Neural networks for option pricing and hedging: a literature review

Johannes Ruf* Weiguan Wang†

May 8, 2020

Abstract

Neural networks have been used as a nonparametric method for option pricing and hedging since the early 1990s. Far over a hundred papers have been published on this topic. This note intends to provide a comprehensive review. Papers are compared in terms of input features, output variables, benchmark models, performance measures, data partition methods, and underlying assets. Furthermore, related work and regularisation techniques are discussed.

1 Introduction

Beginning with Malliaris and Salchenberger [1993b] and Hutchinson et al. [1994], more than one hundred papers in the academic literature concern the use of artificial neural networks (ANNs) for option pricing and hedging. This work provides a review of this literature. The motivation for this summary arose from our companion paper Ruf and Wang [2020]. There we continue the discussions of this note; in particular, of potentially problematic data leakage when training ANNs to historical financial data.

A linear regression model can be thought of as an affine function that maps some input x to an output y. Similarly, an ANN can be thought of as a (possibly repeated) composition of linear and nonlinear functions, again mapping some input x to an output y. Training an ANN usually corresponds to choosing the linear components so that this mapping is optimal, in some sense, for (a subset of) a given dataset (the *training set*) $(x_i, y_i)_i$. Optimality is usually measured by means of a *loss function*, which measures the distance between the ANN output and the given data.

The Stone-Weierstrass theorem asserts that any continuous function on a compact set can be approximated by polynomials. Similarly, the *universal approximation* theorems ensure that ANNs approximate continuous functions in a suitable way. In particular, ANNs are able to capture nonlinear dependencies between input and output.

With this understanding, an ANN can be used for many applications related to option pricing and hedging. In the most common form, an ANN learns the price of an option as a function of the underlying price, strike price, and possibly other relevant option characteristics. Similarly, ANNs might also be trained to learn implied volatility surfaces or optimal hedging ratios. In the pricing task, the corresponding loss function is often chosen to be the squared distance of the observed (simulated) option prices and the ANN predicted prices. In the hedging task, one would compare observed (simulated) option prices and the values of the ANN hedging portfolios.

Let us provide a formal example in the context of the pricing task, namely a two-hidden layer ANN with linear output. Such an architecture maps an input x (usually a vector consisting of several features, such as moneyness, contract-specific implied volatility, etc.) to an output y (the option price) as follows:

$$y = w_2 \cdot \phi(w_1 \cdot x).$$

We thank Agostino Capponi, Marc Chataigner, Stéphane Crépey, Antoine Jacquier, and Martin Larsson for comments on an early version of this note.

*Department of Mathematics, London School of Economics and Political Science. Email: j.ruf@lse.ac.uk

†Department of Mathematics, London School of Economics and Political Science. Email: w.wang34@lse.ac.uk

Here ϕ is a nonlinear function (the so called *activation function*), w_1, w_2 are weight vectors, and the dot denotes the scalar product. Training such an ANN corresponds to finding weight vectors $\hat{w}_1, \hat{w}_2$ such that the output $\hat{y}$ of the ANN is close to the option price y, for all samples in a subset of the data (the *training set*). As already mentioned, a widely used criterion to measure what 'close' means is the mean squared error.

The papers discussed here mostly study how well such an approximation by an ANN works on either simulated or real datasets. Different performance measures are employed, and often the ANNs are compared to a variety of benchmarks, the simplest one being the Black-Scholes formula. We shall also summarize how the individual papers choose the training data.

The universal approximation theorems allow a 'model-based' usage of ANNs. Imagine a data-generating process, along with a computationally involved pricing algorithm, which relies, for example, on solving partial differential equations or Monte-Carlo simulations. When facing such a situation, ANNs can be used to learn directly the pricing formula. We review this literature in Section 4.

This paper is organised in the following way. Section 2 features Table 1, a summary of the literature that concerns the use of ANNs for nonparametric pricing (and hedging) of options. Section 3 provides a list of recommended papers from Table 1. Section 4 provides an overview of related work where ANNs are applied in the context of option pricing and hedging, but not necessarily as nonparametric estimation tools. Section 5 briefly discusses various regularisation techniques used in the reviewed literature.

2 ANN based option pricing and hedging in the literature

Bennell and Sutcliffe [2004], Chen and Sutcliffe [2012], and Hahn [2013][1] provide extensive literature surveys on the application of ANNs to option pricing and hedging problems. Here we complement these surveys with additional and more recent papers.

Table 1 summarises a large part of the literature and compares six relevant characteristics. They are features (or so-called explanatory variables), outputs of the ANN, benchmark models, data partition between training and test sets, and the underlyings along with the time span of the data. In Table 1, we only list papers that study an ANN's performance for the option pricing and hedging problem with a somehow statistical perspective. Other papers have different approaches, e.g., a computational perspective, and hence do not fit naturally in the table. These papers are discussed separately in Section 4.

We have not included a comparison of methodologies for the parameter estimation or of ANN architectures, such as number of nodes and layers, activation functions, etc. These specifications vary strongly between the papers summarized here. As an overall trend let us only remark that more recent papers use more complex architectures, in line with improved availability of computational resources. We also do not include a paper-by-paper summary of specific conclusions been drawn. However, more than half of the paper abstracts explicitly emphasize the positive performance of ANNs in the option pricing and hedging task.

Let us explain how to read Table 1. It summarises six relevant characteristics that describe how each paper treats the pricing/hedging problem. The columns 'Features' and 'Outputs' show explanatory features given to the ANN as inputs and outputs, respectively. Table 2 explains notations and abbreviations used for these columns. The 'Benchmarks' column lists non ANN-based techniques with which an ANN is compared. Table 3 explains the corresponding abbreviations. Table 4 presents abbreviations and definitions for the 'Performance measures' column, which summarises how an ANN (and its benchmarks) are evaluated in each paper. The performance measures marked bold are related to evaluations along multiple periods. Table 5 explains abbreviations for the underlying assets used in each study and listed in the 'Underlyings' column.

Here an 'executive summary' of Table 1:

- There exist two ways of using the stock price and option strike as inputs to an ANN. Sometimes they are used as two separate features. Other times, only their ratio (the so-called moneyness) is used as

[1]Hahn [2013] also surveys the use of ANNs to predict realised volatility. Here we do not aim to do so.

an input. In the previous ten years, the second approach is used more often. See also Subsection 2.1 for a discussion of this point.

- There are many different choices of volatility estimates concerning input features and benchmarks. The conclusions drawn often depend on this choice. Subsections 2.1 and 2.3 provide more details on this point.

- Most papers focus on estimating option prices, around fifteen papers (10% of all papers listed) on estimating implied volatilities, and very few deal with the hedging problem directly; see also Subsection 2.2.

- In some studies, data is partitioned into a training and a test set in a way that violates the underlying time series structure. This introduces information leakage and underestimates the generalization error of the ANN. This is further discussed in Subsection 2.4.

For the reader interested in a small selection of all these papers, we refer to Section 3.

After reading about 150 papers and creating Table 1, we would like to offer three pieces of (personal) advice when implementing ANNs as nonparametric estimation tool of option prices and hedges. First, stationary features should be used as input. Secondly, the ANN performance should be appropriately benchmarked. Third, the time series structure should not be violated when partitioning the data set into training and test sets.

Authors & year	Features	Outputs	Benchmarks	Performance measures	Partition method	Underlyings
Malliaris and Salchenberger [1993a,b]	S, K, τ, σ_{IM}, r, lagged C and S	C	BS-IM	MAE, MAPE, MSE	Chronological	S&P100. 6M
Hutchinson et al. [1994]	S/K, τ	C/K	BS-H, Linear	**MATE**, **PE**, R^2	Chronological	Simulation (BS); S&P500. 5Y
Kelly [1994]	S, K, τ, σ_{H}	C	CRR	MAE, **MTE**, MSE, R^2	?	Individual stocks. 6M
Boek et al. [1995]	S/K, τ, σ_{H}, r	$(C - C_{\mathrm{BS-H}})/K$	BS-H	MAPE, R^2	?	AOSPI. 2Y
Miranda and Burgess [1995]	?	$\Delta\sigma_{\mathrm{I}}$	Linear	?	?	IBEX35. ?
Krause [1996]	$C_{\mathrm{BS-H}}$, S, K, τ, σ_{H}	C	BS-H	R^2	Chronological	DAX. 3Y
Lachtermacher and Rodrigues Gaspar [1996]	S, K, τ, σ_{H}, r	C	BS-H	MAE, MAPE, MPE, MSE	Random	Individual stocks. 2M
Lajbcygier and Flitman [1996]	S/K, τ, σ_{IH}	$(C - C_{\mathrm{BS-IH}})/K$	BS-IH, KR, Linear	MAE, R^2	Chronological	AOSPI. 3Y
Lajbcygier et al. [1996a][2]	?	?	BS-?, BW	?	?	AOSPI. ?
Lajbcygier et al. [1996b], Lajbcygier [2002]	S/K, τ, σ_{H}, r	C/K	BS-H, BW, Linear	MAPE/MAE, MSE, R^2	Random	AOSPI. 2Y
Liu [1996]	S	S[3]	BS-H	MAE, MAX, MSE	Chronological	S&P500. 5Y[4]
Malliaris and Salchenberger [1996]	τ, lagged σ_{IM}, and others	σ_{IM}	None	MAE, MSE	Chronological	S&P100. 1Y
Niranjan [1996]	S/K, τ	C/K	BS-H	MSE	?	FTSE100. 11M
Qi and Maddala [1996][5]	S, K, τ, r, open interest	C	BS-H	MAE, MSE, R^2	Random	S&P500. 2M
Hanke [1997]	S/K, τ, σ_{G},[6] r	C/K, $(C - C_{\mathrm{BS-G}})/K$	None	MSE	Chronological	Simulation (SV)
Herrmann and Narr [1997]	S, K, τ, σ_{I}, σ_{V}, r	C	BS-V	MAE, ME, MSE, R^2	?	Simulation (BS); DAX. 1Y
Karaali et al. [1997]	S, K, σ_{H}	C	None	None	Chronological	DEM volatility. 5Y

[2] We were not able to obtain a copy of this paper.
[3] The network learns the dynamics of the underlying iteratively and then relies on Monte-Carlo to determine option prices.
[4] The network is trained on a five-year long stock price path, but uses only one day's option price data.
[5] This paper relies on the PhD thesis Qi [1996].
[6] Additional GARCH parameters are also added as features.

