X-OT
ON VARIANTS OF THE OT PROBLEM AND UNDERSTANDING MODEL ROBUSTNESS
MATHEMATICAL SNAPSHOTS

JAN OBŁÓJ
Mathematical Institute
University of Oxford

CREST Doctoral Course
ENSAE, Paris
May 2025

UNIVERSITY OF OXFORD
Mathematical Institute

Oxford Mathematics
St John's College
Oxford

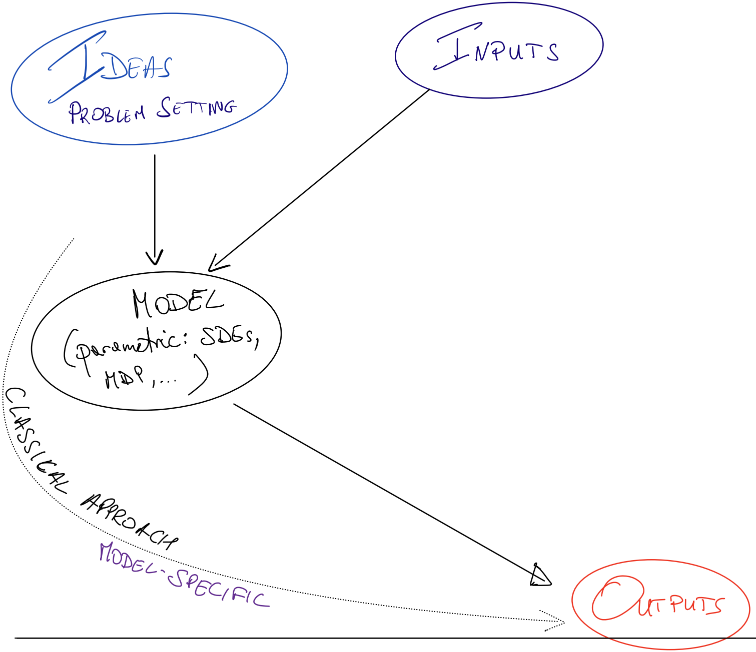
"ALL MODELS ARE WRONG BUT SOME ARE USEFUL"

G. BOX (1976)

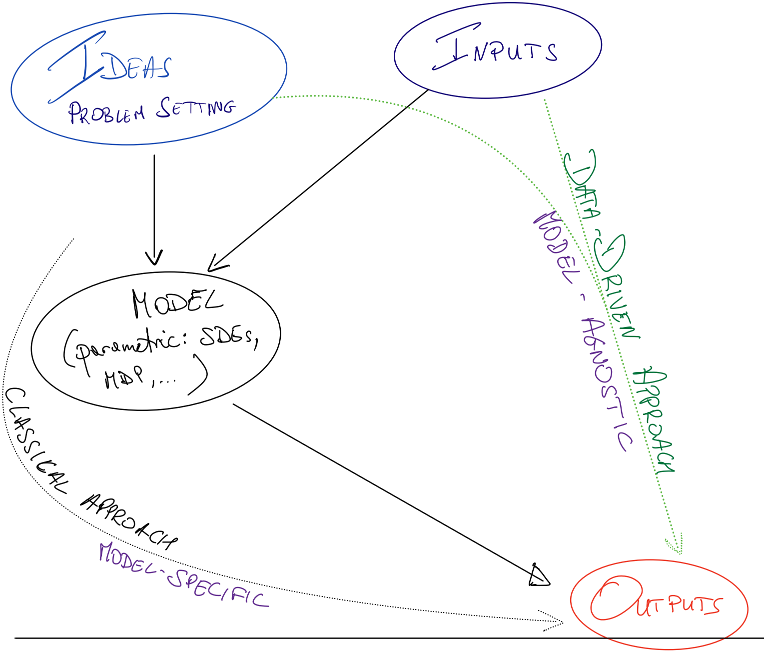
Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. Just as the ability to devise simple but evocative models is the signature of the great scientist so overelaboration and overparameterization is often the mark of mediocrity.

UNIVERSITY OF OXFORD

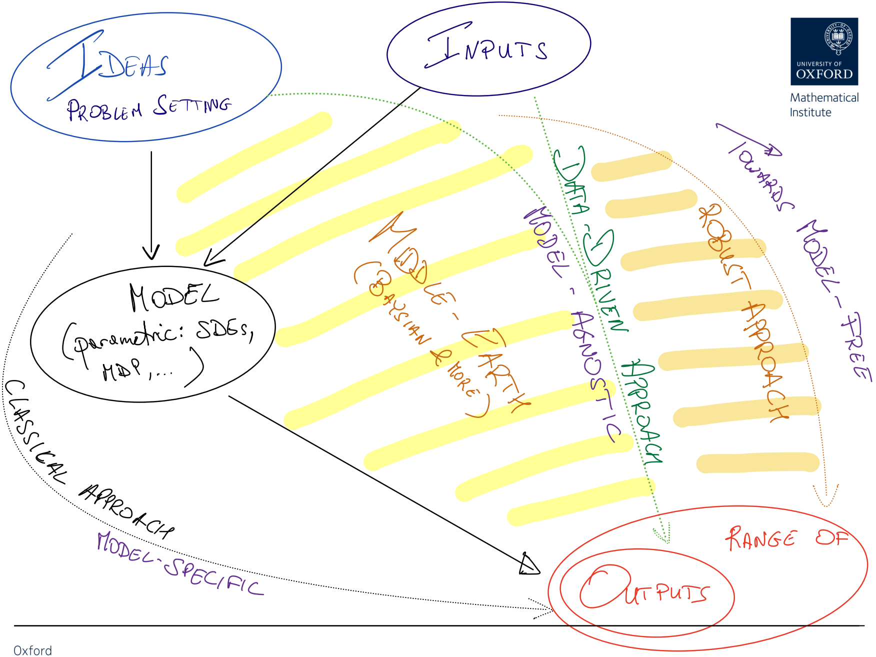
Mathematical Institute

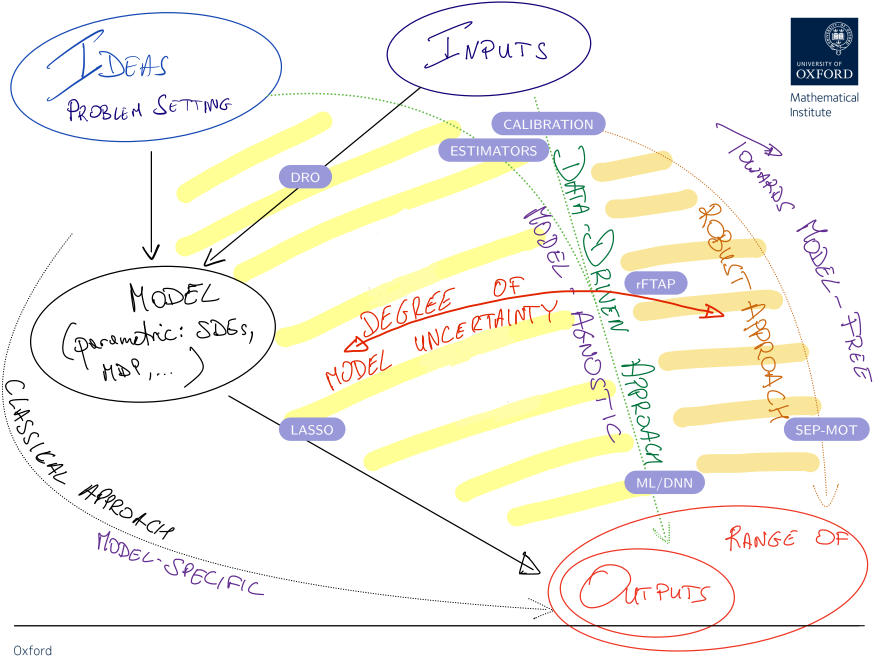


UNIVERSITY OF OXFORD
Mathematical Institute



UNIVERSITY OF OXFORD
Mathematical Institute




IDEAS
PROBLEM SETTING
INPUTS
MODEL (parametric: SDEs, MDP, ...)
CALIBRATION
ESTIMATORS
DRO
rFTAP
LASSO
SEP-MOT
ML/DNN
DEGREE OF MODEL UNCERTAINTY
DATA-DRIVEN APPROACH
MODEL-AGNOSTIC
ROBUST APPROACH
TOWARDS MODEL-FREE
CLASSICAL APPROACH
MODEL-SPECIFIC
RANGE OF
OUTPUTS

Mathematical
Institute

A VERY SHORT INTRODUCTION TO

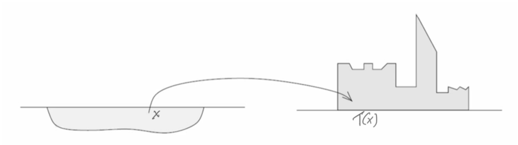
Optimal Transport

Optimal Transport – Monge's Problem

Mathematical Institute

Consider a state space $\mathcal{S}$ and two fixed distributions μ, ν.

Gaspard Monge (1781)



Monge's problem:

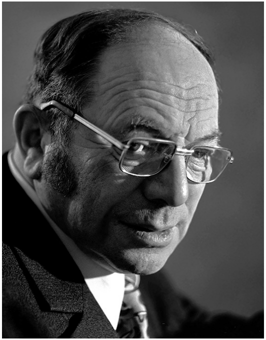
$$\inf\left\{ \int_{\mathcal{S}} \xi(x, T(x)) d\mu(x) \;\middle|\; \mu \circ T^{-1} = \nu \right\}$$

Optimal Transport – Relaxation and Duality

Relaxed problem:

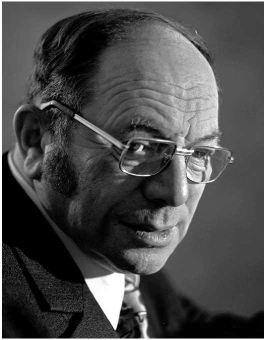
$$\inf\left\{\int_{\mathcal{S}\times\mathcal{S}} \xi(x,y)d\pi(x,y) \;\Big|\; \pi \in \mathrm{Cpl}(\mu,\nu)\right\},$$

where $\mathrm{Cpl}(\mu,\nu)$ are couplings – distributions on $\mathcal{S}^2$ with marginals μ,ν.



Leonid Kantorovich (1948)

Optimal Transport – Relaxation and Duality



Leonid Kantorovich (1948)

Relaxed problem:

$$\inf\Big\{\int_{\mathcal{S}\times\mathcal{S}} \xi(x,y)d\pi(x,y) \;\Big|\; \pi\in \mathrm{Cpl}(\mu,\nu)\Big\},$$

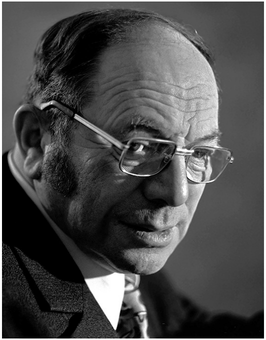
where $\mathrm{Cpl}(\mu,\nu)$ are couplings – distributions on $\mathcal{S}^2$ with marginals μ,ν.

Dual problem:

$$\sup\Big\{\int_{\mathcal{S}} \varphi(x)d\mu(x) + \int_{\mathcal{S}} \psi(y)d\nu(y)\Big\}$$

where $\varphi(x)+\psi(y)\leq \xi(x,y)$.

Optimal Transport – Relaxation and Duality



Leonid Kantorovich (1948)

Relaxed problem:

$$\inf\left\{\int_{\mathcal{S}\times\mathcal{S}} \xi(x,y)d\pi(x,y)\ \Big|\ \pi\in \mathrm{Cpl}(\mu,\nu)\right\},$$

where $\mathrm{Cpl}(\mu,\nu)$ are couplings – distributions on $\mathcal{S}^2$ with marginals μ,ν.

Dual problem:

$$\sup\left\{\int_{\mathcal{S}} \varphi(x)d\mu(x)+\int_{\mathcal{S}}\psi(y)d\nu(y)\right\}$$

where $\varphi(x)+\psi(y)\leq \xi(x,y)$.

Geometric insights: π^* optimal iff

$$\sum_{i=1}^{n}\xi(x_i,y_i)\leq \sum_{i=1}^{n}\xi(x_i,y_{i+1}),\quad \pi^*-a.s.$$

Wasserstein (Kantorovich-Rubinstein) distance

University of Oxford
Mathematical Institute

For $p \geq 1$, μ, ν p-ty measures on $(\mathcal{S}, \xi)$ with p^{th} moments, set

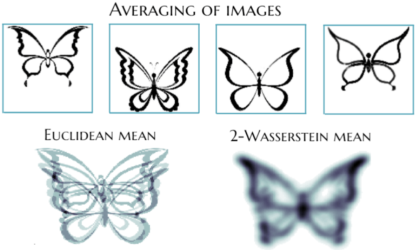
$$W_p(\mu,\nu) = \inf\left\{ \int_{\mathcal{S}\times\mathcal{S}} \xi(x,y)^p \, \pi(dx,dy) \colon \pi \in \mathrm{Cpl}(\mu,\nu) \right\}^{1/p},$$

where $\mathrm{Cpl}(\mu,\nu) = \{\pi : \pi(\cdot \times \mathcal{S}) = \mu \text{ and } \pi(\mathcal{S}\times \cdot) = \nu\}$.

An image

a vector of 0&1s for B&W pixels *OR* a probability measure on the square

Source: J. Ebert, V. Spokoiny, A. Suvorikova arXiv:1703.03658

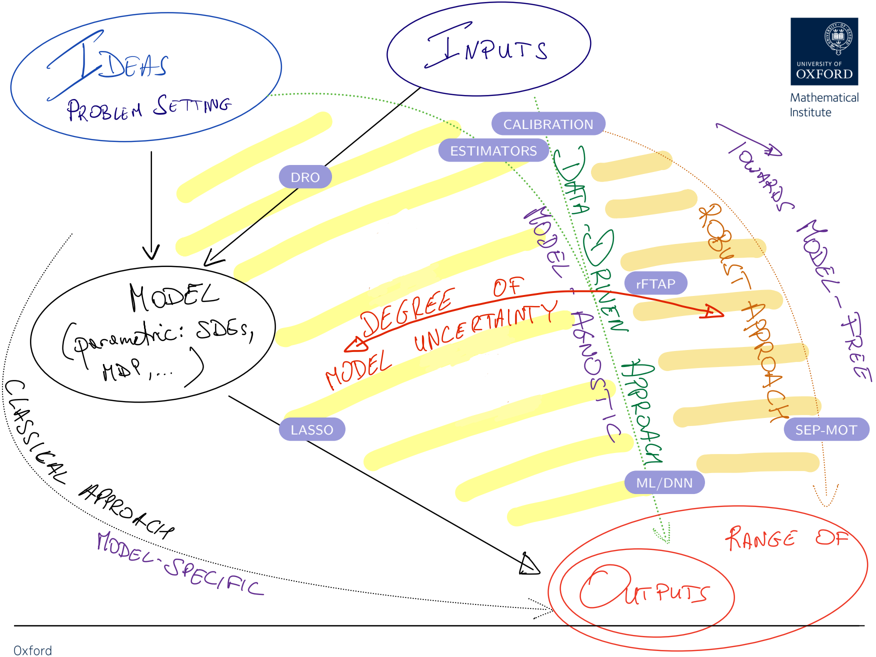


See also Michael Snow & Jan Van lent arXiv:1612.00181.

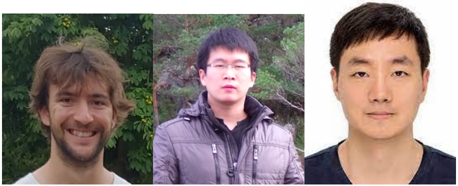
MNIST Digits: Wasserstein vs Euclidean mean

University of Oxford
Mathematical Institute

MNIST Digits: Wasserstein vs Euclidean mean


IDEAS
PROBLEM SETTING
INPUTS
MODEL
(parametric: SDEs, MDP, ...)
OUTPUTS
RANGE OF
CALIBRATION
ESTIMATORS
DRO
rFTAP
LASSO
SEP-MOT
ML/DNN
DATA-DRIVEN APPROACH
MODEL-AGNOSTIC
DEGREE OF MODEL UNCERTAINTY
ROBUST APPROACH
TOWARDS MODEL-FREE
CLASSICAL APPROACH
MODEL-SPECIFIC

Data: Market Prices of Options



based on joint works with Stephan Eckstein, Gaoyue Guo, Tongseok Lim
see *SIAM J. Financial Math. (2021), Ann. App. Probab. (2019)*.

An (idealised) case study: the MOT problem

- market data: prices of call options, $K > 0$,

$$\text{price } C(K) \quad \text{for a T-call with strike K: } (S_T - K)^+$$

An (idealised) case study: the MOT problem

- market data: prices of call options, $K > 0$,

$$\text{price } C(K) \quad \text{for a T-call with strike K: } (S_T - K)^+$$

- feasible pricing model $\leftrightsquigarrow$ probability measure $\mathbb{Q}$ s.t.

$$S \text{ is a } \mathbb{Q}\text{-martingale} \quad \text{and} \quad \mathbb{E}_{\mathbb{Q}}[(S_T - K)^+] = C(K), \; K \geq 0,$$

which is equivalent to

$$S \text{ is a } \mathbb{Q}\text{-martingale} \quad \text{and} \quad S_T \sim_{\mathbb{Q}} \nu, \text{ with } \nu(dK) = C''(dK).$$

An (idealised) case study: the MOT problem

- market data: prices of call options, $K > 0$,

$$\text{price } C(K) \quad \text{for a T-call with strike K: } (S_T - K)^+$$

- feasible pricing model $\leftrightsquigarrow$ probability measure $\mathbb{Q}$ s.t.

$$S \text{ is a } \mathbb{Q}\text{-martingale} \quad \text{and} \quad \mathbb{E}_{\mathbb{Q}}[(S_T - K)^+] = C(K), \; K \geq 0,$$

which is equivalent to

$$S \text{ is a } \mathbb{Q}\text{-martingale} \quad \text{and} \quad S_T \sim_{\mathbb{Q}} \nu, \text{ with } \nu(dK) = C''(dK).$$

- Robust pricing of an exotic option with payoff ξ

$$\rightsquigarrow \quad \sup \mathbb{E}_{\mathbb{Q}}[\xi(S_t : t \leq T)] \quad \text{over such } \mathbb{Q}s.$$

Robust hedging is its dual problem.

from MOT to SEP and back

UNIVERSITY OF OXFORD
Mathematical Institute

- Robust pricing of an exotic option with payoff ξ
  $\rightsquigarrow$ $\sup \mathbb{E}_{\mathbb{Q}}[\xi(S_t : t \leq T)]$ over $\mathbb{Q}$ s.t. S is a mg & $S_T \sim \nu$.
- S cont., so a time change of a BM: $S_t = B_{\tau_t}$, $t \geq 0$.
  Suppose $\mathbb{E}[\xi(S_t : t \leq T)] = \mathbb{E}[\xi(B_u : u \leq \tau)]$.
  This leads to

$$(OSEP) \qquad \sup_{\tau : B_\tau \sim \nu} \mathbb{E}[\xi(B_u : u \leq \tau_T)]$$

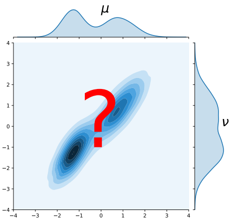
  which is an optimal transport problem *along* Brownian paths.
- Geometry of OT optimizers $\rightsquigarrow$ novel characterisation of (OSEP) in BEIGLBÖCK, COX, HUESMANN '17
- The dual leads to martingale inequalities.

