Predicting AAPL's Stock Price Based on Seasonal Financial Report

Ruocheng Jiang

Abstract

The stock market is a critical component of the economic framework in the United States, serving as a vital investment channel. While offering higher potential returns, stocks also carry a greater risk compared to traditional investments such as bank deposits and bonds. However, with careful management, the rewards can be substantial. This project specifically focuses on understanding the complex dynamics influencing the stock prices of Apple Inc. (AAPL), one of the most prominent companies in the U.S. stock market. To establish a benchmark, the S&P 500 Index is utilized as a comparative measure. Additionally, the stock prices and quarterly financial reports of Microsoft (MSFT), Alibaba (BABA), and NVIDIA (NVDA) are thoroughly analyzed to uncover patterns and correlations in the exploratory data analysis phase of this study.

Following a detailed exploratory analysis, which includes sophisticated visualizations, the project aims to develop a comprehensive Long Short-Term Memory (LSTM) model. This model is designed to predict future stock price movements, offering valuable contributions to the fields of financial analysis and investment strategy. The effectiveness of the model is assessed through rigorous evaluations, comparing four different models based on their R-squared, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) metrics. This systematic approach not only enhances our understanding of stock market dynamics but also aids in the development of robust investment strategies.

Data Source & Data Preprocessing

0.1 U.S. Stock Data from Finviz

The initial dataset was sourced from Finviz via the `pyfinviz.screener` module and includes comprehensive details on 4,412 U.S. stocks as recorded on April 18, 2024. The data was downloaded and aggregated into a single file named `USstock_2024-04-18.csv`.

```
All pages have been merged and saved to 'USstock_2024-04-18.csv' as of 2024-04-18.
```

No	Ticker	MarketCap	PE	FwdPE	PEG	PS	PB	PC	PFCF	EPSthisY	EPSnextY	EPSpast5Y	EPSnext5Y	Salespast5Y	Price	Change	Volume
1	A	38.81B	31.55	21.70	3.92	5.76	6.27	22.20	23.54	1.13%	10.94%	33.93%	8.05%	7.00%	132.44	-1.57%	1,936,563
2	AA	6.37B	-	17.22	-	0.60	1.49	4.59	-	98.37%	5670.50%	-	44.30%	-2.77%	35.47	-0.23%	11,601,635
3	AACI	90.14M	-	-	-	-	5.42	1802.84	-	-	-	-	-	0.00%	11.17	0.00%	118
4	AACT	664.38M	39.28	-	-	-	1.32	347.84	-	-	-	-	-	0.00%	10.63	0.19%	804,210
5	AADI	44.19M	-	-	-	1.81	0.42	0.41	-	-1.07%	-18.00%	11.87%	-	225.26%	1.80	-1.10%	150,100

Figure 1: First Five Rows of Stock_2024-04-18.csv

Several cleaning tasks were performed, firstly, entries with a hyphen ('-') were identified as missing values and replaced with NaN for more accurate processing. Next, rows containing any missing data were removed to maintain data quality.Meanwhile, there is no duplicated data detected. Finally, the cleaned dataset was saved to a new CSV file, tagged with the processing date to differentiate it from the original raw dataset. Each of these steps was crucial in preparing the dataset for the analytical phase, ensuring the data was reliable and representative of the current market conditions.

To enhance the dataset's analytical readiness and for further heatmap and other analysis, 'object' type data should be transfered to appropriate 'float' or 'integer' type. Market cap values were standardized to numeric format by replacing 'B' with 'e9' and 'M' with 'e6' for billions and millions, respectively, then converted to floats. Textual percentage fields for earnings and sales growth were stripped of '%'and converted to decimal form. Comma separators in the volume data were removed to convert the field from string to integer, ensuring numeric consistency. Percentage changes in stock prices were similarly adjusted by removing '%' and converting to floats for uniform processing.

Ticker	MarketCap	PE	FwdPE	PEG	PS	PB	PC	PFCF	EPSthisY	EPSnextY	EPSpast5Y	EPSnext5Y	Salespast5Y	Price	Change	Volume
A	3.881000e+10	31.55	21.70	3.92	5.76	6.27	22.20	23.54	0.0113	0.1094	0.3393	0.0805	0.0700	132.44	-0.0157	1936563
AAON	7.000000e+09	39.98	30.76	3.33	5.99	9.45	775.85	129.12	0.0417	0.2093	0.3182	0.1200	0.2388	85.24	-0.0123	395662
AAP	4.600000e+09	104.27	17.46	1.65	0.41	1.82	9.14	102.36	6.7135	0.1464	-0.3863	0.6330	0.0337	77.20	0.0425	2123063
AAPL	2.579410e+12	25.99	23.46	2.36	6.69	34.85	35.29	24.14	0.0709	0.0848	0.1555	0.1100	0.0834	167.04	-0.0057	42517557
ABBV	2.915600e+11	60.53	13.56	9.64	5.37	28.07	22.75	13.24	0.0027	0.0902	-0.0577	0.0628	0.1176	164.66	0.0025	4480015

Figure 2: The First Five Rows of Data After Cleaned

0.2 S&P 500 Closing Prices from Yahoo Finance

The closing prices of the S&P 500 index over a 5-year period were fetched using the yfinance library, with the data marked from the date of retrieval, April 19, 2024. The data, originally indexed by date, was saved as 'sp500_closing_prices_20240419.csv'. For ease of analysis, the 'Date' column was converted to the datetime format and set as the index of the dataframe, allowing for time-series manipulations and analyses directly within Python's pandas library.

	Date	Close
0	2019-04-22	2907.969971
1	2019-04-23	2933.679932
2	2019-04-24	2927.250000
3	2019-04-25	2926.169922
4	2019-04-26	2939.879883

Figure 3: First Five Rows of S&P 500 Stock Prices

0.3 Closing Prices of Select Four Technology Stocks from Yahoo Finance

Closing prices for selected technology stocks (AAPL, MSFT, NVDA, BABA) were also retrieved using yfinance for the same 5-year period and saved under 'technology_sector_stocks_closing_prices _20240419.csv'. This dataset includes daily closing prices, aiding in a comparative study across major tech companies.

