Robustness and data-driven methods in mathematical finance

Jan Obłój
Mathematical Institute
University of Oxford

Winter Seminar on Mathematical Finance
Universiteit van Amsterdam and the Universe, 24/01/22

University of Oxford
Mathematical Institute
European Research Council
Established by the European Commission
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.

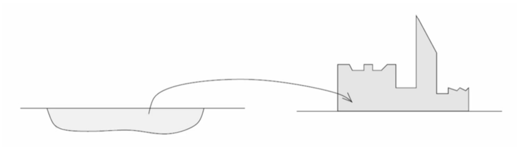
A VERY SHORT INTRODUCTION TO

Optimal Transport

Optimal Transport – Monge's Problem

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

Gaspard Monge (1781)



Monge's problem:

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

$\rightsquigarrow$ see Gabriel Peyré's `optimaltransport.github.io` (with thanks!)

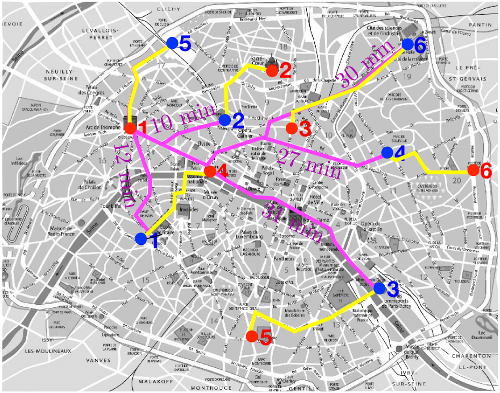
From bakeries to cafés BY G. PEYRÉ

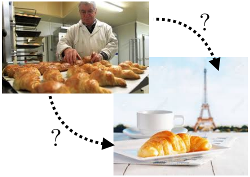
Mathematical
Institute

c_{ij}	y_1	y_2	y_3	y_4	y_5	y_6
x_1	12	10	31	27	10	30
x_2	22	7	25	15	11	14
x_3	19	7	19	10	15	15
x_4	10	6	21	19	14	24
x_5	15	23	14	24	31	34
x_6	35	26	16	9	34	15

From bakeries to cafés by G. Peyré

University of Oxford
Mathematical Institute

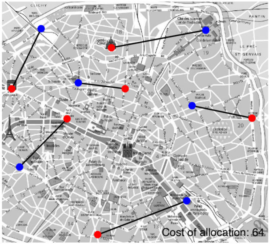


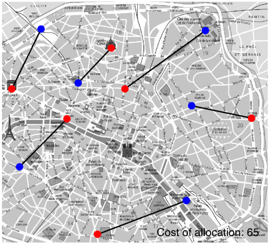


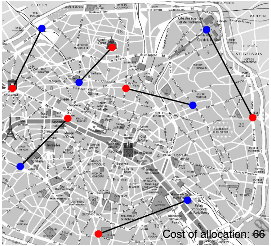
c_{ij}	y_1	y_2	y_3	y_4	y_5	y_6
x_1	12	10	31	27	10	30
x_2	22	7	25	15	11	14
x_3	19	7	19	10	15	15
x_4	10	6	21	19	14	24
x_5	15	23	14	24	31	34
x_6	35	26	16	9	34	15

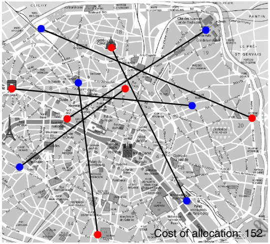
Optimal Cost: $10+7+15+10+14+9 = 65$ min

Allocations: optimal . . . or less so BY G. PEYRÉ





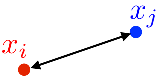


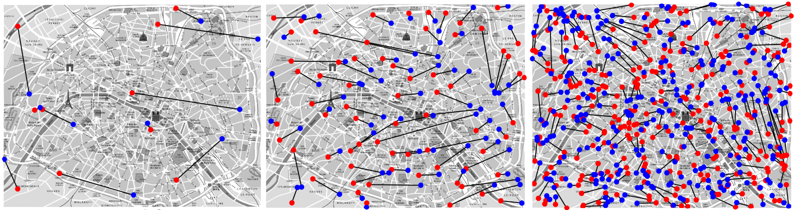


2D: discrete optimal transport BY G. PEYRÉ

$x_i, y_j \in \mathbb{R}^2$

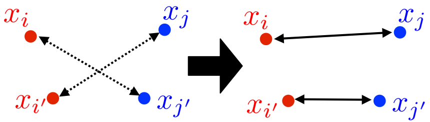
$$c_{i,j} = \|x_i - y_j\| = \sqrt{(x_i^1 - y_j^1)^2 + (x_i^2 - y_j^2)^2}$$

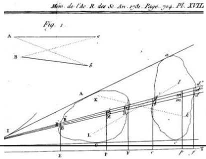


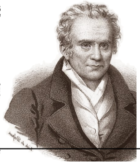


$n = 10$ $n = 70$ $n = 300$

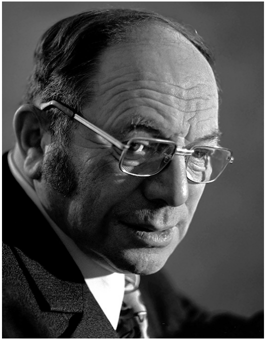
Optimality principle: *two segments do NOT cross*







Optimal Transport – Relaxation and Duality



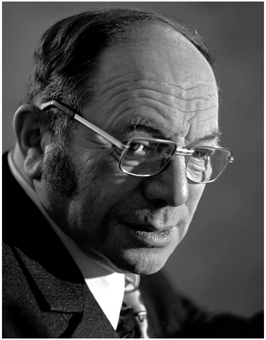
Leonid Kantorovich (1948)

Relaxed problem:

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

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

Optimal Transport – Relaxation and Duality



Leonid Kantorovich (1948)

Relaxed problem:

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

where $\mathrm{Cpl}(\mu,\nu)$ are 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 $\varphi(x) + \psi(y) \leq c(x,y)$.

Wasserstein (Kantorovich-Rubinstein) distance

Mathematical Institute

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

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

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

Wasserstein (Kantorovich-Rubinstein) distance

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

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

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

Define the Wasserstein-barycentre of measures $\mu_1, \ldots, \mu_n$ as

$$\mathrm{argmin}_\nu \sum_{i=1}^{n} W_p(\mu_i, \nu)$$

Image Classification Problem

Sample images from MNIST dataset

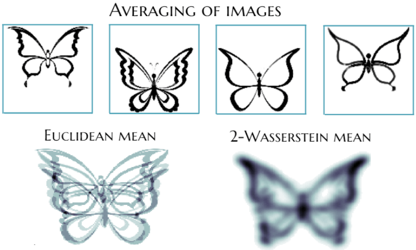
LeCun's version: The digits are size-normalized and centered in a fixed-size image. Training set of 60,000 examples + a test set of 10,000 examples.

