Received: 20 August 2021 | Accepted: 29 June 2022

DOI: 10.1111/mafi.12363

ORIGINAL ARTICLE

MATHEMATICAL FINANCE An International Journal of Mathematics, Statistics and Financial Economics

WILEY

Neural network approximation for superhedging prices

Francesca Biagini | **Lukas Gonon** | **Thomas Reitsam**

Workgroup Financial and Insurance Mathematics, Department of Mathematics, Ludwig-Maximilians Universität, Munich, Germany

Correspondence
Francesca Biagini, Workgroup Financial and Insurance Mathematics, Department of Mathematics, Ludwig-Maximilians Universität, Theresienstraße 39, 80333 Munich, Germany.
Email: biagini@math.lmu.de

Funding information
Verein zur Versicherungswissenschaft München e.V.

Abstract

This article examines neural network-based approximations for the superhedging price process of a contingent claim in a discrete time market model. First we prove that the α-quantile hedging price converges to the superhedging price at time 0 for α tending to 1, and show that the α-quantile hedging price can be approximated by a neural network-based price. This provides a neural network-based approximation for the superhedging price at time 0 and also the superhedging strategy up to maturity. To obtain the superhedging price process for $t > 0$, by using the Doob decomposition, it is sufficient to determine the process of consumption. We show that it can be approximated by the essential supremum over a set of neural networks. Finally, we present numerical results.

KEYWORDS
deep learning, quantile hedging, superhedging

1 | INTRODUCTION

In this paper, we study neural network approximations for the superhedging price process for a contingent claim in discrete time.

Superhedging was first introduced in El Karoui and Quenez (1995) and then thoroughly studied in various settings and market models. It is impossible to cover the complete literature here, but we name just a few references. For instance, in continuous time, for general càdlàg processes we mention Kramkov (1996), for robust superhedging Nutz (2015), Touzi (2014), for pathwise superhedging on prediction sets Bartl et al. (2020), Bartl et al. (2019), or for superhedging under proportional transaction costs (Campi & Schachermayer, 2006), Cvitanić and Karatzas (1996), Kabanov and Last (2002), Schachermayer (2014), Soner et al. (1995). Also in discrete time, there are various studies in the literature, like the standard case (Föllmer & Schied, 2016), robust superhedging Carassus et al. (2019), Obłój and Wiesel (2021), superhedging under volatility uncertainty (Nutz & Soner, 2012), or model-free superhedging Burzoni et al. (2017). The superhedging price provides an opportunity to secure a claim, but it may be too high or reduce the probability to profit from the option. In order to solve this problem, quantile hedging was introduced in Föllmer and Leukert (1999), where the authors propose to either fix the initial capital and maximize the probability of superhedging with this capital or fix a probability of superhedging and minimize the required capital. In this way, a trader can determine the desired trade-off between costs and risk.

In certain situations, it is possible to calculate explicitly or recursively the superhedging or quantile hedging price, see for example, Carassus et al. (2007), but in general incomplete markets, it may be complicated to determine superhedging prices or quantile hedging prices. In this article, we investigate neural network-based approximations for quantile- and superhedging prices. Neural network-based methods have been recently introduced in financial mathematics, for instance for hedging derivatives, see Buehler et al. (2019), determining stopping times, see Becker et al. (2019), or calibration of stochastic volatility models, see Cuchiero et al. (2020), and many more. For an overview of applications of machine learning to hedging and option pricing, we refer to Ruf and Wang (2020) and the references therein.

This paper contributes to the literature on hedging in discrete time market models in several ways. First, we prove that the α-quantile hedging price converges to the superhedging price for α tending to 1. Further, we show that it is feasible to approximate the α-quantile hedging and thus also the superhedging price for $t = 0$ by neural networks. We extend our machine learning approach also to approximate the superhedging price process for $t > 0$. By the first step, we obtain an approximation for the superhedging strategy on the whole interval up to maturity. By using the uniform Doob decomposition, see Föllmer and Schied (2016), we then only need to approximate the process of consumption B to generate the superhedging price process. We prove that B can be obtained as the the essential supremum over a set of neural networks. Finally, we present and discuss numerical results for the proposed neural network methods.

The paper is organized as follows. In Section 2, we present the discrete time market model of Föllmer and Schied (2016) and recall essential definitions and results on superhedging. Section 3 contains the study of the superhedging price for $t = 0$. More specifically, in Section 3.1, we prove in Theorem 3.4 that the α-quantile hedging price converges to the superhedging price as α tends to 1. We also present a similar result in Corollary 3.9 in Section 3.1.2, where α-quantile hedging is given in terms of success ratios. In Section 3.2, we show in Theorem 3.14 that the superhedging price can be approximated by neural networks. This concludes the approximation for $t = 0$. Then, we consider the case for $t > 0$ in Section 4. In Section 4.1, we explain how the uniform Doob decomposition can be used to approximate the superhedging price process. In that account, we prove an explicit representation of the process of consumption, see Proposition 4.1. Proposition 4.3 and Theorem 4.4 show that the process of consumption and thus the superhedging price process can be approximated by neural networks. The numerical results are presented in Section 5. The section is divided in the case $t = 0$, see Section 5.1, and $t > 0$, see Section 5.2. We present details on the algorithm and the implementation. Appendix B contains a version of the universal approximation theorem, derived from Hornik (1991).

2 | PRELIMINARIES

In this section, we introduce the discrete time financial market model from Föllmer and Schied (2016) and recall some basic notions on superhedging.

Consider a finite time horizon $T > 0$. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space endowed with a filtration $\mathbb{F} := (\mathcal{F}_t)_{t=0,1,\ldots,T}$. Further, we suppose that $\mathcal{F} = \mathcal{F}_T$ and that Y_0 is constant $\mathbf{P}$-a.s. Then $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

In our market model on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$, the asset prices are modeled by a non-negative, adapted, stochastic process $\bar{S} = (S^0, S) = (S^0_t, S^1_t, \ldots, S^d_t)_{t=0,1,\ldots,T}$, with $d \geq 1$, $d \in \mathbb{N}$. Further, we assume that $S^0_t > 0$ $\mathbf{P}$-a.s. for all $t = 0, 1, \ldots, T$. The discounted price process $\bar{X} = (X^0, X) = (X^0_t, X^1_t, \ldots, X^d_t)_{t=0,1,\ldots,T}$ is given by

$$X^i_t := \frac{S^i_t}{S^0_t}, \quad t = 0, 1, \ldots, T, \; i = 0, \ldots, d.$$

We denote by $\mathcal{P}$ the set of all equivalent martingale measures for X and assume $\mathcal{P} \neq \emptyset$. By Theorem 5.16 of Föllmer and Schied (2016), this is equivalent to the market model being arbitrage-free.

Definition 2.1. A trading strategy is a predictable $\mathbb{R}^{d+1}$-valued stochastic process $\bar{\xi} = (\xi^0, \xi) = (\xi^0_t, \xi^1_t, \ldots, \xi^d_t)_{t=1,\ldots,T}$ with (discounted) value process $V = (V_t)_{t=0,\ldots,T}$ given by

$$V_0 := \bar{\xi}_1 \cdot \bar{X}_0 \quad \text{and} \quad V_t := \bar{\xi}_t \cdot \bar{X}_t \quad \text{for } t = 1, \ldots, T.$$

A trading strategy $\bar{\xi}$ is called self-financing if

$$\bar{\xi}_t \cdot \bar{S}_t = \bar{\xi}_{t+1} \cdot \bar{S}_t \quad \text{for } t = 1, \ldots, T-1.$$

A self-financing trading strategy is called an *admissible strategy* if its value process satisfies $V_T \geq 0$.

By $\mathcal{A}$, we denote the set of all admissible strategies $\bar{\xi}$ and by $\mathcal{V}$ the associated value processes, that is,

$$\mathcal{V} := \left\{ V = (V_t)_{t=0,1,\ldots,T} \,:\, V_0 = \bar{\xi}_1 \cdot \bar{X}_0, V_t = \bar{\xi}_t \cdot \bar{X}_t \text{ for } t = 1, \ldots, T, \text{ and } \bar{\xi} \in \mathcal{A} \right\}$$

A discounted European contingent claim is represented by a non-negative, $\mathcal{F}_T$-measurable random variable H such that $\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] < \infty$.

Definition 2.2. Let H be a European contingent claim. A self-financing trading strategy $\bar{\xi}$ whose value process V satisfies

$$V_T \geq H \quad \mathbf{P}\text{-a.s.}$$

is called a *superhedging strategy* for H. In particular, any superhedging strategy is admissible since $H \geq 0$ by definition.

The upper Snell envelope for a discounted European claim H is defined by

$$U_t^{\uparrow}(H) = U_t^{\uparrow} := \operatorname*{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_t], \quad \text{for } t = 0, 1, \dots, T.$$

Set

$$\mathcal{U}_t := \left\{ \tilde{U}_t \in L^0(\Omega, \mathcal{F}_t, \mathbf{P}) : \exists \tilde{\xi} \text{ pred. s.t. } \tilde{U}_t + \sum_{k=t+1}^{T} \tilde{\xi}_k \cdot (X_k - X_{k-1}) \geq H \quad \mathbf{P}\text{-a.s.} \right\}. \tag{1}$$

In the sequel, ess inf $\mathcal{U}_t$ denotes the essential infimum of the set of random variables $\mathcal{U}_t$, which is defined by Theorem A.37 and Definition A.38 of Föllmer and Schied (2016), see also Appendix C.

Corollary 2.3 Corollary 7.3, Theorem 7.5, Corollary 7.15, Föllmer and Schied (2016). *The process*

$$\left(\operatorname*{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_t] \right)_{t=0,1,\dots,T},$$

is the smallest $\mathcal{P}$-supermartingale whose terminal value dominates H. Furthermore, there exists an adapted increasing process $B = (B_t)_{t=0,\dots,T}$ with $B_0 = 0$ and a d-dimensional predictable process $\xi = (\xi_t)_{t=1,\dots,T}$ such that

$$\operatorname*{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_t] = \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{t} \xi_k \cdot (X_k - X_{k-1}) - B_t \quad \mathbf{P}\text{-a.s. for all } t = 0, \dots, T. \tag{2}$$

Moreover, $\operatorname{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_t] = \operatorname{ess\,inf} \mathcal{U}_t$ *and*

$$\operatorname*{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_t] + \sum_{k=t+1}^{T} \xi_k \cdot (X_k - X_{k-1}) \geq H, \quad \text{for all } t = 0, \dots, T. \tag{3}$$

The process B in Equation (2) is sometimes called process of consumption, see Kramkov (1996). Equations (2) and (3) yield

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \geq B_t \geq B_{t-1} \geq 0 \quad \text{for all } t = 1, \dots, T. \tag{4}$$

Corollary 7.18 of Föllmer and Schied (2016) and Equation (3) guarantee that $U_t^{\uparrow}$ is the minimal amount needed at time t to start a superhedging strategy and thus there exists a predictable process ξ such that

$$\operatorname*{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_t] + \sum_{k=t+1}^{T} \xi_k \cdot (X_k - X_{k-1}) \geq H.$$

Further, $U_0^{\uparrow}$ is called the *superhedging price* at time $t = 0$ of H and coincides with the upper bound of the set of arbitrage-free prices.

3 | SUPERHEDGING PRICE FOR $t = 0$

In this section, we approximate the superhedging price for $t = 0$ in two steps. In the first part, we introduce the theory of quantile hedging, see Föllmer and Leukert (1999). In Theorem 3.4, we prove that the quantile hedging price for $\alpha \in (0, 1)$ converges to the superhedging price as α tends to 1. Analogously, in Corollary 3.9, we prove that for α tending to 1 also the success ratios for $\alpha \in (0, 1)$ converge to the superhedging price.

In the second part, we prove in Theorem 3.14 that the superhedging price and the associated strategies can be approximated by neural networks.

3.1 | Quantile hedging

3.1.1 | Success sets

In incomplete markets, perfect replication of a contingent claim may not be possible. Superhedging offers an alternative hedging method but it presents two main disadvantages. On the one hand, investing by using the superhedging strategy may reduce the possibility to profit. On the other hand, the superhedging price may result to be too high.

Quantile hedging was proposed for the first time in Föllmer and Leukert (1999) to address these problems. Fix $\alpha \in (0, 1)$. Given probability of success $\alpha \in (0, 1)$, we consider the minimization problem

$$\begin{aligned}\inf \mathcal{U}_0^\alpha := \inf\{u \in \mathbb{R} : \exists \xi = (\xi_t)_{t=1,\dots,T} &\text{ predictable process with values in } \mathbb{R}^d \text{ such that} \\ (u, \xi) \text{ is admissible and } &\mathbf{P}\left(u + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) \geq H\right) \geq \alpha\}. \end{aligned} \tag{5}$$

Here $1 - \alpha$ is called the shortfall probability. Quantile hedging may be considered as a dynamic version of the value at risk concept.

For an admissible strategy (u, ξ) with associated value process V, we call

$$\{V_T \geq H\}$$

the *success set*.

Remark 3.1. Note that in Equation (5), we need to require that (u, ξ) is admissible since this is not automatically implied by the definition of quantile hedging as in the case of superhedging strategies in Definition 2.2.

Proposition 3.2 below provides an equivalent formulation of quantile hedging (5), see also Föllmer and Leukert (1999). The proof is given in Appendix A.

Proposition 3.2. *Fix* $\alpha \in (0, 1)$*. Then*

$$\inf \mathcal{U}_0^\alpha = \inf \left\{ \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mathbb{1}_A] : A \in \mathcal{F}_T, \mathbf{P}(A) \geq \alpha \right\}.$$

Corollary 7.15 of Föllmer and Schied (2016) guarantees that there exists a superhedging strategy with initial value $\inf \mathcal{U}_0$. In contrast, there might be no explicit solution to the quantile hedging approach (5). If a solution to the quantile hedging approach exists, then Proposition 3.2 states that it is given by the solution of the classical hedging formulation for the knockout option $H\mathbb{1}_A$ for some suitable $A \in \mathcal{F}_T$. However, such a set $A \in \mathcal{F}_T$ does not always exist. In particular, quantile hedging does not always admit an explicit solution in general. The Neyman–Pearson lemma suggests to consider so-called success ratios instead of success sets. We will briefly discuss success ratios below. For further information, we refer the interested reader to Föllmer and Leukert (1999).

We now show that the superhedging price $\inf \mathcal{U}_0$, can be approximated by the quantile hedging price $\inf \mathcal{U}_0^\alpha$ for α tending to 1.

Definition 3.3. For $\alpha \in (0,1)$, we define

$$\mathcal{F}^\alpha := \{A \in \mathcal{F}_T \,:\, \mathbf{P}(A) \geq \alpha\}.$$

Theorem 3.4. *The α-quantile hedging price converges to the superhedging price as α tends to 1, that is,*

$$\inf \mathcal{U}_0^\alpha \xrightarrow{\alpha \uparrow 1} \inf \mathcal{U}_0.$$

We assume $\mathcal{F}_t = \sigma(Y_0, \dots, Y_t)$ for $t = 0, \dots, T$ and for some $\mathbb{R}^m$-valued process $Y = (Y_t)_{t=0,\dots,T}$ for some $m \in \mathbb{N}$, and write $\mathcal{Y}_t = (Y_0, \dots, Y_t)$ for $t \geq 0$.

