Comparing Machine Learning Methods with the Black–Scholes–Merton Model for Option Pricing

Aiden Marin   Mentor: Rafiq Islam

FSU FLORIDA STATE UNIVERSITY

FSU UNDERGRADUATE RESEARCH OPPORTUNITY PROGRAM
CENTER FOR UNDERGRADUATE RESEARCH & ACADEMIC ENGAGEMENT

Introduction

Professionals have attempted to find the best way to determine a stock option's value for years. Traditional models, such as the Black-Scholes-Merton (BSM) model, have been used to estimate an option's theoretical value. These models often rely on constant assumptions and are not very reactive to constantly changing market dynamics. To address these issues, this study explores the effectiveness of using machine learning models to determine an option's value.

Background

The traditional option pricing formula for call options given by the BSM model:

$$C = S_0 N(d_1) - K e^{-rT} N(d_2)$$

where K is the strike price, S is the initial stock price, r is the risk-free interest rate, T is the duration until the option matures, and both N(d1) and N(d2) represent the probability that the option will finish in-the-money or not.

- This model assumes constant volatility, and is unable to adapt to abnormal volatility or changes in a market .
- In reality, market volatility is constantly changing, and prices can behave extremely unpredictably, leading to inaccurating pricing.
- Machine learning models are capable of capturing more complex relationships in data, potentially improving price predictions as they are able to respond to abnormal data.

Objectives

- Evaluate the ML models' accuracies in pricing stock options compared to the BSM model.
- Determine the best ML model that is commonly used in quantitative finance to use in option pricing among Linear Regression, Random Forest, Support Vector Machine, XGBoost, and Nueral Networks.
- Determine if ML models outperform the BSM model in determining an option's value

Methodology

1. Collected historical option and stock price data using Yahoo Finance.
2. Cleaned data: removed empty values, and expired options.
3. Split data 80/20 : 80 real values, 20 artifical values in order to avoid overfitting .
4. Calculated option prices using the BSM formula and trained multiple machine learning models (Linear Regression, Random Forest, SVM , XGBoost, Gradient Boosting).
5. Compared model predictions to the actual market data using metrics RMSE and R², and visualized results with graphs.

Visualization

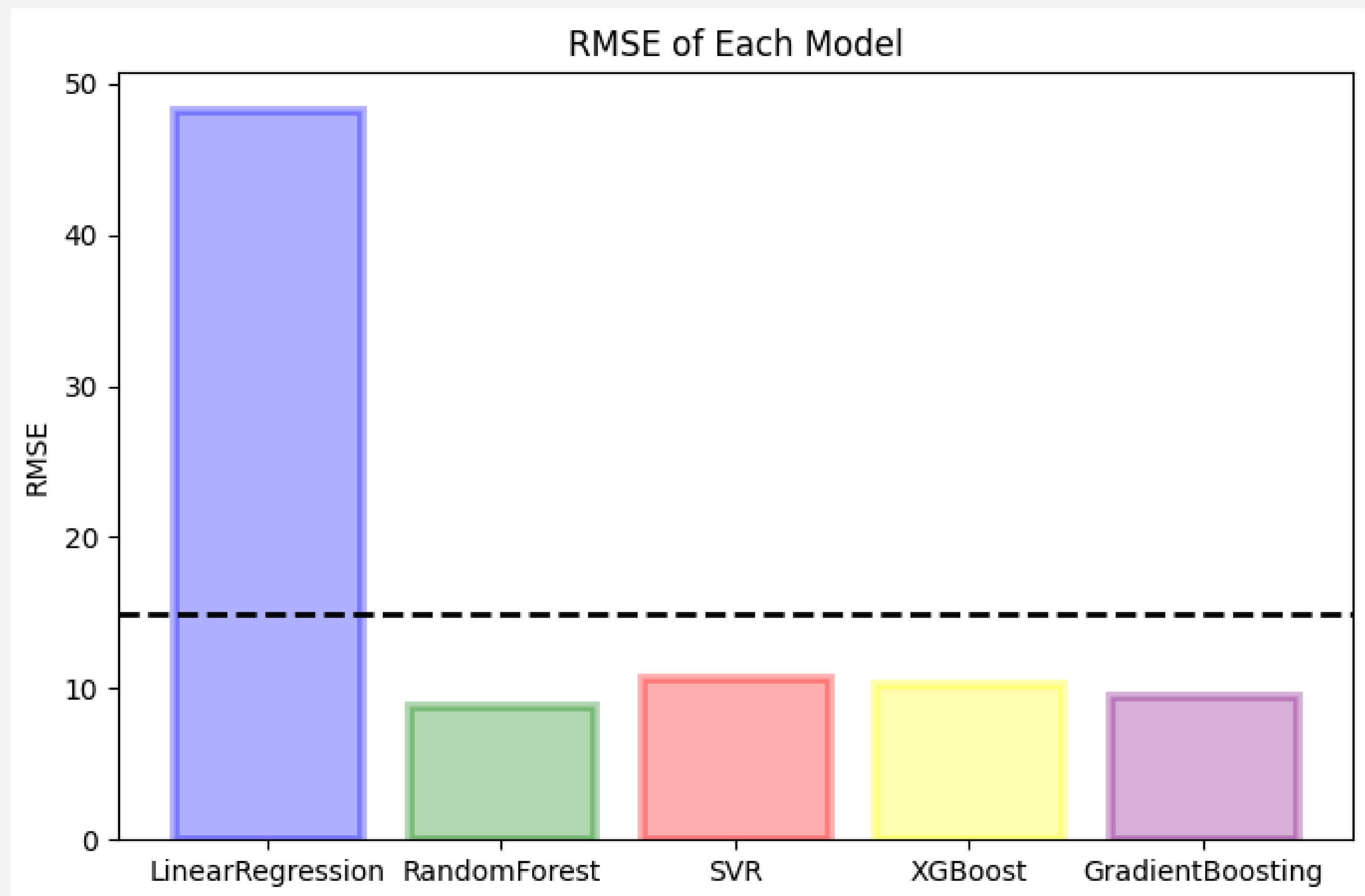


Figure 1: RMSE of each model (SPY)