Image Classification - a simple solution

- Use the training set to construct archetypes of each digit
- Given a new image, check its distance to all the archetypes and label according to the Nearest Neighbour.

Key idea: think of an image *not* as collection of b/w pixels (a matrix) but rather as 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

MNIST Digits: Wasserstein vs Euclidean mean

UNIVERSITY OF OXFORD
Mathematical Institute

A Wasserstein digits blend

Non-parametric sensitivities

based on joint works with Daniel Bartl, Samuel Drapeau and Johannes Wiesel
see *Proc. R. Soc. Lond. A (2021)*.

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.

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

$$B_\delta(\mu) \quad \text{is a } \delta\text{–neighbourhood of the model } \mu.$$

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 $B_\delta(\mu)$ being Wasserstein balls 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.

Model neighbourhood

Mathematical Institute

Measure μ is our model, such as

- $\mu = \hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{x^i}$ is the empirical measure of the observations/test set.
- μ comes from a mathematical modelling effort, e.g., an SDE;

Model neighbourhood

UNIVERSITY OF OXFORD
Mathematical Institute

Measure μ is our model, such as

- $\mu = \hat{\mu}_N = \frac{1}{N}\sum_{i=1}^N \delta_{x^i}$ is the empirical measure of the observations/test set.
- μ comes from a mathematical modelling effort, e.g., an SDE;

There are MANY ways to build a neighbourhood $B(\mu)$ of μ:

- data perturbation
- support estimates
- moments or density contraints
- Prokhorov of Hellinger distances
- Kullback–Leibler divergence/entropy bounds
- and more...

We propose use Wasserstein distances!

Wasserstein distance

For $p \geq 1$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with p^{th} moments, set

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

where $\mathrm{Cpl}(\mu,\nu) = \{\pi : \pi(\cdot \times \mathcal{S}) = \mu \text{ and } \pi(\mathcal{S}\times\cdot) = \nu\}$.
Denote the Wasserstein ball of size $\delta \geq 0$ around μ

$$B_\delta(\mu) = \{\nu \in \mathcal{P}(\mathcal{S}) : W_p(\mu,\nu) \leq \delta\}.$$

Wasserstein distance

For $p \geq 1$, $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with p^{th} moments, set

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

where $\mathrm{Cpl}(\mu,\nu) = \{\pi : \pi(\cdot \times \mathcal{S}) = \mu \text{ and } \pi(\mathcal{S}\times\cdot) = \nu\}$.
Denote the Wasserstein ball of size $\delta \geq 0$ around μ

$$B_\delta(\mu) = \{\nu \in \mathcal{P}(\mathcal{S}) : W_p(\mu,\nu) \leq \delta\}.$$

Note that, for a random variable $X \sim \mu$, on some $(\Omega, \mathbb{F}, \mathbb{P})$ we have

$$\sup_{\nu \in B_\delta(\mu)} \int_{\mathcal{S}} f(x)\, \nu(dx) = \sup_Z \mathbb{E}[f(X+Z)]$$

over all Z satisfying $\mathbb{E}[|Z|^p]^{1/p} \leq \delta$ and $X + Z \in \mathcal{S}$ a.s.

Small uncertainty limit

UNIVERSITY OF OXFORD
Mathematical Institute

Key property: $\hat{\mu}_N \xrightarrow{W_p} \mu$ + cnv rates, see Fournier & Guillin '14

Esfahani & Kuhn '18 argue that using Wasserstein balls gives

- finite sample guarantees,
- asymptotic consistency,
- tractability (see also Eckstein & Kupper '19)

Large uncertainty limit

UNIVERSITY OF OXFORD
Mathematical Institute

Pflug, Pichler & Wozabal '12 use Wasserstein balls for DRO in portfolio selection:

$$\inf_{a:\langle a,1\rangle=1} \sup_{\nu\in B_\delta(\mu)} \left(\mathbb{E}_\nu[\langle a,R\rangle] + \gamma \mathrm{Var}_\nu[\langle a,R\rangle]\right)$$

and show that

$$a^*(\delta) \xrightarrow{\delta\to\infty} \left(\frac{1}{N},\dots,\frac{1}{N}\right)$$

which may not be true for weaker or stronger metrics.

Main results

Part I: Sensitivity of the value function

Uncertainty Sensitivity of DRO problems

Mathematical Institute

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

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

Uncertainty Sensitivity of DRO problems

UNIVERSITY OF OXFORD
Mathematical Institute

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

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

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 $a^*(\delta)$.

Example 1: AV@R minimisation

Consider $X \sim \mu$ vector of returns in $\mathbb{R}^d$ and $a \in \mathcal{A} \subset \mathbb{R}^d$ portfolio

$$V(0) = \inf_{a\in\mathcal{A}} \mathrm{AV@R}_\alpha\left(a\cdot X\right) = \inf_{a\in\mathcal{A}, m\in\mathbb{R}} \left\{ m + \frac{1}{\alpha}\int (a\cdot x - m)^+ \mu(dx) \right\}$$

And its robust version reads

$$V(\delta) = \inf_{a\in\mathcal{A}} \mathcal{R}\mathrm{AV@R}_\alpha\left(a\cdot X\right) = \inf_{a\in\mathcal{A}, m\in\mathbb{R}} \sup_{\nu\in B_\delta(\mu)} \left\{ m + \frac{1}{\alpha}\int (a\cdot x - m)^+ \nu(dx) \right\},$$

where $B_\delta(\mu) = \{\nu \in \mathcal{P}(\mathcal{S}) : W_p(\mu,\nu) \le \delta\}$. A direct computation gives

$$\Upsilon = |a^*| \left(\frac{1}{\alpha^q} \int 1_{\{a^*\cdot x \ge V@R_\alpha(a^*\cdot L)\}} \right)^{\frac{1}{q}} \mu(dx) = \frac{|a^*|}{\alpha^{1/p}} \quad \text{, or}$$

$$\inf_{a\in\mathcal{A}} \mathcal{R}\mathrm{AV@R}_\alpha\left(a\cdot X\right) = \mathrm{AV@R}_\alpha\left(a^*\cdot X\right) + \frac{|a^*|}{\alpha^{1/p}}\delta + o(\delta).$$

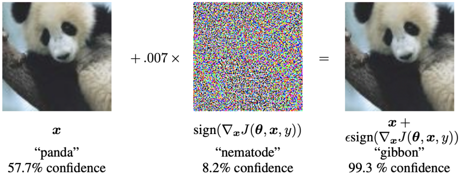
Example 2: NN & adversarial data

Most works focus on explaining the effects and creating algorithms to build adversarial examples.

