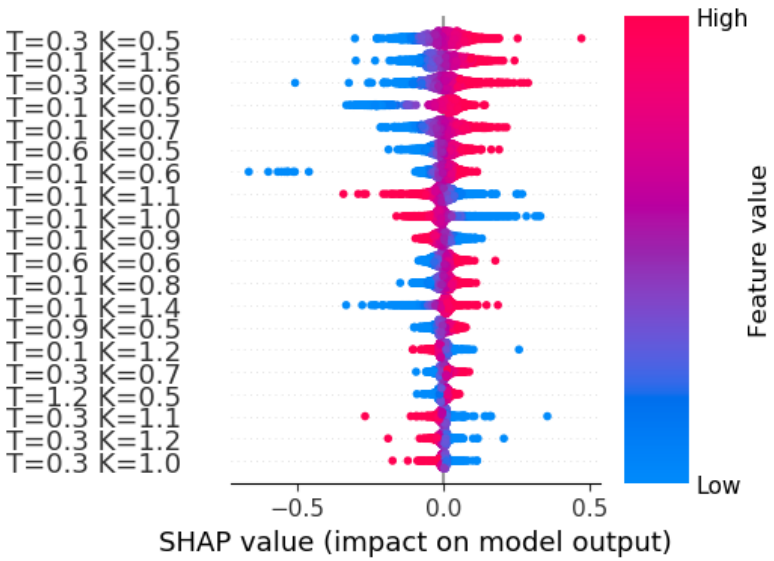
Interpretability in Deep Learning for Finance: A Case Study for the Heston Model



Andrea Pallavicini

Intesa Sanpaolo & Imperial College

joint work with

Damiano Brigo

Xiaoshan Huang

Haitz Sáez de Ocáriz Borde

Imperial College

POLITECNICO
MILANO 1863

Big Data and Machine Learning in Finance
*M*ilan, 10-11 June 2021

Problem Statement and Motivations – I

- More details in the online paper Brigo et al. (2021).
- Artificial neural networks (NN) behave as black boxes and this hinders validation and accountability processes.
  $\longrightarrow$ Being able to interpret the input-output relationship of these NN has become key for the acceptance of such tools.
- We focus on the calibration process of a stochastic volatility model, a subject recently tackled by deep learning algorithms.
  $\longrightarrow$ We analyze the Heston model in particular, as this model properties are well known, resulting in an ideal benchmark case.
  $\longrightarrow$ This topic has been explored by several authors such as Hernandez (2017), Bayer et al. (2018), Horvath et al. (2021), Bloch (2019), and Roeder et al. (2020).

Problem Statement and Motivations – II

- We investigate the capability of local vs. global methods to explain the trained NN.
  - $\longrightarrow$ We find that global methods, coming from cooperative game theory such as Shapley values, can be effectively used in practice.
  - $\longrightarrow$ In Chakraborty et al. (2017) a survey of prior works on interpretability in deep learning models is presented.
- Interpretability methods are a hot topic in the machine learning literature, but marginally in the financial literature.
  - $\longrightarrow$ Applications in trading strategies, credit scoring, and business analytics can be found in Wang et al. (2019), Demajo et al. (2020), Moehle et al. (2021), and Kraus et al. (2020).
- Our analysis also highlights that Shapley values may help to choose the NN architecture.

Talk Outline

Neural Network Calibrations

- When a pricing model is conceived, the performance of the calibration process is of paramount importance.
- A possible way to deal with this issue, and ensuring a fair level of accuracy, speed, and robustness, is using a NN as part of the calibration to speed up the process.
- A first example is given by Horvath et al. (2021) where:
  - $\longrightarrow$ a NN learns the pricing map from model parameters to market quotes,
  - $\longrightarrow$ the calibration is performed by using the trained NN to efficiently find the market quotes from the model parameters.
- In this approach the NN is used only to obtain a faster version of the pricing map, without learning the whole calibration procedure.
- We take a different way and use the NN to learn the whole calibration procedure, as in Roeder et al. (2020) and Hernandez (2017).

Interpreting the Calibration Results – I

- In our calibration problem interpretability methods allow us to understand which input volatilities affect the most our model parameters.
- These methods can be useful in two different situations:
  - $\longrightarrow$ if we have good knowledge of our model, we can test if the NN architecture matches our understanding, while
  - $\longrightarrow$ if we lack such knowledge, we can use these methods precisely to improve our understanding of the model behaviour.

Interpreting the Calibration Results – II

- We can classify interpretability methods in two main categories according to Molnar (2020): global and local.
- Global methods aim at recognizing
  - $\longrightarrow$ how the model makes decisions based on its overall structure, and
  - $\longrightarrow$ how the distribution of the target outcomes occur given the features.
- In contrast, local methods employs simpler surrogate models to explain a single prediction at a time.
- Here, we compare the local method LIME with the global method SHAP.
  - $\longrightarrow$ In the online paper Brigo et al. (2021) also the local methods DeepLIFT and LRP are analysed.