Proof. We first note that using Proposition 3.2, it suffices to prove

$$\inf_{A \in \mathcal{F}^\alpha} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] \xrightarrow{\alpha \uparrow 1} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H].$$

Let $(\alpha_n)_{n \in \mathbb{N}} \subset (0,1)$ be an increasing sequence such that α_n converges to 1 as n tends to infinity. Note that

$$\inf_{A \in \mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] \leq \inf_{A \in \mathcal{F}^{\alpha_{n+1}}} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] \leq \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H], \tag{6}$$

because $\mathcal{F}^{\alpha_{n+1}} \subset \mathcal{F}^{\alpha_n}$. Therefore, the limit of $(\inf_{A \in \mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A])_{n \in \mathbb{N}}$ exists because the sequence is monotone and bounded. Let $\varepsilon > 0$ be arbitrary. For each $n \in \mathbb{N}$ there exists $A_n \in \mathcal{F}^{\alpha_n}$ such that

$$\inf_{A \in \mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] \leq \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_{A_n}] < \inf_{A \in \mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] + \varepsilon. \tag{7}$$

Then, by Lemma 1.70 of Föllmer and Schied (2016), there exists a sequence $\psi_n \in \operatorname{conv}\{\mathbb{1}_{A_n}, \mathbb{1}_{A_{n+1}}, \dots\}$, $n \in \mathbb{N}$, which converges $\mathbf{P}$-a.s. to some $\psi \in L^\infty([\Omega, \mathcal{F}_T, \mathbf{P}; [0,1])$. Note that it is not clear if ψ is an indicator function of some $\mathcal{F}_T$-measurable set. We will show that

$\psi = 1$ **P**-a.s. For $n \in \mathbb{N}$, ψ_n is of the form

$$\psi_n = \sum_{k=n}^{\infty} \lambda_k^n \mathbb{1}_{A_k}, \tag{8}$$

for some $(\lambda_k^n)_{k=n}^{\infty} \geq 0$ such that $\sum_{k=n}^{\infty} \lambda_k^n = 1$, where for each $n \in \mathbb{N}$ only finitely many λ_k^n are nonzero. By dominated convergence and Equation (8), we obtain

$$\mathbb{E}_{\mathbf{P}}[\psi] = \lim_{n\to\infty} \mathbb{E}_{\mathbf{P}}[\psi_n] = \lim_{n\to\infty} \mathbb{E}_{\mathbf{P}}\left[\sum_{k=n}^{\infty} \lambda_k^n \mathbb{1}_{A_k}\right] = \lim_{n\to\infty}\left(\sum_{k=n}^{\infty} \lambda_k^n \mathbb{E}_{\mathbf{P}}\left[\mathbb{1}_{A_k}\right]\right). \tag{9}$$

Because $\sum_{k=n}^{\infty} \lambda_k^n = 1$ and by the definition of the limes inferior, Equation (9) yields

$$\begin{aligned}\mathbb{E}_{\mathbf{P}}[\psi] &\geq \lim_{n\to\infty}\left(\sum_{k=n}^{\infty} \lambda_k^n \inf_{l\geq n} \mathbb{E}_{\mathbf{P}}\left[\mathbb{1}_{A_l}\right]\right) = \lim_{n\to\infty}\left(\inf_{l\geq n} \mathbb{E}_{\mathbf{P}}\left[\mathbb{1}_{A_l}\right]\right) \\ &= \liminf_{n\to\infty} \mathbb{E}_{\mathbf{P}}[\mathbb{1}_{A_n}] = \liminf_{n\to\infty} \mathbf{P}(A_n) \geq \liminf_{n\to\infty} \alpha_n = 1. \end{aligned} \tag{10}$$

Since $0 \leq \psi \leq 1$, it follows that $\psi = 1$ **P**-a.s. By Equation (7) and with similar arguments as in Equations (9) and (10) using the supremum instead of the infimum, we obtain by dominated convergence for any $\bar{\mathbf{P}}^* \in \mathcal{P}$ that

$$\begin{aligned}\limsup_{n\to\infty}\left(\inf_{A\in\mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] + \varepsilon\right) &\geq \limsup_{n\to\infty} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}_{[}H\mathbb{1}_{A_n}] \\ &\geq \lim_{n\to\infty} \mathbb{E}_{\bar{\mathbf{P}}^*}[H\psi_n] = \mathbb{E}^*[H\psi] = \mathbb{E}_{\bar{\mathbf{P}}^*}[H]. \end{aligned} \tag{11}$$

Since the limit on the left hand side in Equation (11) exists by Equations (6) and (11) holds for all $\bar{\mathbf{P}}^* \in \mathcal{P}$, we get

$$\lim_{n\to\infty}\left(\inf_{A\in\mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] + \varepsilon\right) \geq \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H]. \tag{12}$$

Thus, we observe that Equations (6) and (12) yields

$$\lim_{n\to\infty}\left(\inf_{A\in\mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A]\right) \leq \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] \leq \lim_{n\to\infty}\left(\inf_{A\in\mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] + \varepsilon\right).$$

As $\varepsilon > 0$ was arbitrary this implies that

$$\lim_{n\to\infty}\left(\inf_{A\in\mathcal{F}^{\alpha_n}} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A]\right) = \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H].$$

□

3.1.2 | Success ratios

Let $\mathcal{R} := L^\infty(\Omega, \mathcal{F}_T, \mathbf{P}; [0,1])$ be the set of randomized tests. For $\alpha \in (0,1)$ we denote by $\mathcal{R}^\alpha$ the set

$$\mathcal{R}^\alpha := \{\varphi \in \mathcal{R} : \mathbb{E}_{\mathbf{P}}[\varphi] \geq \alpha\}.$$

We now consider the following minimization problem

$$\inf\left\{\sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\varphi] : \varphi \in \mathcal{R}^\alpha\right\}. \tag{13}$$

In a first step, we prove that this problem admits an explicit solution. In a second step, we show that the solution is given by the so-called success ratio, see Definition 3.6 below. In particular, Equation (13) can be formulated in terms of success ratios, see also Föllmer and Leukert (1999). In Propositions 3.5 and 3.8, we provide a proof for some result of Föllmer and Leukert (1999) for the sake of completeness.

Proposition 3.5. *There exists a randomized test $\tilde{\varphi} \in \mathcal{R}$ such that*

$$\mathbb{E}_{\mathbf{P}}[\tilde{\varphi}] = \alpha,$$

and

$$\inf_{\varphi\in\mathcal{R}^\alpha} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\varphi] = \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\tilde{\varphi}]. \tag{14}$$

Proof. Take a sequence $(\varphi_n)_{n\in\mathbb{N}} \subset \mathcal{R}^\alpha$ such that

$$\lim_{n\to\infty} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\varphi_n] = \inf_{\varphi\in\mathcal{R}^\alpha} \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H\varphi]. \tag{15}$$

By Lemma 1.70 of Föllmer and Schied (2016), there exists a sequence of convex combinations $\tilde{\varphi}_n \in \operatorname{conv}\{\varphi_n, \varphi_{n+1}, \dots\}$ converging $\mathbf{P}$-a.s. to a function $\tilde{\varphi} \in \mathcal{R}$ because $\varphi_n \in [0,1]$ for all $n \in \mathbb{N}$. Clearly $\tilde{\varphi}_n \in \mathcal{R}^\alpha$ for each $n \in \mathbb{N}$. Hence, dominated convergence yields that

$$\mathbb{E}_{\mathbf{P}}[\tilde{\varphi}] = \lim_{n\to\infty} \mathbb{E}_{\mathbf{P}}[\tilde{\varphi}_n] \geq \alpha, \tag{16}$$

and we get that $\tilde{\varphi} \in \mathcal{R}^\alpha$. In the following, we use similar arguments as in the proof of Theorem 3.4. In particular, $\tilde{\varphi}_n$ is of the form

$$\tilde{\varphi}_n = \sum_{k=n}^{\infty} \lambda_k^n \varphi_k, \tag{17}$$

for some $(\lambda_k)_{k=n}^{\infty}$ such that $\sum_{k=n}^{\infty}\lambda_k^n = 1$, where for each $n\in\mathbb{N}$ only finitely many λ_k^n are nonzero. By Equation (17), we obtain for any $\bar{\mathbf{P}}^*\in\mathcal{P}$ that

$$\begin{aligned}\limsup_{n\to\infty}\mathbb{E}_{\bar{\mathbf{P}}^*}[H\varphi_n] &= \lim_{n\to\infty}\left(\sup_{k\geq n}\mathbb{E}_{\bar{\mathbf{P}}^*}[H\varphi_k]\right)\geq\lim_{n\to\infty}\left(\sum_{k=n}^{\infty}\lambda_k^n\mathbb{E}_{\bar{\mathbf{P}}^*}[H\varphi_k]\right)\\ &= \lim_{n\to\infty}\mathbb{E}_{\bar{\mathbf{P}}^*}[H\tilde{\varphi}_n] = \mathbb{E}_{\bar{\mathbf{P}}^*}[H\tilde{\varphi}],\end{aligned}\tag{18}$$

where we used monotone convergence. Moreover, we obtain by Equations (15), (18) and dominated convergence that

$$\inf_{\varphi\in\mathcal{R}^\alpha}\sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\varphi] = \limsup_{n\to\infty}\sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\varphi_n]\geq\limsup_{n\to\infty}\mathbb{E}^*[H\varphi_n]\geq\lim_{n\to\infty}\mathbb{E}^*[H\tilde{\varphi}_n] = \mathbb{E}^*[H\tilde{\varphi}].\tag{19}$$

Since Equation (19) holds for all $\mathbf{P}^*\in\mathcal{P}$ we obtain

$$\inf_{\varphi\in\mathcal{R}^\alpha}\sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\varphi]\geq\sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\tilde{\varphi}].$$

Furthermore, $\tilde{\varphi}\in\mathcal{R}^\alpha$ by Equation (16) yields

$$\inf_{\varphi\in\mathcal{R}^\alpha}\sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\varphi] = \sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\tilde{\varphi}].$$

So $\tilde{\varphi}$ is the desired minimizer.

We now show that $\mathbb{E}_{\mathbf{P}}[\tilde{\varphi}]=\alpha$ holds. If $\mathbb{E}_{\mathbf{P}}[\tilde{\varphi}]>\alpha$, then we can find $\varepsilon>0$ such that $\varphi_\varepsilon := (1-\varepsilon)\tilde{\varphi}\in\mathcal{R}^\alpha$, and

$$\sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\varphi_\varepsilon] = (1-\varepsilon)\sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\tilde{\varphi}] < \sup_{\mathbf{P}^*\in\mathcal{P}}\mathbb{E}^*[H\tilde{\varphi}],\tag{20}$$

which contradicts the minimality property of $\tilde{\varphi}$. Thus,

$$\mathbb{E}_{\mathbf{P}}[\tilde{\varphi}] = \alpha.$$

□

Definition 3.6. For an admissible strategy with value process $V\in\mathcal{V}$, we define its success ratio by

$$\varphi_V := \mathbb{1}_{\{V_T\geq H\}} + \frac{V_T}{H}\mathbb{1}_{\{V_T<H\}}.\tag{21}$$

For $\alpha\in(0,1)$, we denote by $\mathcal{V}^\alpha$ the set

$$\mathcal{V}^\alpha := \{\varphi_V\in\mathcal{R} : V\in\mathcal{V},\ \mathbb{E}_{\mathbf{P}}[\varphi_V]\geq\alpha\}.$$

Remark 3.7. Note that for $V\in\mathcal{V}$, we have that $V_T\geq 0$ $\mathbf{P}$-a.s. In particular, $\mathbf{P}(\{H=0\}\cap\{V_T<H\})=0$ and hence Equation (21) is well-defined.

In the following, we formulate the optimization problem (5) in terms of success ratios and prove that it is equivalent to Equation (13), see Proposition 3.8 below.

Consider the minimization problem

$$\inf \left\{ \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[\varphi_V] \,:\, V \in \mathcal{V}^\alpha \right\}. \tag{22}$$

Proposition 3.8. *There exists an admissible strategy with value process $\tilde{V}$ such that*

$$\mathbb{E}_{\mathbf{P}}[\varphi_{\tilde{V}}] = \alpha,$$

and

$$\inf_{V \in \mathcal{V}^\alpha} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi_V] = \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi_{\tilde{V}}], \tag{23}$$

where φ_V denotes the success ratio associated to a portfolio $V \in \mathcal{V}$ as in Equation (21). Moreover, $\varphi_{\tilde{V}}$ coincides with the solution $\tilde{\varphi}$ from Proposition 3.5.

Proof. Note that

$$\{\varphi_V \in \mathcal{R} \,:\, V \in \mathcal{V}^\alpha\} \subseteq \mathcal{R}^\alpha,$$

and thus

$$\inf_{\varphi \in \mathcal{R}^\alpha} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi] \le \inf_{V \in \mathcal{V}^\alpha} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi_V]. \tag{24}$$

By Proposition 3.5, we know that the left-hand side of Equation (24) admits a solution $\tilde{\varphi} \in \mathcal{R}$. We prove that there exists $\tilde{V} \in \mathcal{V}^\alpha$ such that

$$\tilde{\varphi} = \varphi_{\tilde{V}} \quad \mathbf{P}\text{-a.s.}$$

Define the the modified claim

$$\tilde{H} := H\tilde{\varphi}.$$

By Theorem 7.13 of Föllmer and Schied (2016), there exists a minimal superhedging strategy $\tilde{\xi}$ with value process $\tilde{V}$ for $\tilde{H}$ such that

$$\tilde{V}_0 = \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[\tilde{H}].$$

First, $\tilde{\xi}$ can be assumed to be admissible by Remark 3.1 and hence $\tilde{V} \in \mathcal{V}$. Now, we show that $\tilde{V} \in \mathcal{V}^\alpha$. We have

$$\varphi_{\tilde{V}} = \mathbb{1}_{\{\tilde{V}_T \ge H\}} + \frac{\tilde{V}_T}{H}\mathbb{1}_{\{\tilde{V}_T < H\}} \ge \tilde{\varphi}\mathbb{1}_{\{\tilde{V}_T \ge H\}} + \frac{H\tilde{\varphi}}{H}\mathbb{1}_{\{\tilde{V}_T < H\}} = \tilde{\varphi}, \tag{25}$$

where we used that $\tilde{V}$ is the value process of the minimal superhedging strategy of $\tilde{H} = H\tilde{\varphi}$ and $0 \le \tilde{\varphi} \le 1$. Therefore, we get

$$\mathbb{E}_{\mathbf{P}}[\varphi_{\tilde{V}}] \ge \mathbb{E}_{\mathbf{P}}[\tilde{\varphi}] \ge \alpha,$$

so $\tilde{V} \in \mathcal{V}^{\alpha}$ and $\varphi_{\tilde{V}} \in \mathcal{R}^{\alpha}$. It is left to show that $\tilde{\varphi} = \varphi_{\tilde{V}}$ **P**-a.s. By Equation (25), we obtain $\varphi_{\tilde{V}} \ge \tilde{\varphi}$. For the reverse direction, we first show that $\varphi_{\tilde{V}}$ is also a minimizer of the problem (14), that is,

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi_{\tilde{V}}] \le \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\tilde{\varphi}].$$

Indeed, since $\tilde{V}$ is the value process of an admissible strategy, V is a $\mathbf{P}^*$-martingale for all $\mathbf{P}^* \in \mathcal{P}$ by Theorem 5.14 of Föllmer and Schied (2016) and thus we get that

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi_{\tilde{V}}] = \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*\left[H\left(\mathbb{1}_{\{\tilde{V}_T \ge H\}} + \frac{\tilde{V}_T}{H}\mathbb{1}_{\{\tilde{V}_T < H\}}\right)\right] \le \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[\tilde{V}_T] = \tilde{V}_0 = \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\tilde{\varphi}], \tag{26}$$

where we used in the last equality that $\tilde{V}_0$ is the superhedging price of $\tilde{H} = H\tilde{\varphi}$. In particular, $\varphi_{\tilde{V}} \in \mathcal{R}^{\alpha}$ is a minimizer. By the same arguments as in Equation (20), it follows that

$$\mathbb{E}_{\mathbf{P}}[\varphi_{\tilde{V}}] = \alpha. \tag{27}$$

Thus, we get by Equations (20) and (27) that

$$\mathbb{E}_{\mathbf{P}}[\varphi_{\tilde{V}}] = \alpha = \mathbb{E}_{\mathbf{P}}[\tilde{\varphi}],$$

that is, $\mathbb{E}[\varphi_{\tilde{V}} - \tilde{\varphi}] = 0$. Together with Equation (25), this implies $\varphi_{\tilde{V}} = \tilde{\varphi}$ **P**-a.s. We have proved that $\tilde{V} \in \mathcal{V}^{\alpha}$ and

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi_{\tilde{V}}] = \inf_{\varphi \in \mathcal{R}^{\alpha}} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi] \le \inf_{V \in \mathcal{V}^{\alpha}} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi_V].$$

In particular, $\varphi_{\tilde{V}}$ solves Equation (23) and the quantile hedging formulations of Equations (13) and (22) are equivalent. □

Corollary 3.9. *The following convergence holds:*

$$\inf_{V \in \mathcal{V}^{\alpha}} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\varphi_V] \xrightarrow{\alpha \uparrow 1} \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H],$$

where φ_V denotes the success ratio associated to a portfolio $V \in \mathcal{V}$ as in Equation (21).

Proof. The proof is similar to the one of Theorem 3.4 and is omitted. □

3.2 | Neural network approximation for $t = 0$

We now study how to approximate the superhedging price at $t = 0$ by using neural networks.

We recall the following definition, see for example, Buehler et al. (2019). Common choices for σ below are $\sigma(x) = \frac{1}{1-e^{-x}}$ and $\sigma(x) = \tanh(x)$.

