UNSW Interdisciplinary Partner Project

Predicting Realised Volatility

1 Outline

Volatility is one of the most prominent terms you'll hear on any trading floor and for good reason. In financial markets, volatility captures the amount of fluctuation in prices. High volatility is associated to periods of market turbulence and to large price swings, while low volatility describes more calm and quiet markets. For trading firms like Optiver, accurately predicting volatility is essential for the trading of options, whose price is directly related to the volatility of the underlying product.

The purpose of this project is to build a model that predicts future volatility. The following deliverables are required:

1. A *technical report* detailing the development of the model, its predictive validity, learnings gained and recommendations towards further model improvements.
2. All objectives condensed into a *slide deck* format suitable as a presentation for colleagues, etc.

2 Trading Theory

The Order Book

For any given financial instrument, there are buyers and sellers. The term *order book* refers to an electronic list of buy and sell orders for a specific security or financial instrument organized by price level. An order book lists the number of shares being bid on or offered at each price point.

Below is a snapshot of an order book of a stock (let's call it stock A). As you can see, all intended buy orders are on the left side of the book displayed as *bid* while all intended sell orders are on the right side of the book displayed as *ask*:

bid #	price	ask #
	151	196
	150	189
	149	148
	148	221
251	147	
321	146	
300	145	
20	144	

Note 1. *The terms ask and offer are used interchangeably to mean the sell price.*

An actively traded (or *liquid*) financial instrument always has a dense order book. As the order book data is a continous representation of market demand/supply it is always considered as the number one data source for market research.

Trades

An order book is a representation of trading intention on the market. However, the market needs a buyer and seller at the same price to make the trade happen. Therefore, sometimes when someone wants to do a trade in a stock, they check the order book and find someone with counter-interest to trade with.

For example, imagine you want to buy 20 shares of a stock A. Consider the order book given earlier. You need to find people who are willing to trade against you by selling 20 shares or more in total. You check the offer side of the book starting from the lowest price: there are 221 shares of selling interest on the level of 148. You can *lift* 20 shares for a price of 148 and guarantee your execution. This will be the resulting order book of stock A after your trade:

bid #	price	ask #
	151	196
	150	189
	149	148
	148	201
251	147	
321	146	
300	145	
20	144	

Note 2. *The word lift comes from the phrase "lift the offer" which means to buy the thing at the (lowest) price that is currently being offered by the market. The word we use for selling the highest bid is hit and this comes from the phrase "hit the bid".*

In this case, the seller(s) sold 20 shares and buyer bought 20 shares; the exchange will match the order between seller(s) and buyer and one trade message will be broadcast to the public:

20 shares of stock A traded on the market at price of 148.

Similar to order book data, trade data is also extremely crucial to Optiver's data scientists, as it reflects how active the market is. Actually, some commonly seen technical signals of the financial market are derived from trade data directly, such as high-low or total traded volume.

Optiver uses this data to optimise its activity as a market maker on the world's financial exchanges.

Market Making

Imagine, on another day, stock A's order book becomes "below shape", and you, again, want to buy 20 shares from all the intentional sellers. As you can see the book is not as dense as the previous one, and one can say, compared with the previous one, this book is less liquid.

bid #	price	ask #
	151	20
	150	12
	149	1
	148	
5	147	
2	146	
	145	
16	144	

You could insert an order to buy at 148. However, there is nobody currently willing to sell to you at 148, so your order will be sitting in the book, waiting for someone to trade against it. If you get unlucky, the price goes up, and others start bidding at 149, and you never get to buy at all. Alternatively, you could insert an order to buy at 155. The exchange would match this order against the outstanding sell order of one share at 149, so you buy 1 lot at 149. Similarly, you'd buy 12 shares at a price of 150, and 7 shares at 151. Compared to trying to buy at 148, there is no risk of not getting the trade that you wanted, but you do end up buying at a higher price.

You can see that in such an inefficient market it is difficult to trade, as trading will be more expensive, and if you want quality execution of your orders, you need to deal with higher market risk.

