
Deep Smoothing of the Implied Volatility Surface

Damien Ackerer
UBS
Zürich, Switzerland
damien.ackerer@epfl.ch

Natasa Tagasovska
Swiss Data Science Center
Lausanne, Switzerland
natasa.tagasovska@sdsc.ch

Thibault Vatter
Department of Statistics
Columbia University
New York, USA
thibault.vatter@columbia.edu

Abstract

We present a neural network (NN) approach to fit and predict implied volatility surfaces (IVSs). Atypically to standard NN applications, financial industry practitioners use such models equally to replicate market prices and to value other financial instruments. In other words, low training losses are as important as generalization capabilities. Importantly, IVS models need to generate realistic arbitrage-free option prices, meaning that no portfolio can lead to risk-free profits. We propose an approach guaranteeing the absence of arbitrage opportunities by penalizing the loss using soft constraints. Furthermore, our method can be combined with standard IVS models in quantitative finance, thus providing a NN-based correction when such models fail at replicating observed market prices. This lets practitioners use our approach as a plug-in on top of classical methods. Empirical results show that this approach is particularly useful when only sparse or erroneous data are available. We also quantify the uncertainty of the model predictions in regions with few or no observations. We further explore how deeper NNs improve over shallower ones, as well as other properties of the network architecture. We benchmark our method against standard IVS models. By evaluating our method on both training sets, and testing sets, namely, we highlight both their capacity to reproduce observed prices and predict new ones.

1 Introduction

The implied volatility surface (IVS) is a key input for computing margin requirements for brokers, quotes for market makers, prices of exotic derivatives for quants, and strategies positions for traders. As a result, tiny prediction errors can lead to dramatic financial losses. But standard IVS models often lack the ability to flexibly reproduce market prices and value other instruments without quotes. In this paper, we merge known ideas from different fields, namely machine learning (ML) and mathematical finance, to build a new solution for a non-trivial and relevant financial problem: the interpolation and extrapolation of the IVS. More specifically, we use a neural network (NN) to correct the IVS produced by any standard model. This lets practitioners plug our method on top of existing approaches while offering an unprecedented trade-off between flexibility and computational complexity. This problem was initially brought to us by a financial institution willing to improve its existing solution and a version of this method is being tested in production by a financial institution, where thousands of models are continuously updated throughout the day and used as inputs to various quantitative tools.

Practitioners have a growing interest to leverage NNs as flexible predictors for applications requiring an understanding of the complex dynamics of financial markets. And the ML community continues

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

to build better tools and understanding deep models. But leveraging domain expertise about a specific problem is often difficult, and still an active field of research. In this paper, we show how knowledge from both ML and mathematical finance can be merged to build a well performing and consistent hybrid model. Our application focuses on options, yet, the same approach could be similarly applied to other financial problems.

Problem setting and background context. An *option* is a financial contract giving the option holder the right to buy (a call option) or the right to sell (a put option) an asset, such as a stock or a commodity, for a predetermined price (the *strike price*) on a predetermined date (the *expiry* date). An initial *premium* must be paid to the option seller in order to acquire today the right to buy or sell an asset in the future at, possibly, a preferential price. The standard, textbook approach to model option pricing is based on the so-called Black-Scholes (BS) formula. The Black-Scholes formula provides a closed-form formula for option prices for a specific stock price model, the geometric Brownian motion (GBM). However, the formula builds on unrealistic assumptions such as continuous price trajectory and trading, absence of market frictions such as bid-ask spread and integer contract size, and normality of log-returns. Options are traded on financial exchanges, and their prices typically invalidate the GBM model. The Black-Scholes formula is a convenient one-to-one mapping between a price and a volatility parameter that is preferred by practitioners for a variety of reasons. For example, it allows them to easily compare the price of options with different contract strike prices and maturities. As a consequence, a lot of domain expertise has been developed for the so-called *implied volatility surface (IVS)*: the continuous representation of this volatility parameter expressed as a function of the strike price and of the expiry at a given point in time. However, only a finite number of option prices are observed in practice. In addition, quoted option prices should not allow *arbitrage opportunities*, that is, constructing a portfolio that may generate profits at a zero initial cost. Therefore the two main challenges when modeling the implied volatility surface are, first, to ensure that the corresponding option prices do not allow arbitrage opportunities and, second, the generalization to an entire surface given a limited number of observations. Such a construction allows the IVS to be used as input to price financial derivatives in a consistent way, and that enables effective risk-management. It is then worth noting that some commonly used models, such as SABR [26] and SVI [18], may not be arbitrage-free for some parameters values, see for example [48, Section 3], which contradicts real-life scenarios.

Short literature review. As neural networks and machine learning in general, prevail almost all aspects in science and industry, finance applications have also been impacted [30, 21, 31, 9, 29]. Deep learning have been studied with applications to option pricing in [49, 27, 17, 6, 37, 42, 43]. These papers exploit the well-known universal approximation property of neural networks [35, 34, 41]. Several applications of NN models for implied volatility smoothing exist, see [14, 53, 44, 38, 52] with more details in Appendix A. [53] in his PhD thesis also studied the use of soft constraints to train an arbitrage-free model. Extensive domain specific knowledge has been built on the IVS. In this work, we will use the no-arbitrage constraints on the IVS derived in [48], and later refined in [24], the moment condition of [40], and the surface SVI (SSVI) model of [20]. A more exhaustive literature review is available in Appendix A.1.

Summary of contributions. We present a new methodology to correct, interpolate, and extrapolate the implied volatility surface in an arbitrage-free way. We achieve this by modeling the implied total variance as a product of a neural network and a prior model, and by penalizing the loss using soft constraints during training so as to prevent arbitrage opportunities. The prior model should be a standard valid model for the total variance that will guide the general shape of the predictions, two examples are the Black-Scholes model and the SSVI model of [20]. The neural network is thus acting as a corrector of the prior model, enhancing its ability to reproduce observed market prices. The soft-constraints specification is guided by theoretical results from mathematical finance. The two elements together allow to build a realistic, flexible, and parsimonious model for the IVS. We benchmark our method against standard models and study its performance both on training and testing sets, as well as on synthetic data, and real market prices of contracts on the S&P 500 index. Numerical experiments suggest that our method appropriately captures the features of standard option pricing models. We show that increasing model capacity generally leads to better fits, and that the soft constraints generally help decreasing the fitting error and the convergence speed. Similar results are obtained when applying our method to real data, where an ablation study shows that constrained learning helps producing better volatility surfaces, both in interpolation and extrapolation.