Talk Outline

The Heston Model

- We assume that the price process S_t follows a Heston model under the risk-neutral measure, namely we write

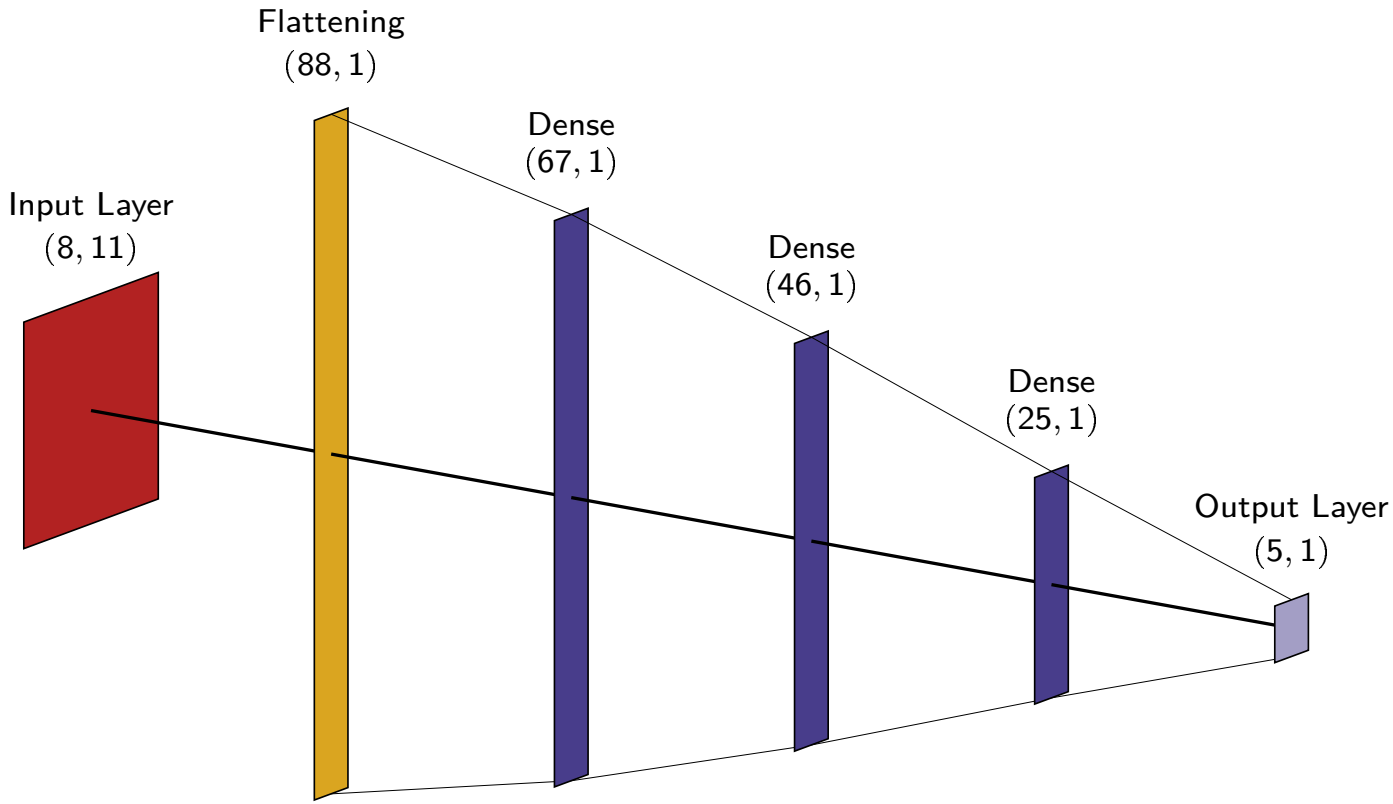
$$dS_t = \sqrt{v_t} S_t dW_t \ , \quad dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t} dW_t^v$$

  where W_t and W_t^v are Brownian motions with correlation ρ.
- All the free parameters are collected in the set $\psi := \{v_0, \rho, \sigma, \theta, \kappa\}$.
- We aim at finding the optimal model parameters $\psi^\star$ such that the model prices best match the market data.

$$C^{\text{mkt}} \longrightarrow \psi^\star := \arg\min_{\psi} \sum_{i=1}^{n} \left(C(\psi; K_i, T_i) - C_i^{\text{mkt}} \right)^2$$

