Deep Hedging

From Trading under Convex Risk Measures to Stochastic Implied Vol

Dr. Hans Buehler

Advances in Mathematics of Randomness for Handling Risks in Finance and Insurance
CIRM, Marseille, France, 15-19 September 2025

http://deep-hedging.com

Motivation

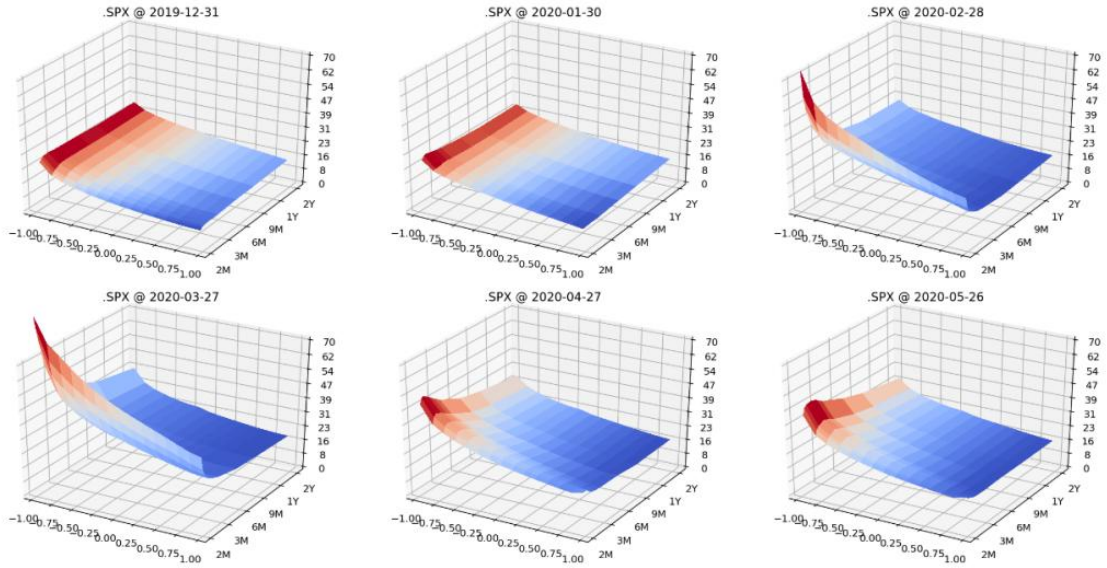
Deep Hedging

- Compute (minimal) price and an optimal risk-adjusted hedging strategy under the real measure P tradding *all* relevant hedging instruments.
- We'll essential do optimization under **convex risk measures** vs the full option surface.
- **Model-free** dynamic programming "reinforcement learning" problem.

Advanced topics:

- Removing the drift to avoid prop trading.
- Bellman version.

Market Simulation



- Our approach relies on training under full market data as opposed to classic made-up derivatives models with ~2 factors.
- Not sufficient daily market data → need to simulate markers using generative methods → different work stream e.g. [1] [2] …

[1] Multi-Asset Spot and Option Market Simulation, Wiese et al 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3980817
[2] Operator Deep Smoothing for Implied Volatility, Wiedeman et al, 2024 https://arxiv.org/abs/2406.11520

Framework

Environment [1]

- We operate in disctete time $t = 0, \dots, \infty$ under the statistical measure P.
- We observe "the market" s_t: time, prices, news, historical trading events etc and assume it generates our filtration.
- Any publicly observable market quantity X_t can be written for some measurable function X as $X_t = X(s_t)$.
- We also assume here that our trading causes no impact.
  - Our framework is easily extended to include impact [2].

[1] Deep Hedging, Buehler et all 2018, https://arxiv.org/pdf/1802.03042.pdf
[2] Lecture Notes Learning to Trade III https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4151043

Framework

Trading Instruments

- We assume that we are trading in a market with liquid **tradable instruments**.
  - These include primary assets such as spot, FX as well as derivatives such as options on equity, indices etc. We assume interest rates are deterministic and cashflows are discounted to today.(1)
- Denote by $H_t \in R^{n_t}$ the mid prices of tradable instruments available at time t, and by $\gamma_t \in R_{\geq 0}^{n_t}$ their bid/ask spreads.
- Our **trading cost** for $a \in R^{n_t}$ are given by a non-negative convex(2) c_t with

$$c_t(a) \downarrow |a|\gamma_t' \quad \text{for} \quad |a| \downarrow 0$$

- The latter condition means that we pay the bid/ask spread for small trades.

(1) If this is not the case, then optimal investment decisions of cash into the available interest rate instruments also need to be considered.
(2) Convexity excludes fixed fee cost which are, in fact, common.

Framework

Risk Limits and other Trading Restrictions

- Convex transaction cost allow defining convex limits to trading capacity by setting $c_t(\neg A) = \infty$ outside a convex set A.

- That means we can use trading cost to impose a wide variety of convex **trading restrictions** of the following type:
  - Maximum liquidity: $A = \{a: \text{askcapacity}^i \leq a^i \leq \text{bidcapacity}^i\}$
  - Total Vega Traded: $A = \{a: |\sum_i a_i \text{Vega}_i| \leq \text{Limit}\}$

- Trading restrictions which refer to the current portfolio are not always convex, as the available capacity in the market might not be sufficient to hedge all our risk:
  - Total Vega Held: $A = \{a: |\text{PortfolioVega} + \sum_i a_i \text{Vega}_i| \leq \text{Limit}\}$

Framework

Monetary Utilities

- We use *Optimized Certainty Equivalents, OCEs:* let u be a concave, increasing *utility function, then*

$$U[X] \coloneqq \sup_{c \in R} E[u(X + c) - c]$$

  is a *monetary utility*: increasing, concave and cash-invariant)$-U$ is a convex risk measure).
  - From an ML perspective, such measures lend themselves into batch-based optimization.