Novelty and significance. Since options and related derivatives are actively exchange traded contracts that are used for risk-management and investment purposes, their fair valuation requires a reliable model for the IVS. The main ambition of this work is bridging the existing gap between a traditional challenge in finance and recent developments from the deep learning community, resulting in a trust-worthy, computationally efficient option pricing framework. This work is the first peer-reviewed published work to use soft-constraints to guide the training of a deep NN model for the IVS. Previous work used hardwired constraints which limits tremendously the neural network flexibility [14], or used constraints on the option price surface which requires additional transformations at each training step [44]. This work also appears to be the first suggesting a hybrid model for the IVS combining a neural network component, and a standard model component.

Paper structure. Section 2 reviews background knowledge such as the implied volatility surface, the no-arbitrage conditions, and formulates the modeling problem. We describe the methodology in Section 3. The numerical experiments and the empirical analysis can be found in Section 4. Section 5 discusses the broader impact. Section 5 concludes. More information on standard option pricing models, on implied volatility models, on the experiments, as well as additional results can be found in the online Appendix.

2 Background and objectives

2.1 The implied volatility surface

Let $\pi(K, \tau)$ denotes the market price of a call option with time to maturity $\tau > 0$ and strike price $K \geq 0$. Without loss of generality, we assume that the initial stock price is S , and that the interest-rate r and dividend yield δ are constants. Denote by $C(\cdot)$ the Black-Scholes formula, namely

$$C(S, \sigma, r, \delta, K, \tau) = S^{-\delta\tau} \Phi(d_+) - e^{-r\tau} K \Phi(d_-), \quad (1)$$

where $d_{\pm} = (\log(S/K) + (r - \delta)\tau) / (\sigma\sqrt{\tau}) \pm (1/2)\sigma\sqrt{\tau}$. More details on this formula can be found in Appendix A.2. The main objective of this paper is modeling the implied volatility whose definition is given below.

Definition 2.1 *The implied volatility $\sigma(k, \tau) > 0$ is given by the equation*

$$\pi(K, \tau) = C(S, \sigma(k, \tau), r, \delta, K, \tau), \quad (2)$$

with the (forward) log moneyness $k = \log(K/Se^{(r-\delta)\tau})$. The implied volatility surface is given by $\sigma(k, \tau)$ for $k \in \mathbb{R}$ and $\tau > 0$.

Interpreting the IVS shape. For a fixed $\tau > 0$, $\sigma(k, \tau)$ with $k \in \mathbb{R}$ defines a volatility smile. If the smile has a *U shape*, then the tail of the log return $\log(S_T/S_t)$ distribution are thicker than the tails of the Gaussian distribution, and vice versa. If the smile exhibits a skew, then one side of the log return distribution is thicker than the other. For example, if the left side of a smile, which is a slice of the surface for a fixed τ , is steeper than the right side, then the log price is more likely to experience large losses than large gains. The implied volatility surface provides a snapshot representation of valid option prices at a given time point. Although option prices fluctuate significantly over time, the shape and level of the implied volatility surface is fairly stable and large movements indicate important changes in market conditions.

2.2 Arbitrage-free surface

A static arbitrage is a static trading strategy that has a value that is both zero initially and always greater than or equal to zero afterwards, and a non-zero probability of having a strictly positive value in the future. In other words, an arbitrage costs nothing to implement while only providing upside potential, that is, it represents a risk-free investment after accounting for transaction costs. Under the assumption that economic agents are rational, any such opportunity should be instantaneously exploited until the market is arbitrage free. Therefore, option pricing models are designed in such a way that their call price surface $\pi(K, T)$ offers no possibility to implement such a strategy. Standard static arbitrage opportunities are described in Appendix A.3.

The absence of arbitrage translates into *constraints* on the call price surface $\pi(K, T)$, which in turn can be expressed as conditions that the implied volatility surface $\sigma(k, \tau)$ must satisfy [48, 20]. To express those conditions, we define the *total variance* of $\sigma(k, \tau)$,

$$\omega(k, \tau) = \sigma^2(k, \tau) \tau. \quad (3)$$

Proposition 2.2 Roper [48, Theorem 2.9] Let $S > 0$, $r = \delta = 0$, and $\omega : \mathbb{R} \times [0, \infty) \mapsto \mathbb{R}$. Let ω satisfy the following conditions:

- C1) (Positivity) for every $k \in \mathbb{R}$ and $\tau > 0$, $\omega(k, \tau) > 0$.
- C2) (Value at maturity) for every $k \in \mathbb{R}$, $\omega(k, 0) = 0$.
- C3) (Smoothness) for every $\tau > 0$, $\omega(\cdot, \tau)$ is twice differentiable.
- C4) (Monotonicity in τ) for every $k \in \mathbb{R}$, $\omega(k, \cdot)$ is non-decreasing, $\ell_{\text{cal}}(k, \tau) = \partial_\tau \omega(k, \tau) \geq 0$, where we have written ∂_τ for $\partial/\partial\tau$.
- C5) (Durrleman's Condition) for every $\tau > 0$ and $k \in \mathbb{R}$,

$$\ell_{\text{but}}(k, \tau) = \left(1 - \frac{k \partial_k \omega(k, \tau)}{2\omega(k, \tau)}\right)^2 - \frac{\partial_k \omega(k, \tau)}{4} \left(\frac{1}{\omega(k, \tau)} + \frac{1}{4}\right) + \frac{\partial_{kk}^2 \omega(k, \tau)}{2} \geq 0,$$

where we have written ∂_k for $\partial/\partial k$ and ∂_{kk} for $\partial^2/(\partial k \partial k)$

- C6) (Large moneyness behaviour) for every $\tau > 0$, $\sigma^2(k, \tau)$ is linear for $k \rightarrow \pm\infty$.

Then, the resulting call price surface is free of static arbitrage.

C1) and C2) are necessary conditions that any sensible model must satisfy. As for C3), it is merely sufficient to prove an absence of arbitrage when C4), C5), and C6) are also satisfied. Note that, assuming C3), C4) (respectively C5) and C6)) is satisfied if and only if the call price surface is free of calendar spread (butterfly) arbitrage [20]. C6) could be refined by imposing that $\sigma^2(k, \tau)/|k| < 2$ when $k \rightarrow \pm\infty$ to guarantee the existence of higher order implied moments, see [40, 7].

Remark 2.3 The Proposition 2.2 is derived under the assumption that $r = \delta = 0$ without loss of generality. Indeed, the static no-arbitrage constraints have the same functional forms as in C4)–C5)–C6) with non-zero parameters and the forward log moneyness k defined in Definition 2.1.

