The Return Predication and Risk Assessment of Tesla

Yongyuan Qu

yq2361@columbia.edu

Jiajun Gu

jg4681@columbia.edu

Yufei Lang

yl5404@columbia.edu

Xuefei Xu

xx2437@columbia.edu

Weitian Xie

wx2298@columbia.edu

Jiayu Chen

jc6116@columbia.edu

I. Introduction

Tesla is a popular company in recent years because of their innovative products, such as Model 3, Model X, Model Y, etc. The notion of Tesla is encouraging zero-carbon emission transportation and focusing on replacing fossil fuels with cleaning energies, like electricity. So, the stock price of Tesla increasing a lot and many investors care about this company. However, due to the COVID-19, the delivery of new vehicles is postponed, and the economy of the US is not very good. it will lead to the decline of the stock price of Tesla and many people begin to lose confidence. In 2023, the business of Tesla begins to rebound, and our team wants to create serval models to predict the future price of Tesla by using the historical volatility and factors. Therefore, we want to investigate the best relationship between **Future Stock prices** with **proper risk**. Based on this situation, several models are established: Model I: Multi-Factor Model; Model II: Time Series Model; Model III: Random Forest Model.

II. Data Source

Table 1: Data Collection

Database	Data Sources	Data Type
Tesla	Yahoo Finance	Float
Benchmark	S&P 500	Float

III. Model I: Multi-Factor Model

Factor analysis is a strategy that find attributes that are associated with higher returns. There are two main types of factors, macroeconomic factors, and style factors. For this part, we will emphasize on utilizing fama french four style factors to give us a broader picture of two stocks factors that have associations with their return performance. We estimate the size of firms, book-to-market values, excess return, and momentum.

OLS Regression Results

Dep. Variable:	TSLA-RF	**R-squared:**	0.365
Model:	OLS	**Adj. R-squared:**	0.363
Method:	Least Squares	**F-statistic:**	146.7
Date:	Sun, 10 Mar 2024	**Prob (F-statistic):**	3.79e-99
Time:	20:04:31	**Log-Likelihood:**	-2712.7
No. Observations:	1026	**AIC:**	5435.
Df Residuals:	1021	**BIC:**	5460.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2027	0.107	1.902	0.057	-0.006	0.412
Mkt-RF	1.4677	0.076	19.404	0.000	1.319	1.616
SMB	0.8373	0.143	5.853	0.000	0.557	1.118
HML	-0.7487	0.092	-8.181	0.000	-0.928	-0.569
Mom	0.1254	0.080	1.566	0.118	-0.032	0.283

Omnibus:	96.936	**Durbin-Watson:**	2.009
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	604.164
Skew:	0.105	**Prob(JB):**	6.42e-132
Kurtosis:	6.753	**Cond. No.**	2.24

Figure 1: Fama-French Model Results

From the initial conditions of multi regression condition check with four factors analysis, we can see the result a little bit different from only including one factors ofmarket risk premium. By estimating past return during the period of covid, both of them has the constant coefficient greater than 0 and it is way of measuring risk adjusted performance (Tesla has 0.2027% abnormal return over the market per day), indicating that it performs not at an equal performance level well over the market. For the coefficient of risk premium, it displayed a high volatility than market which makes sense due to a higher return and it displays a higher accuracy when we have incorporated more factors. Tesla increase the expected return by 1.4677% change with 1 unit change in the market risk premium in a daily basis. We have a more accurate approximation of beta relative to market compared by incorporating more variables compared with one factor in CAPM model, and it also suggests that tesla has a higher magnitude of volatility relative to market.