- Examples for risk aversion $\lambda > 0$:
  - Expectation: $u(x) = x$
  - Entropy: $u(x) \coloneqq (1 - \exp(-\lambda x))/\lambda$ in which case $U[X] = -\frac{1}{\lambda} \log E\left[e^{-\lambda X}\right]$.
  - CVaR: $u(x) = (1 + \lambda) \min\{0, x\}$

Vanilla Deep Hedging

- In Vanilla Deep Hedging [1] we are given a fixed horizon T to hedge until, and a fixed current portfolio Z with terminal value $Z_T \in R$.
- Let a with $a_t \equiv a(s_t)$ be a trading strategy. The **gains** of trading a are given as:

$$G(Z;a) \coloneqq Z_T + \sum_{t=0}^{T-1} a_t(H_T - H_t)' - c_t(a_t)$$

- Let $a \star H_T \coloneqq \sum_t a_t(H_T - H_t)'$ and $C_T(a) \coloneqq \sum_t c_t(a_t)$ such that

$$G(Z;a) \coloneqq Z_T + a \star H_T - C_T(a)$$

- We notice that we do not require any value for Z outside its terminal payoff in T.
- If the number of assets or their type (moneyness, time to expiry) vary per time step we need a “set invariant” network for representing a.

Vanilla Deep Hedging

- Let U be a monetary utility. Then the **Vanilla Deep Hedging** problem for selling a product Z is given as

$$U^*(Z) \coloneqq \sup_a : U(\, G(Z;a)\,)$$

$$G(Z;a) \coloneqq \Pi_T - Z_T + a \star H_T - C_T(a)$$

- The sup is taken over all admissible strategies.

Vanilla Deep Hedging

- Practically write a for an option with relative strike x and time to expiry τ as *set-invariant* **neural network** with network weights θ:

$$a_t^\theta := a^\theta(x, \tau; s_t).$$

- The problem

$$\max_\theta: U\left(\Pi_T - Z_T + a^\theta \star H_T - C_T\left(a^\theta\right)\right)$$

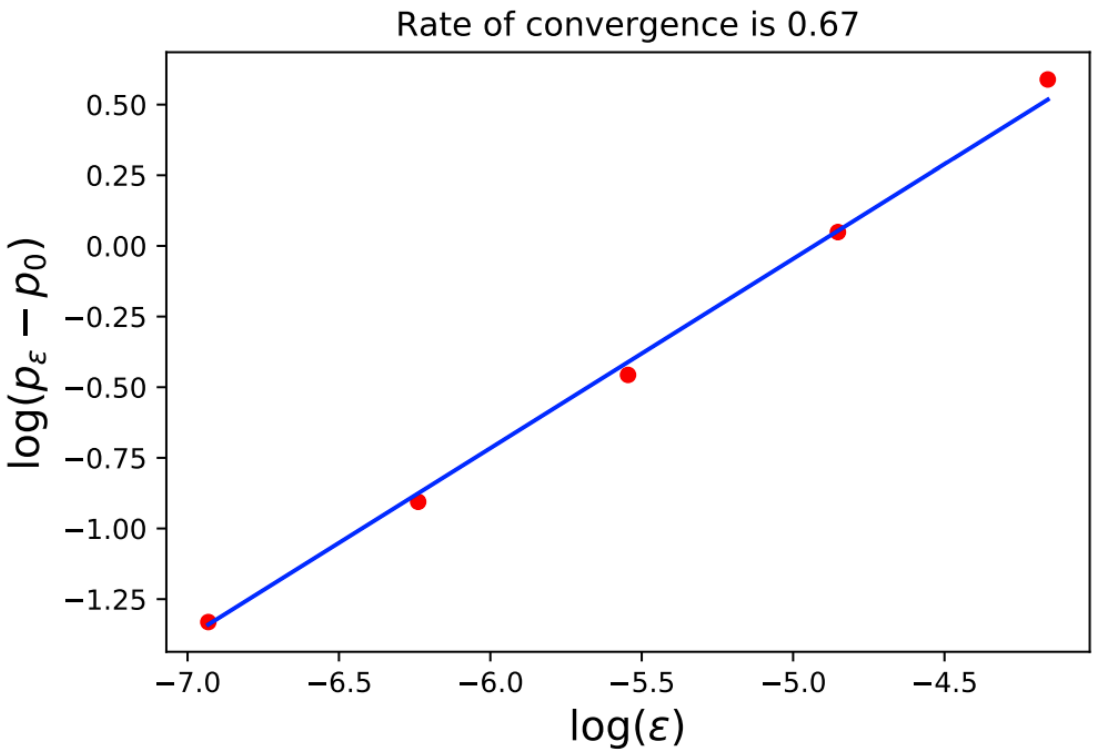
- is a then a classic Monte Carlo problem: if we use N paths it has the following form which is very accessible within modern ML frameworks such as *pyTorch* or *jax*:

$$\max_{\theta,y}: \frac{1}{N}\sum_{\omega=1}^{N} u\left(Z_T + a^\theta \star H_T - C_T\left(a^\theta\right) + y\right) - y$$