Authors & year	Features	Outputs	Benchmarks	Performance measures	Partition method	Underlyings
Lajbcygier and Connor [1997a,b]	S/K, τ, σ_{IH}	$(C - C_{\mathrm{BS-IH}})/K$	BS-IH	MAE, SR	Chronological	AOSPI. 1Y
Lajbcygier et al. [1997][2]	S/K, ?	$(C - C_{\mathrm{BS-?}})/K$	?	?	?	AOSPI. ?
Ahmed and Swidler [1998]	S/K, τ, σ_{H}, volume	σ_{I}	None	MAE, MSE	Random	Individual stocks. 3Y
Anders et al. [1998]	S/K, S, τ, σ_{H}, σ_{V}, r	C/K, $(C - C_{\mathrm{BS-V}})/K$	BS-H, BS-V	MAE, MAPE, ME, MSE, R^2	?	DAX. 3Y
Avellaneda et al. [1998]	S/K, τ	σ_{I}	None	%E	?	USD-DEM. Several days
Garcia and Gençay [1998, 2000]	S/K, τ	C/K	BS-H, Linear	DM, **MATE**, MSE	Chronological	Simulation (BS); S&P500. 8Y
White [1998]	?	C	None	MAE, MSE	Random	Simulation (BS)
Chen and Lee [1999]	S, τ, σ_{H}, Γ, Δ, ρ, $\mathcal{V}$, volume	C	BS-H, CRR	MAE, MAPE, MSE	Chronological	Individual stocks. 1Y
Geigle and Aronson [1999][7]	S/K, τ, σ_{H}, r	C/K	BS-H	MAE, MAPE	Chronological	S&P500. 6Y
Hanke [1999a]	S/K	$(C - C_{\mathrm{BS-H}})/K$	BS-H	MSE	Chronological	DAX. 1Y
Hanke [1999b]	S/K, τ, σ_{Cal}	C/K, $(C - C_{\mathrm{BS-Cal}})/K$	BS-Cal	MSE	Chronological	DAX. 10M
Ormoneit [1999]	S/K	C/K	BS-H, BS-IH	**MATE**, MSE, R^2	?	DAX. 9M
Tsaih [1999]	S, K, τ, σ_{I}, r	C	BS-IH	Sensitivity analysis	Chronological	Simulation (BS)
Briegel and Tresp [2000]	S, τ	C	BS-?, lagged C	MSE	?	FTSE100. 10M
Carelli et al. [2000]	K, τ	σ_{I}	None	%E	?	USD-DEM. Several days
de Freitas et al. [2000a,b]	S/K, τ	C/K	BS-H	R^2	?	FTSE100. 11M
Galindo-Flores [2000]	S, K, τ	C	Decision tree, Linear, Nearest neighbour	MSE	?	Simulation (BS)
Ghaziri et al. [2000]	S, K, τ, σ_{H}, r, open interest	C	BS-H	MSE	?	S&P500. 2M
Raberto et al. [2000]	S/K, τ, $\lvert S-K\rvert/\tau$	C/K	None	None	?	BUND. ?
Saito and Jun [2000][2]	?	?	BS-?	?	?	S&P500. ?

[7] This paper relies on the PhD thesis Geigle [1999].

Authors & year	Features	Outputs	Benchmarks	Performance measures	Partition method	Underlyings
White [2000]	S, K, τ, σ_{H}	C	BS-H	MAE, MSE	Random	Simulation (BS); Eurodollar. 7M
Yao et al. [2000]	S, K, τ	C	BS-H	R^2	Chronological	NIKKEI225. 1Y
Dugas et al. [2001, 2009]	S/K, τ	C/K	None	MSE	Chronological	S&P500. 5Y
Gençay and Qi [2001]	$S/K, \tau$	C/K	BS-H	DM, **MATE**, MSE	Chronological	S&P500. 6Y
le Roux and du Toit [2001]	S, K, τ, σ_{I}, r	C	None	MSE	Chronological	Simulation (BS)
Meissner and Kawano [2001]	S/K, τ, σ_{G}	C/K	BS-G	MAE, MAPE, ME, MSE, R^2	?	Individual stocks. 8M
Schittenkopf and Dorffner [2001]	τ	Gaussian parameters[8]	BS-H, CS	MAE, **MATE**, ME, MSE	Chronological	FTSE100. 5Y
Andreou et al. [2002]	$S/K, \tau, \sigma_{\mathrm{H}}, \sigma_{\mathrm{V}}$, r, and others	C/K, $(C-C_{\mathrm{BS-H}})/K$, $(C-C_{\mathrm{BS-V}})/K$	BS-H, BS-V	MdAE	Chronological	S&P500. 3Y
Billio et al. [2002]	S/K, τ, σ_{I}, r	C/K	BS-?	MSE	Chronological	FTSE100. 1Y
Ghosn and Bengio [2002]	S/K, τ	C/K	None	MSE	Chronological	S&P500. 6Y
Healy et al. [2002]	S, K, τ, σ_{I}, r, spread, open interest, volume	C	None	MAE, ME, R^2	Random	FTSE100. 5Y
Zapart [2002, 2003b]	Lagged wavelet coefficients	Wavelet coefficients[9]	BS-?	MAE	Chronological	Individual stocks. 6M/1Y
Amilon [2003]	S/K, τ, σ_{H}, r, lagged S	C_{Ask}/K, C_{Bid}/K	BS-H, BS-IM	ME, **MTE**, MSE	Chronological	OMX. 2Y
Carverhill and Cheuk [2003]	K/S, τ, σ_{I}, r	C/K, HR	CRR	**?TE**	Chronological	S&P500. 11Y
Gençay and Salih [2003]	$S/K, \tau, \sigma_{\mathrm{H}}, r$	C/K	BS-H	DM, MSE	Chronological	S&P500. 6Y
Healy et al. [2003, 2004][10]	S/K, τ	C/K	None	MSE, R^2	Random	FTSE100. 6Y
Lajbcygier [2003, 2004]	$S/K, \tau$	$(C-C_{\mathrm{BS-IH}})/K$	None	MAE, MSE, R^2	Chronological	AOSPI. 3Y
Montagna et al. [2003]	S, τ	C	None	None	?	Simulation (BS)
Zapart [2003a][11]	S/K, τ, σ_{H}, r	C/K	BS-?	MAE	Chronological	Individual stocks. ?

[8] ANNs output parameters for a Gaussian mixture density as a model for the risk-neutral density.
[9] An ANN is used to predict the future volatility of the underlying. The volatility is represented in terms of wavelets and the underlying modelled as a binomial tree.
[10] These papers also derive prediction intervals for ANN estimates of option prices.
[11] This paper also treats the setup of Zapart [2002].

Authors & year	Features	Outputs	Benchmarks	Performance measures	Partition method	Underlyings
Bennell and Sutcliffe [2004]	$S, K, S/K, \tau$, σ_{IM}, open interest, volume	C, C/K	BS-IM	MAE, ME, MPE, MSE	Chronological	FTSE100. 1Y
Choi et al. [2004]	S, K, τ, $\sigma_{?}$	C	BS-?	?	Random	KOSPI200. 1Y
Dindar and Marwala [2004]	$S, \tau, \sigma_{\mathrm{H}}, r$	K/C	None	?	Random	South Africa Foreign Exchange. 3Y
Morelli et al. [2004]	$S, K, \tau, \sigma_{\mathrm{I}}, r$	C	None	?	?	Simulation (BS)
Pires and Marwala [2004a,b]	$K, \tau, \sigma_{\mathrm{H}}$	C	SVM	MAX, ME	?	Johannesburg Stock Exchange. 3Y
Xu et al. [2004]	S, K, τ, σ_{I}, r	C	None	R^2	Random	FTSE100. 5Y
Charalambous and Martzoukos [2005]	$S, K, \sigma_{\mathrm{H}}, r$, correlations[12]	$C - C_{\mathrm{LA}}^{10}$	LA-10	MAE, MAX, MSE	Chronological	Simulation (BS)
Hamid and Habib [2005]	$S, \tau, \sigma_{\mathrm{H}}, r$	C	None	MAE, MSE	?	S&P500. 12Y
Kakati [2005][2]	?	?	?	?	?	Individual stocks. ?
Ko et al. [2005], Ko [2009]	S, K, τ, σ_{H}	Coefficients[13]	BS-H	**MATE**	?	TAIEX. 1Y/2Y
Lin and Yeh [2005]	$S, K, \tau, \sigma_{\mathrm{H}}, r$	C	BS-H	MAE, MSE	?	TAIEX. 2Y
Pires and Marwala [2005]	$K, \tau, \sigma_{\mathrm{H}}$	C	SVM	MAX, ME, MSE	?	ALSI. 3Y
Tung and Quek [2005]	$S - K$, τ, σ_{H}	C	None	MSE, Correlation[14]	Random	GBP-USD. 1Y
Andreou et al. [2006][15]	S/K, τ, σ_{Cal}, σ_{H}, σ_{V}, r	C/K, $(C - C_{\mathrm{BS-Cal}})/K$, $(C-C_{\mathrm{BS-H}})/K$, $(C-C_{\mathrm{BS-V}})/K$	BS-Cal, BS-H, BS-V	MAE, MSE	Chronological	S&P500. 3Y
Blynski and Faseruk [2006]	S/K, τ, σ_{H}, σ_{IH}	C/K, $(C-C_{\mathrm{BS-H}})/K$, $(C-C_{\mathrm{BS-N}})/K$	BS-H, BS-IH	MAE, MAPE, ME, MSE, R^2	?	S&P100. 7Y
Huang and Wu [2006], Huang [2008]	S/K, τ, σ_{K}	$(C-C_{\mathrm{BS-K}})/K$	SVM	MAE, MAPE, MSE	Chronological	TAIEX. 9M
Jung et al. [2006]	S, K, τ, σ_{IH}	C	BS-IH	MSE	?	KOSPI200. 1Y
Kim et al. [2006]	K, τ	σ_{Cal}	SI	MSE	?	S&P500. 1M

[12] Correlations between underlyings.
[13] Coefficients for a linear regression that returns option prices.
[14] Pearson correlation coefficient, a statistical measure to verify the goodness-of-fit between the predicted and desired function.
[15] This paper relies on the PhD thesis Andreou [2008].

Authors & year	Features	Outputs	Benchmarks	Performance measures	Partition method	Underlyings
Liang et al. [2006]	$\hat{C}$[16]	C	BS-?, CRR	MAE	Chronological	Individual stocks. 5M
Mitra [2006]	$S, K, \tau, \sigma_{\mathrm{H}}, r$	C	None	MAE, MSE	Chronological	NIFTY50. 1Y
Pande and Sahu [2006]	S/K, τ, σ_{PCA}, r	C or(?) C/K	None	ME, MSE, Correlation[17]	?	Individual stocks. 1Y
Teddy et al. [2006]	$S-K$, τ, σ_{H}	C	None	MSE, Correlation[14]	Random	GBP-USD. 1Y
Tzastoudis et al. [2006]	$S, K, \sigma_{\mathrm{H}}$	C	BS-H	MAE, R^2	Chronological	S&P500. Several days
Wang [2006]	S/K, σ_{IH}, $(S-K)^+$, $C-(S-K)^+$, $CS/\sqrt{K}$	σ_{I}	BS-IH	MAE, MSE, R^2	?	Individual stocks. 2M
Amornwattana et al. [2007]	S, K, τ, r	$C-C_{\mathrm{BS-N}}$, σ_{I}	BS-H, BS-N	MAE, MSE	Chronological	Individual stocks. 3M
Gençay and Gibson [2007]	S, K, τ, σ_{G}, r	C	BS-G, BS-H, SV, SVJ	MAE, MSE	?	S&P500. 3Y
Gregoriou et al. [2007]	S_{Ask}, S_{Bid}, S_{Mid}, K, τ, σ_{I}, r	C	None	None	Random	FTSE100. 5Y
Healy et al. [2007]	S, K, τ, σ_{I}, r	C	None	R^2	Chronological	FTSE100. ?
Thomaidis et al. [2007]	S, K, τ	C	BS-G, BS-H	MAE, MSE	Chronological	S&P500. Several days
Zhou et al. [2007]	S/K, S, K, τ, r	C/K	BS-?, CRR	MAE, MAPE, ME, MSE, R^2	Chronological	Convertible bonds. 2Y
Andreou et al. [2008][15]	S/K, σ_{Cal}, σ_{H}, σ_{V}, r, kurtosis, skewness	C/K, $(C-C_{\mathrm{BS-Cal}})/K$, $(C-C_{\mathrm{BS-H}})/K$, $(C-C_{\mathrm{BS-V}})/K$	BS-Cal, BS-H, BS-V, CS	MAE, **MATE**, MdAE, MSE, **MTE**	Chronological	S&P500. 4Y
Chiu and Lin [2008]	S, C_{BS}, volume, and others	C	None	MSE	Chronological	Individual stocks. 1Y
Kakati [2008]	S/K, τ, σ_{G}, σ_{H}, σ_{IH}, r	C/K	BS-G, BS-H, BS-IH	MSE	?	Individual stocks. Several days
Mostafa and Dillon [2008][18]	$S/K, \tau, \sigma_{\mathrm{H}}$	C/K, σ_{I}	BS-H, SV	MAPE, **MATE**, MPE	?	FTSE100. 2Y

[16] Various price estimations from parametric option pricing models.
[17] Correlation between the actual and computed prices.
[18] This paper relies on the PhD thesis Mostafa [2011].

