

SPX Volatility surface estimation

Ergun Kacar, Amey Kaloti, Yiyang Liu, Martin Molina, Dushyanth Sirivolu

December 2, 2023

Goal of the project

Estimate the volatility surface of SPX index options.

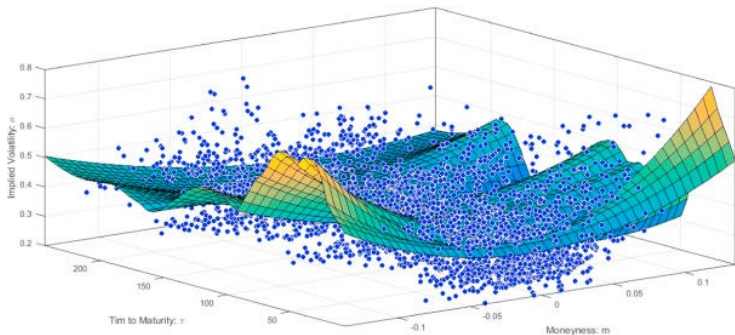


Figure: Picture from <https://quant-galore.medium.com/the-volatility-surface-everything-you-need-to-know-6f454e92ec9a>

So what is a volatility surface?

- Associated to a stock, there are two hedging instruments available to traders, call and put options.
- Call option gives a right, but not the obligation to buy the stock at a certain price (strike) by a certain date (expiry date/maturity date).
- There are numerous models to price options. Black-Scholes, Stochastic volatility, etc.
- All of them have an input called the volatility of the stock, which should be thought off as standard deviation of returns.
- Options are priced by the market, so we can invert the pricing model to get implied volatility of the option.

What is a volatility surface?

- Example: If stock is trading at \$100, a call option with strike of \$101 is \$1 out of the money, whereas call option with strike \$99 is \$1 in the money.
- How far away from the current stock price an option is, is called moneyness of the option.
- There are different ways to measure moneyness, the one which we use is called the delta.

What is a volatility surface?

- On any given trading day, for a stock if we plot moneyness on x -axis, time to expiry on y -axis and the implied volatility corresponding to the option on z -axis, we get a surface called volatility surface of the stock.
- Geometry of this volatility surface is constrained by so called arbitrage conditions. So not any random surface can be a volatility surface of a stock.
- Since options are only available for a discrete set of expiries and strikes, most trading firms have to complete this surface using various methods. Common method is to do an interpolation. Computationally time consuming.
- Alternate approaches have been proposed. One of them is to use machine learning.

Data Acquisition and Cleaning

- We got the data for options on SPX for days February 28, 2018 to February 28, 2023.
- Since the data contained a lot of unuseful features and missing values, we reduced the feature set to include only the relevant information. That is, the date of the option, the expiration date of the option, delta that indicates the moneyness, and the calculated implied volatility. We also add the variable time to expiry to the data set.
- We sorted all the relevant data according to the date, and we converted all the implied volatility to logarithmic scale to reduce skewness of the data.

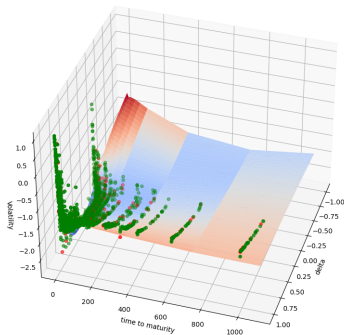
Training and Hyper-parameter Tuning

- We use 20 percent of the selected data as testing data. The train-test split is done time-wise to prevent data leakage.
- We tested three tree-based models (random forest, LGBM, and XGboost) and one neural network (Multilayer-Perceptron Regressor, MLPR) for the task.
- For the tree-based models, the optimal parameters are decided by a brute-force grid search on a smaller data set that consists of 15 percents of the training data. The smaller data set is constructed with stratified random sampling with respect to the dates of the data points.
- For the MLP Regressor with three hidden layers, the optimal parameters are chosen using a grid search on a small data set of the same size and constructed in the same manner.

Conclusion

MSE for these models were

- Linear Regression : 0.2943
- Random Forest: 0.1511
- Multi-layer Perceptron Regressor (MLPR): 0.08816669325587453
- LightGBM: 0.08622938826013876
- XGBoost: 0.08618972208865966



Limitations and future directions

- The size of the dataset meant we could only perform experiments on smaller datasets. Computational power was a big factor.
- The dataset contained only the end of day prices. To get a better idea of volatility surface we would like to have more data containing intraday prices of options as well.