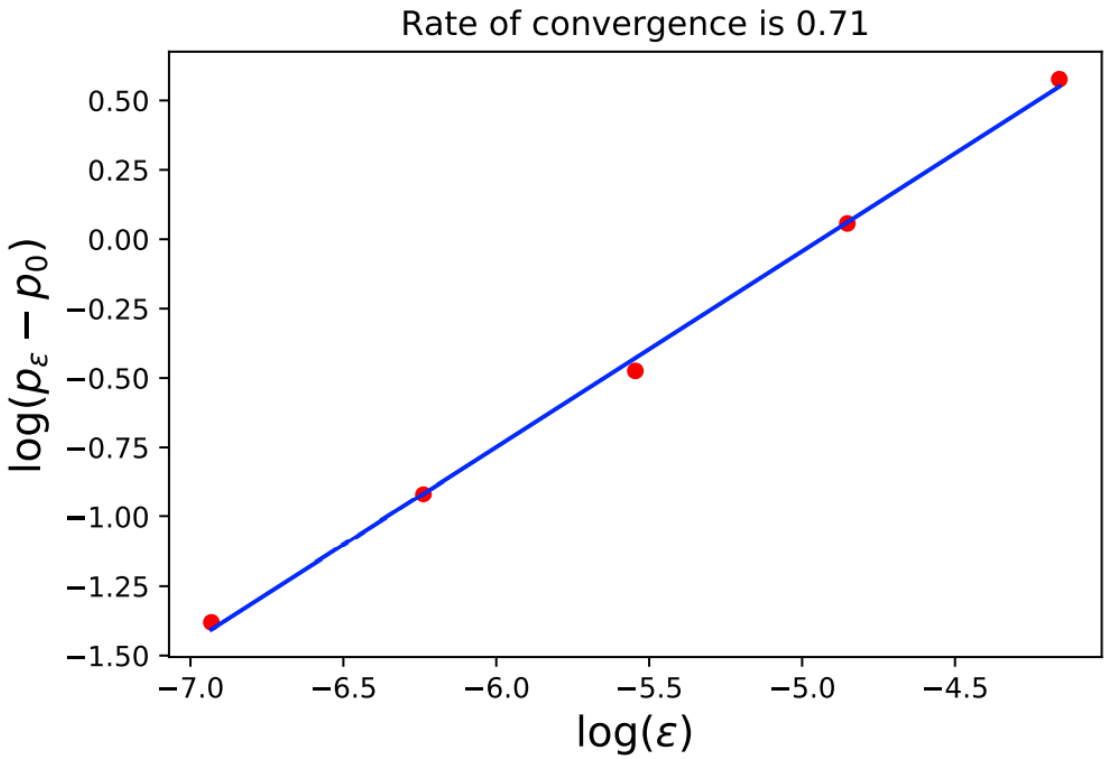
- We note that values for Z_T, $(H_T - H_t)$ etc can be pre-computed before running the optimization.
- In Reinforcement Learning this is called “Periodic Policy Search”.

Vanilla Deep Hedging

- Test first vs cases where we know or guess the theoretical answer [1]



Black-Scholes model price asymptotics.

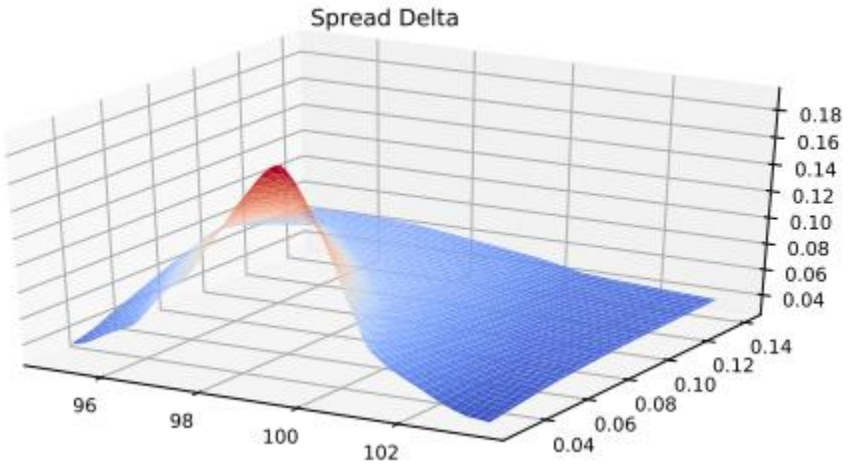


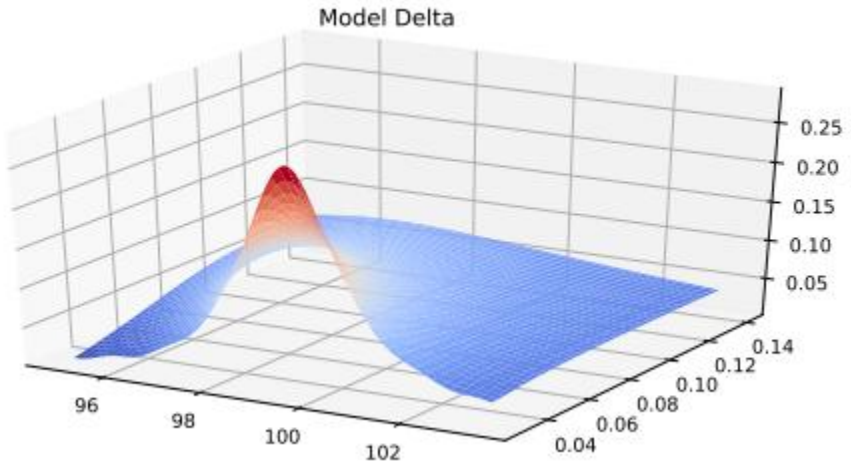
Heston model price asymptotics

[1] Deep Hedging, Buehler et all 2018, https://arxiv.org/pdf/1802.03042.pdf

Vanilla Deep Hedging

- Delta of a call spread in Black & Scholes [1]





