Machine Learning-Based Modeling of Ultra Short-Term Implied Volatilities

1. Introduction

The price of an option is usually represented in terms of its implied volatility.[1] This one-to-one mapping is useful as implied volatilities are easier to compare in the cross-section (across options with different moneyness and tenor) and in the time series (across different dates). The implied volatilities plotted against moneyness and tenor constitute the so-called implied volatility surface (IVS), which conveys information about market expectations of the underlying future return distribution at different horizons.

The IVS plays an important role for practitioners and researchers. Institutional investors manage their options positions through implied volatility and often rely on models for the IVS (Carr and Wu, 2016). Moreover, option pricing models are usually estimated minimizing pricing errors in terms of implied volatility (Andersen, Fusari and Todorov, 2015), and more recently are being designed to fit directly the IVS (Ait-Sahalia, Li and Li, 2021). Therefore, accurately modeling and predicting the IVS is critical.

While many papers study how S&P 500 index options of relatively long maturities are priced (e.g., Goncalves and Guidolin, 2006; Andersen, Fusari and Todorov, 2015; Almeida et al., 2023), little is known about ultra short-term options expiring in less than a week, the so-called weeklies (Andersen, Fusari and Todorov, 2017). This is an important gap in the literature as these options are now the most traded contracts in the option market. More specifically, since the introduction by CBOE in May 2022 of weekly options expiring every day of the week, these contracts account for more than 60% of total trading volume (Bandi, Fusari and Reno, 2023; Almeida, Freire and Hizmeri, 2024). This is because investors are mainly interested in hedging against and/or betting on market movements over daily horizons. The tremendous growth experienced by ultra short-tenor options has made them a trending topic in financial media outlets and trader forums.

Given the new landscape of the option market, the goal of this topic is to expand our knowledge about the dynamics of ultra short-term implied volatilities.

[1] The implied volatility of a given option is the volatility parameter that makes the Black and Scholes (1973) formula deliver the observed option price.

2. Thesis Topic

This thesis topic consists of the general idea of using machine learning techniques to model the (dynamics of the) IVS of ultra short-term options. The motivation for these techniques is their flexibility to handle potentially high dimensional data and ability to approximate nonlinear relations. There are several possible directions to follow, including (but not restricted to) the examples below:

1. An assessment of how predictable ultra short-term IVs are, applying different machine learning methods (in the spirit of Gu, Kelly and Xiu, 2020) using a large set of covariates. The relative importance of different covariates can be analyzed. Investigating how predictability patterns compare to IVs of longer horizons would be interesting. Exploiting predictive information to construct profitable trading strategies would be another possibility.

2. Arguably, the latent factors driving ultra short-term IVs are different from those driving the more well-studied long-horizon IVs. Modeling the factor structure of the options across maturities using Principal Component Analysis techniques or parametric factor models (Goncalves and Guidolin, 2006; François et al., 2022) would be an interesting venue to pursue. A nonlinear alternative, in the spirit of Gu, Kelly and Xiu (2021), would be to use Autoencoders.

3. Options contain forward-looking information about investors' expectations of future market conditions. Therefore, ultra short-term IVs could be useful in predicting excess market returns or realized volatilities over daily horizons. Extracting this information using machine learning and comparing to popular prediction methods in the literature would be interesting.

3. Relevant data

Data on the implied volatilities and option prices of weekly options on the S&P 500 index (SPXW) and options of longer maturities are readily available from OptionMetrics, in WRDS (Wharton Research Data Services). Historical stock price data can be obtained from the Center for Research in Security Prices (CRSP), also in WRDS. A comprehensive macroeconomic database can be downloaded through the FRED (Federal Reserve Economic Data).

References

Ait-Sahalia, Y., Li, C., Li, C. X., 2021. Implied stochastic volatility models. The Review of Financial Studies 34, 394-450.

Almeida, C., Fan, J., Freire, G., Tang, F., 2023. Can a machine correct option pricing models? Journal of Business and Economic Statistics 41 (3), 995-1009, 2023

Almeida, C., Freire, G., Hizmeri, R., 2024. 0dte asset pricing. SSRN Working Paper.

Andersen, T. G., Fusari, N., Todorov, V., 2015. The risk premia embedded in index options. Journal of Financial Economics 117, 558-584.

Andersen, T. G., Fusari, N., and Todorov, V., 2017. Short-term market risks implied by weekly options. The Journal of Finance 72 (3), 1335–1386.

Bandi, F. M., Fusari, N., and Renò, R. (2023). 0dte option pricing. SSRN Working Paper.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637-654.

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

François, P., Galarneau-Vincent, R., Gauthier, G., Godin, F., 2022. Venturing into uncharted territory: An extensible implied volatility surface model. Journal of Futures Markets 42, 1912-1940.

Goncalves, S., Guidolin, M., 2006. Predictable dynamics in the S&P 500 index options implied volatility surface. Journal of Business 79, 1591–1635.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223–2273.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429-450.