In order to use the neural network approximation, we assume from now on that $\mathcal{F}_t = \sigma(Y_0, \dots, Y_t)$ for $t = 0, \dots, T$ and for some $\mathbb{R}^m$-valued process $Y = (Y_t)_{t=0,\dots,T}$ for some $m \in \mathbb{N}$, and write $\mathcal{Y}_t = (Y_0, \dots, Y_t)$ for $t \geq 0$.

Definition 3.10. Consider $L, N_0, N_1, \dots, N_L \in \mathbb{N}$ with $L \geq 2$, $\sigma : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ measurable and for any $\ell = 1, \dots, L$, let $W_\ell : \mathbb{R}^{N_{\ell-1}} \to \mathbb{R}^{N_\ell}$ be an affine function. A function $F : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ defined as

$$F(x) = W_L \circ F_{L-1} \circ \cdots \circ F_1 \text{ with } F_\ell = \sigma \circ W_\ell \text{ for } \ell = 1, \dots, L-1,$$

is called a *(feed forward) neural network*. Here the *activation function* σ is applied componentwise. L denotes the number of layers, $N_1, \dots, N_{L-1}$ denote the dimensions of the hidden layers and N_0, N_L the dimension of the input and output layers, respectively. For any $\ell = 1, \dots, L$, the affine function W_ℓ is given as $W_\ell(x) = A^\ell x + b^\ell$ for some $A^\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and $b^\ell \in \mathbb{R}^{N_\ell}$. For any $i = 1, \dots N_\ell$, $j = 1, \dots, N_{\ell-1}$, the number A^ℓ_{ij} is interpreted as the weight of the edge connecting the node i of layer $\ell - 1$ to node j of layer ℓ. The number of nonzero weights of a network is $\sum_{\ell=1}^L \|A^\ell\|_0 + \|b^\ell\|_0$, that is, the sum of the number of nonzero entries of the matrices A^ℓ, $\ell = 1, \dots, L$, and vectors b^ℓ, $\ell = 1, \dots, L$.

For $k = 1, \dots, T+1$, we denote the set of all possible neural network parameters corresponding to neural networks mapping $\mathbb{R}^{mk} \to \mathbb{R}^d$ by

$$\Theta_k = \cup_{L\geq 2} \cup_{(N_0,\dots,N_L)\in\{mk\}\times\mathbb{N}^{L-1}\times\{d\}} \left(\times_{\ell=1}^L \mathbb{R}^{N_\ell \times N_{\ell-1}} \times \mathbb{R}^{N_\ell} \right).$$

With F^{θ_k}, we denote the neural network with parameters specified by $\theta_k \in \Theta_k$, see Definition 3.10. Recall that $\mathcal{F}_t = \sigma(Y_0, \dots, Y_t) = \sigma(\mathcal{Y}_t)$ for $t = 0, \dots, T$, and for some $\mathbb{R}^m$-valued stochastic process Y. Then, any $\mathcal{F}_t$-measurable random variable Z can be written as $Z = f_t(\mathcal{Y}_t)$ for some measurable function f_t. Using Theorem B.1, f_t can be approximated by a deep neural network in a suitable metric.

The approximate superhedging price is then

$$\inf \mathcal{U}_0^{\Theta} = \inf \left\{ u \in \mathbb{R} \ : \ \exists\, \theta_{k,\xi} \in \Theta_k, k = 1, \dots, T, \text{ s.t. } u + \sum_{k=1}^T F^{\theta_{k,\xi}}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1}) \geq H \ \mathbf{P}\text{-a.s.} \right\}. \quad (28)$$

For $\alpha \in (0, 1)$, the approximate α-quantile hedging price is then

$$\inf \mathcal{U}_0^{\Theta,\alpha} = \inf \left\{ u \in \mathbb{R} \ : \ \exists\, \theta_{k,\xi} \in \Theta_k, k = 1, \dots, T \text{ s.t. } \mathbf{P}\left(u + \sum_{k=1}^T F^{\theta_{k,\xi}}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1}) \geq H \right) \geq \alpha \right\}. \quad (29)$$

For $C > 0$, we also define the truncated approximate superhedging price $\inf \mathcal{U}_0^{\Theta,C}$ and the truncated approximate α-quantile hedging price $\inf \mathcal{U}_0^{\Theta,C,\alpha}$ with

$$\mathcal{U}_0^{\Theta,C} := \Bigg\{ u \in \mathbb{R} : \exists \theta_{k,\xi} \in \Theta_k, k = 1, \dots, T \text{ s.t. } u + \sum_{k=1}^{T} \Big(\Big(F^{\theta_{k,\xi}} \wedge C \Big) \vee (-C) \Big) (\mathcal{Y}_{t-1}) \cdot (X_k - X_{k-1}) \geq H \ \mathbf{P}\text{-a.s.} \Bigg\} \tag{30}$$

and

$$\mathcal{U}_0^{\Theta,C,\alpha} := \Bigg\{ u \in \mathbb{R} : \exists \theta_{k,\xi} \in \Theta_k, k = 1, \dots, T \text{ s.t. } \mathbf{P} \Bigg(u + \sum_{k=1}^{T} \Big(\Big(F^{\theta_{k,\xi}} \wedge C \Big) \vee (-C) \Big) (\mathcal{Y}_{t-1}) \cdot (X_k - X_{k-1}) \geq H \Bigg) \geq \alpha \Bigg\}, \tag{31}$$

where the maximum and minimum are taken component-wise.

In the definitions of the (truncated) approximate superhedging price, in Equations (28) and (30), and of the (truncated) approximate α-quantile hedging price, in Equations (29) and (31), it cannot be guaranteed that the trading strategies given by neural networks are admissible. For this reason, we need to impose Assumption 3.11 below.

Assumption 3.11. Suppose that

$$\inf \mathcal{U}_0 = \inf \mathcal{U}_0^{\text{bdd}} := \inf \Bigg\{ u \in \mathbb{R} : \exists \xi \text{ pred. s.t. } \xi_k \in L^\infty \ \forall k \in \{1, \dots, T\},\ u + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) \geq H \ \mathbf{P}\text{-a.s.} \Bigg\}.$$

Remark 3.12. Under Assumption 3.11, the superhedging price does not change if we restrict to bounded strategies. This hypothesis is needed for the approximation result in Theorem 3.14, as shown in Example 3.15. However, Assumption 3.11 is not very restrictive and holds in a wide range of settings, for example, in all models with a finite probability space, since in this case, all random variables are bounded. In particular, Assumption 3.11 is satisfied in any scenario-based model as considered in the section on numerical results. Furthermore, Assumption 3.11 holds in the case of $d = 1$ and bounded, independent price increments, that is, if for each t, the price increment $X_t - X_{t-1}$ is bounded and independent of $\mathcal{F}_{t-1}$ under $\mathbf{P}$, since in this case, any admissible strategy is necessarily bounded. This can be seen by contraposition. Let $t \in \{1, \dots, T\}$ be the first time at which ξ is unbounded, that is, $\mathbf{P}(\xi_t > n) > 0$ for all $n \in \mathbb{N}$ (the case $\mathbf{P}(\xi_t < -n) > 0$ for all $n \in \mathbb{N}$ can be treated analogously) and $\|\xi_k\|_\infty < \infty$ for $k < t$. Note that $\mathcal{P} \neq \emptyset$, and for $\mathbf{P}^* \in \mathcal{P}$, the property $\mathbb{E}^*[X_t - X_{t-1}] = 0$ and $\mathbf{P} \approx \mathbf{P}^*$ imply that $\mathbf{P}(X_t - X_{t-1} \leq -C) > 0$ for some $C > 0$. Therefore, for $n \geq C^{-1} \sum_{k=0}^{t-1} \|\xi_k\|_\infty \|X_k - X_{k-1}\|_\infty$, it follows that $V_{t-1} \leq Cn$ $\mathbf{P}$-a.s. and

$$\mathbf{P}(V_t < 0) \geq \mathbf{P}(V_{t-1} \leq Cn, \xi_t(X_t - X_{t-1}) < -Cn) \geq \mathbf{P}(\xi_t > n)\mathbf{P}(X_t - X_{t-1} < -C) > 0.$$

Hence, the strategy associated to ξ can not be admissible.

Finally, in a large class of models and for many commonly considered options, the optimal superhedging strategy is static, as shown in Carassus et al. (2007). Consequently, Assumption 3.11 also holds in these settings.

Theorem 3.14 shows that $\inf \mathcal{U}_0^{\Theta,C,\alpha}$ can be used as an approximation of the superhedging price $\inf \mathcal{U}_0$. We start with a preliminary result.

For $\alpha \in (0,1]$, $C > 0$ define $\mathcal{U}_0^{C,\alpha}$ by

$$\mathcal{U}_0^{C,\alpha} := \left\{ u \in \mathbb{R} : \exists \xi \text{ pred. s.t. } \sup_{1\le k\le T} \|\xi_k\|_\infty \le C,\ \mathbf{P}\left(u + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) \ge H \right) \ge \alpha \right\}.$$

Lemma 3.13. *Suppose Assumption 3.11 holds. Then for any $\varepsilon > 0$, there exists $C \in (0,\infty)$ such that $\lim_{\alpha\to 1} \inf \mathcal{U}_0^{C,\alpha}$ exists and*

$$\inf \mathcal{U}_0^{bdd} \le \lim_{\alpha\to 1} \inf \mathcal{U}_0^{C,\alpha} \le \inf \mathcal{U}_0^{bdd} + \varepsilon. \tag{32}$$

Proof. Let $\varepsilon > 0$ be fixed. Then, there exists a predictable strategy $\tilde{\xi}$ such that $\sup_{1\le k\le T} \|\tilde{\xi}_k\|_\infty < \infty$ and $\inf \mathcal{U}_0^{\mathrm{bdd}} + \frac{\varepsilon}{2} + \sum_{k=1}^T \tilde{\xi}_k \cdot (X_k - X_{k-1}) \ge H$, $\mathbf{P}$-a.s. Define $C = C(\varepsilon)$ by

$$C := \sup_{1\le k\le T} \|\tilde{\xi}_k\|_\infty + 1. \tag{33}$$

Let $(\alpha_n)_{n\in\mathbb{N}} \subset (0,1)$ be a sequence such that $\alpha_n \uparrow 1$ as n tends to infinity. Then, $\inf \mathcal{U}_0^{C,\alpha_n} \le \inf \mathcal{U}_0^{C,\alpha_{n+1}} \le \inf \mathcal{U}_0^{C,1} =: \inf \mathcal{U}_0^{C}$ since

$$\mathcal{U}_0^{C,\alpha_n} \supset \mathcal{U}_0^{C,\alpha_{n+1}},$$

and therefore, $u_n \le u_{n+1}$, where $u_n := \inf \mathcal{U}_0^{C,\alpha_n}$, for $n \in \mathbb{N}$. Thus, the limit $u^C = \lim_{n\to\infty} u_n$ is well-defined and $u^C \le \inf \mathcal{U}_0^C$. Furthermore, for $n \in \mathbb{N}$ and $\delta > 0$, there exists $\xi^{(n)}$ predictable such that $\sup_{1\le k\le T} \|\xi_k^{(n)}\|_\infty \le C$ and

$$\mathbf{P}\left(u_n + \delta + \sum_{k=1}^T \xi_k^{(n)} \cdot (X_k - X_{k-1}) \ge H \right) \ge \alpha_n. \tag{34}$$

For $n \in \mathbb{N}$, define $A_n \in \mathcal{F}_T$ by

$$A_n := \left\{ u_n + \delta + \sum_{k=1}^T \xi_k^{(n)} \cdot (X_k - X_{k-1}) \ge H \right\}.$$

Then $\mathbf{P}(A_n) \ge \alpha_n$ and hence $\mathbf{P}(A_n) \uparrow 1$ as n tends to infinity. Since $\sup_{1\le k\le T} \|\xi_k^{(n)}\|_\infty \le C$ for all $n \in \mathbb{N}$, we get by Theorem 5.14 of Föllmer and Schied (2016) for any $\bar{\mathbf{P}}^* \in \mathcal{P}$ that

$$u_n + \delta = \mathbb{E}_{\bar{\mathbf{P}}^*}\left[u_n + \delta + \sum_{k=1}^T \xi_k^{(n)} \cdot (X_k - X_{k-1}) \right] \tag{35}$$

$$\geq \mathbb{E}_{\bar{\mathbf{P}}^*}[H\mathbb{1}_{A_n}] + \mathbb{E}_{\bar{\mathbf{P}}^*}\left[\left(u_n + \delta + \sum_{k=1}^{T} \xi_k^{(n)} \cdot (X_k - X_{k-1})\right)\mathbb{1}_{A_n^c}\right]$$

$$\geq \mathbb{E}_{\bar{\mathbf{P}}^*}[H\mathbb{1}_{A_n}] + \mathbb{E}_{\bar{\mathbf{P}}^*}\left[\left(u_n + \delta - \sum_{k=1}^{T}\sum_{i=1}^{d} |\xi_k^{i,(n)}||X_k^i - X_{k-1}^i|\right)\mathbb{1}_{A_n^c}\right]$$

$$\geq \mathbb{E}_{\bar{\mathbf{P}}^*}[H\mathbb{1}_{A_n}] + \mathbb{E}_{\bar{\mathbf{P}}^*}\left[\left(u_n + \delta - C\sum_{k=1}^{T}\sum_{i=1}^{d} |X_k^i - X_{k-1}^i|\right)\mathbb{1}_{A_n^c}\right]. \tag{36}$$

Recall that $X = (X^1, \dots, X^d)$ is a d-dimensional $\mathbf{P}^*$-martingale, $u_n \leq u^C \leq \inf \mathcal{U}_0^{\mathrm{bdd}} + \frac{\varepsilon}{2}$, $n \in \mathbb{N}$ and thus for all $n \in \mathbb{N}$

$$\left|u_n + \delta - C\sum_{k=1}^{T}\sum_{i=1}^{d} |X_k^i - X_{k-1}^i|\right| \leq \left(|u^C + \delta| + |C|\sum_{k=1}^{T}\sum_{i=1}^{d} \|X_k - X_{k-1}\|\right) \in L^1(\Omega, \mathcal{F}_T, \mathbf{P}^*).$$

Furthermore, $\mathbb{1}_{A_n}$ converges to 1 in probability as n tends to infinity, since for any $\gamma \in (0,1)$, we have

$$\mathbf{P}\left(\left|\mathbb{1}_{A_n} - 1\right| > \gamma\right) = \mathbf{P}\left(\mathbb{1}_{A_n^c} > \gamma\right) = \mathbf{P}(A_n^c) \xrightarrow{n\to\infty} 0,$$

because of Equation (34). By dominated convergence, we obtain that

$$\lim_{n\to\infty} \mathbb{E}^*[H\mathbb{1}_{A_n}] = \mathbb{E}^*[H],$$

and

$$\lim_{n\to\infty} \mathbb{E}^*\left[\left(u_n + \delta - C\sum_{k=1}^{T}\sum_{i=1}^{d} |X_k^i - X_{k-1}^i|\right)\mathbb{1}_{A_n^c}\right] = 0.$$

Note that for dominated convergence, it is sufficient that $\mathbb{1}_{A_n}$ converges only in probability. Taking n to infinity in Equations (35) and (36) yields

$$\lim_{n\to\infty} u_n + \delta = u^C + \delta \geq \lim_{n\to\infty}\left(\mathbb{E}^*[H\mathbb{1}_{A_n}] + \mathbb{E}^*\left[\left(u_n + \delta - C\sum_{k=1}^{T}\sum_{i=1}^{d} |X_k^i - X_{k-1}^i|\right)\mathbb{1}_{A_n^c}\right]\right) = \mathbb{E}^*[H]. \tag{37}$$

As Equation (37) holds for all $\mathbf{P}^* \in \mathcal{P}$, we get by the superhedging duality that

$$\lim_{n\to\infty} \inf \mathcal{U}_0^{C,\alpha_n} + \delta = u^C + \delta \geq \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] = \inf \mathcal{U}_0 = \inf \mathcal{U}_0^{\mathrm{bdd}}.$$

Because $\delta > 0$ was arbitrary, this implies

$$\lim_{\alpha\to 1} \inf \mathcal{U}_0^{C,\alpha} \geq \inf \mathcal{U}_0 = \inf \mathcal{U}_0^{\mathrm{bdd}}.$$

To conclude the proof of Equation (32), we note that $(\inf \mathcal{U}_0^{\mathrm{bdd}} + \frac{\varepsilon}{2}) \in \mathcal{U}_0^C$ by definition. This implies that

$$\lim_{\alpha\to 1} \inf \mathcal{U}_0^{C,\alpha} \le \inf \mathcal{U}_0^C \le \inf \mathcal{U}_0^{\mathrm{bdd}} + \varepsilon,$$

hence Equation (32) follows. □

We can now prove the main result.

