Data driven robustness and uncertainty sensitivity analysis

Jan Obłój
Mathematical Institute
University of Oxford

joint work with
Daniel Bartl, Samuel Drapeau and Johannes Wiesel

BFS One World Seminar, 21 May 2020

Oxford Mathematics
St John's College
Oxford

University of Oxford
Mathematical Institute

European Research Council
Established by the European Commission

Problem Setting

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: $\inf_{a\in\mathcal{A}} \mathbb{E}_{\mathbb{P}}[-U(\langle a, R_T\rangle)]$,
- optimised certainty equivalents: $\inf_{a\in\mathbb{R}} \mathbb{E}_{\mathbb{P}}[a - U(X+a)]$
- OLS regression: $\inf_{a\in\mathbb{R}^d} \frac{1}{N}\sum_{i=1}^{d}(y^i - \langle a, x^i\rangle)^2$,
- 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.

We propose to compute

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

with $B_\delta(\mu)$ being Wasserstein balls around μ.

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

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) log-normal.

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

Parametric Approach

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

Then

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

Non-parametric Approach

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

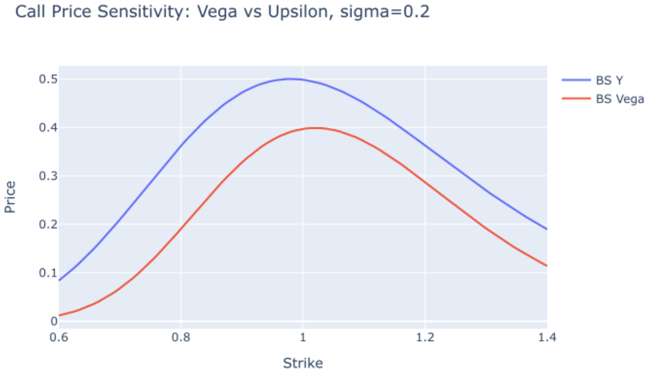
Then

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

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.



DRO & Wasserstein distances

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 contraints
- density constraints
- Prokhorov distance
- Hellinger distance
- Kullback–Leibler divergence/entropy bounds
- and more...

Wasserstein distance

Mathematical Institute

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

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

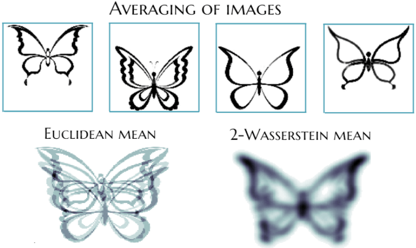
Note that, for a random variable $X \sim \mu$, on some $(\Omega, \mathbb{F}, \mathbb{P})$ we have

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

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

Observe historical returns $r^1, \dots, r^N$ assumed to follow a time-homogeneous ergodic Markov chain on $\mathbb{R}^d$ with an invariant distribution μ. Should we work with

the data points $(r^i)_{i=1}^N$ or the empirical measure $\hat{\mu}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{r^i}$?



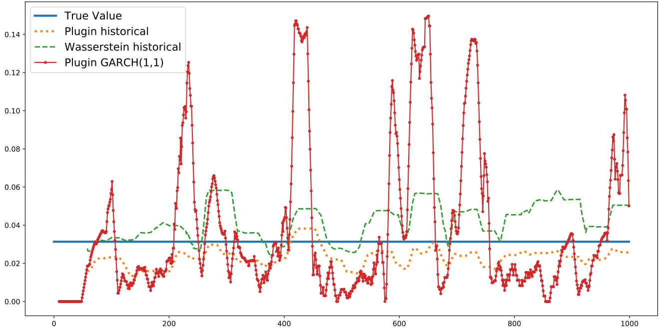
Source: J. Ebert, V. Spokoiny, A. Suvorikova arXiv:1703.03658

We want to estimate

$$\pi_\alpha^\mu(\xi) = \inf_{H \in \mathbb{R}^d} \mathrm{AV@R}_\alpha^\mu(\xi - \langle H, r - 1 \rangle).$$

Estimates for $\pi^{\mu}_{0.95}((r-1)^+)$

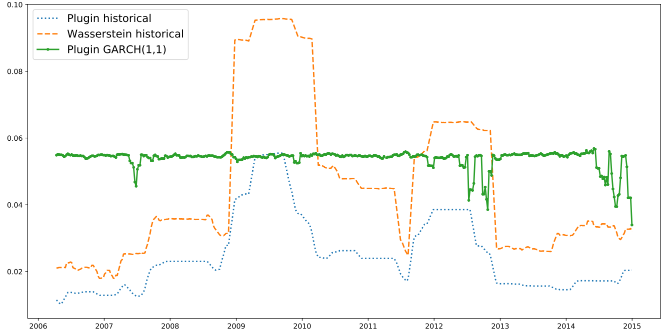
UNIVERSITY OF OXFORD
Mathematical Institute