Authors & year	Features	Outputs	Benchmarks	Performance measures	Partition method	Underlyings
Quek et al. [2008]	lagged C	C	None	None	?	GBP-USD, Gold, Oil. 2Y
Saxena [2008]	$S/K, \tau, \sigma_{\mathrm{H}}, r$	$(C - C_{\mathrm{BS-H}})/K$	BS-H	MAE, ME, MPE, MSE, R^2	?	NIFTY50. 1Y
Teddy et al. [2008]	$S - K, \tau, \sigma_{\mathrm{H}}$	C	BS-H	MSE, R^2	Random	GBP-USD. 9M
Tseng et al. [2008]	$S, K, \tau, \sigma_{\mathrm{G}}, r$	C	None	MAE, MAPE, MSE	?	TAIEX. 2Y
Chen [2009]	$S, K, \tau, \sigma_{\mathrm{H}}, r$	C	BS-H, SVM	MAE, MSE	Chronological	S&P500. Several days
Gradojevic et al. [2009]	$S/K, \tau$	C/K	BS-H	DM, MSE, MSPE	Chronological	S&P500. 8Y
Leung et al. [2009]	σ_{H}, σ_{IH}, volume, open interest	σ_{I}	BS-IH, Linear, Polynomial	ME	Chronological	Several currencies. 17Y
Liang et al. [2009]	$\hat{C}$[16]	C	CRR, SVM	MAE, MAPE	Chronological	Individual stocks. 2Y
Martel et al. [2009]	$S/K, \tau, \sigma_{\mathrm{H}}, r$	C_{Bid}/K, C_{Ask}/K	BS-H	ME, MSE, **MTE**	Chronological	IBEX35. 2Y
Samur and Temur [2009]	$S, K, \tau, \sigma_{\mathrm{H}}, r$	C	None	MAE, MSE, R^2	?	S&P100. Several days
Wang [2009a]	$S/K, \tau, \sigma_{\mathrm{G}}, r$	C/K	None	MAE, MAPE, MSE	?	TAIEX. 2Y
Wang [2009b]	$S/K, \tau, \sigma_{\mathrm{G}}, \sigma_{\mathrm{H}}, \sigma_{\mathrm{IH}}, r$	C/K	None	MAE, MAPE, MSE	?	TAIEX. 2Y
Andreou et al. [2010][15]	$S/K, \tau$	σ_{I}	BS-Cal, CS, SV, SVJ	MAE, **MATE**, MdAE, MSE	Chronological	S&P500. 3Y
Barunikova and Barunik [2011]	S, K, τ	C	BS-H	MAE, MAPE, MSE	Random	S&P500. 3Y
Gradojevic and Kukolj [2011]	$S/K, \tau, \sigma_{\mathrm{IH}}, r$	C/K	BS-H	DM, MAPE, MSE	Chronological	S&P500. 7Y
Liu and Zhang [2011]	$S/K, \tau, \sigma_{\mathrm{H}}$,[19] r	C/K	BS-H	MAE, MSE	Chronological	Individual stocks. 2Y
Phani et al. [2011]	S, K, τ	C	BS-?, SVM	MAE	?	NIFTY50. 2Y
Tung and Quek [2011]	σ_{IH}	σ_{I}	None	MAPE, MSE, R^2	Chronological	HSI. 5Y
Wang [2011]	$S/K, S, \tau, \sigma_{\mathrm{Cal}}, r$	C	SV, SVJ, SVM	MAE, MAPE	Chronological	Several currencies. 7M

[19] More precisely, a Markov regime switching model is used to estimate the volatility.

Authors & year	Features	Outputs	Benchmarks	Performance measures	Partition method	Underlyings
Ahn et al. [2012]	Lagged σ_{I}, Greeks	Sign($\Delta\sigma_{\mathrm{I}}$)	None	Accuracy	Chronological	KOSPI200. 2Y
Chen and Sutcliffe [2012]	$S/K, \tau$	C/K, $(C - C_{\mathrm{BS-H}})/K$, HR	BS-H	MAE, ME, MSE	Random	Sterling futures. 2Y
Mitra [2012]	S, K, τ, σ_{H}, r	C	BS-H	ME, MSE	Chronological	NIFTY50. 3Y
Shin and Ryu [2012]	S, K, τ, r	HR	None	MPE	Chronological	KOSPI200. 10Y
Wang et al. [2012]	S, K, τ, σ_{Cal}, σ_{G}, σ_{H}, σ_{IH}	C	None	MAE, MAPE, MSE	Chronological	TAIEX. 2Y
Chang et al. [2013]	$S/K, \tau, \sigma_{\mathrm{G}}, r$	C or(?) C/K	None	MAE, MAPE	?	TAIEX. 2Y
Hahn [2013]	S/K, τ, σ_{H}, r	C/K	SV	MAE, MAPE, MSE	Chronological	Individual stocks. 10Y
Can and Fadda [2014]	$S/K, S, \tau, r$	C/K	BS-H	MAE	Chronological	S&P100. Several days
Lai [2014]	$S/K, \tau, r$	σ_{I}	KR, SI	KS	?	Simulation (BS, SV, SVJ)
Park et al. [2014]	$S/K, \tau$	C/K	BS-H, SV	MSE	Chronological	KOSPI200. 10Y
von Spreckelsen et al. [2014]	S/K, K, τ	C/K	None	MSE, R^2	Chronological	EUR-USD. 1M
Ludwig [2015]	$S/K, \tau$	σ_{I}	Quadratic	MSE, R^2	?	S&P500. 12Y
Liu and Huang [2016]	$S/K, \tau$	$(C - C_{\mathrm{BS-H}})/K$	BS-H	MAE, MAPE, ME, MSE	?	HSI. 6Y
Montesdeoca and Niranjan [2016]	S/K, τ, σ_{H}, volume	C/K	None	MSE	Chronological	FTSE100. ?; Individual stocks. ?
Culkin and Das [2017]	$S/K, \tau, \sigma_{\mathrm{I}}, r$	C/K	None	MSE, R^2	Chronological	Simulation (BS)
Das and Padhy [2017]	S/K, τ, $\widehat{C}$[16]	C	BS-H, SVM	MAE, MSE	Chronological	NIFTY50. 2Y
Fang and George [2017]	σ_{H}	σ_{I}	None	MSE, R^2	Chronological	Simulation (BS); WTI. 1M
Palmer and Gorse [2017]	S, K, σ_{I}, r	C	None	MAE, MdAE, MAPE	Chronological	Simulation (BS)
Yang et al. [2017][20]	$K/S, \tau$	C/S	BS-?, Kou, VG	MAPE, MSE	?	S&P500. 10Y
Ferguson and Green [2018]	S, τ, σ_{I}, correlations[12]	C	None	MSE	Chronological	Simulation (BS)
Ackerer et al. [2019]	$\log(K/S), \tau$, $\log(K/S)\tau^{-0.5}$, $\log(K/S)\tau^{-0.95}$	σ_{I}	None	MAPE, MSE	Random	S&P500. 1M

[20] This paper relies on the PhD thesis Zheng [2017].

Authors & year	Features	Outputs	Benchmarks	Performance measures	Partition method	Underlyings
Buehler et al. [2019a,b]	$\log(S)$	HR	BS-I	**CVaR**	Chronological	Simulation (BS, SV); S&P500. 5Y
Cao et al. [2019]	$S/K, \tau, \sigma_{\mathrm{V}}$, underlying return	σ_{I}	HW	MSE	Random	S&P500. 8Y
Jang and Lee [2019]	?	C	BS-Cal, BW, KR, LSM, LV, SVJ, SVM	MAE, MAPE, MPE, MSE	?	S&P100. 9Y
Liu et al. [2019b]	$S/K, \tau$	σ_{I}	None	MAE, MAPE, MSE	Chronological	Simulation (BS)
Liu et al. [2019c]	$S/K, \tau, \sigma_{\mathrm{Cal}}, r$	$(C - C_{\mathrm{BS-H}})/K$	BS-Cal, SVJ	MAE, **MATE**, MPE, MSE	Chronological	DAX. 4Y
Karatas et al. [2019]	$S/K, \tau, r$, ?	C/K	None	MSE, R^2	Chronological	Simulation (BS, SV, VG)
Palmer [2019]	$S/K, \sigma_{\mathrm{I}}\sqrt{\tau}, r$	C/K	BS-I, LSM	MAE, MAPE	Chronological	Simulation (BS)
Zheng et al. [2019]	$S/K, \tau$	σ_{I}	SSVI	MAPE	?	S&P500. 10Y
Ruf and Wang [2020]	$S/K, \sigma_{\mathrm{I}}\sqrt{\tau}, \Delta$, $\mathcal{V}$, Vanna	HR	BS-I, HW, Linear	MSE	Chronological	Simulation (BS, SV); S&P500. 8Y; STOXX50. 3Y

Table 1: This table summarises more than 150 papers that use ANNs as a nonparametric option pricing or hedging tool. These papers are compared in terms of features (or so-called explanatory variables), outputs of the ANN, benchmark models, data partition between training and test sets, and the underlyings along with the time span of the data. The performance measures marked bold are related to evaluations along multiple periods. We refer to Tables 2–5 for a dictionary of all abbreviations used here.

C	Option price
$C_{\mathrm{BS-X}}$	Option price given by the Black-Scholes formula; see Table 3 for the different meanings of X
C^n_{LA}	Option price given by n-step multi-dimensional lattice scheme
HR	Hedging ratio
K	Strike price
S	Stock price
r	Interest rate
Γ	Gamma: second-order sensitivity of option price with respect to underlying price
Δ	Delta: sensitivity of option price with respect to underlying price
$\mathcal{V}$	Vega: sensitivity of option price with respect to volatility
ρ	Rho: sensitivity of option price with respect to interest rate
σ_{Cal}	Volatility from calibration (e.g., constant across strikes and maturities)
σ_{G}	GARCH–generated volatility
σ_{H}	Historical volatility
σ_{I}	Implied volatility
σ_{IH}	Implied historical volatility
σ_{IM}	At-the-money implied volatility
σ_{K}	Volatility obtained from Kalman filter
σ_{PCA}	Macroeconomic variables that contribute the most to volatility, determined by principle component analysis
σ_{V}	Volatility index such as VIX and DVAX
τ	Time to maturity

Table 2: This table presents notations and abbreviations for features and outputs, used in Table 1.