[1] Deep Hedging, Buehler et all 2018, https://arxiv.org/pdf/1802.03042.pdf

Deep Hedging

Marginal Pricing:

- We wish to sell a derivative D to our client while we already have a position Z.
- The marginal price of selling D in the presence of Z is given as

$$p(Z) \coloneqq U^*(Z) - U^*(Z - D)$$

- It represents the **minimal price** we should charge to not be worse off vs our existing position Z.
- We note that cash-invariance satisfies the invariance condition

$$U^*\big(Z - D + p(Z)\big) = U^*(Z)$$

Statistical Arbitrage

Statistical Arbitrage

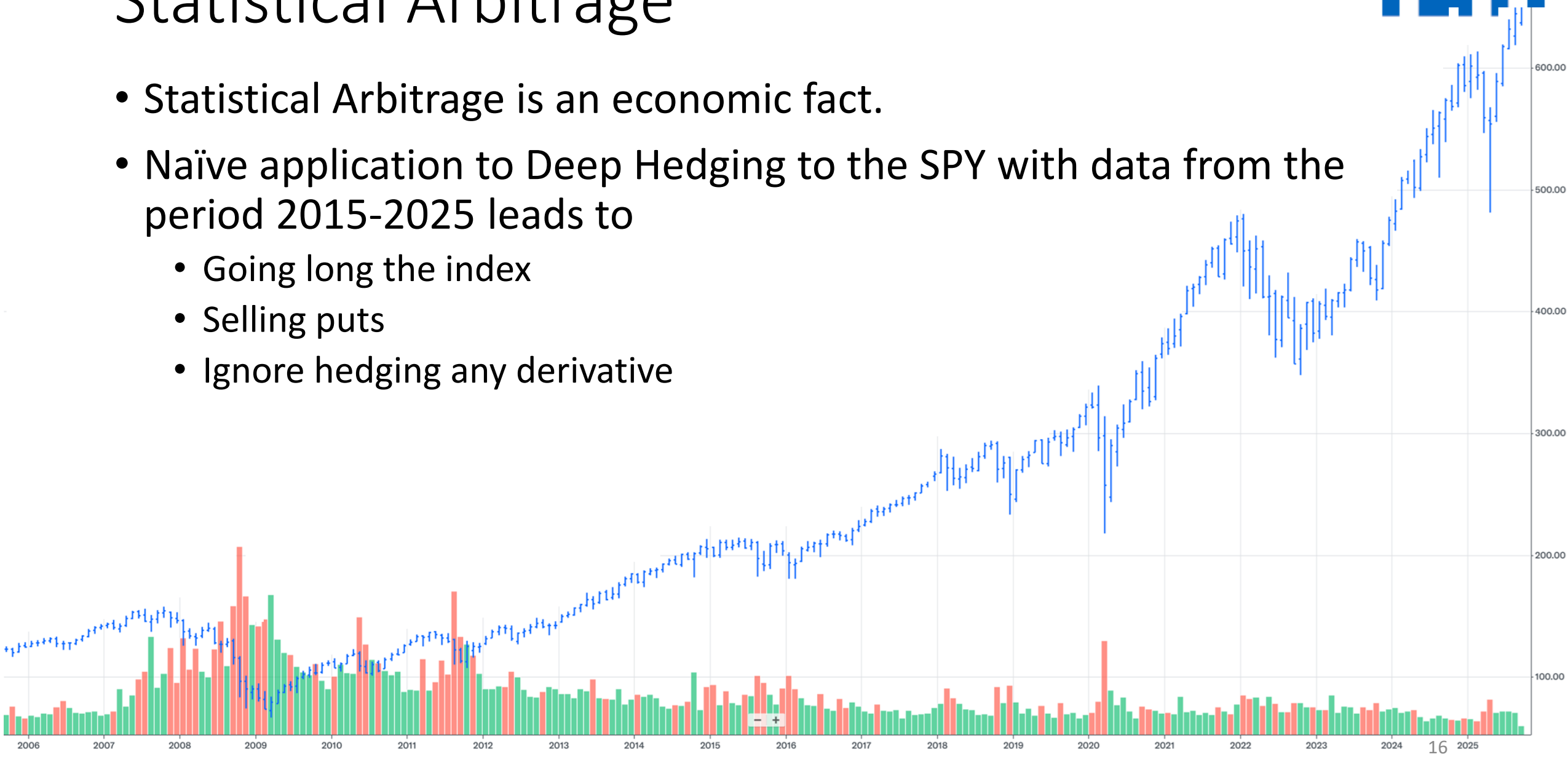
- We have earlier defined

$$U^*(Z) \coloneqq \sup_a : U\left(Z_T + a \star H_T - C_T(a)\right)$$

- What about $U^*(0)$ which represents the value of an empty portfolio?
- We say that the market exhibits **statistical arbitrage** if $U^*(0) > 0$.
- Happens naturally (contrary to static arbitrage).