	Date	AAPL	MSFT	NVDA	BABA
0	2019-04-22	51.132500	185.380005	123.760002	47.117500
1	2019-04-23	51.869999	187.289993	125.440002	47.667500
2	2019-04-24	51.790001	185.669998	125.010002	47.792500
3	2019-04-25	51.320000	187.880005	129.149994	46.727501
4	2019-04-26	51.075001	187.089996	129.889999	44.522499

Figure 4: First Five Rows of Closing Prices of Select Technology Stocks

0.4 Seasonal Financial Reports from Yahoo Finance

Seasonal financial reports for Apple, Microsoft, NVIDIA, and Alibaba were downloaded from Yahoo Finance, capturing financial performance on a quarterly basis. This data was merged with historical stock prices to assess correlations between financial health and stock performance. The data included 39 columns representing various financial metrics, structured to align with the corresponding stock price data by setting dates as the index.

	Tax Effect Of Unusual Items	Tax Rate For Calcs	Normalized EBITDA	Net Income From Continuing Operation Net Minority Interest	Reconciled Depreciation	Reconciled Cost Of Revenue	EBITDA	EBIT	Net Interest Income	Interest Expense	...	In Expens Ope
2023-12-31	0.0	0.159	43221000000.0	33916000000.0	2848000000.0	64720000000.0	43221000000.0	40373000000.0	NaN	NaN	...	
2023-09-30	0.0	0.149715	30653000000.0	22956000000.0	2653000000.0	49071000000.0	30653000000.0	28000000000.0	-18000000.0	1002000000.0	...	1002000
2023-06-30	0.0	0.125	26783000000.0	19881000000.0	3052000000.0	45384000000.0	26783000000.0	23731000000.0	-18000000.0	998000000.0	...	99800
2023-03-31	0.0	0.149	32210000000.0	24160000000.0	2898000000.0	52860000000.0	32210000000.0	29312000000.0	-12000000.0	930000000.0	...	93000

4 rows × 39 columns

Figure 5: First Five Rows of AAPL's Quarterly Financial Reports

0.5 Combination of Closing Prices and Financial Reports of Technology Stocks

In the last step, the dataset of closing prices of select four technology stocks and their seasonal financial reports are combined for further exploratory analysis and model building. It should also be noticed that due to the data of quarterly reports would only be updated every 3 months, which is the launch time of next season's report, the financial factor data would also be changed every 3 months.

Exploratory Data Analysis

0.6 Correlation of Each Stock with S&P 500

The correlation between individual stock prices and the S&P 500 index was investigated to understand how closely movements in these stocks are related to the broader market. The analysis revealed a strong positive correlation for Apple (AAPL) and NVIDIA (NVDA), with correlation coefficients of 0.9195 and 0.9550 respectively, suggesting that their stock prices tend to move in tandem with the S&P 500. On the contrary, Microsoft (MSFT) exhibited a negative correlation of -0.5457, indicating an inverse relationship with the market trend. Alibaba (BABA), while also positively correlated, showed a lower correlation coefficient of 0.8062, indicating a moderate alignment with the S&P 500's movements.

```
AAPL    0.919505
MSFT   -0.545658
NVDA    0.954972
BABA    0.806221
Name: SP500, dtype: float64
```

Figure 6: Correlation of Each Stock with S&P 500

The dual y-axis graph illustrates the historical performance of the S&P 500 index alongside selected technology stocks—Apple (AAPL), Microsoft (MSFT), NVIDIA (NVDA), and Alibaba (BABA)—from January 2020 to January 2024. The primary y-axis, representing the S&P 500, displays a significantly higher range compared to the individual tech stocks, necessitating the use of a double y-axis for clearer comparison and visibility.

The S&P 500 shows a pronounced upward trend with periodic volatilities, reflecting broader market sentiments and economic factors.

Meanwhile, the tech stocks exhibit varying growth trajectories, with Alibaba showing a steep increase, suggesting strong company performance and possibly sector-specific growth factors.

The use of dual y-axes allows for direct visual comparison of trends between the broad market and individual tech stocks.

0.7 Dual Y-axhis Graph of 5 Stocks

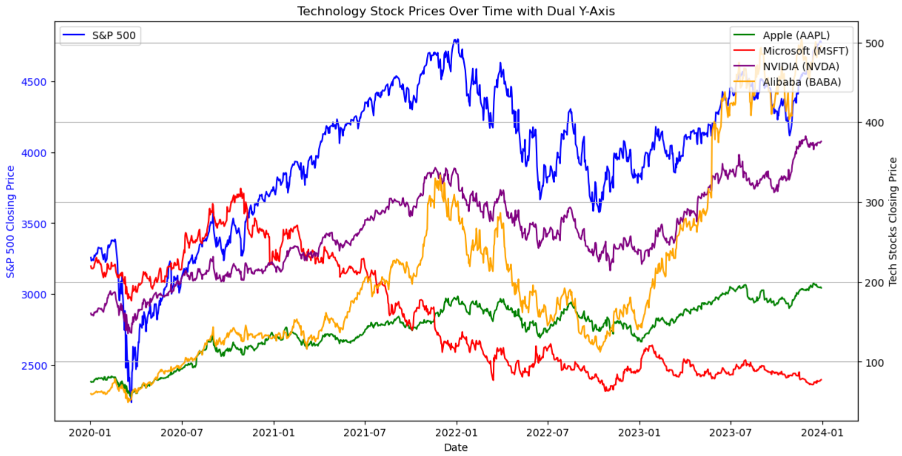

Figure 7: Dual Y-axis Graph of 5 Stocks

0.8 Seasonal Change Points Detection

Changepoint Detection Analysis was applied in the exploratory data analysis (EDA) for the AAPL, MSFT, and BABA stocks. This analysis highlighted potential changepoints in the stock price trends which coincided with the publication of yearly and seasonal financial reports. The red and yellow markers indicate the release of annual and quarterly financial statements, respectively.

For Apple's stock, the overall trend of AAPL shows a notable increase in stock price over time, with sharp rises often following the release of yearly financial reports (marked with red circles).