Theorem 3.14. *Assume σ is bounded and nonconstant. Further, suppose Assumption 3.11 is fulfilled. Then, for any $\varepsilon > 0$, there exists $\alpha_0 = \alpha_0(\varepsilon) \in (0,1)$ and $C = C(\varepsilon) \in (0,\infty)$ such that for all $\alpha \in (\alpha_0, 1)$*

$$\inf \mathcal{U}_0 + \varepsilon \ge \inf \mathcal{U}_0^{\Theta,C,\alpha} \ge \inf \mathcal{U}_0 - \varepsilon. \tag{38}$$

Proof. By Assumption 3.11, we can consider $\inf \mathcal{U}_0^{\mathrm{bdd}}$ instead of $\inf \mathcal{U}_0$. Then for $\varepsilon > 0$, there exists a predictable strategy $\tilde{\xi}$ such that $\sup_{1\le k\le T} \|\tilde{\xi}_k\|_\infty < \infty$ and $\inf \mathcal{U}_0^{\mathrm{bdd}} + \frac{\varepsilon}{2} + \sum_{k=1}^T \tilde{\xi}_k \cdot (X_k - X_{k-1}) \ge H$, $\mathbf{P}$-a.s. Define $C = C(\varepsilon)$ as in Equation (33) in the proof of Lemma 3.13. Then, we observe that $\mathcal{U}_0^{\Theta,C,\alpha} \subset \mathcal{U}_0^{C,\alpha}$ for all $\alpha \in (0,1)$ and Equation (32) implies that there exists $\alpha_0 = \alpha_0(\varepsilon) \in (0,1)$ such that for all $\alpha \in (\alpha_0, 1)$

$$\inf \mathcal{U}_0 - \varepsilon = \inf \mathcal{U}_0^{\mathrm{bdd}} - \varepsilon \le \inf \mathcal{U}_0^{C,\alpha} \le \inf \mathcal{U}_0^{\Theta,C,\alpha}, \tag{39}$$

which proves the second inequality in Equation (38).

To prove the first inequality in Equation (38), let α be given. Consider

$$M_n = \left\{ \inf \mathcal{U}_0^{\mathrm{bdd}} + \frac{\varepsilon}{2} + \sum_{k=1}^T \tilde{\xi}_k \cdot (X_k - X_{k-1}) \ge H \right\} \cap \{\|X_k - X_{k-1}\| \le n \text{ for } k = 1, \ldots, T\},$$

for $n \in \mathbb{N}$. Then $M_n \subset M_{n+1}$ and therefore, by continuity from below

$$1 = \mathbf{P}\left(\inf \mathcal{U}_0^{\mathrm{bdd}} + \frac{\varepsilon}{2} + \sum_{k=1}^T \tilde{\xi}_k \cdot (X_k - X_{k-1}) \ge H \right) = \mathbf{P}(\cup_{n\in\mathbb{N}} M_n) = \lim_{n\to\infty} \mathbf{P}(M_n).$$

Thus, we may choose $n \in \mathbb{N}$ such that $\mathbf{P}(M_n) \ge \frac{\alpha+1}{2}$. As $\tilde{\xi}$ is predictable, for each $k = 1, \ldots, T$, there exists a measurable function $f_k : (\mathbb{R}^{mk}, \mathcal{B}(\mathbb{R}^{mk})) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that $\tilde{\xi}_k = f_k(\mathcal{Y}_{k-1})$. By the universal approximation theorem (Hornik, 1991, Theorem 1 and Section 3), see also Theorem B.1 in the appendix, with measure μ given by the law of $\mathcal{Y}_{k-1}$ under $\mathbf{P}$, for each $k = 1, \ldots, T$, there exists $\theta_{k,\tilde{\xi}} \in \Theta$ such that

$$\mathbf{P}(D_k) < \frac{1-\alpha}{2T}, \quad \text{where } D_k = \left\{ \omega \in \Omega : \|f_k(\mathcal{Y}_{k-1}(\omega)) - F^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega))\| > \left(\frac{\varepsilon}{2nT} \wedge \frac{1}{2} \right) \right\}. \tag{40}$$

Define

$$\tilde{F}^{\theta_{k,\xi}} := \left(F^{\theta_{k,\xi}} \wedge C\right) \vee (-C), \quad k = 1, \dots, T.$$

By the definition of C in Equation (33), we get that

$$\|\tilde{\xi}_k\|_\infty + \left(\frac{\varepsilon}{2nT} \wedge \frac{1}{2}\right) < C \quad \text{for all } k = 1, \dots, T.$$

On D_k^c, we have for $i \in \{1, \dots, d\}$ that

$$\left|F_i^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1})\right| \le \left\|F^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1})\right\| \le \left\|\tilde{\xi}_k\right\|_\infty + \left(\frac{\varepsilon}{2nT} \wedge \frac{1}{2}\right) < C,$$

and hence $\tilde{F}_i^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}) = F_i^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1})$ on D_k^c. Conversely, for $\omega \in \Omega$ such that

$$\|f_k(\mathcal{Y}_{k-1}(\omega)) - \tilde{F}^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega))\| \le \left(\frac{\varepsilon}{2nT} \wedge \frac{1}{2}\right),$$

we get for $i \in \{1, \dots, d\}$ that

$$\left|\tilde{F}_i^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega))\right| \le \left\|\tilde{F}^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega))\right\| \le \left\|\tilde{\xi}_k\right\|_\infty + \left(\frac{\varepsilon}{2nT} \wedge \frac{1}{2}\right) < C,$$

and hence $\tilde{F}_i^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega)) = F_i^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega))$. In particular,

$$\begin{aligned} &\left\{\omega \in \Omega : \|f_k(\mathcal{Y}_{k-1}(\omega)) - \tilde{F}^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega))\| \le \left(\frac{\varepsilon}{2nT} \wedge \frac{1}{2}\right)\right\} \\ &= \underbrace{\left\{\omega \in \Omega : \|f_k(\mathcal{Y}_{k-1}(\omega)) - F^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega))\| \le \left(\frac{\varepsilon}{2nT} \wedge \frac{1}{2}\right)\right\}}_{=D_k^c}, \end{aligned}$$

for all $k = 1, \dots, T$. Therefore, we get that $D_k = \tilde{D}_k$ with

$$\tilde{D}_k = \left\{\omega \in \Omega : \|f_k(\mathcal{Y}_{k-1}(\omega)) - \tilde{F}^{\theta_{k,\tilde{\xi}}}(\mathcal{Y}_{k-1}(\omega))\| > \left(\frac{\varepsilon}{2nT} \wedge \frac{1}{2}\right)\right\},$$

and

$$\mathbf{P}(\tilde{D}_k) < \frac{1-\alpha}{2T}.$$

On $M_n \cap \tilde{D}_1^c \cap \ldots \cap \tilde{D}_T^c$, we have

$$\begin{aligned}\sum_{k=1}^{T} \tilde{\xi}_k \cdot (X_k - X_{k-1}) &= \sum_{k=1}^{T} (\tilde{\xi}_k - \tilde{F}^{\theta_{k,\xi}}(\mathcal{Y}_{k-1})) \cdot (X_k - X_{k-1}) + \sum_{k=1}^{T} \tilde{F}^{\theta_{k,\xi}}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1}) \\ &\leq \sum_{k=1}^{T} \|f_k(\mathcal{Y}_{k-1}) - \tilde{F}^{\theta_{k,\xi}}(\mathcal{Y}_{k-1})\| \|X_k - X_{k-1}\| \\ &+ \sum_{k=1}^{T} \tilde{F}^{\theta_{k,\xi}}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1}) \\ &\leq \frac{\varepsilon}{2} + \sum_{k=1}^{T} \tilde{F}^{\theta_{k,\xi}}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1})\end{aligned}$$

and therefore,

$$M_n \cap \tilde{D}_1^c \cap \ldots \cap \tilde{D}_T^c \subset \left\{ \inf \mathcal{U}_0^{\text{bdd}} + \varepsilon + \sum_{k=1}^{T} \tilde{F}^{\theta_{k,\xi}}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1}) \geq H \right\}.$$

This inclusion and the Fréchet inequalities yield

$$\begin{aligned}&\mathbf{P}\left(\inf \mathcal{U}_0^{\text{bdd}} + \varepsilon + \sum_{k=1}^{T} \tilde{F}^{\theta_{k,\xi}}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1}) \geq H \right) \\ &\geq \mathbf{P}(M_n \cap \tilde{D}_1^c \cap \ldots \cap \tilde{D}_T^c) \\ &\geq \mathbf{P}(M_n) + \mathbf{P}(\tilde{D}_1^c) + \cdots + \mathbf{P}(\tilde{D}_T^c) - T \geq \frac{\alpha+1}{2} + T\left(1 - \frac{1-\alpha}{2T}\right) - T = \alpha.\end{aligned}$$

This proves the left inequality of Equation (38). □

We now provide a counterexample, which shows that Assumption 3.11 is necessary for Equation (38) to hold. In particular, this implies that the approximation result in Theorem 3.14 cannot be applied without Assumption 3.11.

Example 3.15. Let Z_1, Z_2 be independent standard uniform random variables under $\mathbf{P}$ and let Φ be the standard normal cumulative distribution function. Consider the market model with $T = 2, d = 1, S_t^0 = 1$ for $t \in \{0, 1, 2\}, X_0 = 1, X_1 = Z_1 + 0.5, X_2 = X_1 + (Z_2 - 0.5)/(1 + |\Phi^{-1}(Z_1)|)$, and $Y = X$. Then, $\mathbf{P} \in \mathcal{P}$ and $H = 1 + \Phi^{-1}(Z_1)(X_2 - X_1)$ is a bounded contingent claim, since $|\Phi^{-1}(Z_1)(X_2 - X_1)| \leq |Z_2 - 0.5| < 1$. The self-financing strategy induced by the predictable process ξ^* with $\xi_1^* = 0, \xi_2^* = \Phi^{-1}(Z_1)$ and initial wealth 1 is a replicating strategy for H, hence $\inf \mathcal{U}_0 = 1$. On the other hand, let $v > 0$ and consider ξ predictable and bounded. If $V_T = v + \sum_{k=1}^{T} \xi_k(X_k - X_{k-1})$ super-replicates H, then

$$(1 + |\Phi^{-1}(Z_1)|)\xi_1(Z_1 - 0.5) + \xi_2(Z_2 - 0.5) + v - 1 \geq |\Phi^{-1}(Z_1)|[1 - v + (Z_2 - 0.5)\text{sign}(\Phi^{-1}(Z_1))]. \tag{41}$$

If $\xi_1 > 0$, then the left-hand side is negative on the event $\{Z_1 < \frac{1}{R}\}$ for $R > 0$ large enough, since ξ_2 is bounded. Thus, for $R > 0$ large enough, on the event $\{Z_1 < \frac{1}{R}\} \cap \{0.3 \le Z_2 \le 0.4\}$, the left-hand side of Equation (41) is negative, whereas the right-hand side is non-negative for $v < 1.1$. This means that necessarily $v \ge 1.1$, otherwise $V_T \ge H$ can not be satisfied $\mathbf{P}$-a.s. In the case $\xi_1 < 0$, the same conclusion can be made using the set $\{Z_1 > 1 - \frac{1}{R}\} \cap \{0.6 \le Z_2 \le 0.7\}$. This proves that $\inf \mathcal{V}_0^{\mathrm{bdd}} \ge 1.1 > \inf \mathcal{V}_0$, that is, Assumption 3.11 is not satisfied.

We now argue that in this example, the first inequality in Equation (38) does not hold, that is, for all $\alpha_0 \in (0,1)$, $C > 0$ there exists $\alpha \ge \alpha_0$ such that $\inf \mathcal{V}_0^{\Theta,C,\alpha} > \inf \mathcal{V}_0 + 0.05$. To prove this, let $v > 0$ and ξ predictable satisfy $\sup_{1\le k\le T} \|\xi_k\|_\infty \le C$ and $\mathbf{P}(v + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) \ge H) \ge \alpha_0$. The super-replication condition can again be rewritten as Equation (41) and same reasoning as above shows that there exists $R > 0$ such that for all $v < 1.1$, we have $\mathbf{P}(v + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) \ge H) \le 1 - \frac{1}{10R}$, since $\mathbf{P}(\{Z_1 < \frac{1}{R}\} \cap \{0.3 \le Z_2 \le 0.4\}) = \mathbf{P}(\{Z_1 > 1 - \frac{1}{R}\} \cap \{0.6 \le Z_2 \le 0.7\}) = \frac{1}{10R}$. Thus, for all $\alpha > 1 - \frac{1}{10R}$, it follows that $\inf \mathcal{V}_0^{C,\alpha} \ge 1.1$. The inclusion $\mathcal{V}_0^{\Theta,C,\alpha} \subset \mathcal{V}_0^{C,\alpha}$ then implies $\inf \mathcal{V}_0^{\Theta,C,\alpha} \ge \inf \mathcal{V}_0^{C,\alpha} \ge 1.1 > \inf \mathcal{V}_0 + 0.05$.

Remark 3.16. Note that in the proof of Theorem 3.14, we compute both the price at $t = 0$ and the superhedging strategy for the complete interval.

Remark 3.17. Thanks to the universal approximation theorem in Hornik (1991), we could in fact restrict our attention to neural networks with one hidden layer and the result in Theorem 3.14 remains valid. Thus, for each $k = 1, \dots, T$, we could fix $L = 2$, $N_0 = mk$, $N_2 = d$ and consider instead the simpler parameter sets

$$\begin{aligned}\Theta_k &= \cup_{N_1\in\mathbb{N}}(\mathbb{R}^{N_1\times mk} \times \mathbb{R}^{N_1}) \times (\mathbb{R}^{d\times N_1} \times \mathbb{R}^d)\\ \Theta_k^C &= ([-C,C]^{C\times mk} \times [-C,C]^C) \times ([-C,C]^{d\times N_1} \times [-C,C]^d).\end{aligned}$$

Note the simpler form of Θ_k^C, which is due to the fact that all one-hidden layer networks with $N_1 \le C$ hidden nodes can be written as one-hidden layer networks with C hidden nodes and appropriate weights set to 0.

4 | SUPERHEDGING PRICE FOR $t > 0$

In this section, we establish a method to approximate superhedging prices for $t > 0$. Using a version of the uniform Doob decomposition, see Theorem 7.5 of Föllmer and Schied (2016), the problem reduces to the approximation of the so-called process of consumption. In the first part, we build the theoretical basis for this approach. In the second part, we prove that this method can be used to approximate the superhedging price for $t > 0$ by neural networks.

4.1 | Uniform Doob decomposition

We briefly summarize some results on superhedging in discrete time in Corollary 2.3 below. For a more detailed overview, we refer to Chapter 7 of Föllmer and Schied (2016).

Recall that H denotes a discounted European claim satisfying

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] < \infty.$$

The superhedging price at $t = 0$, $\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H]$ and the associated strategy ξ can be calculated as in Section 3 and so we consider them as known. The remaining unknown component is the process of consumption B given by Equation (2). By Corollary 2.3,

$$\left(\operatorname*{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_t] \right)_{t=0,1,\dots,T}$$

is the smallest $\mathcal{P}$-supermartingale whose terminal value dominates H. Consider the stochastic process $\tilde{B} = (\tilde{B}_t)_{t=0,\dots,T}$ defined as $\tilde{B}_0 := 0$ and for $t = 1, \dots, T$,

$$\tilde{B}_t := \operatorname{ess\,sup} \mathcal{B}_t, \tag{42}$$

where

$$\mathcal{B}_t := \left\{ D_t \in L^0(\Omega, \mathcal{F}_t, \mathbf{P}) : \tilde{B}_{t-1} \le D_t \le \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ \mathbf{P}\text{-a.s.} \right\}. \tag{43}$$

Proposition 4.1. *We have that*

$$B_t = \tilde{B}_t \quad \mathbf{P}\text{-a.s., for all } t = 0, \dots, T,$$

where B is given in Equation (2) and $\tilde{B}$ in Equation (42), respectively.