The MOT problem

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$P(\mu,\nu) := \sup_{\mathbb{Q}\in\mathcal{M}(\mu,\nu)} \mathbb{E}_{\mathbb{Q}}\left[\xi(S_1,S_2)\right],$$

where $\mathcal{M}(\mu,\nu) := \left\{\mathbb{Q} : S_1 \sim \mu,\ S_2 \sim \nu \text{ and } \mathbb{E}_{\mathbb{Q}}\left[S_2 \middle| S_1\right] = S_1\right\}$.

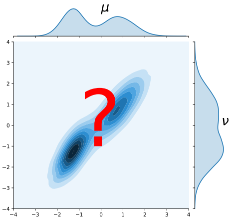


The MOT problem

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$P(\mu,\nu) := \sup_{\mathbb{Q}\in\mathcal{M}(\mu,\nu)} \mathbb{E}_{\mathbb{Q}}\left[\xi(S_1,S_2)\right],$$

where $\mathcal{M}(\mu,\nu) := \left\{\mathbb{Q} : S_1 \sim \mu,\ S_2 \sim \nu \text{ and } \mathbb{E}_{\mathbb{Q}}\left[S_2 \middle| S_1\right] = S_1\right\}$.



$$P(\mu,\nu) = \inf\left\{\int \phi d\mu + \int \psi d\nu : \phi(S_1) + \psi(S_2) + h(S_1)(S_2 - S_1) \geq \xi(S_1,S_2)\right\} =: D$$

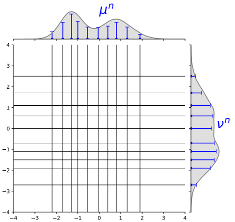
MOT Numerics: take I (Guo & O. '19)

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\sup_{\mathbb{Q} \in \mathcal{M}_\varepsilon(\mu^n, \nu^n)} \mathbb{E}_{\mathbb{Q}}\left[\xi(X, Y)\right],$$

where

$$\mathcal{M}_\varepsilon(\mu^n, \nu^n) := \left\{\mathbb{Q} : S_1 \sim \mu^n,\ S_2 \sim \nu^n \text{ and } \mathbb{E}_{\mathbb{Q}}\left[\left|\mathbb{E}_{\mathbb{Q}}\left[S_2 | S_1\right] - S_1\right|\right] \leq \varepsilon\right\}.$$



MOT Numerics: take I (Guo & O. '19)

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\sup_{\mathbb{Q}\in\mathcal{M}_\varepsilon(\mu^n,\nu^n)} \mathbb{E}_{\mathbb{Q}}\left[\xi(X,Y)\right],$$

where

$$\mathcal{M}_\varepsilon(\mu^n,\nu^n) := \left\{\mathbb{Q} : S_1 \sim \mu^n,\ S_2 \sim \nu^n \text{ and } \mathbb{E}_{\mathbb{Q}}\left[\left|\mathbb{E}_{\mathbb{Q}}\left[S_2 \middle| S_1\right] - S_1\right|\right] \leq \varepsilon\right\}.$$

- If $\mu = \sum_{i=1}^m \alpha_i \delta_{x_i}(dx)$ and $\nu = \sum_{j=1}^n \beta_j \delta_{y_j}(dy)$, then $P(\mu,\nu)$ is an LP problem;
- Discretisation $(\mu,\nu) \rightsquigarrow (\mu^n,\nu^n)$ typically does NOT preserve the convex order, see Alfonsi et al. (2017).
- Further, continuity of $(\mu,\nu) \to P(\mu,\nu)$ is unclear.
- $\rightsquigarrow$ we propose to look at a suitable relaxation!

MOT Numerics: take I

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\sup_{\mathbb{Q}\in\mathcal{M}_\varepsilon(\mu^n,\nu^n)} \mathbb{E}_{\mathbb{Q}}\big[\xi(X,Y)\big],$$

where

$$\mathcal{M}_\varepsilon(\mu^n,\nu^n) := \Big\{\mathbb{Q} : S_1 \sim \mu^n,\ S_2 \sim \nu^n \text{ and } \mathbb{E}_{\mathbb{Q}}\Big[\Big|\mathbb{E}_{\mathbb{Q}}\big[S_2\big|S_1\big] - S_1\Big|\Big] \le \varepsilon\Big\}.$$

Theorem

Assume ξ is L-Lipschitz. Let $(\mu^n, \nu^n)_{n\geq 1}$ be a sequence converging to (μ,ν): $r_n := \mathcal{W}(\mu^n,\mu) + \mathcal{W}(\nu^n,\nu) \to 0$. Then, $\mathcal{M}_{r_n}(\mu^n,\nu^n) \neq \emptyset$ and $\lim_{n\to\infty} P_{r_n}(\mu^n,\nu^n) = P(\mu,\nu)$.

MOT Numerics: take I

Given marginal laws $\mu, \nu \in$ on $\mathbb{R}^d$, consider

$$\sup_{\mathbb{Q}\in\mathcal{M}_\varepsilon(\mu^n,\nu^n)} \mathbb{E}_\mathbb{Q}\big[\xi(X,Y)\big],$$

where

$$\mathcal{M}_\varepsilon(\mu^n,\nu^n) := \Big\{\mathbb{Q} : S_1 \sim \mu^n,\ S_2 \sim \nu^n \text{ and } \mathbb{E}_\mathbb{Q}\Big[\Big|\mathbb{E}_\mathbb{Q}\big[S_2\big|S_1\big] - S_1\Big|\Big] \le \varepsilon\Big\}.$$

Theorem

Assume ξ is L-Lipschitz. Let $(\mu^n,\nu^n)_{n\ge1}$ be a sequence converging to (μ,ν): $r_n := \mathcal{W}(\mu^n,\mu) + \mathcal{W}(\nu^n,\nu) \to 0$. Then, $\mathcal{M}_{r_n}(\mu^n,\nu^n) \neq \emptyset$ and $\lim_{n\to\infty} P_{r_n}(\mu^n,\nu^n) = P(\mu,\nu)$. More precisely,

$$P(\mu,\nu) \quad \le \quad P_{\varepsilon_n}(\mu^n,\nu^n) \ + \ L\varepsilon_n \quad \le \quad P_{2\varepsilon_n}(\mu,\nu) \ + \ 2L\varepsilon_n$$

holds for any sequence $(\varepsilon_n) \subset \mathbb{R}_+$ converging to zero s.t. $\varepsilon_n \ge r_n$.

Further results

UNIVERSITY OF OXFORD
Mathematical Institute

- Strategy: discretise μ, e.g., generic method gives μ^n with

$$\mathcal{W}_1(\mu^n, \mu) \leq \sqrt{d}/n + \int_{|x|\geq n} |x|\mu(dx) =: \varepsilon_n.$$

  Solve LP for atomic μ^n. Take limits.
- Results/methods extend to T-periods.
- For $T = 2$, $d = 1$:
  - bespoke discretisation
  - convergence rates
  - entropic regularisation + iterative Bregman projection method $\rightsquigarrow$ efficient numerics.

MOT Numerics: take II (Eckstein & Kupper '19)

UNIVERSITY OF OXFORD
Mathematical Institute

- Numerics on the dual (superhedging) problem
- $\rightsquigarrow$ optimisation over functions
- $\rightsquigarrow$ Deep Neural Network implementation
  - hedging strategies $\in \mathcal{H}^m$ (a deep NN)
  - superhedging "$\leq$" replaced by a smooth penalisation w.r.t. a reference measure allowing for gradient descent algorithms:

$$D^m_{\theta,\gamma} = \inf_{h \in \mathcal{H}^m} \varphi(h) + \int \beta_\gamma(\xi - h) d\theta$$

  - Dual optimiser $\hat{h}$ allows to recover the primal one $\hat{\mathbb{Q}}$ via

$$\frac{d\hat{\mathbb{Q}}}{d\theta} = \beta'_\gamma(\xi - \hat{h})$$

is an optimiser of $P_{\theta,\gamma}$.

Market data: reality check

University of Oxford
Mathematical Institute

- For $d > 1$ we do NOT have full marginals.
  Only marginals of marginals (the MMOT problem):

$$S_1^i \sim \mu_i, \quad S_2^i \sim \nu_i$$

- Some interesting cases:
  - $d = 2$, $\xi(S) = (S_T^1 - \alpha S_T^2 - K)^+$ spread options
    $\rightsquigarrow$ *both LP and NN methods work*
  - $d = 30, 50, 100, \ldots, 500$ and $\xi(S) = \left(\sum_{i=1}^d \lambda_i S_T^i - K\right)^+$,
    i.e., calls/puts on an index

A Toy Example

Mathematical Institute

INPUTS:

- Data recorded on 16/11/2018:
  - Spot prices $F_0 = 140$, $A_0 = 194$ for Facebook and Apple
  - Call/Puts prices for Facebook and Apple maturing $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$
- Beliefs: bounds on correlation between Facebook and Apple

A Toy Example

University of Oxford
Mathematical Institute

INPUTS:

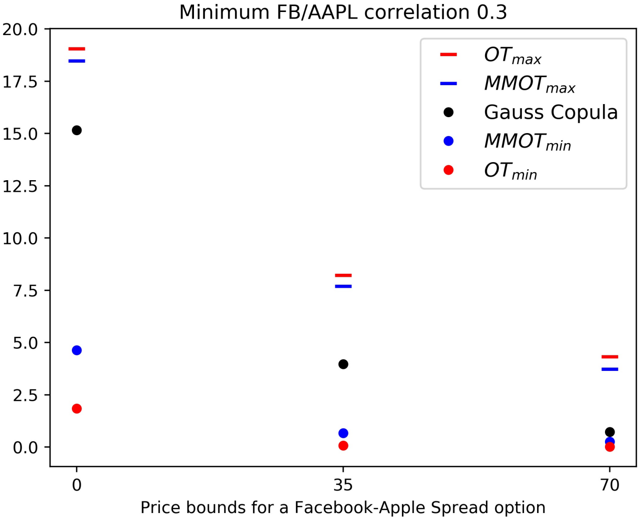
- Data recorded on 16/11/2018:
  - Spot prices $F_0 = 140$, $A_0 = 194$ for Facebook and Apple
  - Call/Puts prices for Facebook and Apple maturing $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$
- Beliefs: bounds on correlation between Facebook and Apple

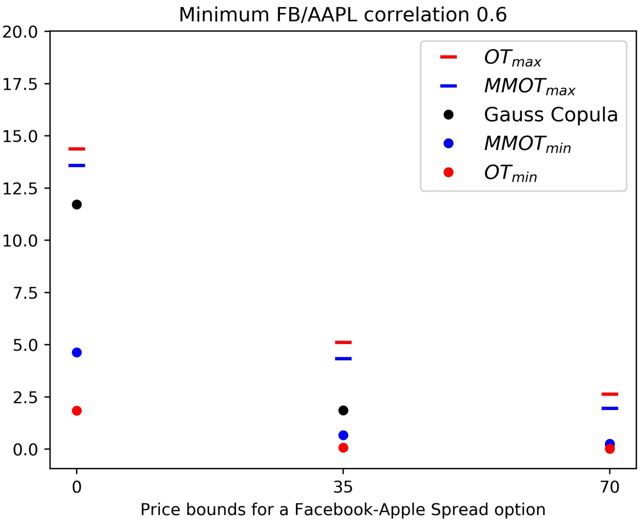
OUTPUTS:

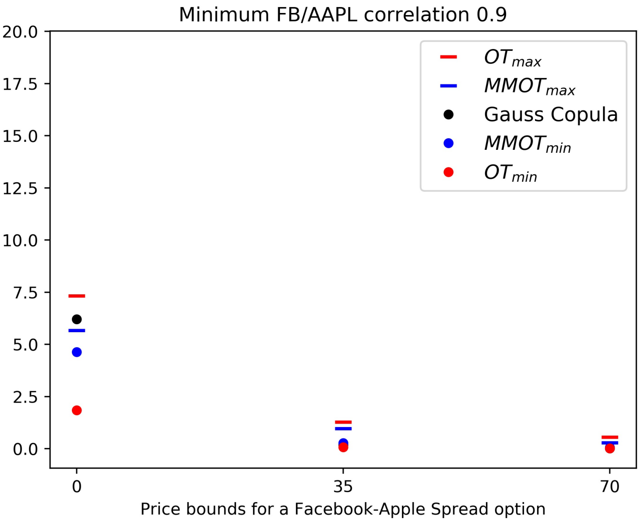
- Range of no-arbitrage prices for a spread option:

$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K \right)^+, \quad K = 0,\ 35,\ 70.$$

- Distribution of (F_{T_2}, A_{T_2}) for the minimiser/maximiser
- Robust hedging strategies

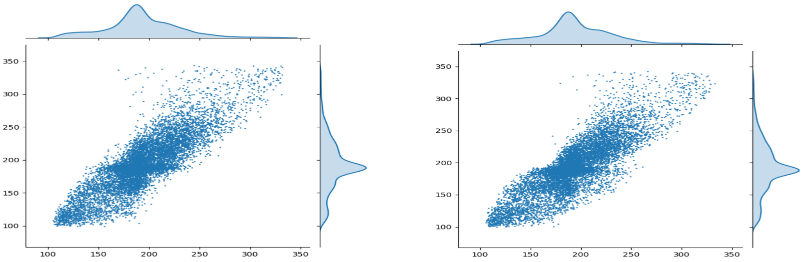

Minimum FB/AAPL correlation 0.3
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
35
70
Price bounds for a Facebook-Apple Spread option
OT_{max}
$MMOT_{max}$
Gauss Copula
$MMOT_{min}$
OT_{min}


Minimum FB/AAPL correlation 0.6
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
35
70
Price bounds for a Facebook-Apple Spread option
OT_{max}
$MMOT_{max}$
Gauss Copula
$MMOT_{min}$
OT_{min}


Minimum FB/AAPL correlation 0.9
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
35
70
Price bounds for a Facebook-Apple Spread option
OT_{max}
$MMOT_{max}$
Gauss Copula
$MMOT_{min}$
OT_{min}

Temporal evolution under Extreme models

UNIVERSITY OF OXFORD
Mathematical Institute

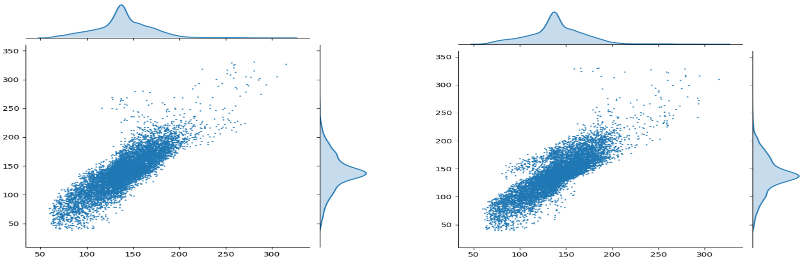


Joint distribution of (A_{T_1}, A_{T_2}), for the Minimiser and Maximiser $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$, $K = 35$ and $\rho \geq 0.6$ and

$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K\right)^+$$

Temporal evolution under Extreme models

UNIVERSITY OF OXFORD
Mathematical Institute

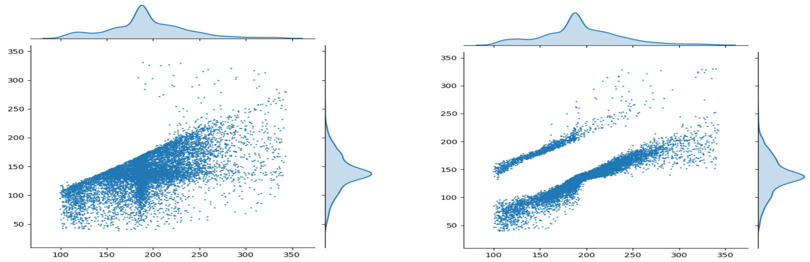


Joint distribution of (F_{T_1}, F_{T_2}), for the Minimiser and Maximiser $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$, $K = 35$ and $\rho \geq 0.6$ and

$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K \right)^+$$

Dependence Structure under Extreme models

Mathematical
Institute



Joint distribution of (A_{T_2}, F_{T_2}), $T_2 = 21/06/2019$, for the Minimiser and Maximiser for $K = 35$ and $\rho \geq 0.6$ and

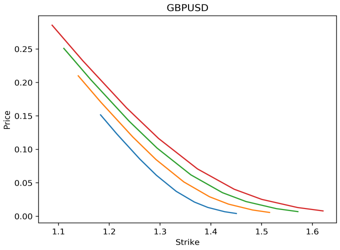
$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K\right)^+$$

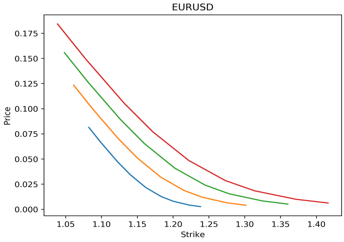
An FX Example

UNIVERSITY OF OXFORD
Mathematical Institute

Inputs:

- GBPUSD, EURUSD, GBPEUR data on 28/01/2019
- Spot + European calls for 0.5y, 1y, 1.5y and 2y for 10 strikes



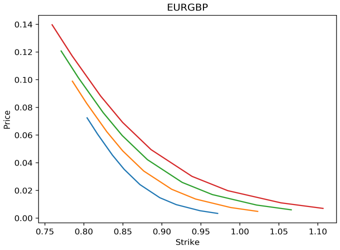


An FX Example

Mathematical Institute

Inputs:

- GBPUSD, EURUSD, GBPEUR data on 28/01/2019
- Spot + European calls for 0.5y, 1y, 1.5y and 2y for 10 strikes



$$\left(\frac{EUR}{GBP} - K\right)^+ = \left(\frac{EUR}{USD}\frac{USD}{GBP} - K\right)^+$$

An FX Example

Inputs:

- GBPUSD, EURUSD, GBPEUR data on 28/01/2019
- Spot + European calls for 0.5y, 1y, 1.5y and 2y for 10 strikes

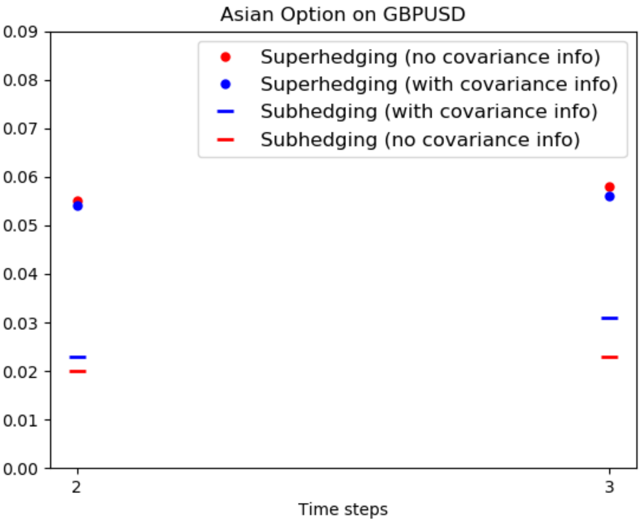
Outputs:

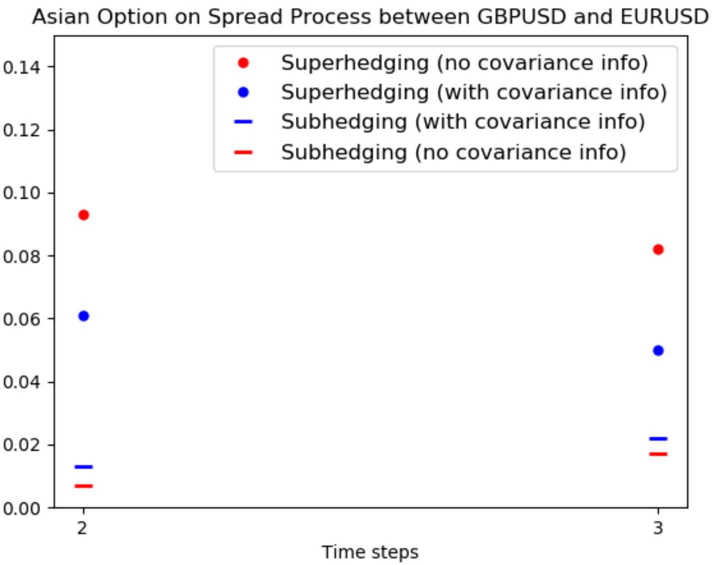
- Range of no-arbitrage prices for:

$$\left(\frac{1}{T}\sum_{t=1}^{T} X_t - X_0\right)^+ \quad \text{and} \quad \left(\sum_{t=1}^{T} X_t - \frac{X_0}{Y_0}\sum_{t=1}^{T} Y_t\right)^+$$

an Asian call on X=GBPUSD and
an Asian spread call on X=GBPUSD & Y=EURUSD

(with T in units of 0.5y).