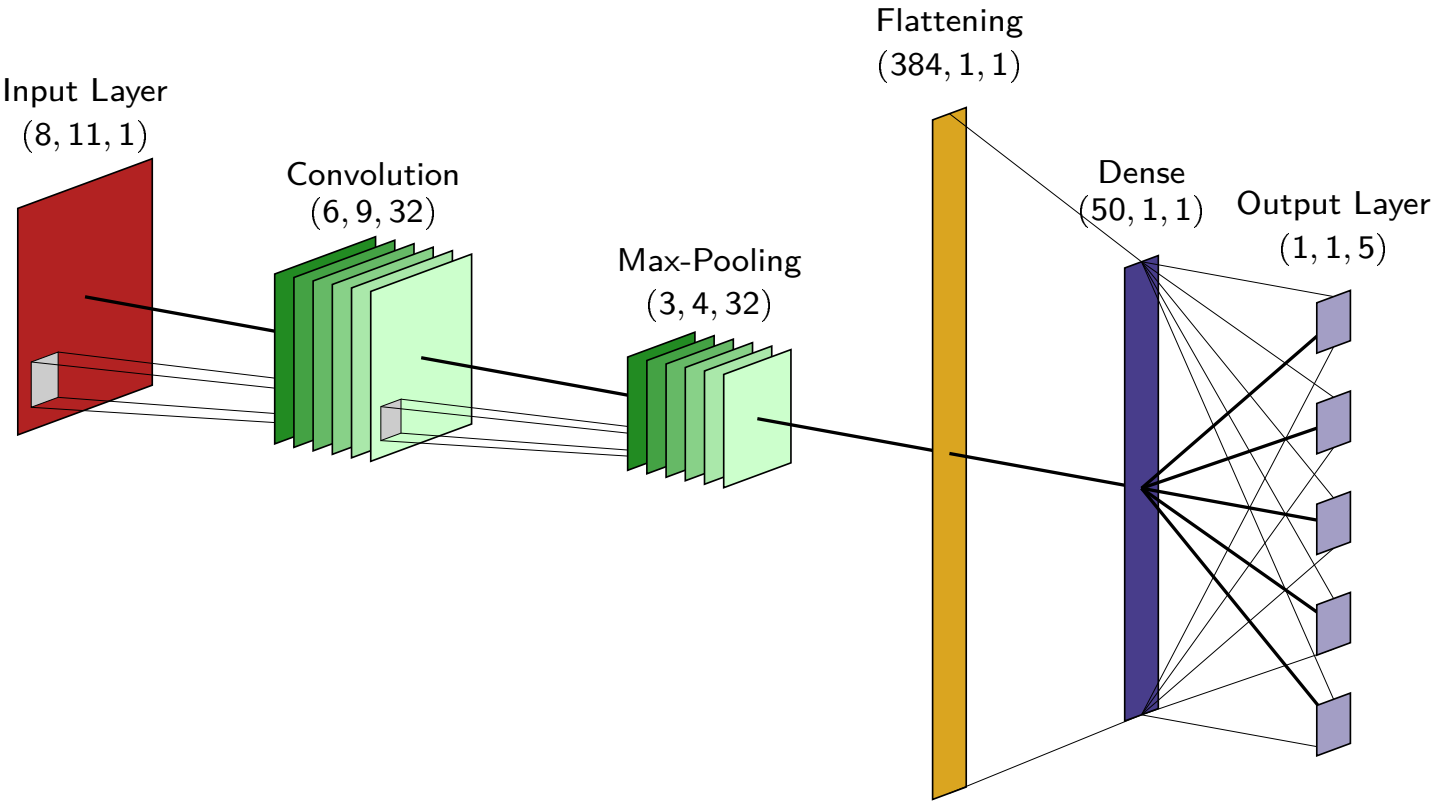
- We proceed by introducing two different NN to approximate the previous map: a fully connected network (FCNN) and a convolutional network (CNN).

Neural Network Architectures – I



FCNN architecture summary. The data flow is from left to right and it shows the flattening and fully-connected layers. The coordinates below each layer name represent the data set dimension.