Proof. The proof follows by induction. For $t = 0$, we have $B_0 = 0 = \tilde{B}_0$ by definition. For the induction step, assume that $B_{t-1} = \tilde{B}_{t-1}$ $\mathbf{P}$-a.s. for some $1 \le t \le T$. First, we observe that $B_t \ge \tilde{B}_{t-1}$ because B is increasing and by the assumption of the induction step. In addition, by Equation (4), we obtain

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ge B_t. \tag{44}$$

In particular, $B_t \in \mathcal{B}_t$ and thus $B_t \le \tilde{B}_t$ $\mathbf{P}$-a.s. Define $\tilde{V} = (\tilde{V}_s)_{s=0,\dots,T}$ by

$$\tilde{V}_s := \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{s} \xi_k \cdot (X_k - X_{k-1}) - \tilde{B}_s.$$

First, we note that

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) \ge H \ge 0 \quad \mathbf{P}\text{-a.s.},$$

and thus by Theorem 5.14 of Föllmer and Schied (2016), we have that

$$\left(\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{s} \xi_k \cdot (X_k - X_{k-1}) \right)_{s=0,\dots,T}$$

is a $\bar{\mathbf{P}}^*$-martingale for all $\bar{\mathbf{P}}^* \in \mathcal{P}$. Further, by Equations (42) and (43), we obtain

$$0 \leq \tilde{B}_s \leq \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \quad \text{for all } s = 0, \dots, T, \ \mathbf{P}\text{-a.s.}$$

and

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \in L^1(\Omega, \mathcal{F}, \bar{\mathbf{P}}^*) \quad \text{for all } \bar{\mathbf{P}}^* \in \mathcal{P},$$

implies that $\tilde{V}_s \in L^1(\Omega, \mathcal{F}_s, \mathbf{P}^*)$ for all $\mathbf{P}^* \in \mathcal{P}$ and all $s = 0, \dots, T$. In particular, since $\tilde{B}$ is increasing and non-negative, we can conclude that $\tilde{V}$ is a $\mathbf{P}^*$-supermartingale for all $\mathbf{P}^* \in \mathcal{P}$. Furthermore, $\tilde{V}_s \geq 0$ $\mathbf{P}$-a.s. for all $s = 0, \dots, T$. To this end, let $\mathbf{P}^* \in \mathcal{P}$ be arbitrary, then we have by the $\mathbf{P}^*$-supermartingale property that

$$\begin{aligned}
\tilde{V}_s &\geq \mathbb{E}^*[\tilde{V}_T \mid \mathcal{F}_s] \\
&= \mathbb{E}^*\left[\sup_{\bar{\mathbf{P}}^* \in \mathcal{P}} \mathbb{E}_{\bar{\mathbf{P}}^*}[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - \tilde{B}_T \mid \mathcal{F}_s \right] \\
&\geq \mathbb{E}^*[H \mid \mathcal{F}_s] \geq 0.
\end{aligned}$$

The terminal value of $\tilde{V}$ dominates H by construction and since $B_s \leq \tilde{B}_s$ for all $s = 0, \dots, T$, we have

$$\tilde{V}_s \leq \operatorname*{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_s] \quad \mathbf{P}\text{-a.s. for all } s = 0, 1 \dots, T.$$

Therefore, we get

$$\tilde{V}_s = \operatorname*{ess\,sup}_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mid \mathcal{F}_s] = \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{s} \xi_k \cdot (X_k - X_{k-1}) - B_s \quad \mathbf{P}\text{-a.s. for all } s = 0, 1 \dots, T,$$

and thus $B_t = \tilde{B}_t$ $\mathbf{P}$-a.s. This concludes the proof. □

Remark 4.2. In the definition of Equation (42), we can equivalently consider $\operatorname{ess\,sup} \widehat{\mathcal{B}}_t$, where

$$\widehat{\mathcal{B}}_t := \left\{ D_t \in L^0(\Omega, \mathcal{F}_t, \mathbf{P}) : 0 \leq D_t \leq \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ \mathbf{P}\text{-a.s.} \right\},$$

for $t = 1, \dots, T$. This is due to the fact that, on the one hand, $\mathcal{B}_t \subset \widehat{\mathcal{B}}_t$ for all $t = 1, \dots, T$. On the other hand, for $D_t \in \widehat{\mathcal{B}}_t$, we have that $\tilde{D}_t := D_t \vee B_{t-1} \in \mathcal{B}_t$ and $D_t \le \tilde{D}_t$ **P**-a.s. Therefore, $\operatorname{ess\,sup} \widehat{\mathcal{B}}_t = \operatorname{ess\,sup} \mathcal{B}_t = B_t$ for all $t = 1, \dots, T$.

4.2 | Neural network approximation for $t > 0$

We now study a neural network approximation for the superhedging price process for $t > 0$. Throughout this section, we use the notation of Section 3. For $\varepsilon, \tilde{\varepsilon} \in (0, 1)$, we define the set

$$\mathcal{B}_t^{\theta_t^*, \varepsilon, \tilde{\varepsilon}} := \Big\{ F^{\theta_t}(\mathcal{Y}_t) : \theta_t \in \Theta_{t+1} \text{ and } \mathbf{P}\Big(B_{t-1} - \tilde{\varepsilon} \le F^{\theta_t}(\mathcal{Y}_t) \le \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H + \tilde{\varepsilon} \Big) > 1 - \varepsilon \Big\},$$

where B is the consumption process for H introduced in Equation (2). We now construct an approximation of B by neural networks.

Proposition 4.3. *Assume σ is bounded and nonconstant. Then, for any $\varepsilon, \tilde{\varepsilon} > 0$, there exist neural networks $(F^{\theta_0, \varepsilon, \tilde{\varepsilon}}, \dots, F^{\theta_T, \varepsilon, \tilde{\varepsilon}})$ such that $F^{\theta_t, \varepsilon, \tilde{\varepsilon}}(\mathcal{Y}_t) \in \mathcal{B}_t^{\theta_t^*, \varepsilon, \tilde{\varepsilon}}$ for all $t = 0, \dots, T$ and*

$$\mathbf{P}\left(\left| F^{\theta_t, \varepsilon, \tilde{\varepsilon}}(\mathcal{Y}_t) - B_t \right| > \tilde{\varepsilon} \right) < \varepsilon, \quad \text{for all } t = 0, \dots, T.$$

In particular, there exists a sequence of neural networks $(F^{\theta_0^n}, \dots, F^{\theta_T^n})_{n \in \mathbb{N}}$ with $F^{\theta_t^n}(\mathcal{Y}_t) \in \mathcal{B}_t^{\theta_t^*, \frac{1}{n}, \frac{1}{n}}$ for all $n \in \mathbb{N}$ and for all $t = 0, \dots, T$ such that

$$\left(F^{\theta_0^n}(\mathcal{Y}_0), \dots, F^{\theta_T^n}(\mathcal{Y}_T) \right) \xrightarrow{\mathbf{P}\text{-a.s.}} (B_0, \dots, B_T) \quad \text{for } n \to \infty.$$

Proof. Fix $\varepsilon, \tilde{\varepsilon} > 0$ and $t \in \{1, \dots, T\}$. Note that $B_0 = 0$ by definition. Let B be given by the representation (42). Observe that the set $\mathcal{B}_t$ from Equation (43) is directed upwards. By Theorem A.33 of Föllmer and Schied (2016), there exists an increasing sequence

$$(B_t^k)_{k \in \mathbb{N}} \subset \mathcal{B}_t,$$

such that B_t^k converges **P**-almost surely to $\tilde{B}_t = B_t$ as k tends to infinity. Since almost sure convergence implies convergence in probability, there exists $K = K(\varepsilon, \tilde{\varepsilon}) \in \mathbb{N}$ such that

$$\mathbf{P}\left(\left| B_t^k - B_t \right| > \frac{\tilde{\varepsilon}}{2} \right) < \frac{\varepsilon}{2}, \quad \text{for all } k \ge K. \tag{45}$$

For all $k \ge K$, there exist measurable functions $f_t^k : \mathbb{R}^{mt} \to \mathbb{R}$ such that $B_t^k = f_t^k(\mathcal{Y}_t)$. Fix $k \ge K$. By the universal approximation theorem (Hornik, 1991, Theorem 1 and Section 3), see also Theorem B.1 in the appendix, (with measure μ given by the law of $\mathcal{Y}_t$ under **P**) there exists

$\theta_t = \theta_t^k \in \Theta_{t+1}$ and $F^{\theta_t} = F^{\theta_t^k,\varepsilon,\tilde{\varepsilon}}$ such that

$$\mathbf{P}\left(\left|f_t^k(\mathcal{Y}_t) - F^{\theta_t}(\mathcal{Y}_t)\right| > \frac{\tilde{\varepsilon}}{2}\right) < \frac{\varepsilon}{2}.$$

By the triangle inequality and by De Morgan's law, we obtain that

$$\left\{\omega \in \Omega : \left|B_t(\omega) - F^{\theta_t}(\mathcal{Y}_t(\omega))\right| > \tilde{\varepsilon}\right\} \subseteq \left\{\omega \in \Omega : \left|B_t(\omega) - B_t^k(\omega)\right| + \left|B_t^k - F^{\theta_t}(\mathcal{Y}_t(\omega))\right| > \tilde{\varepsilon}\right\}$$
$$\subseteq \left\{\omega \in \Omega : \left|B_t(\omega) - B_t^k(\omega)\right| > \frac{\tilde{\varepsilon}}{2}\right\} \cup \left\{\omega \in \Omega : \left|B_t^k(\omega) - F^{\theta_t}(\mathcal{Y}_t(\omega))\right| > \frac{\tilde{\varepsilon}}{2}\right\}.$$

In particular, we obtain by subaddidivity that

$$\mathbf{P}\left(\left|B_t - F^{\theta_t}(\mathcal{Y}_t)\right| > \tilde{\varepsilon}\right) \le \mathbf{P}\left(\left|B_t - B_t^k\right| > \frac{\tilde{\varepsilon}}{2}\right) + \mathbf{P}\left(\left|B_t^k - F^{\theta_t}(\mathcal{Y}_t)\right| > \frac{\tilde{\varepsilon}}{2}\right) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Next, we show that $F^{\theta_t} \in \mathcal{B}_t^{\theta_t^*,\varepsilon,\tilde{\varepsilon}}$. For this purpose, we note that

$$B_{t-1} \le B_t \le \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \quad \mathbf{P}\text{-a.s.}$$

Therefore, we have that

$$\mathbf{P}\left(B_{t-1} - \tilde{\varepsilon} \le F^{\theta_t}(\mathcal{Y}_t) \le \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H + \tilde{\varepsilon}\right) \ge \mathbf{P}\left(\left|B_t - F^{\theta_t}(\mathcal{Y}_t)\right| \le \tilde{\varepsilon}\right) > 1 - \varepsilon,$$

which implies that $F^{\theta_t}(\mathcal{Y}_t) = F^{\theta_t^k,\varepsilon,\tilde{\varepsilon}}(\mathcal{Y}_t) \in \mathcal{B}_t^{\theta_t^*,\varepsilon,\tilde{\varepsilon}}$. We set $\varepsilon = \frac{1}{n} = \tilde{\varepsilon}$ for $n \in \mathbb{N}$ and consider the neural network

$$F^{\theta_t^n} := F^{\theta_t^{K(n)},\frac{1}{n},\frac{1}{n}}, \quad t \in \{1, \dots, T\},\ n \in \mathbb{N},$$

where $K(n) = K(\frac{1}{n}, \frac{1}{n})$ is given by Equation (45). Then, $F^{\theta_t^n} \in \mathcal{B}_t^{\theta_t^*,\frac{1}{n},\frac{1}{n}}$ for all $n \in \mathbb{N}$ and for all $t = 1, \dots, T$. Further, we have

$$\mathbf{P}\left(\left|F^{\theta_t^n}(\mathcal{Y}_t) - B_t\right| > \frac{1}{n}\right) < \frac{1}{n} \quad \text{for all } t = 1, \dots, T,$$

which implies convergence in probability, that is,

$$F^{\theta_t^n}(\mathcal{Y}_t) \xrightarrow{\mathbf{P}} B_t \quad \text{for } n \to \infty, \text{ for all } t = 0, \dots, T.$$

By passing to a suitable subsequence, convergence also holds $\mathbf{P}$-a.s. simultaneously for all $t = 0, \dots, T$. $\square$

Let $\tilde{\varepsilon} > 0$. Recursively, we define the set

$$\begin{aligned}\tilde{\mathcal{B}}_t^{\theta_t^*,\tilde{\varepsilon}} := \{&F^{\theta_t}(\mathcal{Y}_t)\mathbb{1}_A + B_{t-1}^{\theta_{t-1}^*,\tilde{\varepsilon}}\mathbb{1}_{A^c} \,:\, \theta_t \in \Theta_{t+1}, A \in \mathcal{F}_t,\\ &B_{t-1}^{\theta_{t-1}^*,\tilde{\varepsilon}} \le F^{\theta_t}(\mathcal{Y}_t)\mathbb{1}_A + B_{t-1}^{\theta_{t-1}^*,\tilde{\varepsilon}}\mathbb{1}_{A^c} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) - H + \tilde{\varepsilon}\}\end{aligned} \tag{46}$$

for $t = 1, \dots, T$, and the approximated process of consumption by $B_0^{\theta_0^*,\tilde{\varepsilon}} = 0$ and

$$B_t^{\theta_t^*,\tilde{\varepsilon}} := \operatorname{ess\,sup} \tilde{\mathcal{B}}_t^{\theta_t^*,\tilde{\varepsilon}} \quad \text{for } t = 1, \dots, T. \tag{47}$$

Theorem 4.4. *Assume σ is bounded and non-constant. Then*

$$\left|B_t^{\theta_t^*,\tilde{\varepsilon}} - B_t\right| \le \tilde{\varepsilon} \quad \mathbf{P}\text{-a.s. for all } t = 0, \dots, T.$$

Proof. We prove the statement by induction. For $t = 0$, we have by definition $B_0^{\theta_0^*,\tilde{\varepsilon}} = B_0 = 0$. Assume now that

$$\left|B_{t-1}^{\theta_{t-1}^*,\tilde{\varepsilon}} - B_{t-1}\right| \le \tilde{\varepsilon} \quad \mathbf{P}\text{-a.s.}$$

for some $t \in \{1, \dots, T\}$. First, we note that $B_s^{\theta_s^*,\tilde{\varepsilon}} \le B_{s+1}^{\theta_{s+1}^*,\tilde{\varepsilon}}$ by Equations (46) and (47), and because $B_0^{\theta_0^*,\tilde{\varepsilon}} = 0$, it follows that $B_s^{\theta_s^*,\tilde{\varepsilon}} \ge 0$ for all $s = 1, \dots, T$. Let $\theta_t \in \Theta_{t+1}$ and $A \in \mathcal{F}_t$ such that

$$0 \le B_{t-1}^{\theta_{t-1}^*,\tilde{\varepsilon}} \le F^{\theta_t}(\mathcal{Y}_t)\mathbb{1}_A + B_{t-1}^{\theta_{t-1}^*,\tilde{\varepsilon}}\mathbb{1}_{A^c} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) - H + \tilde{\varepsilon}.$$

Then, we can easily see that

$$F^{\theta_t}(\mathcal{Y}_t)\mathbb{1}_A + B_{t-1}^{\theta_{t-1}^*,\tilde{\varepsilon}}\mathbb{1}_{A^c} \in \left\{ D_t \in L^0(\Omega, \mathcal{F}_t, \mathbf{P}) \,:\, 0 \le D_t \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) - H + \tilde{\varepsilon}\ \mathbf{P}\text{-a.s.} \right\}.$$

We now prove that

$$B_t + \tilde{\varepsilon} = \operatorname{ess\,sup} \left\{ D_t \in L^0(\Omega, \mathcal{F}_t, \mathbf{P}) \,:\, -\tilde{\varepsilon} \le D_t - \tilde{\varepsilon} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) - H\ \mathbf{P}\text{-a.s.} \right\}. \tag{48}$$

On the one hand, we have

$$\begin{aligned}&\left\{ \tilde{D}_t \in L^0(\Omega, \mathcal{F}_t, \mathbf{P}) \,:\, 0 \le \tilde{D}_t \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) - H\ \mathbf{P}\text{-a.s.} \right\} + \tilde{\varepsilon}\\ =&\left\{ D_t \in L^0(\Omega, \mathcal{F}_t, \mathbf{P}) \,:\, 0 \le D_t - \tilde{\varepsilon} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^T \xi_k \cdot (X_k - X_{k-1}) - H\ \mathbf{P}\text{-a.s.} \right\}\end{aligned}$$

$$\subseteq\left\{D_t \in L^0(\Omega,\mathcal{F}_t,\mathbf{P}) : -\tilde{\varepsilon} \le D_t - \tilde{\varepsilon} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ \mathbf{P}\text{-a.s.}\right\},$$

which by Remark 4.2 implies that

$$B_t + \tilde{\varepsilon} \le \operatorname{ess\,sup}\left\{D_t \in L^0(\Omega,\mathcal{F}_t,\mathbf{P}) : -\tilde{\varepsilon} \le D_t - \tilde{\varepsilon} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ \mathbf{P}\text{-a.s.}\right\}.$$

