Machine Learning for Stochastics
Financial Mathematics

Thorsten Schmidt

Abteilung für Mathematische Stochastik

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de

SS 2025

A very short introduction to Financial mathematics

In this lecture we will shortly visit Financial mathematics.

- For all who read and understand German, I uploaded my scriptum together with the first introductory lecture.
- There is an excellent book available: Hans Föllmer and Alexander Schied (2011). **Stochastic finance: an introduction in discrete time**. Walter de Gruyter.
- We are only interested in Chapters 5.1, 5.2, 5.3, 5.5 and later in chapter 8.1 and 8.2

The multi-period market model

- We consider 1 bank account and d stocks, making up $d+1$ assets.
- They are modelled through their price processes $S^0, \ldots, S^d$ given by

$$S^i = (S_0^i, S_1^i, \ldots, S_T^i)$$

- Our time horizon is therefore $\mathbb{T} = \{0, \ldots, T\}$. $S^0 > 0$ is the bank account. We can not loose all money here.
- Trading is done as follows: I buy today, say at time $t-1$ a number of shares, say $\bar{H}$. Tomorrow, at t I sell these shares. My gain (or loss) is obviously

$$\sum_{i=0}^{d} \bar{H}^i (S_t^i - S_{t-1}^i).$$

- We now make a convention: we denote $\bar{H}$ as $\bar{H}_t$. This makes notation easy and simple. However, $\bar{H}_t$ is already known at $t-1$! Such processes are called predictable.
- This is made precise via a filtration $\mathbb{F}$. $\mathbb{F}$ is an increasing family of σ-algebras $(\mathscr{F}_t)_{t=0,\ldots,T}$. A process X is called adapted if $X_t \in \mathscr{F}_t$ and predictable if $X_t \in \mathscr{F}_{t-1}$.

Definition

A trading strategy H is a predictable, $d+1$-dimensional process. It is called **self-financing** if

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

Self-financing simply means that by selling and buying at each time t (and switching from $\bar{H}_t$ to $\bar{H}_{t+1}$) we do not gain or loose money. We can only buy shares for as much money as we have.
But of course it is possible to borrow money (through the bank account S^0).
Often we will have

$$S_t^0 = \prod_{s=1}^{t} (1 + r_s)$$

with some interest rate r.

Lemma

For a self-financing trading strategy $\bar{H}$ and $t \geq 1$

$$\bar{H}_t \cdot \bar{S}_t = \bar{H}_1 \cdot \bar{S}_0 + \sum_{k=1}^{t} \bar{H}_k \cdot \Delta \bar{S}_k.$$

This is easy to show. Intuitively, the wealth at time t consists of initial wealth plus gains from trade.

Discounted prices

- A very nice trick is to consider discounted prices (note that 1 EUR tomorrow is of course different in value compared to 1 EUR today). We define
$$X_t^i := \frac{S_t^i}{S_t^0}, \quad t = 0, \dots, T, \quad i = 0, \dots, d.$$
Then $X^0 \equiv 1$ and the representations simplify. In particular, if $\bar{H}$ is self-financing it is already defined by specifying $H = (H^1, \dots, H^d)$!
- We introduce the discounted wealth process $V = V^{\bar{H}}$
$$V_t := \bar{H}_t \cdot \bar{X}_t, \quad t = 1, \dots, T,$$
$V_0 := \bar{H}_1 \cdot \bar{X}_0$. and the discounted gains process $G = G^{\bar{H}}$ by
$$G_t := \sum_{k=1}^{t} H_k \cdot \Delta X_k, \quad t = 1, \dots, T$$
with $G_0 = 0$.

Proposition

For a trading strategy $\bar{H}$ t.f.a.e.

1. $\bar{H}$ is self-financing
2. $\bar{H}_t \cdot \bar{X}_t = \bar{H}_{t+1} \cdot \bar{X}_t, \quad t = 1, \ldots, T-1,$
3. $V_t = V_0 + G_t$ for $0 \leqslant t \leqslant T$.

The central concept is **arbitrage**. It is a risk-less gain through trading.