Neural Network Architectures – II

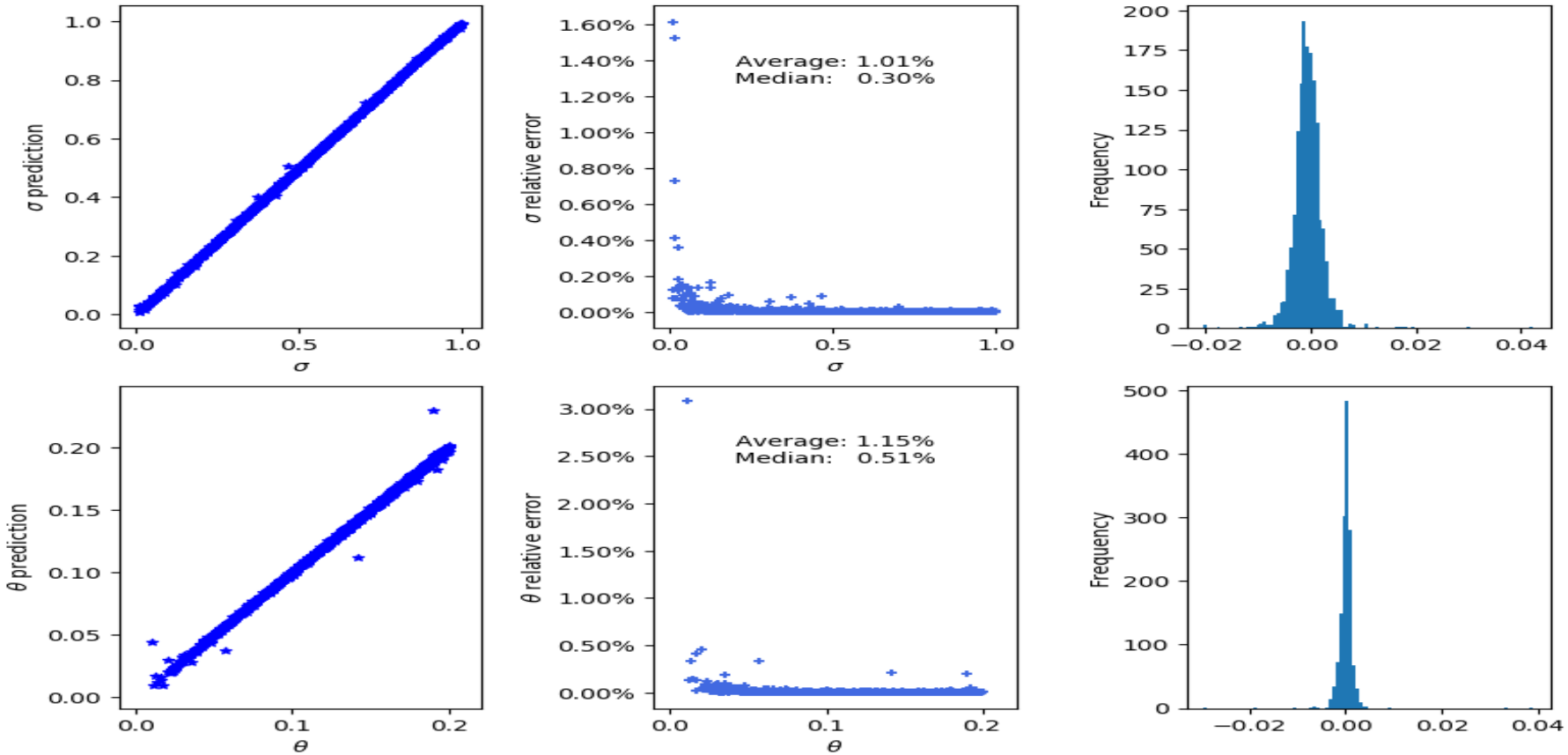


CNN architecture summary. The data flow is from left to right and it shows the convolution, max-pooling, flattening, and fully-connected layers. The coordinates below each layer name represent the data set dimension.

Calibration Results – I

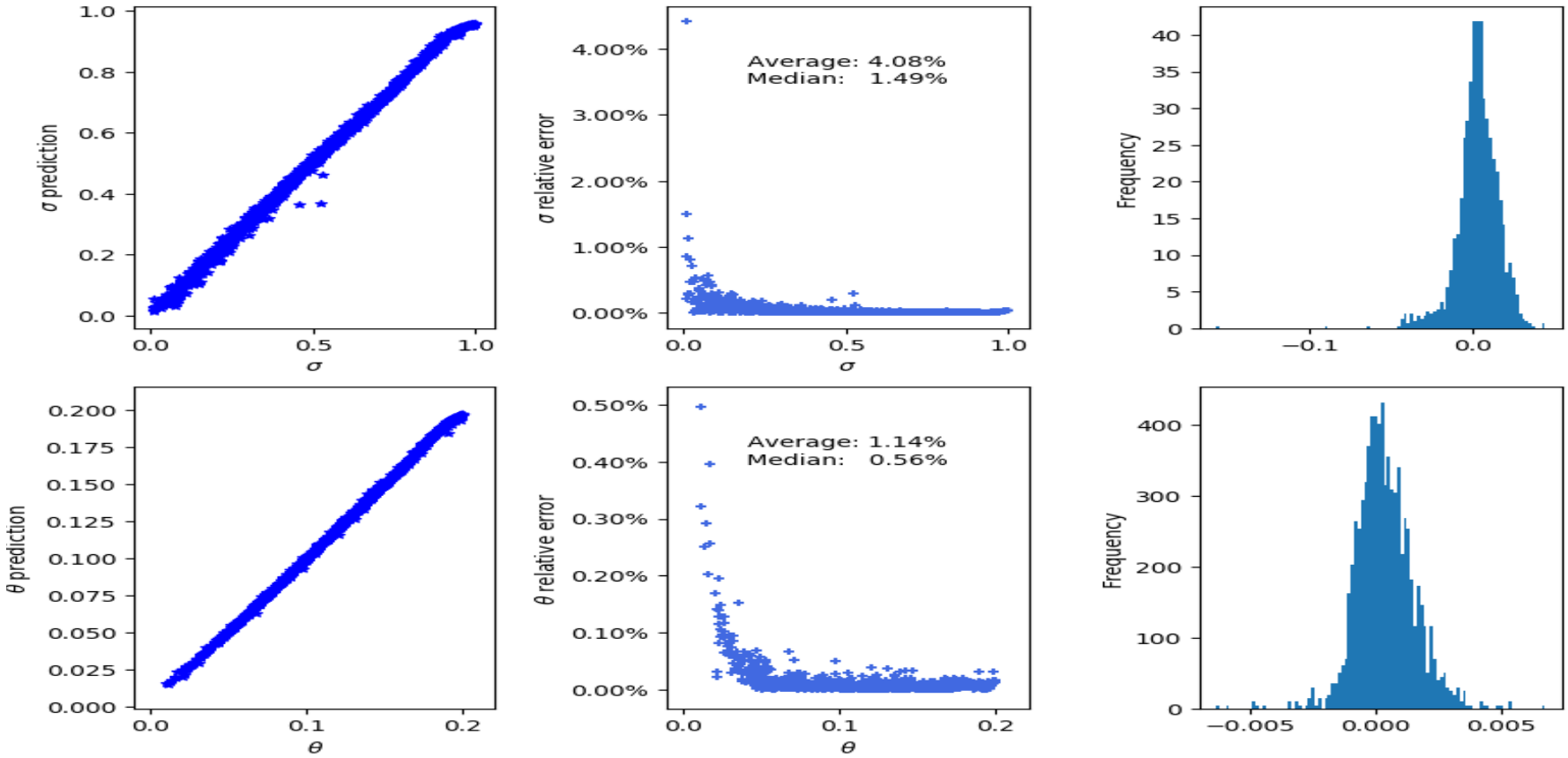
- We randomly generate the parameters ψ for different moneyness and time-to-maturities, and we calculate the implied-volatility surface.
  - $\longrightarrow$ Here, we use a data set of 10000 elements to train and of 1500 to test the networks.
  - $\longrightarrow$ Additional details, along with results using a larger data set, can be found in the online paper.
- Both the FCNN and the CNN shows good results both in the training and in testing data set.
- Yet, the relative errors obtained by the FCNN are significantly smaller than the CNN ones, expecially for ρ, σ and κ.
  - $\longrightarrow$ The CNN filters and max-pooling might cause information loss.
- As found in Roeder et al. 2020 some outliers occurs in the predictions, since there are parameter sets with very different values leading to similar volatility surfaces.