2.3 Problem formulation

Data availability. In practice, market data may need to be validated. This could be for a variety of reasons. First, exchange traded securities have at least two quotes, a bid and a ask price, and there is no guarantee that the average prices are arbitrage-free. Second, the observed prices may not be refreshed and thus not actionable, which may translate into notable input data noise. In addition, market data is typically sparse away from the money and dense close to the money. Indeed, far out of the money options are less likely to be exercised and are thus less likely to be used by their buyers. There is typically more demand and supply for contracts that are around the money.

Modeling objectives. The goal is to construct an IVS model $\sigma(k, \tau)$ that (I) generates options prices that are in line with market data, (II) is free of static arbitrage opportunities in the sense of Proposition 2.2, and (III) generalizes to unobserved data regions in a controlled fashion.

3 Methodology

3.1 Model and loss function

Explanatory and target variables. At a given time we observe triplets $(\sigma_i, k_i, \tau_i) \in \mathcal{I}_0$ where σ_i is the market implied volatility (the target/response), and (k_i, τ_i) are the log moneyness and the time to maturity (the features/explanatory variables). In addition, we complement the sample with synthetic pairs $(k_i, \tau_i) \in \mathcal{I}_{C45} \cup \mathcal{I}_{C6}$ used to control the arbitrage opportunities and the model asymptotic behavior (see Section 3.2).

Implied volatility model. Our model for the total variance and implied volatility is given by

$$\omega_\theta(k, \tau) = \omega_{\text{nn}}(k, \tau; \theta_1) \times \omega_{\text{prior}}(k, \tau; \theta_2) \quad \text{and} \quad \sigma_\theta(k, \tau) = \sqrt{\omega_\theta(k, \tau)/\tau} \quad (4)$$

for the parameters $\theta = \{\theta_1, \theta_2\}$, where ω_{nn} and ω_{prior} are the NN and prior models described below.

Prior model. $\omega_{\text{prior}} : \mathbb{R}^2 \mapsto \mathbb{R}$ is a prior model with parameters θ_2 . An implied volatility model without prior is obtained by setting $\omega_{\text{prior}} \equiv 1$ so that $\omega_\theta \equiv \omega_{\text{nn}}$. The prior model choice is useful to ensure that the model generalization is compliant with a prescribed preferred behavior. Essentially, the BS prior is a parameter-free model matching the implied volatility's at-the-money (ATM) term-structure. As for the SSVI prior, it improves on the BS model by capturing both the smile and the skew of the surface. Both models are described in Appendix B.1 and B.2 respectively.

NN model. $\omega_{\text{nn}} : \mathbb{R}^2 \mapsto \mathbb{R}$ is a standard feedforward multilayer neural network, namely

$$\omega_{\text{nn}}(k, \tau; \theta_1) = \bigcirc_{i=1}^{n+1} f_i^{W_i, b_i}(k, \tau) \text{ with } f_i(x) = \begin{cases} g_i(W_i x + b_i) & i < n + 1 \\ \alpha(1 + \tanh(W_{n+1} x + b_{n+1})) & i = n + 1 \end{cases} \quad (5)$$

with g_i an activation function, $\theta_1 = \{W_1, b_1, W_2, b_2, \dots, \alpha\}$ the set of weight matrices and bias vectors, α a scaling parameter letting ω_{nn} take values in $[0, \alpha]$, and n the number of hidden layers. The functional choice $1 + \tanh$ in the last layer is not critical, yet it appeared to perform slightly better than the sigmoid function on a limited number of trials. Note that one can initialize the NN to take output value one, that is no NN correction, by setting $\alpha = 1$ and $W_{n+1} = b_{n+1} = 0$.

Loss function. We fit the network parameters and prior parameters θ by minimizing the loss function

$$\mathcal{L}(\theta) = \mathcal{L}_0(\theta) + \sum_{j=1}^6 \lambda_j \mathcal{L}_{C_j}(\theta) \quad (6)$$

where the term $\mathcal{L}_0(\theta)$ is a prediction error cost, the terms $\mathcal{L}_{C_j}(\theta)$ for $j = 1, \dots, 6$ materialize soft constraints aiming to ensure that the shape of $\{\omega_\theta(k, \tau); (k, \tau) \in \mathbb{R} \times \mathbb{R}_+\}$ is indeed a sensible implied volatility surface, and λ_i for $j = 1, \dots, 6$ are the corresponding penalty weights. Note that some parameters of the prior model may also be calibrated.

We let the prediction error be the sum of the root-mean-squared-error (RMSE) and the mean-absolute-percentage-error (MAPE),

$$\mathcal{L}_0(\theta) = \sqrt{\frac{1}{|\mathcal{I}_0|} \sum_{(\sigma_i, k_i, \tau_i) \in \mathcal{I}_0} (\sigma_i - \sigma_\theta(k_i, \tau_i))^2} + \frac{1}{|\mathcal{I}_0|} \sum_{(\sigma_i, k_i, \tau_i) \in \mathcal{I}_0} |\sigma_i - \sigma_\theta(k_i, \tau_i)| / \sigma_i,$$

so as to penalize both absolute and relative errors. The IV values far out of the money can take values significantly larger than at the money. We found that the above loss function results in general a balanced allocation of prediction errors across the different moneynesses.

Remark 3.1 (Soft versus hard constraints) *An alternative approach to impose shape constraints on the mapping ω_θ is to hard-wire them into the neural network architecture, as in [14] for example. However, hard constraints are difficult to impose on multilayer neural networks, may reduce the neural network's flexibility, and may lead to more challenging leaning routines, see [46].*

3.2 No-arbitrage conditions and synthetic grid

We explain how each of the constraints/conditions in Proposition 2.2 can be handled either by refining the architecture of the neural network, or by adding a penalty term to the loss function (6). Note that the conditions **C1**–**C2** are in principle satisfied by design of the NN. The mapping ω_θ is twice differentiable as long as the activation functions g_i and g_{n+1} (as well as the prior model) are twice differentiable, in which case **C3** is satisfied. An example of valid activation function is the SoftPlus given by $\ln(1 + \exp(x))$ ¹. Hence we set $\mathcal{L}_{C1} \equiv \mathcal{L}_{C2} \equiv \mathcal{L}_{C3} \equiv 0$.

