

Liquidity in the Options Market: A Machine Learning Perspective

Brian Ward*

November 19, 2020

Abstract

We explore the liquidity in the options market.

Contents

1	Introduction	2
2	Data Overview	2
3	Basic Relationships	2
4	Machine Learning Model	2
5	Conclusion	2

*Email: bmw2150@columbia.edu. Corresponding author.

1 Introduction

There are many well-known strategies that utilize index options. In particular, there is the well-known volatility risk premium that attempts to harvest the fact that options are typically overpriced relative to historical volatility.

Our purpose here is not to explore 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.

Whether or not you can apply machine learning in the financial industry is a somewhat hotly debated topic. 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 some 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 holdings in order to meet their target holdings based on the strategy. In a similar way, variables like marketwide events, changes in macroeconomic variables, etc. all have an intuitive reason to be linked to something like volume: they all could generate a change in an investment manager's signal and thus, cause them to trade.

However, the exact form of the 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. In other words, there clearly exist rules relating the input predictors to the output variable, but seems incredibly difficult to actually write down those rules. 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 massive. 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 combined with 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