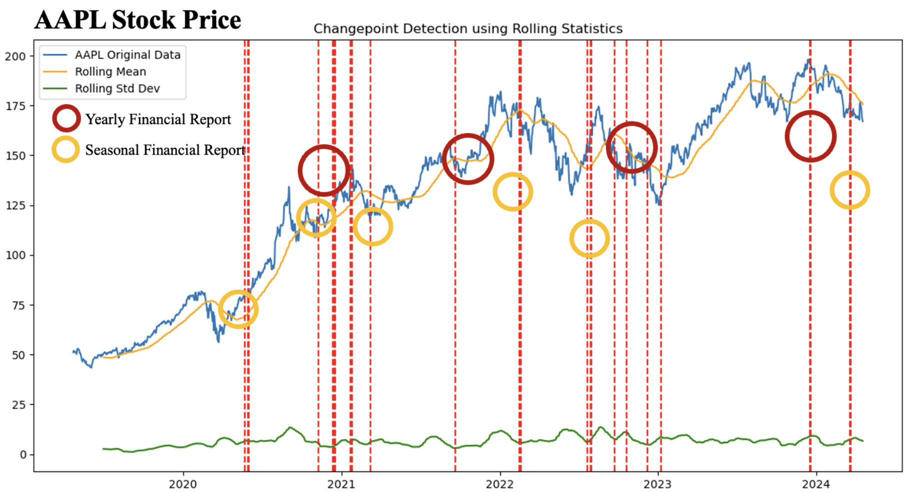

Figure 8: AAPL Change Points Detection

Similar to Apple, Microsoft's stock price movements are aligned with the release of financial reports, where seems most changepoints coincide with the publication dates, especially reports in the end of the year.

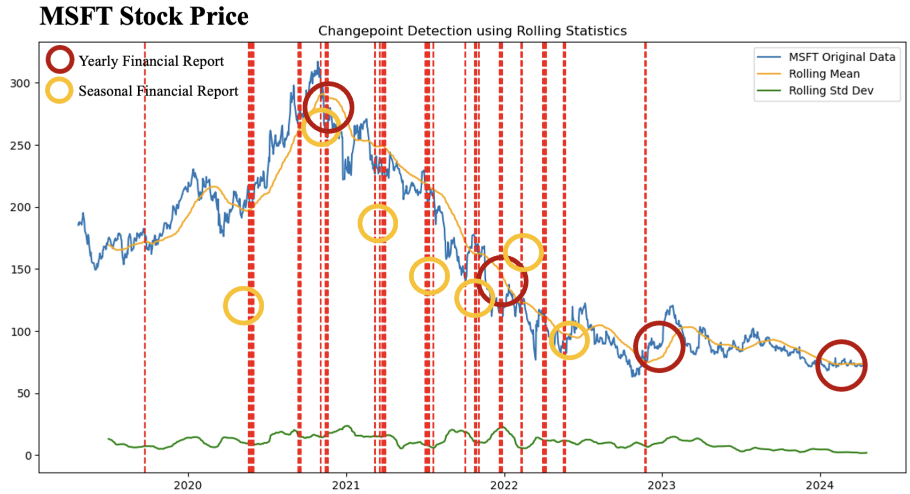

Figure 9: MSFT Change Points Detection

NVIDIA shows several key changepoints around the times financial reports are released, suggesting these events might have a substantial impact on stock performance.

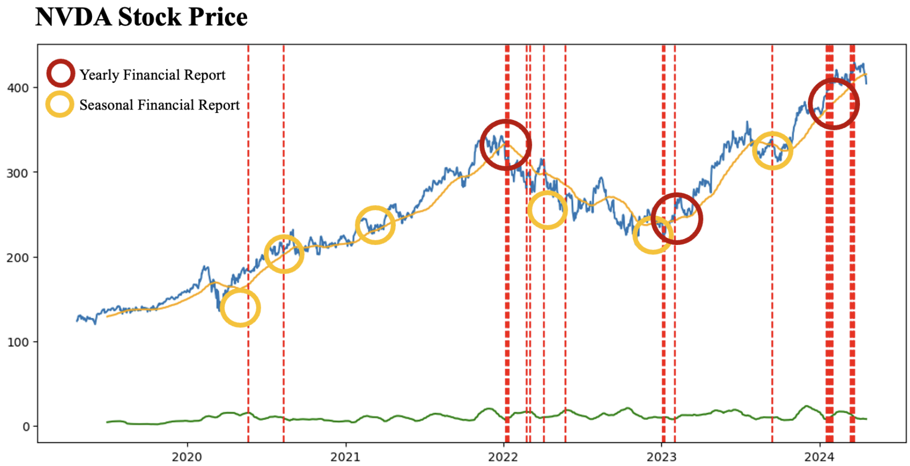

Figure 10: NVDA Change Points Detection

Alibaba's stock price is the least volatile among the four, with gradual ascents and declines rather than sharp changes. After a period of relative stability, there is a sharp increase in late 2023, which might align with a financial report release.

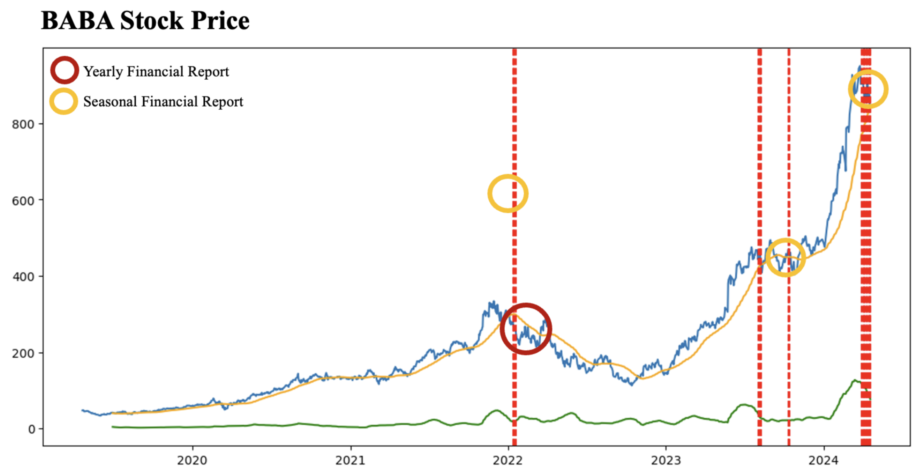

Figure 11: BABA Change Points Detection

Notably, significant price movements often followed these reports, suggesting that financial results may have a measurable impact on stock prices. This pattern is particularly evident in the graphs where price trends change direction or volatility increases around the dates marked by red and yellow circles.

