# Bachelor's degree defense

## Option Analysis With Advanced Artificial Intelligence Techniques

Aleksandr Algazinov

Higher School of Economics

May 28, 2024

NATIONAL RESEARCH UNIVERSITY

# Introduction

## What is the problem?

We wanted to come up with a universal, applicable, efficient, scalable, and open-source method for estimating (and forecasting) option prices. It turns out that AI tools seem to be sufficient for such purposes.

## Why is it relevant?

Options are popular instruments. However, they are generally considered as advanced investment tools (high risks and uncertainty). Therefore, precise, reliable, open, and available to public methods will be beneficial both for students, companies, and traders.

## What we propose?

The classic way to solve this problem is to apply well-known analytic solutions (such as Black-Scholes formula and its extensions). However, this family of methods has several limitations. To eliminate most of them, we will switch to using AI tools.

# Benchmark (Black-Scholes model)

The Black-Scholes formula is used to calculate the theoretical price of European-style call options. It is given by:

$$C(S_0, t) = N(d_1)S_0 - N(d_2)Xe^{-r(T-t)}. \tag{1}$$

$$d_1 = \frac{\log\left(\frac{S_0}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}. \tag{2}$$

$$d_2 = d_1 - \sigma\sqrt{T-t}. \tag{3}$$

$S_0$ - current price of the underlying asset, $N(x)$ - cumulative distribution function of the standard normal distribution, X - strike price of the option, r - risk-free interest rate, T - time to expiration, t - current time, $\sigma$ - volatility of the underlying asset.

# Significance of Black-Scholes model

## Historical significance

Model brought a new (systematic, clear, and rather easy to implement) quantitative approach to pricing options. People developed option trading strategies (using the principle of dynamic hedging) and started doing more research related to options.

## The way it is applied today

While the Black-Scholes model is unlikely to be used in its original form, the modern approach to options trading is based on the principle of dynamic hedging (at least partially). Moreover, Black-Scholes model is used to calculate implied volatility.

## Extensions of the model

Several "more realistic"models were created (e.g. Heston, Jump-Diffusion). Moreover, researchers continue to offer various modifications of the model. However, the main attention turns to the use of neural networks.

# Assumptions of the Black-Scholes model

- Stock prices follow a random walk in continuous time. So, prices of financial assets cannot be predicted with consistent accuracy.
- Stock prices are log normally distributed.
- Stock returns are normally distributed.
- Volatility is constant over time.
- Stocks do not pay dividends.
- Risk-free interest rate does not fluctuate over time.
- Options can only be exercised at expiration.
- Markets are frictionless (no transaction costs).
- There are no arbitrage opportunities in the market.

# Assumptions we make

- The data for modeling is of high quality and is large enough.
- We do not put prior distributions on the evolvement of prices of underlying assets and returns on them.
- Risk-free is constant. In the process of modeling, the rate on US treasuries (which we consider risk-free) is unlikely to change.
- Companies do not pay dividends at all (since we do not have data representing dividends).
- Market can be inefficient, there may be transactions costs, and the short-selling opportunities are likely to be restricted.
- Volatility can and should change over time. However, for the simplicity of analysis, we will calculate it only once.
- Options are not required to be European-style. However, we analyze only call options.
- Additional features (besides the ones used in Black-Scholes model and its extensions) affect the prices of options and should be included in the analysis.

# Why using AI-tools is justified

Limitations of the classical methods:

- 1. Rely on strong assumptions.
- 2. Do not fully take into account the characteristics of individual markets and companies.
- 3. Cannot take certain features as an input (for example, categorical).
- 4. Work slow on inference.
- 5. Are complicated in terms of forecasting.

With implicit modeling we can:

- 1. Estimate prices of all types of options with weaker assumptions.
- 2. Learn directly from data (cannot learn irrelevant things).
- 3. Use unlimited number of features.
- 4. Perform hypothesis testing.
- 5. Combine different models, experiment with design.

# Justifications and restrictions of the method

## Justifications

- 1. Intuitive: our assumptions imply that there cannot be an analytic solution (given as a formula).
- 2. Theoretical: by Universal Approximation Theorem, neural networks can estimate any continuous function. Linear Regression provides us with BLUE estimates.
- 3. Practical: approach is applicable for big data, high frequency trading, and forecasting.

## Restrictions

- We face both time and resource limits in terms of acquiring data. It does not mean our data is bad, but there always can be better.
- We may not choose the best available model (since there are too many options). Moreover, each model has both advantages and disadvantages (assumptions, work speed, tendency to overfit).