Definition

An arbitrage is a self-financing trading strategy H, s.t.

1. $V_0 \leqslant 0,$
2. $V_T \geqslant 0$ and
3. $P(V_T > 0) > 0.$

A market without arbitrages is called arbitrage-free.

- We can now show that every market is free of arbitrage, if and only if every single-period market (S_t, S_{t+1}) is free of arbtirage (a key result).
- Note that Hans Föllmer and Alexander Schied require positivity in their book for this step, although it is not really necessary.
- We are able to classify arbitrage-free marktes through martingales.

Definition

A stochastic process M is a Q-**martingale**, if

1. M is adapted
2. $E_Q[|M_t|] < \infty$ for $t = 0, \dots, T$,
3. $M_s = E_Q[M_t|\mathscr{F}_s]$ for $0 \leqslant s \leqslant t \leqslant T$.

We call two measures P and Q equivalent ($P \sim Q$) if for all $F \in \mathscr{F}$

$$P(F) = 0 \Leftrightarrow Q(F) = 0.$$

We call a measure Q a martingale measure if X is a martingale under Q.

The first fundamental theorem

Theorem

A financial market is free of arbitrage, if and only if there exist an equivalent martingale measure

European claims

- We are most interested in derivatives like calls and puts. A call on a stock S^i offers the payoff
$$(S_T - K)^+$$
at maturity T when the so-called strike is K.
- More generally, we call any $\mathscr{F}_T$-measurable random variable C an European contingent claim.
- How can we define arbitrage-free prices for C? We choose an equivalent martingale measure Q and price the claim via
$$C_t = E_Q[C_T|\mathscr{F}_t].$$
- Then, the price process (S, C) is again a Q-martingale and the market is still free of arbtirage !
- Please check the Black-Scholes formula in the scriptum or in the book.

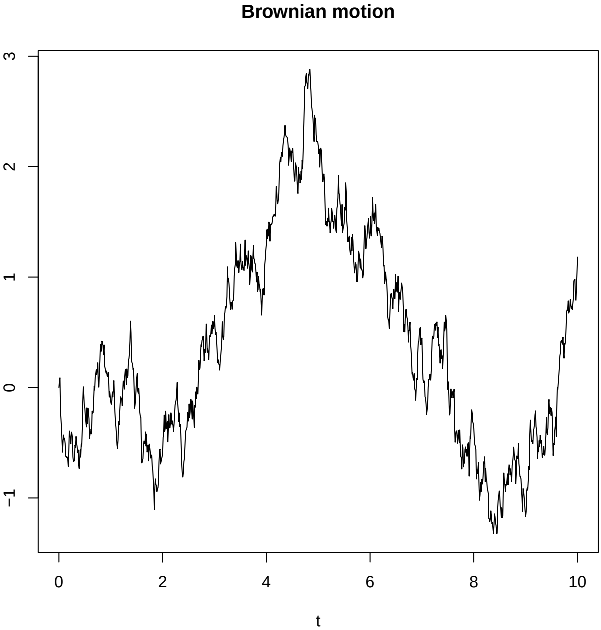
The Black-Scholes model

- Recall the definition of a Brownian motion: this is a continuous process B with independent and stationary (and, by the central limit theorem) normally distributed increments, starting in 0.
- Hence, $B_t - B_s$ is independent from $B_s = B_s - 0$.
- We call B **standard** when $B_t \sim \mathscr{N}(0, t)$. We say B has volatility σ if $B_t \sim \mathscr{N}(0, \sigma^2 t)$.
- Intuitively, we can discretize B and see it as a limit of the discretization

$$B_{t_n} = \sum_{i=1}^{n} \sqrt{t_i - t_{i-1}}\, \xi_i,$$

with some time-points $t_0 < t_1 < \cdots$ and (ξ_i) i.i.d. $\mathscr{N}(0, 1)$.

