Data-Driven Optimization with Distributionally Robust Second-Order Stochastic Dominance Constraints

Erick Delage

HEC Montréal & GERAD, Montréal, Canada
erick.delage@hec.ca

(Joint work with Chun Peng)

INFORMS Annual Meeting
October 18th, 2022
Stochastic Dominance

Definition (Second-Order Stochastic Dominance, SOSD)

Given any two random variables $X$ and $Y$ capturing some earnings, $X$ stochastically dominates $Y$ in the second-order, $X \succeq (2) Y$, if and only if

$$
\int_{-\infty}^{\eta} F_X(t) \, dt \leq \int_{-\infty}^{\eta} F_Y(t) \, dt, \forall \eta \in \mathbb{R},
$$

where $F_X(t) = P(X \leq t)$.

Equivalent representations:

$\>
X \succeq (2) Y \Leftrightarrow \mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all non-decreasing concave functions $u$.

$\>
X \succeq (2) Y \Leftrightarrow \mathbb{E}[\eta - X]^+] \leq \mathbb{E}[\eta - Y]^+], \forall \eta \in \mathbb{R}.$
Optimization with SOSD Constraints

Consider the SOSD Constrained Problem\(^1\):

\[
\begin{align*}
[SOSDCP] \quad & \text{minimize} \quad c^\top x \\
& \text{subject to} \quad f(x, \xi) \succeq^p f_0(\xi). 
\end{align*}
\]

► \(f(x, \xi)\) is the random controlled performance function, and \(f_0(\xi)\) is the random reference performance function: e.g., \(f_0(\xi) := f(x_0, \xi)\) with \(x_0 \in \mathcal{X}\).

► E.g. SOSD constrained portfolio optimization problem:

\[
\begin{align*}
\text{maximize} \quad & \mathbb{E}_p[\xi]^\top x, \quad \text{s.t.} \quad \xi^\top x \succeq^p \xi^\top x_0. \\
\text{subject to} \quad & x: 1^\top x = 1, x \geq 0
\end{align*}
\]

\(^1\)Dentcheva and Ruszczyński. 2003
Distributionally Robust Stochastic Dominance

Definition (Distributionally Robust Second-Order Stochastic Dominance, DRSOSD)

Given two random variables $X$ and $Y$, we say that $X$ robustly stochastically dominates $Y$ in the second order if and only if:

$$X \succeq^P (2) Y \quad \forall P \in \mathcal{P},$$

where $X \succeq^P (2) Y$ refers to the fact that $X$ stochastically dominates $Y$ in the second-order when the probability measure is $P$.  

---

\footnote{[Dentcheva and Ruszczyński 2010]}
Data-Driven DRSOSD using Wasserstein ambiguity

Consider DRSOSD constraint under a type-1 Wasserstein Ambiguity Set:

\[
\begin{align*}
[WDRSOSDCP] & \quad \min_{x \in X} \quad c^\top x \\
& \quad \text{subject to } f(x, \xi) \succeq^p f_0(\xi) \quad \forall P \in \mathcal{P}_W^1(\hat{P}, \epsilon),
\end{align*}
\]

where $\hat{P}$ is the empirical distribution of $M$ i.i.d. observations.

**Definition (Type-1 Wasserstein Ambiguity Set)**

The type-1 Wasserstein ambiguity set\(^a\) of radius $\epsilon$ centered at $\hat{P}$ is defined by

\[
\mathcal{P}_W^1(\hat{P}, \epsilon) := \left\{ P \in \mathcal{M}(\Xi) \mid d_W^1(P, \hat{P}) \leq \epsilon \right\},
\]

where $\mathcal{M}(\Xi)$ is the space of all distributions supported on $\Xi$ and $d_W^1$ is the Wasserstein metric.

\(^a\)[Esfahani and Kuhn. 2018]
Special Cases of WDRSOSDCP

Proposition (Reduction to SOSDCP)

\[
\text{WDRSOSDCP with } \mathcal{P} := \mathcal{P}_W^{1}(\hat{\mathcal{P}}, 0) \text{ reduces to SOSDCP with } \mathcal{P} := \hat{\mathcal{P}}.
\]
Special Cases of WDRSOSDCP

Proposition (Reduction to SOSDCP)

\[ WDRSOSDCP \text{ with } \mathcal{P} := \mathcal{P}^1_{\mathcal{W}}(\hat{\mathcal{P}}, 0) \text{ reduces to } SOSDCP \text{ with } \mathbb{P} := \hat{\mathbb{P}}. \]