0.9 Heatmap Analysis

Furthermore, this project also delve into the interrelationships between various financial indicators and their correlation with stock prices through heatmap visualizations. The initial correlation matrix, represented through a range of colors from blue (low correlation) to red (high correlation), provides an overview of how different financial metrics like P/E ratio, market capitalization, and earnings per share interact within the dataset of U.S. stocks.

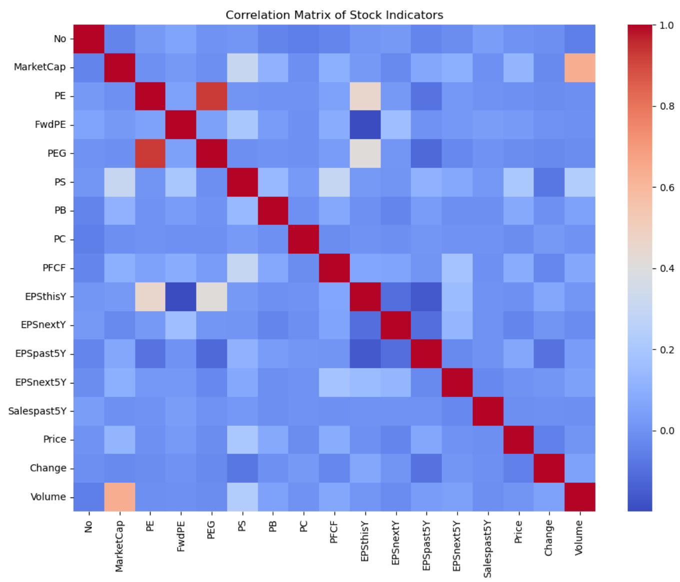

Figure 12: US Stock Heatmap Analysis

Next, a threshold of 0.2 was applied to filter out weaker correlations, resulting in a more refined and readable correlation matrix. This filtered matrix highlights only the most substantial relationships, aiding in clearer visual interpretation and analysis.

Notable findings include a high correlation between the P/E and PEG ratios (0.93), and between volume and market cap (0.64), suggesting that larger companies tend to have higher trading volumes. Additionally, earnings this year (EPSthisY) shows significant correlations with both P/E and PEG ratios, indicating that earnings performance could be a driving factor in valuing stocks.

Meanwhile, for factor that directly correlated with stock prices, the price-to-sales (P/S) ratio(0.2) is spotlighted due to its inherent inclusion of stock price in its calculation, presenting a direct but simplistic view of stock valuation.

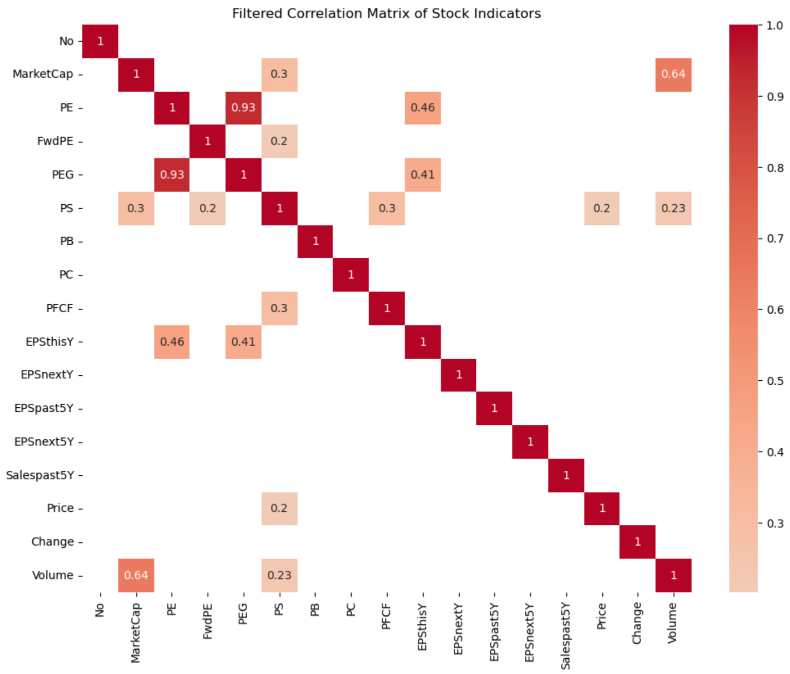

Figure 13: US Stock Correlation Heatmap Analysis with Threshold = 0.2

It has to be admitted that this information is far from enough, thus, deeeper heatmap analysis was applied to focus on four technology stocks.

The correlation matrix for Apple Inc. (AAPL) stock prices and a comprehensive array of financial metrics provides a visual exploration into how various financial factors are interrelated and their impact on the stock price. Significantly, the heatmap highlights the positive correlation between AAPL stock price and interest income and expense (both interest income and expense showing approximately 0.50), suggesting that financial income generated outside of Apple's core business activities has a notable relationship with stock price movements.

Conversely, there are notable negative correlations with diluted and basic average shares, indicating that increases in the number of shares (potentially due to stock dilution) tend to negatively affect the stock price, a common outcome in equity markets as dilution typically spreads earnings across a greater number of shares, potentially reducing earnings per share.