As for the other three constraints, we control for

- calendar arbitrages with $\mathcal{L}_{C4}(\theta) = 1/|\mathcal{I}_{C45}| \sum_{(k_i, \tau_i) \in \mathcal{I}_{C45}} \max(0, -\ell_{\text{cal}}(k_i, \tau_i))$,
- butterfly arbitrages with $\mathcal{L}_{C5}(\theta) = 1/|\mathcal{I}_{C45}| \sum_{(k_i, \tau_i) \in \mathcal{I}_{C45}} \max(0, -\ell_{\text{but}}(k_i, \tau_i))$,
- and the asymptotic behavior with $\mathcal{L}_{C6}(\theta) = 1/|\mathcal{I}_{C6}| \sum_{(k_i, \tau_i) \in \mathcal{I}_{C6}} |\partial^2 \omega_\theta(k_i, \tau_i) / \partial k \partial k|$,

where

$$\begin{aligned} \mathcal{I}_{C45} &= \{(k, \tau) : k \in \{x^3 : x \in [-(2k_{\min})^{1/3}, (2k_{\max})^{1/3}]_{100}\}, \tau \in \mathcal{T}\} \\ \mathcal{I}_{C6} &= \{(k, \tau) : k \in \{6k_{\min}, 4k_{\min}, 4k_{\max}, 6k_{\max}\}, \tau \in \mathcal{T}\} \\ \mathcal{T} &= \{\exp(x) : x \in [\log(1/365), \max(\log(\tau_{\max} + 1))]_{100}\} \end{aligned}$$

$k_{\min} = \min(\mathcal{I}_0^k)$, $k_{\max} = \max(\mathcal{I}_0^k)$, $\tau_{\max} = \max(\mathcal{I}_0^\tau)$, $[a, b]_x$ indicates an equidistant set of x points between a and b , and \mathcal{I}_0^τ and \mathcal{I}_0^k are the sets of unique time to maturity and forward log

¹We conjecture that one could also use ReLU activation functions with adjusted constraint conditions.

moneyness in \mathcal{I}_0 . Note that it should always be that $\min(\mathcal{I}_0^k) < 0$ and $\max(\mathcal{I}_0^k) > 0$. The motivation for the above transformations is to obtain a denser grid around the money and for short maturities. The particular parametric choices for the above sets do not appear critical as long as they are sufficiently dense and cover the regions of interest.

Note that the loss \mathcal{L}_{C6} guarantees that ω is asymptotically linear and should in principle be refined to control the coefficient value to be less than 2, as discussed in [40]. However \mathcal{L}_{C6} has better training performance and proved to be compliant with C6) in our post training verifications. Furthermore, we only work with finite moneyness in practice and imposing an asymptotic condition on large but finite moneyness values may unnecessarily constraint the model.

3.3 Model training

The training procedure and the parameters choice are described in details in Appendix C.1.

4 Results

The first part presents results on synthetic data where it is easier to study and compare model predictions. The second part presents results on real-world data to modeling implied volatility surfaces extracted from S&P500 options prices. The synthetic and market data are described in Appendix C.2 and C.3 respectively. We always set $\lambda_4 = \lambda_5 = \lambda_6 = \lambda$ for some $\lambda \geq 0$.

4.1 Numerical experiments

Smoothing behavior. Figure 1 displays the trained model on synthetic data with a strong penalty $\lambda = 10$. A dot indicates the positioning (k, τ) of an observation used for training on figures in the top row. The figures on the first and second rows show that σ_{prior} does not succeed to reproduce the training data, while σ_θ does so perfectly. The figures in the last row display σ_{prior} and σ_θ on an extended grid of log moneyness, and also shows the NN correction before the scaling by α (center figure). As σ_{prior} fits at the money (ATM) term structure, σ_θ makes little correction to it on the region $k \approx 0$. On the other hand, σ_θ makes important corrections to σ_{prior} for out of the money options.

The same figure for additional scenarios (λ value, prior model, and data limits) can be found in Appendix D.1. These additional results show that the model generalization is significantly better with a SSVI prior and with soft-constraints. Indeed, the IVS tends to flatten for large log moneyness values with a BS prior, whereas the present data is V shaped. Also, it is shown that the absence of soft-constraints can translate into non-realistic and possibly absurd out of sample predictions.

Losses and convergence.

Figure 2 displays statistics on the losses in (6) for trained models with different number of layers, neurons per layer, and penalty value λ . A total of 50 models with different random seeds for parameter initialization have been trained for 5'000 epochs for each configuration. Recall that, with $\lambda = x$, the weight on the soft constraints is x times larger than the loss on the predictions errors. We see that increasing the number of layers or of neurons per layers generally decreases the RMSE and MAPE. We also see that the arbitrage opportunities vanish as λ increases. The models prediction errors typically worsen as λ increases, but the effect appears stronger for the lesser flexible models. We also observe that, when the number of layers is smaller, increasing the neurons per layer without enforcing more strictly the constraints can generate more severe arbitrage opportunities. But such opportunities disappear completely when the absence of arbitrage penalty is given a higher weight. Additional results can be found in Appendix D.2 for the BS prior where, in particular, we observe that the prediction errors are worse for this choice of prior model.

Additional results. In Appendix D.3 we show that the NN model suggested in [14] cannot capture the IVS on its full range of maturities and moneynesses for the synthetic data. We also illustrate the impact of increasing arbitrage opportunities on the losses in Appendix D.4.

4.2 Empirical results

Smoothing market data. Figure 3 displays S&P500 options data and trained model predictions with a strong penalty $\lambda = 10$ for the April 13-th 2018. This market data contains several thousands observations, and the corresponding IVS has a fairly complex shape compared to the previous synthetic data. The first two rows highlight that σ_{prior} fails to reproduce the market data, but that