Asian Option on GBPUSD
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Superhedging (no covariance info)
Superhedging (with covariance info)
Subhedging (with covariance info)
Subhedging (no covariance info)
2
3
Time steps



Asian Option on Spread Process between GBPUSD and EURUSD
Superhedging (no covariance info)
Superhedging (with covariance info)
Subhedging (with covariance info)
Subhedging (no covariance info)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
2
3
Time steps

UNIVERSITY OF OXFORD

A $d = 10$ Example (Pricilia '21)

Mathematical Institute

Inputs:

- Data recorded on 07/05/2021:
  - Spot prices for 10 constituents in the NYSE FAANG+ Index
  - Call/Puts prices for all 10 constituents with $T = 70$ days
  - historical time series $\rightsquigarrow$ 20^{th} quantile for pathwise covariances

A $d = 10$ Example (Pricilia '21)

UNIVERSITY OF OXFORD
Mathematical Institute

Inputs:

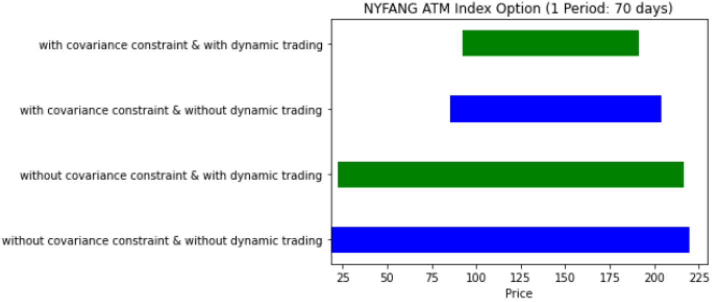
- Data recorded on 07/05/2021:
  - Spot prices for 10 constituents in the NYSE FAANG+ Index
  - Call/Puts prices for all 10 constituents with $T = 70$ days
  - historical time series $\rightsquigarrow 20^{th}$ quantile for pathwise covariances

Outputs:

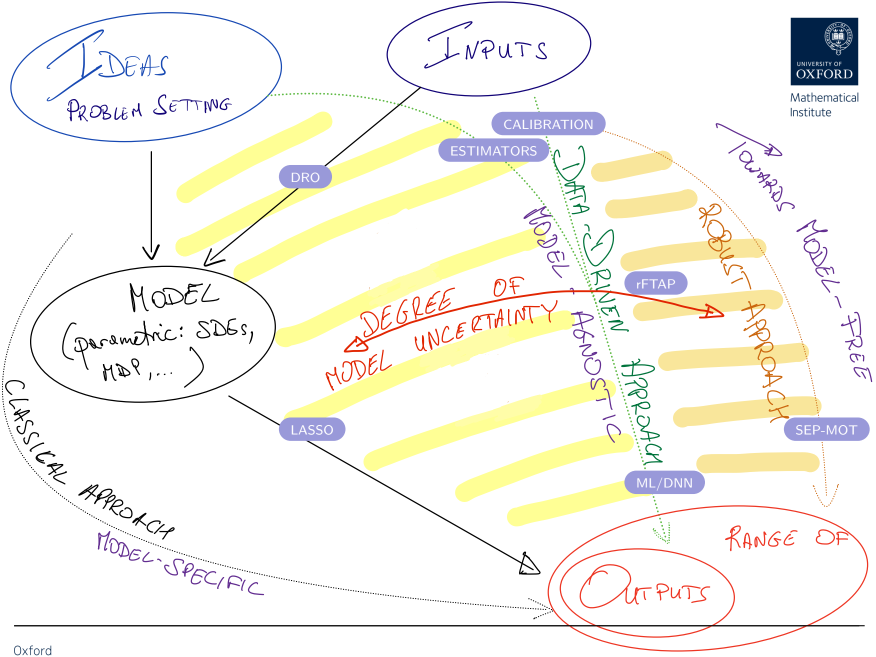
- Range of no-arbitrage prices for the ATM basket option:

$$\xi = (S_T - S_0)^+ , \quad S_t = \sum_{i=1}^{10} \lambda_i S_t^i$$

- Robust hedging strategies

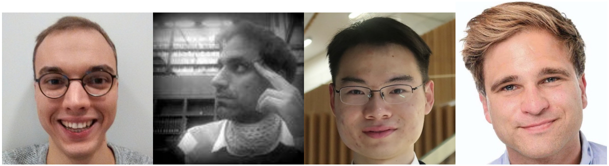


Range of no-arbitrage prices for ATM Call on NYSE FAANG+ after (Pricilia '21)



UNIVERSITY OF OXFORD
Mathematical Institute

Non-parametric sensitivities



works with Daniel Bartl, Samuel Drapeau, Yifan Jiang and Johannes Wiesel
see *Proc. R. Soc. Lond. A (2021), Math. Fin. (2021), arXiv:2408.17109*.

Consider the following optimisation problem

$$V = \inf_{a \in \mathcal{A}} \int_{\mathcal{S}} f(a, x) \mu(dx),$$

where $\mathcal{A}$ is the set of controls, $\mathcal{S}$ is the state space and μ is the model.

Consider the following optimisation problem

$$V = \inf_{a \in \mathcal{A}} \int_{\mathcal{S}} f(a, x)\mu(dx),$$

where $\mathcal{A}$ is the set of controls, $\mathcal{S}$ is the state space and μ is the model.

Examples:

- risk neutral pricing: $\mathbb{E}_{\mathbb{Q}}[f(S_T)]$,
- optimal investment: $\inf_{a \in \mathcal{A}} \mathbb{E}_{\mathbb{P}}[-U(x + \langle a, S_T - S_0 \rangle)]$,
- optimised certainty equivalents: $\inf_{a \in \mathbb{R}} \mathbb{E}_{\mathbb{P}}[a - U(X + a)]$
- marginal utility pricing (Davis' price)...

Consider the following optimisation problem

$$V = \inf_{a \in \mathcal{A}} \int_{\mathcal{S}} f(a,x)\mu(dx),$$

where $\mathcal{A}$ is the set of controls, $\mathcal{S}$ is the state space and μ is the model.

Examples:

- risk neutral pricing: $\mathbb{E}_{\mathbb{Q}}[f(S_T)]$,
- optimal investment: $\inf_{a\in\mathcal{A}} \mathbb{E}_{\mathbb{P}}[-U(x + \langle a, S_T - S_0\rangle)]$,
- optimised certainty equivalents: $\inf_{a\in\mathbb{R}} \mathbb{E}_{\mathbb{P}}[a - U(X+a)]$
- marginal utility pricing (Davis' price)...
- OLS regression: $\inf_{a\in\mathbb{R}^d} \frac{1}{N}\sum_{i=1}^N (y^i - \langle a, x^i\rangle)^2$,
- ML/NN: $\inf \frac{1}{N}\sum_{i=1}^N |y^i - ((A_2(\cdot)+b_2)\circ\sigma\circ(A_1(\cdot)+b_1))(x^i)|^p$ over $a = (A_1, A_2, b_1, b_2) \in \mathcal{A} = \mathbb{R}^{k\times d}\times\mathbb{R}^{d\times k}\times\mathbb{R}^k\times\mathbb{R}^d$, where $(x^i, y^i)_{i=1}^N$ is the training set.
-

Given our optimisation problem

$$V = \inf_{a \in \mathcal{A}} \int_{\mathcal{S}} f(a, x) \mu(dx),$$

we want to understand its dependence on the "model" μ.

We are interested in computing

$$\frac{\partial V}{\partial \mu}$$ – the uncertainty sensitivity of the problem

- parametric programming and statistical inference
  see Armacost & Fiacco '76 ... Bonnans & Shapiro '13;
- qualitative/quantitative stability in μ
  see Dupačová '90, Römisch '03
- robust optimisation
  see Bertsimas, Gupta & Kallus '18

Distributionally Robust Optimisation (DRO) considers

$$V(\delta) = \inf_{a \in \mathcal{A}} \sup_{\nu \in B_\delta(\mu)} \int_{\mathcal{S}} f(a, x)\nu(dx),$$

see Scarf '58, ... ,Rahimian & Mehrotra '19, where

$B_\delta(\mu)$ is a δ–neighbourhood of the model μ.

UNIVERSITY OF OXFORD
Mathematical Institute

Distributionally Robust Optimisation (DRO) considers

$$V(\delta) = \inf_{a \in \mathcal{A}} \sup_{\nu \in B_\delta(\mu)} \int_{\mathcal{S}} f(a, x)\nu(dx),$$

see Scarf '58, ... ,Rahimian & Mehrotra '19, where

$B_\delta(\mu)$ is a δ–neighbourhood of the model μ.

We propose to compute

$$\Upsilon := V'(0) = \lim_{\delta \searrow 0} \frac{V(\delta) - V(0)}{\delta} \quad \text{and} \quad \beth := \lim_{\delta \searrow 0} \frac{a^*(\delta) - a^*(0)}{\delta},$$

with $B^p_\delta(\mu)$ a p-Wasserstein ball around μ.

Υ *the sensitivity of the value w.r.t.* $\Upsilon\pi o\delta\varepsilon\gamma\mu\alpha$, *the Model.*

$\beth$ *the sensitivity of* בקרה, the control, w.r.t. the Model.

Uncertainty Sensitivity of DRO problems

UNIVERSITY OF OXFORD
Mathematical Institute

Recall our DRO problem (for simplicity $\mathcal{A} = \mathbb{R}^k$, $\mathcal{S} = \mathbb{R}^d$)

$$V(\delta) = \inf_{a \in \mathbb{R}^k} \sup_{\nu \in B_\delta^p(\mu)} \int_{\mathbb{R}^d} f(x, a)\, \nu(dx).$$

Uncertainty Sensitivity of DRO problems

Mathematical Institute

Recall our DRO problem (for simplicity $\mathcal{A} = \mathbb{R}^k$, $\mathcal{S} = \mathbb{R}^d$)

$$V(\delta) = \inf_{a \in \mathbb{R}^k} \sup_{\nu \in B_\delta^p(\mu)} \int_{\mathbb{R}^d} f(x, a)\, \nu(dx).$$

Theorem

For $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, and under suitable assumptions, we have

$$\Upsilon := V'(0) = \lim_{\delta \to 0} \frac{V(\delta) - V(0)}{\delta} = \inf_{a^* \in A^{\mathrm{opt}}(0)} \left(\int_{\mathbb{R}^d} |\nabla_x f(x, a^*)|^q \, \mu(dx) \right)^{1/q},$$

where $A^{\mathrm{opt}}(\delta)$ denotes the set of optimisers for $V(\delta)$.

Υ: uncertainty sensitivity of the value function

We can restate the result as

$$\inf_{a\in\mathbb{R}^k} \sup_{\nu\in B_\delta(\mu)} \int_{\mathbb{R}^d} f(x,a)\,\nu(dx) \approx \inf_{a\in\mathbb{R}^k} \int_{\mathbb{R}^d} f(x,a)\,\mu(dx) + \Upsilon\delta + o(\delta)$$

where

$$\Upsilon = \inf_{a^*\in A^{\mathrm{opt}}(0)} \left(\int_{\mathbb{R}^d} |\nabla_x f(x,a^*)|^q\,\mu(dx) \right)^{1/q}.$$

- extends to general semi-norms;
- extends to sensitivity at a fixed $\delta > 0$: $V'(\delta+)$;
- extends to DRO problems with linear constraints, e.g., martingale;
- no first order loss from using $a^*(0)$ instead of $a^*(\delta)$.

Sketch of the proof (1)

Mathematical Institute

Sensitivity of the value function: "$\leq$"

$$V(\delta) - V(0) \leq \sup_{\pi \in C_\delta(\mu)} \int f(y, a^*) - f(x, a^*)\, \pi(dx, dy)$$

$$= \sup_{\pi \in C_\delta(\mu)} \int \int_0^1 \langle \nabla_x f(x + t(y-x), a^*), (y-x) \rangle \, dt\, \pi(dx, dy)$$

$$\leq \delta \sup_{\pi \in C_\delta(\mu)} \int_0^1 \left(\int |\nabla_x f(x + t(y-x), a^*)|^q \pi(dx, dy) \right)^{1/q} dt.$$

$+$ growth conditions $+$ DCT.

Sketch of the proof (2)

Mathematical
Institute

Sensitivity of the value function: "$\geq$"

$$T(x) := \frac{\nabla_x f(x, a^*)}{|\nabla_x f(x, a^*)|^{2-q}} \left(\int |\nabla_x f(z, a^*)|^q \, \mu(dz) \right)^{1/q-1}$$

$$\pi^\delta := [x \mapsto (x, x + \delta T(x))]_\# \mu \in C_\delta(\mu)$$

We can use π^δ to get a lower bound:

$$\begin{aligned} \frac{V(\delta) - V(0)}{\delta} &\geq \frac{1}{\delta} \int f(x + \delta T(x), a^\delta) - f(x, a^\delta) \, \mu(dx) \\ &= \int \int_0^1 \langle \nabla_x f(x + t\delta T(x), a^\delta), T(x) \rangle \, dt \, \mu(dx) \\ &\xrightarrow{\delta \to 0} \int \langle \nabla_x f(x, a^*), T(x) \rangle \, \mu(dx) = \left(\int |\nabla_x f(x, a^*)|^q \, \mu(dx) \right)^{1/q}. \end{aligned}$$

Ex: Decision making & prefs representation

Mathematical Institute

Let X be agent's wealth/consumption. Savage '51, von Neuman & Morgenstern '53 give

$$\mathbb{P} \succeq \check{\mathbb{P}} \quad \Leftrightarrow \quad \mathbb{E}_{\mathbb{P}}[u(X)] \geq \mathbb{E}_{\check{\mathbb{P}}}[u(X)].$$

Ex: Decision making & prefs representation

Mathematical Institute

Let X be agent's wealth/consumption. Savage '51, von Neuman & Morgenstern '53 give

$$\mathbb{P} \succeq \check{\mathbb{P}} \quad \Leftrightarrow \quad \mathbb{E}_{\mathbb{P}}[u(X)] \geq \mathbb{E}_{\check{\mathbb{P}}}[u(X)].$$

An ambiguity averse agent of Gilboa & Schmeidler '89, might instead consider

$$\mathbb{P} \succeq_{\rho} \check{\mathbb{P}} \quad \Leftrightarrow \quad \min_{\tilde{\mathbb{P}} \in B_\delta(\mathbb{P})} \mathbb{E}_{\tilde{\mathbb{P}}}[u(X)] \geq \min_{\tilde{\mathbb{P}} \in B_\delta(\check{\mathbb{P}})} \mathbb{E}_{\tilde{\mathbb{P}}}[u(X)].$$

for $B_\delta(\mathbb{P})$ a δ-ball around $\mathbb{P}$ in some metric ρ,

(also called *constraint preferences* by Hansen & Sargent '01).

Variational prefs: relative entropy vs Wasserstein

UNIVERSITY OF OXFORD
Mathematical Institute

The variational/constraint preferences with ρ-ball $B_\delta(\mathbb{P})$

$$\mathcal{U}(X) := \min_{\tilde{\mathbb{P}} \in B_\delta(\mathbb{P})} \mathbb{E}_{\tilde{\mathbb{P}}}[u(X)]$$

up to $o(\delta)$ are equivalent to:

$\rho =$ REL. ENTROPY

$$\mathcal{U}(X) \approx \mathbb{E}_{\mathbb{P}}[u(X))] - \delta\sqrt{2\mathrm{Var}_{\mathbb{P}}(u(X))}$$

(cf. Lam '16)

$\rho = W_2$ WASSERSTEIN

$$\mathcal{U}(X) \approx \mathbb{E}_{\mathbb{P}}[u(X))] - \delta\sqrt{\mathbb{E}_{\mathbb{P}}\left[|u'(X)|^2\right]}$$

(cf. our Υ-sensitivity)

Ex: Robust call pricing (martingale constraint)

Mathematical Institute

We optimise over measures $\nu \in B_\delta(\mu)$ satisfying $\int x\,\nu(dx) = 1$.
A constrained version of our main results gives, for $p = 2$,

$$\Upsilon = \inf_{a^* \in \mathrm{A}^{\mathrm{opt}}(0)} \left(\int \left(\nabla_x f(x, a^*) - \int \nabla_x f(y, a^*)\,\mu(dy) \right)^2 \mu(dx) \right)^{1/2},$$

i.e., Υ is the standard deviation of $\nabla_x f(\cdot, a^*)$ under μ.

Ex: Robust call pricing (martingale constraint)

University of Oxford
Mathematical Institute

We optimise over measures $\nu \in B_\delta(\mu)$ satisfying $\int x\, \nu(dx) = 1$.
A constrained version of our main results gives, for $p = 2$,

$$\Upsilon = \inf_{a^* \in \mathrm{A}^{\mathrm{opt}}(0)} \left(\int \left(\nabla_x f(x, a^*) - \int \nabla_x f(y, a^*)\, \mu(dy) \right)^2 \mu(dx) \right)^{1/2},$$

i.e., Υ is the standard deviation of $\nabla_x f(\cdot, a^*)$ under μ.

Let $\mu \sim S_T/S_0$ with (S_t) from the BS(σ) model and

$$\mathcal{R}BS(\delta) = \sup_{\nu \in B_\delta(\mu)} \left\{ \int (S_0 x - K)^+ \nu(dx) \colon \int x\nu(dx) = 1 \right\}$$

so that $\mathcal{R}BS(0) = BSCall(S_0, K, \sigma)$. For $p = 2$ we find

$$\Upsilon(K) = S_0 \sqrt{\Phi(d_-)(1 - \Phi(d_-))}.$$

Robust call: numerics

Exact value $\mathcal{R}BS(\delta)$, first-order (FO) approximation and the model (BS) price.