This is why investors love liquidity, and market makers like Optiver are there to provide it, no matter how extreme the market conditions are.

A *market maker* is a firm or individual who actively quotes two-sided markets in a security, providing bids and offers along with the market size of each. As a market maker will show both bid and offer orders, an order book with the presence of market maker will be more liquid, therefore a more efficient market will be provided to end investors to trade freely without concern on executions.

Making a market involves providing both a buy and a sell price for an instrument.

Order Book Statistics

There are a lot of statistics we can derive from raw order book data to reflect market liquidity and stock valuation. These stats are proven to be fundamental inputs of any market prediction algorithms. Below we would like to list some common stats; you might find that these are valuable signals that you can get from the order book data. You might come up with some of your own during this project too!

Let's come back to the original order book of stock A:

bid #	price	ask #
	151	196
	150	189
	149	148
	148	221
251	147	
321	146	
300	145	
20	144	

Bid/Ask Spread

As different stocks trade on different level on the market we take the ratio of best offer price and best bid price to calculate the BidAskSpread.

The formula for BidAskSpread can be written in below form:

$$\text{BidAskSpread} = \frac{\text{BestOffer}}{\text{BestBid}} - 1$$

Exercise 1. *What does a BidAskSpread of 0.5 mean? What about 0.02?*

Weighted Average Price (WAP)

The order book is also one of the primary source for stock valuation. A fair book-based valuation must take two factors into account: the level and the size of orders. In this competition we used weighted averaged price, or WAP, to calculate the instantaneous stock valuation and calculate realized volatility as our target.

The formula of WAP can be written as below, which takes the top level price and volume information into account:

$$\text{WAP} = \frac{\text{BidPrice}_1 \cdot \text{AskSize}_1 + \text{AskPrice}_1 \cdot \text{BidSize}_1}{\text{BidSize}_1 + \text{AskSize}_1}$$

As you can see, if two books have both bid and ask offers on the same price level respectively, the one with more offers in place will generate a lower stock valuation, as there are more intended seller in the book, and more seller implies a fact of more supply on the market resulting in a lower stock valuation.

Note that in most of cases, during the continuous trading hours, an order book should not have the scenario when bid order is higher than the offer, or ask, order. In another word, most likely, the bid and ask should never be in cross.

In this competition the target is constructed from the WAP. The WAP of the order book snapshot is about 147.53.

Exercise 2. *Consider a stock that currently has a best bid of 100 and a best offer of 105. There are 20 lots bid at 100 and 5 lots offered at 105. What is the BidAskSpread and the WAP of this order book?*

Exercise 3. *What is a big problem with the current definition of WAP? Hint: Think about how you can manipulate the WAP of the previous example by inserting only a single 1-lot order.*

Log Returns

How can we compare the price of a stock between yesterday and today?

The easiest method would be to just take the difference. This is definitely the most intuitive way, however price differences are not always comparable across stocks.

For example, let's assume that we have invested \$1000 dollars in both stock A and stock B and that stock A moves from \$100 to \$102 and stock B moves from \$10 to \$11. We had a total of 10 shares of A (\$1000/\$100=10 shares) which led to a profit of

$$\$10 \cdot (\$102 - \$100) = \$20$$

and a total of 100 shares of B that yielded \$100. So the price increase was larger for stock A, although the move was proportionally much larger for stock B. Price difference is clearly the wrong measure.

We can solve the above problem by dividing the move by the starting price of the stock, effectively computing the percentage change in price, also known as the stock return.

In our example, the return for stock A was

$$\frac{\$102 - \$100}{\$100} = 2\%$$

while for stock B it was

$$\frac{\$11 - \$10}{\$10} = 10\%.$$

The stock return coincides with the percentage change in our invested capital.