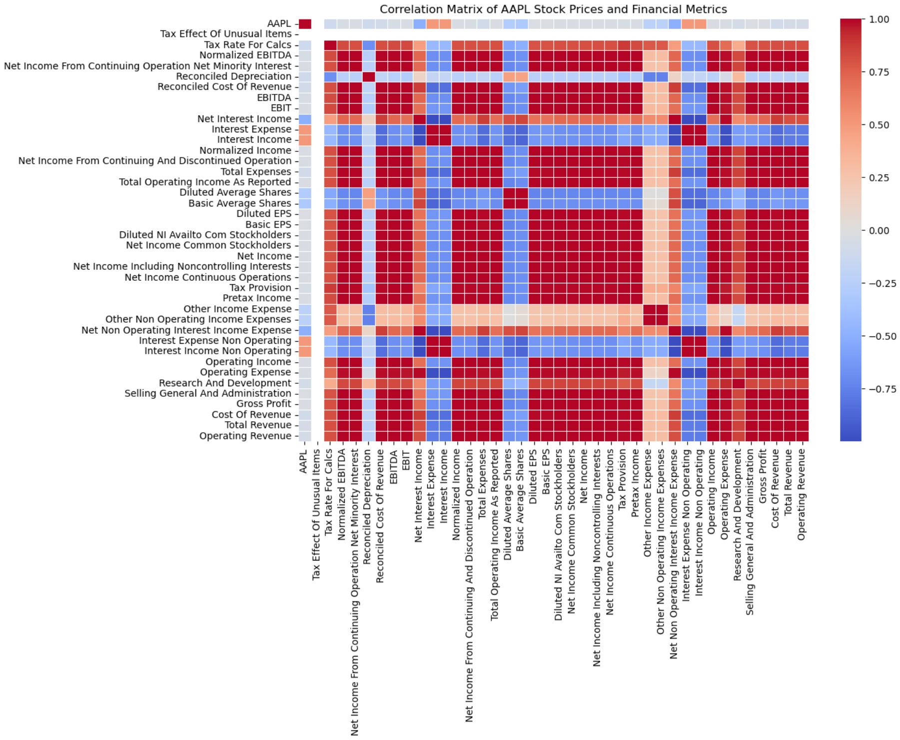

Figure 14: Correlation Heatmap of AAPL Stock Prices and Financial Factors

To support the correlation analysis, heatmaps from other 3 stocks are also included.