III. ARIMA Model

Condition: p=1, d=0, q=1

Our team first plot the monthly volatility to see whether the data is stationary or not. We find that the volatility has strong seasonality and decide to remove the seasonality by using decentralizing method. Then we apply ADF (Augumented Dicky Fuller Test) to conclude that the time series is stationary.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  250
Model:                 ARIMA(1, 0, 1)   Log Likelihood                -843.144
Date:                Mon, 22 Jan 2024   AIC                           1694.289
Time:                        17:20:21   BIC                           1708.374
Sample:                             0   HQIC                          1699.958
                                - 250
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const        158.8567     51.002      3.115      0.002      58.894     258.819
ar.L1          0.9958      0.006    174.267      0.000       0.985       1.007
ma.L1         -0.0011      0.051     -0.022      0.982      -0.102       0.099
sigma2        48.8251      2.677     18.239      0.000      43.578      54.072
===================================================================================
Ljung-Box (L1) (Q):                   0.17   Jarque-Bera (JB):               162.80
Prob(Q):                              0.68   Prob(JB):                         0.00
```

Figure 2: ARIMA Model Result

From the SARIMA model results, we can see that constant and AR coefficients are both significant at a 5% significance level, we can see that the overall volatility level of tesla, As the SARIMA coefficient value is higher than the other model's coefficient value, we can conclude that the volatility is highly persistent and clustering. Also, the graph at the bottom can better illustrate the volatility level of tesla stock as well. The relative low AIC and BIC socre means the better predict result and fit the regression of the model.

95% Confidence Interval:

	lower y	upper y
250	185.023602	212.414092
251	179.233861	217.865585
252	174.778897	221.983737
253	171.020093	225.407157
254	167.709249	228.384039
255	164.718717	231.042028
256	161.972467	233.457144
257	159.420840	235.679043
258	157.029252	237.742302
259	154.772461	239.672156
260	152.631364	241.487704
261	150.591083	243.203817
262	148.639754	244.832353
263	146.767736	246.382948
264	144.967061	247.863563
265	143.231054	249.280869
266	141.554053	250.640521
267	139.931206	251.947365

Figure 3: Forecast Stock Price

IV. Random Forest Model

The original data set contains four dependent variables: Open, High, Low, and Volume. Considering the characteristics of the financial markets, the researchers of this program added two features: 5-day moving average and 10-day moving average. Moving averages help in identifying the underlying trend in stock prices by smoothing out daily price fluctuations.

We use MSE and R^2 to assess the accuracy of regression model predictions. A lower MSE indicates that the model's predictions are closer to the actual closing prices, while a higher R^2 score (closer to 1) suggests that the model explains a larger portion of the variance in the closing prices. In this project, the Training MSE is 4.1028, suggesting that, on average, the model's predictions deviate from the actual values by a relatively small margin. The Training R^2 is 0.9972, which is exceptionally high, indicating that 99.72% of the variance in the dependent variable is predictable from the independent variables. This suggests that the model fits the training data well. The low MSE and high R^2 indicate that this model has learned the training data with a high degree of accuracy, capturing the underlying relationships between the independent variables and the target variable very well. However, while excellent training performance is desirable, such high R^2 and low MSE values might also indicate a potential risk of overfitting, which can negatively impact the model's ability to generalize to unseen data (test data). Therefore, it's essential to evaluate the model on a separate test dataset using the same metrics to confirm that it generalizes well.