# What researchers found out before us

- Neural networks perform better (more accurately) than the Black-Scholes model on the same set of features.
- Neural networks make predictions faster (compared to classical option pricing methods) on new data.
- It is useful to use neural networks that work with sequences.
- Transformer architecture outperforms other (less complex) neural network architectures.

| | Model | train-MSE | MSE | Bias | AAPE | MAPE | PE5 | PE10 | PE20 |
|---|---|---|---|---|---|---|---|---|---|
| Call | BS | 322.95 | 321.37 | -0.05 | 78.79 | 4.81 | 50.52 | 59.33 | 67.43 |
| | MLP1 | 23.71 | 24.00 | 0.01 | 24.49 | 2.12 | 61.04 | 68.39 | 74.33 |
| | MLP2 | 7.70 | 15.21 | 0.09 | 23.45 | 1.73 | 63.03 | 70.10 | 75.54 |
| | LSTM | 30.61 | 30.97 | 0.13 | 26.58 | 2.33 | 58.94 | 66.35 | 72.42 |
| Put | BS | 543.48 | 533.25 | 97.37 | 68.00 | 97.46 | 12.87 | 18.22 | 23.58 |
| | MLP1 | 15.65 | 15.66 | 5.03 | 43.73 | 18.48 | 30.46 | 40.51 | 51.13 |
| | MLP2 | 2.03 | 8.84 | 3.85 | 39.59 | 14.32 | 33.74 | 44.25 | 55.01 |
| | LSTM | 22.81 | 23.15 | 6.01 | 48.32 | 26.05 | 27.45 | 36.24 | 46.17 |

Table 1: Error metrics comparing MLP1 price and MLP2 equilibrium price with Black-Scholes prices. Note all metrics beside MSE are percentages

# Benefits and expectations

- Benefits of the research
  - We used a data-driven approach. Hence, features which indeed affect option pricing were identified.
  - We performed a large number of experiments (to assess models, their speed, and feature importance).
  - We used a significant number of additional features and also tested hypotheses.
  - We performed an open research (public data, available code, reproducible results).
- Our expectations
  - Comparatively high performance. It is unlikely that a benchmark will not be outperformed, since we have more features.
  - Scalability. In case of need for high-frequency forecasting of a large number of options, it will be very useful.
  - Ability to perform high quality forecasting (in combination with additional models).
  - Additional features will be useful (models will pay attention to them).

# Data overview

- Dataset was constructed using Yahoo Finance (open-source) and Python.
- A script has been written that collects fresh data on options for 35 US companies.
- The dataset was updated every day for over 2 months. Overall, it has a panel data structure and around 1.25 million observations.
- Around 35 additional features were created (lags, market capitalization, weight in S&P 500 index, etc.).
- Fortunately, the historical data related to underlying assets (stocks) was available on Yahoo Finance.

# Data preparation and additional features

- Time to maturity was calculated as the difference between the date of expiration and the date of data collection, divided by 365.
- We took historical data on stock prices of particular companies, and trading volumes (from Yahoo).
- We calculated the standard deviations of contracts (will be used as volatilities) and of stock prices.
- We decided to keep only contracts which occurred at least 11 times. Later, we eliminated contracts that were cheaper than $1.
- 10 lag features were created for the price of contracts, the price of underlying assets, and the trading volume.
- The features of the company's age, the company's market capitalization, and the company's weight in the index were added.

# Quality metrics for model evaluation

- Mean Squared Error (MSE) and Median Squared Error (MedSE):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2.$$

$$\mathrm{MedSE} = \mathrm{median}\left((y_i - \hat{y}_i)^2\right).$$

- Mean Absolute Error (MAE) and Median Absolute Error (MedAE):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|; \mathrm{MedAE} = \mathrm{median}\left(|y_i - \hat{y}_i|\right).$$

- Mean Absolute Percentage Error (MAPE) and MedAPE:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|.$$

$$\mathrm{MedAPE} = \mathrm{median}\left(\left|\frac{y_i - \hat{y}_i}{y_i}\right|\right).$$

# Baseline (Black-Scholes and Linear Regression)

*Table 1: Comparison of different models.*

| Model | MSE | MAE | MAPE | Time |
|---|---|---|---|---|
| Black-Scholes (implied) | 4681 | 17.66 | 0.33 | 3min 36s |
| Black-Scholes (calculated) | 136866 | 233.72 | 15.62 | Not available |
| Linear Regression (implied) | 12377 | 57.5 | 88.73 | 286 ms |
| Linear Regression (calculated) | 828245 | 445 | 132 | Not available |

- The Black-Scholes model with implied volatility was the best model.
- Ones volatility was changed from implied (not available on inference) to calculated, the Black-Scholes model deteriorated more.
- Linear Regression works much faster.
- Further data filtering was required, additional features were needed.

# Linear Regression (advanced modeling + hypotheses testing)

- Let the level of statistical significance equal to 5%.
- All features of the Black-Scholes model were statistically significant with p-value = 0.
- However, when we add other features, time to maturity is no longer significant on 5% level.
- Meanwhile, the adjusted $R^2$ increased from 0.462 to 0.997.
- The market capitalization and the company's share in the index turned out to be insignificant features at the 5% level.
- Almost all lag features turned out to be statistically significant with p-value = 0. The insignificance of the remaining ones is explained by strong multicollinearity.
- Linear Regression with additional features outperformed Black-Scholes on the same set of observations (MAPE = 2.94 and MAPE = 15.62, respectively).