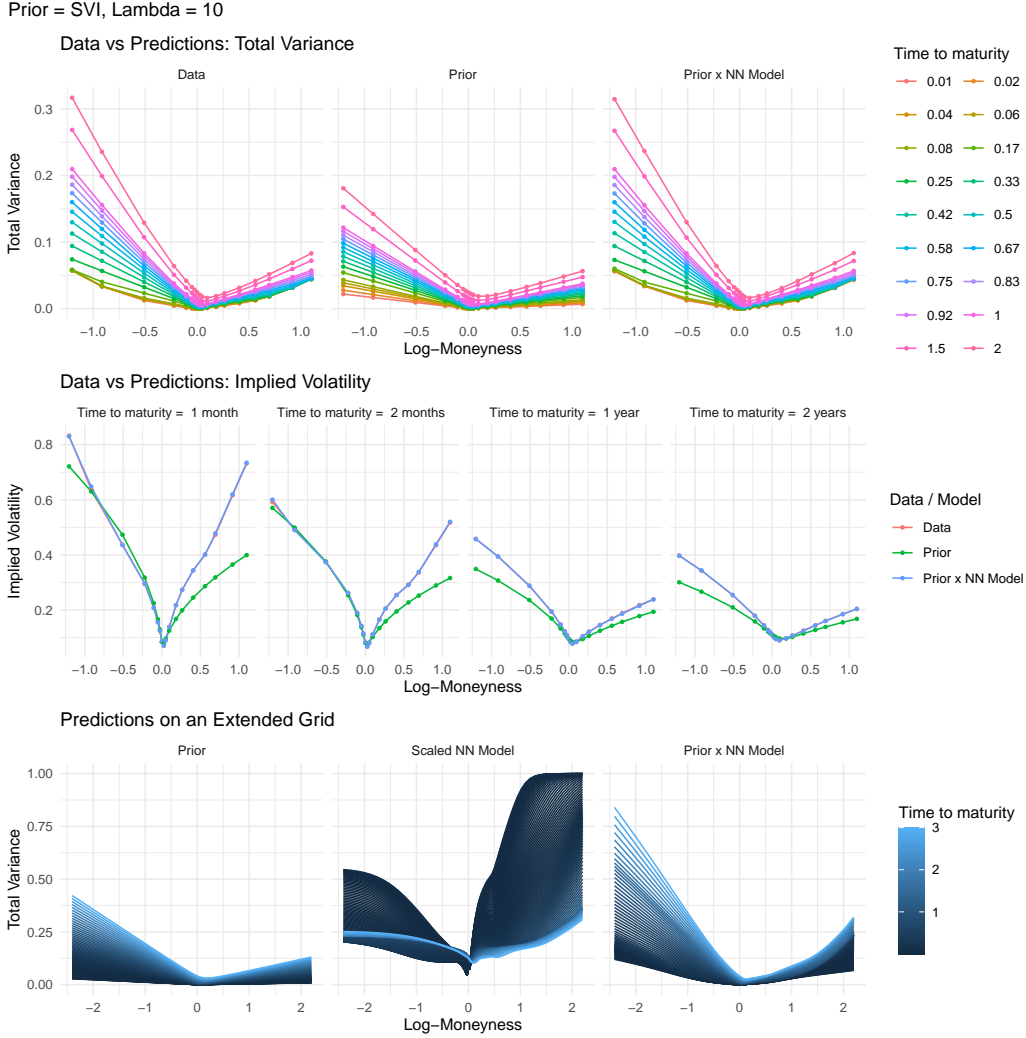


Figure 1: Synthetic data and trained model predictions for a specific configuration (scenario 12).

σ_θ is able to make accurate predictions. Notice that, because of data noise, the target IVS seem not to be arbitrage-free as some total variance slices cross each others, see upper-left figure. Yet, the predictions generated by σ_θ are perfectly smooth and the slices do not cross. The same figure for additional scenarios can be found in Appendix D.5.

Backtests. We perform the following exercise for each day in the sample. First, we split the daily sample into a training and a testing set. Second, we fit the model on the training set and evaluate its performance on the testing set. We use two different configurations for training and testing. In the interpolation setting, for each maturity, we randomly select half of the contracts. As such, we also sample options that are far out or in the money for training, and the testing error represents the approximation error for the range of moneyness that are actually observed. In the extrapolation setting, for each maturity, we select half of the contracts whose log moneyness is between the 10% and 90% of the log moneyness in the corresponding slice. This second filter therefore contains more observations around the money. Thus, we do not select options that are far out or in the money for training, and the testing error measures how well our model extrapolates. Finally, we again use three values for λ in order to study how the arbitrage-related penalties affect the results.

In Table 1 we present our results for model trained daily on the S&P500 options data between January and April 2018, and for the subprime crisis period between September and December 2008, with a SSVI prior. First, we describe the RMSE and MAPE. As expected, training errors are generally below testing errors. Increasing the no-arbitrage penalty λ leads to worse RMSE and MAPE metrics on the training set. However, larger λ also implies similar or even less RMSE and MAPE metrics on

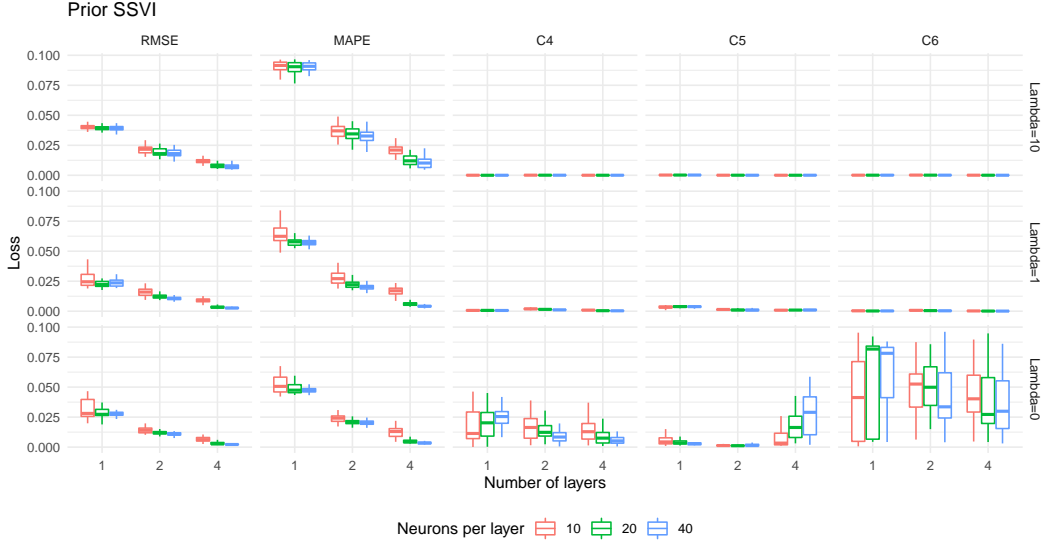


Figure 2: Losses for different number of layers, neurons per layer, and penalty value.