Simulation



- The Black-Scholes model is a geometric Brownian motion.
- It is often described through a stochastic differential equation

$$dS_t = S_t \mu dt + S_t \sigma dB_t,$$

  with initial value $S_0 = 0$.
- This is equivalent to

$$S_t = S_0 + \int_0^t S_s \mu ds + \int_0^t S_s \sigma dB_s.$$

- The first integral is a classical integral. The second integral is a **stochastic integral**. It is obtained as an appropriate limit of the elementary sums

$$\sum_{i=1}^{n} S_{t_{i-1}} \sigma (B_{t_i} - B_{t_{i-1}}).$$

- An important consequence is that a function $f \in C^2$ of Brownian motion can be represented via the stochastic integral. This is the important Ito-formula.

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds.$$

- With this formula, we can show that the solution of the above SDE is

$$S_t = S_0 \exp\left(\sigma B_t + \frac{(\mu - \sigma^2)t}{2}\right).$$

- You have already looked at the Black-Scholes formula. It says that

$$E[(S_t - K)^+] = S_0\Phi(d_1) - K\Phi(d_2),$$

where Φ is the cdf of a standard normal distribution and

$$d_{1/2} = \frac{\log(\frac{S}{K}) \pm \frac{\sigma^2 t}{2}}{\sqrt{\sigma^2 t}}$$

(we have assumed zero interest rate here).

- Our central question will be: how to apply deep learning in this setting ?

Motivation

- Affine processes have been considered in many variants (in particular in insurance markets)
- In applications, the parameters of those processes have to be estimated, thus leading to a certain amount of model risk
- What can be done to incorporate this model risk into our models ?
- How should it be incorporated in a general (but not too restrictive) way ?

Philosphical aspects

Uncertainty can be motivated in a number of situations:

1. **Statistics:** We have some past data and are able to estimate the parameters. **But** the estimation is not perfect and still carries uncertainty $\rightarrow$ (confidence intervals, Bayesian Statistics)
2. We have **no idea** about the distribution of the future evolution of X, except some rough guesses about intervals of parameters
3. We believe that the future evolution is close to the observed evolution, but not exactly like it.

Where do we place ourselves? Certainly, this depends on the task we want to achieve!

1. While in insurance we often have a large sample of independent data, stationarity might be questioned (both in life insurance and non-life insurance), at least if we consider a longer time horizon (longevity, dynamic mortality tables, etc)!
2. In any case, if statistical errors (confidence bounds) shall be incorporated in a **dynamic** framework, Markov processes are a useful tool.

Affine processes (no parameter uncertainty)

- Consider the state space $\mathcal{X} = \mathbb{R}$ or $\mathcal{X} = \mathbb{R}^+$ (canonical state space)
- A (time-homogeneous) Markov processes X is called **affine**, if

$$\mathbb{E}[e^{iuX_T} \mid \mathscr{F}_t] = e^{\phi(T-t,u)+\psi(T-t,u)X_t}$$

  for all $u \in i\mathbb{R}$, $0 \leq t \leq T$ with appropriate functions ϕ and ψ.
- Then, $X = X^x$ is the strong solution of

$$dX_t = (b^0 + b^1 X_t)dt + \sqrt{a^0 + a^1 X_t}\, dW_t, \qquad X_0 = x. \tag{1}$$

  where the parameter vector $\theta := (b^0, b^1, a^0, a^1)^\top$ satisfies certain admissibility conditions and W is a standard Brownian motion.
- Most prominent examples:

$$dX_t = \kappa(\theta - X_t)dt + \sigma\, dW_t, \tag{Vasiček}$$
$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}\, dW_t, \tag{CIR}$$

  To incorporate seasonalities, θ can be replaced by a function $\theta(t)$.
- Note that there are also affine processes with discontinuities (affine semimartingales $\rightarrow$ Martin Keller-Ressel, Thorsten Schmidt, Robert Wardenga, et al. (2019). „Affine processes beyond stochastic continuity“. In: **The Annals of Applied Probability** 29.6, pp. 3387–3437)

Affine processes under uncertainty

The concept of **uncertainty** goes back to the influental work of Frank Knight (1921): "Risk, uncertainty and profit."