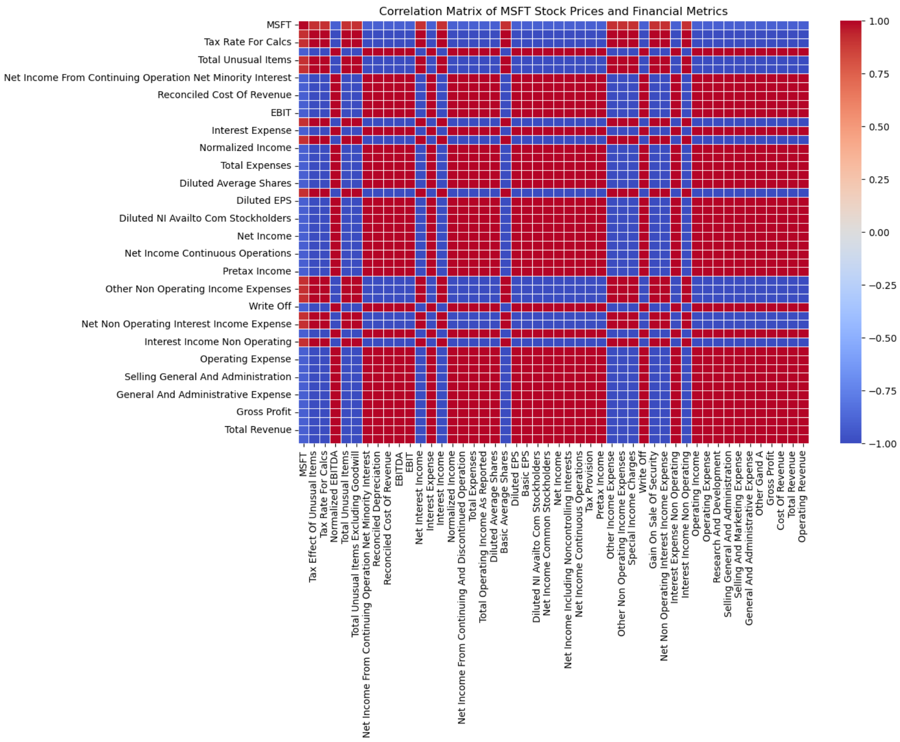

Figure 15: Correlation Matrix of MSFT Stock Prices and Financial Factors

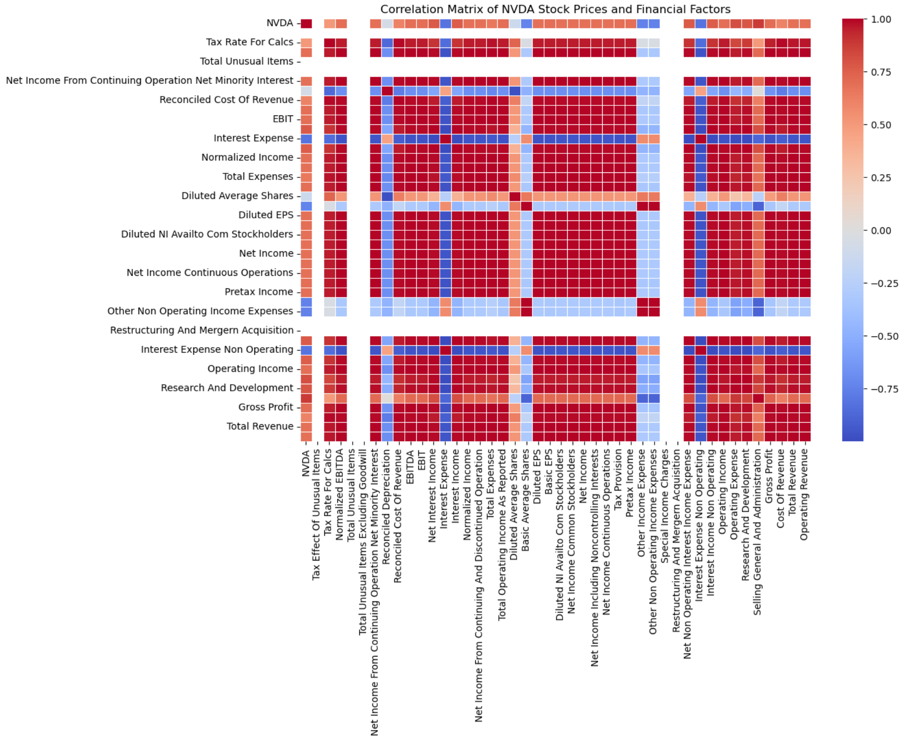

Figure 16: Correlation Heatmap Matrix of NVDA

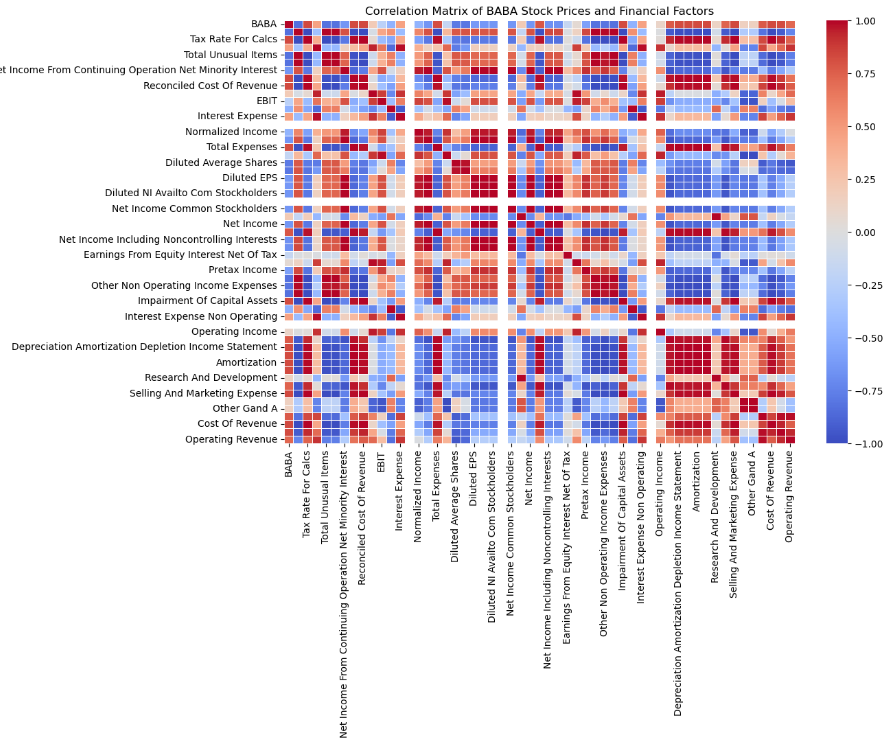

Figure 17: Correlation Matrix Heatmap of Alibaba Prices and Financial Factors

0.10 Financial Factors Selection

Based on analysis of these correlation matrices, several factors from the financial statement are selected to build the LSTM model: 'Interest Income Non Operating', 'Interest Income', 'Interest Expense Non Operating', 'Interest Expense', 'Basic EPS', 'Net Interest Income', 'Net Non Operating Interest Income Expense', 'Basic Average Shares', 'Diluted Average Shares', 'Net Income', 'Operating Income'.

Long Short-Term Memory (LSTM) Model

The following series of Long Short-Term Memory(LSTM) models are produced based on these chosen factors, which are applied to forecast AAPL's stock price based on various datasets reveal distinct insights and challenges in using time series prediction techniques for stock price prediction. Below is a summary and analysis of the models and their respective outputs:

0.11 LSTM Model Based on Selected AAPL Financial Factors

The initial model utilizes LSTM to predict AAPL stock price based solely on selected financial factors from AAPL's reports. However, the quarterly changing nature of the input seasonal report leads to an unrealistic prediction pattern, with three change points and other parts are all horizontal line, where the predicted values for both training and test sets show substantial deviation from actual stock prices, indicating that the model struggles to capture the underlying trends and volatilities of AAPL's stock price based on the selected financial factors alone.

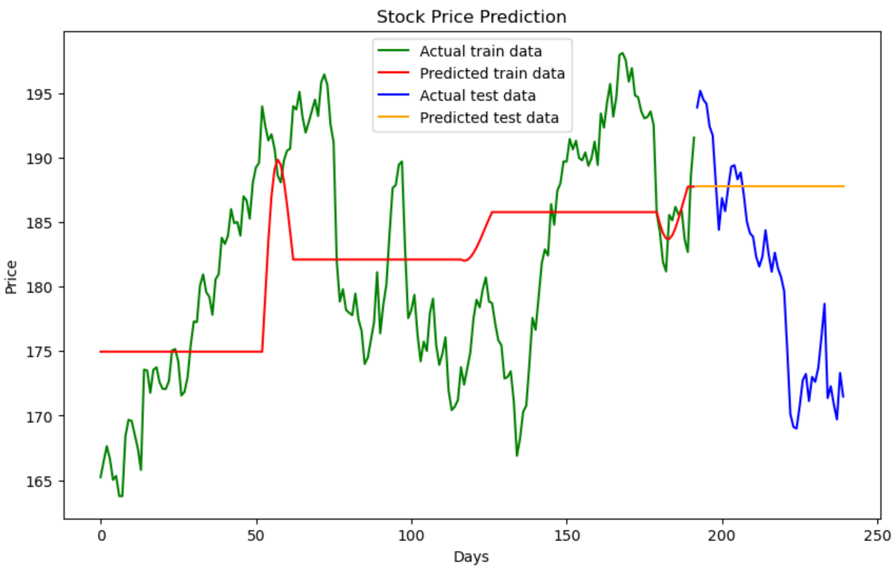

Figure 18: LSTM Based on SelectedAAPL Financial Factors

```
R-squared: -0.6401930760730441
Mean Squared Error: 109.6262931273911
Root Mean Squared Error: 10.470257548283666
Mean Absolute Error: 8.529051144917808
```

Figure 19: Evaluation

0.12 Add Moving Averages and Bidirectional Adjustments to LSTM Model Based on AAPL Financial Factors

To improve the unnatural prediction pattern, enhancements with moving averages and bidirectional LSTM layers were introduced to better capture trends and potential reversals in the stock price movement. Despite these adjustments, the model's performance did not improve, as indicated by the unchanged unrealistic prediction pattern in the figure.