Consider data (x, y) from μ and a 1-layer NN: $(A_1^*, A_2^*, b_1^*, b_2^*)$ solve

$$\inf \int \underbrace{\left|y - \left(\left(A_2(\cdot) + b_2\right) \circ \sigma \circ \left(A_1(\cdot) + b_1\right)\right)(x)\right|^p}_{=:f(x,y;A,b)} \mu(dx, dy),$$

where the inf is taken over $(A_1, A_2, b_1, b_2) \in \mathbb{R}^{k\times d} \times \mathbb{R}^{d\times k} \times \mathbb{R}^{k} \times \mathbb{R}^{d}$.



Source: Goodfellow, Shlens & Szegedy ICLR 2015

Example 2: NN & adversarial data

Mathematical Institute

Most works focus on explaining the effects and creating algorithms to build adversarial examples.
Consider data (x, y) from μ and a 1-layer NN: $(A_1^*, A_2^*, b_1^*, b_2^*)$ solve

$$\inf \int \underbrace{\left|y - \left((A_2(\cdot) + b_2) \circ \sigma \circ (A_1(\cdot) + b_1)\right)(x)\right|^p}_{=:f(x,y;A,b)} \mu(dx, dy),$$

where the inf is taken over $(A_1, A_2, b_1, b_2) \in \mathbb{R}^{k\times d} \times \mathbb{R}^{d\times k} \times \mathbb{R}^k \times \mathbb{R}^d$.

Then, sensitivity to adversarial data examples from $\hat{\mu} \in B_\delta(\mu)$ given by:

$$\left(\int |\nabla_{(x,y)} f(x, y; A^*, b^*)|^q \, \mu(dx, dy)\right)^{1/q}.$$

Ex 3: Robust call pricing (martingale constraint)

UNIVERSITY OF OXFORD
Mathematical Institute

We optimise over measures $\nu \in B_\delta(\mu)$ satisfying $\int x\, \nu(dx) = S_0$.
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 3: Robust call pricing (martingale constraint)

UNIVERSITY OF OXFORD
Mathematical Institute

We optimise over measures $\nu \in B_\delta(\mu)$ satisfying $\int x\,\nu(dx) = S_0$.
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

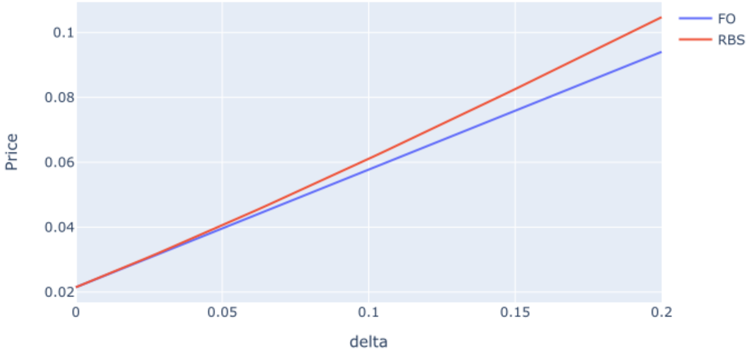
$$\mathcal{R}BS(\delta) = \sup_{\nu \in B_\delta(\mu)} \left\{ \int (S_0 x - K)^+ \nu(dx) \colon \int x\nu(dx) = 1 \right\}$$

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 $\mathcal{R}BS(\delta)$, from Bartl, Drapeau & Tangpi '19, vs first-order (FO) approximation $\mathcal{R}BS(0) + \Upsilon\delta$



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

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

UNIVERSITY OF OXFORD
Mathematical Institute

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

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

PARAMETRIC APPROACH

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

Then

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

NON-PARAMETRIC APPROACH

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

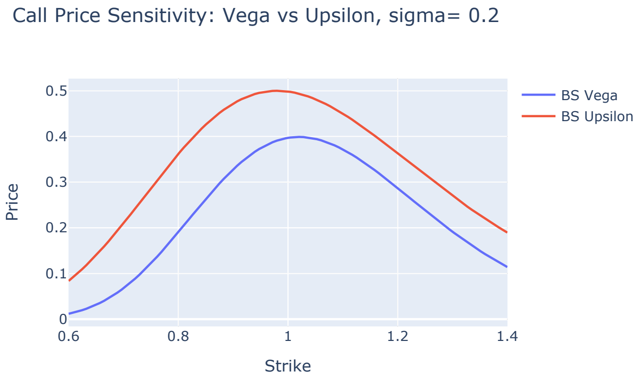
Then

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

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

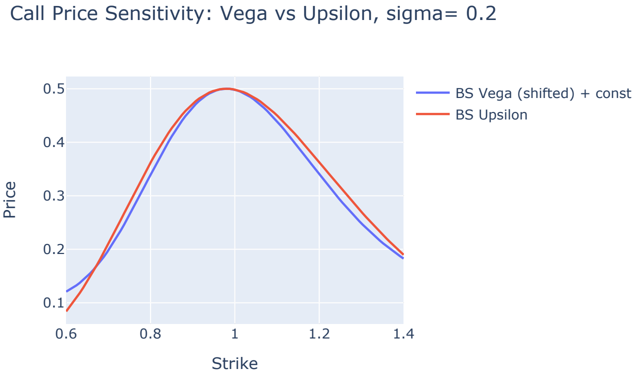
Consider the simple example of a call option pricing.

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



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

It turns out that $\Upsilon(S_0, K, \sigma) \approx \mathcal{V}(S_0, Ke^{\sigma^2}, \sigma) + \frac{1}{2} - \frac{1}{\sqrt{2\pi}}$.



Main results

Part II: Sensitivity of the optimisers

Sensitivity of optimisers

UNIVERSITY OF OXFORD
Mathematical Institute

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

The results extends to general $p > 1$ and semi-norms.

Example 1: Square-root LASSO

Consider $\|(x,y)\|_* = |x|_r 1_{\{y=0\}} + \infty 1_{\{y\neq 0\}}$, $r>1$, $(x,y)\in\mathbb{R}^k\times\mathbb{R}$.

Then (see Blanchet, Kang & Murthy '19)

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

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: $a^*=\frac{1}{N}D^{-1}\int yx d\mu$.

