Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis Quantitative Finance

Thesis pre-proposal

Liu Nuo Su (510846)

Supervisor:	Dr. M. Grith
Second assessor:	TBD
Date final version:	10th April 2024

Contents

1 Introduction

The market for ultra short tenor has grown tremendously in recent years. Since the introduction of the so-called 'weeklies', investors and traders have gained the ability to hedge against, or take advantage of short-term market trends and specific events without committing to longer-term contracts. In 2022, it is estimated that around 60% of the daily option trading volume belonged to the ultra short term tenor options.

Although making up the majority of the trading volume, limited research has been done on ultra short tenor options. For longer maturity derivative, researches modeling the Implied Volatility Surface (IVS) were often centered around the fundamentals of a company and the macro-economic trends at the time (e.g. Yan (2011), Bernales and Guidolin (2014)). For those derivatives, out-of-the money options are less affected by sudden market events, as volatility tends to stabilize over longer time horizons. In contrast, ultra short tenor options tend to expire within immediate market conditions, where the sudden volatility spike often ends up determining the payoff of the derivative. Therefore, ultra short tenor options are much more affected by immediate market conditions, making the predictability of the derivative much more challenging.

In this work, we investigate the implied volatility of ultra short tenor options, and plot the options against different strike prices and times to maturity using the IVS. We employ machine learning algorithms, as they are suited for capturing complex, non-linear relations in the process of implied volatility estimation. Furthermore, as quarterly numbers do not immediately reflect the current market conditions, we incorporate news data obtained from The New York Times as explored by (Kim, Kim & Choi, 2023) as input. We make use of the FinBERT language model developed by (Yang, Uy & Huang, 2020), and is specialized to obtain the semantics of the text, use the embedding, and transform them into a format compatible for machine learning for Financial news data.

From an academic standpoint, we contribute to literature investigating ultra short tenor options. We demonstrate the use of text, and we employ state of the art machine learning techniques in the context of short tenor derivatives. From an economic standpoint, the IVS is used to gauge the market sentiment at the time, and the inclusion of text data can reveal the relation between specific news events and the sentiment at the time.

In this work, we aim to answer the following questions: 1. How useful is news data in the prediction of implied volatility surfaces of ultra short tenor options?

2. To what extend can machine learning algorithms predict the implied volatility surface of ultra short tenor derivatives?

3. What differences exist between the estimation of individual stock options and the S&P500 index options?

[4. (Ideally) we want to make an improvement on one of the machine learning algorithms, too, and see if it is superior to the other algorithms] [Short answer to questions]

In section 2, previous works on the IVS and short tenor options are investigated. Section 3 gives an overview of the data, while 4 discusses the methods used in this work.

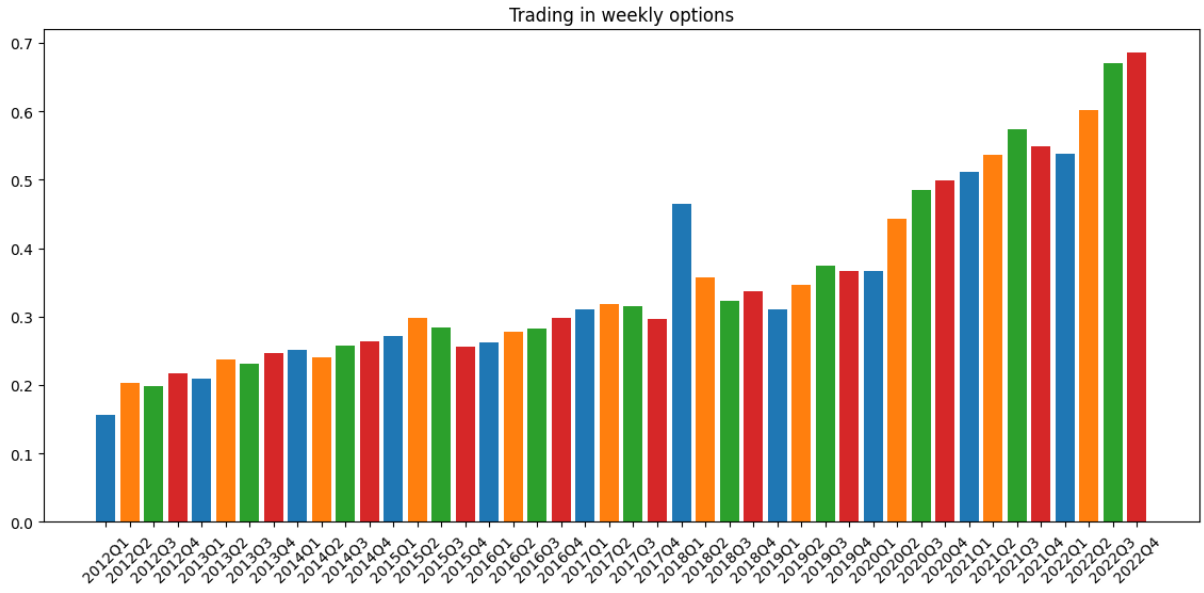

Figure 1: *The number of contracts traded in Weekly SPX options (with a tenor less or equal to 5 trading days) per quarter as a fraction of total volume for options with a maturity up to one year. The data covers the period from 2012 to 2022. Data source: OptionMetrics*

2 Literature Review

[Describe gap in literature - weekly options not researched much]

Contribution to literature:

1. Investigate the IVS on ultra short tenor options
2. Demonstrate the use of machine learning methods on the estimation of implied volatility
3. Explore the use of text data to gauge the sentiment around the stocks

3 Data

3.1 Ultra Short Tenor Options

Figure 1 shows the rising trend in the scale of weekly options trading. In the past few years, weekly options have made up the majority of total trading volume. There are several reasons that can explain this phenomenon: first, the past years were filled with economic downturns and periods of recovery and economic growth, which has led to increased demands for options, for both hedging and speculative purposes. Second, the federal reserve implemented policies of quantitative easing, allowing US citizens easy access to capital. This has led to a rise in the amount of retail investors, whom are also active in weekly options trading. Third, by 2022, the CBOE has expanded their zero-day-to-expiration (0DTE) options to every trading day of the week, which increased the volume traded in 0DTE options.

For our study, we consider the time period 2020 till 2022. The period is rich in economic events: with the pandemic prompting global supply chain disruptions and lockdowns, to the implementation quantitative easing by central banks, and reports of rising inflation.

[For firms specifically: oil prices, war, rise of AI, among others]

Company Name	Ticker	Sector
Apple inc	AAPL	Technology
NVIDIA	NVDA	Technology
Shell PLC	SHELL	Energy
Exxonmobil	XOM	Energy
Boeing	BA	Industrial

Table 1: Different stocks along with their ticker and sector, selected for our IVS model

Macro variable	Source
Gross Domestic Product	FRED
Unemployment rate	FRED
Inflation rate	FRED
Risk free Interest rate	FRED
M1 money supply	FRED
M2 money supply	FRED
Nominal house prices	[...]
...	

Table 2: Overview of the quarterly macro-economic data considered in the study

We obtain European options data using the database of OptionMetrics. For every trading day, we obtain the data of the ultra short tenor options by filtering for options with less or equal to 5 business days. We obtain the strike price, time to maturity, trading volume, best bid, and best ask price of the options. The mid price is then taken as a measure of the option price. We omit the options that are far in, or far out-of-the money, that show low trading volumes, as the price for those option are either the pay-off itself or close to 0, adding little information to the implied volatility. For the SPX options, we obtain a total of 823,118 observations. [Decide on a training, validation, and test set (initial idea: 70% train, 10% validation, 20% test, while retaining the time-series structure)]

In addition to SPX options, we consider several individual stocks in Table 2 to evaluate the effectiveness of our methods. We argue that the stocks chosen can be affected by news data, by both articles on the company itself, and sector-wide news. S&P500 first, then individual stocks

3.2 Macro Economic Variables and Events

News and text data - obtain amount of articles through the New York Times API (have to figure how to go smart about this, but searching using keywords will likely be the approach)

FinBERT - demonstrate the embeddings

3.3 Idiosyncratic Variables and News

In addition to the macro events, we consider firm-specific data for the estimation of the ultra short tenor option of the singular stocks. We obtain the daily stock price data using the CRSP database, and retrieve accounting variables by going through the quarterly report of each individual company. An overview of the accounting variables is given in Table 3

Singular stock news data; rumors and announcements can influence the volatility of the stock.

Variable
Total assets
Current assets
Cash flow
Total liabilities
Current liabilities
Sales
EBITDA

Table 3: Accounting variables considered for each stock

Examples: Apple supply chain disruptions, advancements in AI improves NVIDIA earnings, Court cases against Shell regarding climate change, reports of Boeing aircrafts falling apart during flights, announcements of Exxonmobil stumbling upon new oil fields. Obtain the text data from financial news (The financial times, or The New York times, (check other literature, what they have done))

[Provide overview of number of articles per stock, and their wordcount in a Table here]

4 Methods

4.1 Implied Volatility

[Implied volatility - how to obtain from option prices?]

4.1.1 Black Scholes

4.1.2 Carr and Wu model

Carr and Wu (2016) specified a model to predict the option price, and is defined as

$$\begin{aligned} \frac{dS_t}{S_t} &= \sqrt{v_t} dW_t, \\ d\sigma_t(K,T) &= \mu_t dt + \omega_t dZ_t, \\ \mathrm{E}[dW_t dZ_t] &= \rho_t dt, \end{aligned} \tag{1}$$

with $[\log(K/S_t)] < \kappa$ and $0 < T - t < \tau$. [work it out..]

4.2 FinBERT

4.3 Machine Learning

4.3.1 Lasso and Ridge regression

4.3.2 Ensemble Methods

4.3.3 Support Vector Machines

4.3.4 Neural Network

4.4 Evaluation Metrics

IVRSME, R-squared, shapley values for determining feature importance

Graphs and tables obtained will depend on which evaluation metrics we want to include

References

Bernales, A. & Guidolin, M. (2014). Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests. *Journal of Banking & Finance*, *46*, 326–342.

Carr, P. & Wu, L. (2016). Analyzing volatility risk and risk premium in option contracts: A new theory. *Journal of Financial Economics*, *120*(1), 1–20.

Kim, J., Kim, H.-S. & Choi, S.-Y. (2023). Forecasting the s&p 500 index using mathematical-based sentiment analysis and deep learning models: a finbert transformer model and lstm. *Axioms*, *12*(9), 835.

Yan, S. (2011). Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics*, *99*(1), 216-233. doi: https://doi.org/10.1016/j.jfineco.2010.08.011

Yang, Y., Uy, M. C. S. & Huang, A. (2020). Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.