To incorporate it in our setting we have to utilize some machinery:

- Let $\Omega = C([0,T];\mathbb{R})$ be the canonical space of continuous paths.
- Let X be the canonical process $X_t(\omega) = \omega_t$, and let $\mathbb{F} = (\mathscr{F}_t)_{t \geq 0}$ with $\mathscr{F}_t = \sigma(X_s, 0 \leq s \leq t)$ be the (raw) filtration generated by X.
- Denote by $\mathfrak{P}(\Omega)$ the Polish space of all probability measures on Ω equipped with the topology of weak convergence.

- Consider $P \in \mathfrak{P}(\Omega)$.
- X will be called a P-$\mathbb{F}$-semimartingale, if there exist $B = B^P$ and $M = M^P$ satisfying $B_0 = M_0 = 0$ such that

$$X = X_0 + B + M,$$

  here B has paths of (locally) finite variation P-a.s. and M is a P-$\mathbb{F}$-local martingale.
- Then X is a P-$\mathbb{F}$-semimartingale if and only if it has this property for the right-continuous filtration $\mathbb{F}_+ = (\mathscr{F}_{t+})_{t\geq 0}$ or for the usual augmentation $\mathbb{F}_+^P$; here $\mathscr{F}_{t+} = \cap_{s>t}\mathscr{F}_s$.
- We focus on semimartingales with **absolutely continuous characteristics**, i.e. where there exist predictable processes β and $\alpha \geq 0$, such that

$$B = \int_0^{\cdot} \beta_s ds, \quad C = \langle M \rangle = \int_0^{\cdot} \alpha_s ds.$$

- A probability measure $P \in \mathfrak{P}(\Omega)$ is called a semimartingale law for X, if X is a $(P, \mathbb{F})$-semimartingale. We denote

$$\mathfrak{P}_{\text{sem}} = \{P \in \mathfrak{P}(\Omega) \,|\, X \text{ is a } (P, \mathbb{F})\text{-semimartingale with a.c. characteristics}\}.$$

- We will consider **model risk** in the sense that there is uncertainty on the parameter vector $\theta = (b^0, b^1, a^0, a^1)$ of the affine process.
- Assume there is additional information on bounds on the parameter vector θ, denoted by

$$\underline{b}_i, \quad \bar{b}_i, \quad \underline{a}_i, \quad \bar{a}_i$$

leading to

$$\Theta = [\underline{b}^0, \bar{b}^0] \times [\underline{b}^1, \bar{b}^1] \times [\underline{a}^0, \bar{a}^0] \times [\underline{a}^1, \bar{a}^1]. \tag{2}$$

- We are interested in the intervals generated by the associated affine functions: let

$$\begin{aligned} b(x) &:= \{b^0 + b^1 x : (b^0, b^1) \in [\underline{b}^0, \bar{b}^0] \times [\underline{b}^1, \bar{b}^1]\}, \\ a(x) &:= \{a^0 + a^1 x : (a^0, a^1) \in [\underline{a}^0, \bar{a}^0] \times [\underline{a}^1, \bar{a}^1]\} \end{aligned} \tag{3}$$

  for $x \in \mathbb{R}$.
- Note that these sets are always intervals, due to the nice structure of Θ: for example,

$$b(x) = [\underline{b}^0 + \min(\underline{b}^1 x, \bar{b}^1 x), \bar{b}^0 + \max(\underline{b}^1 x, \bar{b}^1 x)].$$

Definition

Consider $t \in [0, T]$ and a semimartingale law $P \in \mathfrak{P}_{\text{sem}}$. We call P **affine-dominated on** $[t, T]$ by Θ, if (β^P, α) satisfy

$$\beta_s^P \in b(X_s), \quad \text{and} \quad \alpha_s \in a(X_s), \tag{4}$$

for dt-almost all $s \in [t, T]$ for P-almost all $\omega \in \Omega$. If $t = 0$, we call P **affine-dominated**.

Definition

A **non-linear affine process** starting at $x \in \mathbb{R}$ is a semimartingale law $P \in \mathfrak{P}_{\text{sem}}$, such that

(i) $P(X_0 = x) = 1$,

(ii) P is affine-dominated by Θ.