BS-Cal	Black-Scholes formula with calibrated volatility
BS-G	Black-Scholes formula with GARCH-generated volatility
BS-H	Black-Scholes formula with historical volatility
BS-I	Black-Scholes formula with contract-specific implied volatility
BS-IH	Black-Scholes formula with historical implied volatility
BS-IM	Black-Scholes formula with at-the-money implied volatility
BS-K	Black-Scholes formula with volatility obtained from Kalman filter
BS-N	Black-Scholes formula with ANN-generated volatility
BS-V	Black-Scholes formula with volatility index, such as VIX or VDAX
BW	Barone-Adesi and Whaley [1987] pricing method
CRR	Cox et al. [1979] model
CS	Corrado and Su [1996] model
HW	Hull and White [2017] model
Kou	Kou [2002]'s jump diffusion model
KR	Kernel regression
LA-n	n-step multi-dimensional lattice scheme
Linear	Linear regression on features
LSM	Longstaff and Schwartz [2001] method
LV	Local volatility model
Quadratic	Quadratic regression on features
SI	Spline interpolation
SSVI	Surface stochastic volatility inspired model, see Gatheral and Jacquier [2014]
SV	Stochastic volatility models, such as Heston [1993] or GARCH
SVJ	Stochastic volatility with jumps model, see Bates [1996] or Carr et al. [2003]
SVM	Support vector machine
VG	Variance Gamma model, see Madan et al. [1998]

Table 3: This table presents abbreviations for various benchmarks, used in Table 1.

DM	Diebold and Mariano test	
KS	Kolmogorov and Smirnov two-sample test	
MAE	Mean absolute error	$\frac{1}{N}\sum\lvert\hat{y}_i - y_i\rvert$
MAPE	Mean absolute percentage error	$\frac{1}{N}\sum\frac{\lvert\hat{y}_i - y_i\rvert}{y_i}$
MAX	Maximum error	$\max_i \lvert\hat{y}_i - y_i\rvert$
MdAE	Median absolute error	$\sup_z\left\{\frac{1}{N}\sum\mathbf{1}_{\lvert\hat{y}_i - y_i\rvert < z} \leq 0.5\right\}$
ME	Mean error	$\frac{1}{N}\sum(\hat{y}_i - y_i)$
MPE	Mean percentage error	$\frac{1}{N}\sum\frac{\hat{y}_i - y_i}{y_i}$
MSE	Mean squared error	$\frac{1}{N}\sum(\hat{y}_i - y_i)^2$
R^2	Coefficient of determination	$1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(\bar{y} - y_i)^2}$
SR	Sharpe ratio of a trading ratio	
%E	Sample-wise percentage error	$\frac{\hat{y}_i - y_i}{y_i}$
CVaR	Conditional value-at-risk	
MATE	Mean absolute tracking error	$\frac{1}{N}\sum e^{-rT_i}\lvert V(T_i)\rvert$
MTE	Mean tracking error	$\frac{1}{N}\sum e^{-rT_i}V(T_i)$
PE	Prediction error	$\sqrt{\mathrm{MTE}^2 + \frac{1}{N}\sum\left(e^{-rT_i}V\left(T_i\right) - \mathrm{MTE}\right)^2}$

Table 4: This table presents abbreviations and definitions for performance measures, used in Table 1. Here, $\hat{y}_i$ is the estimated option price / implied volatility / portfolio value, y_i is the target value, $\bar{y}$ is the average of target values, and N denotes the number of samples. Moreover, $V(T)$, also called tracking error, denotes the terminal value at T of a hedged option portfolio starting with zero wealth. All performance measures marked bold are related to evaluations along multiple periods.

ALSI	South African All Share Index
AOSPI	Australian All Ordinaries Share Price Index
BUND	German treasury bond
DAX	German stock index
DEM	Deutsche Mark
FTSE100	UK Financial Times Stock Exchange 100 index
HSI	Hong Kong Heng Seng Index
IBEX35	Spanish stock index
KOSPI200	Korea Composite Stock Price Index
NIFTY50	Indian National Stock Exchange Fifty
NIKKEI225	Japanese stock index
OMX	Swedish stock index
S&P100	US Standard & Poor's 100
S&P500	US Standard & Poor's 500
STOXX50	Eurozone stock index
TAIEX	Taiwanese stock index
WTI	US Light Sweet Crude Oil Futures

Table 5: This table presents abbreviations for various stock market indices and other underlyings, used in Table 1. For the shortcuts used to describe simulation data, we refer to Table 3.

In the following, we compare and classify papers listed in Table 1 in terms of features, outputs, performance measures and benchmarks, data partition methods, underlying assets and time span.

2.1 Features

To estimate the option price, the underlying price and the strike price are two indispensable variables. Two ways of feeding these two variables into an ANN as input have been suggested. One way is to use the underlying price and strike price separately. An alternative is to use a ratio (i.e., moneyness) instead. Several arguments are formulated in the literature in favor of using moneyness:

- Using moneyness instead of the stock price and the strike price separately reduces the number of inputs and thus makes the training of the ANN easier; see Hutchinson et al. [1994].
- Many parametric models assume that the statistical distribution of the underlying asset's return is independent of the level of the underlying. Hence, the option pricing function is homogeneous of degree one with respect to the underlying stock price and the strike price, so that only moneyness is needed to learn the function. Incorporating this assumption into the ANN can potentially reduce overfitting; see Hutchinson et al. [1994], Lajbcygier and Connor [1997a,b], Anders et al. [1998], and Garcia and Gençay [1998, 2000].
- Moneyness is a stationary input feature in contrast to the stock price and the strike price. Using it helps generalisation and reduces overfitting; see Ghysels et al. [1998] and Garcia and Gençay [1998, 2000]. Our own experiments also confirm that the use of moneyness can significantly improve the generalisation.

Bennell and Sutcliffe [2004] undertake a systematic experiment on various choices of input features, including underlying price, strike price, moneyness, and on choices of outputs, including option price and option price divided by strike.

Apart from the underlying price and the strike price, volatilities are also widely used as input features. This can be done in several different ways. The most relevant ones are the following:

- Using historical volatility estimates as features.
- Using volatility indices such as VIX as features.

- Using implied volatilities as features.
- Using GARCH forecasts of (realised or implied) volatility as features.

Table 2 lists further volatility features. The choices of features by the different papers are worked out in the 'Features' column of Table 1. There exist also several papers that do not use any volatility-type feature as input for their ANNs.

A few papers, e.g., Blynski and Faseruk [2006], Andreou et al. [2008], or Wang [2009b], compare different volatility features. Here we summarize their results. Blynski and Faseruk [2006] show an ANN outperforms the conventional Black-Scholes when using historical volatility as input, but underperforms when using implied volatility. Andreou et al. [2008] show that replacing historical by implied volatility improves the performance of ANNs. Wang [2009a] argue that an ANN with a GARCH volatility forecast outperforms that with historical and implied volatility as features.

Some papers investigate whether additional features can help the ANN with prediction. To name a few, Ghaziri et al. [2000] and Healy et al. [2002] incorporate option open interests. Samur and Temur [2009] study whether the inclusion of variance improves the performance of the ANN. Montesdeoca and Niranjan [2016] explore the potential prediction power of trading volume, option interest, and other variables. Cao et al. [2019] investigate the benefit from using the underlying return.

2.2 Outputs

The papers of Table 1 can also be categorised in terms of their outputs:

- The most common output is the option price. Depending on whether moneyness is used, or underlying price and strike price are used separately, the output can be the option price or the option price divided by the strike price. Some papers also investigate the ANN's ability when it is trained to learn the so-called bias; i.e., the difference between market price and a price estimated by a parametric model. Such an ANN is called hybrid ANN; see, for example, Boek et al. [1995] or Lajbcygier and Connor [1997a,a]. While most of the early papers train their ANNs to fit prices, Garcia and Gençay [2000] train to prices, but validate to hedging errors in order to determine the network size that gives the lowest hedging error. Andreou et al. [2010] emphasize the relevance of choosing the right loss function when interested in the hedging task.
- Another type of output is the implied volatility. The obtained implied volatilities can be converted to option prices by the Black-Scholes formula. Mostafa and Dillon [2008] compare ANNs that output option prices to ANNs that output implied volatilities. More recently, Liu et al. [2019b] evaluate an ANN's ability to approximate the inverse of the Black-Scholes formula.
- The third kind of output (always denoted by HR in Table 1) is a sensitivity or a hedging ratio. Only a few papers discuss such an architecture for an ANN. The first papers are Carverhill and Cheuk [2003], Chen and Sutcliffe [2012], and Shin and Ryu [2012]. More recently, Buehler et al. [2019a,b] and Ruf and Wang [2020] follow up on this line of research. Buehler et al. [2019b] consider also the hedging of exotic options such as barrier options.

We could have also added the so-called calibration papers to Table 1, which construct ANNs to map prices to specific model parameters or vice versa. Instead we decided to dedicate Section 4.1 below to these papers.

2.3 Performance measures and benchmarks

When evaluating the performance of ANNs, common statistical measures are mean absolute error (MAE), mean absolute percentage error (MAPE), and mean squared error (MSE).[21] These are related to evaluations

[21] Several papers use equivalent versions of the measures in Table 4. For example, sometimes root mean squared error is used instead of mean squared error. For consistency, in Table 1, we have made the corresponding adjustments.

over a single period, in terms of pricing or hedging. Some papers also propose to evaluate the ANN's performance over multiple periods. For instance, Hutchinson et al. [1994] introduce the mean absolute tracking error (MATE) and prediction error (PE), which appear also in many later papers. Buehler et al. [2019a] introduce the conditional value-at-risk (CVaR) for evaluating hedging strategies.

An ANN's performance should also be compared to a benchmark, for example, a parametric pricing model. The most widely used benchmark is the Black-Scholes formula, which requires a volatility as input. As Table 1 summarises a historic volatility estimate is used the most often. Also certain implied volatilities (e.g., historical or at-the-money) appear in the literature. Blynski and Faseruk [2006] compare historical realised and historical implied volatility for the Black-Scholes benchmark.

The Black-Scholes formula with contract-specific implied volatility is a valid benchmark for the hedging task. For the pricing task, however, such a benchmark would lead to zero error as by definition of implied volatility it prices options without errors. Thus, for the pricing task, the Black-Scholes formula with contract-specific implied volatility is not a suitable benchmark .

In addition to the Black-Scholes formula, other widely used parametric benchmarks are stochastic volatility pricing models; e.g., used in Gençay and Gibson [2007], Jang and Lee [2019], or Liu et al. [2019b]. Ruf and Wang [2020] observe that if a benchmark is chosen that incorporates both delta and vega hedging then an ANN does not outperform even a simple two-factor regression model.

For American type options, benchmarks used are the Barone-Adesi and Whaley [1987] pricing method (e.g., Lajbcygier [2002]), and the Cox-Ross-Rubinstein model (e.g., Chen and Lee [1999]).