On the other hand, let

$$D_t \in \left\{D_t \in L^0(\Omega,\mathcal{F}_t,\mathbf{P}) : -\tilde{\varepsilon} \le D_t - \tilde{\varepsilon} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ \mathbf{P}\text{-a.s.}\right\},$$

and define $\tilde{D}_t := D_t \vee \tilde{\varepsilon}$. Then $D_t \le \tilde{D}_t$ **P**-a.s. and

$$\tilde{D}_t \in \left\{\bar{D}_t \in L^0(\Omega,\mathcal{F}_t,\mathbf{P}) : 0 \le \bar{D}_t - \tilde{\varepsilon} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ \mathbf{P}\text{-a.s.}\right\},$$

which implies that

$$B_t + \tilde{\varepsilon} \ge \operatorname{ess\,sup}\left\{D_t \in L^0(\Omega,\mathcal{F}_t,\mathbf{P}) : -\tilde{\varepsilon} \le D_t - \tilde{\varepsilon} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ \mathbf{P}\text{-a.s.}\right\},$$

and hence Equation (48) follows. Further, we also have that

$$\left\{D_t \in L^0(\Omega,\mathcal{F}_t,\mathbf{P}) : 0 \le D_t \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H + \tilde{\varepsilon} \ \mathbf{P}\text{-a.s.}\right\}$$
$$=\left\{D_t \in L^0(\Omega,\mathcal{F}_t,\mathbf{P}) : -\tilde{\varepsilon} \le D_t - \tilde{\varepsilon} \le \sup_{\mathbf{P}^*\in\mathcal{P}} \mathbb{E}^*[H] + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) - H \ \mathbf{P}\text{-a.s.}\right\}.$$

Therefore, we obtain by Equation (48) that

$$F^{\theta_t}(\mathcal{Y}_t)\mathbb{1}_A + B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}\mathbb{1}_{A^c} \le B_t + \tilde{\varepsilon} \quad \mathbf{P}\text{-a.s.},$$

and hence

$$B_t^{\theta^*_t,\tilde{\varepsilon}} \le B_t + \tilde{\varepsilon} \quad \mathbf{P}\text{-a.s.} \tag{49}$$

For the converse direction let $\varepsilon \in (0,1)$. By the proof of Proposition 4.3, there exists a neural network $F^{\tilde{\theta}_t} = F^{\tilde{\theta}_t,\varepsilon,\tilde{\varepsilon}}$ such that

$$\mathbf{P}\left(\left|F^{\tilde{\theta}_t}(\mathcal{Y}_t) - B_t\right| > \tilde{\varepsilon}\right) < \varepsilon.$$

Define the sets $A_1, A_2 \in \mathcal{F}_t$ by

$$A_1 := \left\{\omega \in \Omega : B_t(\omega) - \tilde{\varepsilon} \le F^{\tilde{\theta}_t}(\mathcal{Y}_t(\omega)) \le B_t(\omega) + \tilde{\varepsilon}\right\},$$

and

$$A_2 := \left\{\omega \in \Omega : B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}(\omega) \le F^{\tilde{\theta}_t}(\mathcal{Y}_t(\omega))\right\}.$$

Then, $\mathbf{P}(A_1) > 1 - \varepsilon$. Note that by the assumption of the induction

$$B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}} \le B_{t-1} + \tilde{\varepsilon} \le B_t + \tilde{\varepsilon} \quad \mathbf{P}\text{-a.s.}$$

For $A := A_1 \cap A_2$, we have by construction,

$$F^{\tilde{\theta}_t}(\mathcal{Y}_t)\mathbb{1}_A + B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}\mathbb{1}_{A^c} = F^{\tilde{\theta}_t}(\mathcal{Y}_t)\mathbb{1}_{A_1\cap A_2} + B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}\mathbb{1}_{A_1^c\cup A_2^c} \in \tilde{\mathcal{B}}_t^{\theta^*_t,\tilde{\varepsilon}}.$$

For $\omega \in A_1 \cap A_2^c$, we get that

$$F^{\tilde{\theta}_t}(\mathcal{Y}_t(\omega))\mathbb{1}_{A_1\cap A_2}(\omega) + B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}(\omega)\mathbb{1}_{A_1^c\cup A_2^c}(\omega) = B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}(\omega)$$

and

$$B_t(\omega) - \tilde{\varepsilon} \le F^{\tilde{\theta}_t}(\mathcal{Y}_t(\omega)) < B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}(\omega) \le B_t(\omega) + \tilde{\varepsilon}.$$

For $\omega \in A_1 \cap A_2$, we have

$$F^{\tilde{\theta}_t}(\mathcal{Y}_t(\omega))\mathbb{1}_{A_1\cap A_2}(\omega) + B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}(\omega)\mathbb{1}_{A_1^c\cup A_2^c}(\omega) = F^{\tilde{\theta}_t}(\mathcal{Y}_t(\omega))$$

and

$$\left|F^{\tilde{\theta}_t}(\mathcal{Y}_t(\omega)) - B_t(\omega)\right| \le \tilde{\varepsilon}.$$

Thus, using that $A_1 = (A_1 \cap A_2) \cup (A_1 \cap A_2^c)$ and $\mathbf{P}(A_1) > 1 - \varepsilon$, we get

$$\mathbf{P}\left(\left|\left(F^{\tilde{\theta}_t}(\mathcal{Y}_t)\mathbb{1}_A + B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}\mathbb{1}_{A^c}\right) - B_t\right| > \tilde{\varepsilon}\right) \le \mathbf{P}(A_1^c) < \varepsilon. \tag{50}$$

Then, Equation (50) implies

$$\mathbf{P}\left(B_t^{\theta^*_t,\tilde{\varepsilon}} < B_t - \tilde{\varepsilon}\right) \le \mathbf{P}\left(F^{\tilde{\theta}_t}(\mathcal{Y}_t)\mathbb{1}_A + B_{t-1}^{\theta^*_{t-1},\tilde{\varepsilon}}\mathbb{1}_{A^c} < B_t - \tilde{\varepsilon}\right) < \varepsilon. \tag{51}$$

Because $\varepsilon \in (0,1)$ was arbitrary, it follows that $B_t \le B_t^{\theta^*_t,\tilde{\varepsilon}} + \tilde{\varepsilon}$ $\mathbf{P}$-a.s. by Equation (51). By Equations (49) and (51), we conclude that $|B_t^{\theta^*_t,\tilde{\varepsilon}} - B_t| \le \tilde{\varepsilon}$ $\mathbf{P}$-a.s. for all $t = 0, \dots, T$. □

5 | NUMERICAL RESULTS

In this section, we present some numerical applications for the results in Sections 3 and 4. Combining Theorems 3.4 and 3.14, we obtain a two-step approximation for the superhedging price at $t = 0$. Then, we use Theorem 4.4 to simulate the superhedging process for $t > 0$.

5.1 | Case $t = 0$

5.1.1 | Algorithm and implementation

Let $N \in \mathbb{N}$ denote a fixed batch size. For fixed $\lambda > 0$, we implement the following iterative procedure: for each iteration step i, we generate i.i.d. samples $Y(\omega_0^{(i)}), \dots, Y(\omega_N^{(i)})$ of Y and consider the empirical loss function

$$L_\lambda^{(i)}(\theta) = \left| F^{\theta_u}\left(\mathcal{Y}_0\left(\omega_0^{(i)}\right)\right)\right|^2 + \frac{\lambda}{N}\sum_{j=1}^{N} l\Bigg(H\left(\omega_j^{(i)}\right) - \left[F^{\theta_u}\left(\mathcal{Y}_0\left(\omega_j^{(i)}\right)\right) + \sum_{k=1}^{T} F^{\theta_{k,\xi}}\left(\mathcal{Y}_{k-1}\left(\omega_j^{(i)}\right)\right)\cdot\left(X_k\left(\omega_j^{(i)}\right) - X_{k-1}\left(\omega_j^{(i)}\right)\right)\right]\Bigg),$$

with $\theta = (\theta_u, \theta_{1,\xi}, \dots, \theta_{T,\xi})$. In the sequel, we consider two possible choices for the function $l : \mathbb{R} \to [0, \infty)$, that is, the squared *rectifier* function

$$l(x) = (\max\{x, 0\})^2, \tag{52}$$

or the *truncated sigmoid function*

$$l(x) = \left(\max\left\{\frac{1}{1+e^{-x}} - \frac{1}{2}, 0\right\}\right)^2, \tag{53}$$

see Figure 1 for the truncated sigmoid function.

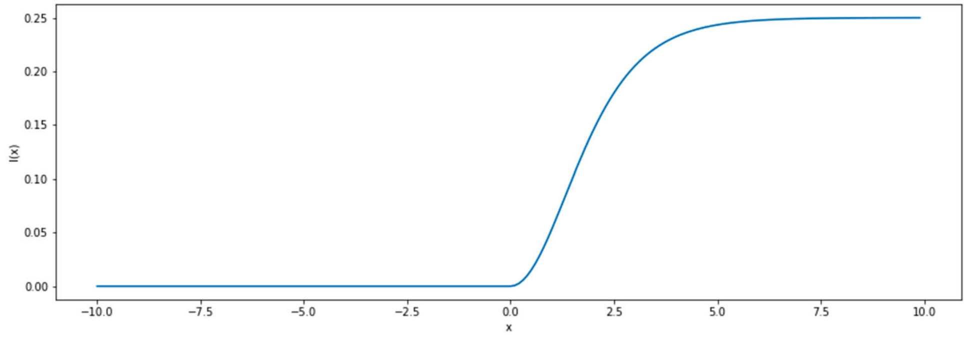


FIGURE 1 Truncated sigmoid function (53) [Color figure can be viewed at wileyonlinelibrary.com]

We then calculate the gradient of $L_\lambda^{(i)}(\theta)$ at $\theta^{(i)}$ and use it to update the parameters from $\theta^{(i)}$ to $\theta^{(i+1)}$ according to the *Adam* optimizer, see Kingma and Ba (2014). After sufficiently many iterations i, we expect (see, e.g., (Goodfellow et al., 2016, Chapter 8) or (Buehler et al., 2019, Section 4.3)) that the loss $L_\lambda(\theta^{(i)})$ is close to the minimal value of the loss function L_λ, where L_λ is given by

$$L_\lambda(\theta) = \left|F^{\theta_u}(\mathcal{Y}_0)\right|^2 + \lambda \mathbb{E}\left[l\left(H - \left(F^{\theta_u}(\mathcal{Y}_0) + \sum_{k=1}^{T} F^{\theta_{k,\xi}}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1}) \right)\right)\right]. \tag{54}$$

Note that $\mathcal{Y}_0$ is constant and hence $F^{\theta_u}(\mathcal{Y}_0)$ is a constant. We obtain a small value for the first term of L_λ if $F^{\theta_u}(\mathcal{Y}_0)$ representing the superhedging price is small. On the other hand, for our choices of l, the second summand in Equation (54) is equal to 0 when the portfolio dominates the claim H. Thus, minimizing the second summand in Equation (54) corresponds to maximizing the superhedging probability. When l is given by the truncated sigmoid function (53), it is considered as an approximation of the indicator function, which corresponds one to one to the method presented in Section 3. The weight λ offers the opportunity to balance between a small initial price of the portfolio and a high probability of superhedging. The superhedging probability can always be estimated on a test set. In order to guarantee a given superhedging probability, we can retrain the network with a larger choice of λ until the desired superhedging probability is achieved. In particular, if θ is the minimum for the loss function $L_\lambda(\theta)$, then $F^{\theta_u}(\mathcal{Y}_0)$ is close to the minimal price required to superhedge the claim H with a certain probability, that is, to the quantile hedging price for a certain $\alpha = \alpha(\lambda)$. In view of Theorem 3.14, we thus expect $F^{\theta_u}(\mathcal{Y}_0) \approx \inf \mathcal{U}_0$ for λ large enough.

Remark 5.1. We note that also other choices for l are possible as for instance the sigmoid function. However, the standard sigmoid function is very sensitive to the choice of λ because it is strictly monotone on the complete real line. In general, for different choices of l, the stability in the learning process and the impact of λ may vary. It appears to be important to choose a loss function that does not reward higher excess of the superhedging portfolio compared to a perfect hedge without excess.

The algorithm is implemented in Python, using Keras with backend TensorFlow to build and train the neural networks. More precisely, we create a *Sequential* object to build the models and compile with a customized loss function.

We use a long-short-term-memory network (LSTM), see Hochreiter and Schmidhuber (1997), with the following architecture: the network has two LSTM layers of size 30, which return sequences and one dense layer of size 1. Between the layers, the *swish* activation function is used. The activation functions within the LSTM layers are set to default, that is, activation between cells is *tanh* and the recurrent activation is the *sigmoid* function. The kernel and bias initializer of the first LSTM layer are set to *truncated normal*, that is, the initial weights are drawn from a standard normal distribution but we discard and redraw values, which are more than two standard deviations from the mean. This gives 11, 191 trainable parameters. The training is performed using the *Adam* optimizer with a learning rate of 0.001 or 0.0001. We generate 1024, 000 samples, which we split in 70% for the training set and 30% for the test set. The batch size is set to 1024. We apply the procedure described above in two examples, which we present in the following. The code is available at https://github.com/tomrtsm/neural_network_approximation_for_superhedging_prices.

5.1.2 | Trinomial model

We consider a discrete time financial market model given by an arbitrage-free trinomial model with $X_0 = 100$ and

$$X_t = X_0 \prod_{k=1}^{t} (1 + R_t), \quad t \in \{0, \dots, T\},$$

where R_t is $\mathcal{F}_t$-measurable for $t \in \{1, \dots, T\}$, and takes values in $\{d, m, u\}$ with equal probability, where $-1 < d < m < u$. Here, we set $d = -0.01$, $m = 0$, and $u = 0.01$ and $T = 29$ yielding 3^{29} possible paths. In this model, we want to superhedge a European Call option $H = (X_T - K)^+$ with strike price $K = 100$. For this choice of parameters, the theoretical superhedging price is 2.17, as it can be easily obtained by Proposition 3.4 of Carassus and Vargiolu (2010).

The network is trained and evaluated for different λ to illustrate the impact of λ in Equation (54) and the relation between $\alpha(\lambda) \in (0, 1)$ and the corresponding $\alpha(\lambda)$-quantile hedging price. For each λ, the network is trained over 40 epochs.