Calibration Results – II



FCNN prediction errors for σ and θ. From left to right: test set prediction, relative errors, absolute error histogram.

Calibration Results – III



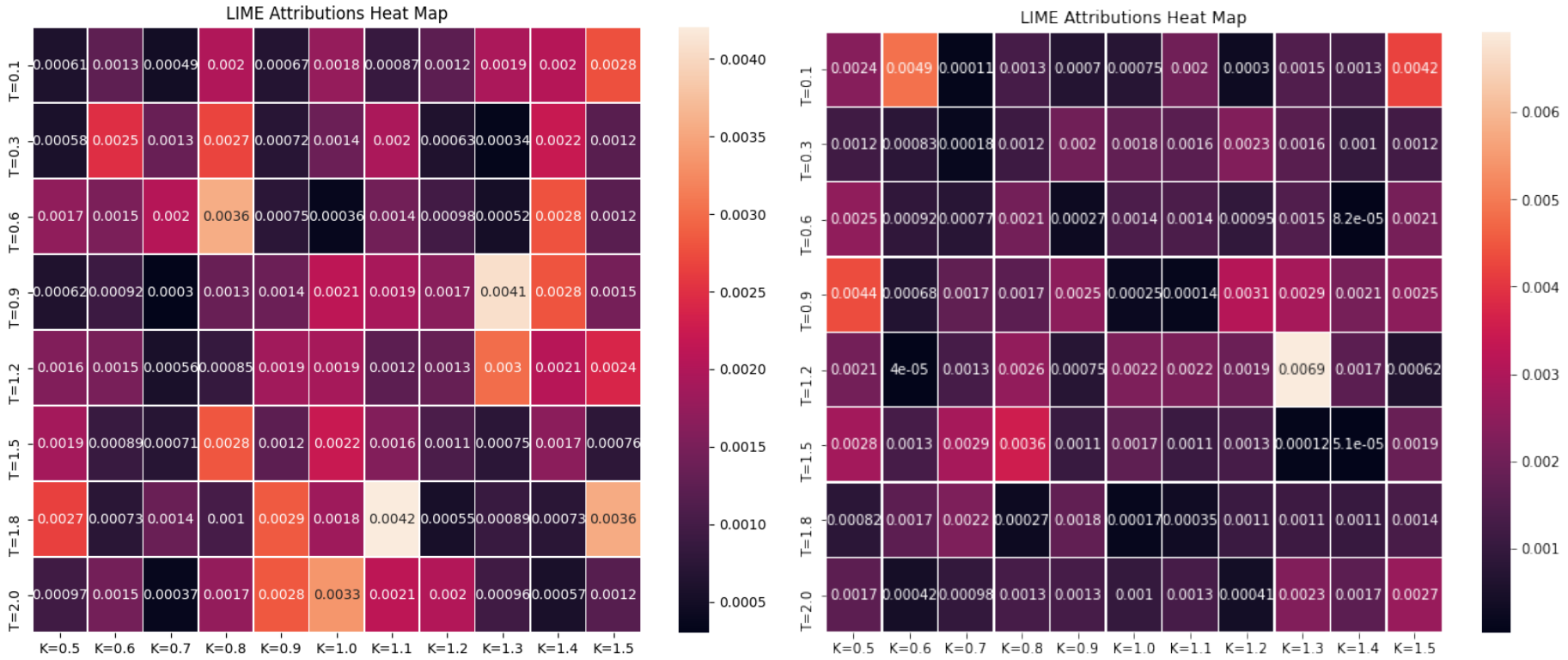
CNN prediction errors for σ and θ. From left to right: test set prediction, relative errors, absolute error histogram.

Talk Outline

LIME – I

- LIME (Local Interpretable Model-agnostic Explanations) is proposed by Ribeiro et al. (2016), and it is primarily designed to explain classifiers and NNs performing image recognition.
- LIME trains a local explanation model around individual predictions.
- In our case we treat the NN as a regressor, and we perform a local approximation around an individual prediction using a linear model.
  $\longrightarrow$ The Huber Regressor is chosen for its robustness against outliers.
- We look at the overall impact on the model output by averaging the absolute importance values of each feature.
- The most influential volatilities seem to be randomly located.
  $\longrightarrow$ This suggests that the mapping function for the Heston model is highly non-linear and it may contain multiple local minima.
  $\longrightarrow$ LIME seems not ideal for this application.

LIME – II



LIME heat map for the FCNN (left) and CNN (right) architecture. Light (dark) colors indicate high (low) attribution values.

SHAP – I

- SHAP (SHapley Additive exPlanations) is proposed by Lundberg et al. (2017), and it is based on optimal Shapley values.
  $\longrightarrow$ They originate from cooperative game theory, see Shapley (1953),
- The Shapley values are the marginal contributions of each player.

$$\phi(x) := \frac{1}{n} \sum_{z \subseteq \mathcal{X} \setminus x} \binom{n-1}{|z|}^{-1} \left(V(z \cup x) - V(z) \right)$$