Statistical Arbitrage

- Statistical Arbitrage is an economic fact.
- Naïve application to Deep Hedging to the SPY with data from the period 2015-2025 leads to
  - Going long the index
  - Selling puts
  - Ignore hedging any derivative



Statistical Arbitrage

Removing the Drift

- In classic portfolio optimization we only have “linear” assets.
- To remove the drift, we simply divide each asset by the mean return over the sample period(1)

$$d\tilde{X}_t^i := \frac{dX_t^i}{\frac{1}{T}(X_T - X_0)}$$

- For markets with complex assets removing the drift distorts the co-dependence of the instruments, e.g. stock and options thereon.
- Instead of changing the paths we aim now to *reweight* the observed paths such that the drift disappears – that means constructing a new equivalent measure Q.

Statistical Arbitrage

- Assume using the entropy $u(x) \coloneqq \frac{1-e^{-\lambda x}}{\lambda}$
- Cost zero for illustration.
- Under a measure Q define

$$U_Q(X) \coloneqq \sup_y : E_Q[u(X+y)-y] = -\frac{1}{\lambda}\log E_Q\left[e^{-\lambda X}\right]$$

- We wish to chose $Q \approx P$ such that

$$0 = \max_a : U_Q(\, a \star H_T \,)$$

Statistical Arbitrage

- Step 1: under P find the optimal strategy a^0 for the empty portfolio

$$a^0 := \text{argsup}: U_P(\, a \star H_T\,) = \text{argsup}_a : \frac{1}{\lambda} \log E_P\left[e^{-\lambda\{a \star H_T\}}\right]$$

- Step 2: define the measure

$$dQ := \frac{e^{-\lambda\{a^0 \star H_T\}}}{E_P\left[e^{-\lambda\{a^0 \star H_T\}}\right]} dP$$

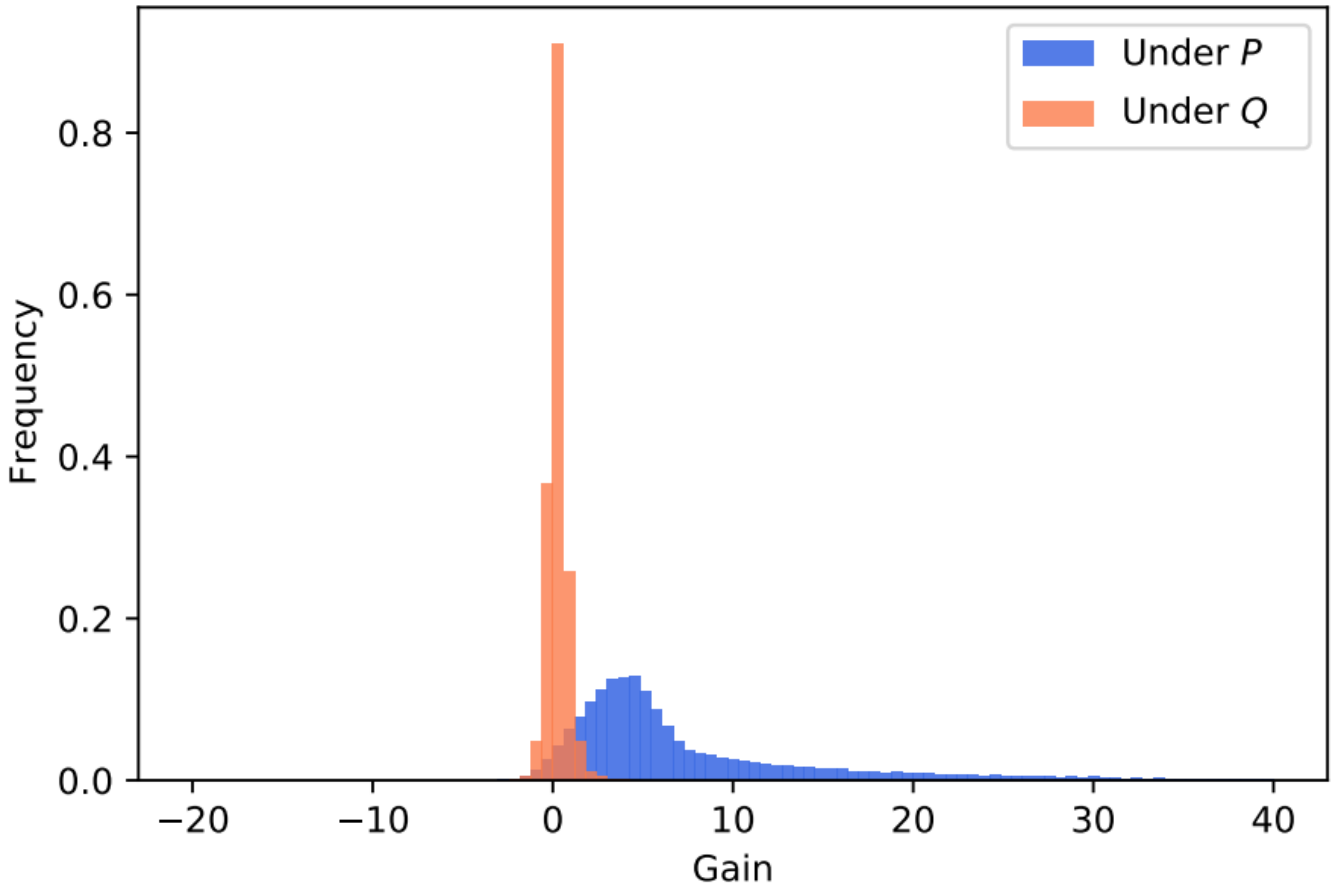
Statistical Arbitrage

- Step 3: under the monetary utility U_Q:

$$\begin{aligned}
&\max_a: U_Q(\, a \star H_T \,) \\
&= \max_a: \frac{1}{\lambda} \log E_P \left[\frac{e^{-\lambda\{(a+a^0)\star H_T\}}}{E_P\left[e^{-\lambda\{a^0 \star H_T\}}\right]} \right] \\
&\sim \max_a: \frac{1}{\lambda} \log E_P \left[e^{-\lambda\{(a+a^0)\star H_T\}} \right] \\
&= 0
\end{aligned}$$

  - Because a_0 was optimally chosen under P.
- This shows that the new optimal a is zero – in other words, the monetary utility U_Q is free of statistical arbitrage

Statistical Arbitrage



Numerical results of reducing the drift

[1] Learning Risk-Neutral Implied Volatility Dynamics, Buehler et al 2021, https://arxiv.org/pdf/2103.11948.pdf

Statistical Arbitrage

- The measure Q is one of many equivalent martingale measures.
- Our specific choice minimizes the entropy of Q with respect to P among all equivalent martingale measures

$$Q \mapsto E_P\left[\frac{dQ}{dP}\log\frac{dQ}{dP}\right]$$

and is called the *minimal entropy martingale measure* [1].

[1] Marco Frittelli. The minimal entropy martingale measure and the valuation problem in incomplete markets . In: Mathematical nance 10.1 (2000), pp. 3952.

Statistical Arbitrage

- In the case of the entropy without transaction cost we have the following intuitive result [1]: for a portfolio Z the optimal strategy a^* under U_P is given as

$$a^* := a^0 + \tilde{a}$$

- Where:
  - a^0 is the optimal "prop trading" strategy for the empty portfolio under P.
  - $\tilde{a}$ is the optimal "pure hedging" strategy under the risk-neutral Q.

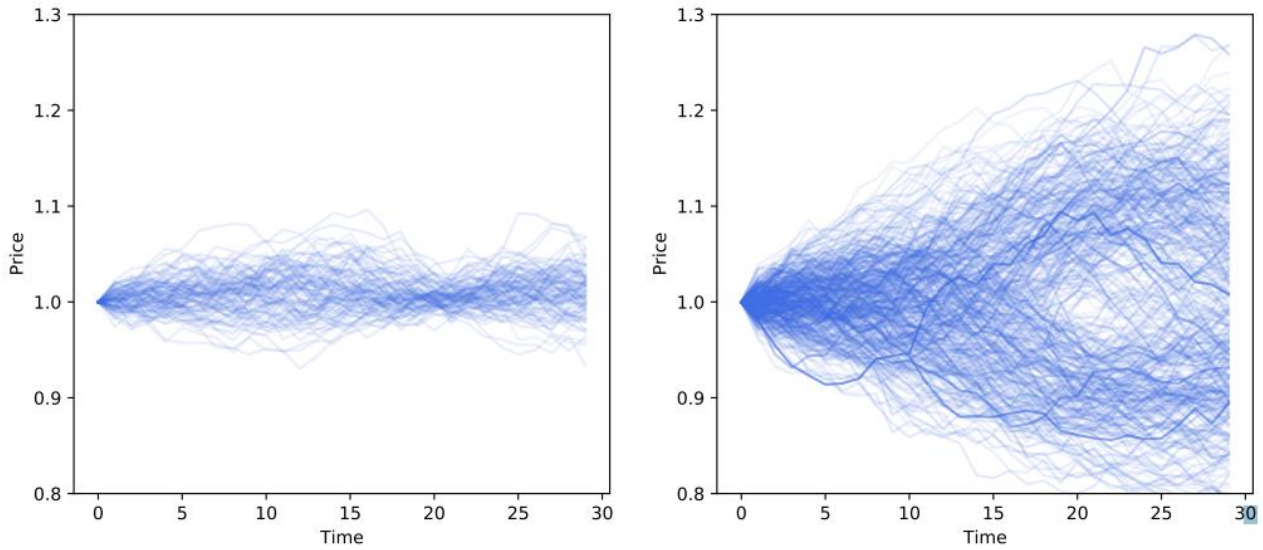
[1] Deep hedging: learning to remove the drift, Buehler et al 2022 https://www.risk.net/cutting-edge/banking/7932226/deep-hedging-learning-to-remove-the-drift and https://arxiv.org/abs/2111.07844

Statistical Arbitrage

- Fun fact: in discrete time, we can change also the *volatility* of a process by changing measure.
  - Experiment: market with 15% annual realized volatility. Option traded with 20% volatility. Statistical arbitrage is selling the option and delta hedge.
  - What happens when we change our measure:
    The measure will put more weight on paths with lower realized (discrete time) variance per path

Statistical Arbitrage

- Experiment [1]: market with 15% annual realized volatility. Option traded with 20% volatility. Statistical arbitrage is selling the option and delta hedge.
- The measure will put more weight on paths with lower realized (discrete time) variance per path