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

Estimates for $\pi^{\mu}_{0.95}((r-1)^+)$

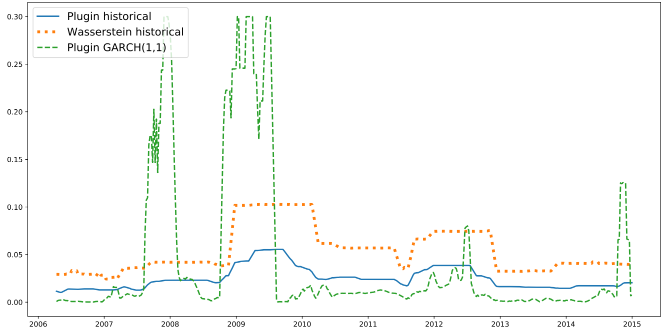
UNIVERSITY OF OXFORD
Mathematical Institute



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

Estimates for $\pi^{\mu}_{0.95}((r-1)^+)$

UNIVERSITY OF OXFORD
Mathematical Institute



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

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)

More on Wasserstein data-driven estimators for risk measures in:
O. & Wiesel, Robust estimation of superhedging prices, *Ann. Stat.* (forthcoming), arXiv:1807.04211

Large uncertainty limit

University of Oxford
Mathematical Institute

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

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

and show that

$$a^*(\delta) \stackrel{\delta\to\infty}{\longrightarrow} \left(\frac{1}{N},\ldots,\frac{1}{N}\right)$$

which may not be true for weaker or stronger metrics.

Main results

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

Oxford Mathematical Institute

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 gain 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

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

And its robust version reads

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

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

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

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

Example 2: Mean-variance optimal investment

UNIVERSITY OF OXFORD
Mathematical Institute

Consider $X \sim \mu$ vector of returns in $\mathbb{R}^d$ and $\mathcal{A} = \{a : \langle a, 1\rangle = 1\}$.

$$V(0) = \inf_{a\in\mathcal{A}} \mathbb{E}[\langle a, X\rangle] + \gamma \mathrm{VAR}_\mu(\langle a, X\rangle) = \inf_{a\in\mathcal{A}} \sup_{Z:\mathbb{E}[Z]=1,\mathbb{E}[Z^2]=1+\gamma^2} \mathbb{E}\Big[\langle a, X\rangle Z\Big]$$

And its robust version, for $p = q = 2$, reads

$$V(\delta) = \inf_{a\in\mathcal{A}} \sup_{(\xi,Z):\mathbb{E}[\langle\xi,\xi\rangle]\leq\delta^2,\mathbb{E}[Z]=1,\mathbb{E}[Z^2]=1+\gamma^2} \mathbb{E}\Big[\langle a, X+\xi\rangle Z\Big]$$

A two-step computation recovers the result in Pflug et al. '12:

$$\Upsilon = |a^*|\sqrt{1+\gamma^2}.$$

Example 3: NN & adversarial examples

Mathematical Institute

Most works focus on explaining the effects and creating algos to create adversarial examples, see SZEGEDY ET AL. '13, BASTANI ET AL. '16

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

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

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 $\nu \in B_\delta(\mu)$ given by:

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

Example 4: UQ: Certification Problem

Consider a system governed by G, μ represents the input parameters. Given a convex set E of undesirable outcomes, one wants to control $\sup_{\nu\in\mathcal{P}} \nu(G(x)\in E)$. Consider a regularised version

$$\max_{\delta} \quad s.t. \quad \inf_{\nu\in B_\delta(\mu)} \int d(G(x),E)\,\nu(dx) \geq \alpha,$$

for a given safety level α. We obtain

$$\inf_{\nu\in B_\delta(\mu)} \int d(G(x),E)\,\nu(dx) = \int d(G(x),E)\,\mu(dx) - \left(\int |\nabla_x d(G(x),E)\nabla_x G(x)|^q\,\mu(dx)\right)^{1/q} \delta + o(\delta).$$

In the special case $\nabla_x G(x) = cI$ this simplifies to

$$\int d(G(x),E)\,\mu(dx) - c\,(\mu(G(x)\notin E))^{1/q}\,\delta + o(\delta)$$

and recovers the intuition of Thm 1 in Chen, Kuhn & Wiesemann '18.