- Denote by $\mathcal{A}(x, \Theta)$ those semimartingale laws $P \in \mathfrak{P}_{\text{sem}}$, satisfying $P(X_0 = x) = 1$ and being dominated by Θ.

Dynamic programming

- We will utilize general results on dynamic programming
- The key to dynamic programming is a certain stability property under stopping (and we skip the details for now)

Conditional non-linear affine processes

- Denote by $\mathcal{A}(x, \Theta)$ those semimartingale laws $P \in \mathfrak{P}_{\text{sem}}$, satisfying $P(X_0 = x) = 1$ and being dominated by Θ.
- By $\mathcal{A}^t(x, \Theta)$ we denote all semimartingale laws $P \in \mathfrak{P}_{\text{sem}}$, such that $P(X_t = x) = 1$ and P is affine-dominated by Θ on $[t, T]$.
- Finally, we are interested in

$$\mathcal{P}(t, \omega) := \mathcal{A}^t(\omega_t, \Theta), \qquad t \geq 0,\ \omega \in \Omega,$$

the laws of non-linear affine processes on $[t, T]$, starting in t at ω_t.

The Kolmogorov equation

Consider the state space $\mathscr{O}$ which will be either $\mathbb{R}$, $\mathbb{R}_{\geq 0}$ or $\mathbb{R}_{>0}$. Fix $\psi : \mathscr{O} \to \mathbb{R}$ and consider the fully non-linear PDE

$$\begin{cases} -\partial_t v(t,x) - G\big(x, \partial_x v(t,x), \partial_{xx} v(t,x)\big) = 0 & \text{on } [0,T) \times \mathscr{O}, \\ \qquad\qquad\qquad\qquad v(T,x) = \psi(x) & x \in \mathscr{O}, \end{cases} \tag{5}$$

where $G : \mathscr{O} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is defined by

$$G(x,p,q) := \sup_{(b^0,b^1,a^0,a^1) \in \Theta} \Big\{ (b^0 + b^1 x) p + \frac{1}{2} (a^0 + a^1 x^+) q \Big\}. \tag{6}$$

The function G satisfies the degenerate ellipticity condition and as Θ is compact, it is also continuous. Observe that the PDE defined in (5) can be seen as non-linear affine PDE, since for $\theta := (b^0, b^1, a^0, a^1)$ it is of the form

$$\begin{cases} -\partial_t v(t,x) - \sup_{\theta \in \Theta} \mathscr{L}^\theta v(t,x) = 0 & \text{on } [0,T) \times \mathscr{O}, \\ \qquad\qquad\qquad v(T,x) = \psi(x) & x \in \mathscr{O}. \end{cases}$$

$$-\partial_t v(t,x) - \sup_{\theta\in\Theta} \mathscr{L}^\theta v(t,x) = 0, \quad v(T,x) = \psi(x) \tag{7}$$

Finally, it will turn out that nonlinear affine processes correspond exactly to a family of semimartingales which generates the value function $v(t,x)$.

Theorem

Let $f \in C_b^{2,3}(\mathbb{R})$. Then,

$$v(t,x) := \mathcal{E}_x[f(X_t)], \quad x \in \mathbb{R},\ 0 \le t \le T$$

is the (unique) viscosity solution of the non-linear PDE in (7).

The proof is a suitable modfication of the non-linear Lévy setting. Uniqueness is more complicated due to the lack of the Lipschitz property.

Uniqueness I

The easy case:

Proposition

Assume that $\underline{a}^0 > 0$ and let $\mathscr{O} = \mathbb{R}$. Then, for any given continuous and bounded function $\psi : \mathbb{R} \to \mathbb{R}$, the non-linear PDE introduced in (7) admits at most one viscosity solution $v(t,x)$ on $[0,T] \times \mathbb{R}$ satisfying the terminal conditions

$$v(T,x) = \psi(x), \quad x \in \mathbb{R}.$$

Uniqueness II

Now we come to the more complicated case, the non-linear CIR case.