Example 1: Square-root LASSO

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

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

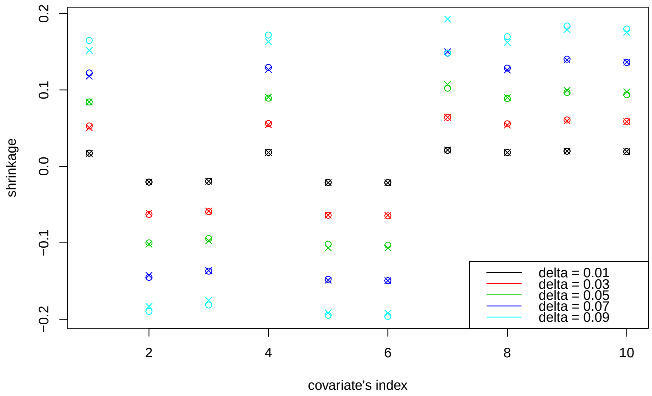
where $1/r+1/s=1$. $\mu=\hat{\mu}_N=\frac{1}{N}\sum_{i=1}^N \delta_{(x^i,y^i)}$ encodes the observations.
System is overdetermined so that $D=\int xx^T\mu(dx)$ is invertible.
$\delta=0$ case is the ordinary least squares regression: $a^*=\frac{1}{N}D^{-1}\int yx d\mu$.
$\delta>0$, $s=1 \rightsquigarrow$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

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

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. $\mathcal{N}(0,1)$)

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



Example 2: a CLT of Blanchet, Murphy and Si '19

Mathematical Institute

Consider the empirical measure μ_N of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \underset{min}{\text{arg}} \sup_{\nu \in B_\delta(\mu_N)} \int f(x,a)\,\nu(dx), \;\; a^{*,N} = \underset{min}{\text{arg}} \int f(x,a)\,\mu_N(dx), \;\; a^* = \underset{min}{\text{arg}} \int f(x,a)\,\mu(dx).$$

Regularity and strict convexity of f gives $a_{1/\sqrt{N}}^{*,N} \to a^*$.

Example 2: a CLT of Blanchet, Murphy and Si '19

Consider the empirical measure μ_N of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \underset{min}{\arg} \sup_{\nu \in B_\delta(\mu_N)} \int f(x,a)\,\nu(dx), \;\; a^{*,N} = \underset{min}{\arg} \int f(x,a)\,\mu_N(dx), \;\; a^* = \underset{min}{\arg} \int f(x,a)\,\mu(dx).$$

Regularity and strict convexity of f gives $a_{1/\sqrt{N}}^{*,N} \to a^*$.

Let $\sigma^2 := \int \nabla_a f(x,a^*)^T \nabla_a f(x,a^*)\,\mu(dx)$. Classical results give

$$\sqrt{N}\left(a^{*,N} - a^*\right) \Longrightarrow (\nabla_a^2 V(0,a^*))^{-1} H, \quad \text{where } H = \mathcal{N}(0,\sigma^2).$$

Example 2: a CLT of Blanchet, Murphy and Si '19

Consider the empirical measure μ_N of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \underset{min}{\arg} \sup_{\nu \in B_\delta(\mu_N)} \int f(x,a)\,\nu(dx), \;\; a^{*,N} = \underset{min}{\arg} \int f(x,a)\,\mu_N(dx), \;\; a^* = \underset{min}{\arg} \int f(x,a)\,\mu(dx).$$

Regularity and strict convexity of f gives $a^{*,N}_{1/\sqrt{N}} \to a^*$.

Let $\sigma^2 := \int \nabla_a f(x,a^*)^T \nabla_a f(x,a^*)\,\mu(dx)$. Classical results give

$$\sqrt{N}\left(a^{*,N} - a^*\right) \Longrightarrow (\nabla_a^2 V(0,a^*))^{-1} H, \quad \text{where } H = \mathcal{N}(0,\sigma^2).$$

Our results show that

$$\sqrt{N}\left(a^{*,N}_{1/\sqrt{N}} - a^{*,N}\right) \approx (\nabla_a^2 V(0,a^*))^{-1} \cdot \nabla_a \sqrt{\int |\nabla_x f(x,a^{*,N})|_s^2\,\mu_N(dx)}.$$

Putting the two together yields the CLT of Blanchet, Murphy and Si '19

$$\sqrt{N}\left(a^{*,N}_{1/\sqrt{N}} - a^*\right) \Longrightarrow (\nabla_a^2 V(0,a^*))^{-1} \left(H - \nabla_a \sqrt{\int |\nabla_x f(x,a^*)|_s^2\,\mu(dx)}\right).$$

Example 2: a CLT of Blanchet, Murphy and Si '19

Consider the empirical measure μ_N of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \underset{min}{arg} \sup_{\nu \in B_\delta(\mu_N)} \int f(x,a)\,\nu(dx), \;\; a^{*,N} = \underset{min}{arg} \int f(x,a)\,\mu_N(dx), \;\; a^* = \underset{min}{arg} \int f(x,a)\,\mu(dx).$$

Regularity and strict convexity of f gives $a_{1/\sqrt{N}}^{*,N} \to a^*$.

Let $\sigma^2 := \int \nabla_a f(x,a^*)^T \nabla_a f(x,a^*)\,\mu(dx)$. Classical results give

$$\sqrt{N}\left(a^{*,N} - a^*\right) \Longrightarrow (\nabla_a^2 V(0,a^*))^{-1} H, \quad \text{where } H = \mathcal{N}(0,\sigma^2).$$

Our results show that

$$\sqrt{N}\left(a_{1/\sqrt{N}}^{*,N} - a^{*,N}\right) \approx (\nabla_a^2 V(0,a^*))^{-1} \cdot \nabla_a \sqrt{\int |\nabla_x f(x,a^{*,N})|_s^2\,\mu_N(dx)}.$$

Putting the two together yields the CLT of Blanchet, Murphy and Si '19

$$\sqrt{N}\left(a_{1/\sqrt{N}}^{*,N} - a^*\right) \Longrightarrow (\nabla_a^2 V(0,a^*))^{-1}\,(H - \Theta\,).$$

Example 3: out-of-sample error

Consider the empirical measure μ_N of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \underset{min}{\arg} \sup_{\nu \in B_\delta(\mu_N)} \int f(x,a)\,\nu(dx),\ \ a^{*,N} = \underset{min}{\arg} \int f(x,a)\,\mu_N(dx),\ \ a^* = \underset{min}{\arg} \int f(x,a)\,\mu(dx).$$

Write $V(0,a) = \int f(x,a)\mu(dx)$. We get

$$V(0,a_\delta^{\star,N}) - V(0,a^\star) = \frac{1}{2}(a_\delta^{\star,N} - a^\star)^T \nabla_a^2 V(0,\tilde{a})(a_\delta^{\star,N} - a^\star),$$

for some $\tilde{a}$ between $a^\star$ and $a_\delta^{\star,N}$. Combining with the CLT gives:

$$V(0,a_\delta^{\star,N}) - V(0,a^\star) \approx \frac{1}{2N}(H-\Theta)^T(\nabla_a^2 V(0,a^\star))^{-1}(H-\Theta)$$

Example 3: out-of-sample error

Consider the empirical measure μ_N of N i.i.d. samples from μ and

$$a_\delta^{*,N} = \operatorname*{arg\,min} \sup_{\nu \in B_\delta(\mu_N)} \int f(x,a)\,\nu(dx), \;\; a^{*,N} = \operatorname*{arg\,min} \int f(x,a)\,\mu_N(dx), \;\; a^{*} = \operatorname*{arg\,min} \int f(x,a)\,\mu(dx).$$

Write $V(0,a) = \int f(x,a)\mu(dx)$. We get

$$V(0,a_\delta^{\star,N}) - V(0,a^\star) = \frac{1}{2}(a_\delta^{\star,N} - a^\star)^T \nabla_a^2 V(0,\tilde{a})(a_\delta^{\star,N} - a^\star),$$

for some $\tilde{a}$ between $a^\star$ and $a_\delta^{\star,N}$. Combining with the CLT gives:

$$V(0,a_\delta^{\star,N}) - V(0,a^\star) \approx \frac{1}{2N}(H - \Theta)^T (\nabla_a^2 V(0,a^\star))^{-1} (H - \Theta)$$

And we get a non-asymptotic approximation for the DRO impact

$$\int f(x,a_\delta^{*,N})\,\mu(dx) - \int f(x,a^{*,N})\,\mu(dx)$$

complementing and generalising the results of Anderson & Philipott '19.