Table 1: Backtesting results for the SSVI prior (quantiles in %, Jan-Apr 2018 / Sep-Dec 2008)

Loss	λ	Interpolation						Extrapolation					
		Train			Test			Train			Test		
		q_{05}	q_{50}	q_{95}	q_{05}	q_{50}	q_{95}	q_{05}	q_{50}	q_{95}	q_{05}	q_{50}	q_{95}
RMSE	10	0.4 / 0.7	0.5 / 3.1	1.3 / 19.9	0.4 / 0.9	0.5 / 3.3	1.4 / 18.8	0.2 / 0.3	0.3 / 1.1	0.4 / 3.5	2.2 / 2.7	5.0 / 6.3	8.0 / 12.0
	1	0.3 / 3.0	0.4 / 6.8	1.0 / 13.9	0.4 / 2.9	0.5 / 7.1	1.3 / 13.5	0.2 / 1.9	0.2 / 4.4	0.4 / 10.6	3.7 / 5.0	6.6 / 10.6	11.7 / 20.5
	0	0.2 / 0.4	0.3 / 1.3	0.5 / 4.0	0.3 / 1.6	0.5 / 3.5	4.8 / 10.8	0.2 / 0.3	0.2 / 0.9	0.3 / 3.3	3.1 / 5.5	7.5 / 11.6	18.1 / 26.7
MAPE	10	0.5 / 0.9	0.7 / 2.1	1.2 / 17.4	0.5 / 1.1	0.8 / 2.4	1.2 / 18.5	0.4 / 0.5	0.6 / 0.9	0.9 / 2.1	1.2 / 2.4	1.7 / 3.3	2.4 / 6.5
	1	0.5 / 4.0	0.6 / 8.2	1.2 / 12.2	0.5 / 4.3	0.7 / 8.1	1.3 / 12.9	0.3 / 2.2	0.5 / 6.3	0.9 / 11.0	1.5 / 5.6	2.3 / 9.7	3.3 / 13.1
	0	0.4 / 0.5	0.6 / 1.0	0.9 / 1.8	0.5 / 1.2	0.7 / 1.9	0.9 / 3.2	0.3 / 0.3	0.5 / 0.7	0.8 / 1.6	1.5 / 4.3	2.2 / 7.4	4.8 / 14.3
C4	10	0.0 / 0.0	0.0 / 0.0	0.0 / 0.2	0.0 / 0.0	0.0 / 0.0	0.0 / 0.5	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 14.1
	1	0.0 / 0.0	0.0 / 0.0	0.2 / 0.1	0.0 / 0.0	0.0 / 0.0	0.2 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.1	0.0 / 0.0	0.0 / 0.0	0.1 / 0.1
	0	1.6 / 1.5	18.0 / 9.8	99+ / 99+	1.6 / 1.2	40.7 / 12.1	99+ / 99+	0.3 / 1.1	2.4 / 3.7	44.2 / 83.7	0.4 / 1.0	2.8 / 4.3	46.7 / 30.4
C5	10	0.0 / 0.0	0.0 / 0.0	0.0 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 1.2
	1	0.0 / 0.0	0.0 / 0.0	0.1 / 0.0	0.0 / 0.0	0.0 / 0.0	0.5 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.5 / 0.0
	0	2.6 / 99+	72.1 / 99+	99+ / 99+	99+ / 99+	2.6 / 99+	63.8 / 99+	99+ / 99+	2.8 / 99+	24.2 / 99+	99+ / 99+	1.7 / 99+	19.0 / 99+
C6	10	0.0 / 0.0	0.0 / 0.0	0.0 / 0.2	0.0 / 0.0	0.0 / 0.0	0.0 / 0.6	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 44.3
	1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.6 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
	0	0.8 / 0.0	48.2 / 1.7	99+ / 99+	0.8 / 0.0	78.4 / 0.4	99+ / 99+	0.3 / 0.1	12.0 / 7.5	99+ / 99+	0.0 / 0.0	2.4 / 0.0	99+ / 99+

the testing set. Note that the RMSE for testing set of the extrapolation appears to be large compared to the MAPE. The extrapolation errors are most of the time larger than interpolation errors. This suggest that the high RMSE value is likely to be caused by deep out of the money options with large IV values. Regarding C4)–C6), we see that the models resulting from $\lambda = 1, 10$ are essentially arbitrage-free. As for the model resulting from no enforcement of the constraints (i.e., $\lambda = 0$), it generates arbitrage opportunities both in interpolation and extrapolation. Note that 99+ indicates a value larger than 99 which can happen for arbitrage losses when $\lambda = 0$.

Additional results for the BS prior, as well as for baselines models (Bates and SSVI models), can be found in Appendix D.6. Our approach surpasses the baseline models in terms of prediction accuracy, which should not be a surprise. Indeed, in Section 4.1 we showed that σ_θ could reproduce the IVS from a typical Bates model, hence it is likely to be at least as flexible. In addition, σ_θ with a SSVI prior σ_{prior} is likely to perform better than the underlying baseline SSVI model.

Additional results. In Appendix D.7 we display the risk-neutral density and the local volatility corresponding to the scenario 12. The experience on Apple options in Appendix D.8 shows that our approach can fit complex shape surfaces, such as W-shaped smiles. We trained a model on the absolute IV errors weighted by the IV spread in Appendix D.9 which appears to converge faster but fails to learn the IVS shape far out of the money. The computational times for the empirical experiments are reported in Appendix D.10 and appear satisfying despite the experimental nature of the implementation.