BS model with $S_0 = T = 1$, $K = 1.2$, $r = q = 0$, $\sigma = 0.2$. $\delta = 0.05$

Robust call: classical vs robust

Mathematical Institute

Take $r = q = 0$, $T = 1$, $S_0 = 1$ and $\mu =$BS(σ) log-normal.

$$\mathcal{R}BS(\delta) = \sup_{\nu \in B_\delta(\mu)} \int_{\mathcal{S}} (s-K)^+ \nu(ds)$$

PARAMETRIC APPROACH

$$B_\delta(\mu) = \{\mathrm{BS}(\tilde{\sigma}) : |\tilde{\sigma} - \sigma| \leq \delta\}$$

Then

$$\mathcal{R}BS'(0) = \mathcal{V} = S_0 \phi(d_+).$$

NON-PARAMETRIC APPROACH

$$B_\delta(\mu) = \{\nu : W_2(\mu,\nu) \leq \delta\}$$

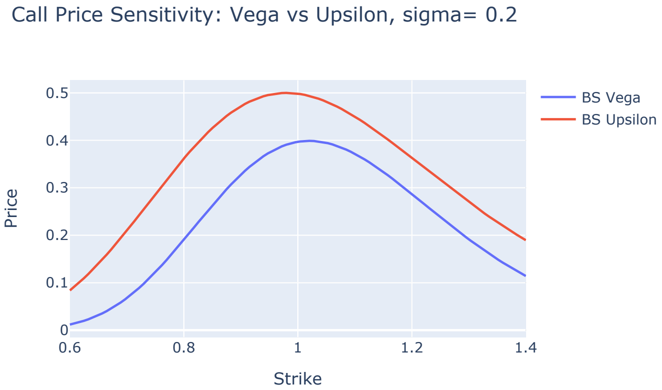
Then

$$\mathcal{R}BS'(0) = \Upsilon = S_0 \sqrt{\Phi(d_-)(1-\Phi(d_-))}$$

BS Call: Vega($\mathcal{V}$) vs Upsilon(Υ)

UNIVERSITY OF OXFORD
Mathematical Institute

Consider the simple example of a call option pricing.
Take $r = q = 0$, $T = 1$, $S_0 = 1$ and $\mu =$BS(σ) model.



Hedging: Δ–Vega vs Δ–Υ (WITH S. MOLINER '22)

University of Oxford
Mathematical Institute

Observe that $\Upsilon[aS_t + b] = 0$, *i.e., cash and stock carry no uncertainty.*

Comparison of two hedging approaches:

- Δ–Vega: at rebalancing buy/sell stock + ATM Call so that $\Delta = 0 = \mathcal{V}$
- Δ–Υ: at rebalancing buy/sell stock + ATM Call so that $\Delta = 0$ and Υ is minimized

	Δ	$\Delta + \mathcal{V}$	$\Delta + \Upsilon$
Mean	-0.015	-0.001	-0.002
Std	0.095	0.01	0.014
$V@R_{0.95}$	-0.190	-0.016	-0.028
$ES_{0.95}$	-0.296	-0.032	-0.045

Table 2: Risk measures with Bates Model $S_0 = T = 1$, $K = 1.05$, $v_0 = 0.04$, $\kappa = 1$, $\theta = 0.09$, $\sigma = 0.6$, $\rho = 0.5$, $\lambda = 15$, $\mu_J = 0$, $\sigma_J = 0.1$

Sensitivity of causal DRO

UNIVERSITY OF OXFORD
Mathematical Institute

Let $p > 1$ and $1/p + 1/q = 1$. Take $c(x, y) = \|\Delta x - \Delta y\|^p$ for $p > 1$, where

$$\Delta(x_1, x_2, \ldots, x_N) = (x_1, x_2 - x_1, \ldots, x_N - x_{N-1}).$$

Write $\mathbb{D} = (\mathbb{D}_1, \ldots, \mathbb{D}_N)$ as the pullback of ∇ under Δ, i.e., $\mathbb{D}_n = \sum_{l \geq n} \partial_l$.

Sensitivity of causal DRO

UNIVERSITY OF OXFORD
Mathematical Institute

Let $p > 1$ and $1/p + 1/q = 1$. Take $c(x, y) = \|\Delta x - \Delta y\|^p$ for $p > 1$, where

$$\Delta(x_1, x_2, \dots, x_N) = (x_1, x_2 - x_1, \dots, x_N - x_{N-1}).$$

Write $\mathbb{D} = (\mathbb{D}_1, \dots, \mathbb{D}_N)$ as the pullback of ∇ under Δ, i.e., $\mathbb{D}_n = \sum_{l \geq n} \partial_l$.

Under suitable assumptions, we have

$$\Upsilon := \lim_{\delta \to 0} \frac{v(\delta) - v(0)}{\delta} = L^* \left(\mathbb{E}_\mu \left[\sum_{n=1}^{N} |\mathbb{E}_\mu [\mathbb{D}_n f(X) | \mathcal{F}_n]|^q \right]^{1/q} \right) = L^*(\| {}^{\mathrm{o}}\mathbb{D} f \|_q).$$

Extensions

- Martingale constraint on the model.

$$\Upsilon_{\mathsf{Mart}} = L^*(\|\,{}^{\mathrm{o}}\mathbb{D}f - {}^{\mathrm{p}}\mathbb{D}f\|_2).$$

Extensions

- Martingale constraint on the model.

$$\Upsilon_{\text{Mart}} = L^*(\| {}^{\mathrm{o}}\mathbb{D}f - {}^{\mathrm{p}}\mathbb{D}f\|_2).$$

- Pass limit to the continuous time!
  - Hyperbolic scaling — drift uncertainty.

$$c(x,y) = \lim_{N\to\infty} N^{p-1} \sum_{n=1}^{N} |\Delta x_n - \Delta y_n|^p = \|\partial_t(x-y)\|^p.$$

A *pathwise* Malliavin derivative leads to $\Upsilon = L^*(\| {}^{\mathrm{o}}\mathbb{D}f\|_q)$.

Extensions

- Martingale constraint on the model.

$$\Upsilon_{\text{Mart}} = L^*(\| {}^{\mathrm{o}}\mathbb{D}f - {}^{\mathrm{p}}\mathbb{D}f \|_2).$$

- Pass limit to the continuous time!
  - Hyperbolic scaling — drift uncertainty.

$$c(x,y) = \lim_{N\to\infty} N^{p-1} \sum_{n=1}^{N} |\Delta x_n - \Delta y_n|^p = \|\partial_t(x-y)\|^p.$$

    A *pathwise* Malliavin derivative leads to $\Upsilon = L^*(\| {}^{\mathrm{o}}\mathbb{D}f\|_q)$.
  - Parabolic scaling — volatility uncertainty. Focus on $p = 2$ and $\mu = \gamma$.

$$c(x,y) = \lim_{N\to\infty} \sum_{n=1}^{N} |\Delta x_n - \Delta y_n|^2 = [x-y]_T.$$

    An *extended* Skorokhod integral gives Υ_{Mart}.

Uncertainty Sensitivity of DRO optimisers

Mathematical Institute

Recall: $V(\delta) = \inf_{a \in \mathbb{R}^k} \sup_{\nu \in B_\delta^p(\mu)} \int_{\mathbb{R}^d} f(x, a) \, \nu(dx).$

Theorem

For $p = q = 2$, under suitable regularity and growth assumptions,

$$\lim_{\delta \to 0} \frac{a^*(\delta) - a^*}{\delta} = -\frac{1}{\Upsilon} (\nabla_a^2 V(0, a^*))^{-1} \int \nabla_x \nabla_a f(x, a^*) \nabla_x f(x, a^*) \, \mu(dx),$$

where $a^* := a^*(0)$.

Extends to general $p > 1$ and semi-norms. Applications to:

- CLT for DRO optimisers
- out-of-sample error estimates
- ...

Ex: Square-root LASSO

Consider $\|(x,y)\|_* = |x|_r \mathbf{1}_{\{y=0\}} + \infty \mathbf{1}_{\{y\neq 0\}}$, $r>1$, $(x,y)\in\mathbb{R}^k\times\mathbb{R}$.
Then (see BLANCHET, KANG & MURTHY '19)

$$\inf_{a\in\mathbb{R}^k} \sup_{\nu\in B^2_\delta(\mu)} \int (y-\langle x,a\rangle)^2\, d\nu = \inf_{a\in\mathbb{R}^k} \left(\delta|a|_s + \sqrt{\int (y-\langle a,x\rangle)^2\, d\mu}\right)^2,$$

where $1/r+1/s=1$. $\mu=\hat{\mu}_N=\frac{1}{N}\sum_{i=1}^N \delta_{(x^i,y^i)}$ encodes the observations.
System is overdetermined so that $D=\int xx^T\,\mu(dx)$ is invertible.
$\delta=0$ case is the ordinary least squares regression: $a^*=\frac{1}{N}D^{-1}\int yx d\mu$.

Ex: Square-root LASSO

Consider $\|(x,y)\|_* = |x|_r \mathbf{1}_{\{y=0\}} + \infty \mathbf{1}_{\{y\neq 0\}}$, $r>1$, $(x,y)\in\mathbb{R}^k\times\mathbb{R}$.
Then (see Blanchet, Kang & Murthy '19)

$$\inf_{a\in\mathbb{R}^k} \sup_{\nu\in B_\delta^2(\mu)} \int (y-\langle x,a\rangle)^2\, d\nu = \inf_{a\in\mathbb{R}^k} \left(\delta|a|_s + \sqrt{\int (y-\langle a,x\rangle)^2\, d\mu}\right)^2,$$

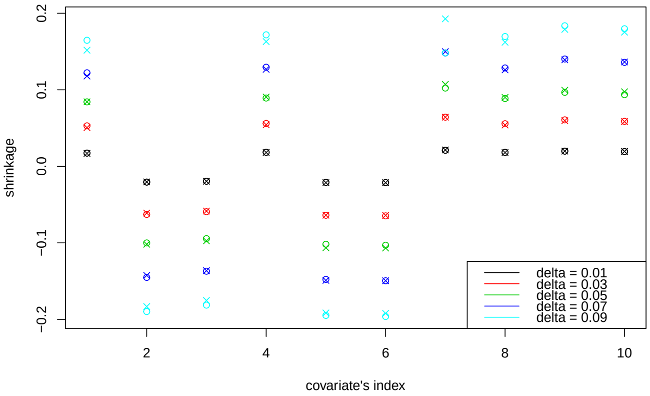
where $1/r+1/s=1$. $\mu=\hat{\mu}_N=\frac{1}{N}\sum_{i=1}^N \delta_{(x^i,y^i)}$ encodes the observations.
System is overdetermined so that $D=\int xx^T\,\mu(dx)$ is invertible.
$\delta=0$ case is the ordinary least squares regression: $a^*=\frac{1}{N}D^{-1}\int yx d\mu$.
$\delta>0$, $s=1 \rightsquigarrow$ RHS $=$ square-root LASSO regression Belloni et al. '11
$\delta>0$, $s=2 \rightsquigarrow$ RHS $\approx$ Ridge regression
Then $a^*(\delta)$ is approximately, for $s=1$ and $s=2$ (cf. Tibshirani '96):

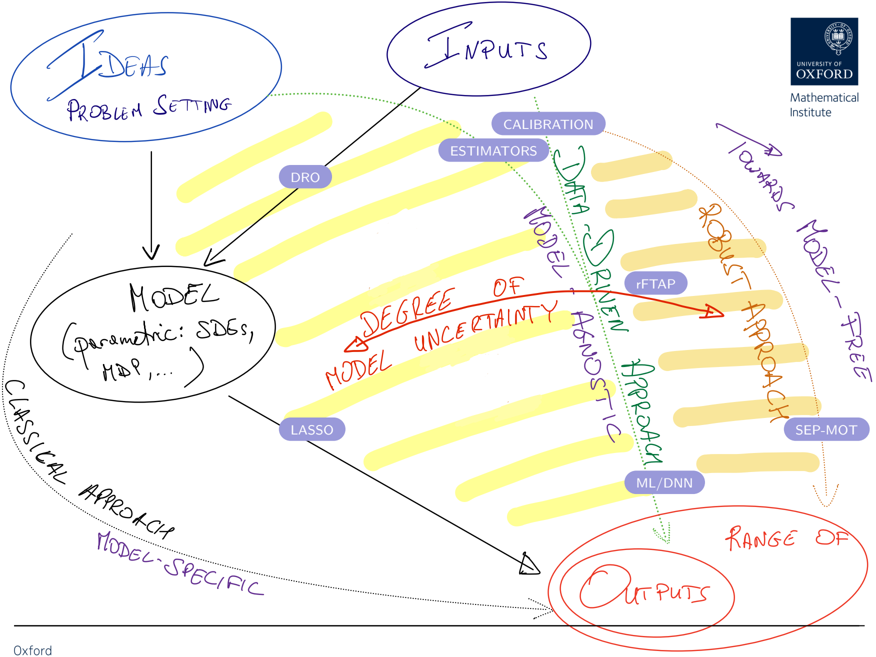
$$a^*-\delta\sqrt{V(0)}D^{-1}\mathrm{sgn}(a^*) \quad \text{and} \quad a^*-\delta a^*\frac{\sqrt{V(0)}}{|a^*|_2}D^{-1}$$

Square-root LASSO: numerics

Comparison of exact (o) and first-order (x) approximation of square-root LASSO coefficients for 2000 data generated from: (with all X_i, ε i.i.d. $\mathcal{N}(0,1)$)

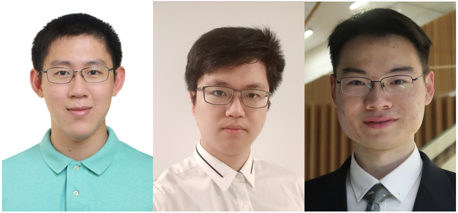
$$Y = 1.5X_1 - 3X_2 - 2X_3 + 0.3X_4 - 0.5X_5 - 0.7X_6 + 0.2X_7 + 0.5X_8 + 1.2X_9 + 0.8X_{10} + \varepsilon.$$




IDEAS
PROBLEM SETTING
INPUTS
MODEL
(parametric: SDEs, MDP, ...)
CALIBRATION
ESTIMATORS
DRO
DATA-DRIVEN APPROACH
MODEL-AGNOSTIC
TOWARDS MODEL-FREE
ROBUST APPROACH
rFTAP
DEGREE OF MODEL UNCERTAINTY
LASSO
SEP-MOT
ML/DNN
RANGE OF
OUTPUTS
CLASSICAL APPROACH
MODEL-SPECIFIC

University of Oxford
Mathematical Institute

W-Distributional Robustness of NNs



with X. Bai, G. He, Y. Jiang

NuerIPS '23 and arXiv:2502:xxxx

GitHub: JanObloj/W-DRO-Adversarial-Methods

Image classification setup

- An image is interpreted as a tuple $(x, y) \in \mathcal{X} \times \mathcal{Y}$, where x denotes the feature vector and y denotes the class.
- W.l.o.g, we take $\mathcal{X} = [0,1]^n$ and $\mathcal{Y} = \{1, \dots, m\}$.
- $\mathbb{P}$ is a given data distribution on $\mathcal{X} \times \mathcal{Y}$.
- A neural network is a map $f_\theta : \mathcal{X} \to \mathbb{R}^m$

$$f_\theta(x) = f^l \circ \cdots \circ f^1(x), \qquad \text{where } f^i(x) = \sigma(w^i x + b^i).$$

- Prediction of x under f_θ is given by $\arg\max_{1 \le i \le m}\{f_\theta(x)_i\}$.

Image classification setup

- An image is interpreted as a tuple $(x, y) \in \mathcal{X} \times \mathcal{Y}$, where x denotes the feature vector and y denotes the class.
- W.l.o.g, we take $\mathcal{X} = [0,1]^n$ and $\mathcal{Y} = \{1, \dots, m\}$.
- $\mathbb{P}$ is a given data distribution on $\mathcal{X} \times \mathcal{Y}$.
- A neural network is a map $f_\theta : \mathcal{X} \to \mathbb{R}^m$

$$f_\theta(x) = f^l \circ \cdots \circ f^1(x), \qquad \text{where } f^i(x) = \sigma(w^i x + b^i).$$

- Prediction of x under f_θ is given by $\arg\max_{1 \le i \le m} \{f_\theta(x)_i\}$.

The aim of image classification is to find a model with high accuracy

$$A := \mathbb{P}(\arg \max_{1 \le i \le m} \{f_\theta(x)_i\} = y) = \mathbb{P}(S).$$

This is achieved by training the network f_θ according to:

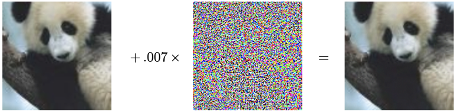
$$\inf_{\theta \in \Theta} \mathbb{E}_{\mathbb{P}}[J(\theta, x, y)] \quad \text{where } J(\theta, x, y) = L(f_\theta(x), y).$$

NN & adversarial attacks

Mathematical Institute

Consider data (x, y) from $\mathbb{P}$ and a NN trained according to:

$$\inf_{\theta} \int |J(\theta, x, y)| \, \hat{\mathbb{P}}(dx, dy).$$

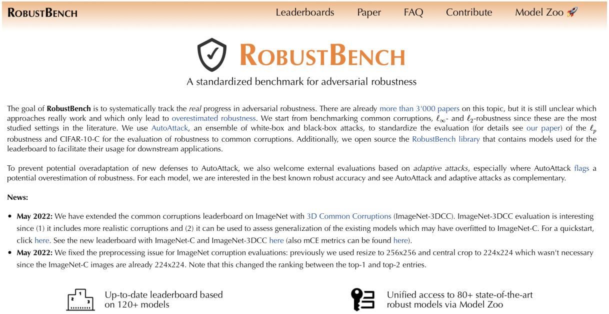


$\boldsymbol{x}$	$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$	$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"panda"	"nematode"	"gibbon"
57.7% confidence	8.2% confidence	99.3 % confidence

Source: Goodfellow, Shlens & Szegedy ICLR 2015

Adversarial robustness dataset and benchmarks

- Adversarial attacks and defence is a large field in ML
- RobustBench tracks over 3000 papers and maintains a leaderboard for CIFAR datasets



RobustBench Leaderboards Paper FAQ Contribute Model Zoo 🚀

RobustBench

A standardized benchmark for adversarial robustness

The goal of **RobustBench** is to systematically track the *real* progress in adversarial robustness. There are already more than 3'000 papers on this topic, but it is still unclear which approaches really work and which only lead to overestimated robustness. We start from benchmarking common corruptions, ℓ_∞- and ℓ_2-robustness since these are the most studied settings in the literature. We use AutoAttack, an ensemble of white-box and black-box attacks, to standardize the evaluation (for details see our paper) of the ℓ_p robustness and CIFAR-10-C for the evaluation of robustness to common corruptions. Additionally, we open source the RobustBench library that contains models used for the leaderboard to facilitate their usage for downstream applications.

To prevent potential overadaptation of new defenses to AutoAttack, we also welcome external evaluations based on *adaptive attacks*, especially where AutoAttack flags a potential overestimation of robustness. For each model, we are interested in the best known robust accuracy and see AutoAttack and adaptive attacks as complementary.

News:

- **May 2022:** We have extended the common corruptions leaderboard on ImageNet with 3D Common Corruptions (ImageNet-3DCC). ImageNet-3DCC evaluation is interesting since (1) it includes more realistic corruptions and (2) it can be used to assess generalization of the existing models which may have overfitted to ImageNet-C. For a quickstart, click here. See the new leaderboard with ImageNet-C and ImageNet-3DCC here (also mCE metrics can be found here).
- **May 2022:** We fixed the preprocessing issue for ImageNet corruption evaluations: previously we used resize to 256x256 and central crop to 224x224 which wasn't necessary since the ImageNet-C images are already 224x224. Note that this changed the ranking between the top-1 and top-2 entries.

Up-to-date leaderboard based on 120+ models

Unified access to 80+ state-of-the-art robust models via Model Zoo

Background on adv attacks/training

Mathematical Institute

Adversarial attack:

- Fast Gradient Sign Method (FGSM), see Goodfellow, Shlens & Szegedy '14
- Projected Gradient Descent (PGD), see Madry et al. '18
- Black-box attacks: Zeroth order optimization (Chen et al. '17), query-limited attack (Ilyas et al. '18) ...
- Autoattack, see Croce & Hein '20

Adversarial training:

- Random data generation by GAN/ diffusion models, see Gowal et al. '21 and Wang et al. '23
- Robustness–accuracy tradeoff, see TRADES Zhang et al. '19, MART Wang et al. '20, SCORE Pang et al. '22
- W-DRO based methods: Staib & Jegelka '17, Sinha, Namkoong & Duchi '18, Trillos & Trillos '22, Bui et al. '22 ...

