Liquidity in the Options Market: A Machine Learning Perspective

Brian Ward[*]

January 1, 2021

Abstract

We explore the liquidity in the options market.

Contents

---
[*]Email: bmw2150@columbia.edu. Corresponding author.

1   Introduction

There are many well-known strategies that utilize index options. In paritcular, there is the well-known volatility risk premium (VRP) that attempts to harvest the fact that options are typically overpriced relative to historical volatility. The strategy profits consistently when these options expire out of the money. On the other end of the spectrum are the strategies that are long volatility, namely tail risk hedging (TRH) strategies. Such strategies seek to pay off heavily in the event of a market downturn, thereby cushioning any drawdown in times of crisis.

In fact, the COVID-19 pandemic seemed to spark a huge debate about the efficacy of both strategies, with either side completely disagreeing with the other. Naturally, those invested in TRH cite the enormous profitability they experienced in March and April of 2020[1]. Critics of TRH counter that the long-term evidence in favor of the VRP indicates that holding out-of-the-money put options is too expensive and cannot be profitable over the long term[2]. In fact, because of this evidence, one pension fund[3] liquidated their TRH *just* prior to the mid-to-late March period when volatility spiked and the turmoil began in the US. Unfortunately, this meant they spent many years holding a losing strategy only to exit just before it would have paid off. Still there are others who defend TRH, saying that proponents of the VRP over-emphasize using short term, slightly out-of-the-money options simply because they are more liquid[4].

Although this debate is fascinating, our purpose here is not to explore or build such strategies. Instead we are focused on the scalability of them in the SPX options market. In particular, we seek to understand the drivers of liquidity metrics in the options market. The debate above comes down to how one might try to implement a TRH. In particular, they differ on how far out-of-the-money and how far away in time the expiry of these options should be. The successful implementation of TRH above used very long-dated and deep out-of-the-money put options, while the unsuccessful one used very short-dated and only slightly out-of-the-money put options. The latter is effectively the opposite of the VRP, where one simply sells short (rather than purchases) the same basket of options. We believe the difference comes down to whether one can actually implement the strategy through the availability in the market and to that end, this paper is about measuring liquidity in the SPX options market. The main goal is to be able to describe the differences in liquidity across the options chain so that a portfolio manager can make an informed decision about which options are available to protect the downside of the portfolio. Our approach uses machine learning for reasons that will be clear in the following paragraphs.

Whether or not you can apply machine learning in the financial industry is a somewhat hotly debated topic as well. The reason for that is the users of machine learning often try to forecast future stock returns and that in and of itself is difficult, regardless of the researcher's choice of a machine learning model or a simple linear model. Thus, using machine learning can easily lead one to overfit the return forecasts and the out-of-sample viability of the model is called into question. On the other hand, liquidity metrics are somewhat persistent and have intuitive reasons to be linked to a host of predictor variables. For example, in equities trading, around earnings announcements, there is a pronounced spike in volume. An intuitive explanation for that is investment managers who rank companies based on earnings per share will have to rebalance their portfolios in order to meet their target holdings based on the strategy's requirements along with the new information that's revealed when earnings are announced. In a similar way, other marketwide events, changes in macroeconomic variables, etc. all have an intuitive reason to be linked to volume: they all could generate a change in an investment manager's signal and thus, cause them to trade.

For the options market studied here, the key features of time to expiry and moneyness certainly play a key role. Although it is reflected in the price (to an extent), the longer dated options provide more protection and that is potentially something to cause volume to shrink (vs. a shorter dated option). That is because market participants are not willing to take on risk for longer periods and

---

[1]See this Bloomberg article regarding the profitability of Universa Investments' TRH strategy.
[2]Researchers from AQR Capital Management attempted to create profitable TRH strategies to no avail.
[3]See this Institutional Investor article about CalPERS's liquidating its position in a TRH.
[4]See this Advisor Perspectives article where the author backtested TRH strategies with long-dated options.

would prefer to provide protection for a shorter period and re-evaluate. A similar story holds for moneyness. The deeper out-of-the-money an option is, the more protection the option seller promises so it is intuititive that near-the-money options are more actively traded and hence, more liquid. Again, moneyness is a key driver of the price of an option, but we believe these effects can also impact the liquidity profile of the options. That is because some of these risks (long time periods and large underlying price movements) are hard to price into the options effectively because they are somewhat unknown or hard to measure.

In any case, it is clear that there are some intuitive drivers of liquidity in the options market, but we face a situation where the exact form of that relationship is unclear. Moreover, the non-linear interactions between these variables are completely unknown, but one could certainly argue for their existence at an intuitive level as well[5]. In other words, there clearly exist rules relating the input predictors to the output variable, but it seems incredibly difficult to actually write down those rules, let alone derive them from first principles. It is in this setting where machine learning appears to shine strongest. On the other hand, there is still the potential to overfit the relationship by using an overly complicated model. That is why we prefer to apply machine learning when the datasets are large. As we show, daily options data for even a single underlying over 20 years leads to a dataset of millions of datapoints. It is the confluence of this massive dataset and the fact that there should be rules governing the traded volume in options that justifies our approach in this paper.

2 Data Overview

3 Basic Relationships

Univariate relationships between liquidity. Decile plots of average/median volume by buckets of some sorting variables.

4 Machine Learning Model

5 Conclusion

---

[5]I.e. the effect of moneyness on a long dated option could in theory be smaller, larger, of even the opposite sign of the effect of moneyness on a short dated option.