Proposition (Reduction to Robust Optimization)

\[ WDRSOSDCP \text{ with } \mathcal{P} := \mathcal{P}^1_{\mathcal{W}}(\hat{\mathcal{P}}, \infty) \text{ reduces to the robust optimization problem,} \]

\[
\begin{align*}
\text{minimize} & \quad c^\top x \\
\text{subject to} & \quad f(x, \xi) \geq f_0(\xi) \quad \forall \xi \in \Xi. 
\end{align*}
\]
Outline

Introduction

Axiomatic Motivation

Statistical Properties

Exact Solution Scheme

Numerical Study

Conclusion
Outline

Introduction

Axiomatic Motivation

Statistical Properties

Exact Solution Scheme

Numerical Study

Conclusion
What is the right extension when $\mathbb{P} \in \mathcal{P}$?

[Montes et al. 2014] propose six different extensions of SOSD:

$$
X \succeq^{(4)} Y \iff F_X^{P_1} \succeq^{(2)} F_Y^{P_2}, \forall P_1, P_2 \in \mathcal{P}
$$

$$
X \succeq^{(5)} Y \iff \exists P_1 \in \mathcal{P}, F_X^{P_1} \succeq^{(2)} F_Y^{P_2}, \forall P_2 \in \mathcal{P}
$$

$$
X \succeq^{(6)} Y \iff \forall P_2 \in \mathcal{P}, \exists P_1 \in \mathcal{P}, F_X^{P_1} \succeq^{(2)} F_Y^{P_2}
$$

$$
X \succeq^{(7)} Y \iff \exists P_1, P_2 \in \mathcal{P}, F_X^{P_1} \succeq^{(2)} F_Y^{P_2}
$$

$$
X \succeq^{(8)} Y \iff \exists P_2 \in \mathcal{P}, F_X^{P_1} \succeq^{(2)} F_Y^{P_2}, \forall P_1 \in \mathcal{P}
$$

$$
X \succeq^{(9)} Y \iff \forall P_1 \in \mathcal{P}, \exists P_2 \in \mathcal{P}, F_X^{P_1} \succeq^{(2)} F_Y^{P_2}.
$$

where:

$$
F_1 \succeq^{(2)} F_2 \triangleq \int_{-\infty}^{\eta} F_1(t) \, dt \leq \int_{-\infty}^{\eta} F_2(t) \, dt, \forall \eta \in \mathbb{R}.
$$

Recall, the extension from [Dentcheva and Ruszczyński 2010]:

$$
[\text{DRSOSD}] \quad X \succeq^{(*)} Y \iff F_X^{P} \succeq^{(2)} F_Y^{P}, \forall P \in \mathcal{P}.
$$
Axiomatic Motivation for DRSOSD

- Consider a non-atomic ambiguous probability space \((\Omega, \mathcal{F}, \mathcal{P})\).
- Let \(\mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P}) = \bigcap_{\mathcal{P} \in \mathcal{P}} \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P})\).
- Let \(\mathcal{U} := \{X \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P}) | \exists F_X, F_X = F_X^\mathcal{P}, \forall \mathcal{P} \in \mathcal{P}\}\).
- Non-atomic \(\Rightarrow \forall X, \forall \mathcal{P} \in \mathcal{P}, \exists X^\mathcal{P} \in \mathcal{U}, F_X^\mathcal{P} = F_X^\mathcal{P}\).

Theorem

If the preference relation \(\succeq\) satisfies:

- \((\text{SOSD on } \mathcal{U})\) If \(\{X, Y\} \subset \mathcal{U}\), then \(X \succeq Y \iff F_X \succeq_{(2)} F_Y\).
- \((\text{Ambiguity Monotonicity})\) If \(X^\mathcal{P} \succeq Y^\mathcal{P}\) for all \(\mathcal{P} \in \mathcal{P}\), then \(X \succeq Y\).
- \((\text{Maximal Ambiguity Indecisiveness})\) If \(\exists \mathcal{P} \in \mathcal{P}\) such that \(X^\mathcal{P} \not\succeq Y^\mathcal{P}\), then \(X \not\succeq Y\).