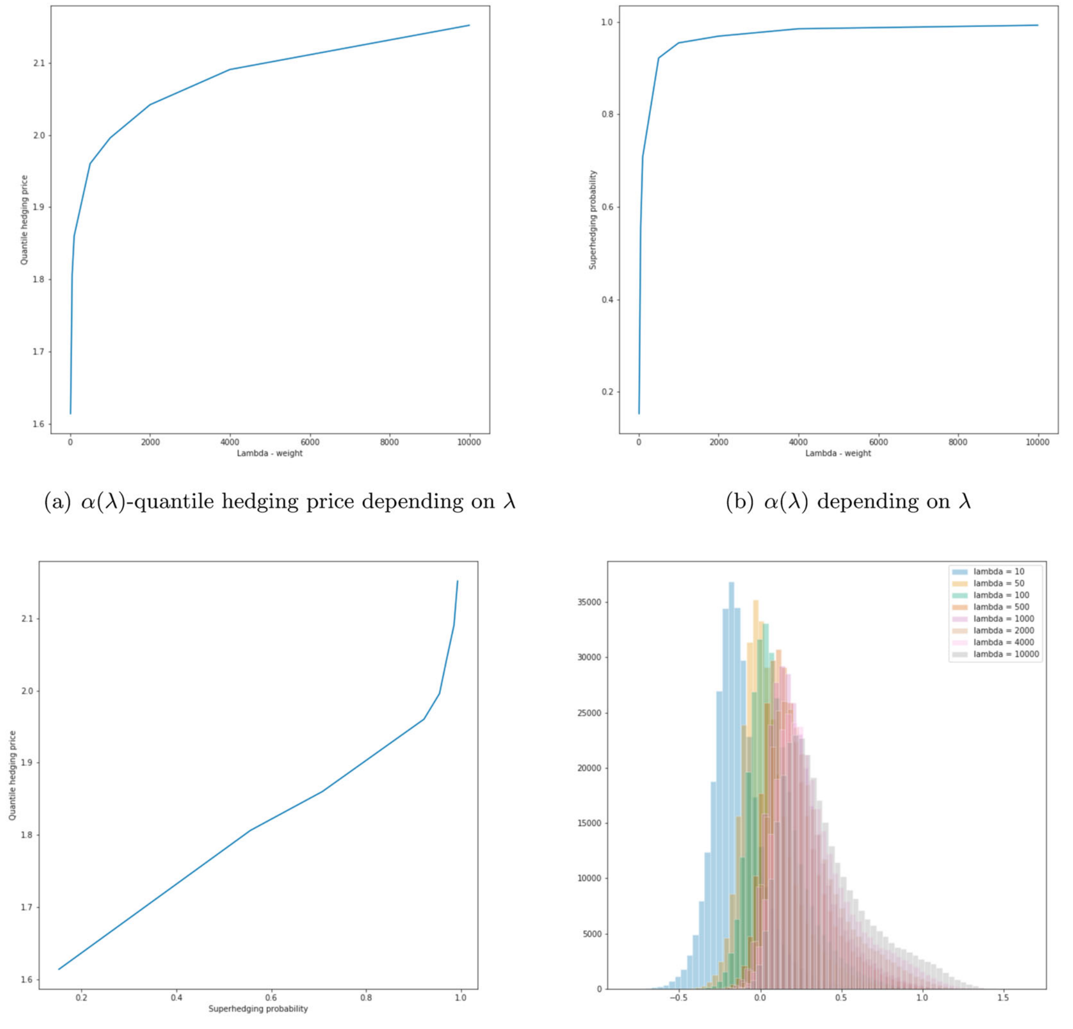
In Figure 2a–c for l given by Equation (52) and Figure 3a–c for l given by Equation (53), we see that $\alpha(\lambda)$ as well as the $\alpha(\lambda)$-quantile hedging price increase in λ, and that the $\alpha(\lambda)$-quantile hedging price increases in $\alpha(\lambda)$. Figures 2d and 3d show the superhedging performance on the test set for all λ's, that is, samples of

$$F^{\theta_u(\lambda)}(\mathcal{Y}_0) + \sum_{k=0}^{T} F^{\theta_{k,\xi}(\lambda)}(\mathcal{Y}_{k-1}) \cdot (X_k - X_{k-1}) - H, \tag{55}$$

for each λ. Tables 1 and 2 summarize the values for λ, $\alpha(\lambda)$ and the $\alpha(\lambda)$-quantile hedging price for l given by Equations (52) and (53), respectively. In particular, for $\lambda = 10,000$, we obtain a numerical price of 2.15 and $\alpha(\lambda) = 99.24\%$ for l given by Equation (52). For $\lambda = 300,000$, we obtain a numerical price of 2.16 and $\alpha(\lambda) = 99.69\%$ for l given by Equation (53). Furthermore, the analysis shows that the impact of λ significantly depends on the choice of l.

TABLE 1 Impact of λ on $\alpha(\lambda)$ and on the $\alpha(\lambda)$-quantile hedging price for $l = (\max\{x, 0\})^2$

λ	$\alpha(\lambda)$	$\alpha(\lambda)$-quantile hedging price
10	15.23%	1.61
50	55.61%	1.81
100	70.75%	1.86
500	92.16%	1.96
1000	95.42%	2.00
2000	96.88%	2.04
4000	98.48%	2.09
10, 000	99.24%	2.15



(a) $\alpha(\lambda)$-quantile hedging price depending on λ

(b) $\alpha(\lambda)$ depending on λ

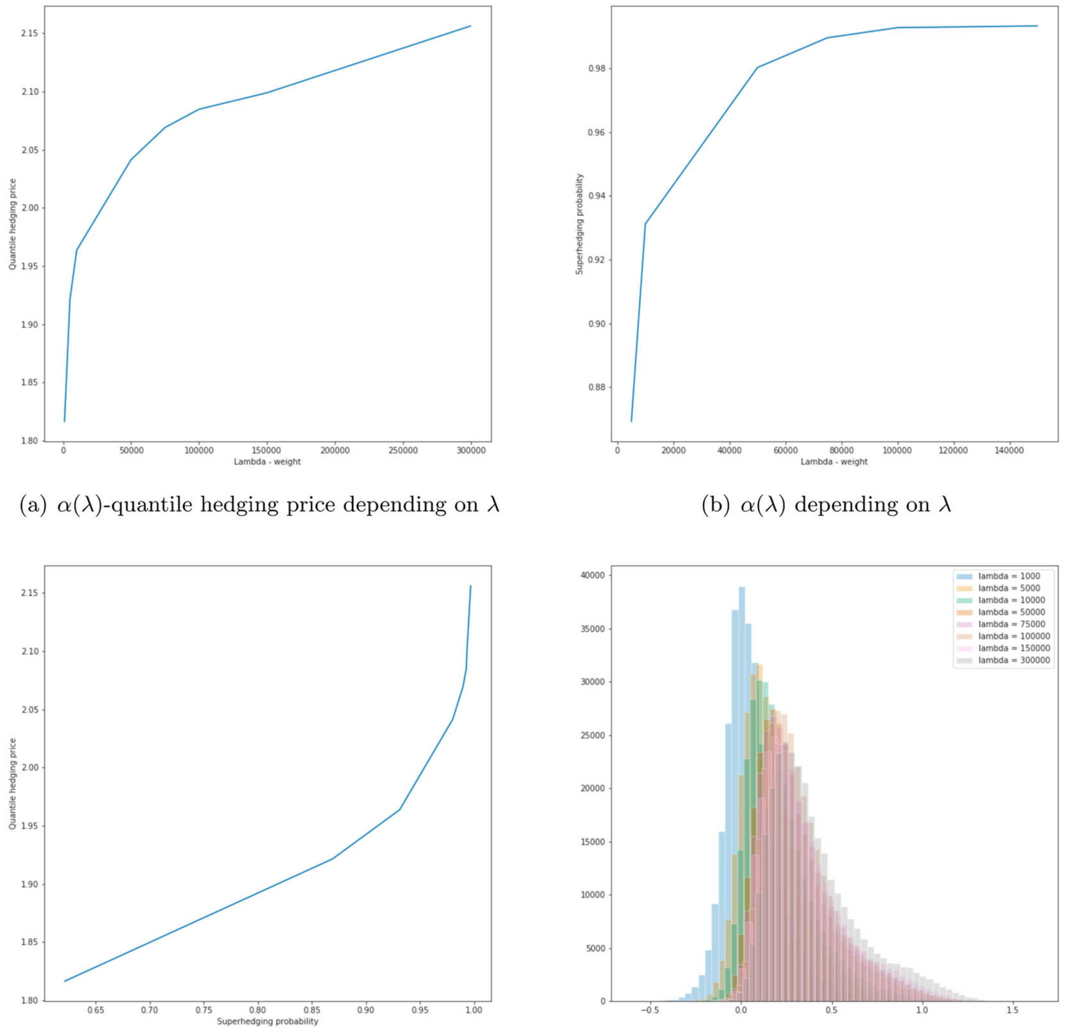
(c) $\alpha(\lambda)$-quantile hedging price depending on $\alpha(\lambda)$

(d) Superhedging performance

FIGURE 2 Impact of λ on the quantile hedging price and on the superhedging probability for $l = (\max\{x, 0\})^2$ [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Impact of λ on $\alpha(\lambda)$ and on the $\alpha(\lambda)$-quantile hedging price for $l = (\max\{\frac{1}{1+e^{-x}} - \frac{1}{2}, 0\})^2$

λ	$\alpha(\lambda)$	$\alpha(\lambda)$-quantile hedging price
1000	62.11%	1.82
5000	86.92%	1.92
10,000	93.12%	1.96
50,000	98.03%	2.04
75,000	98.97%	2.07
100,000	99.28%	2.08
150,000	99.34%	2.10
300,000	99.69%	2.16



(a) $\alpha(\lambda)$-quantile hedging price depending on λ

(b) $\alpha(\lambda)$ depending on λ

(c) $\alpha(\lambda)$-quantile hedging price depending on $\alpha(\lambda)$

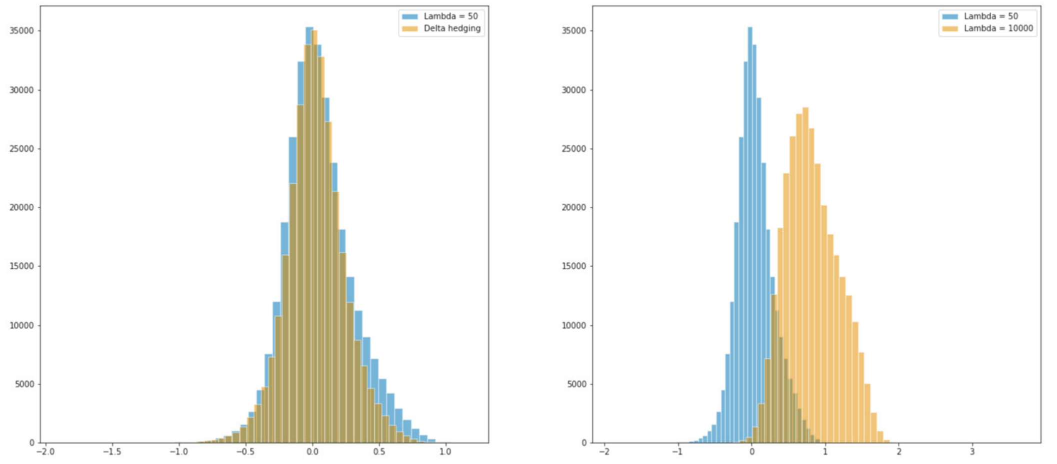
(d) Superhedging performance

FIGURE 3 Impact of λ on the quantile hedging price and on the superhedging probability for $l = (\max\{\frac{1}{1+e^{-x}} - \frac{1}{2}, 0\})^2$ [Color figure can be viewed at wileyonlinelibrary.com]

5.1.3 | Discretized Black–Scholes model

Here, we consider a discrete time financial market for the asset price X given by a Black–Scholes model considered at discrete times th for $t \in \{0, 1, \dots, 30\}$ and $h = \frac{1}{250}$. We consider a Barrier Up and Out Call option $H = \prod_{t=0}^{T} \mathbb{1}_{\{X_t < U\}}(X_T - K)^+$ with strike $K = 100$ and upper bound $U = 105$ such that $K < U$ and $X_0 < U$. We set $X_0 = 100$, $\sigma = 0.3$, and $\mu = 0$. We assume to have 250 trading days per year and a time horizon T of 30 trading days with daily rebalancing. In particular, for a European contingent claim, the time until expiration for the option is $\tau = 30/250$.

The weight λ of the loss function is set to $10,000,000$ in order to obtain a high superhedging probability. Indeed, we obtain a superhedging probability of 100% on the training set as well as



(a) δ-hedging strategy compared to approximate strategy for $\lambda = 50$

(b) Approximate strategy for $\lambda = 50$ and $\lambda = 10000$

FIGURE 4 Hedging losses for $\lambda = 50$, $\lambda = 10000$ and for the δ-hedging strategy [Color figure can be viewed at wileyonlinelibrary.com]

on the test set with an approximate price of 3.73. By tab. 1 in Carassus et al. (2007), the theoretical superhedging price π^H is given by

$$\pi^H = X_0\left(1 - \frac{K}{U}\right) \approx 4.76.$$

In the Black–Scholes model, the asset price process at time $t > 0$ has unbounded support and thus, the additional error, which arises from the discretization of the probability space, is non-negligible. Although the Barrier option artificially bounds the support of the model, the numerical price still significantly deviates from the theoretical price.

Finally, we consider a European call option $H = (X_T - K)^+$ with strike $K = 100$ and parameters $X_0 = 100$, $\sigma = 0.1$ and $\mu = 0$. By Carassus et al. (2007), the theoretical price of H for the discrete time version of the Black–Scholes model is equal to $X_0 = 100$. The theoretical price of H in a standard Black–Scholes model in the continuous time is 1.38, and by following the δ-hedging strategy, we superhedge H with a probability of 53.69%. We first choose l given by Equation (52). Here, we consider $\lambda = 50$ in Equation (54) in order to compare the result to the discretized δ-hedging strategy of the Black–Scholes model, and $\lambda = 10,000$ in order to obtain a high superhedging probability. For $\lambda = 50$, we obtain an approximate price of 1.41 and a superhedging probability of 54.43%. In Figure 4a, we compare the δ-hedging strategy with the approximated superhedging strategy obtained for $\lambda = 50$. Further, in Figure 4b, we compare the results for $\lambda = 50$ and $\lambda = 10,000$, respectively. For $\lambda = 10,000$, the superhedging probability on the test set is 99.79% with an approximated price of 2.18.

For l given by Equation (53), we obtain for $\lambda = 100,000$, a superhedging probability of 99.66% and a price of 2.10. Again, we can see the different impact of λ for different choices of l in Equation (54).

5.2 | Case $t > 0$

In this section, we approximate the process of consumption by neural networks as proposed in Section 4.2. We implement the same iterative procedure as introduced in Section 5.1.1. We define $G^{(i)}$ as the difference of the approximated superhedging strategy obtained from Section 5.1 and the claim H, that is,

$$G_j^{(i)}(\theta^*) := \left[F^{\theta_u^*}\left(\mathcal{Y}_0\left(\omega_j^{(i)}\right)\right) + \sum_{k=1}^{T} F^{\theta_{k,\xi}^*}\left(\mathcal{Y}_{k-1}\left(\omega_j^{(i)}\right)\right) \cdot \left(X_k\left(\omega_j^{(i)}\right) - X_{k-1}\left(\omega_j^{(i)}\right)\right) - H\left(\omega_j^{(i)}\right)\right].$$

Then, the empirical loss function is given by

$$\tilde{L}_{t,\beta}^{(i)}(\theta_t) = \frac{1}{N} \sum_{j=1}^{N} -\left| B_t^{\theta_t}\left(\omega_j^{(i)}\right)\right|^2 + \beta \max\left\{\left(B_t^{\theta_t}\left(\omega_j^{(i)}\right) - G_j^{(i)}(\theta^*)\right), 0\right\},$$

where $B_t^{\theta_t}$ is given by

$$B_t^{\theta_t}\left(\omega_j^{(i)}\right) := \max\left\{F^{\theta_t}\left(\mathcal{Y}_t\left(\omega_j^{(i)}\right)\right), B_{t-1}^{\theta_{t-1}}\left(\omega_j^{(i)}\right)\right\}.$$

At a local minimum, the two terms of $\tilde{L}$ guarantee that F^{θ_t} is as big as possible but less or equal than $G(\theta^*)$.

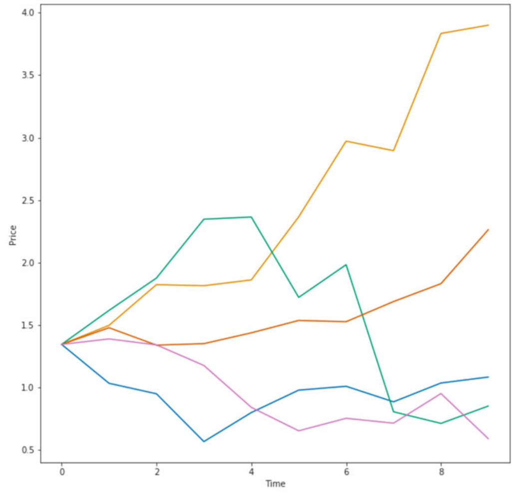
Here, we also consider a discretized Black–Scholes model as in Section 5.1.3 but only a time horizon of 10 trading days and set $X_0 = 100$, $\sigma = 0.1$, and $\mu = 0$. For each $t > 0$, the neural network consists of two LSTM layers of size 30 and 20, respectively, which return sequences, one LSTM layer of size 20 providing one single value and one dense layer of size 1. The remaining parameters are chosen as in Section 5.1.1.

As in Section 5.1.3, we compute an approximated superhedging price and strategy for the complete interval. Setting $\lambda = 1024$ yields an approximated price of 1.35 and a superhedging probability of 98.87% for $t = 0$. For $t \geq 1$, we choose $\beta = 500$ and then obtain a superhedging probability of 98.78%. In Figure 5a, we show trajectories of the approximated superhedging price process generated by this method. Figure 5b illustrates paths given by the δ-hedging strategy of the discretized Black–Scholes model. Finally, we plot the difference of the approximated superhedging price processes and the corresponding price process obtained by the δ-hedging strategy in Figure 5c.