UNIVERSITY OF OXFORD
Mathematical Institute

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)*.

Some math finance jargon

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).

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

Some math finance jargon

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).

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

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

Some math finance jargon

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).

$$\pi^{\mathcal{Q}}(\xi) = \inf\{x : \exists H \in \mathcal{A} \text{ s.t. } x + H \circ S \geq \xi \quad \mathcal{Q}\text{-q.s.}\}$$
$$\stackrel{?}{=} \sup_{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \text{ s.t. } \ldots} \mathbb{E}_{\mathbb{Q}}[\xi]$$

- Model-specific approach: postulate a probability measure $\mathbb{P}$.
- Model-independent/robust approach: weaker/no assumptions.
  - Quasi-sure approach considers a large set of measures $\mathcal{Q}$ (Super-)hedging is required $\mathcal{Q}$-q.s. ($\mathbb{P}$-a.s. $\forall \mathbb{P} \in \mathcal{Q}$)
  - Pathwise approach requires hedging property to hold for each scenario $\omega \in \Omega$ on for each $\omega \in \Omega' \subseteq \Omega$

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

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

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. $\mathcal{W}^\infty$.

$\mathcal{W}^p$-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):

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

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

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

$\mathcal{W}^p$-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.

$\mathcal{W}^p$-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

$$\mathcal{W}^p(\mathfrak{P}, \tilde{\mathfrak{P}}) = \max\left(\sup_{\mathbb{P}\in\mathfrak{P}} \inf_{\tilde{\mathbb{P}}\in\tilde{\mathfrak{P}}} \mathcal{W}^p(\mathbb{P}, \tilde{\mathbb{P}}), \sup_{\tilde{\mathbb{P}}\in\tilde{\mathfrak{P}}} \inf_{\mathbb{P}\in\mathfrak{P}} \mathcal{W}^p(\mathbb{P}, \tilde{\mathbb{P}})\right).$$

Theorem

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

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

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

Convergence of Wasserstein estimators

UNIVERSITY OF OXFORD
Mathematical Institute

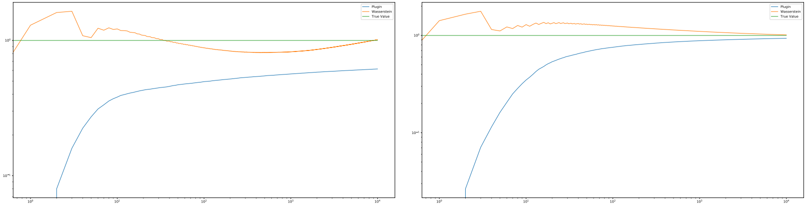


Figure: Wasserstein estimators with $\xi(r) = (1-r)\mathbb{1}_{\{r\leq 1\}} - \sqrt{r-1}\mathbb{1}_{\{r>1\}}$, $\mathbb{P} = \text{Exp}(1)$ *(left)* and $\xi(r) = (r-2)^+$, $\mathbb{P} = \exp(\mathcal{N}(0,1))$ *(right)*.

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

UNIVERSITY OF OXFORD
Mathematical Institute

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

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

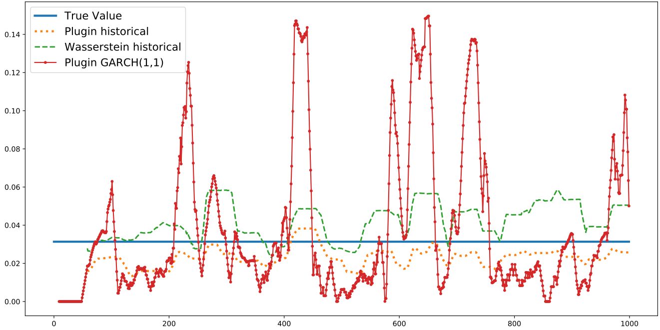
Under mild assumption, this is consistently estimated using:

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

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

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

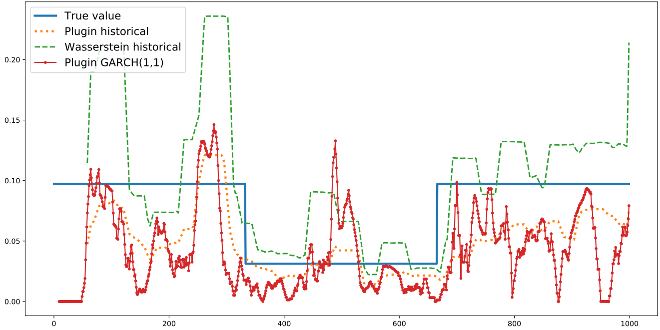
UNIVERSITY OF OXFORD
Mathematical Institute



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

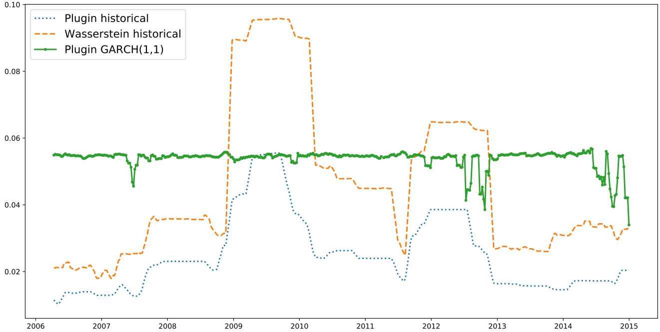
Mathematical Institute



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

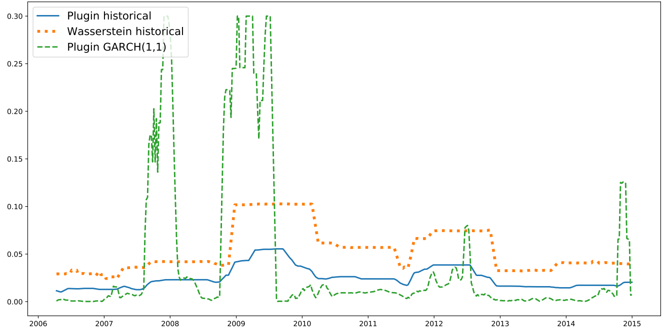
UNIVERSITY OF
OXFORD
Mathematical
Institute