2.4 Data partition methods

An ANN needs to be trained on a training set (in-sample) and then tested on a test set (out-of-sample). There exist several ways to partition a data set into such a training and test set. The first way is chronologically. That is, the early data constitutes the training set, and the late data constitutes the test set. Table 1 indicates that most of the papers follow this approach. However, some studies violate this time structure in the data by choosing a different way to partition the data. Violations can be introduced by randomly partitioning the data into a training and a test set or by using a so-called 'odd-even split.'

Random partitioning breaks the time structure and introduces information leakage between the training set and the test set. When an ANN is trained on a training set constructed in such a way, the error on the test set underestimates the generalisation error of the ANN. Yao et al. [2000] and our companion paper Ruf and Wang [2020] provide more discussion on this point.

Some papers only work with independent draws from various distributions, and therefore do not involve any time series structure. Although these papers randomly partition the whole data set into a training and test set, no time structure is violated. Hence, in Table 1, we classify this approach as chronological partition.

A related issue is the existence of time-inhomogeneity in financial data; in particular, volatility changes over time. When working with real data, some papers use a rolling window method to tackle this issue, especially when the time range is long and volatilities are not included as input features. Such papers include Hutchinson et al. [1994], Dugas et al. [2009], and others. However, it remains an open question how big window sizes need to be.

2.5 Underlying assets and time span

Both simulation data and real data can be used to train an ANN for a specific problem. Simulation data is much easier to work with, since it is free of noise and sometimes a close-to-optimal solution is available as a benchmark, such as for the Black-Scholes and Heston models. For instance, le Roux and du Toit [2001], Morelli et al. [2004], and Karatas et al. [2019] investigate an ANN's performance on simulation data. Most other papers use either both simulation and real data or only real data. Options on S&P500 have been studied by the largest number of papers, since they are the most liquidly traded options. Options on FTSE100 and S&P100 have also been studied in several papers. We refer to Table 5 for a more complete list of all the underlyings being used.

Some papers focus on American option pricing and hedging. Underlyings for American options are usually individual stocks. Papers involving American options include Kelly [1994], Chen and Lee [1999], Meissner and Kawano [2001], Pires and Marwala [2004a], Pires and Marwala [2005], and Amornwattana et al. [2007]. As elaborated in Subsection 4.3, American options can also be priced differently by ANNs, via learning the value function or optimal stopping rule in a dynamic programming setting; see Kohler et al. [2010] and Becker et al. [2019].

3 Recommended papers

Among the many papers of Table 1, we would like to highlight a few. Such a selection is clearly personal and subjective. Despite the subjective selection, we believe that this list might serve as a good starting point to get an overview of this field. We also provide a Google Scholar citation count.[22] As mentioned before, Table 1 focuses only on those papers that use ANNs to estimate option prices and related variables. Recently there have been many interesting and promising developments in the use of ANNs for calibration purposes or as computational tools. These papers are not included here, but Section 4 provides some pointers to this literature.

Among the following highlighted papers, some are the first to propose innovative solutions. Others investigate the problem in a systematic way.

- Hutchinson et al. [1994] (# citations: 749) is one of the first papers and the most highly cited one to use ANNs to estimate option prices. They introduce a methodology to evaluate the hedging performance over multiple periods, applied by many papers later on.
- Lajbcygier and Connor [1997a] (# citations:[23] 51) is one of the first papers that propose to learn the difference between model prices and observed market option prices.
- Anders et al. [1998] (# citations: 106) compare the performance of ANNs and of the Black-Scholes benchmark when using different volatility estimates.
- Garcia and Gençay [2000] (# citations:[24] 210) incorporate a homogeneity hint for the ANN. Hence, this is one of the first papers that embed financial domain knowledge into the construction of an ANN.
- Carverhill and Cheuk [2003] (# citations: 15) first propose an ANN that outputs hedging strategies directly, instead of option prices.
- Bennell and Sutcliffe [2004] (# citations: 83), Chen and Sutcliffe [2012] (# citations: 12), and Hahn [2013] (# citations: 9) provide three extensive literature surveys.
- Dugas et al. [2009] (# citations:[25] 172) first design an ANN architecture that enforces no-arbitrage conditions such as convexity of option prices.
- Andreou et al. [2010] (# citations: 19) combines an ANN with parametric models to learn functions that return implied model parameters. Such an ANN essentially calibrates parametric models.
- Buehler et al. [2019a] (# citations: 23) develop a novel framework for hedging a portfolio of derivatives in the presence of market frictions, and allow convex risk measures as loss functions. Their framework allows pricing and hedging without observing option prices.

As this is a subjective selection, we also would like to highlight our companion paper Ruf and Wang [2020], which provides a new benchmark based on delta-vega hedging and discusses data leakage issues.

[22] As of October 3, 2019.
[23] This count includes the number of citations for Lajbcygier and Connor [1997b].
[24] This count includes the number of citations for Garcia and Gençay [1998].
[25] This count includes the number of citations for Dugas et al. [2001].

4 Related papers

In the last few years, many novel techniques have been developed to apply ANNs to tasks arising in option pricing beyond the nonparametric estimation of prices and hedging ratios. In this section we provide a few pointers to this rapidly developing literature.[26]

4.1 Calibration

As already mentioned in Section 3, Andreou et al. [2010] propose an ANN that returns implied model parameters. Hence, the ANN essentially calibrates parametric models. We observe a recent surge of the application of ANN to calibration. In this approach option prices are first mapped to a parametric model, which is then used to determine option prices. This approach can move the computationally heavy calibration off-line, thus significantly accelerating option pricing.

Abu-Mostafa [2001] use neural networks to calibrate the Vasicek model with a consistency hint to produce valid parameters. More recently, Hernandez [2017] uses an ANN to calibrate a single-factor Hull-White model. Dimitroff et al. [2018], McGhee [2018] and Liu et al. [2019a] calibrate stochastic volatility models, and Stone [2019] and Bayer et al. [2019][27] calibrate rough volatility models. Itkin [2019] highlights some pitfalls in the existing approaches and proposes resolutions that improve both performance and accuracy of calibration.

Going the 'indirect' way via first calibrating a model and then using it to determine the hedging ratio has at least two advantages. First, it provides additional interpretability as only the calibration step is replaced by an ANN. This can be important for a financial entity subject to regulatory requirements. Second, it provides an arguably strong tailor-made regularisation effect as it replaces a nonparametric estimation task by the task of estimating a model with usually less than 5-10 parameters.

4.2 Solving partial differential equations

The option pricing problem sometimes involves solving a partial differential equation (PDE). Barucci et al. [1996, 1997] use the Galerkin method and ANNs for solving the Black-Scholes PDE. E et al. [2017], Han et al. [2018], and Beck et al. [2019] utilize ANNs to solve high-dimensional semilinear parabolic PDEs. They propose to reformulate the PDEs using backward stochastic differential equations, and the gradient of the unknown solutions is approximated by ANNs. Their numerical results suggest that the method is effective for a wide variety of (possibly high-dimensional) problems. One case study involves the pricing of European options on 100 defaultable underlying assets. There are several recent papers, such as Henry-Labordère [2017], Sirignano and Spiliopoulos [2018], Chan-Wai-Nam et al. [2019], Huré et al. [2019] , Jacquier and Oumgari [2019], and Vidales et al. [2019], who have developed this application of ANNs further.

4.3 Approximating value functions in optimal control problems

ANNs can be used to approximate value functions that appear in dynamic programming, for example arising in the American option pricing problem; see for example Ye and Zhang [2019]. Kohler et al. [2010] use ANNs to estimate continuation values for high-dimensional American option pricing. Becker et al. [2019] use ANNs for optimal stopping problems by learning the optimal stopping rule from Monte Carlo samples. ANNs have also been proposed to approximate the value function of a dynamic program for real option pricing, see Taudes et al. [1998].

In this context, we also mention Fecamp et al. [2019], who use an ANN as a computational tool to solve the pricing and hedging problem under market frictions such as transaction costs.

[26] At times it was not always clear cut to us whether a paper should be included in Table 1 or in this section. For example, the calibration papers of Section 4.1 could have been put into Table 1 as mentioned in Section 2.2. Similarly, Barucci et al. [1996, 1997], discussed in Section 4.2, learn the Black-Scholes model and hence could have been put into Table 1.

[27] For more details, see also Bayer and Stemper [2018] and Horvath et al. [2019].

4.4 Further work

Albanese et al. [2019] use an ANN to compute the conditional value-at-risk and expected shortfall necessary for certain XVA computations, by solving a quantile regression.

We would like to also mention Halperin [2017] and Kolm and Ritter [2019] who suggest a reinforcement learning methodology to take market frictions into account for the option pricing task.

Finally, generative ANNs have been suggested recently as a non-parametric simulation tool for stock prices; see, for example, Henry-Labordère [2019], Kondratyev and Schwarz [2019], and Wiese et al. [2019b]. Such simulation engines could then be used for option pricing and hedging, a direction still to be explored systematically. Just after finishing this survey, Wiese et al. [2019a] proposed a generative ANN for option prices (instead of stock prices).

5 Digression: regularisation techniques

As the advance of hardware allows for bigger ANNs to be built, regularization techniques have become more important as part of the ANN training. Such techniques include L^2, dropout, early stopping, etc.; see Ormoneit [1999], Gençay and Qi [2001], Gençay and Salih [2003], and Liu et al. [2019b]. Complementing these universal regularisations, several papers embed financial domain knowledge into ANNs, either at the stage of architecture design or training. Let us here also mention the suggested feature design by Lu and Ohta [2003a,b], who consider the pricing of exotic options and suggest to use digital option prices as features.

For the architecture design the following has been suggested:

- Homogeneity hint. Garcia and Gençay [1998, 2000] incorporate a homogeneity hint by considering an ANN consisting of two parts, one controlled by moneyness and the other controlled by time-to-maturity.
- Shape-restricted outputs. Dugas et al. [2001, 2009], Lajbcygier [2004], Yang et al. [2017], Huh [2019], and Zheng et al. [2019] enforce certain no-arbitrage conditions such as monotonicity and convexity of the ANN pricing function by fixing an appropriate architecture.

At the training state the following techniques are being used:

- Data augmentation. Yang et al. [2017] and Zheng et al. [2019] create additional synthetic options to help with the training of ANNs.
- Loss penalty. Itkin [2019] and Ackerer et al. [2019] add various penalty terms to the loss function. Those terms present no-arbitrage conditions. For example, parameter configurations that allow for calendar arbitrage are being penalised.

In the context of ANN training, we would like also to mention Niranjan [1996], de Freitas et al. [2000a,b], and Palmer [2019]. These papers propose and examine novel training algorithms for ANNs and illustrate them in the context of option hedging; these algorithms include the extended Kalman filter, sequential Monte Carlo, and evolutionary algorithms.

References

Y. S. Abu-Mostafa. Financial model calibration using consistency hints. *IEEE transactions on neural networks*, 12(4):791–808, 2001.

D. Ackerer, N. Tagasovska, and T. Vatter. Deep smoothing of the implied volatility surface. SSRN 3402942, 2019.

P. Ahmed and S. Swidler. Forecasting properties of neural network generated volatility estimates. In *Decision Technologies for Computational Finance*, pages 247–258, 1998.

J. J. Ahn, D. H. Kim, K. J. Oh, and T. Y. Kim. Applying option Greeks to directional forecasting of implied volatility in the options market: an intelligent approach. *Expert Systems with Applications*, 39 (10):9315–9322, 2012.