W-DRO formulation

Clean training:

$$\inf_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}}[L(f_\theta(x), y)].$$

Adversarial training (Madry et al. '18):

$$\inf_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}}\left[\max_{\|x-x'\|_r\leq\delta} L(f_\theta(x'), y)\right].$$

W-DRO formulation

Clean training:

$$\inf_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}}[L(f_\theta(x), y)].$$

Adversarial training (Madry et al. '18):

$$\inf_{\theta\in\Theta} \mathbb{E}_{\mathbb{P}}\left[\max_{\|x-x'\|_r\leq\delta} L(f_\theta(x'), y)\right].$$

W-DRO adversarial training:

$$\inf_{\theta\in\Theta} \sup_{\mathbb{Q}\in B_\delta(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}[L(f_\theta(x), y)],$$

where $B_\delta(\mathbb{P})$ is the **p**-Wasserstein ball induced by a 'distance' d on $\mathcal{X}\times\mathcal{Y}$ defined by, $\mathbf{r}>1$,

$$d((x,y),(x',y')) = \|x-x'\|_r + \infty\mathbf{1}_{\{y\neq y'\}}.$$

Taking the ∞-Wasserstein ball reduces W-DRO to Madry et al..

UNIVERSITY OF OXFORD
Mathematical Institute

First order approximation

UNIVERSITY OF OXFORD
Mathematical Institute

Let $J_\theta(x,y) = L(f_\theta(x),y)$ and $V(\delta) = \sup_{\mathbb{Q}\in B_\delta(\mathbb{P})} \mathbb{E}_{\mathbb{Q}}[L(f_\theta(x),y)]$.

Theorem

Assuming J_θ is Lipschitz, the following first order approximations hold:

(i) $V(\delta) = V(0) + \delta\Upsilon + o(\delta)$, *where*

$$\Upsilon = \left(\mathbb{E}_{\mathbb{P}}\|\nabla_x J_\theta(x,y)\|_s^q\right)^{1/q}.$$

(ii) $V(\delta) = \mathbb{E}_{\mathbb{Q}_\delta}[J_\theta(x,y)] + o(\delta)$, *where*

$$\mathbb{Q}_\delta = \left[(x,y) \mapsto \left(x + \delta h(\nabla_x J_\theta(x,y))\|\Upsilon^{-1}\nabla_x J_\theta(x,y)\|_s^{q-1}, y\right)\right]_{\#}\mathbb{P},$$

and h is uniquely determined by $\langle h(x), x\rangle = \|x\|_s$.

Wasserstein distributionally adversarial attacks

Based on the first order approximation, we propose W-FGSM attack given by

$$x' = x + \delta h(\nabla_x J_\theta(x^t, y)) \|\Upsilon^{-1} \nabla_x J_\theta(x, y)\|_s^{q-1}, \qquad (1)$$

In particular, under the case $(\mathcal{W}_\infty, \ell_\infty)$ we retrieve FGSM attack given by

$$x' = x + \delta \, \mathrm{sgn}(\nabla_x J_\theta(x, y)).$$

Similarly, we propose W-PGD attack as

$$x^{t+1} = \mathrm{proj}_\delta \big(x^t + \alpha h(\nabla_x J_\theta(x^t, y)) \|\Upsilon^{-1} \nabla_x J_\theta(x^t, y)\|_s^{q-1} \big), \qquad (2)$$

where α is the stepsize, proj_δ is the projection onto Wasserstein ball $B_\delta(\mathbb{P})$ and $t = 1, \ldots, t_{max}$.

Define the adversarial accuracy A_δ as

$$A_\delta := \inf_{\mathbb{Q}\in B_\delta(\mathbb{P})} \mathbb{Q}(S)$$

We write $\mathcal{R}_\delta := A_\delta / A$ as a metric of robustness for neural networks. Any admissible attack gives an upper bound on adversarial accuracy:

$$\mathcal{R}_\delta \leq \mathcal{R}^u_\delta := \mathbb{Q}_\delta(S)/A.$$

Define the adversarial accuracy A_δ as

$$A_\delta := \inf_{\mathbb{Q} \in B_\delta(\mathbb{P})} \mathbb{Q}(S)$$

UNIVERSITY OF OXFORD
Mathematical Institute

We write $\mathcal{R}_\delta := A_\delta / A$ as a metric of robustness for neural networks. Any admissible attack gives an upper bound on adversarial accuracy:

$$\mathcal{R}_\delta \leq \mathcal{R}_\delta^u := \mathbb{Q}_\delta(S)/A.$$

To obtain a lower bound we impose:

- $0 < \mathbb{Q}(S) < 1$,
- $\mathcal{W}_p(\mathbb{P}(\cdot \mid S), \mathbb{Q}(\cdot \mid S)) + \mathcal{W}_p(\mathbb{P}(\cdot \mid S^c), \mathbb{Q}(\cdot \mid S^c)) = o(\delta)$,

for any $\mathbb{Q} \in B_\delta(\mathbb{P})$.

Define the adversarial accuracy A_δ as

$$A_\delta := \inf_{\mathbb{Q}\in B_\delta(\mathbb{P})} \mathbb{Q}(S)$$

We write $\mathcal{R}_\delta := A_\delta / A$ as a metric of robustness for neural networks. Any admissible attack gives an upper bound on adversarial accuracy:

$$\mathcal{R}_\delta \leq \mathcal{R}_\delta^u := \mathbb{Q}_\delta(S)/A.$$

Theorem (lower bound)

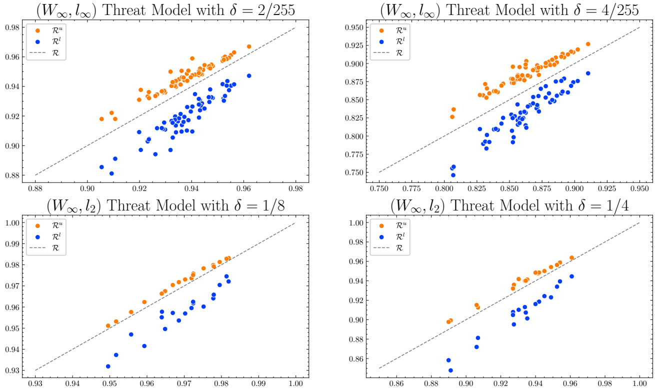
We write $W(0) = \mathbb{E}_{\mathbb{P}}[J_\theta(x, y)|S^c]$. *Under suitable assumptions, we have an asymptotic lower bound as* $\delta \to 0$

$$\mathcal{R}_\delta \geq \frac{W(0) - V(\delta)}{W(0) - V(0)} + o(\delta) = \mathcal{R}_\delta^l + o(\delta) \qquad (3)$$

where $\mathcal{R}_\delta^l = \min\{\widetilde{\mathcal{R}}_\delta^l, \overline{\mathcal{R}}_\delta^l\}$ *and the first order approximations are given by*

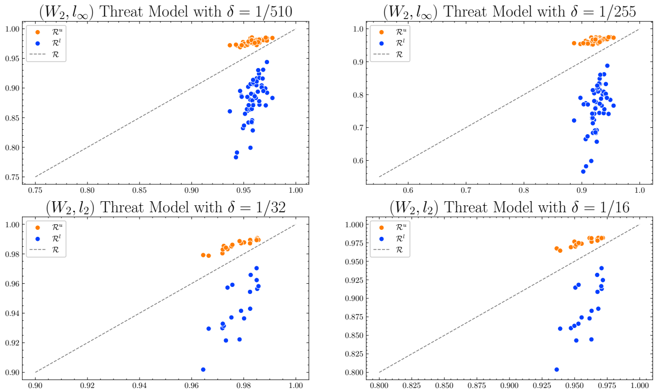
$$\widetilde{\mathcal{R}}_\delta^l = \frac{W(0) - \mathbb{E}_{\mathbb{Q}_\delta}[J_\theta(x, y)]}{W(0) - V(0)} \quad \text{and} \quad \overline{\mathcal{R}}_\delta^l = \frac{W(0) - V(0) - \delta\Upsilon}{W(0) - V(0)}. \qquad (4)$$

Bounds on $\mathcal{W}_\infty$-adversarial accuracy



$\mathcal{R}^l$ computed using CE loss. Blue dot takes around $1-2\%$ of computational time compared to the diagonal.

Bounds on $\mathcal{W}_2$-adversarial accuracy

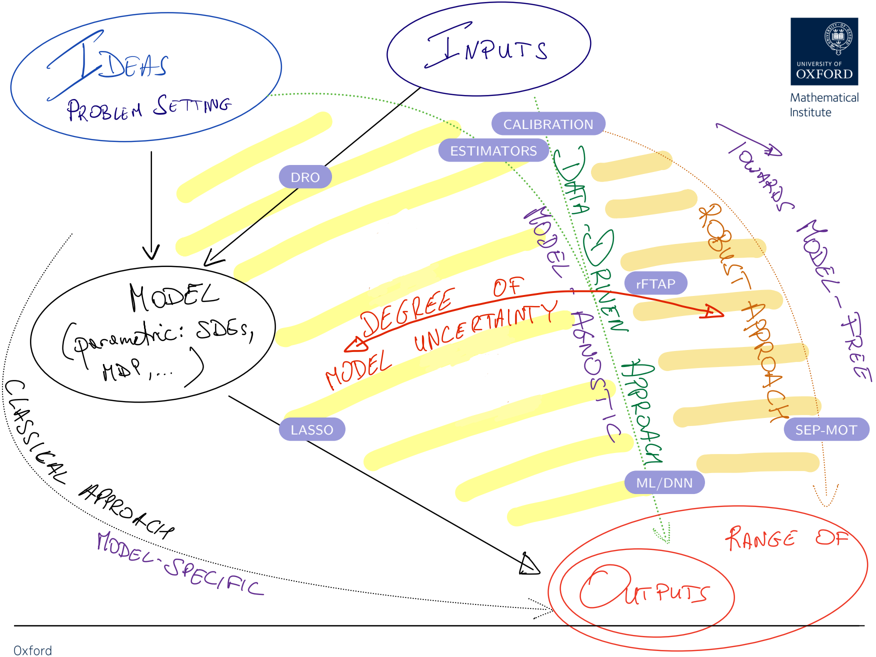


$\mathcal{R}^l$ computed using Rectified DLR loss. Blue dot takes 2% of computational time compared to the diagonal.

W-DRO Training as fine-tunning

Mathematical Institute

Networks	Clean Acc	$\mathcal{W}_\infty$ Adversarial Acc	$\mathcal{W}_2$ Adversarial Acc
Zhang et al. '19	83.71	59.99 (+2.95)	50.53 (+7.54)
Chen et al. '24	85.44	62.12 (+1.98)	53.42 (+9.66)
Gowal et al. '20	85.93	63.39 (-3.05)	52.14 (+1.15)
Cui et al. '23	88.88	68.71 (-2.21)	58.02 (+4.86)
Wang et al. '23	91.45	69.19 (-1.43)	55.93 (+3.79)


IDEAS
PROBLEM SETTING
INPUTS
MODEL
(parametric: SDEs, MDP, ...)
OUTPUTS
RANGE OF
CALIBRATION
ESTIMATORS
DRO
rFTAP
LASSO
SEP-MOT
ML/DNN
DATA-DRIVEN APPROACH
MODEL-AGNOSTIC
ROBUST APPROACH
TOWARDS MODEL-FREE
DEGREE OF MODEL UNCERTAINTY
CLASSICAL APPROACH
MODEL-SPECIFIC

Data: Historical Financial Returns

$$(r_1, \ldots, r_N) \in \mathbb{R}^{dN} \quad \text{v.s.} \quad \hat{\mathbb{P}}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{r_i} \in \mathcal{P}(\mathbb{R}^d)$$

base on a joint work with Johannes Wiesel, *Ann. Stat. (2021)*.

Superhedging price recalled

UNIVERSITY OF OXFORD
Mathematical Institute

Prices are seen as a stochastic process (S_t) in $\mathbb{R}^d_+$.
Dynamic trading $H \circ S$ with $H \in \mathcal{A}$ (admissible strategies).

$$\pi(\xi) = \inf\{x : \exists H \in \mathcal{A} \text{ s.t. } x + H \circ S \geq \xi \quad \text{in some sense}\}$$

Superhedging price recalled

Mathematical Institute

Prices are seen as a stochastic process (S_t) in $\mathbb{R}^d_+$.
Dynamic trading $H \circ S$ with $H \in \mathcal{A}$ (admissible strategies).

$$\begin{aligned}\pi^{\mathbb{P}}(\xi) &= \inf\{x : \exists H \in \mathcal{A} \text{ s.t. } x + H \circ S \geq \xi \quad \mathbb{P}\text{-a.s.}\} \\ &= \sup_{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \mathbb{P}} \mathbb{E}_{\mathbb{Q}}[\xi] \quad \text{pricing-hedging duality}\end{aligned}$$

- Model-specific approach: postulate a probability measure $\mathbb{P}$.

A simple setting: d assets, one-period, no other traded options.
Information: historical returns $r_1, \dots, r_N$ assumed i.i.d. from $\mathbb{P}$.

Aim: Build an estimator for

$$\pi^{\mathbb{P}}(\xi) = \inf \left\{ x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } \; x + H(r-1) \geq \xi(r) \; \mathbb{P}\text{-a.s.} \right\}$$

A simple setting: d assets, one-period, no other traded options. Information: historical returns $r_1, \ldots, r_N$ assumed i.i.d. from $\mathbb{P}$.

Aim: Build an estimator for

$$\pi^{\mathbb{P}}(\xi) = \inf \left\{ x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } \; x + H(r-1) \geq \xi(r) \; \mathbb{P}\text{-a.s.} \right\}$$

Theorem (Plugin estimator)

Let $\xi : \mathbb{R}^d_+ \to \mathbb{R}$ be Borel-measurable. Define the empirical measure $\hat{\mathbb{P}}_N = \frac{1}{N}\sum_{i=1}^N \delta_{r_i}$. Then

$$\lim_{N\to\infty} \pi^{\hat{\mathbb{P}}_N}(\xi) = \pi^{\mathbb{P}}(\xi) \qquad \mathbb{P}^\infty\text{-a.s.},$$

where $\mathbb{P}^\infty$ denotes the product measure on $\Pi_{i=1}^\infty \mathbb{R}^d_+$.

Problems with the plugin estimator

UNIVERSITY OF OXFORD
Mathematical Institute

The plugin estimator $\pi^{\hat{\mathbb{P}}_N}(\xi)$ is not robust!

- Not Financially: it underestimates the superhedging price $\pi^{\hat{\mathbb{P}}_N} \leq \pi^{\mathbb{P}}$.
- Not Statistically: (in the sense of Hampel). This applies to any estimator in fact (cf. Krätschmar, Schied and Zähle '11).
  $\Longrightarrow$ need to control the support $\Longrightarrow$ robustness w.r.t. W_∞.

W_p-approach

University of Oxford
Mathematical Institute

Fix $p \geq 1$. Assume we can find confidence bounds for the Glivenko-Cantelli theorem (see Dereich, Scheutzow & Schottstedt '11; Fournier & Guilllin '13):

$$\mathbb{P}^N(W_p(\mathbb{P}, \hat{\mathbb{P}}_N) \geq \varepsilon_N(\beta_N)) \leq \beta_N.$$

Definition

For a sequence $(k_N)_{N\in\mathbb{N}}$ such that $k_N \to \infty$ and $k_N\varepsilon_N(\beta_N) \to 0$ we define

$$\hat{\mathcal{Q}}_N = \left\{ \mathbb{Q} \in \mathcal{M} \;\middle|\; \exists \nu \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N), \; \left\| \frac{d\mathbb{Q}}{d\nu} \right\|_\infty \leq k_N \right\}.$$

Motivation/Intuition: $\hat{\mathcal{Q}}_N \rightsquigarrow \{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \mathbb{P}\}$.

W_p-approach: Consistency

Mathematical Institute

Theorem

Let g be Lipschitz continuous and bounded from below or continuous and bounded and $p \geq 1$. Then

$$\lim_{N \to \infty} \sup_{\mathbb{Q} \in \hat{\mathcal{Q}}_N} \mathbb{E}_{\mathbb{Q}}[\xi] = \pi^{\mathbb{P}}(\xi) \qquad \mathbb{P}^{\infty} - a.s.,$$

if $NA(\mathbb{P})$ holds.

W_p-approach: Robustness

UNIVERSITY OF OXFORD
Mathematical Institute

Definition

Let $\mathfrak{P}, \tilde{\mathfrak{P}} \subseteq \mathcal{P}(\mathbb{R}^d_+)$. We define p-Wasserstein-Hausdorff metric

$$W_p(\mathfrak{P}, \tilde{\mathfrak{P}}) = \max\left(\sup_{\mathbb{P}\in\mathfrak{P}} \inf_{\tilde{\mathbb{P}}\in\tilde{\mathfrak{P}}} W_p(\mathbb{P}, \tilde{\mathbb{P}}), \sup_{\tilde{\mathbb{P}}\in\tilde{\mathfrak{P}}} \inf_{\mathbb{P}\in\mathfrak{P}} W_p(\mathbb{P}, \tilde{\mathbb{P}}) \right).$$

Theorem

The estimator $\sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N} \mathbb{E}_{\mathbb{Q}}[\xi]$ *is robust with respect to the* W_p *in the sense that*

$$\sup_{\xi\in\mathcal{L}_1} \left| \sup_{\mathbb{Q}\in\hat{\mathcal{Q}}^1_N} \mathbb{E}_{\mathbb{Q}}[\xi] - \sup_{\mathbb{Q}\in\hat{\mathcal{Q}}^2_N} \mathbb{E}_{\mathbb{Q}}[\xi] \right| \leq W_p(\hat{\mathcal{Q}}^1_N, \hat{\mathcal{Q}}^2_N),$$

where $\hat{\mathcal{Q}}^i_N$ *are defined corresponding to* $\mathbb{P}^i \in \mathcal{P}(\mathbb{R}^d_+)$*,* $i = 1, 2$*.*

Robust AV@R hedging

UNIVERSITY OF OXFORD
Mathematical Institute

$$\pi_{\hat{\mathcal{Q}}_N}(\xi) = \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \sup_{\mathbb{Q}\in\mathcal{M}:\|d\mathbb{Q}/d\mathbb{P}\|_\infty\le k_N} \mathbb{E}_{\mathbb{Q}}[\xi]$$

$$= \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \sup_{\|d\mathbb{Q}/d\mathbb{P}\|_\infty\le k_N} \inf_{H\in\mathbb{R}^d} \mathbb{E}_{\mathbb{Q}}[\xi - H(r-1)]$$

$$= \inf_{H\in\mathbb{R}^d} \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \sup_{\|d\mathbb{Q}/d\mathbb{P}\|_\infty\le k_N} \mathbb{E}_{\mathbb{Q}}[\xi - H(r-1)]$$

$$= \inf_{H\in\mathbb{R}^d} \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} AV@R^{\mathbb{P}}_{\frac{k_N-1}{k_N}}(\xi - H(r-1))$$

$$= \inf\left\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } \sup_{\mathbb{P}\in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} AV@R^{\mathbb{P}}_{\frac{k_N-1}{k_N}}(\xi - H(r-1) - x)\le 0\right\}$$

Superhedging with respect to risk measures

Mathematical Institute

Consider risk evaluation which takes into account the capacity to trade in the liquidly traded assets:

$$\pi^{\rho_{\mathbb{P}}}(\xi) = \inf\left\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } \rho_{\mathbb{P}}(\xi - x - H(r-1)) \le 0 \ \mathbb{P}\text{-a.s.}\right\}$$

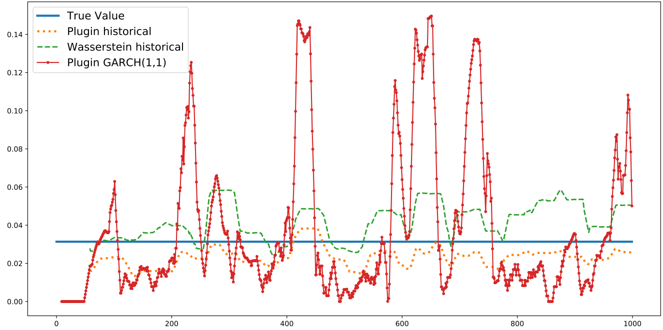
Under mild assumption, this is consistently estimated using:

$$\pi^{\rho}_{B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)}(\xi)$$

$$:= \inf\left\{x \in \mathbb{R}^d \,\middle|\, \exists H \in \mathbb{R}^d \text{ s.t. } \sup_{\nu \in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)} \rho_\nu(\xi - x - H(r-1)) \le 0\right\}.$$

Estimates for $\pi^{\mathrm{AV@R}_{0.95}^{\tilde{\mathbb{P}}}}((r-1)^+)$

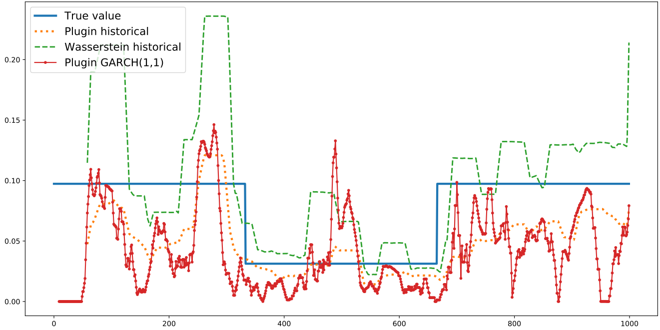
University of Oxford
Mathematical Institute



Rolling window of 50 data points, average of the last 10 estimates.
The data is from $\mathbb{P} \sim \mathrm{GARCH}(1,1)$.