Proposition

Consider $x > 0$ and assume that $\underline{a}^0 = \bar{a}^0 = 0$ and that $\underline{b}^0 \geq {}^{\bar{a}^1}/_2 > 0$. Then for any $P \in \mathcal{A}(x, \Theta)$ it holds that

$$P(X_t > 0,\ 0 \leq t \leq T) = 1.$$

For the proof we adopt the method in Ilya Gikhman (2011). „A Short Remark on Feller's Square Root Condition“. In: **Available on SSRN** to our setting.

In this regard, let

$$h_\epsilon(x) = \begin{cases} \max\left(\frac{1}{\log(x)}, 1\right) & \epsilon = 0 \\ \max\left(\frac{1}{x^\epsilon}, 1\right) & \text{otherwise.} \end{cases} \tag{8}$$

Proposition

Assume that $\underline{a}^0 = \bar{a}^0 = 0$, $\underline{b}^0 \geq {}^{\bar{a}^1}/_2 > 0$, let $\mathscr{O} = \mathbb{R}_{>0}$ and consider a Lipschitz-continuous $\psi : \mathbb{R}_{>0} \to \mathbb{R}$. Then, (5)–(6) admits at most one viscosity solution $v(t,x)$ on $[0,T] \times \mathbb{R}_{>0}$ such that

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}_{>0}} \frac{|v(t,x)|}{1+x} < \infty. \tag{9}$$

Applications

Towards applications we provide a number of examples.

1. Explicit examples ?
2. Valuation of insurance claims
3. Robust pricing of options (under model risk).
4. Calibration I

Explicit examples?

- Towards our search for explicit formulae we tried the simplest case: the payoff $f(x) = e^x$.
- Consider for example the Vasiček-case and recall the generator

$$\mathscr{L}^\theta = (b_0 + b_1 x)\partial_x + \frac{1}{2} a_0^2 \partial_x x.$$

- If the payoff is monotone and convex, this carries over to v and the first and second derivative will be positive. Hence,

$$\sup_{\theta\Theta} \mathscr{L}^\theta = \begin{cases} (\bar{b}_0 + \bar{b}_1 x)\partial_x + \frac{1}{2}(\overline{a}_0)^2 \partial_x x & x \geq 0, \\ (\bar{b}_0 + \underline{b}_1 x)\partial_x + \frac{1}{2}(\overline{a}_0)^2 \partial_x x & x < 0. \end{cases}$$

which opens the door for fast numerical solutions.

The non-linear Vasicek-CIR model

$$G(x,p,q) := \sup_{(b^0,b^1,a^0,a^1)\in\Theta} \Big\{ (b^0 + b^1 x)p + \frac{1}{2}(a^0 + a^1 x^+)q \Big\}.$$

- If a_0 and a_1 are subject to parameter uncertainty we are in the non-linear Vasicek-CIR model. State space is $\mathbb{R}$!
- As long as $\underline{a}^0 > 0$ with $\mathscr{O} = \mathbb{R}$ niqueness for the non-linear PDE (5) follows.
- In this general case, there will be no explicit solutions and we have to rely on numerical techniques.

Robust Deep Hedging

- In the work Eva Lütkebohmert, Thorsten Schmidt, and Julian Sester (2021). „Robust deep hedging“. In: **arXiv preprint arXiv:2106.10024**, we generalize the affine setting and study a machine learning approach to hedging (deep hedging).
- A **generalized affine diffusion** is a continuous semimartingale X which is a unique strong solution of the stochastic differential equation (SDE)

$$dX_t = (b_0 + b_1 X_t)dt + (a_0 + a_1 X_t)^\gamma dW_t, \tag{10}$$

  $\gamma \in [1/2, 1]$.
- As above, we set

$$b(x) := \{b_0 + b_1 x : b_0, b_1 \in \Theta\}, \quad a(x) := \{(a_0 + a_1 x^+)^{2\gamma} : a_0, a_1, \gamma \in \Theta\}$$

- **nonlinear generalized affine process (NGA)** starting in $x \in E$ at time $t \in [0, T]$ is the family of all absolutely continuous semimartingale laws $\mathcal{A}(t, x, \Theta)$, such that for each $P \in \mathcal{A}(t, x, \Theta)$ giving rise to the differential characteristics (β^P, α) we have

$$\beta_s^P \in b(X_s), \qquad \alpha_s \in a(X_s)$$

  $dt \otimes dP$-almost surely on $(t, T] \times \Omega$ and $P(X_t = x) = 1$.