Then, for any random variables \(X, Y \in \mathcal{L}_\infty(\Omega, \mathcal{F}, \mathcal{P})\), we have that \(X \succeq Y\) if and only if \(X \succeq_{(*)} Y\), i.e. \(X \succeq_{(2)} Y\) \(\forall \mathcal{P} \in \mathcal{P}\).
Outline

Introduction

Axiomatic Motivation

Statistical Properties

Exact Solution Scheme

Numerical Study

Conclusion
Data-Driven WDRSOSDCP

- Assumption 1: The feasible set $\mathcal{X}$ is a non-empty convex set and the outcome space $\Xi$ is a non-empty compact convex set.

- Assumption 2: $f(x, \xi)$ and $f_0(\xi)$ are piecewise linear concave in $x$ and $\xi$.

- Assumption 3: $P^1_W(\hat{P}, \epsilon)$ uses the $\ell_1$-norm or $\ell_\infty$-norm as the reference metric.
Finite sample guarantee of WDRSOSDCP solutions

Proposition

Suppose that Assumption 1 holds and that each observations in \(\{\hat{\xi}_i\}_{i=1}^M\) are drawn i.i.d. from some \(\bar{P}\), with \(M \geq 1\) and \(m > 2\). Given some \(\beta \in (0, 1)\), let \(\hat{x}_M\) be the optimal solution of the WDRSOSDCP with ambiguity set \(\mathcal{P}^1_W(\hat{P}, \epsilon_M(\beta))\) where

\[
\epsilon_M(\beta) := \begin{cases} 
\left( \frac{\log(c_1\beta^{-1})}{c_2 M} \right)^{1/\max(m,2)} & \text{if } M \geq \frac{\log(c_1\beta^{-1})}{c_2} \\
\left( \frac{\log(c_1\beta^{-1})}{c_2 M} \right)^{1/a} & \text{otherwise},
\end{cases}
\]

and where \(c_1, c_2,\) and \(a > 1\) are positive constants (see [Esfahani and Kuhn. 2018] for details). One has the guarantee that, with probability larger than \(1 - \beta\), \(\hat{x}_M\) satisfies the SOSD constraint under \(\bar{P}\), i.e., \(f(\hat{x}_M, \xi) \succeq_{(2)}^\bar{P} f_0(\xi)\).
Asymptotic consistency of WDRSOSDCP solutions

**Proposition**

Suppose that assumptions 1 and 2 hold, that \( \mathcal{X} \) is bounded, and that \( \beta_M \in (0, 1) \) satisfies \( \sum_{M=1}^{\infty} \beta_M < \infty \) and \( \lim_{M \to \infty} \epsilon_M (\beta_M) = 0 \). Consider

\[
[\phi\text{-SOSDCP}] \quad \text{minimize} \quad c^\top x
\]

subject to \( \mathbb{E}_{\hat{P}} [(t - f(x, \xi))^+] \leq \mathbb{E}_{\hat{P}} [(t - f_0(\xi))^+] + \phi \quad \forall t \in \mathbb{R}, \)

with \( \phi > 0 \), and assume that Slater’s condition is satisfied. Let \( \{\hat{\xi}_i\}_{i=1}^M \) be i.i.d. from \( \hat{P} \), \( x_M \) be an optimal solution of:

\[
\text{minimize} \quad c^\top x
\]

subject to \( \mathbb{E}_P [(t - f(x, \xi))^+] \leq \mathbb{E}_P [(t - f_0(\xi))^+] + \phi \quad \forall t \in \mathbb{R} \quad \forall P \in \mathcal{P}_W(\hat{P}, \epsilon_M (\beta_M)) \)

and \( \mathcal{X}^* \) be the set of optimal solutions to the \( \phi\)-SOSDCP under the true distribution \( \hat{P} \). Then \( x_M \) converges almost surely to \( \mathcal{X}^* \) as \( M \) goes to infinity.
Outline

Introduction

Axiomatic Motivation

Statistical Properties

Exact Solution Scheme

Numerical Study

Conclusion
Optimization with DRSOSD Constraints

- Relevant studies (continue):
  - [Guo et al. 2017]: use a discretization scheme to approximate DRSOSD constrained problem under a moment-based ambiguity set.
  - [Kozmík. 2019]: study a portfolio optimization problem with DRSOSD constraints under type-1 Wasserstein ball over the space of $M$-points distributions, and derive a conservative approximation.
  - [Sehgal and Mehra. 2020]: study a robust portfolio optimization problem with SOSD where scenario perturbations lie within a budgeted uncertainty set.
  - [Mei et al. 2022] study independently the WDRSOSDCP and propose a novel split-and-dual decomposition framework
Multistage Robust Optimization Reformulation