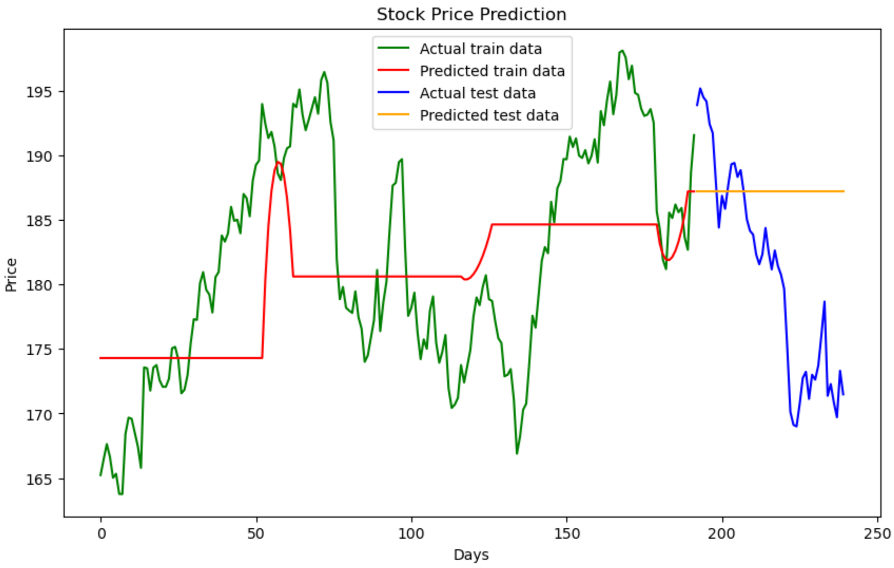

Figure 20: Moving Average & Bidirectional Adjustments Added LSTM Model

```
R-squared: -17.500849936446876
Mean Squared Error: 102.37864391500386
Root Mean Squared Error: 10.118233240788822
Mean Absolute Error: 8.25001621246338
```

Figure 21: Evaluation

0.13 LSTM Model Based on S&P 500 Index

To further try to avoid the huge impact caused by quarterly financial report, this model attempts to predict AAPL stock price based on the historical prices of the S&P 500 index.

With this adjustment, the predicted training and testing data begin to fluctuate. However, the predicting accuracy is far from enough, with an opposite trend from the actual testing set.

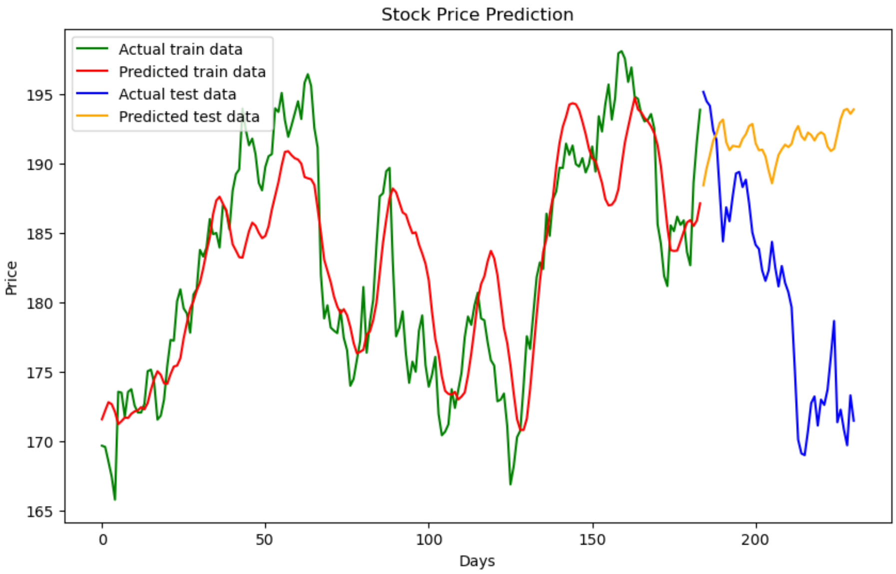

Figure 22: LSTM Model Based on S&P 500 Index

```
Mean Squared Error: 192.51676273132443
Root Mean Squared Error: 13.875040999266432
Mean Absolute Error: 11.676761383705955
R-squared: -2.181036351331445
```

Figure 23: Evaluation

0.14 LSTM Model Combining S&P 500 and AAPL's Seasonal Financial Reports

Consequently, a further attempt combined S&P 500 Index with AAPL's Quarterly Financial Report are generated.

From this figure, although the trend seems similar, but predicted change point differs too much from the actual change point.

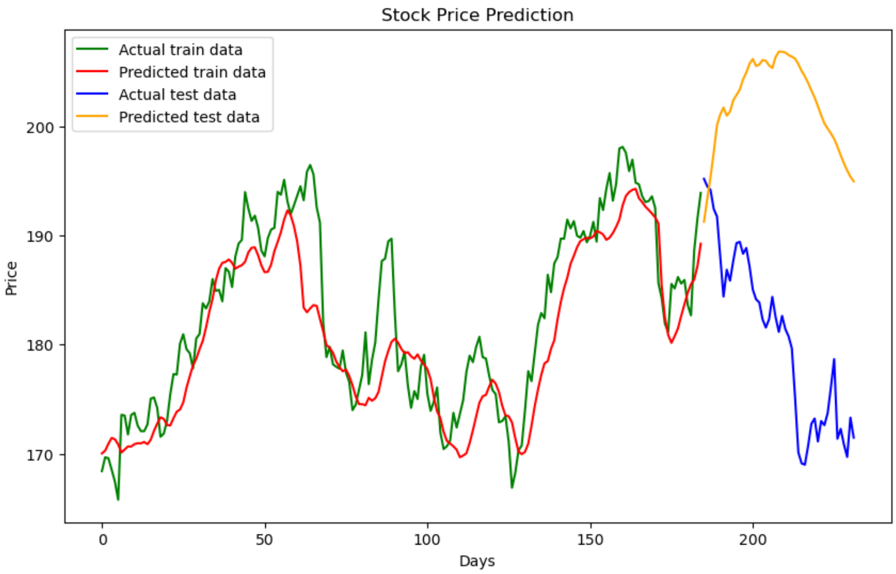

Figure 24: LSTM Model Combining S&P 500 and AAPL's Seasonal Financial Reports

```
Mean Squared Error: 539.7354124588645
Root Mean Squared Error: 23.23220636226496
Mean Absolute Error: 21.644342625394785
R-squared: -7.91827777889992
```

Figure 25: Evaluation

0.15 Combining Stock Prices of MSFT, NVDA, BABA to LSTM Model with AAPL's Financial Reports

Another attempt is to build the model by combining AAPL's financial report factors with AAPL, MSFT, NVDA and BABA these four stocks' price history. In this way, the figure and evaluation all reaches the best precision.