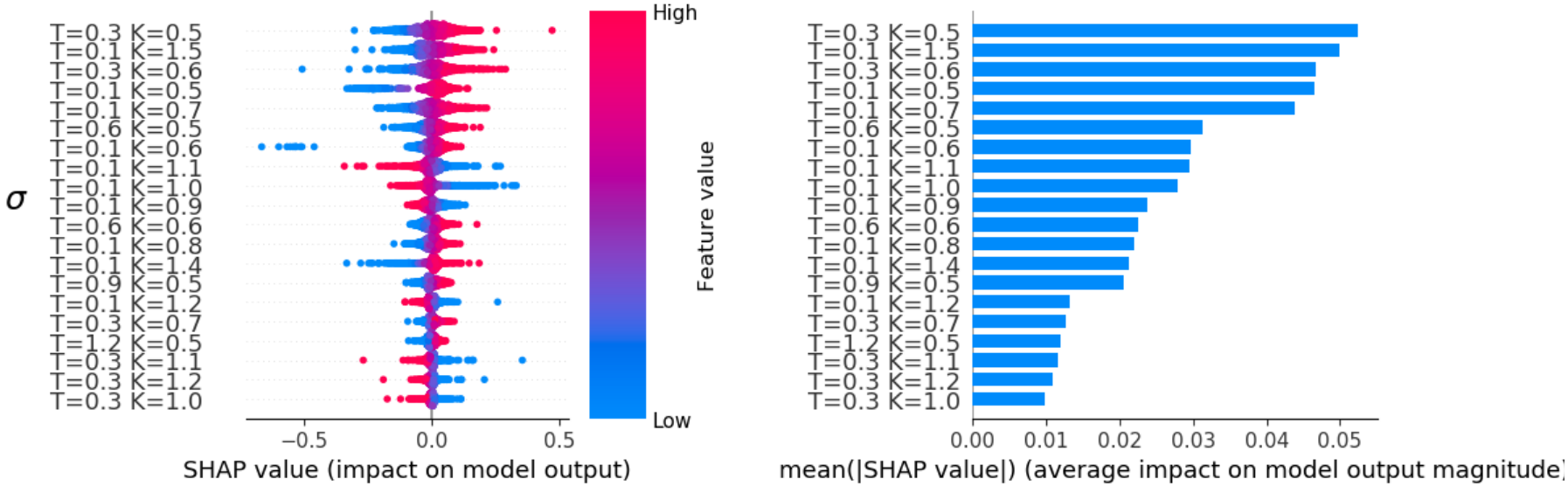
where $\mathcal{X}$ is the set of all the n players, and V is the coalition gain.

- In our case the game is predicting the Heston parameters ψ, and the plain-vanilla quotes are the players. Thus, we can define

$$V(z) := \mathbb{E}[\, \psi \mid z \,] - \mathbb{E}[\, \psi \,]$$

with the expectation taken over the volatilities of the test data set.

SHAP – II

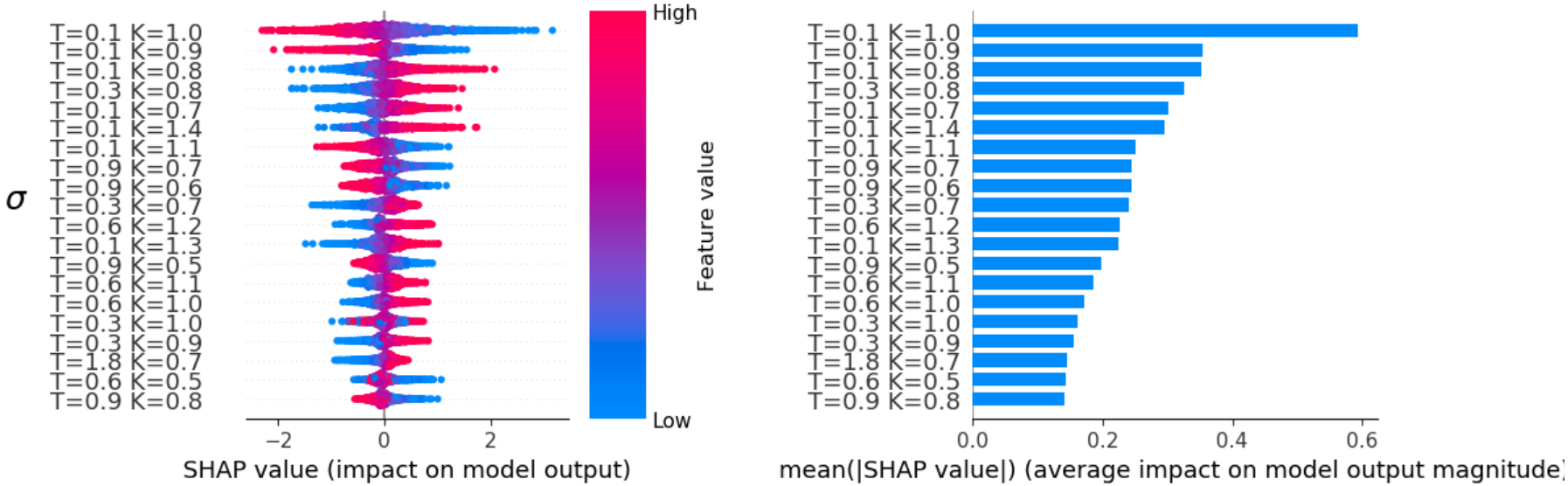


FCNN Shapley values ϕ along with feature importances $\mathbb{E}[\,|\phi|\,]$ for the volatility of volatility σ.

Left panel: on the x-axis the Shapley values, each row is a market quote ordered according to feature importance, the color represents the magnitude of σ, overlapping points are jittered in y-axis direction.

Right panel: the feature importance each market quote.