Proposition

**WDRSOSDCP** with $\epsilon \in (0, \infty)$ coincides with the optimal value of the following multistage robust optimization problem:

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad L(x, t) \leq 0 \quad \forall t \in \bar{T},
\end{align*}
\]

where

\[
\begin{align*}
L(x, t) & := \inf_{\lambda, q} \lambda \epsilon + \frac{1}{M} \sum_{i=1}^{M} q_i \\
\text{s.t.} & \quad g(x, \xi, t) - \lambda \|\xi - \hat{\xi}_i\| \leq q_i \quad \forall i \in [M], \xi \in \Xi \\
& \quad \lambda \geq 0, q \in \mathbb{R}^M,
\end{align*}
\]

where

\[
g(x, \xi, t) := (t - f(x, \xi))^+ - (t - f_0(\xi))^+ \quad \text{and} \quad \bar{T} := [\inf_{\xi \in \Xi} f_0(\xi), \sup_{\xi \in \Xi} f_0(\xi)].
\]

- Multistage robust optimization problem: $\min_{x} \sup_{t} \inf_{\lambda, q} \sup_{\xi}$.
- Multistage robust linear optimization problem under Assumption 3 and when $\mathcal{X}$ and $\Xi$ are polyhedral.
An Exact Solution Scheme
Inspired by [Postek and Hertog. 2016] and [Bertsimas and Dunning. 2016]

**Algorithm** Iterative Partition based Solution Algorithm

1: **Initialize:** $LB^0 = -\infty$, $UB^0 = +\infty$, $\ell = 1$, $\hat{T}^0 := \emptyset$, $\varepsilon$.
2: **Initialize:** $\mathcal{P}^1 := \{\overline{T}\}$, $\overline{T} := [\inf_{\xi \in \Xi} f_0(\xi), \sup_{\xi \in \Xi} f_0(\xi)]$.
3: **while** $|(UB^{\ell-1} - LB^{\ell-1})/UB^{\ell-1}| > \varepsilon$ **do**
4: Solve an upper bound problem with the partition $\mathcal{P}^\ell$ and linear decision rules.
5: Identify the optimal solution $(x^*, \lambda^*, q^*, \overline{q}^*)$ and optimal objective $UB^\ell$.
6: Calculate an active scenarios set $\hat{A}^\ell$.
7: Construct the finite scenarios set $\hat{T}^\ell := \hat{A}^\ell \cup \hat{T}^{\ell-1}$.
8: Solve a lower bound problem with $\hat{T}^\ell$ and identify the new $LB^\ell$.
9: **Update the partitions** $\mathcal{P}^{\ell+1} := \nu(\mathcal{P}^\ell, \hat{A}^\ell)$, and $\ell := \ell + 1$.
10: **end while**
11: **return** optimal objective value $z^*$ and optimal solution $(x^*, \lambda^*, q^*, \overline{q}^*)$. 
Outline

Introduction

Axiomatic Motivation

Statistical Properties

Exact Solution Scheme

Numerical Study

Conclusion
Application to Portfolio Optimization

DRSOSD constrained portfolio optimization problem with uncertain returns $\xi$:

$$\max_{x \in \mathcal{X}} \mathbb{E}_P[\xi]^\top x$$

subject to $\xi^\top x \succeq_P (2) \xi^\top x_0$ \quad $\forall P \in \mathcal{P}_W(\hat{P}, \epsilon),$

where $\mathcal{X} := \left\{ x \in \mathbb{R}^m \mid \sum_{j=1}^m x_j = 1, x_j \geq 0, \forall j \in \{1, \ldots, m\} \right\}.$

- $x_0$ is a reference portfolio.
- Assume $\Xi$ to be a box, i.e. $\Xi := \{\xi^- \leq \xi \leq \xi^+\}.$

We experiment with both synthetic data and real stock market data.
Real Stock Data: Calibration of $M$ and $\epsilon$

- In-sample data: Weekly stock returns from 335 companies from S&P 500 over Jan 1994 - Dec 2013
- Portfolio optimization over 5 randomly picked stocks with $x_0$ as the equally weighted portfolio
- Cross-validation of look-back period and Wasserstein ball size based on distribution of next 26 weekly returns

![Graph showing average expected return vs. size of Wasserstein ball for different M values.

