Data-driven mathematical modeling of implied volatility surfaces

Shuaiqiang Liu, Cornelis Vuik

January 2023

1 Introduction

Implied volatility is an important quantity in finance (e.g., option pricing and risk management), which represents a specific measure of the future price uncertainty from the viewpoint of market practitioners. In the setting of European options, a holder has the right (but not the obligation) to exercise the contract at the maturity time T, and the fair price $V(t, S)$ can be determined by the Black-Scholes (BS) equation,

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0, \tag{1}$$

with stock price S, risk-free interest rate r, volatility σ. Given the pay-off function at T, for instance, an European put option $H^P(T, S) = \max(K - S, 0)$, where K is the strike price, the analytic solution exists for European options. Provided the option price V^{mkt} observed in the market, the Black-Scholes implied volatility σ^* is calculated by

$$BS(\sigma^*; S_0, K, T - t_0, r) = V^{mkt}. \tag{2}$$

where t_0, S_0, K are obtained from the financial contract and r is known. In other words, computing the implied volatility is to search for a volatility for the Black-Scholes pricing model, so that the model can provide an option price which exactly matches the observed option price.

A closed-form expression of the implied volatility in Eq. (2) is unavailable, and the paper [8] proposed a data-driven numerical method for fast computation of implied volatility. In reality, the implied volatility also varies over strike prices and time to maturity, see Figure 1, which forms so-called implied volatility surfaces $\sigma^* := \sigma^*(T - t, K)$. Volatility surfaces are widely used for pricing financial derivatives, as well as hedging and risk management. These surfaces have complex patterns, such as volatility smile/skew and term structure. Unfortunately, there are no explicit formulas to describe such volatility surface.

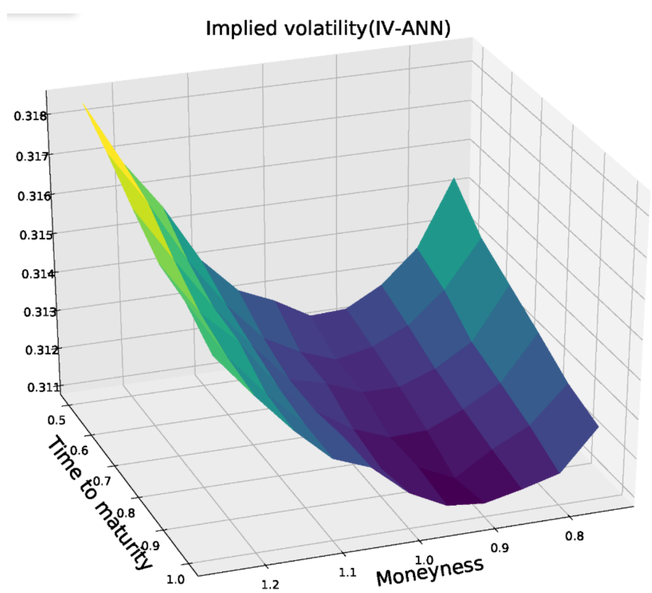

Figure 1: Implied volatility surface, from [8]. Moneyness equals S/K.

The approaches of modeling implied volatility surfaces generally fall into two categories. The first one is assuming the underlying asset follows specific dynamics [3] (e.g., Heston stochastic volatility model), here called mathematical modeling, which requires a calibration procedure [7] to recover the value of model parameters. The second approach is the data-driven modeling [1, 2, 9], where deep learning methods are used to treat the volatility surfaces in a non-parametric way. With the increasing complexity of volatility surfaces data (for example, Financial institutions frequently hold a large number of derivatives with various underlying assets, such as FX, stock, commodities, and so on), it has been becoming difficult to develop suitable assumption-based models, especially considering fast-moving markets. On the other hand, the data-driven non-parametric technique can manage massive data of volatility surfaces with flexibility and efficiency, but it lacks explainability and physical meanings.

This thesis will investigate a novel modeling approach to tackle the aforementioned challenge, i.e., data-driven mathematical modeling of implied volatility surfaces. We will start with the data-driven approach, and employ machine learning techniques (here generative models), without any assumptions, to extract as many as possible features (volatility smile, term structure, etc) from volatility surfaces data available. Next we will look into the ways to incorporate prior knowledge (e.g., arbitrage-free conditions) by combining the already learnt machine learning model with mathematical models (e.g., differential equations). Eventually we hope to apply the developed method for asset pricing and risk management, for example, generating synthetic volatility surfaces when data is missing.

2 Objectives

1. Literature review about modeling implied volatility surfaces and generative deep learning models, for instance, Diffusion probabilistic models [5], Variational Autoencoders [6] or Generative Adversarial Networks [4].

2. Training generative models to learn implied volatility surfaces, tuning hyper-parameters, providing best practices, etc.

3. Introducing appropriate mathematical tools to guarantee financial constraints (e.g. no arbitrage) and proper dynamics of volatility surfaces.

4. Exploring the interpretability of the developed data-driven mathematical models and generalizing the methodology (if feasible).

5. Application to option pricing and risk management, e.g., generating synthetic volatility surfaces, detecting anomalies.

References

[1] Bergeron, M., Fung, N., Hull, J., Poulos, Z., and Veneris, A. (2022). Variational autoencoders: A hands-off approach to volatility. *The Journal of Financial Data Science*, 4(2):125–138.

[2] Dierckx, T., Davis, J., and Schoutens, W. (2023). Towards data-driven volatility modeling with variational autoencoders. In *Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pages 97–111, Cham. Springer Nature Switzerland.

[3] Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*. John Wiley Sons 2006.

[4] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. *arXiv*, 1406.2661.

[5] Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *arXiv*, 2006.11239.

[6] Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv*, 1312.6114.

[7] Liu, S., Borovykh, A., Grzelak, L. A., and Oosterlee, C. W. (2019a). A neural network-based framework for financial model calibration. *Journal of Mathematics in Industry*, 9(1):9.

[8] Liu, S., Oosterlee, C. W., and Bohte, S. M. (2019b). Pricing Options and Computing Implied Volatilities using Neural Networks. *Risks*, 7(1).

[9] Ning, B., Jaimungal, S., Zhang, X., and Bergeron, M. (2021). Arbitrage-Free Implied Volatility Surface Generation with Variational Autoencoders. *arXiv*, 2108.04941.