Estimates for $\pi^{\mathrm{AV@R}^{\tilde{\mathbb{P}}}_{0.95}}((r-1)^+)$

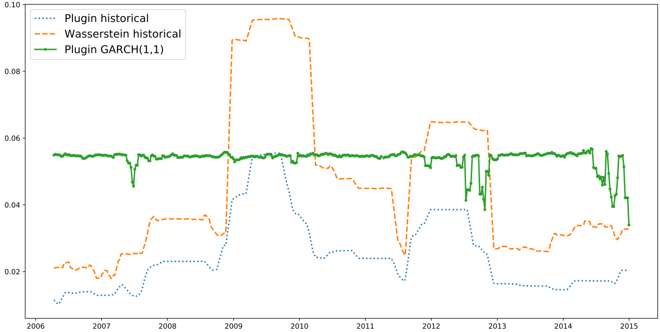
University of Oxford
Mathematical Institute



Rolling window of 50 data points, average of the last 10 estimates.
The data is from $\mathbb{P} \sim \mathrm{GARCH}(1,1)$.

Estimates for $\pi^{\mathrm{AV@R}_{0.95}^{\tilde{\mathbb{P}}}}((r-1)^+)$

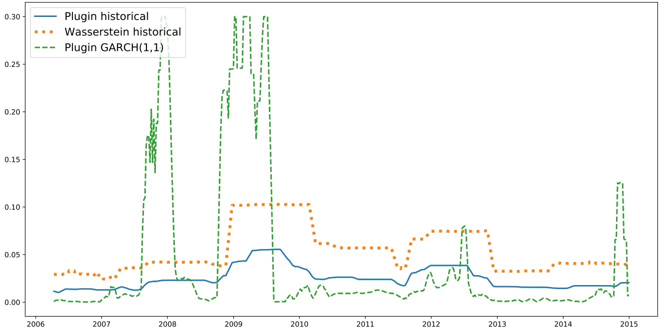
Mathematical Institute



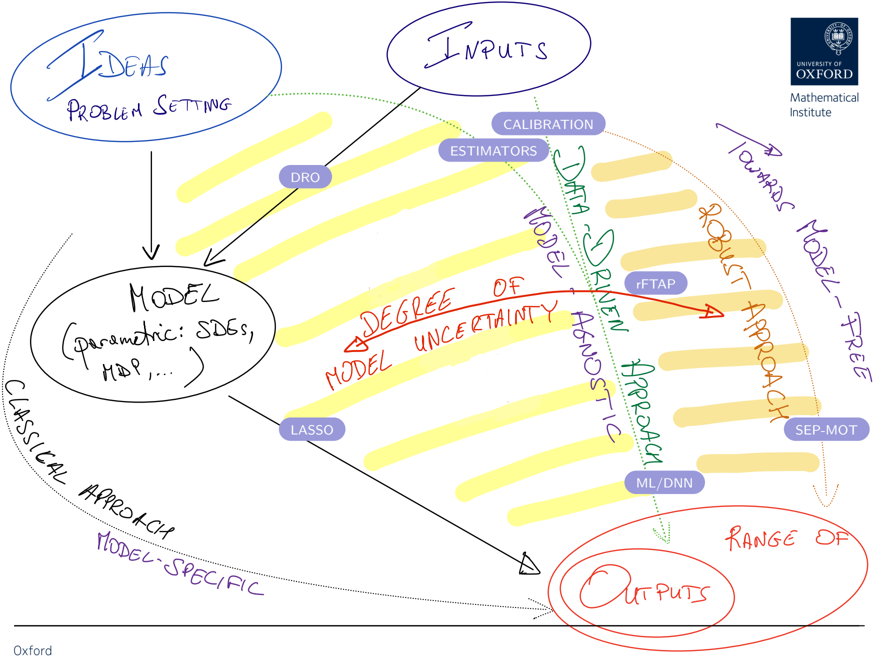
Rolling window of 50 data points, average of the last 5 estimates.
Weekly S&P500 returns.

Estimates for $\pi^{\mathrm{AV@R}_{0.95}^{\tilde{\mathbb{P}}}}((r-1)^+)$

UNIVERSITY OF OXFORD
Mathematical Institute



Rolling window of 50 data points, average of the last 5 estimates.
Weekly S&P500 log-returns.


IDEAS
PROBLEM SETTING
INPUTS
MODEL
(parametric: SDEs, MDP, ...)
CALIBRATION
ESTIMATORS
DRO
rFTAP
LASSO
SEP-MOT
ML/DNN
DATA-DRIVEN APPROACH
MODEL-AGNOSTIC
ROBUST APPROACH
TOWARDS MODEL-FREE
DEGREE OF MODEL UNCERTAINTY
CLASSICAL APPROACH
MODEL-SPECIFIC
RANGE OF
OUTPUTS

Oxford Mathematical Institute

Non-parametric calibration

with Julio Backhoff, Benjamin Joseph, Ivan Guo, Grégoire Loeper, Leo Wang
see *SIAM J. Financial Math. (2021), Risk Magazine (2022), Quantitative Finance (2024), Proc. AMS (forthcoming), arXiv:2310.13797*

Optimal transport – Fluid mechanics formulation

Mathematical Institute

OT: (Benamou-Brenier '00) continuous-time formulation

Minimising the cost function F under given initial density ρ_0 and final density ρ_1

$$\inf_{\rho, v} \int_0^1 \int_{\mathbb{R}^d} \rho(t,x) F(v(t,x))\, dxdt,$$

subject to the continuity equation

$$\partial_t \rho(t,x) + \nabla \cdot (\rho(t,x) v(t,x)) = 0,$$

and the initial and final distributions

$$\rho(0,x) = \rho_0, \quad \rho(1,x) = \rho_1.$$

Stochastic optimal transport

Tan & Touzi (2013) (also Mikami & Thieullen (2006), Huesmann & Trevisan (2017), Backhoff et al. (2017)): Consider probability measures $\mathbb{P}$ such that X is a semimartingale,

$$dX_t = \beta_t^{\mathbb{P}}\, dt + (\alpha_t^{\mathbb{P}})^{1/2}\, dW_t^{\mathbb{P}}.$$

We want to minimise

$$V(\mu_0, \mu_1) = \inf_{\mathbb{P} \in \mathcal{P}(\mu_0, \mu_1)} \mathbb{E}^{\mathbb{P}} \int_0^1 F(\alpha^{\mathbb{P}}, \beta^{\mathbb{P}})\, dt,$$

where $\mathcal{P}(\mu_0, \mu_1)$ contains probability measures satisfying

$$\mathbb{P} \circ X_0^{-1} = \mu_0, \quad \mathbb{P} \circ X_1^{-1} = \mu_1.$$

Stochastic optimal transport

Tan & Touzi (2013) (also Mikami & Thieullen (2006), Huesmann & Trevisan (2017), Backhoff et al. (2017)): Consider probability measures $\mathbb{P}$ such that X is a semimartingale,

$$dX_t = \beta_t^{\mathbb{P}}\,dt + (\alpha_t^{\mathbb{P}})^{1/2}\,dW_t^{\mathbb{P}}.$$

We want to minimise

$$V(\mu_0,\mu_1) = \inf_{\mathbb{P}\in\mathcal{P}(\mu_0,\mu_1)} \mathbb{E}^{\mathbb{P}} \int_0^1 F(\alpha^{\mathbb{P}},\beta^{\mathbb{P}})\,dt,$$

where $\mathcal{P}(\mu_0,\mu_1)$ contains probability measures satisfying

$$\mathbb{P}\circ X_0^{-1} = \mu_0, \quad \mathbb{P}\circ X_1^{-1} = \mu_1.$$

General constraint version: replace with $\mathcal{P}(\mu_0,\mu_1)$

$$\mathbb{P}\circ X_0^{-1} = \delta_{X_0} \quad \text{and} \quad \mathbb{E}^{\mathbb{P}} G_i(X_{\tau_i}) = c_i, \quad i = 1,\dots,m.$$

Stochastic OT & Calibration

UNIVERSITY OF OXFORD
Mathematical Institute

SOT induces a projection onto a subset of (semi)-martingales.

Use for calibration:

- Gather market data $\mathcal{G}$
- Fix a favourite reference model $\bar{\mathbb{P}}$
- Consider a cost F given by

$$F(\mathbb{P}) = \begin{cases} \mathrm{dist}(\mathbb{P}, \bar{\mathbb{P}}) & \text{if } \mathbb{P} \text{ is calibrated to } \mathcal{G}, \\ +\infty & \text{otherwise.} \end{cases}$$

- ensuring convexity to get duality
- Solve the dual via a non-linear (P)PDE
- $\mathbb{P}^*$ recovered via $\nabla F^*(\ldots)$.

- S&P 500 Index (SPX): a stock market index that measures the stock performance of 500 large companies listed in the US stock market.
- CBOE Volatility Index (VIX): a volatility index that measures the market's expectation of the volatility of SPX over the following 30 days.

UNIVERSITY OF OXFORD
Mathematical Institute

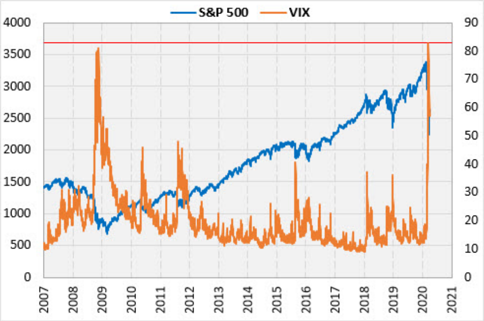


Figure: Historical SPX and VIX data. (Source: Schaeffer's Investment Research)

Underlying assets:

$$S_t = S_0 + \int_0^t \sigma_s S_s \, dW_s$$

$$VIX(t_0, T) = \sqrt{\mathbb{E}\left(\frac{100^2}{T - t_0} \int_{t_0}^{T} \sigma_t^2 \, dt \Big| \mathcal{F}_{t_0} \right)}$$

since the *realised variance* of S_t during $[t_0, T]$:

$$AF \sum_{i=1}^{n} \left(\log \frac{S_{t_i}}{S_{t_i - 1}} \right)^2 \to \frac{100^2}{T - t_0} \int_{t_0}^{T} \sigma_t^2 \, dt, \quad a.s.$$

Market traded instruments:

$$\begin{aligned} \text{SPX calls:} \quad & u^{SPX,c} = \mathbb{E}((S_T - K)^+) \\ \text{SPX puts:} \quad & u^{SPX,p} = \mathbb{E}((K - S_T)^+) \\ \text{VIX futures:} \quad & u^{VIX,f} = \mathbb{E}(VIX_{t_0}) \\ \text{VIX calls:} \quad & u^{VIX,c} = \mathbb{E}((VIX_{t_0} - K)^+) \\ \text{VIX puts:} \quad & u^{VIX,p} = \mathbb{E}((K - VIX_{t_0})^+) \end{aligned}$$

Why joint calibration?

Mathematical Institute

- VIX futures and options are very popular hedging instruments. e.g., Szado (2009) shows that VIX call options are better than S&P 500 put options as a hedging instrument against the financial crisis in 2008.
- An arbitrage argument (Guyon 2020): existence of a liquid market
  $\Rightarrow$ need for models that jointly calibrate to the option prices of SPX and VIX
  $\Rightarrow$ avoid arbitrage between financial institutions (or even within the same institution)
- Joint calibration problem: build a (stochastic volatility) model that jointly calibrates to the prices of SPX options, VIX futures and VIX options.
- Very challenging problem, especially for short maturities.

We consider a two dimensional stochastic process $X = (X^1, X^2)$ with

$$X_t^1 := \log S_t = X_0^1 - \frac{1}{2}\int_0^t \sigma_s^2\, ds + \int_0^t \sigma_s\, dW_s.$$

$$X_t^2 = \mathbb{E}\left(\frac{1}{2}\int_t^T \sigma_s^2\, ds \Big| \mathcal{F}_t\right).$$

Calibrating instruments: for $\tau \leq T$,

$$\begin{aligned}
\text{SPX calls:} \quad & u^{SPX,c} = \mathbb{E}((\exp(X_\tau^1) - K)^+) =: \mathbb{E}(G^{SPX,c}(X_\tau)) \\
\text{SPX puts:} \quad & u^{SPX,p} = \mathbb{E}((K - \exp(X_\tau^1))^+) =: \mathbb{E}(G^{SPX,p}(X_\tau)) \\
\text{VIX futures:} \quad & u^{VIX,f} = \mathbb{E}(100\sqrt{2X_{t_0}^2/(T-t_0)}) =: \mathbb{E}(G^{VIX,f}(X_{t_0})) \\
\text{VIX calls:} \quad & u^{VIX,c} = \mathbb{E}((100\sqrt{2X_{t_0}^2/(T-t_0)} - K)^+) =: \mathbb{E}(G^{VIX,c}(X_{t_0})) \\
\text{VIX puts:} \quad & u^{VIX,p} = \mathbb{E}((K - 100\sqrt{2X_{t_0}^2/(T-t_0)})^+) =: \mathbb{E}(G^{VIX,p}(X_{t_0}))
\end{aligned}$$

Dynamics of X are captured via drift and volatility:

$$(\alpha_t^{\mathbb{P}}, \beta_t^{\mathbb{P}}) = \left(\begin{bmatrix} -\frac{1}{2}\sigma_t^2 \\ -\frac{1}{2}\sigma_t^2 \end{bmatrix}, \begin{bmatrix} \sigma_t^2 & (\beta_t)_{12} \\ (\beta_t)_{12} & (\beta_t)_{22} \end{bmatrix} \right), \quad 0 \le t \le T,$$

where $(\beta_t)_{12} = d\langle X^1, X^2\rangle_t \,/\, dt$ and $(\beta_t)_{22} = d\langle X^2\rangle_t \,/\, dt$ and with the additional property that $X_T^2 = 0$ $\mathbb{P}$-a.s.

Given $\bar{\beta}$, a reference for β, define the cost function:

$$F(\alpha, \beta) = \begin{cases} \sum_{i,j=1}^{2} (\beta_{ij} - \bar{\beta}_{ij})^2 & \text{if } \alpha_1 = \alpha_2 = -\frac{1}{2}\beta_{11}, \\ +\infty & \text{otherwise.} \end{cases}$$

The cost function plays a regularisation role to ensure that X has the correct dynamics.

It is enough to consider diffusions! (Krylov / Gyongy / Brunick & Shreve)

Numerical method: solving the dual formulation

Dual formulation (via Fenchel–Rockafellar):

$$\text{maximise} \quad V = \sup_{\lambda \in \mathbb{R}^{m+n+2}} \lambda \cdot c - \phi^{\lambda}(0, X_0),$$

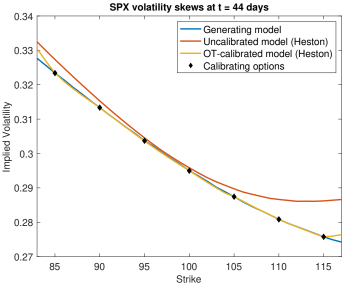
$$\text{subject to} \quad \partial_t \phi^{\lambda} + F^*(\nabla_x \phi^{\lambda}, \frac{1}{2}\nabla_x^2 \phi^{\lambda}) = -\sum_{i=1}^{m+n+2} \lambda_i \mathcal{G}_i \delta(t - \mathcal{T}_i), \quad \phi(T, \cdot) = 0.$$

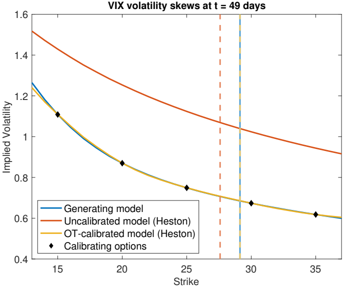
Numerical solution:

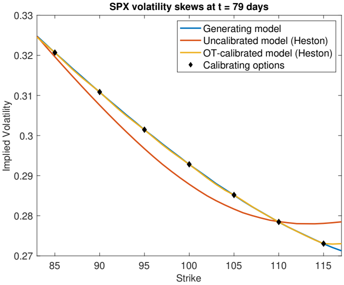
1. Set an initial λ (e.g., $\lambda = \mathbf{0}$),
2. Solve the HJB equation backward to get $\phi^{\lambda}(0, X_0)$,
3. Solve the linear PDEs and calculate all gradients,
4. Update λ by gradient descent.

This is analogous to the one dimensional case in Avellaneda, Friedman, Holmes and Samperi (1997)! Therein motivated by minimising a relative entropy–like functional.