![Graph showing estimated feasibility frequency vs. size of Wasserstein ball for different M values.]}
Real Stock Data: Out-of-sample performance

- Out-of-sample data: Weekly stock returns from 257 companies from S&P 500 over Jan 2014 - Dec 2019
- SOSDCP uses lookback of $M = 52$ weeks
- WDRSOSDCP uses $M = 52$ and $\epsilon = 0.01$
- Performance is averaged over 1000 runs.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average expected return</td>
<td>0.183</td>
<td>0.190</td>
<td>0.184</td>
<td>-</td>
</tr>
<tr>
<td>Average standard deviation</td>
<td>0.032</td>
<td>0.029</td>
<td>0.022</td>
<td>-</td>
</tr>
<tr>
<td>Average CVaR (conf. 90%)</td>
<td>0.054</td>
<td>0.048</td>
<td>0.037</td>
<td>-</td>
</tr>
<tr>
<td>SOSD feasibility frequency</td>
<td>87%</td>
<td>96%</td>
<td>100%</td>
<td>94%</td>
</tr>
</tbody>
</table>
Outline

Introduction

Axiomatic Motivation

Statistical Properties

Exact Solution Scheme

Numerical Study

Conclusion
Concluding Remarks

- We provide an axiomatic motivation for distributionally robust version of the second-order stochastic dominance ordering
- We show that the data-driven WDRSOSDCP can provide finite sample guarantees and asymptotic consistency
- We develop an efficient exact solution scheme, an iterative partition-based solution algorithm.
- We show how out-of-sample SOSD feasibility can be improved by carefully adjusting the level of robustification without sacrificing much objective performance.
Questions & Comments...

Our paper is available on Optimization Online via

...Thank you!
Bibliography


Consider 3 assets (asset #1, #2 and #3) that are independent from each other, while their respective marginal distribution of return is such that $\xi_3 \succ (\bar{P}) \xi_1 \succ (\bar{P}) \xi_2$ and $E[\bar{P}[\xi_3]] > E[\bar{P}[\xi_2]] = E[\bar{P}[\xi_1]]$.

$x_0 := [1, 0, 0]$, invest all the resources in project #1.

Generate $M \in \{10, 100, 1000\}$ i.i.d. in-sample data.

Results are averaged over 100 runs.
- The estimated out-of-sample feasibility frequency shows the probability of obtaining a WDRSOSDCP solution that satisfies the SOSD constraint in out-of-sample test.
Synthetic Data: Out-of-Sample Feasibility of SOSD

- The estimated out-of-sample feasibility frequency shows the probability of obtaining a WDRSOSDCP solution that satisfies the SOSD constraint in out-of-sample test.
Real Stock Data: Computational Performance

Table: The average computational performance with respect to different $\epsilon$, number of stocks ($m \in \{10, 50, 100\}$) and in-sample sizes ($M \in \{50, 100\}$), in terms of average CPU time (Time, in seconds), proportion of unsolved instances (prop, in %), and average number of iterations (Iter) over 20 runs.

<table>
<thead>
<tr>
<th>$M$</th>
<th>$\epsilon$</th>
<th>$m=10$</th>
<th></th>
<th>$m=50$</th>
<th></th>
<th>$m=100$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Time</td>
<td>prop</td>
<td>Iter</td>
<td>Time</td>
<td>prop</td>
<td>Iter</td>
</tr>
<tr>
<td>50</td>
<td>0.0100</td>
<td>242</td>
<td>0</td>
<td>6.0</td>
<td>1997</td>
<td>0.05[1.9]</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>0.0464</td>
<td>103</td>
<td>0</td>
<td>5.0</td>
<td>3418</td>
<td>0.10[1.4]</td>
<td>6.0</td>
</tr>
<tr>
<td>100</td>
<td>0.0100</td>
<td>1528</td>
<td>0</td>
<td>6.0</td>
<td>6259</td>
<td>0.05[3.0]</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>0.0464</td>
<td>506</td>
<td>0</td>
<td>5.0</td>
<td>5966</td>
<td>0.25[3.8]</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>595</td>
<td>0</td>
<td>5.5</td>
<td>4410</td>
<td>0.11[2.5]</td>
<td>5.8</td>
</tr>
</tbody>
</table>

[···]: the average sub-optimality gap (in %) for the unsolved instances within 2 hours limit.

Based on our test from the previous experiment, the midrange values of $\epsilon$ (i.e., $\epsilon \in \{0.0100, 0.0464\}$) appeared to be the hardest to handle.