Returns are widely used in finance, however *log returns* are preferred whenever some mathematical modelling is required. Let S_t be the price of the stock S at time t. We define the log return between t_1 and t_2 as:

$$r_{t_1,t_2} = \log\left(\frac{S_{t_2}}{S_{t_1}}\right).$$

Note 3. *Here,* log *refers to the logarithm with base* e *where* e *is the exponential number. This log is sometimes written as* ln *and referred to as the natural log.*

Usually, we look at log returns over fixed time intervals, so with 10-minute log return we could set the notation $r_t = r_{t-10\text{minutes},t}$.

Log returns present several advantages, for example:

- They are additive across time: $r_{t_1,t_2} + r_{t_2,t_3} = r_{t_1,t_3}$.
- regular returns cannot go below -100%, while log returns are not bounded.

Note 4. *Note that taking the log essentially pulls out the percentage change in the stock (this is rough and only works for small moves in the stock). For example, if a stock was trading at $2 yesterday and is trading at $2.10 today, then there has been a 5% increase in the stock. We can get this in an explicit way by taking the log-return*

$$\begin{aligned} \log\left(\frac{2.10}{2}\right) &= \log(1.05) \\ &= \log(1 + 0.05) \\ &\approx 0.05 \end{aligned}$$

where the last line follows from the Taylor approximation $\ln(1 + x) \approx x$ *which holds for small* x*.*

2.1 Volatility

Realised Volatility

When we trade options, a valuable input to our models is the standard deviation of the stock's log returns. The standard deviation will be different for log returns computed over longer or shorter intervals, for this reason it is usually normalized to a 1-year period and the annualized standard deviation is called *volatility*.

For this project, you will be given 10 minutes of book data and we want you to predict what the volatility will be in the following 10 minutes. Volatility will be measured in the following way.

Calculating Realised Volatility

We will compute the log returns over all consecutive book updates and we define the realized volatility σ as the square root of the sum of squared log returns:

$$\sigma = \sqrt{\sum_t r_{t-1,t}^2}.$$

There are a few notes we should make here:

- In the above $r_{t-1,t}$ refers to the log return between two consecutive order book updates and *not* a one minute return. So, over the entire 10 minutes, you will look at all changes to the order book. You will compute the log return between any two consecutive order updates. You will square each of these log-returns, add them together and then take the square root.
- You can use the order book WAP for the stock price S_t.
- For those with some financial knowledge, we are keeping things simple. We are not annualizing the volatility and we are assuming that log returns have 0 mean.

Predicting Realised Volatility

For this section, we will use some notation. We write σ_{time} to mean the historical volatility over whatever time is specified. We write $\hat{\sigma}_{\text{time}}$ to mean our *estimate* for future volatility over the next whatever time is specified.

So, let's assume that you have just computed the realised volatility $\sigma_{10\text{ minutes}}$ and you want to generate $\hat{\sigma}_{10\text{ minutes}}$. You might take the schoolboy approach and just set

$$\hat{\sigma}_{10\text{ minutes}} = \sigma_{10\text{ minutes}}.$$

That is, your prediction for volatility for the next 10 minutes is precisely what you've just observed the last 10 minutes.

This is an OK starting point. But, maybe, the previous 5 minutes is a better estimate for the next 10 minutes:

$$\hat{\sigma}_{10\text{ minutes}} = \sigma_{5\text{ minutes}}.$$

Now we have an estimate that is perhaps more responsive but is more forgetful. Perhaps we need to do something like this:

$$\hat{\sigma}_{10\text{ minutes}} = 0.8 \cdot \sigma_{5\text{ minutes}} + 0.2 \cdot \sigma_{10\text{ minutes}}.$$

Now our prediction remembers the last 10 minutes but weights heavily (90% - don't forget that the last five minutes is included in the last ten minutes) towards the last 5 minutes.

This is all a bit naive, but at least it's intuitive. You want to build a sensible model! Sometimes, when we use the latest and greatest techniques, we lose our simple ideas. I would suggest coming up with a way that captures the behaviour of volatility and build your ideas and model around that. You have the entire orderbook at your disposal!