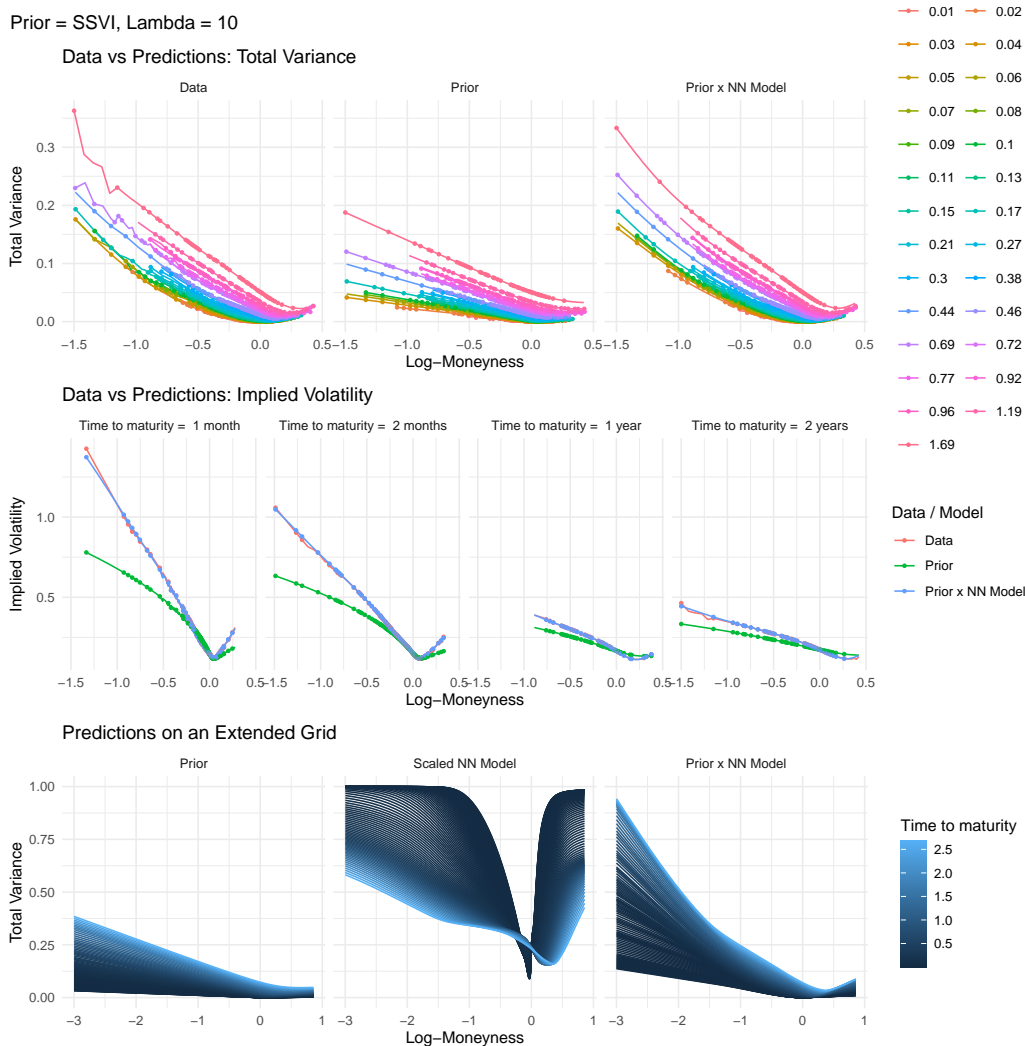


Figure 3: Market data and trained model predictions for a specific configuration (scenario 12).

5 Conclusion

We described a flexible methodology the price financial derivatives in an economically sensible way. This is achieved by modeling the implied volatility surface with a multilayer neural network and shaping it by penalizing the loss. We validate our approach with various numerical and empirical applications. The presented approach could be used as a building block to construct arbitrage-free models for multiple stocks, and for the IVS dynamics, as further discussed in Appendix E. Additionally, the loss function could be improved by avoiding to penalize models for predictions inside the spread, or using vega-weighting, as is commonly done by practitioners.

Broader Impact

Financial markets play a central role in our economy, and our work could be used to generalize in a robust way the information available to participants. Options are actively traded securities that can be used for multiple reasons such as protecting pension portfolios against future losses, hedging against future fluctuations in crop prices for agricultural farmers, or speculation for hedge funds. Options are also used by economist and financial expert to extract market implied sentiment measures such as the VIX “fear” index. The financial risk created by options is often held by financial intermediaries, such as banks or brokers, which need appropriate tools to monitor and control their financial risk.

Acknowledgments and Disclosure of Funding

The authors would like to thank Michael Roper for providing detailed comments and suggestions, as well as Serge Kassibrakis, Charles-Albert Lehalle, Johannes Wiesel, and participants at the 2019 SIAM conference on Financial Mathematics and Engineering in Toronto for helpful discussions.

The statements and opinions expressed in this article are those of the authors and do not represent the views of UBS (and/or any branches) and/or their affiliates.

References

- [1] *Abadi Martín, Barham Paul, Chen Jianmin, Chen Zhifeng, Davis Andy, Dean Jeffrey, Devin Matthieu, Ghemawat Sanjay, Irving Geoffrey, Isard Michael, others* . Tensorflow: A system for large-scale machine learning // 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). 2016. 265–283.
- [2] *Abi Jaber Eduardo, Larsson Martin, Pulido Sergio*. Affine volterra processes // The Annals of Applied Probability. 2019. 29, 5. 3155–3200.
- [3] *Ackerer Damien, Filipović Damir, Pulido Sergio*. The Jacobi stochastic volatility model // Finance and Stochastics. 2018. 22, 3. 667–700.
- [4] *Barndorff-Nielsen Ole E*. Processes of normal inverse Gaussian type // Finance and stochastics. 1997. 2, 1. 41–68.
- [5] *Bates David S*. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options // The Review of Financial Studies. 1996. 9, 1. 69–107.
- [6] *Becker Sebastian, Cheridito Patrick, Jentzen Arnulf*. Deep optimal stopping // Journal of Machine Learning Research. 2019. 20. 1–25.
- [7] *Benaim Shalom, Friz Peter*. Regular variation and smile asymptotics // Mathematical Finance. 2009. 19, 1. 1–12.
- [8] *Buehler Hans, Gonon Lukas, Teichmann Josef, Wood Ben*. Deep hedging // Quantitative Finance. 2019. 8. 1271–1291.
- [9] *Cao Li-Juan, Tay Francis Eng Hock*. Support vector machine with adaptive parameters in financial time series forecasting // IEEE Transactions on Neural Networks. 2003. 14, 6. 1506–1518.
- [10] *Carr Peter, Geman Hélyette, Madan Dilip B, Yor Marc*. The fine structure of asset returns: An empirical investigation // The Journal of Business. 2002. 75, 2. 305–332.
- [11] *Carr Peter, Madan Dilip*. Option valuation using the fast Fourier transform // Journal of Computational Finance. 1999. 2, 4. 61–73.
- [12] *Caruana Rich, Lawrence Steve, Giles C Lee*. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping // Advances in Neural Information Processing Systems. 2001. 402–408.
- [13] *Corlay Sylvain*. B-spline techniques for volatility modeling // Journal of Computational Finance. 2016. 19, 3. 97–135.
- [14] *Dugas Charles, Bengio Yoshua, Bélisle François, Nadeau Claude, Garcia René*. Incorporating second-order functional knowledge for better option pricing // Advances in Neural Information Processing Systems. 2001. 472–478.
- [15] *El Euch Omar, Rosenbaum Mathieu*. The characteristic function of rough Heston models // Mathematical Finance. 2019. 29, 1. 3–38.
- [16] *Fengler Matthias R*. Arbitrage-free smoothing of the implied volatility surface // Quantitative Finance. 2009. 9, 4. 417–428.
- [17] *Fujii Masaaki, Takahashi Akihiko, Takahashi Masayuki*. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs // Asia-Pacific Financial Markets. 2019. 26. pages391–408.
- [18] *Gatheral Jim*. A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives // Presentation at Global Derivatives & Risk Management, Madrid. 2004.

