

Volatility Surface Completion

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

November 2023

1 Introduction

Goal of this project is to model the volatility surface of S & P 500 index (SPX). Volatility surface is constructed using options on the stock. Options are derivative contract which derive their value from the underlying instrument, here SPX.

There are two types of options. A call option, gives a right, but not the obligation to buy the stock a certain price (strike price) by a certain date (expiry date). A put option gives a right, but not the obligation to sell the stock at a certain price by a certain date. Assuming efficient market hypothesis one can derive models to find a fair value of the option. There are various models for pricing options, the most well known being the Black-Scholes model. These days options on stocks are traded on open exchanges and hence are priced by market participants.

An option pricing model such as Black-Scholes has as inputs, the stock price, time to expiry, risk free interest rate and volatility of the stock. Out of the above mentioned inputs only volatility of the stock is an unobserved quantity. Since Black-scholes gives a closed form formula for the option price in terms of the inputs, one can invert this price to get the only unobserved quantity the volatility of the stock. This is called the implied volatility (IV). One can think of option price as the price of the volatility. It is common for traders to quote option prices in terms of their volatilities rather than actual dollar amount. So modelling and predicting implied volatility is an important task.

Given a stock, there are various different strike prices for which options are traded. For a put option, any strike below the current stock price is said to be out of the money, whereas any strike above the current stock price is said to be in the money. The opposite is true for the call options. This so called moneyness of the option also determines the price of the option.

For a given strike and given expiry date one has both the call and the put options at that strike. For each of them one can find their IV. Using a principle called the put-call parity, one can calculate IV of the call option from the IV of the put option at the same strike.

Depending on the stock there are various options with different expiry dates available for traders to trade on. Usually there are weekly expiry dates on stocks and at any given time, one can see expiry up to a couple of months.

The goal of this project was to model the implied volatility surface. This is a three dimensional surface that you get by plotting implied volatility against time to expiry, and the moneyness of option. This surface cannot take any arbitrary shapes since it is constrained by non arbitrage restrictions. We collected data for SPX option prices for February 28, 2018 to February 28, 2023 from a data provider OptionMetrics.

2 Stakeholders

- Trading firms.
- Investment banks.
- Financial institutions with a need to actively hedge their portfolio.

3 Model

1. We found out that a lot of variables in the dataset were unnecessary/redundant. In the end we used as features: last date, cp flag, strike price, best bid, best offer, volume, open interest, impl volatility, delta, gamma, vega, theta.
2. We found out that the strike prices were all multiplied by 1000. After digging in the documentation we realized that this is how OptionMetrics desimates this information.
3. Since implied volatility captures the price of an option to an extent we did not use price as one of the. Gamma, vega and theta are parameters associated to options. Mathematically they are described as follows: gamma is the derivative of delta with respect to stock price, vega is the derivative of option price with respect to volatility, theta is the derivative of option price with respect to time.
4. After experimenting with the data, we chose to use delta, and implied volatility provided by OptionsMetrics for our prediction.
5. We added a variable time to expiry for each option based on the trading day.
6. The dataset contains implied volatility computation for puts as well as calls. Data for calls seemed to be better behaved than for the puts. To estimate the volatility surface, implied volatility computation for either puts or calls should suffice due to put call parity from options theory. We decided to use calls to estimate the implied volatility surface.

7. We treated this first as a supervised learning problem and tried Linear Regression, XGBoost, Random Forest, and Light GBM algorithms on the data.
8. We used 20% of the data for validation.
9. We found through our analysis, that both XGBoost and Light Gradient Boosting machine performed equally well for this giving mean squared error of 0.08622938826013876 for Light GBM and 0.08618972208865966 for XGBoost on the dataset. These were the best models to use based on the data.