# Further filtering + using a Gradient Boosting algorithm

As a result of additional data filtering, we had about 987,000 observations left (compared to 1.168 million). Dataset was divided into train, validation, and test samples. Quality metrics were calculated on the test sample:

*Table 6: Comparison of the results.*

| Model | MSE | MAE | MAPE |
|---|---|---|---|
| Linear regression | 19.21 | 1.99 | 0.11 |
| CatBoost | 61.2 | 2.87 | 0.073 |
| Black-Scholes | 164210 | 274 | 10.93 |

- Based on MSE and MAE, the best model was Linear Regression.
- Probably this means that the prices of options are determined rather not by complex nonlinear dependencies, but by the right features.
- Linear Regression: MedSE = 0.94, MedAE = 0.97, MedAPE = 0.018 (which means a good model).
- According to the MDI criterion, lag features were the most important.
- CatBoost paid little attention to many features that were earlier determined as statistically significant.

# Summary of modeling

- Machine Learning models can indeed work faster on inference than the Black-Scholes model.
- Implied volatility is calculated using a model very close to the Black-Scholes model (the Black-Scholes model demonstrated a relatively very good result with the presence of implied volatility).
- The Black-Scholes model works better than Linear Regression on the features that are available for the Black-Scholes model.
- Additional features significantly improve the quality of the model. Linear Regression with additional features turned out to be the best model in the context of our task.
- Not all the features were statistically significant.
- The use of more complex Machine Learning models did not lead to an improvement in quality metrics. However, Machine Learning was still needed to use additional features.
- Cleaning the data and calculating the median values can make the quality metrics more adequate (an adequate interpretation of the results will become possible).

# Plans for the future

- Improve the quality of data.
  - More observations (contracts, time period).
  - Higher frequency (every hour, every 5 minutes).
  - Additional features (recalculate volatility, quantification of market positivity).
- Try more models.
  - Neural networks (including advanced architectures).
  - Advanced ensemble (blending).
- Experiment with design.
  - Hierarchical regression (clustering + regression).
  - Individual model for each company / contract.
  - Forecast volatilities and stock prices.

# Conclusion

- We started with a brief introduction to the problem, and talked about the historical significance, advantages, and limitations of the Black-Scholes model (which is often a benchmark in option pricing).
- Based on the experience of other researchers, we found out that using neural networks allows us to get better and faster solutions with fewer unrealistic assumptions.
- Chosen strategy - the data-driven approach. We focus not on the complexity and architecture of the model, but on the data and generation of additional features that can affect the pricing of options.
- During the modeling, we tested a large number of hypotheses, and also tested three methods. Linear Regression model with additional features demonstrated the best results.
- Finally, we discussed further possible directions within the framework of our research, having previously touched upon the limitations.
- The code was made publicly available, and the dataset is about to be uploaded as well.

Thank you for your attention!

# Derivation (p. 1)

Dynamics of stocks are given as a Geometric Brownian motion:

$$dS_t = \mu S_t dt + \sigma S_t dZ_t. \tag{4}$$

Change in the price of an option $f(t, S_t)$ is described by Ito's Lemma:

$$df = (\frac{\partial f}{\partial t} + \mu S_t \frac{\partial f}{\partial S_t} + \frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial {S_t}^2})dt + \sigma S_t \frac{\partial f}{\partial S_t} dZ_t. \tag{5}$$

In order to eliminate risk, we apply the principle of dynamic hedging to construct a portfolio with deterministic dynamics:

$$\Pi = -f(t, S_t) + \frac{\partial f}{\partial S_t} S_t \Rightarrow d\Pi = (-\frac{\partial f}{\partial t} - \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 f}{\partial {S_t}^2})dt. \tag{6}$$

# Derivation (p. 2)

Since portfolio is deterministic, it must grow exactly at a risk-free rate, therefore, $d\Pi = r\Pi dt$. Combining equations and plugging everything into:

$$rf = \frac{\partial f}{\partial t} + rS_t \frac{\partial f}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 f}{\partial {S_t}^2}. \tag{7}$$

We obtained the Black-Scholes PDE. The Black-Scholes formula is a solution to this equation given the following boundaries:

- We never execute the option when the underlying asset costs nothing, therefore, $f(t, 0) = 0$.
- At maturity (T), the asset will cost $S_T$, and the strike price remains being $X$. If the deal is profitable ($S_T > X$), we will exercise the option, and the profit will be exactly $S_T - X$. Otherwise, we will not exercise the option, so its value will be 0. Hence, $f(T, S_T) = max(S_T - X, 0)$.
- If $S_t \to +\infty$, the probability of exercising the option approaches 1.

# Extensions of Black-Scholes

The assumption of constant volatility has been a subject of criticism. There were invented methods that model volatility as a function of time. One example is the family of stochastic volatility models.

**Heston model**:

$$dS_t = rS_t dt + \sqrt{V_t} S_t dZ_{1t}. \tag{8}$$

$$dV_t = k(\theta - V_t)dt + \sigma\sqrt{V_t} dZ_{2t}. \tag{9}$$

$V_t$ is the volatility of an asset price, while $\sigma$ is the volatility of the volatility. The Heston model does not have analytic solutions, and is thus solved numerically (for example, using Monte Carlo simulation).

Another possible approach is to use GARCH, where volatility is a deterministic function, which is estimated using historical data and parameters.