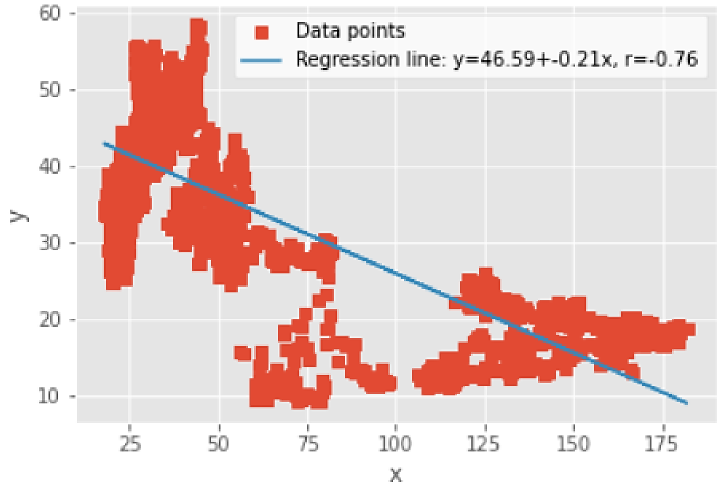


Figure 10: Scatterplot diagram (2)

For the last test, AAPL (Apple stock) and PPC (Pilgrim's Pride Corporation stock) were analyzed. Calculating coefficient, we get the value (0.08994049). Thus, we made sure that the stocks of the grocery retailer have no linear relationship with the stocks of Apple (Figure 11).

```
In [30]:   1 r = np.corrcoef(array_AAPL_new, array_PPC_new)
           2 r

Out[30]: array([[1.        , 0.08994049],
                [0.08994049, 1.        ]])
```

Figure 11: Calculation of the coefficient of correlations (3)

For a more illustrative example, a graph was built, where it is clear that the growth of one stock does not affect the change in another stock. Thus, it can be seen that the stocks change price without depending on each other (Figure 12).

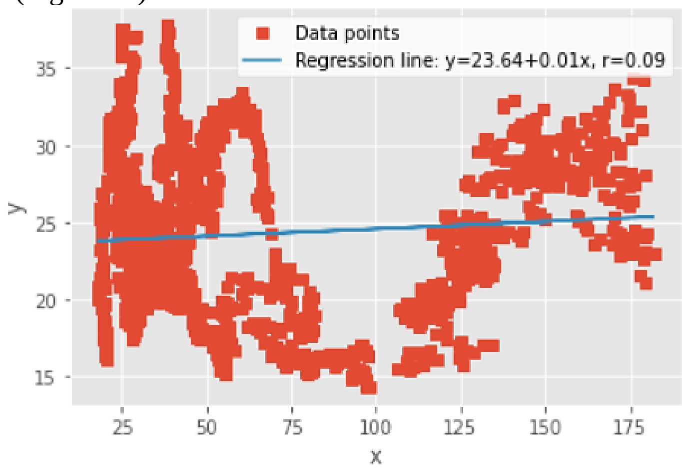


Figure 12: Scatterplot diagram (3)

Source code and all datasets used in the algorithm can be found at the link in github - https://github.com/LauraKarimova/correlation_coefficient.git.

5. Conclusion

The grade of association depends on a high level of correlation ratio. In particular, a coefficient of 0.2 between values is considered to have a correlation, but most likely not essential. According to the observations of various research in different areas, statistical general trends are observed, since a coefficient of less than 0.8 does not matter in the presence of higher indicators. However, a coefficient above 0.9 indicates a very powerful association. The financial industry shows best practice in the use of variable correlation, especially in investment research. For instance, association is valuable in defining how properly a joint fund is distinguished with its base index or different asset classes. At the same time, low data correlation rates can also be useful in a given field, where a low ratio leads to increased asset diversification in a joint fund. Weakly correlated assets are mainly used for portfolio hedging, resulting in reduced market risk due to price volatility. Given approach efficiently decreases losses, since the correlation found that all the negative sides are in the same behavior. Correlation statistics can determine the moment of deviation of the association between variables, which allows investors to quickly respond to market changes. For example, the situation in the banking sector, when there is a change in the price of a bank share and interest rate. Bank stocks commonly have high correlation with interest rates, because lending rates are calculated based on market interest rates. Based on this, we can conclude that the bank has left the field of standard behavior and it is necessary to study the correlation indicators of other banks. In the case when the correlations in other banks are identical, it can be concluded that the decrease or increase in the bank's shares is not related to interest rates. Otherwise, the bank under study has a fundamental problem in the banking system.