The non-linear Kolmogorov PDE

- The infinitesimal generator of the generalized affine process is given by

$$\mathscr{L}^{\theta} f(x) = (b_0 + b_1 x)\partial_x f(x) + \frac{1}{2}(a_0 + a_1 x^+)^{2\gamma} \partial_{xx} f(x),$$

  with $f \in C^2(\mathbb{R})$.
- The non-linear Kolmogorov equation reads as above,

$$\begin{cases} -\partial_t v(t,x) - \sup_{\theta \in \Theta} \mathscr{L}^{\theta} v(t,x) = 0 & \text{on } [0,T) \times \mathscr{O}, \\ v(T,x) = \psi(x) & x \in \mathscr{O}. \end{cases}$$

Robust Deep Hedging

- Now to the deep hedging: Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a non-constant function, called **activation function**.
- A (feed-forward) neural network with input dimension $d_{\text{in}} \in \mathbb{N}$, output dimension $d_{\text{out}} \in \mathbb{N}$, $l \in \mathbb{N}$ layers, and activation function φ is a function of the form

$$\begin{aligned} \mathbb{R}^{d_{\text{in}}} &\to \mathbb{R}^{d_{\text{out}}} \\ x &\mapsto A_l \circ \varphi \circ A_{l-1} \circ \cdots \circ \varphi \circ A_0(x), \end{aligned}$$

where $(A_i)_{i=0,\dots,l}$ are affine functions $A_i : \mathbb{R}^{h_i} \to \mathbb{R}^{h_{i+1}}$, and where the activation function is applied component-wise.

- A **derivative** is given by its payoff Φ_T (which also could be path-dependent)
- Our aim is to determine hedging strategies $(h_t)_{0 \le t \le T}$ and cash positions $d \in \mathbb{R}$ such that the quadratic error is minimized

$$\min_{(h_t)_{0 \le t \le T}, d \in \mathbb{R}} \mathbb{E}^P \left[\left(d + \int_0^T h_t \, dX_t - \Phi_T \right)^2 \right] \tag{11}$$

  for all $P \in \mathcal{A}(0, x_0, \Theta)$.
- This formulation is a consequence of the considered model ambiguity, under which every measure from $\mathcal{A}(0, x_0, \Theta)$ is taken into account.

Numerical Experiments

- To gather some experience we provide a number of numerical experiments in the paper.
- Consider a nonlinear generalized affine process with parameters specified through

$$\begin{aligned} x_0 &= 10 \\ a_0 &\in [0.3, 0.7], \quad a_1 \in [0.4, 0.6], \\ b_0 &\in [-0.2, 0.2], \quad b_1 \in [-0.1, 0.1], \\ \gamma &\in [0.5, 1.5]. \end{aligned} \tag{12}$$

Algorithm 1: Computation of Optimal Hedging Strategies

for $\text{iter} = 1, \ldots, N_{\text{iter}}$ **do**

  for $b = 1, \ldots, B$ **do**

    Generate paths of the generalized affine process:

    $X_0^b := x_0$, $\Delta t_i := t_{i+1} - t_i$

    for $i = 0, \ldots, n-1$ **do**

      Generate $\Delta W_i \sim N(0, \Delta t_i)$;