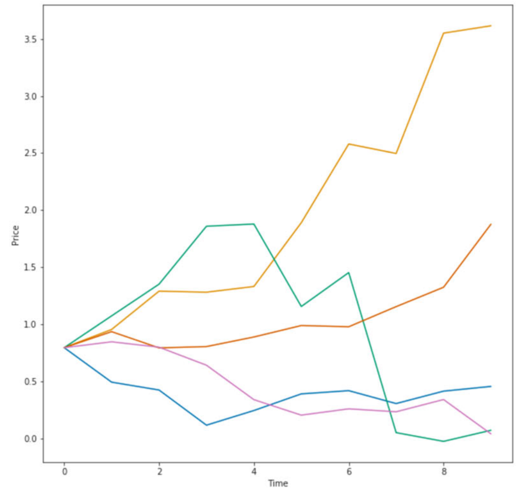
5.3 | Discussion

In finite market models as in Section 5.1.2, our methodology delivers an approximation of α-quantile hedging and approximated superhedging prices with small approximation error. It is also worth noting that the predicted superhedging price and the corresponding superhedging probability of the training set are consistent with the values on the test set.

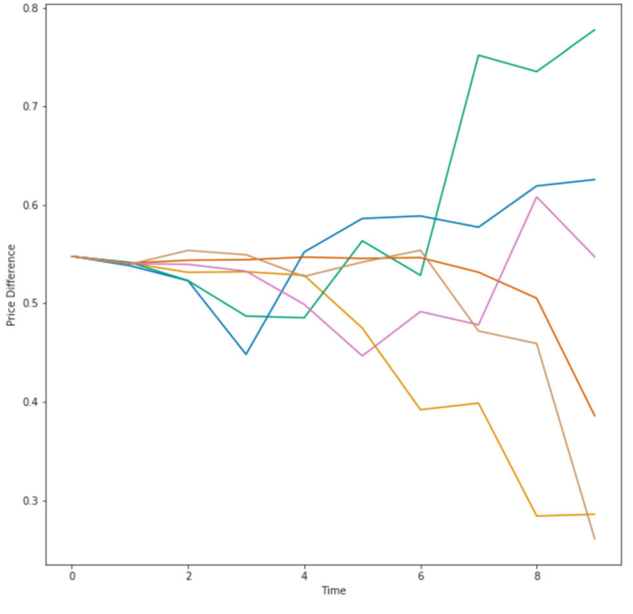
In contrast, in models in which the price process has unbounded support, our numerical results indicate that the additional error caused by the discretization of the probability space cannot be ignored. However, we obtain consistent results of the α-quantile hedging price for the training set



(a) Superhedging price process



(b) δ-hedging price process



(c) Difference of the price processes

FIGURE 5 Superhedging price process compared to the δ-hedging price process [Color figure can be viewed at wileyonlinelibrary.com]

and test set. Note also that, in Section 5.1.3, the Barrier option can be superhedged with 100% on the training and on the test set.

A further possible application of our methodology is given by superhedging in a model-free setting on prediction sets, see Bartl et al. (2020), Bartl et al. (2019), Hou and Obłój (2018), where prediction sets offer the opportunity to include beliefs in price developments or to select relevant price paths.

ACKNOWLEDGMENT

Thomas Reitsam gratefully acknowledges the financial support of the Verein zur Versicherungswissenschaft München e.V.

Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in "neural_network_approximation_for_superhedging_prices" at http://doi.org/10.5281/zenodo.6378421.

ORCID

Francesca Biagini https://orcid.org/0000-0001-9801-5259

REFERENCES

Bartl, D., Kupper, M., & Neufeld, A. (2020). Pathwise superhedging on prediction sets. *Finance and Stochastics*, *24*(1), 215–248.

Bartl, D., Kupper, M., Prömel, D. J., & Tangpi, L. (2019). Duality for pathwise superhedging in continuous time. *Finance and Stochastics*, *23*(3), 697–728.

Becker, S., Cheridito, P., & Jentzen, A. (2019). Deep optimal stopping. *Journal of Machine Learning Research*, *20*, 74.

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

Burzoni, M., Frittelli, M., & Maggis, M. (2017). Model-free superhedging duality. *Annals of Applied Probability*, *27*(3), 1452–1477.

Campi, L., & Schachermayer, W. (2006). A super-replication theorem in Kabanov's model of transaction costs. *Finance and Stochastics*, *10*(4), 579–596.

Carassus, L., Gobet, E., & Temam, E. (2007). A class of financial products and models where super-replication prices are explicit. In *Stochastic processes and applications to mathematical finance* (pp. 67–84). World Scientific.

Carassus, L., Obłój, J., & Wiesel, J. (2019). The robust superreplication problem: A dynamic approach. *SIAM Journal on Financial Mathematics*, *10*(4), 907–941.

Carassus, L., & Vargiolu, T. (2010). Super-replication price for asset prices having bounded increments in discrete time.

Cuchiero, C., Khosrawi, W., & Teichmann, J. (2020). A generative adversarial network approach to calibration of local stochastic volatility models. *Risks*, *8*(4), 101.

Cvitanić, J., & Karatzas, I. (1996). Hedging and portfolio optimization under transaction costs: A martingale approach 1 2. *Mathematical Finance*, *6*(2), 133–165.

El Karoui, N., & Quenez, M.-C. (1995). Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM Journal on Control and Optimization*, *33*(1), 29–66.

Föllmer, H., & Leukert, P. (1999). Quantile hedging. *Finance and Stochastics*, *3*, 251–273.

Föllmer, H., & Schied, A. (2016). *Stochastic finance: An introduction in discrete time* (4th rev. ed.). Walter de Gruyter.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. http://www.deeplearningbook.org

Hochreiter, S., & Schmidhuber, J. (1997). Lstm can solve hard long time lag problems. In *Proceedings of the Advances in Neural Information Processing Systems*, 473–479.

Hornik, K. (1991). Approximation capabilities of muitilayer feedforward networks. *Neural Networks*, *4*(1989), 251–257.

Hou, Z., & Obłój, J. (2018). Robust pricing–hedging dualities in continuous time. *Finance and Stochastics*, *22*(3), 511–567.

Kabanov, Y. M., & Last, G. (2002). Hedging under transaction costs in currency markets: A continuous-time model. *Mathematical Finance*, *12*(1), 63–70.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kramkov, D. O. (1996). Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets. *Probability Theory and Related Fields*, *105*(4), 459–479.

Nutz, M. (2015). Robust superhedging with jumps and diffusion. *Stochastic Processes and their Applications*, *125*(12), 4543–4555.
Nutz, M., & Soner, H. M. (2012). Superhedging and dynamic risk measures under volatility uncertainty. *SIAM Journal on Control and Optimization*, *50*(4), 2065–2089.
Obłój, J., & Wiesel, J. (2021). Robust estimation of superhedging prices. *The Annals of Statistics*, *49*(1), 508–530.
Ruf, J., & Wang, W. (2020). Neural networks for option pricing and hedging: A literature review. *Journal of Computational Finance*, Forthcoming.
Schachermayer, W. (2014). The super-replication theorem under proportional transaction costs revisited. *Mathematics and Financial Economics*, *8*(4), 383–398.
Soner, H. M., Shreve, S. E., & Cvitanic, J. (1995). There is no nontrivial hedging portfolio for option pricing with transaction costs. *The Annals of Applied Probability*, *5*(2), 327–355.
Touzi, N. (2014). Martingale inequalities, optimal martingale transport, and robust superhedging. *ESAIM: Proceedings and Surveys*, *45*, 32–47.

How to cite this article: Biagini, F., Gonon, L., & Reitsam, T. (2023). Neural network approximation for superhedging prices. *Mathematical Finance*, *33*, 146–184.
https://doi.org/10.1111/mafi.12363

APPENDIX A: PROOF OF PROPOSITION 3.2

Proof. "$\leq$": Take $A \in \mathcal{F}_T$ such that $\mathbf{P}(A) \geq \alpha$. We prove that

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] \in \left\{ u \in \mathbb{R} \,:\, \exists \xi \text{ adm. s.t. } \mathbf{P}\left(u + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) \geq H \right) \geq \alpha \right\}. \tag{A.1}$$

By the well-known superhedging duality, see Theorem 7.13 of Föllmer and Schied (2016), we have that

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] = \inf \left\{ u \in \mathbb{R} \,:\, \exists \xi \text{ pred. s.t. } u + \sum_{k=1}^{T} \xi_k \cdot (X_k - X_{k-1}) \geq H\mathbb{1}_A \ \mathbf{P}\text{-a.s.} \right\},$$

and that there exists a superhedging strategy $\hat{\xi}$ for $H\mathbb{1}_A$ with initial value $\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A]$, that is,

$$\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] + \sum_{k=1}^{T} \hat{\xi}_k \cdot (X_k - X_{k-1}) \geq H\mathbb{1}_A \geq 0 \ \mathbf{P}\text{-a.s.} \tag{A.2}$$

In particular, by Equation (A.2) we get for $\hat{\xi}$ that

$$\mathbf{P}\left(\sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] + \sum_{k=1}^{T} \hat{\xi}_k \cdot (X_k - X_{k-1}) \geq H \right) \geq \mathbf{P}(A) \geq \alpha.$$

This implies Equation (A.1) and hence

$$\inf \mathcal{U}_0^{\alpha} \leq \inf \left\{ \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H\mathbb{1}_A] \,:\, A \in \mathcal{F}_T,\ \mathbf{P}(A) \geq \alpha \right\}.$$

"$\geq$": Take $\tilde{u} \in \mathcal{U}_0^\alpha$ and denote by $\tilde{\xi} = (\tilde{\xi}_k)_{k=1}^T$ the corresponding strategy such that

$$\mathbf{P}\left(\tilde{u} + \sum_{k=1}^{T} \tilde{\xi}_k \cdot (X_k - X_{k-1}) \geq H \right) \geq \alpha.$$

Define the set $\tilde{A}$ by

$$\tilde{A} := \left\{ \omega \in \Omega : \tilde{u} + \sum_{k=1}^{T} \tilde{\xi}_k(\omega) \cdot (X_k(\omega) - X_{k-1}(\omega)) \geq H(\omega) \right\}.$$

Clearly $\tilde{A} \in \mathcal{F}_T$ and $\mathbf{P}(\tilde{A}) \geq \alpha$. By construction we have that

$$\left(\tilde{u} + \sum_{k=1}^{T} \tilde{\xi}_k \cdot (X_k - X_{k-1}) \right) \mathbb{1}_{\tilde{A}} \geq H \mathbb{1}_{\tilde{A}} \quad \mathbf{P}\text{-a.s.}$$

and because $\tilde{\xi}$ is assumed to be admissible, we have

$$\left(\tilde{u} + \sum_{k=1}^{T} \tilde{\xi}_k \cdot (X_k - X_{k-1}) \right) \mathbb{1}_{\tilde{A}^c} \geq 0 \quad \mathbf{P}\text{-a.s.}$$

In particular, $\tilde{u} \in \mathcal{U}_0(H \mathbb{1}_{\tilde{A}})$ and by Theorem 7.13 of Föllmer and Schied (2016) we obtain

$$\tilde{u} \geq \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mathbb{1}_{\tilde{A}}] \in \left\{ \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mathbb{1}_A] : A \in \mathcal{F}_T, \mathbf{P}(A) \geq \alpha \right\}. \tag{A.3}$$

That is, for an arbitrary $\tilde{u} \in \mathcal{U}_0^\alpha$, we have constructed a set $\tilde{A}$ such that Equation (A.3) holds. Therefore,

$$\inf \mathcal{U}_0^\alpha \geq \inf \left\{ \sup_{\mathbf{P}^* \in \mathcal{P}} \mathbb{E}^*[H \mathbb{1}_A] : A \in \mathcal{F}_T, \mathbf{P}(A) \geq \alpha \right\}.$$

□

APPENDIX B: NEURAL NETWORKS

For the reader's convenience, we recall some results on neural networks. The following result essentially follows from Hornik (1991, Theorem 1). For completeness, we include its proof here.

Theorem B.1. *Assume σ is bounded and nonconstant. Let $f : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$ be a measurable function and μ be a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then for any $\varepsilon, \tilde{\varepsilon} > 0$ there exists a neural network g such that*

$$\mu(\{x \in \mathbb{R}^d : \|f(x) - g(x)\| > \tilde{\varepsilon}\}) < \varepsilon.$$

Proof. Let $\varepsilon, \tilde{\varepsilon} > 0$ be given and let $C > 0$ satisfy that

$$\mu(\{x \in \mathbb{R}^d : \|f(x)\| > C\}) < \frac{\varepsilon}{2}. \tag{B.1}$$

Define $\tilde{f} = \mathbb{1}_{\{x \in \mathbb{R}^d : \|f(x)\| \leq C\}} f$. Then $\tilde{f} \in L^1(\mathbb{R}^d, \mu)$ and hence (Hornik, 1991, Theorem 1) shows that there exists a neural network g with

$$\int_{\mathbb{R}^d} \|\tilde{f}(x) - g(x)\| \mu(dx) < \frac{\varepsilon\tilde{\varepsilon}}{4}.$$

Markov's inequality thus proves that

$$\mu(\{x \in \mathbb{R}^d : \|\tilde{f}(x) - g(x)\| > \frac{\tilde{\varepsilon}}{2}\}) \leq \frac{2}{\tilde{\varepsilon}} \int_{\mathbb{R}^d} \|\tilde{f}(x) - g(x)\| \mu(dx) < \frac{\varepsilon}{2}. \tag{B.2}$$

Combining Equations (B.1) and (B.2) and recalling $f - \tilde{f} = f \mathbb{1}_{\{x \in \mathbb{R}^d : \|f(x)\| > C\}}$ yields

$$\begin{aligned} \mu\big(\{x \in \mathbb{R}^d : \|f(x) - g(x)\| > \tilde{\varepsilon}\}\big) &\leq \mu\Big(\{x \in \mathbb{R}^d : \|f(x) - \tilde{f}(x)\| > \frac{\tilde{\varepsilon}}{2}\} \cup \{x \in \mathbb{R}^d : \|\tilde{f}(x) - g(x)\| > \frac{\tilde{\varepsilon}}{2}\}\Big) \\ &< \mu\big(\{x \in \mathbb{R}^d : \|f(x)\| > C\}\big) + \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

□

APPENDIX C: FURTHER RESULTS

Theorem C.1 Theorem A.37, Föllmer and Schied (2016). *Let Φ be any set of random variables on* $(\Omega, \mathcal{F}, \mathbf{P})$.

1. *There exists a random variable φ^* with the following two properties.*
   (a) $\varphi^* \geq \varphi$ $\mathbf{P}$-*a.s. for all* $\varphi \in \Phi$.
   (b) $\varphi^* \leq \psi$ $\mathbf{P}$-*a.s. for every random variable ψ satisfying* $\psi \geq \varphi$ $\mathbf{P}$-*a.s. for all* $\varphi \in \Phi$.
2. *Suppose in addition that Φ is directed upwards, that is, for $\varphi, \tilde{\varphi} \in \Phi$ there exists $\psi \in \Phi$ with $\psi \geq \varphi \vee \tilde{\varphi}$. Then there exists an increasing sequence $\varphi_1 \leq \varphi_2 \leq \ldots$ in Φ such that* $\varphi^* = \lim_{n\to\infty} \varphi_n$ $\mathbf{P}$-*a.s.*

Definition C.2 Definition A.38, Föllmer and Schied (2016). The random variable φ^* in Theorem C.1 is called the essential supremum of Φ with respect to $\mathbf{P}$, and we write

$$\operatorname{ess\,sup} \Phi = \operatorname*{ess\,sup}_{\varphi \in \Phi} \varphi := \varphi^*.$$

The essential infimum of Φ with respect to $\mathbf{P}$ is defined as

$$\operatorname{ess\,inf} \Phi = \operatorname*{ess\,inf}_{\varphi \in \Phi} := -\operatorname*{ess\,sup}_{\varphi \in \Phi}(-\varphi).$$

For random variables $(\xi_n)_{n\in\mathbb{N}} \subset L^0(\Omega, \mathcal{F}, \mathbf{P}; \mathbb{R}^d)$ we denote by $\mathrm{conv}\{\xi_1, \xi_2, \ldots\}$ the convex hull of $\xi_1, \xi_2, \ldots$ which is defined ω-wise.

Lemma C.3 Lemma 1.70, Föllmer and Schied (2016). *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence in $L^0(\Omega, \mathcal{F}, \mathbf{P}; \mathbb{R}^d)$ such that* $\sup_{n\in\mathbb{N}} |\xi_n| < \infty$ $\mathbf{P}$*-a.s. Then, there exists a sequence of convex combinations*

$$\eta_n \in conv\{\xi_n, \xi_{n+1}, \ldots\}, \quad n \in \mathbb{N},$$

which converges $\mathbf{P}$-almost surely to some $\eta \in L^0(\Omega, \mathcal{F}, \mathbf{P}; \mathbb{R}^d)$.