C. Albanese, S. Crépey, R. Hoskinson, and B. Saadeddine. XVA analysis from the balance sheet. Retrieved on October 25, 2019 from `https://math.maths.univ-evry.fr/crepey/`, 2019.

H. Amilon. A neural network versus Black–Scholes: a comparison of pricing and hedging performances. *Journal of Forecasting*, 22:317–335, 2003.

S. Amornwattana, D. Enke, and C. H. Dagli. A hybrid option pricing model using a neural network for estimating volatility. *International Journal of General Systems*, 36(5):558–573, 2007.

U. Anders, O. Korn, and C. Schmitt. Improving the pricing of options: a neural network approach. *Journal of Forecasting*, 17(5-6):369–388, 1998.

P. C. Andreou. *Parametric and Nonparametric Functional Estimation for Options Pricing with Applications in Hedging and Trading*. PhD thesis, University of Cyprus, 2008.

P. C. Andreou, C. Charalambous, and S. H. Martzoukos. Critical assessment of option pricing methods using artificial neural networks. In *International Conference on Artificial Neural Networks*, pages 1131–1136, 2002.

P. C. Andreou, C. Charalambous, and S. H. Martzoukos. Robust artificial neural networks for pricing of European options. *Computational Economics*, 27(2-3):329–351, 2006.

P. C. Andreou, C. Charalambous, and S. H. Martzoukos. Pricing and trading European options by combining artificial neural networks and parametric models with implied parameters. *European Journal of Operational Research*, 185(3):1415–1433, 2008.

P. C. Andreou, C. Charalambous, and S. H. Martzoukos. Generalized parameter functions for option pricing. *Journal of Banking & Finance*, 34(3):633–646, 2010.

M. Avellaneda, A. Carelli, and F. Stella. Following the Bayes path to option pricing. *Journal of Computational Intelligence in Finance*, 1998.

G. Barone-Adesi and R. E. Whaley. Efficient analytic approximation of American option values. *The Journal of Finance*, 42(2):301–320, 1987.

E. Barucci, U. Cherubini, and L. Landi. No-arbitrage asset pricing with neural networks under stochastic volatility. In *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, pages 3–16, 1996.

E. Barucci, U. Cherubini, and L. Landi. Neural networks for contingent claim pricing via the Galerkin method. In *Computational Approaches to Economic Problems*, pages 127–141. 1997.

M. Barunikova and J. Barunik. Neural networks as semiparametric option pricing tool. *Bulletin of the Czech Econometric Society*, 18, 2011.

D. S. Bates. Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options. *The Review of Financial Studies*, 9(1):69–107, 1996.

C. Bayer and B. Stemper. Deep calibration of rough stochastic volatility models. arXiv:1810.03399, 2018.

C. Bayer, B. Horvath, A. Muguruza, B. Stemper, and M. Tomas. On deep calibration of (rough) stochastic volatility models. arXiv:1908.08806, 2019.

C. Beck, S. Becker, P. Cheridito, A. Jentzen, and A. Neufeld. Deep splitting method for parabolic PDEs. arXiv:1907.03452, 2019.

S. Becker, P. Cheridito, and A. Jentzen. Deep optimal stopping. *Journal of Machine Learning Research*, 20 (74):1–25, 2019.

J. Bennell and C. Sutcliffe. Black–Scholes versus artificial neural networks in pricing FTSE 100 options. *Intelligent Systems in Accounting, Finance & Management: International Journal*, 12(4):243–260, 2004.

M. Billio, M. Corazza, and M. Gobbo. Option pricing via regime switching models and multilayer perceptrons: a comparative approach. *Rendiconti per gli Studi Economici Quantitativi*, 2002:39–59, 2002.

L. Blynski and A. Faseruk. Comparison of the effectiveness of option price forecasting: Black–Scholes vs. simple and hybrid neural networks. *Journal of Financial Management & Analysis*, 19(2):46–58, 2006.

C. Boek, P. Lajbcygier, M. Palaniswami, and A. Flitman. A hybrid neural network approach to the pricing of options. In *Proceedings of ICNN'95–International Conference on Neural Networks*, volume 2, pages 813–817. IEEE, 1995.

T. Briegel and V. Tresp. Dynamic neural regression models. Retrieved on August 29, 2019 from `https://epub.ub.uni-muenchen.de/1571/`, 2000.

H. Buehler, L. Gonon, J. Teichmann, and B. Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019a.

H. Buehler, L. Gonon, J. Teichmann, B. Wood, B. Mohan, and J. Kochems. Deep hedging: hedging derivatives under generic market frictions using reinforcement learning. SSRN 3355706, 2019b.

M. Can and Š. Fadda. A nonparametric approach to pricing options learning networks. *Southeast Europe Journal of Soft Computing*, 3(1), 2014.

J. Cao, J. Chen, and J. C. Hull. A neural network approach to understanding implied volatility movements. SSRN 3288067, 2019.

A. Carelli, S. Silani, and F. Stella. Profiling neural networks for option pricing. *International Journal of Theoretical and Applied Finance*, 3(02):183–204, 2000.

P. Carr, H. Geman, D. B. Madan, and M. Yor. Stochastic volatility for Lévy processes. *Mathematical Finance*, 13(3):345–382, 2003.

A. P. Carverhill and T. H. Cheuk. Alternative neural network approach for option pricing and hedging. SSRN 480562, 2003.

Q. Chan-Wai-Nam, J. Mikael, and X. Warin. Machine learning for semi linear PDEs. *Journal of Scientific Computing*, 79(3):1667–1712, 2019.

T.-Y. Chang, Y.-H. Wang, and H.-Y. Yeh. Forecasting of option prices using a neural network model. *Journal of Accounting, Finance & Management Strategy*, 8(1):123–136, 2013.

C. Charalambous and S. H. Martzoukos. Hybrid artificial neural networks for efficient valuation of real options and financial derivatives. *Computational Management Science*, 2(2):155–161, 2005.

F. Chen and C. Sutcliffe. Pricing and hedging short sterling options using neural networks. *Intelligent Systems in Accounting, Finance and Management*, 19(2):128–149, 2012.

J. Chen. Learning the Black–Scholes formula via support vector machines. In *Recent Advances in Statistics Application and Related Areas, 2nd Conference of the International Institute of Applied Statistics Studies*, volume 1&2, pages 756–760, 2009.

S.-H. Chen and W.-C. Lee. Pricing call warrants with artificial neural networks: the case of the Taiwan derivative market. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)*, volume 6, pages 3877–3882. IEEE, 1999.

D.-Y. Chiu and C.-C. Lin. Exploring internal mechanism of warrant in financial market with a hybrid approach. *Expert Systems with Applications*, 35(3):1237–1245, 2008.

H.-J. Choi, H.-S. Lee, G.-S. Han, and J. Lee. Efficient option pricing via a globally regularized neural network. In *International Symposium on Neural Networks*, pages 988–993, 2004.

C. J. Corrado and T. Su. Skewness and kurtosis in S&P 500 index returns implied by option prices. *Journal of Financial Research*, 19(2):175–192, 1996.

J. C. Cox, S. A. Ross, and M. Rubinstein. Option pricing: a simplified approach. *Journal of Financial Economics*, 7(3):229–263, 1979.

R. Culkin and S. R. Das. Machine learning in finance: the case of deep learning for option pricing. *Journal of Investment Management*, 15(4):92–100, 2017.

S. P. Das and S. Padhy. A new hybrid parametric and machine learning model with homogeneity hint for European-style index option pricing. *Neural Computing and Applications*, 28(12):4061–4077, 2017.

J. F. G. de Freitas, M. Niranjan, and A. H. Gee. Hierarchical Bayesian models for regularization in sequential learning. *Neural Computation*, 12(4):933–953, 2000a.

J. F. G. de Freitas, M. Niranjan, A. H. Gee, and A. Doucet. Sequential Monte Carlo methods to train neural network models. *Neural Computation*, 12(4):955–993, 2000b.

G. Dimitroff, D. Röder, and C. Fries. Volatility model calibration with convolutional neural networks. SSRN 3252432, 2018.

Z. A. Dindar and T. Marwala. Option pricing using a committee of neural networks and optimized networks. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, volume 1, pages 434–438. IEEE, 2004.

C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia. Incorporating second-order functional knowledge for better option pricing. In *Advances in Neural Information Processing Systems*, pages 472–478, 2001.

C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia. Incorporating functional knowledge in neural networks. *Journal of Machine Learning Research*, 10(Jun):1239–1262, 2009.

W. E, J. Han, and A. Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, 5(4):349–380, 2017.

Z. Fang and K. George. Application of machine learning: an analysis of Asian options pricing using neural network. In *2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)*, pages 142–149. IEEE, 2017.

S. Fecamp, J. Mikael, and X. Warin. Risk management with machine-learning-based algorithms. arXiv:1902.05287, 2019.

R. Ferguson and A. Green. Deeply learning derivatives. SSRN 3244821, 2018.

J. Galindo-Flores. A framework for comparative analysis of statistical and machine learning methods: an application to the Black–Scholes option pricing model. *Computational Finance 1999*, pages 635–660, 2000.

R. Garcia and R. Gençay. Option pricing with neural networks and a homogeneity hint. In *Decision Technologies for Computational Finance*, pages 195–205, 1998.

R. Garcia and R. Gençay. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, 94(1-2):93–115, 2000.

J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.

D. S. Geigle. *An Artificial Neural Network Approach to the Valuation of Options and Forecasting of Volatility*. PhD thesis, Nova Southeastern University, 1999.

D. S. Geigle and J. E. Aronson. An artificial neural network approach to the valuation of options and forecasting of volatility. *Journal of Computational Intelligence in Finance*, 7(6):19–25, 1999.

R. Gençay and R. Gibson. Model risk for European-style stock index options. *IEEE Transactions on Neural Networks*, 18(1):193–202, 2007.

R. Gençay and M. Qi. Pricing and hedging derivative securities with neural networks: Bayesian regularization, early stopping, and bagging. *IEEE Transactions on Neural Networks*, 12(4):726–734, 2001.

R. Gençay and A. Salih. Degree of mispricing with the Black–Scholes model and nonparametric cures. *Annals of Economics and Finance*, 4:73–101, 2003.

H. Ghaziri, S. Elfakhani, and J. Assi. Neural networks approach to pricing options. *Neural Network World*, 10(1):271–277, 2000.

J. Ghosn and Y. Bengio. Multi-task learning for option pricing. Retrieved on October 29, 2019 from `https://cirano.qc.ca/files/publications/2002s-53.pdf`, 2002.

E. Ghysels, V. Patilea, É. Renault, and O. Torrès. Nonparametric methods and option pricing. In D. Hand and S. Jacka, editors, *Statistics in Finance*, chapter 13, pages 261–282. John Wiley & Sons, 1998.

N. Gradojevic and D. Kukolj. Parametric option pricing: a divide-and-conquer approach. *Physica D: Nonlinear Phenomena*, 240(19):1528–1535, 2011.

N. Gradojevic, R. Gençay, and D. Kukolj. Option pricing with modular neural networks. *IEEE Transactions on Neural Networks*, 20(4):626–637, 2009.