      Generate $\gamma^{(i)} \sim U\left([\underline{\gamma}, \overline{\gamma}]\right)$, $a_0^{(i)} \sim U\left([\underline{a_0}, \overline{a_0}]\right)$, $a_1^{(i)} \sim U\left([\underline{a_1}, \overline{a_1}]\right)$, $b_0^{(i)} \sim U\left([\underline{b_0}, \overline{b_0}]\right)$, $b_1^{(i)} \sim U\left([\underline{b_1}, \overline{b_1}]\right)$;

      set $X_{i+1}^b := X_i^b + (b_0^{(i)} + b_1^{(i)} X_i^b)\Delta t_i + (a_0^{(i)} + a_1^{(i)} X_i^+)^{\gamma^{(i)}} \Delta W_i$

    end

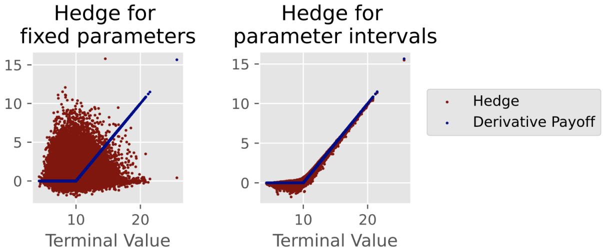
  end

  Apply stochastic gradient descent to minimize the loss

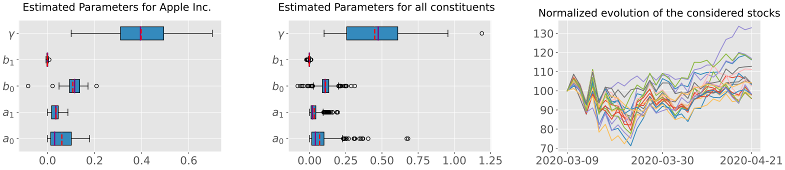
$$\sum_{b=1}^{B} \left(d + \sum_{i=0}^{n-1} h(t_i, X_i^b)(X_{i+1}^b - X_i^b) - \Phi\left((X_i^b)_{i=1,\ldots,n} \right) \right)^2$$

  w.r.t. the parameters of h and w.r.t. d

end



- Hedging of an at-the-money call option.
- left: The optimal hedge for fixed parameters $a_0 = 0.5$, $a_1 = 0.5$, $b_0 = 0$, $b_1 = 0$, $\gamma = 1$ evaluated on $50,000$ paths.
- right: The robust deep hedge evaluated on $50,000$ paths.



- Left: The maximum-likelihood-estimated parameters of a generalized affine process when assuming the stock of Apple Inc. is modelled through a generalized affine process. The estimations are performed every 100 days for a lookback window of 250 days.
- Middle: The maximum-likelihood-estimated parameters of all considered 20 constituents of the $S\&P$ 500.
- Right: The normalized (to initial value 100) evolution of the considered 20 constituents of the $S\&P$ 500-index in the considered time period from 09 March 2020 until 21 April 2020.

Parameters	fixed	robust	a_0 fixed	a_1 fixed	b_0 fixed	b_1 fixed	γ fixed	Black–Scholes
mean	2.1328	1.6657	1.7426	1.7091	1.6632	1.5820	1.5742	3.2860
std. dev.	2.3723	1.5512	2.2838	1.6384	1.3461	1.2306	2.3638	2.3752
min.	0.0975	0.1493	0.0084	0.1194	0.3532	0.1160	0.0261	0.0247
max.	8.6534	6.4227	9.9224	6.5793	5.8297	4.9860	9.3054	9.7704

- Absolute values of hedging errors (in $\%$) for an Asian at-the money put option $\left(x_0 - \frac{1}{30}\sum_{t=1}^{30} X_t\right)^+$ of trained hedging strategies of the considered 20 constituents.
- Each column represents another trained strategy which considers either fixed parameters (from the most recent maximum-likelihood estimation), robust parameter intervals (determined by the most extreme maximum-likelihood estimations), or robust intervals except for a single parameter which is still fixed.
- The rightmost column shows the hedging error when assuming an underlying Black–Scholes model.

- Summarizing, the change to uncertainty might seem like a small step, but it is a change in paradigm:
- Completely different techniques have to be used
- Where do we get the uncertainty intervals from ? A still open question.
- We showed that recent deep hedging techniques can be applied in this setting as well and outperform classical strategies on some data examples.