SHAP – III

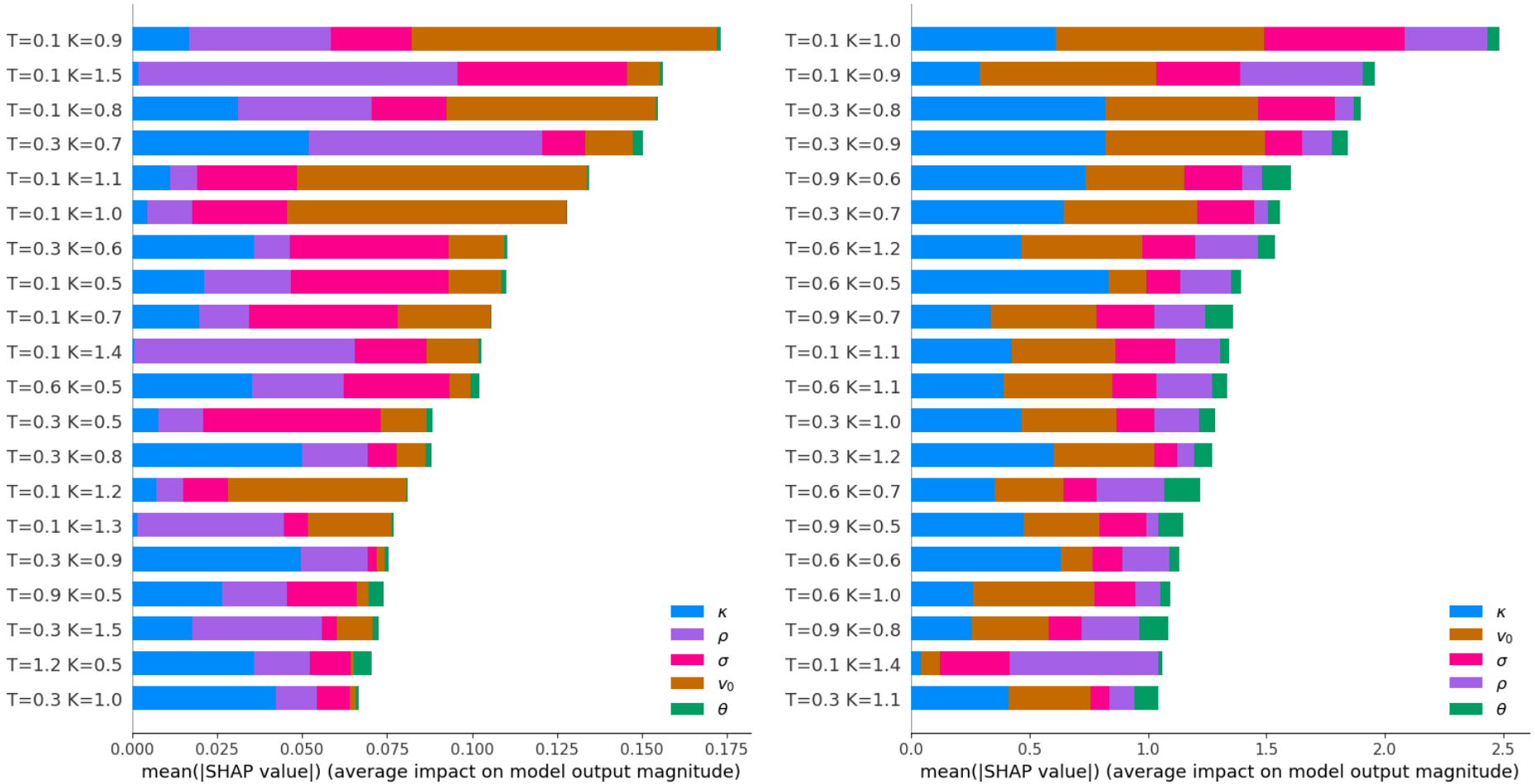


CNN Shapley values ϕ along with feature importances $\mathbb{E}[\,|\phi|\,]$ for the volatility of volatility σ.

Left panel: on the x-axis the Shapley values, each row is a market quote ordered according to feature importance, the color represents the magnitude of σ, overlapping points are jittered in y-axis direction.