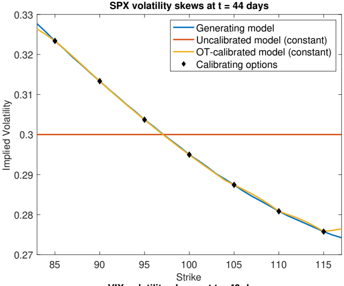
Mathematical
Institute

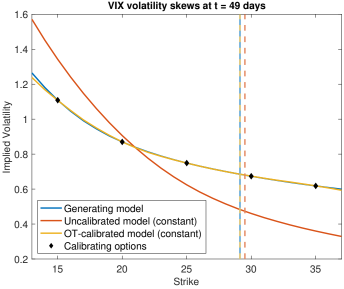


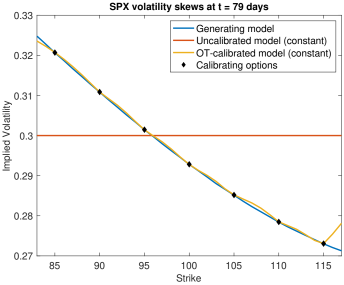




Simulated data example — Calibration results for Heston reference



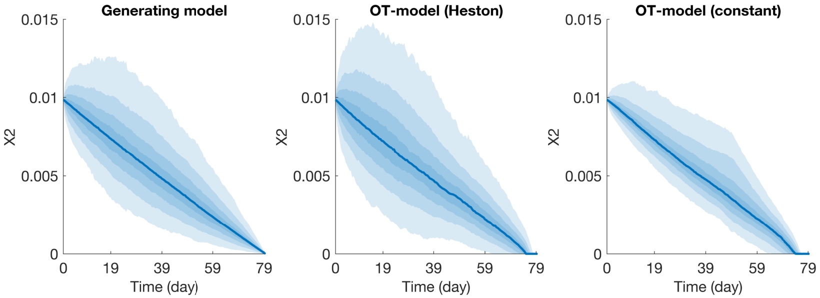




Simulated data example — Calibration results for constant reference

Simulated data example — Simulation of X^2

Mathematical
Institute



Market data example

UNIVERSITY OF OXFORD
Mathematical Institute

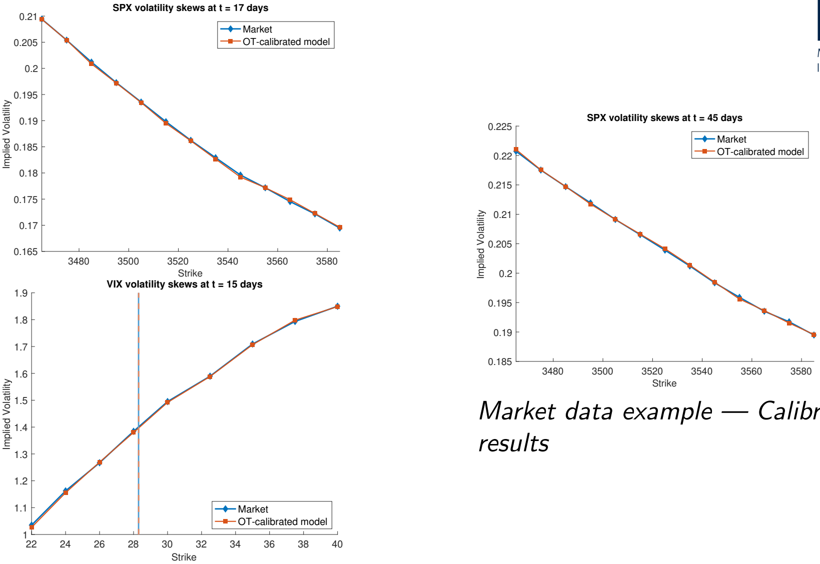
Market data as of 1st September 2020:

- 26 SPX call options maturing at 17 days and 45 days
- 1 VIX futures maturing at 15 days
- 9 VIX call option maturing at 15 days

These are the shortest maturities, which is known as the most challenging case!

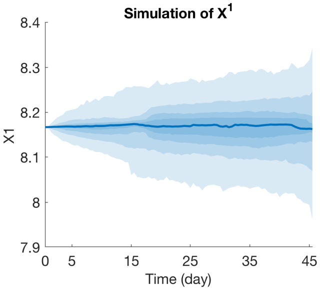
We calibrate the OT-model with a Heston reference $\bar{\beta}$. The parameters $(\bar{\kappa}, \bar{\theta}, \bar{\omega}, \bar{\eta}) = (4.99, 0.038, 0.52, -0.99)$ are obtained by (roughly) calibrating a standard Heston model to the SPX option prices.

Remark. Interest rates and dividends are NOT zero
$\Rightarrow$ model X^1 as the log of T-forward SPX price (instead of the spot price)
$\Rightarrow$ $\mathbb{P}$ are T-forward measures under which $\exp(X^1)$ is still a martingale.



Market data example — Calibration results

Market data example — Simulation of X^1 and X^2



SPX and Stochastic Interest Rates

We also look at the joint calibration of SPX with stochastic interest rates. This requires a modification to work with "discounted densities". We took the SPX as the underlying and the 1M US LIBOR for a proxy of the short rate. We obtained the following data on 23/05/2022 from a Bloomberg terminal:

- Calls on the SPX with expiry 19/08/2022,
- Caps on the one month LIBOR with notional $10,000,000 and expiry 23/08/2022,
- Calls on the SPX with expiry 18/11/2022,
- Caps on the one month LIBOR with notional $10,000,000 and expiry 23/11/2022.

Best of both: OT calibration with a parametric reference

We first calibrate a parametric model to obtain our reference guess. The then use OT to improve it and ensure perfect calibration. We consider CEV-Vasicek model:

$$X_t^1 = X_0^1 + X_t^2 - \frac{1}{2}\int_0^t \sigma_s^2 \,\mathrm{d}s + \int_0^t \sigma_s \,\mathrm{d}W_s^1,$$
$$X_t^2 = \int_0^t a(b - X_s^2)\mathrm{d}s + \int_0^t \sigma_r \mathrm{d}W_s^2,$$
$$\langle W_\cdot^1, W_\cdot^2\rangle_t = \int_0^t \rho \,\mathrm{d}s.$$

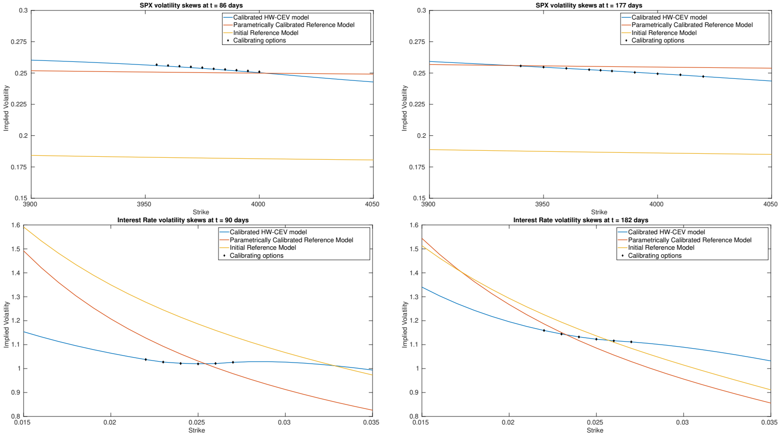
We calibrate via the usual LSE:

$$\min_{\sigma,\sigma_r,\gamma,\rho,a,b} \frac{1}{n}\sum_{i=1}^n \left(\tilde{u}_i(\sigma,\sigma_r,\gamma,\rho,a,b) - u_i\right)^2,$$

Where the minimisation is taken over $\sigma, \sigma_r, a, b > 0$ and $\rho \in [-1,1]$, u_i are the observed prices and $\tilde{u}_i$ are the model prices.

Market Data Example — CEV-Vasicek

Mathematical
Institute



Market Data Example — Plots of Characteristics

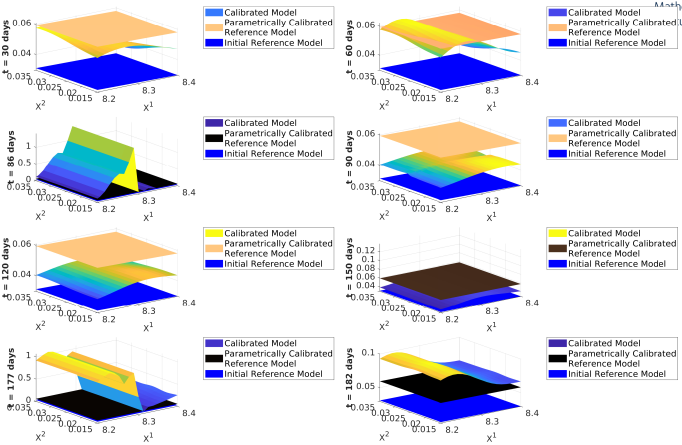


Figure: Compatison of $\beta_{11} = \sigma_X^2$ for the calibrated and generating model

Market Data Example — Plots of Characteristics

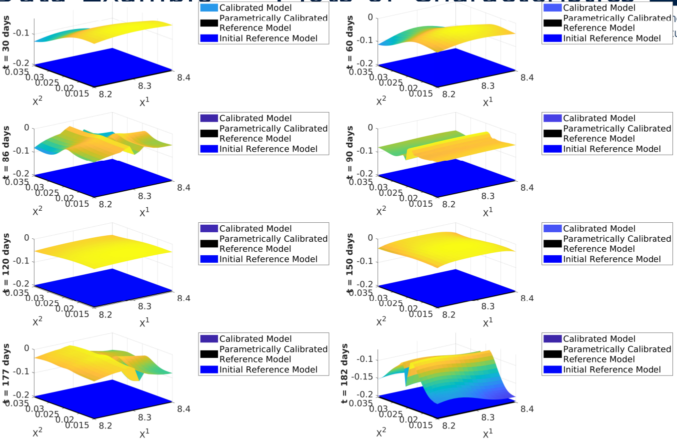


Figure: Comparison of the correlation $\frac{\beta_{12}}{\sqrt{\beta_{11}\beta_{22}}} = \rho$ for the calibrated and generating model

Market Data Example — Plots of Characteristics

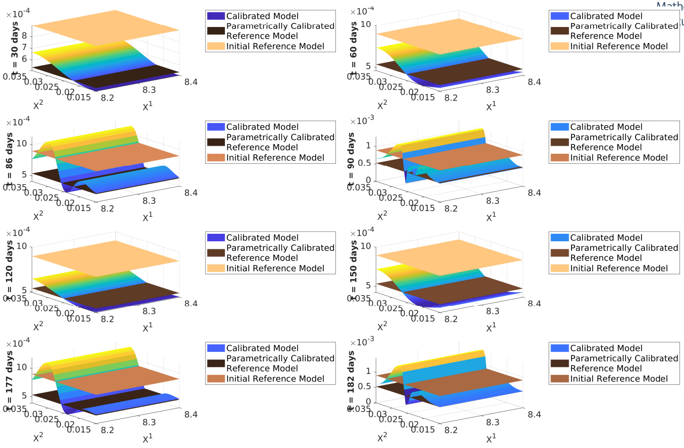


Figure: Comparison of $\beta_{22} = \sigma_r^2$ for the calibrated and generating model

Market Data Example — Plots of Characteristics

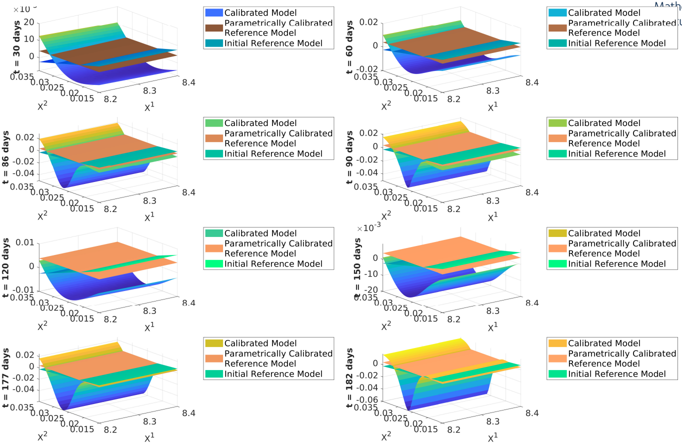


Figure: Comparison of $\alpha_2 = \mu_r$ for the calibrated and generating model

G-Bass calibration problem

Mathematical Institute

We want to solve

$$\mathbf{GmBB}_{\mu_0,\mu_1} = \inf_{\substack{S_0\sim\mu_0, S_1\sim\mu_1 \\ S_t = S_0 + \int_0^t \sigma_u S_u dB_u}} \mathbb{E}\left[\int_0^1 (\sigma_t - \bar{\sigma})^2 dt\right], \qquad \text{(G-mBB)}$$

That is, we want to find

- a calibrated model,
- which is the closest to the $\bar{\sigma}$-Black-Scholes model.

(M)OT Motivation

The celebrated Benamou-Brenier reformulation of the classical OT problem is:

$$\inf_{\substack{X_0 \sim \nu_0, X_1 \sim \nu_1 \\ X_t = X_0 + \int_0^t V_s ds}} \mathbb{E}\left[\int_0^1 |V_t|^2 dt\right],$$

i.e., we look for a particle with velocity as close as possible to a constant one, with given initial and terminal distributions.

(M)OT Motivation

Mathematical Institute

The celebrated Benamou-Brenier reformulation of the classical OT problem is:

$$\inf_{\substack{X_0 \sim \nu_0, X_1 \sim \nu_1 \\ X_t = X_0 + \int_0^t V_s ds}} \mathbb{E}\left[\int_0^1 |V_t|^2 dt\right],$$

i.e., we look for a particle with velocity as close as possible to a constant one, with given initial and terminal distributions.

More recently, Backhoff et al. '20 and Huesmann & Trevisan '19, considered the martingale analogue of this problem:

$$\mathbf{AmBB}_{\nu_0,\nu_1} = \inf_{\substack{M_0 \sim \nu_0, M_1 \sim \nu_1 \\ M_t = M_0 + \int_0^t \Sigma_s dB_s}} \mathbb{E}\left[\int_0^1 (\Sigma_t - \bar{\Sigma})^2 dt\right], \qquad \text{(A-mBB)}$$

where the optimisation is taken over filtered probability spaces with a Brownian motion $(B_t)_{t \geq 0}$, possibly starting from a non-trivial position B_0.

A-mBB: martingale Benamou-Brenier problem

We rewrite $\mathbf{AmBB}_{\nu_0,\nu_1}$ as

$$\inf_{\substack{M_0\sim\nu_0, M_1\sim\nu_1\\ M_t=M_0+\int_0^t \Sigma_s dB_s}} \mathbb{E}\left[\int_0^1 (\Sigma_t - \bar{\Sigma})^2 dt\right] = \bar{\Sigma}^2 + \int x^2 d\nu_1 - \int x^2 d\nu_0 - 2\bar{\Sigma}\mathbf{AP}_{\nu_0,\nu_1},$$

since

$$\mathbb{E}\left[\int_0^1 \Sigma_t^2 dt\right] = \mathbb{E}[\langle M\rangle_1] = \int x^2 d\nu_1 - \int x^2 d\nu_0$$

A-mBB: martingale Benamou-Brenier problem

Mathematical Institute

We rewrite $\mathbf{AmBB}_{\nu_0,\nu_1}$ as

$$\inf_{\substack{M_0\sim\nu_0,M_1\sim\nu_1\\ M_t=M_0+\int_0^t \Sigma_s dB_s}} \mathbb{E}\left[\int_0^1 (\Sigma_t-\bar{\Sigma})^2 dt\right] = \bar{\Sigma}^2 + \int x^2 d\nu_1 - \int x^2 d\nu_0 - 2\bar{\Sigma}\mathbf{AP}_{\nu_0,\nu_1},$$

since

$$\mathbb{E}\left[\int_0^1 \Sigma_t^2 dt\right] = \mathbb{E}[\langle M\rangle_1] = \int x^2 d\nu_1 - \int x^2 d\nu_0$$

and where the new problem is

$$\mathbf{AP}_{\nu_0,\nu_1} = \sup_{\substack{M_0\sim\nu_0,M_1\sim\nu_1\\ M_t=M_0+\int_0^t \Sigma_s dB_s\\ M \text{ martingale}}} \mathbb{E}\left[\int_0^1 \Sigma_t dt\right] = \sup_{\substack{M_0\sim\nu_0,M_1\sim\nu_1\\ M_t=M_0+\int_0^t \Sigma_s dB_s\\ M \text{ martingale}}} \mathbb{E}\left[M_1(B_1-B_0)\right]. \tag{AP}$$

A-mBB: martingale Benamou-Brenier problem

Mathematical Institute

We rewrite $\mathbf{AmBB}_{\nu_0,\nu_1}$ as

$$\inf_{\substack{M_0\sim\nu_0,M_1\sim\nu_1\\ M_t=M_0+\int_0^t \Sigma_s dB_s}} \mathbb{E}\left[\int_0^1 (\Sigma_t-\bar{\Sigma})^2 dt\right] = \bar{\Sigma}^2 + \int x^2 d\nu_1 - \int x^2 d\nu_0 - 2\bar{\Sigma}\mathbf{AP}_{\nu_0,\nu_1},$$

since

$$\mathbb{E}\left[\int_0^1 \Sigma_t^2 dt\right] = \mathbb{E}[\langle M\rangle_1] = \int x^2 d\nu_1 - \int x^2 d\nu_0$$

and where the new problem is

$$\mathbf{AP}_{\nu_0,\nu_1} = \sup_{\substack{M_0\sim\nu_0,M_1\sim\nu_1\\ M_t=M_0+\int_0^t \Sigma_s dB_s\\ M \text{ martingale}}} \mathbb{E}\left[\int_0^1 \Sigma_t dt\right] = \sup_{\substack{M_0\sim\nu_0,M_1\sim\nu_1\\ M_t=M_0+\int_0^t \Sigma_s dB_s\\ M \text{ martingale}}} \mathbb{E}\left[M_1(B_1-B_0)\right]. \tag{AP}$$

$\Longrightarrow$ it follows that $M_1 = F^{B_0}(B_1)$ is optimal with F^x increasing.

$\Longrightarrow$ in fact, we can find $\alpha\sim B_0$, such that $F^x\equiv F$.

$\Longrightarrow$ and hence $M_t = \mathbb{E}[F(B_1)|\mathcal{F}_t] = (F*\gamma_{1-t})(B_t)$, with $\gamma_t\sim\mathcal{N}(0,t)$.

G-mBB calibration problem

University of Oxford
Mathematical Institute

Similarly, in our calibration problem

$$\mathbf{GmBB}_{\mu_0,\mu_1} = \inf_{\substack{S_0 \sim \mu_0, S_1 \sim \mu_1 \\ S_t = S_0 + \int_0^t \sigma_u S_u dB_u}} \mathbb{E}\left[\int_0^1 (\sigma_t - \bar{\sigma})^2 dt\right],$$

for any such martingale S we have

$$\mathbb{E}\left[\int_0^1 \sigma_t^2 dt\right] = 2\mathbb{E}[\log(S_0/S_1)] = 2\int \log(x) d\mu_0 - 2\int \log(x) d\mu_1$$

and hence $\mathbf{GmBB}_{\mu_0,\mu_1}$ is equivalent to the following problem:

$$\mathbf{GP}_{\mu_0,\mu_1} = \sup_{\substack{S_0 \sim \mu_0, S_1 \sim \mu_1 \\ S_t = S_0 + \int_0^t \sigma_u S_u dB_u \\ S \text{ martingale}}} \mathbb{E}\left[\int_0^1 \sigma_t dt\right], \qquad \text{(GP)}$$

where $(B_t)_{t \geq 0}$ is a Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$.

From A-mBB to G-mBB and back

It turns out G-mBB can be mapped 1-1 to A-mBB for different marginals!
W.l.o.g., suppose $\int x\mu_0(dx) = \int x\mu_1(dx) = 1$.

From A-mBB to G-mBB and back

Mathematical Institute

It turns out G-mBB can be mapped 1-1 to A-mBB for different marginals! W.l.o.g., suppose $\int x\mu_0(dx) = \int x\mu_1(dx) = 1$.