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

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

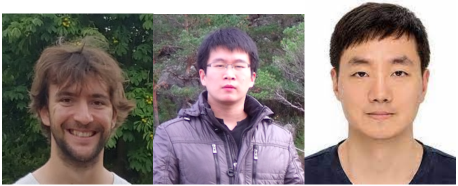
UNIVERSITY OF OXFORD
Mathematical Institute



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

UNIVERSITY OF OXFORD
Mathematical Institute

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

Mathematical Institute

- suppose you observe prices of call options, $K > 0$,

price $C(K)$ for a T-call with strike K: $(S_T - K)^+$

An (idealised) case study: the MOT problem

Mathematical Institute

- suppose you observe prices of call options, $K > 0$,

  price $C(K)$ 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

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

An (idealised) case study: the MOT problem

UNIVERSITY OF OXFORD
Mathematical Institute

- suppose you observe prices of call options, $K > 0$,

  price $C(K)$ 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

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

- Robust pricing of an exotic option with payoff ξ
  $\rightsquigarrow$ $\sup \mathbb{E}_{\mathbb{Q}}[\xi(S_t : t \leq T)]$ over such $\mathbb{Q}$s.
  Robust hedging is its dual problem.

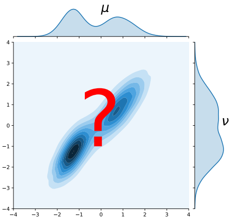
The MOT problem

UNIVERSITY OF OXFORD
Mathematical Institute

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

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

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



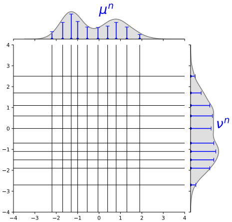
MOT Numerics: take I

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

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



MOT Numerics: take I

Mathematical
Institute

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

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

- Discretisation $(\mu,\nu) \rightsquigarrow (\mu^n,\nu^n)$ typically does NOT preserve the convex order, see Alfonsi et al. (2017).
- Relaxation of the martingale condition needed.
- Thm: Convergence under suitable choice of $\varepsilon_n \to 0$.

MOT Numerics: take II (Eckstein & Kupper '17)

University of Oxford
Mathematical Institute

- Numerics on the dual (superhedging) problem
- $\rightsquigarrow$ optimisation over functions
- $\rightsquigarrow$ Deep Neural Network implementation
  - hedging strategies $\in \mathcal{H}_n$ (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
    $\rightsquigarrow$ *LP fails, NN work for $dT \leq 30$ and then harder, sampling the superhedging condition tricky!*

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

A Toy Example

University of Oxford
Mathematical Institute

INPUTS:

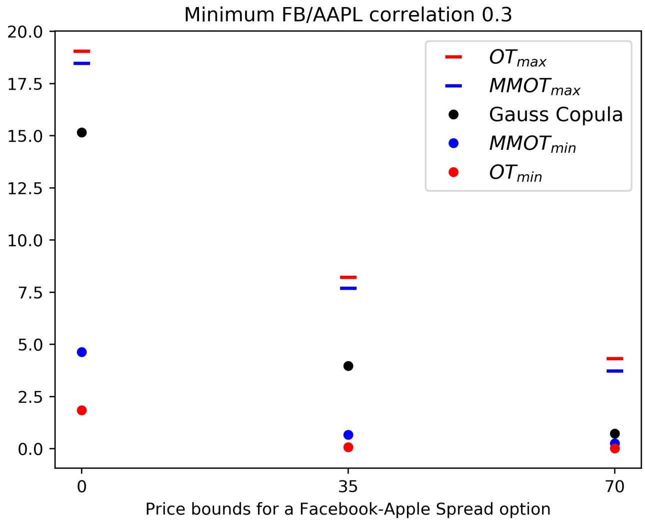
- Data recorded on 16/11/2018:
  - Spot prices $F_0 = 140$, $A_0 = 194$ for Facebook and Apple
  - Call/Puts prices for Facebook and Apple maturing $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$
- Beliefs: bounds on correlation between Facebook and Apple

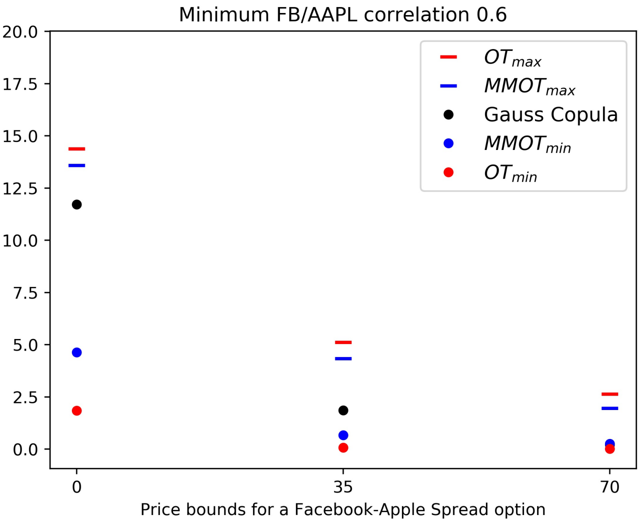
OUTPUTS:

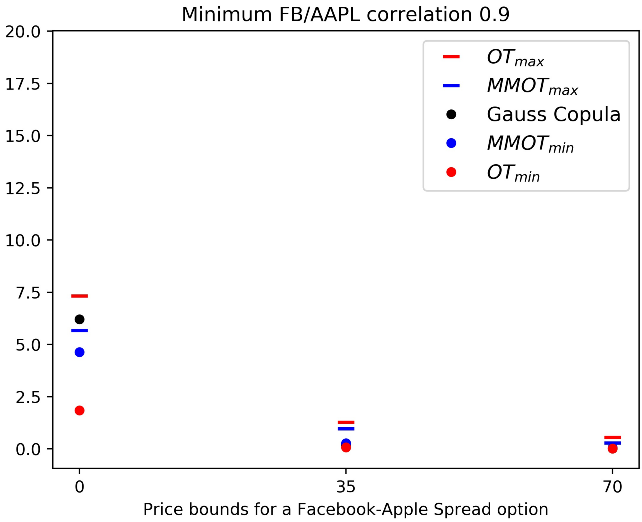
- Range of no-arbitrage prices for a spread option:

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

- 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}

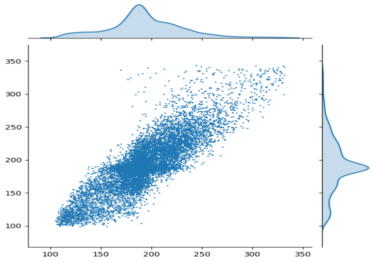

Minimum FB/AAPL correlation 0.6
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
OT_{max}
$MMOT_{max}$
Gauss Copula
$MMOT_{min}$
OT_{min}
0
35
70
Price bounds for a Facebook-Apple Spread option

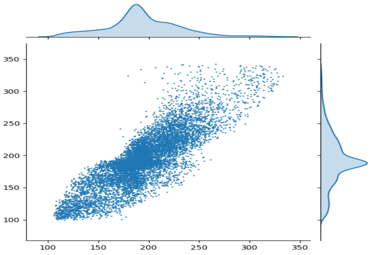


Minimum FB/AAPL correlation 0.9
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
35
70
Price bounds for a Facebook-Apple Spread option
OT_{max}
$MMOT_{max}$
Gauss Copula
$MMOT_{min}$
OT_{min}

Temporal evolution under Extreme models

UNIVERSITY OF OXFORD
Mathematical Institute



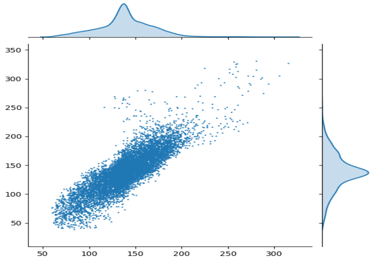


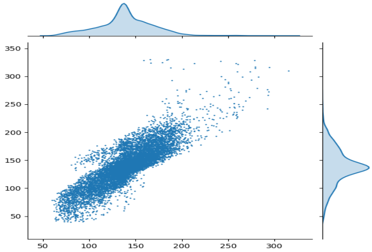
Joint distribution of (A_{T_1}, A_{T_2}), for the Minimiser and Maximiser $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$, $K = 35$ and $\rho \geq 0.6$ and

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

Temporal evolution under Extreme models

UNIVERSITY OF OXFORD
Mathematical Institute



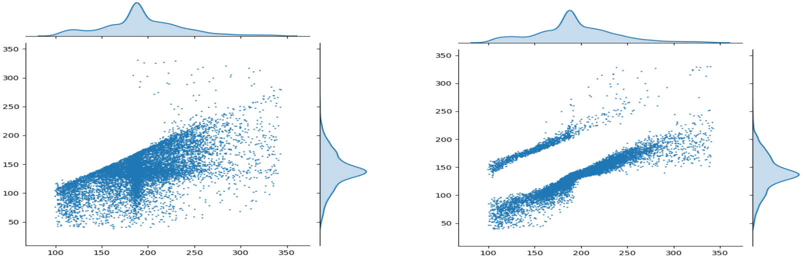


Joint distribution of (F_{T_1}, F_{T_2}), for the Minimiser and Maximiser $T_1 = 18/04/2019$ and $T_2 = 21/06/2019$, $K = 35$ and $\rho \geq 0.6$ and

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

Dependence Structure under Extreme models

UNIVERSITY OF OXFORD
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

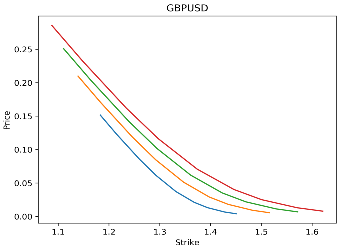
$$\xi = \left(F_{T_2} - \frac{F_0}{A_0} A_{T_2} - K\right)^+$$

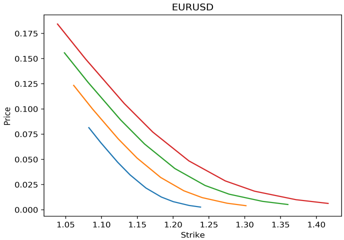
An FX Toy Example

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



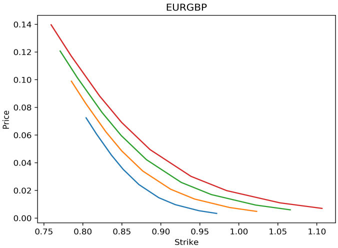


An FX Toy Example

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



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

An FX Toy 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

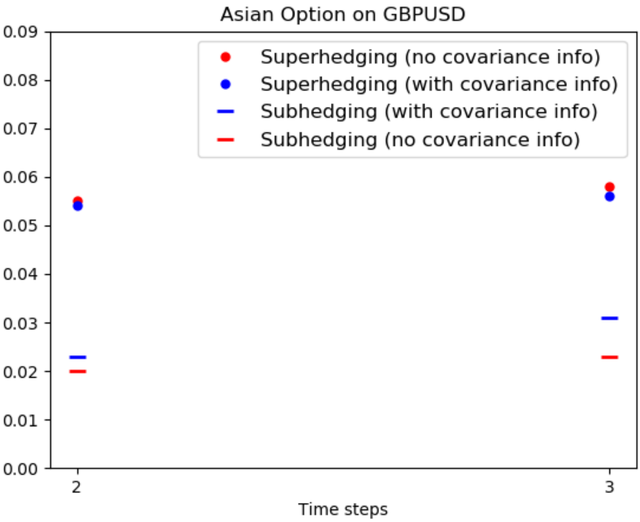
OUTPUTS:

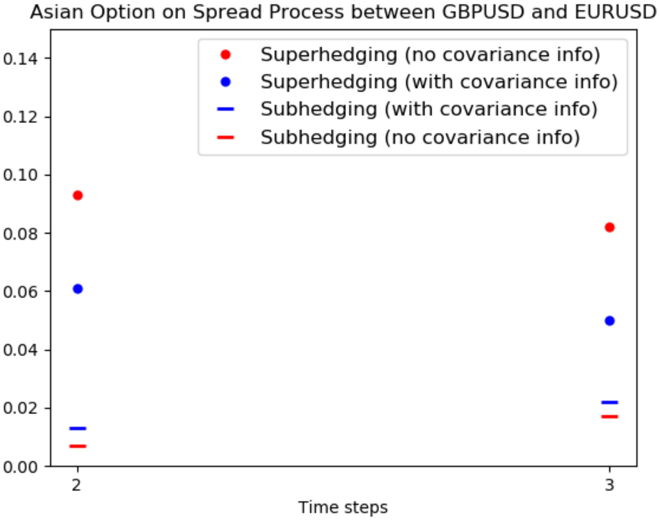
- Range of no-arbitrage prices for:

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

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
Superhedging (no covariance info)
Superhedging (with covariance info)
Subhedging (with covariance info)
Subhedging (no covariance info)
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
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

Mathematical
Institute

Non-parametric calibration

based on joint works with Ivan Guo, Grégoire Loeper and Leo Wang
see *SIAM J. Financial Math. (2021), Risk Magazine (2022)*.