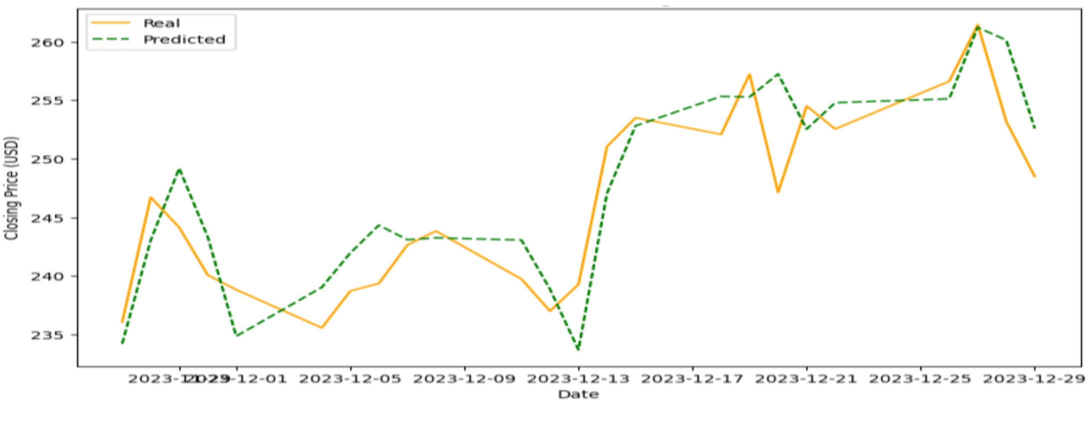

Figure 4: Predicated Trend

The MSE on the test set is 15.4604, indicating that the model's predictions deviate from the actual values to a certain extent. Compared to the training MSE of 4.1028, the higher test MSE suggests that the model does not predict as accurately on unseen data as it does on the training data. The R^2 value on the test set is 0.7238, indicating that approximately 72.38% of the variance in the dependent variable is predictable from the independent variables in your model when applied to unseen data. While this is a significant proportion, the drop from the training R^2 of approximately 99.72% to 72.38% on the test set indicates a decrease in predictive accuracy when the model encounters new data.

V. Sensitivity Analysis

MSE Across ARIMA Parameters: This plot shows how the Mean Squared Error varies with different values of the AR term p for combinations of differencing d and moving average term q. Lower MSE values are better, indicating that the model's predictions are closer to the actual data.

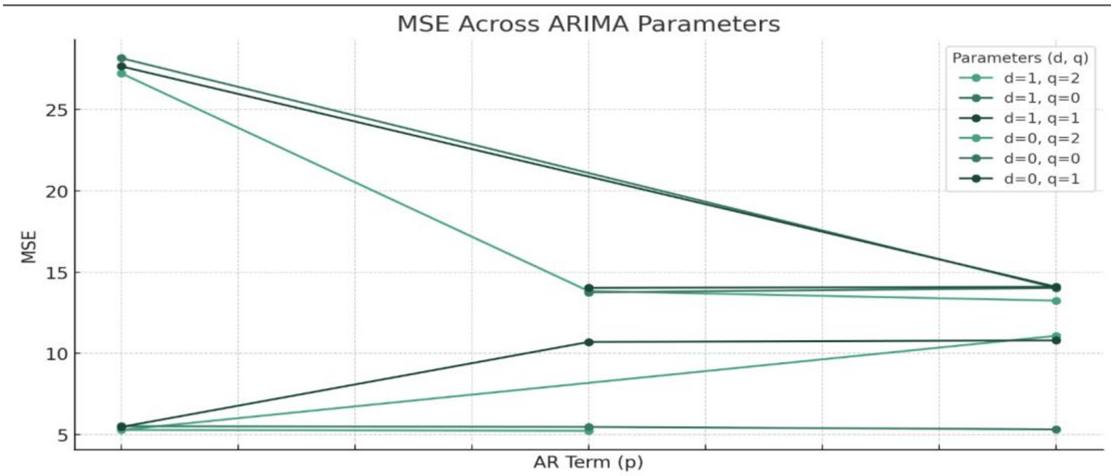

Figure 5: MSE Across ARIMA Parameters

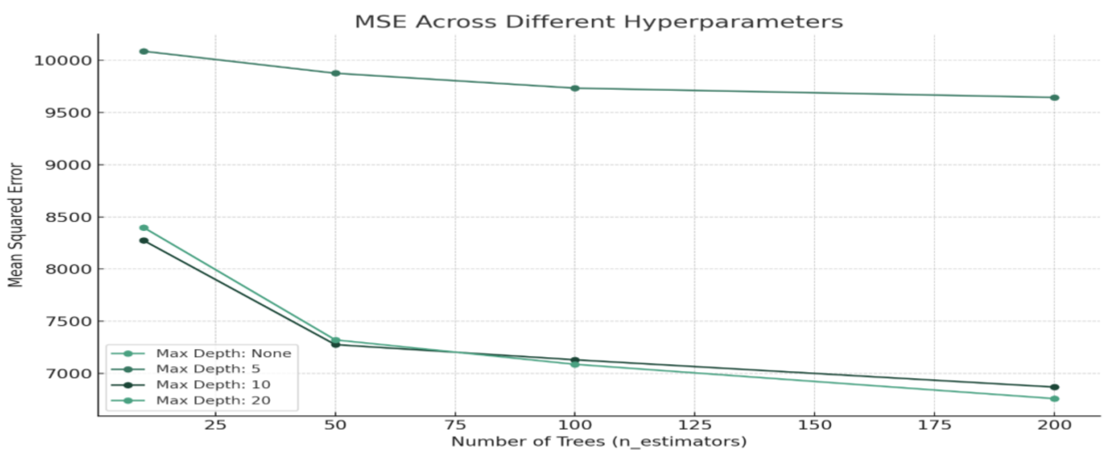

Figure 6: MSE Across Different Hyperparameter

This plot illustrates how MSE changes with different numbers of trees (n_estimators) for various maximum tree depths (max_depth). Lower MSE values indicate better model performance, where predictions are closer to the actual data. Generally, more trees lead to lower MSE, particularly when no maximum depth is set, allowing the trees to fully grow.

VI. Conclusion

Strengths

1. Comprehensive Approach: The use of diverse modeling techniques allows for a robust analysis from various angles, ensuring that different aspects of stock behavior are thoroughly examined.
2. In-depth Model Evaluation: Each model is carefully evaluated for its predictive accuracy and relevance to Tesla's stock, with sensitivity analyses adding depth to the understanding of each model's strengths and limitations.
3. Practical Implications: The findings are relevant for investors looking for high-risk, high-return opportunities, providing them with a detailed risk assessment and return prediction for Tesla.

Weakness

1. Complexity and Interpretability: The complexity of models like Random Forest may hinder their interpretability, which is crucial for financial decision-making where understanding the model's decision process is as important as the outcomes it predicts.
2. Market Volatility and Model Sensitivity: The high volatility of Tesla's stock and the external market conditions significantly impact the models' predictions, suggesting a potential for reduced predictive performance under unexpected market shifts.

The analysis underscores Tesla as a potentially lucrative but highly volatile investment. The high volatility reflects Tesla's sensitivity to both internal developments and external economic conditions. For investors, this means a need for a balanced approach that considers both the promising growth potential and the significant risks. Effective risk management strategies are recommended, emphasizing the need for continuous monitoring of market conditions and regular model updates to adapt to new data. The insights gained from these models equip investors with a deeper understanding of how Tesla's stock might behave, enabling more informed investment decisions in a landscape shaped by innovation and market dynamics.