Left: paths given the highest 0.1% of probabilities under Q; right: lowest 0.1%

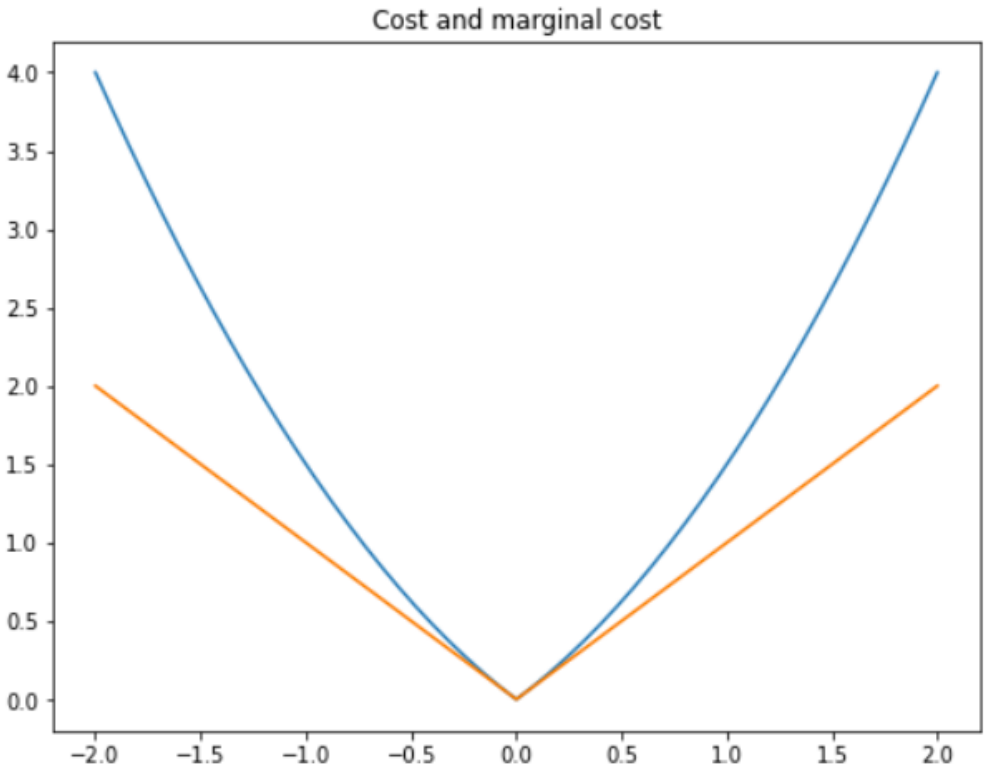
[1] Learning Risk-Neutral Implied Volatility Dynamics, Buehler et al 2021, https://arxiv.org/pdf/2103.11948.pdf

Statistical Arbitrage

Generalization to cost and arbitrary u

- Recall that c_t converges for marginal transaction sizes to its marginal cost, $c_t(a) \downarrow |a|\gamma_t'$ for $|a| \downarrow 0$.
- Define

$$M_T(a) \coloneqq \sum_t |a_t|\gamma_t{}'$$



Statistical Arbitrage

- Apply similar idea to before [1]: find a^0, y^0 as solution to

$$\sup_{a,y} E_P[\, u(\, a \star H_T - M_T(a) + y^0) - y^0 \,]$$

- Define the measure Q via

$$dQ := u'(\, a^0 \star H_T - M_T(a^0) + y^0 \,)\, dP$$

[1] Deep hedging: learning to remove the drift, Buehler et al 2022 https://www.risk.net/cutting-edge/banking/7932226/deep-hedging-learning-to-remove-the-drift and https://arxiv.org/abs/2111.07844

Statistical Arbitrage

- Under the new measure the expected returns of all instruments are within bid/ask spread in the following sense:
  - The measure Q is a ***near-martingale measure*** [1] in the sense that

$$bid_t^i = H_t^i - \gamma_t^i \leq E_Q\left[H_T^i | s_t\right] \leq H_t^i + \gamma_t^i = ask_t^i$$

  - Therefore, there is no statistical arbitrage under Q with full (or marginal) transaction cost for *any* OCE utility:

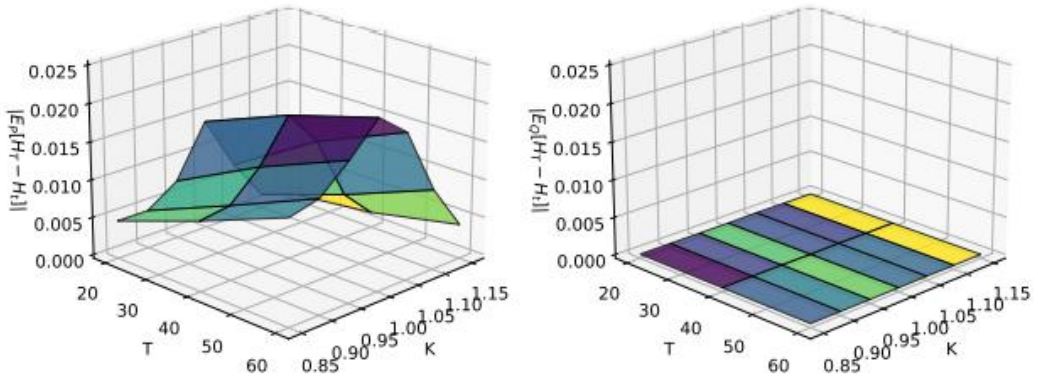
$$0 = \max_a : \tilde{U}_Q(a \star H_T - C_T(a))$$

  - The measure Q minimizes the $\tilde{u}$-divergence to P among all near-martingale measures [1].