- 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.

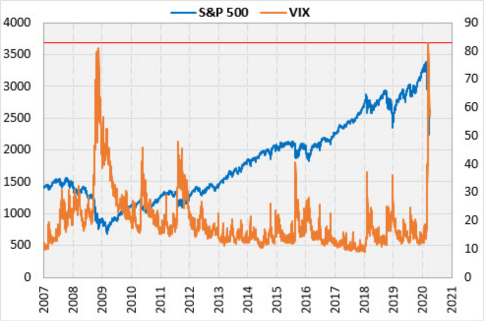


Figure: Historical SPX and VIX data. (Source: Schaeffer's Investment Research)

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.

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

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

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 $\mathcal{P}(\mu_0, \mu_1)$

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

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 $\tau \le T$,

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

UNIVERSITY OF OXFORD
Mathematical Institute

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:

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

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

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

Numerical method: solving the dual formulation

Dual formulation (via Fenchel–Rockafellar):

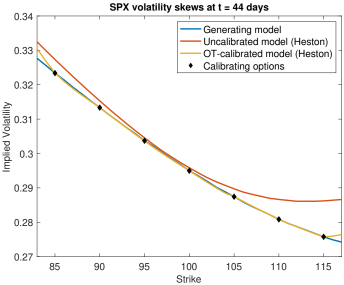
$$\text{maximise} \quad V = \sup_{\lambda \in \mathbb{R}^{m+n+2}} \lambda \cdot c - \phi^{\lambda}(0, X_0),$$

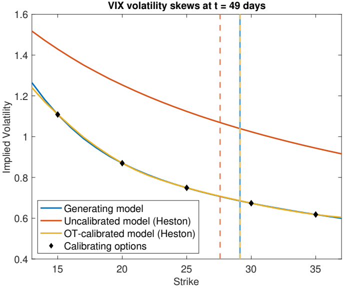
$$\text{subject to} \quad \partial_t \phi^{\lambda} + F^*(\nabla_x \phi^{\lambda}, \frac{1}{2}\nabla_x^2 \phi^{\lambda}) = -\sum_{i=1}^{m+n+2} \lambda_i \mathcal{G}_i \delta(t - \mathcal{T}_i), \quad \phi(T, \cdot) = 0.$$

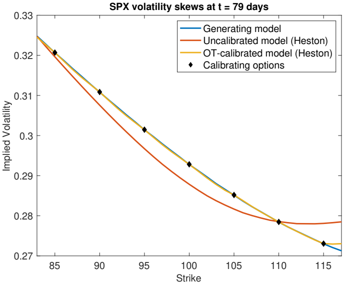
Numerical solution:

1. Set an initial λ (e.g., $\lambda = 0$),
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.

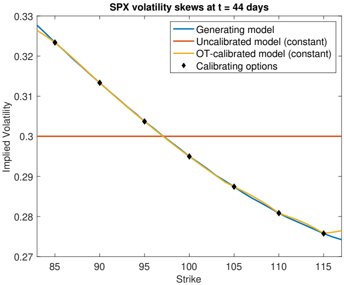
This is analogous to the one dimensional case in Avellaneda, Friedman, Holmes and Samperi (1997)! Therein motivated by minimising a relative entropy–like functional.

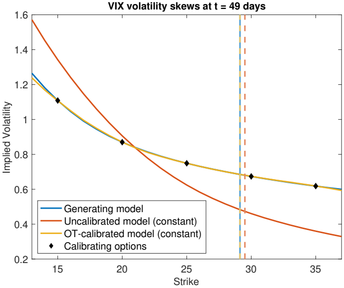


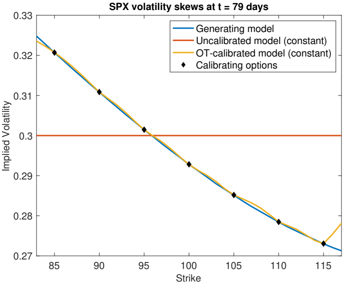




Simulated data example — Calibration results for Heston reference



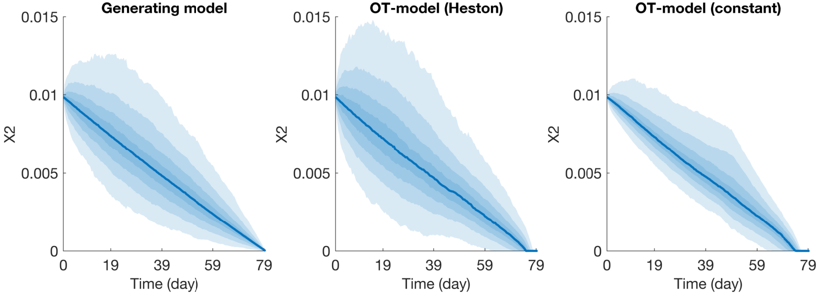




UNIVERSITY OF OXFORD
Mathematical Institute

Simulated data example — Calibration results for constant reference

Simulated data example — Simulation of X^2



Market data example

Mathematical Institute

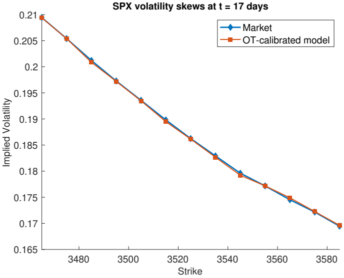
Market data as of 1st September 2020:

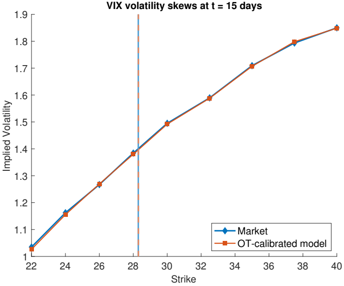
- 26 SPX call options maturing at 17 days and 45 days
- 1 VIX futures maturing at 15 days
- 9 VIX call option maturing at 15 days

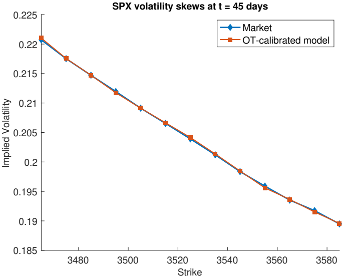
These are the shortest maturities, which is known as the most challenging case!

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.



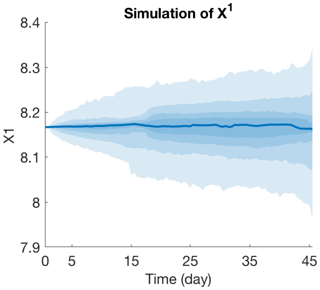




Market data example — Calibration results

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

UNIVERSITY OF OXFORD
Mathematical Institute



Conclusion & Outlook

University of Oxford
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. Worth exploring!

Thank You

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