- [19] *Gatheral Jim*. The volatility surface: A practitioner's guide // Book, John Wiley & Sons. 2011.
- [20] *Gatheral Jim, Jacquier Antoine*. Arbitrage-free SVI volatility surfaces // *Quantitative Finance*. 2014. 14, 1. 59–71.
- [21] *Gençay Ramazan, Qi Min*. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging // *IEEE Transactions on Neural Networks*. 2001. 12, 4. 726–734.
- [22] *Glorot Xavier, Bengio Yoshua*. Understanding the difficulty of training deep feedforward neural networks // *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*. 2010. 249–256.
- [23] *Grasselli Martino*. The 4/2 stochastic volatility model: a unified approach for the Heston and the 3/2 model // *Mathematical Finance*. 2017. 27, 4. 1013–1034.
- [24] *Guo Gaoyue, Jacquier Antoine, Martini Claude, Neufcourt Leo*. Generalized arbitrage-free SVI volatility surfaces // *SIAM Journal on Financial Mathematics*. 2016. 7, 1. 619–641.
- [25] *Hagan Martin T, Menhaj Mohammad B*. Training feedforward networks with the Marquardt algorithm // *IEEE Transactions on Neural Networks*. 1994. 5, 6. 989–993.
- [26] *Hagan Patrick S, Kumar Deep, Lesniewski Andrew S, Woodward Diana E*. Managing smile risk // *The Best of Wilmott*. 2002. 1. 249–296.
- [27] *Han Jiequn, Jentzen Arnulf, Weinan E*. Solving high-dimensional partial differential equations using deep learning // *Proceedings of the National Academy of Sciences*. 2018. 115, 34. 8505–8510.
- [28] *Haykin Simon S*. *Neural networks and learning machines* // Book, Prentice Hall. 2009.
- [29] *Heaton James B, Polson Nick G, Witte Jan Hendrik*. Deep learning for finance: Deep portfolios // *Applied Stochastic Models in Business and Industry*. 2017. 33, 1. 3–12.
- [30] *Hernández-Lobato José Miguel, Hernández-Lobato Daniel, Suárez Alberto*. GARCH processes with non-parametric innovations for market risk estimation // *International Conference on Artificial Neural Networks*. 2007. 718–727.
- [31] *Hernández-Lobato José Miguel, Lloyd James R, Hernández-Lobato Daniel*. Gaussian process conditional copulas with applications to financial time series // *Advances in Neural Information Processing Systems*. 2013. 1736–1744.
- [32] *Heston Steven L*. A closed-form solution for options with stochastic volatility with applications to bond and currency options // *The review of financial studies*. 1993. 6, 2. 327–343.
- [33] *Hinton Geoffrey, Vinyals Oriol, Dean Jeff*. Distilling the knowledge in a neural network // Unpublished manuscript. 2015.
- [34] *Hornik Kurt*. Approximation capabilities of multilayer feedforward networks // *Neural Networks*. 1991. 4, 2. 251–257.
- [35] *Hornik Kurt, Stinchcombe Maxwell, White Halbert*. Multilayer feedforward networks are universal approximators // *Neural Networks*. 1989. 2, 5. 359–366.
- [36] *Hull John, White Alan*. The pricing of options on assets with stochastic volatilities // *The Journal of Finance*. 1987. 42, 2. 281–300.
- [37] *Hutchinson James M, Lo Andrew W, Poggio Tomaso*. A nonparametric approach to pricing and hedging derivative securities via learning networks // *The Journal of Finance*. 1994. 49, 3. 851–889.
- [38] *Itkin Andrey*. To sigmoid-based functional description of the volatility smile // *The North American Journal of Economics and Finance*. 2015. 31. 264–291.
- [39] *Kingma Diederik P, Ba Jimmy*. Adam: A method for stochastic optimization // *Proceedings of the 3rd International Conference on Learning Representations*. 2015.
- [40] *Lee Roger W*. The moment formula for implied volatility at extreme strikes // *Mathematical Finance*. 2004. 14, 3. 469–480.
- [41] *Leshno Moshe, Lin Vladimir Ya, Pinkus Allan, Schocken Shimon*. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function // *Neural Networks*. 1993. 6, 6. 861–867.

- [42] *Liu Shuaiqiang, Borovykh Anastasia, Grzelak Lech A, Oosterlee Cornelis W.* A neural network-based framework for financial model calibration // *Journal of Mathematics in Industry.* 2019. 9, 1. 9.
- [43] *Liu Shuaiqiang, Oosterlee Cornelis W., Bohte Sander M.* Pricing Options and Computing Implied Volatilities using Neural Networks // *Risks.* 2019. 7, 1.
- [44] *Ludwig Markus.* Robust estimation of shape constrained state price density surfaces // *The Journal of Derivatives.* 2015. 22, 3. 56–72.
- [45] *Madan Dilip B, Carr Peter P, Chang Eric C.* The variance gamma process and option pricing // *Review of Finance.* 1998. 2, 1. 79–105.
- [46] *Márquez-Neila Pablo, Salzmann Mathieu, Fua Pascal.* Imposing hard constraints on deep networks: Promises and limitations // *Unpublished manuscript.* 2017.
- [47] *Prechelt Lutz.* Early stopping-but when? // *Neural Networks: Tricks of the trade.* 1998. 55–69.
- [48] *Roper Michael.* Arbitrage free implied volatility surfaces // *Unpublished manuscript.* 2010.
- [49] *Sirignano Justin, Spiliopoulos Konstantinos.* DGM: A deep learning algorithm for solving partial differential equations // *Journal of Computational Physics.* 2018. 375. 1339–1364.
- [50] *Stein Elias M, Stein Jeremy C.* Stock price distributions with stochastic volatility: An analytic approach // *The Review of Financial Studies.* 1991. 4, 4. 727–752.
- [51] *Sutskever Ilya, Martens James, Dahl George, Hinton Geoffrey.* On the importance of initialization and momentum in deep learning // *Proceedings of the 30th International Conference on Machine Learning.* 2013. 1139–1147.
- [52] *Yang Yongxin, Zheng Yu, Hospedales Timothy M.* Gated neural networks for option pricing: Rationality by design // *Proceedings of the 31th AAAI Conference on Artificial Intelligence.* 2017.
- [53] *Zheng Yu.* Machine learning and option implied information // *PhD thesis, Imperial College London.* 2018.