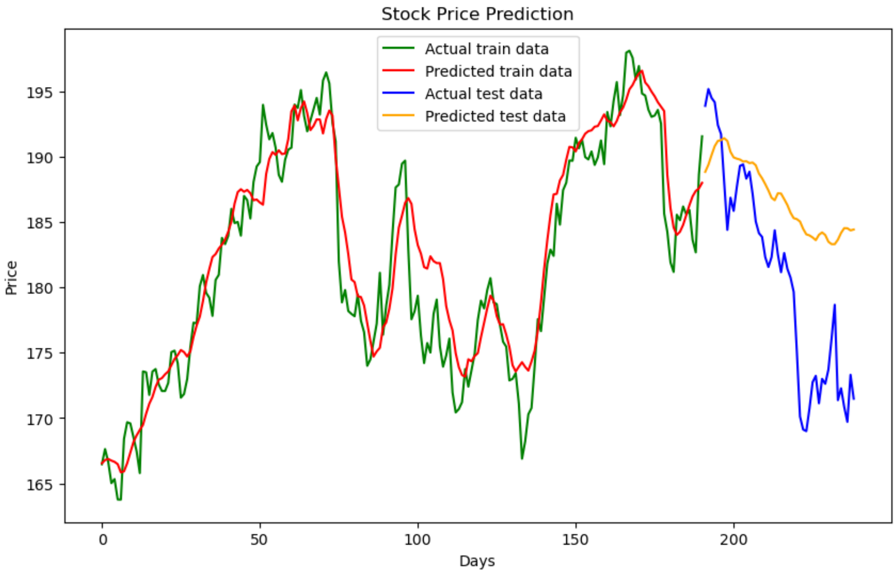

Figure 26: MSFT, NVDA, BABA Prices being Added to LSTM

```
Mean Squared Error: 68.12444958262014
Root Mean Squared Error: 8.253753666218792
Mean Absolute Error: 6.900003115336102
R-squared: -0.08404271767100058
```

Figure 27: Evaluation

Comparison of Four Models Based on Evaluations and Figures

0.16 R-Squared Value Analysis

The R-Squared value of the Model Combining Stock Prices of MSFT, NVDA, BABA (Model 5) reaches the best R-Squared value of -0.08, the fisrt model, based on simply selected AAPL financial factors (Model 1)reaches the second best value of -0.64.

However, according to According to Pierce and David A[1], it is not the most appropriate method to evaluate model's accuracy through R-Squared value, and other factors would be considered more in this project.

Moreover, it is also obvious that although the R-Squared value of Model 1 seems good, the predicted trend is far from real.

0.17 Mean Squared Error(MSE) Analysis

The MSE value of 68.12, which is the lowest among the models evaluated, indicates that Model 5 has the best prediction accuracy. While LSTM Model Combining S&P 500 and AAPL's Seasonal Financial Reports(Model 4) presents the worst value. The relatively better performance of this model could indicate that Apple's stock price movements are more closely correlated with these technology peers than with broader market indices like the S&P 500.

Although the first two models, which rely solely on Apple's seasonal financial reports, show relatively low MSE values of 109.63 and 102.38, the patterns they predict are unrealistic. This inconsistency suggests that despite the seemingly adequate error metrics, these models fail to capture the true dynamics of Apple's stock price movements. Therefore, they cannot be considered the best models for accurately forecasting stock prices. The discrepancy between the MSE values and the observed prediction patterns underscores the importance of evaluating models not just on numerical metrics, but also on how realistically they reflect market behaviors and trends.

0.18 Root Mean Squared Error(RMSE) Analysis

Similar with the first two accuracy analysis, Root Mean Squared Error(RMSE) of Model 5(8.25) is also the lowest and best, Model 4(23.23) being highest and worst one. Model1(10.47) and Model2(10.12) also has lower scores, but not being considered due to unnatural patterns.

0.19 Mean Absolute Error(MAE) Analysis

The analysis of Mean Absolute Error (MAE) corroborates the earlier findings, indicating that Model 5 is the most accurate model given its lowest MAE value of 6.90. Conversely, Model 4 exhibits the least precision, recording the highest MAE at 21.64. This further supports the effectiveness of Model 5 in providing the most reliable stock price predictions.

Summary

0.20 Final Selected Model

After comprehensive analysis using MAE, MSE, RMSE, R-squared values, and figure-based evaluation, it is evident that the most appropriate model is the one where Apple's financial report data is combined with the historical stock prices of Apple, Microsoft, NVIDIA, and Alibaba. This model achieves the highest precision, confirming its effectiveness in forecasting stock prices more accurately than other tested models.

0.21 Further Improvements of This Model

Looking ahead, there are several ways to enhance the predictive power and reliability of this model. Incorporating additional data dimensions such as trading volumes, macroeconomic indicators, and industry news could provide deeper insights and improve the accuracy of the model. Regularization techniques, like L1 or L2, can be applied to prevent overfitting, ensuring the model remains robust even when exposed to new or varied data sets. Furthermore, extending the analysis to cover a longer historical period may capture more cyclical trends and anomalies, providing a richer dataset for training and validation.

0.22 Practical Applications

In practical applications, this refined model can be an invaluable tool for various financial strategies. It can aid in risk management by estimating potential price volatility and associated risks more accurately. Additionally, the model can enhance timing strategies for buying and selling stocks, helping investors capitalize on predicted price increases and sell at anticipated peaks. It can also inform portfolio diversification efforts, allowing investors to analyze correlations between different stocks or assets and construct a well-balanced investment portfolio. Lastly, the model can support long-term investment strategies by identifying stocks with high potential for long-term growth, based on predicted long-term trends.

0.23 Conclusion

To sum, this approach not only underscores the potential of machine learning in financial modeling but also highlights the importance of continuous model refinement and the integration of diverse data sources to capture the complex dynamics of the stock markets more effectively.

References

[1] David A Pierce. R 2 measures for time series. *Journal of the American Statistical Association*, 74(368):901–910, 1979.