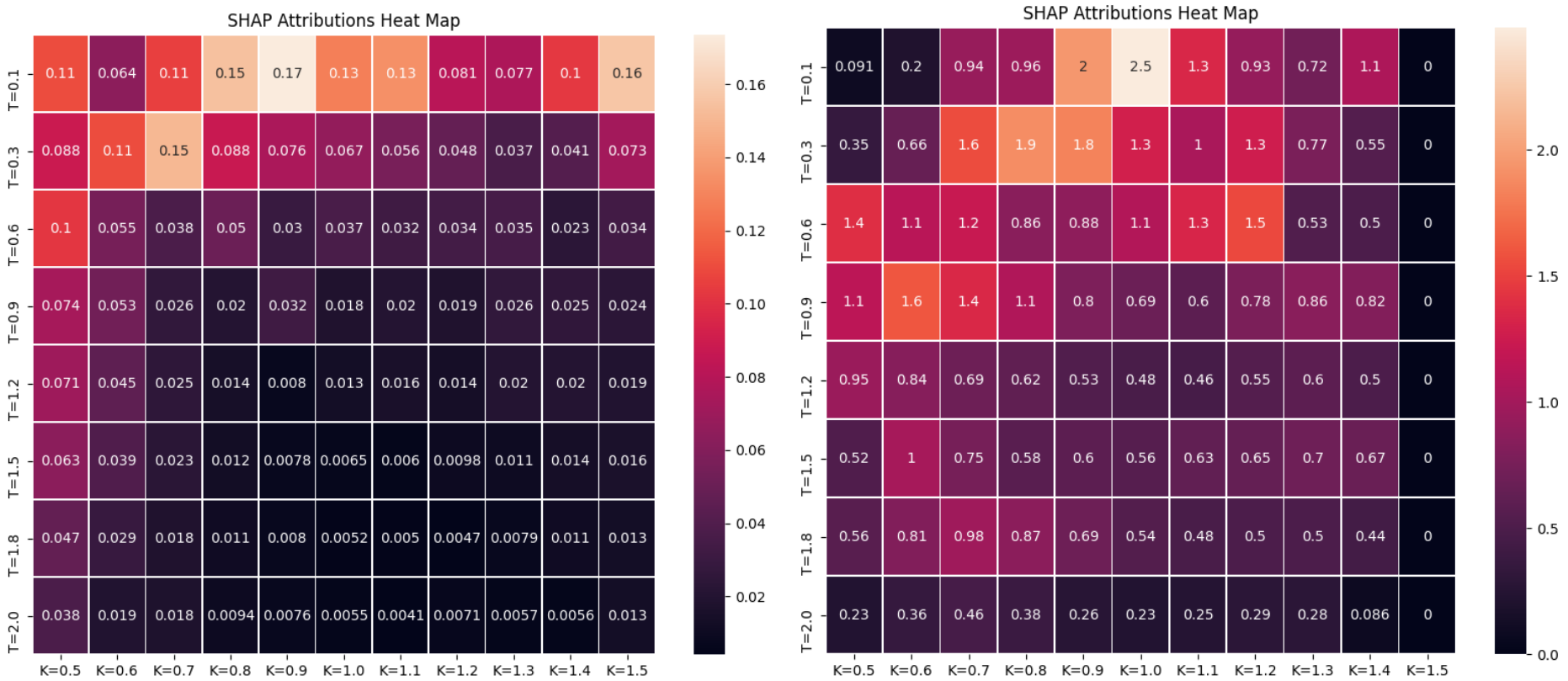
Right panel: the feature importance each market quote.

SHAP – IV



Overall feature importance $\mathbb{E}[\,|\phi|\,]$ for all the Heston parameters. Left panel FCNN, right panel CNN.

SHAP – V



SHAP heat map for the FCNN architecture. Light (dark) colors indicate high (low) attribution values.

Conclusion and Further Developments

- In this paper we have explored, using deep learning, the map from plain-vanilla options to model parameters in the Heston model.
- We conclude that the FCNN architecture outperforms the CNN one, since they require less trainable parameters and obtain better results.
- Then, we apply interpretability methods, and we find that global methods seem to be most reliable, and they substantially align with the common intuition behind the Heston model.
- They may also help in choosing the most convenient type of NN, favouring FCNNs over CNNs in our case.
- In a future work we wish to consider different pricing models and to analyse the impact of calibration uncertainties occurring on real market data.

Selected References – I

- Bayer, C. and B. Stemper (2018). *Deep calibration of rough stochastic volatility models*. Preprint arXiv:1810.03399.
- Bloch, D. (2019). *Option Pricing With Machine Learning*. Preprint SSRN:3486224.
- Brigo, D., X. Huang, A. Pallavicini, and H. Sáez de Ocáriz Borde (2021). *Interpretability in deep learning for finance: a case study for the Heston model*. Preprint arXiv:2104.09476.
- Chakraborty, S., R. Tomsett, R. Raghavendra, D. Harborne, M. Alzantot, F. Cerutti, M. Srivastava, A. Preece, S. Julier, R. M. Rao, T. D. Kelley, D. Braines, M. Sensoy, C. J. Willis, and P. Gurram (2017). "Interpretability of deep learning models: a survey of results". In: *2017 IEEE SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*. IEEE, pp. 1–6. DOI: 10.1109/UIC-ATC.2017.8397411.
- Demajo, L. M., V. Vella, and A. Dingli (2020). *Explainable AI for Interpretable Credit Scoring*. Preprint arXiv:2012.03749.
- Hernandez, A. (2017). "Model Calibration with Neural Networks". In: *Risk Magazine*.
- Horvath, B., A. Muguruza, and M. Tomas (2021). "Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility modelsOpen Data". In: *Quantitative Finance* 21.1, pp. 11–27. DOI: 10.1080/14697688.2020.1817974.
- Kraus, M., S. Feuerriegel, and A. Oztekin (2020). "Deep learning in business analytics and operations research: Models, applications and managerial implications". In: *European Journal of Operational Research* 281.3, pp. 628–641. DOI: 10.1016/j.ejor.2019.09.018.
- Lundberg, S. and S. Lee (2017). "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30*.

Selected References – II

- Moehle, N., S. Boyd, and A. Ang (2021). *Portfolio Performance Attribution via Shapley Value*. Preprint arXiv:2102.05799.
- Molnar, C. (2020). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. lulu.com.
- Ribeiro, M. T., S. Singh, and C. Guestrin (2016). ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144.
- Roeder, D. and G. Dimitroff (2020). *Volatility model calibration with neural networks a comparison between direct and indirect methods*. Preprint arXiv:2007.03494.
- Shapley, L. (1953). "A value for n-person games". In: *Contributions to the Theory of Games* 2.28, pp. 307–317.
- Wang, J., Y. Zhang, K. Tang, J. Wu, and Z. Xiong (2019). "Alphastock: A buying-winners-and-selling-losers investment strategy using interpretable deep reinforcement attention networks". In: *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1900–1908.

The opinions expressed in this work are solely those of the authors and do not represent in any way those of their current and past employers.