A. Gregoriou, J. Healy, and C. Ioannidis. Hedging under the influence of transaction costs: an empirical investigation on FTSE 100 index options. *Journal of Futures Markets*, 27(5):471–494, 2007.

J. T. Hahn. *Option Pricing Using Artificial Neural Networks: An Australian Perspective*. PhD thesis, Bond University, 2013.

I. Halperin. QLBS: Q-learner in the Black-Scholes (-Merton) worlds. arXiv:1712.04609, 2017.

S. A. Hamid and A. Habib. Can neural networks learn the Black-Scholes model?: A simplified approach. Retrieved on September 9, 2019 from `https://academicarchive.snhu.edu/bitstream/handle/10474/1662/cfs2005-01.pdf`, 2005.

J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.

M. Hanke. Neural network approximation of option pricing formulas for analytically intractable option pricing models. *Journal of Computational Intelligence in Finance*, 5(5):20–27, 1997.

M. Hanke. Adaptive hybrid neural network option pricing. *Journal of Computational Intelligence in Finance*, 7(5):33–39, 1999a.

M. Hanke. Neural networks versus Black-Scholes: an empirical comparison of the pricing accuracy of two fundamentally different option pricing methods. *Journal of Computational Intelligence in Finance*, 5: 26–34, 1999b.

J. Healy, M. Dixon, B. Read, and F. Cai. A data-centric approach to understanding the pricing of financial options. *The European Physical Journal B*, 27(2):219–227, 2002.

J. V. Healy, M. Dixon, B. J. Read, and F. F. Cai. Confidence in data mining model predictions: a financial engineering application. In *IECON'03. 29th Annual Conference of the IEEE Industrial Electronics Society (IEEE Cat. No. 03CH37468)*, volume 2, pages 1926–1931. IEEE, 2003.

J. V. Healy, M. Dixon, B. J. Read, and F. F. Cai. Confidence limits for data mining models of options prices. *Physica A: Statistical Mechanics and its Applications*, 344(1-2):162–167, 2004.

J. V. Healy, M. Dixon, B. J. Read, and F. F. Cai. Non-parametric extraction of implied asset price distributions. *Physica A: Statistical Mechanics and its Applications*, 382(1):121–128, 2007.

P. Henry-Labordère. Deep primal-dual algorithm for BSDEs: applications of machine learning to CVA and IM. SSRN 3071506, 2017.

P. Henry-Labordère. Generative models for financial data. SSRN 3408007, 2019.

A. Hernandez. Model calibration with neural networks. *Risk Magazine*, pages 1–5, June 2017.

R. Herrmann and A. Narr. Neural networks and the evaluation of derivatives: some insights into the implied pricing mechanism of german stock index options. Retrieved on August 29, 2019 from `http://finance.fbv.kit.edu/download/dp202.pdf`, 1997.

S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

B. Horvath, A. Muguruza, and M. Tomas. Deep learning volatility. arXiv:1901.09647, 2019.

S.-C. Huang. Online option price forecasting by using unscented Kalman filters and support vector machines. *Expert Systems with Applications*, 34(4):2819–2825, 2008.

S.-C. Huang and T.-K. Wu. A hybrid unscented Kalman filter and support vector machine model in option price forecasting. In *International Conference on Natural Computation*, pages 303–312, 2006.

J. Huh. Pricing options with exponential Lévy neural network. *Expert Systems with Applications*, 127: 128–140, 2019.

J. Hull and A. White. Optimal delta hedging for options. *Journal of Banking & Finance*, 82:180–190, 2017.

C. Huré, H. Pham, and X. Warin. Some machine learning schemes for high-dimensional nonlinear PDEs. arXiv:1902.01599, 2019.

J. M. Hutchinson, A. W. Lo, and T. Poggio. A nonparametric approach to pricing and hedging derivative securities via learning networks. *The Journal of Finance*, 49(3):851–889, 1994.

A. Itkin. Deep learning calibration of option pricing models: some pitfalls and solutions. arXiv:1906.03507, 2019.

A. Jacquier and M. Oumgari. Deep PPDEs for rough local stochastic volatility. arXiv:1906.02551, 2019.

H. Jang and J. Lee. Generative Bayesian neural network model for risk-neutral pricing of American index options. *Quantitative Finance*, 19(4):587–603, 2019.

K.-H. Jung, H.-C. Kim, and J. Lee. A novel learning network for option pricing with confidence interval information. In *International Symposium on Neural Networks*, pages 491–497, 2006.

M. Kakati. Pricing and hedging performances of artificial neural net in Indian stock option market. *The ICFAI Journal of Applied Finance*, 11(1):62–73, 2005.

M. Kakati. Option pricing using Adaptive Neuro-Fuzzy System (ANFIS). *ICFAI Journal of Derivatives Markets*, 5(2), 2008.

O. Karaali, W. Edelberg, and J. Higgins. Modelling volatility derivatives using neural networks. In *Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering*, pages 280–286. IEEE, 1997.

T. Karatas, A. Oskoui, and A. Hirsa. Supervised deep neural networks (DNNs) for pricing/calibration of vanilla/exotic options under various different processes. arXiv:1902.05810, 2019.

D. L. Kelly. Valuing and hedging American put options using neural networks. Retrieved on August 29, 2019 from `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.721.8497&rep=rep1&type=pdf`, 1994.

B.-H. Kim, D. Lee, and J. Lee. Local volatility function approximation using reconstructed radial basis function networks. In *International Symposium on Neural Networks*, pages 524–530, 2006.

P. Ko, P. Lin, W. Chien, and Y. Cheng. Hedging derivative securities based on the neural network coefficient model. In *Proceedings of the Eighth Joint Conference on Information Sciences*, pages 1163–1166, 2005.

P.-C. Ko. Option valuation based on the neural regression model. *Expert Systems with Applications*, 36(1): 464–471, 2009.

M. Kohler, A. Krzyżak, and N. Todorovic. Pricing of high-dimensional American options by neural networks. *Mathematical Finance*, 20(3):383–410, 2010.

P. N. Kolm and G. Ritter. Dynamic replication and hedging: A reinforcement learning approach. *The Journal of Financial Data Science*, 1(1):159–171, 2019.

A. Kondratyev and C. Schwarz. The market generator. SSRN 3384948, 2019.

S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101, 2002.

J. Krause. Option pricing with neural networks. In *Proceedings of the Fourth European Congress on Intelligent Techniques and Soft Computing*, volume 3, pages 2206–2210, 1996.

G. Lachtermacher and L. Rodrigues Gaspar. Neural networks in derivative securities pricing forecasting in Brazilian capital markets. In *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, pages 92–97, 1996.

W.-N. Lai. Comparison of methods to estimate option implied risk-neutral densities. *Quantitative Finance*, 14(10):1839–1855, 2014.

P. R. Lajbcygier. Comparing conventional and artificial neural network models for the pricing of options. In *Neural Networks in Business: Techniques and Applications*, pages 220–235. IGI Global, 2002.

P. R. Lajbcygier. Improving option pricing with the product constrained hybrid neural network. In *Artificial Neural Networks and Neural Information Processing*, pages 615–621, 2003.

P. R. Lajbcygier. Improving option pricing with the product constrained hybrid neural network. *IEEE Transactions on Neural Networks*, 15(2):465–476, 2004.

P. R. Lajbcygier and J. T. Connor. Improved option pricing using artificial neural networks and bootstrap methods. *International Journal of Neural Systems*, 8(04):457–471, 1997a.

P. R. Lajbcygier and J. T. Connor. Improved option pricing using bootstrap methods. In *Proceedings of International Conference on Neural Networks*, volume 4, pages 2193–2197. IEEE, 1997b.

P. R. Lajbcygier and A. Flitman. A comparison of non-parametric regression techniques for the pricing of options using an optimal implied volatility. In *Decision Technologies for Financial Engineering: Proceedings of the Fourth International Conference on Neural Networks in Capital Markets*, pages 201–213, 1996.

P. R. Lajbcygier, C. Boek, A. Flitman, and M. Palaniswami. Comparing conventional and artificial neural network models for the pricing of options on futures. *NeuroVe$t Journal*, 4(5):16–24, 1996a.

P. R. Lajbcygier, C. Boek, M. Palaniswami, and A. Flitman. Neural network pricing of all ordinaries SPI options on futures. In *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, 1996b.

P. R. Lajbcygier, A. Flitman, A. Swan, and R. J. Hyndman. The pricing and trading of options using a hybrid neural network model with historical volatility. *NeuroVe$t Journal*, pages 27–41, 1997.

L. J. le Roux and G. S. du Toit. Emulating the Black & Scholes model with a neural network. *Southern African Business Review*, 5(1):54–57, 2001.

M. T. Leung, A.-S. Chen, and R. Mancha. Making trading decisions for financial-engineered derivatives: a novel ensemble of neural networks using information content. *Intelligent Systems in Accounting, Finance & Management*, 16(4):257–277, 2009.

X. Liang, H. Zhang, and J. Yang. Pricing options in Hong Kong market based on neural networks. In *International Conference on Neural Information Processing*, pages 410–419, 2006.

X. Liang, H. Zhang, J. Xiao, and Y. Chen. Improving option price forecasts with neural networks and support vector regressions. *Neurocomputing*, 72(13-15):3055–3065, 2009.

C.-T. Lin and H.-Y. Yeh. The valuation of Taiwan stock index option price—comparison of performances between Black-Scholes and neural network model. *Journal of Statistics and Management Systems*, 8(2): 355–367, 2005.

D. Liu and S. Huang. The performance of hybrid artificial neural network models for option pricing during financial crises. *Journal of Data Science*, 14(1):1–18, 2016.

D. Liu and L. Zhang. Pricing Chinese warrants using artificial neural networks coupled with Markov regime switching model. *International Journal of Financial Markets and Derivatives*, 2(4):314–330, 2011.

M. Liu. Option pricing with neural networks. In *Progress in Neural Information Processing*, volume 2, pages 760–765, 1996.

S. Liu, A. Borovykh, L. A. Grzelak, and C. W. Oosterlee. A neural network-based framework for financial model calibration. *Journal of Mathematics in Industry*, Forthcoming, 2019a.

S. Liu, C. W. Oosterlee, and S. M. Bohte. Pricing options and computing implied volatilities using neural networks. *Risks*, 7(1):1–22, 2019b.

X. Liu, Y. Cao, C. Ma, and L. Shen. Wavelet-based option pricing: an empirical study. *European Journal of Operational Research*, 272(3):1132–1142, 2019c.

F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *The Review of Financial Studies*, 14(1):113–147, 2001.

J. Lu and H. Ohta. A data and digital-contracts driven method for pricing complex derivatives. *Quantitative Finance*, 3(3):212–219, 2003a.

J. Lu and H. Ohta. Digital contracts-driven method for pricing complex derivatives. *Journal of the Operational Research Society*, 54(9):1002–1010, 2003b.

M. Ludwig. Robust estimation of shape-constrained state price density surfaces. *The Journal of Derivatives*, 22(3):56–72, 2015.

D. B. Madan, P. P. Carr, and E. C. Chang. The variance Gamma process and option pricing. *Review of Finance*, 2(1):79–105, 1998.

M. Malliaris and L. Salchenberger. Beating the best: a neural network challenges the Black-Scholes formula. In *Proceedings of 9th IEEE Conference on Artificial Intelligence for Applications*, pages 445–449. IEEE, 1993a.