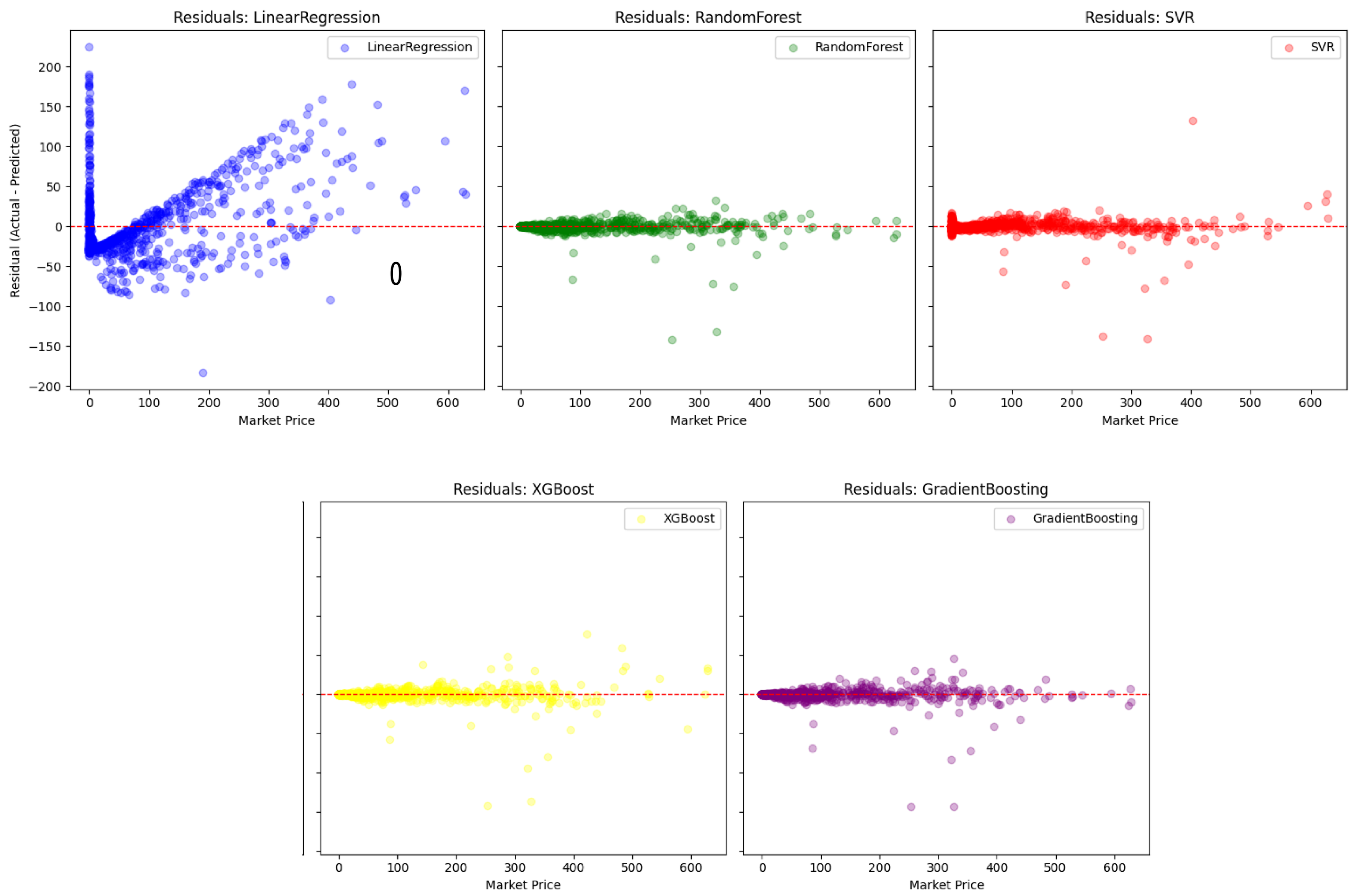
** Black Dotted Line Represents the Black-Schole Merton Formula's RMSE **

Root Mean Square Error (RMSE) is metric that magnifies the average difference between predicted and observed values, measuring a model's accuracy.

The R2 Score (R2) is metric that determines how well a model explains data

- 0 = Explains none of the data
- 1 = Explains all of it

Individual Model Residual Plots



Results

Graph	RMSE	R2	Ranking
BSM	14.91		5
LR	48.3	0.83	6
RF	8.85	0.99	1
SVM	10.65	0.99	4
XGB	10.28	0.99	3
GB	9.5	0.99	2

Table 1: Comparison of model performance on 1000 in-the-money options on SPY (Nearest Hundreth)

Conclusion

- Machine learning models can predict option prices with high accuracy, even outperforming the traditional Black-Scholes-Merton model.
- Random Forest performed the best consistently in every trial, so is the best model to be used to determine an options value.
- Data could be potentially overfit, even after splitting.

2