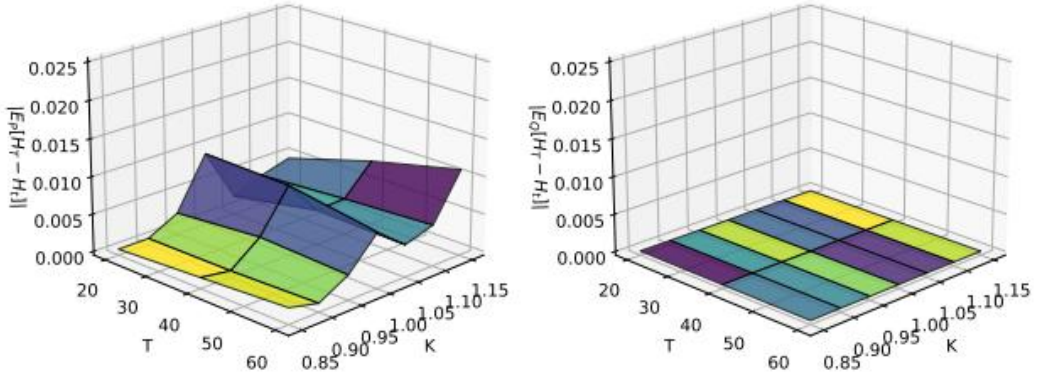
[1] Deep hedging: learning to remove the drift, Buehler et al 2022 https://www.risk.net/cutting-edge/banking/7932226/deep-hedging-learning-to-remove-the-drift and https://arxiv.org/abs/2111.07844

Statistical Arbitrage

Calls



Puts



Full market simulation results [1]: left are expected returns under P, right under Q under transaction cost

[1] Deep hedging: learning to remove the drift, Buehler et al 2022 https://www.risk.net/cutting-edge/banking/7932226/deep-hedging-learning-to-remove-the-drift and https://arxiv.org/abs/2111.07844

Stochastic Implied Volatility

- Assume we have built a market simulator for implied volatilities which does not have static arbitrage.
- We then removed the drift … such that the price processes become (near-)martingales
- We have created with machine learning a *stochastic implied volatility model* … a task not achieved through years of quantitative finance research !

Deep Bellman Hedging
(on-going research)

Deep Bellman Hedging [1]

- Bellman relationship for an optimal value V^*.

$$V^*\left(Z_t^{(t)}, s_t\right) \coloneqq \sup_a : U\left[\beta_t\, V^*\left(Z_{t+1}^{(t+1)}; s_{t+1}\right)\right] + R(Z_t, a, ; s_t) - C(a; s_t)$$

- Here
  - $Z_t^{(r)}$ is a linear representation at time t of the **portfolio at time** r. Its update rule is:

$$Z_t^{(t+1)} \coloneqq Z^{(t)} \oplus a\, H^{(t)\prime}$$

  - The representation is linear in the payoff space; $H^{(t)}$ denotes the hedges available at time t. Practically Z is represented as (term structure a matrices of) greeks.
  - $R(Z_t, a, s_t)$ are the **rewards** from holding Z_t (expiry cash flows, dividends, coupons)
  - $C(a; s_t) \coloneqq -aH_t^{(t)\prime} - c(a; s_t)$ are the **cost** of trading a.
  - $\beta_t < 1$ is a **discount factor**.

[1] Deep Bellman Hedging https://arxiv.org/abs/2207.00932

Deep Bellman Hedging

- Risk-adjusted Bellman equation:

$$V^*\left(Z_t^{(t)}, s_t\right) \coloneqq \sup_{\mathrm{a}} U\left[\beta_t\, V^*\left(Z_{t+1}^{(t)} \oplus a\, H_{t+1}^{(t)\prime}; s_{t+1}\right)\right] + R(Z_t, a, ; s_t) - C(a; s_t)$$

- Theorem
  - If U is an OCE monetary utility, if $\beta_t < 1 - \epsilon$, if rewards R are finite (e.g. if a_t is limited to a compact set) then the above has a unique solution V^*.

Deep Bellman Hedging

- Hard to learn since most cashflows (non-zero rewards) are very sparse.
- Alternative: assume B is a valuation model such as LV which captures cashflows and provides a baseline value. Then estimate

$$\tilde{V}(Z, s) \coloneqq V^*(Z, s) - B(Z, s)$$

- Because of the linearity of baseline models we can shift the change of value of the baseline model into rewards $\tilde{R}$ which now contain mark-to-market changes [1]:

- $\tilde{V}\left(Z_t^{(t)}, s_t\right) \coloneqq \sup_{\mathrm{a}} U\left[\beta_t \tilde{V}\left(Z_t^{(t)} \oplus a\, H_t^{(t)\prime}; s_{t+1}\right)\right] + \tilde{R}(Z_t, a, ; s_t) - C(a; s_t)$

[1] Deep Bellman Hedging https://arxiv.org/abs/2207.00932

Deep Bellman Hedging

- Implementation with fixed T: worked https://arxiv.org/abs/2207.07467 but is overkill

- Actual actor/critic: pretty unstable with results so far for only simple cases (such as portfolios of vanillas)

- More work to be done

Please ask questions