M. Malliaris and L. Salchenberger. A neural network model for estimating option prices. *Journal of Applied Intelligence*, 3(3):193–206, 1993b.

M. Malliaris and L. Salchenberger. Using neural networks to forecast the S&P100 implied volatility. *Neurocomputing*, 10(2):183–195, 1996.

C. G. Martel, M. D. G. Artiles, and F. F. Rodriguez. A financial option pricing model based on learning algorithms. In *Proceedings of the World Multiconference on Applied Economics, Business and Development*, pages 153–157, 2009.

W. A. McGhee. An artificial neural network representation of the SABR stochastic volatility model. SSRN 3288882, 2018.

G. Meissner and N. Kawano. Capturing the volatility smile of options on high-tech stocks—a combined GARCH-neural network approach. *Journal of Economics and Finance*, 25(3):276–292, 2001.

F. G. Miranda and N. Burgess. Intraday volatility forecasting for option pricing using a neural network approach. In *Proceedings of 1995 Conference on Computational Intelligence for Financial Engineering*, page 31. IEEE, 1995.

S. K. Mitra. Improving accuracy of option price estimation using artificial neural networks. SSRN 876881, 2006.

S. K. Mitra. An option pricing model that combines neural network approach and Black Scholes formula. *Global Journal of Computer Science and Technology*, 12(4), 2012.

G. Montagna, M. Morelli, O. Nicrosini, P. Amato, and M. Farina. Pricing derivatives by path integral and neural networks. *Physica A: Statistical Mechanics and its Applications*, 324(1-2):189–195, 2003.

L. Montesdeoca and M. Niranjan. Extending the feature set of a data-driven artificial neural network model of pricing financial options. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6. IEEE, 2016.

M. J. Morelli, G. Montagna, O. Nicrosini, M. Treccani, M. Farina, and P. Amato. Pricing financial derivatives with neural networks. *Physica A: Statistical Mechanics and its Applications*, 338(1-2):160–165, 2004.

F. Mostafa. *Applications of Neural Networks in Market Risk*. PhD thesis, Curtin University, 2011.

F. Mostafa and T. Dillon. A neural network approach to option pricing. *WIT Transactions on Information and Communication Technologies*, 41:71–85, 2008.

M. Niranjan. Sequential tracking in pricing financial options using model based and neural network approaches. In *Advances in Neural Information Processing Systems*, pages 960–966, 1996.

D. Ormoneit. A regularization approach to continuous learning with an application to financial derivatives pricing. *Neural Networks*, 12(10):1405–1412, 1999.

S. Palmer. *Evolutionary Algorithms and Computational Methods for Derivatives Pricing*. PhD thesis, University College London, 2019.

S. Palmer and D. Gorse. Pseudo-analytical solutions for stochastic options pricing using Monte Carlo simulation and breeding PSO-trained neural networks. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 365–370, 2017.

A. Pande and R. Sahu. A new approach to volatility estimation and option price prediction for dividend paying stocks. In *WEHIA 2006–1st International Conference on Economic Sciences with Heterogeneous Interacting Agents; 15–17 June 2006, University of Bologna, Italy*, 2006.

H. Park, N. Kim, and J. Lee. Parametric models and non-parametric machine learning models for predicting option prices: empirical comparison study over KOSPI 200 index options. *Expert Systems with Applications*, 41(11):5227–5237, 2014.

B. Phani, B. Chandra, and V. Raghav. Quest for efficient option pricing prediction model using machine learning techniques. In *The 2011 International Joint Conference on Neural Networks*, pages 654–657. IEEE, 2011.

M. M. Pires and T. Marwala. American option pricing using multi-layer perceptron and support vector machine. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, volume 2, pages 1279–1285. IEEE, 2004a.

M. M. Pires and T. Marwala. Option pricing using Bayesian neural networks. In *Fifteenth Annual Symposium of the Pattern Recognition Association of South Africa*, pages 161–166, 2004b.

M. M. Pires and T. Marwala. American option pricing using Bayesian multi-layer perceptrons and Bayesian support vector machines. In *IEEE 3rd International Conference on Computational Cybernetics*, pages 219–224. IEEE, 2005.

M. Qi. *Financial Applications of Generalized Nonlinear Nonparametric Econometric Methods (Artificial Neural Networks)*. PhD thesis, Ohio State University, 1996.

M. Qi and G. Maddala. Option pricing using artificial neural networks: the case of S&P 500 index call options. In *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets*, pages 78–91, 1996.

C. Quek, M. Pasquier, and N. Kumar. A novel recurrent neural network-based prediction system for option trading and hedging. *Applied Intelligence*, 29(2):138–151, 2008.

M. Raberto, G. Cuniberti, M. Riani, E. Scales, F. Mainardi, and G. Servizi. Learning short-option valuation in the presence of rare events. *International Journal of Theoretical and Applied Finance*, 3(03):563–564, 2000.

J. Ruf and W. Wang. Hedging with neural networks. SSRN 3580132, 2020.

S. Saito and L. Jun. Neural network option pricing in connection with the Black and Scholes model. In *Proceedings of the Fifth Conference of the Asian Pacific Operations Research Society*, 2000.

Z. I. Samur and G. T. Temur. The use of artificial neural network in option pricing: the case of S&P 100 index options. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 3(6):644–649, 2009.

A. Saxena. Valuation of S&P CNX Nifty options: comparison of Black-Scholes and hybrid ANN model. In *Proceedings SAS Global Forum*, 2008.

C. Schittenkopf and G. Dorffner. Risk-neutral density extraction from option prices: improved pricing with mixture density networks. *IEEE Transactions on Neural Networks*, 12(4):716–725, 2001.

H. J. Shin and J. Ryu. A dynamic hedging strategy for option transaction using artificial neural networks. *International Journal of Software Engineering and its Applications*, 6(4):111–116, 2012.

J. Sirignano and K. Spiliopoulos. DGM: a deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018.

H. Stone. Calibrating rough volatility models: a convolutional neural network approach. *Quantitative Finance*, pages 1–14, 2019.

A. Taudes, M. Natter, and M. Trcka. Real option valuation with neural networks. *Intelligent Systems in Accounting, Finance & Management*, 7(1):43–52, 1998.

S. D. Teddy, E.-K. Lai, and C. Quek. A brain-inspired cerebellar associative memory approach to option pricing and arbitrage trading. In *International Conference on Neural Information Processing*, pages 370–379, 2006.

S. D. Teddy, E.-K. Lai, and C. Quek. A cerebellar associative memory approach to option pricing and arbitrage trading. *Neurocomputing*, 71(16-18):3303–3315, 2008.

N. S. Thomaidis, V. S. Tzastoudis, and G. Dounias. A comparison of neural network model selection strategies for the pricing of S&P500 stock index options. *International Journal on Artificial Intelligence Tools*, 16(06):1093–1113, 2007.

R. Tsaih. Sensitivity analysis, neural networks, and the finance. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No. 99CH36339)*, volume 6, pages 3830–3835. IEEE, 1999.

C.-H. Tseng, S.-T. Cheng, Y.-H. Wang, and J.-T. Peng. Artificial neural network model of the hybrid EGARCH volatility of the Taiwan stock index option prices. *Physica A: Statistical Mechanics and its Applications*, 387(13):3192–3200, 2008.

W. L. Tung and C. Quek. GenSo-OPATS: a brain-inspired dynamically evolving option pricing model and arbitrage trading system. In *2005 IEEE Congress on Evolutionary Computation*, volume 3, pages 2429–2436. IEEE, 2005.

W. L. Tung and C. Quek. Financial volatility trading using a self-organising neural-fuzzy semantic network and option straddle-based approach. *Expert Systems with Applications*, 38(5):4668–4688, 2011.

V. S. Tzastoudis, N. S. Thomaidis, and G. D. Dounias. Improving neural network based option price forecasting. In *Hellenic Conference on Artificial Intelligence*, pages 378–388, 2006.

M. S. Vidales, D. Siska, and L. Szpruch. Unbiased deep solvers for parametric PDEs. arXiv:1810.05094, 2019.

C. von Spreckelsen, H.-J. von Mettenheim, and M. H. Breitner. Steps towards a high-frequency financial decision support system to pricing options on currency futures with neural networks. *International Journal of Applied Decision Sciences*, 7(3):223–238, 2014.

C.-P. Wang, S.-H. Lin, H.-H. Huang, and P.-C. Wu. Using neural network for forecasting TXO price under different volatility models. *Expert Systems with Applications*, 39(5):5025–5032, 2012.

H.-W. Wang. Dual derivatives spreading and hedging with evolutionary data mining. *Journal of American Academy of Business*, 9:45–52, 2006.

P. Wang. Pricing currency options with support vector regression and stochastic volatility model with jumps. *Expert Systems with Applications*, 38(1):1–7, 2011.

Y.-H. Wang. Nonlinear neural network forecasting model for stock index option price: hybrid GJR–GARCH approach. *Expert Systems with Applications*, 36(1):564–570, 2009a.

Y.-H. Wang. Using neural network to forecast stock index option price: a new hybrid GARCH approach. *Quality & Quantity*, 43(5):833–843, 2009b.

A. White. *Pricing Options with Futures-Style Margining: a Genetic Adaptive Neural Network Approach*. Garland Publishing, 2000.

A. J. White. A genetic adaptive neural network approach to pricing options: a simulation analysis. *Journal of Computational Intelligence in Finance*, 6(2):13–23, 1998.

M. Wiese, L. Bai, B. Wood, and H. Buehler. Deep hedging: learning to simulate equity option markets. arXiv:1911.01700, 2019a.

M. Wiese, R. Knobloch, R. Korn, and P. Kretschmer. Quant GANs: deep generation of financial time series. arXiv:1907.06673, 2019b.

L. Xu, M. Dixon, B. A. Eales, F. F. Cai, B. J. Read, and J. V. Healy. Barrier option pricing: modelling with neural nets. *Physica A: Statistical Mechanics and its Applications*, 344(1-2):289–293, 2004.

Y. Yang, Y. Zheng, and T. M. Hospedales. Gated neural networks for option pricing: rationality by design. In *Association for the Advancement of Artificial Intelligence*, pages 52–58, 2017.

J. Yao, Y. Li, and C. L. Tan. Option price forecasting using neural networks. *Omega*, 28(4):455–466, 2000.

T. Ye and L. Zhang. Derivatives pricing via machine learning. SSRN 3352688, 2019.

C. Zapart. Stochastic volatility options pricing with wavelets and artificial neural networks. *Quantitative Finance*, 2(6):487–495, 2002.

C. Zapart. Beyond Black–Scholes: a neural networks-based approach to options pricing. *International Journal of Theoretical and Applied Finance*, 6(05):469–489, 2003a.

C. Zapart. Statistical arbitrage trading with wavelets and artificial neural networks. In *2003 IEEE International Conference on Computational Intelligence for Financial Engineering*, pages 429–435. IEEE, 2003b.

Y. Zheng. *Machine Learning and Option Implied Information*. PhD thesis, Imperial College London, 2017.

Y. Zheng, Y. Yang, and B. Chen. Gated deep neural networks for implied volatility surfaces. arXiv:1904.12834, 2019.

W. Zhou, M. Yang, and L. Han. A nonparametric approach to pricing convertible bond via neural network. In *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007)*, volume 2, pages 564–569. IEEE, 2007.