The relationship between the problems can be deduced using PDE arguments on the dual side.

From A-mBB to G-mBB and back

It turns out G-mBB can be mapped 1-1 to A-mBB for different marginals! W.l.o.g., suppose $\int x\mu_0(dx) = \int x\mu_1(dx) = 1$.

The relationship between the problems can be deduced using PDE arguments on the dual side.

On the primal side, it is a change of measure argument, akin to Campi, Laachir and Martini '17; see also Beiglböck, Pammer and Riess '24.

Define $d\tilde{\mathbb{P}} := S_1 d\mathbb{P}$ and let $R_t = 1/S_t$, a $\tilde{\mathbb{P}}$-martingale. Then

$$\mathbb{E}\left[\int_0^1 \sigma_t dt\right] = \tilde{\mathbb{E}}\left[R_1 \int_0^1 \sigma_t dt\right] = \tilde{\mathbb{E}}\left[\int_0^1 R_t\sigma_t dt\right] = \tilde{\mathbb{E}}\left[\int_0^1 \Sigma_t dt\right],$$

where $\Sigma_t := R_t\sigma_t$ and Itô gives $dR_t = \Sigma_t d\tilde{W}_t$, for a $\tilde{\mathbb{P}}$-BM W.

$$\int g d\nu_1 := \tilde{\mathbb{E}}[g(R_1)] = \mathbb{E}\left[\frac{g(R_1)}{R_1}\right] = \mathbb{E}\left[g\left(\frac{1}{S_1}\right) S_1\right] = \int g\left(\frac{1}{y}\right) y\mu_1(dy).$$

From A-mBB to G-mBB and back

UNIVERSITY OF OXFORD
Mathematical Institute

For a μ-integrable $f : \mathbb{R}_+ \to \mathbb{R}_+$, we consider the f-reflected measure

$$f_\dagger \mu = \left(y \to \frac{1}{f(y)}\right)_\# \left(\frac{f(y)}{\int f(x)\mu(dx)}\mu(dy)\right).$$

Theorem

Let $\mu_0, \mu_1 \in \mathcal{P}_{-1,1}(\mathbb{R}_+)$ *satisfy* $\mu_0 \preccurlyeq_{cx} \mu_1$. *Let* $\nu_i = Id_\dagger \mu_i$, $i = 0, 1$. *Then*

$$\mathbf{GP}_{\mu_0,\mu_1} = \mathbf{AP}_{\nu_0,\nu_1},$$

and (GP) *admits a unique optimiser in distribution characterised by*

$$\mathbb{E}\Big[g\big(\{S_t : t \in [0,1]\}\big)\Big] = \mathbb{E}\Big[g\big(\{1/F(t,B_t) : t \in [0,1]\}\big) \cdot F(1,B_1)\Big],$$

for any measurable functional $g : C([0,1];\mathbb{R}) \to \mathbb{R}_+$, *where* $(F(t,B_t), t \in [0,1])$ *is an optimiser for* $\mathbf{AP}_{\nu_0,\nu_1}$.

Numerics

Mathematical Institute

Solution to $\textbf{AmBB}_{\nu_0,\nu_1}$, given by $M_1 = F(B_1)$ with $B_0 \sim \alpha$ is characterised by the Martingale Sinkhorn system:

$$\nu_0 = (\gamma_1 * F)_{\#}\alpha,$$
$$\nu_1 = F_{\#}(\gamma_1 * \alpha),$$

which is another way to write the fixed-point problem of Conze & Henry-Labordère '21, see also Acciaio, Marini and Pammer '23.

The above immediately allows us to solve also $\textbf{GmBB}_{\mu_0,\mu_1}$.

Furthermore, we can do this across many maturities. Note that the resulting local volatility surface will likely be discontinuous across maturities. We now test and compare A-mBB and G-mBB martingale on market data.

Market Data

Mathematical Institute

- BNP Paribas data on SPX options as of 27/10/2023, with many strikes and maturities: 27/11/2023, 29/12/2023, 19/01/2024, 29/02/2024, 15/03/2024, 28/03/2024, 19/04/2024 and 17/05/2024.
- CDFs built via Breeden Litzenberger formula and interpolated/extrapolated implied vols.
- Rescale variables $S_t \to S_t/S_0$. Com domain $(-0.5, 3) \times (T_k, T_{k+1})$ with 1001 spatial gridpoints and 1t01 time gridpoints.
- Solve heat equation using Crank-Nicolson.
- For G-Bass, we do CDF $\to$ density $\to$ reflected density $\to$ reflected cdf. Reflected density via numerical derivative over 250 gridpoints
- whole numerics took ca 5 min on a laptop.

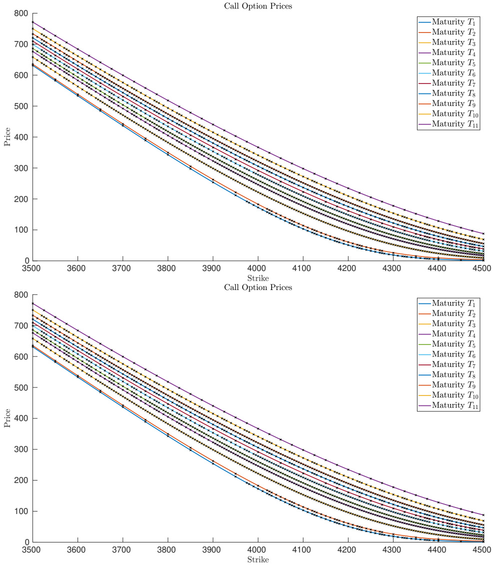
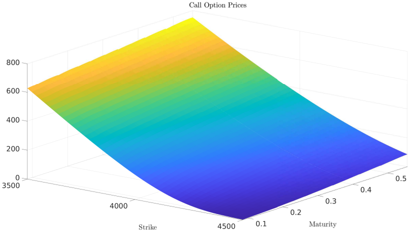


Figure: Call prices: Bass and Geometric Bass models.



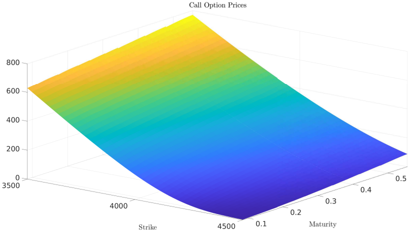


Figure: Call price surface: Bass and Geometric Bass models.

Conclusion & Outlook

Mathematical Institute

- Understanding and quantifying model uncertainty is a key problem in finance and across applied mathematics.
- Wasserstein distances offer a natural lift of the geometry
- and allow us to think in terms of probability measures instead of data points.
- Ideas from optimal transport offer a novel point of view on many classical problems.
- Both large-uncertainty and small-uncertainty regimes interesting and possible.
- Numerical methods available.

Thank You

papers and more available at
http://people.maths.ox.ac.uk/obloj/.

Robust approach to mathematical finance

Specific Model $\longleftrightarrow$ **Universal Model**

Specific Model	Universal Model
− strong assumptions, significant model risk	+ few assumptions, wide universe of scenarios
+ unique outputs	− non-unique outputs
− often takes limited inputs	+ sharpened by adding inputs

Robust approach to mathematical finance

Specific Model	⟷	Universal Model
– strong assumptions, significant model risk		+ few assumptions, wide universe of scenarios
+ unique outputs		– non-unique outputs
– often takes limited inputs		+ sharpened by adding inputs

> *The first part of this paper concentrates on laying the foundations for a rational theory of warrant pricing [...] to derive theorems about the properties of option prices based on assumptions sufficiently weak to gain universal support. [...]*
> *As one might expect, assumptions weak enough to be accepted by all are not sufficient to determine uniquely a rational theory of warrant pricing. To do so, more structure must be added to the problem through additional assumptions at the expense of loosing some agreement.*
> *[...] the second part of the paper examines their [B-S] model in detail.*
>
> (R. Merton, *Theory of rational option pricing*, 1973)

AIM: Develop a framework interpolating between the two modelling settings and quantifying the impact/risk of assumptions.

SPECIFIC MODEL $\longleftrightarrow$ UNIVERSAL MODEL

University of Oxford
Mathematical Institute

$\longrightarrow$ model uncertainty, quasi-sure approach...: $\mathbb{P} \rightsquigarrow \{\mathbb{P}_i : i \in \mathcal{I}\}$

- keep expanding the universe of scenarios
- $\mathcal{I}$ likely defined in terms of some model parameters
- natural to consider only $\mathbb{P}_i$ which are arbitrage-free
- weak notion of arbitrage, strong no-arbitrage condition

Avellaneda, Bayraktar, Bouchard, Carr, Deng, Fahim, Galichon, Hansen, Huang, Lyons, Neufeld, Nutz, Possamaï, Sargent, Tan, Touzi, Zhang, Zhou...

Specific Model $\longleftrightarrow$ Universal Model

University of Oxford — Mathematical Institute

$\longrightarrow$ model uncertainty, quasi-sure approach...: $\mathbb{P} \rightsquigarrow \{\mathbb{P}_i : i \in \mathcal{I}\}$

- keep expanding the universe of scenarios
- $\mathcal{I}$ likely defined in terms of some model parameters
- natural to consider only $\mathbb{P}_i$ which are arbitrage-free
- weak notion of arbitrage, strong no-arbitrage condition

$\longleftarrow$ pathwise approach...: $\Omega_0 \supseteq \Omega_1 \supseteq \ldots \supseteq \Omega_n \supseteq \ldots$

- keep shrinking the universe of scenarios
- Ω_n defined in terms of market observables
- as long as it supports *some rational pricing rule*
- strong notion of arbitrage, weak no-arbitrage condition

Avellaneda, Bayraktar, Bouchard, Carr, Deng, Fahim, Galichon, Hansen, Huang, Lyons, Neufeld, Nutz, Possamaï, Sargent, Tan, Touzi, Zhang, Zhou... Acciaio, Aksamit, Bartl, Beiglböck, Cheridito, Davis, Dupire, Dolinsky, Guo, Henry-Labordère, Hobson, Hou, Kupper, Mykland, Nadtochiy, Penkner, Prömel, Raval, Schachermayer, Schied, Siorpaes, Spoida, Tangpi, Vovk...

University of Oxford
Mathematical Institute

⟶ model uncertainty, quasi-sure approach...: $\mathbb{P} \rightsquigarrow \{\mathbb{P}_i : i \in \mathcal{I}\}$

- keep expanding the universe of scenarios
- $\mathcal{I}$ likely defined in terms of some model parameters
- natural to consider only $\mathbb{P}_i$ which are arbitrage-free
- weak notion of arbitrage, strong no-arbitrage condition

⟵ pathwise approach...: $\Omega_0 \supseteq \Omega_1 \supseteq \ldots \supseteq \Omega_n \supseteq \ldots$

- keep shrinking the universe of scenarios
- Ω_n defined in terms of market observables
- as long as it supports *some rational pricing rule*
- strong notion of arbitrage, weak no-arbitrage condition

O. & Wiesel ('21): the two yield essentially equivalent arbitrage pricing/hedging theory

Specific Model $\longleftrightarrow$ Universal Model

Mathematical Institute

$\longrightarrow$ model uncertainty, quasi-sure approach...: $\mathbb{P} \rightsquigarrow \{\mathbb{P}_i : i \in \mathcal{I}\}$

- keep expanding the universe of scenarios
- $\mathcal{I}$ likely defined in terms of some model parameters
- natural to consider only $\mathbb{P}_i$ which are arbitrage-free
- weak notion of arbitrage, strong no-arbitrage condition

$\longleftarrow$ pathwise approach...: $\Omega_0 \supseteq \Omega_1 \supseteq \ldots \supseteq \Omega_n \supseteq \ldots$

- keep shrinking the universe of scenarios
- Ω_n defined in terms of market observables
  Information *endogenously specifies* the (robust) modelling setup
- as long as it supports *some rational pricing rule*
- strong notion of arbitrage, weak no-arbitrage condition

O. & Wiesel ('21): the two yield essentially equivalent arbitrage pricing/hedging theory

Robust framework for pricing and hedging

University of Oxford
Mathematical Institute

- no frictions, prices in discounted units...
- d dynamically traded assets (primary or derivative)
  $S_t : \Omega_0 \to \mathbb{R}^d_+$, $t \leq T$, on a Polish space Ω_0, $\mathbb{F}$ natural filtration
- k statically traded European options $\phi_1, \dots, \phi_k$; initial prices $= 0$,

Robust framework for pricing and hedging

UNIVERSITY OF OXFORD
Mathematical Institute

- no frictions, prices in discounted units...
- d dynamically traded assets (primary or derivative) $S_t : \Omega_0 \to \mathbb{R}^d_+$, $t \leq T$, on a Polish space Ω_0, $\mathbb{F}$ natural filtration
- k statically traded European options $\phi_1, \ldots, \phi_k$; initial prices $= 0$,
- Model $=$ set of paths $\Omega \subset \Omega_0$
  - Special case 1: $\Omega = \Omega_0 \rightsquigarrow$ Universal Model
  - Special case 2: $\Omega = \Omega^{\mathbb{P}} \approx \text{support}(\mathbb{P}) \rightsquigarrow$ Specific Model

Robust framework for pricing and hedging

University of Oxford
Mathematical Institute

- no frictions, prices in discounted units...
- d dynamically traded assets (primary or derivative)
  $S_t : \Omega_0 \to \mathbb{R}^d_+$, $t \le T$, on a Polish space Ω_0, $\mathbb{F}$ natural filtration
- k statically traded European options $\phi_1, \ldots, \phi_k$; initial prices $= 0$,
- Model $=$ set of paths $\Omega \subset \Omega_0$
  - Special case 1: $\Omega = \Omega_0 \rightsquigarrow$ Universal Model
  - Special case 2: $\Omega = \Omega^{\mathbb{P}} \approx \text{support}(\mathbb{P}) \rightsquigarrow$ Specific Model
- Superhedging price: $\pi(\xi) = \pi_{\Omega,\Phi}(\xi) :=$

$$\inf \{x : \exists (\alpha_i) \;\&\; H \in \text{Pred}(\mathbb{F}) \text{ s.t. } x + \alpha \cdot \Phi + (H \circ S)_T \ge \xi \text{ on } \Omega\},$$

where $\alpha \cdot \Phi = \sum_{i=1}^k \alpha_i \phi_i$ and $(H \circ S)_T = \sum_{t=0}^{T-1} H_t(S_{t+1} - S_t)$.

Robust framework for pricing and hedging

UNIVERSITY OF OXFORD
Mathematical Institute

- no frictions, prices in discounted units...
- d dynamically traded assets (primary or derivative) $S_t : \Omega_0 \to \mathbb{R}^d_+$, $t \leq T$, on a Polish space Ω_0, $\mathbb{F}$ natural filtration
- k statically traded European options $\phi_1, \ldots, \phi_k$; initial prices $= 0$,
- Model $=$ set of paths $\Omega \subset \Omega_0$
  - Special case 1: $\Omega = \Omega_0 \rightsquigarrow$ Universal Model
  - Special case 2: $\Omega = \Omega^{\mathbb{P}} \approx \mathrm{support}(\mathbb{P}) \rightsquigarrow$ Specific Model
- Superhedging price: $\pi(\xi) = \pi_{\Omega,\Phi}(\xi) :=$

$$\inf \{x : \exists(\alpha_i) \;\&\; H \in \mathrm{Pred}(\mathbb{F}) \text{ s.t. } x + \alpha \cdot \Phi + (H \circ S)_T \geq \xi \text{ on } \Omega\},$$

where $\alpha \cdot \Phi = \sum_{i=1}^k \alpha_i \phi_i$ and $(H \circ S)_T = \sum_{t=0}^{T-1} H_t(S_{t+1} - S_t)$.

- Study evolution of $[-\pi(-\xi), \pi(\xi)]$ as $\Omega_0 \searrow \Omega^{\mathbb{P}}$

Robust framework — FTAP

University of Oxford
Mathematical Institute

Robust FTAP question: which Ω's are a dead-end and which can be used for "rational warrant pricing"?

Robust framework — FTAP

Robust FTAP question: which Ω's are a dead-end and which can be used for "rational warrant pricing"?

"dead end": everyone agrees there is an arbitrage (but may disagree on *what* it is, cf. Davis & Hobson '07)

Robust framework — FTAP

Mathematical Institute

Robust FTAP question: which Ω's are a dead-end and which can be used for "rational warrant pricing"?

"dead end": everyone agrees there is an arbitrage (but may disagree on *what* it is, cf. Davis & Hobson '07)

(Robust) Strong Arbitrage: (α, H) s.t. $\alpha \cdot \Phi + (H \circ S)_T > 0$ on Ω

$\mathcal{M}_{\Omega,\Phi} = \mathbb{F}$–martingale measures for S on Ω calibrated to Φ

Theorem (Robust FTAP, Burzoni et al. '19)

For any analytic $\Omega \subset \Omega_0$ there exists a filtration $\tilde{\mathbb{F}} \supseteq \mathbb{F}$ s.t.

$$\textit{no Strong Arbitrage on } \Omega \textit{ w.r.t. } \tilde{\mathbb{F}} \iff \mathcal{M}_{\Omega,\Phi} \neq \emptyset$$

$\Longrightarrow$ (r)FTAPs of Acciaio et al. '13, Bouchard and Nutz '15, DMW '90

Robust framework – Pricing-Hedging duality

Recall $\Omega^*_\Phi = \{\omega \in \Omega : \exists \mathbb{Q} \in \mathcal{M}_{\Omega,\Phi} \text{ s.t. } \mathbb{Q}(\{\omega\}) > 0\}$.
Redefine the superhedging price: $\pi(\xi) = \pi_{\Omega^*_\Phi,\Phi}(\xi) :=$

$$\inf \{x : \exists (\alpha_i) \;\&\; H \in \text{Pred}(\mathbb{F}^{pr}) \text{ s.t. } x + \alpha \cdot \Phi + (H \circ S)_T \geq \xi \text{ on } \Omega^*_\Phi\},$$

where $\alpha \cdot \Phi = \sum_{i=1}^k \alpha_i \phi_i$ and $(H \circ S)_T = \sum_{t=0}^{T-1} H_t(S_{t+1} - S_t)$.

Robust framework – Pricing-Hedging duality

UNIVERSITY OF OXFORD
Mathematical Institute

Recall $\Omega^*_\Phi = \{\omega \in \Omega : \exists \mathbb{Q} \in \mathcal{M}_{\Omega,\Phi} \text{ s.t. } \mathbb{Q}(\{\omega\}) > 0\}$.
Redefine the superhedging price: $\pi(\xi) = \pi_{\Omega^*_\Phi,\Phi}(\xi) :=$

$$\inf \{x : \exists (\alpha_i) \ \& \ H \in \text{Pred}(\mathbb{F}^{pr}) \text{ s.t. } x + \alpha \cdot \Phi + (H \circ S)_T \geq \xi \text{ on } \Omega^*_\Phi\},$$

where $\alpha \cdot \Phi = \sum_{i=1}^k \alpha_i \phi_i$ and $(H \circ S)_T = \sum_{t=0}^{T-1} H_t(S_{t+1} - S_t)$.

Theorem (Pricing–hedging duality, Burzoni et al. '19)

Suppose $\Omega \subset \Omega_0$ is such that $\mathcal{M}_{\Omega,\Phi} \neq \emptyset$ and $\pi_{\Omega^*_\Phi}(\pm\phi_j) < \infty$, $j \leq k$. Then for any measurable ξ

$$\pi_{\Omega^*_\Phi,\Phi}(\xi) = \sup_{\mathbb{Q} \in \mathcal{M}_{\Omega,\Phi}} \mathbb{E}^{\mathbb{Q}}[\xi]$$