6. References

[1] W. Jiang, Applications of deep learning in stock market prediction: Recent progress, Expert Systems with Applications 184 (2021) 115537.
[2] D. Kumar. P.K. Sarangi, R. Verma, A systematic review of stock market prediction using machine learning and statistical techniques, in: Materials Today: Proceedings, 2021.
[3] Q. Liu, Z. Tao, Y. Tse, C. Wang, Stock market prediction with deep learning: The case of China, Finance Research Letters (2021) 102209.
[4] P. Chhajer, M. Shah, A. Kshirsagar, The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction, Decision Analytics Journal 2 (2022) 100015. URL: https://doi.org/10.1016/j.dajour.2021.100015.
[5] M. M. Akhtar, A. S. Zamani, S. Khan, A. S. Shatat, S. Dilshad, F. Samdani, Stock market prediction based on statistical data using machine learning algorithms, Journal of King Saud University – Science 34.4 (2022) 101940. URL: https://doi.org/10.1016/j.jksus.2022.101940.
[6] O. Ben-Salha, K. Mokni, Detrended cross-correlation analysis in quantiles between oil price and the US stock market, Energy 242 (2022) 122918. URL: https://doi.org/10.1016/j.energy.2021.122918.
[7] W. Zhao, Y.D. Wang, On the time-varying correlations between oil-, gold-, and stock markets: The heterogeneous roles of policy uncertainty in the US and China, Petroleum Science 19.3 (2022) 1420–1432. URL: https://doi.org/10.1016/j.petsci.2021.11.015.
[8] A. Maghyereh, H. Abdoh, Extreme dependence between structural oil shocks and stock markets in GCC countries, Resources Policy 76 (2022) 102626. URL: https://doi.org/10.1016/j.resourpol.2022.102626.
[9] X. Dong, L. Song, S.M. Yoon, How have the dependence structures between stock markets and economic factors changed during the COVID-19 pandemic? The North American Journal of Economics and Finance 58 (2021) 101546. URL: https://doi.org/10.1016/j.najef.2021.101546.
[10] Y. M. Zhang, Influence of stock market factors on the natural resources dependence for environmental change: Evidence from China, Resources Policy 77 (2022) 102711. URL: https://doi.org/10.1016/j.resourpol.2022.102711.

[11] L. Ding, Conditional correlation between exchange rates and stock prices, The Quarterly Review of Economics and Finance 80 (2021) 452–463. URL: https://doi.org/10.1016/j.qref.2021.02.004.
[12] Y. Wang, H. Li, J. Guan, N. Liu, Similarities between stock price correlation networks and co-main product networks: Threshold scenarios, Physica A: Statistical Mechanics and Its Applications 516 (2019) 66–77. URL: https://doi.org/10.1016/j.physa.2018.09.154.
[13] W.Liu, Q. Ma, X. Liu, Research on the dynamic evolution and its influencing factors of stock correlation network in the Chinese new energy market, Finance Research Letters 45 (2022) 102138. URL: https://doi.org/10.1016/j.frl.2021.102138.
[14] J. Zhang, Y.H. Shao, L.W. Huang, J.Y. Teng, Y.T. Zhao, Z.K. Yang, X.Y. Li, Can the Exchange Rate Be Used to Predict the Shanghai Composite Index? IEEE Access 8 (2020) 2188–2199. URL: https://doi.org/10.1109/access.2019.2962221.
[15] I. Yagi, Y. Masuda, T. Mizuta Analysis of the Impact of High-Frequency Trading on Artificial Market Liquidity, IEEE Transactions on Computational Social Systems 7.6 (2020) 1324–1334. URL: https://doi.org/10.1109/tcss.2020.3019352.
[16] P. S. Chavan, S. T. Patil, Parameters for stock market prediction, International Journal of Computer Technology and Applications 4.2 (2013) 337.
[17] D. Enke, S. Thawornwong, The use of data mining and neural networks for forecasting stock market returns, Expert Systems with applications 29.4 (2005) 927-940.
[18] E. Chong, C. Han, F.C. Park, Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies, Expert Systems with Applications 83 (2017) 187-205.
[19] G. S. Navale, N. Dudhwala, K. Jadhav, P. Gabda, B.K. Vihangam, Prediction of Stock Market using Data Mining and Artificial Intelligence, International Journal of Computer Applications 134.12 (2016) 9–11.
[20] T. Jasic, D. Wood, The profitability of daily stock market indices trades based on neural network predictions: Case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965–1999, Applied Financial Economics 14.4 (2004) 285-297.