University of Oxford Mathematical Institute

Robust call pricing: martingale constraint

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

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

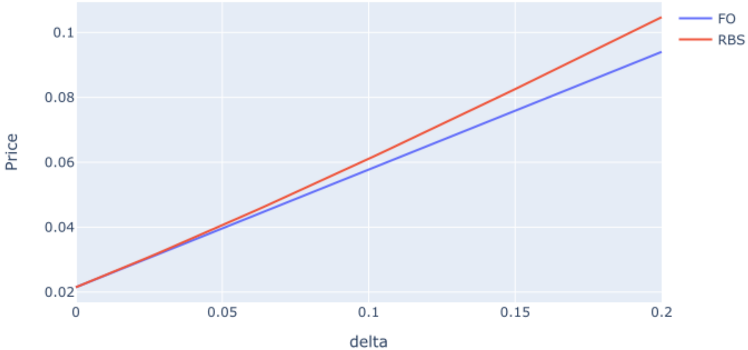
so that $\mathcal{R}BS(0) = BSCall(S_0, K, \sigma)$. For $p = 2$ we find

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

Robust call: numerics

Mathematical Institute

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

Sensitivity of optimisers

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)

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

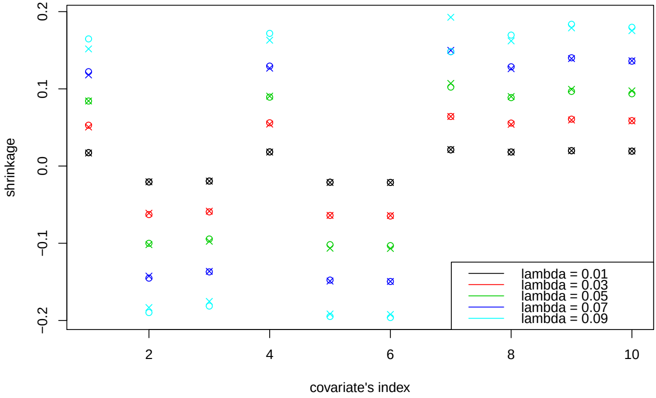
where $1/r+1/s=1$. $\mu=\hat{\mu}_N=\frac{1}{N}\sum_{i=1}^N \delta_{(x^i,y^i)}$ encodes the observations.
System is overdetermined so that $D=\int xx^T\mu(dx)$ is invertible.
$\delta=0$ case is the ordinary least squares regression: $a^*=\frac{1}{N}D^{-1}\int yxd\mu$.
$\delta>0$, $s=1 \rightsquigarrow$ RHS = square-root LASSO regression
$\delta>0$, $s=2 \rightsquigarrow$ RHS $\approx$ Ridge regression
Then $a^*(\delta)$ is approximately, for $s=1$ and $s=2$:

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

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: Influence curves (IC)

For a functional $\mu \mapsto T(\mu)$ its IC is defined as

$$\mathrm{IC}(y) = \lim_{t \to 0} \frac{T(t\delta_y + (1-t)\mu) - T(\mu)}{t}.$$

For M-estimators, defined as optimisers for $V(0)$ for some f,

$$T(\mu) := \mathrm{argmin}_a \int f(x,a)\mu(dx) \quad \Rightarrow \quad \mathrm{IC}(y) = \frac{\nabla_a f(y, T(\mu))}{-\int \nabla_a^2 f(s, T(\mu))\, \mu(ds)}.$$

Taylor-expand $\mathrm{IC}(y)$, take $y = x + \delta \nabla f_x(x, T(\mu))$ and integrate:

$$\int \frac{\mathrm{IC}(x + \delta \nabla_x f(x, T(\mu)) - \mathrm{IC}(x)}{\delta}\, \mu(dx) \approx \frac{T^\delta - T(\mu)}{\delta}.$$

Example 3: Robust Marginal Utility Price

Fix $\delta \geq 0$, $p > 1$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$. Define

$$W_P(\varepsilon) := \inf_{h \in \mathbb{R}^k} \sup_{\nu \in B_\delta(\mu)} \int l\left(x_0 - \varepsilon + h(x - x_0) + \frac{\varepsilon}{P} B\right) d\nu(x),$$

for an initial wealth $x_0 \in \mathbb{R}$ and an option payoff $B : \mathbb{R}^d \to \mathbb{R}$.

The unique $\hat{P}$, which satisfies

$$\nabla_\varepsilon W_{\hat{P}}(0) = 0.$$

is called the robust marginal utility price. For $\delta = 0$ it is the Davis' price. For $\delta > 0$, we show it solves

$$\inf_{h^* \in H^{\mathrm{opt}}(0)} \sup_{\nu \in MB_0^p(\mu, h^*)} \int l'(x_0 + h^*(x - x_0)) \left(-1 + \frac{B}{\hat{P}}\right) \nu(dx) = 0.$$

Conclusion & Outlook

University of Oxford
Mathematical Institute

- Wasserstein balls capture model uncertainty well, small and large uncertainty alike
- DRO conceptually appealing
- We provide first order sensitivities to model uncertainty
  - of the value function
  - and of the optimisers
- Results for general action and state spaces, generic semi-norms and also under further linear constraints on measures in the Wasserstein balls
- Applications in finance, statistics, UQ, ML and more!

Thank You

preprint coming soon at

www.maths.ox.ac.uk/people/jan.obloj