(Data-Driven) Conditional Robust Optimization

Abhilash Chenreddy   Erick Delage

HEC Montréal & GERAD
Montréal, Canada

Seminar in honour of Yinyu's career achievements
Stanford University
July 29th, 2024

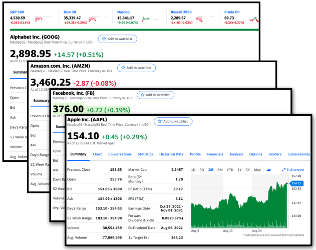
Canada Research Chairs   Chaires de recherche du Canada
Canada

Portfolio optimization with contextual information

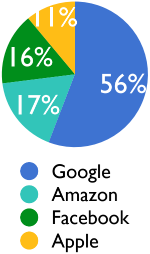
Problem: How to invest wealth among a set of assets?

$$x^*(\psi) := \arg \min_{x:\sum_{i=1}^n x_i=1,\, x\geq 0} \mathrm{VaR}_{1-\epsilon}(\xi^\top x|\psi)$$

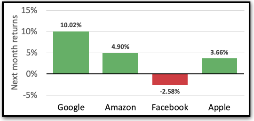
Context (ψ)



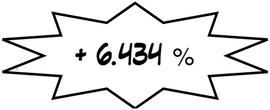
Decision (x)



Outcome (ξ)



Performance $c(x, \xi)$



What is contextual stochastic optimization?

- Optimization problems arising in practice almost always involve **unknown parameters** $\xi \in \mathbb{R}^{m_\xi}$
- Oftentimes, there is a relationship between unknown parameters and some **contextual data** $\psi \in \mathbb{R}^{m_\psi}$

What is contextual stochastic optimization?

- Optimization problems arising in practice almost always involve **unknown parameters** $\xi \in \mathbb{R}^{m_\xi}$
- Oftentimes, there is a relationship between unknown parameters and some **contextual data** $\psi \in \mathbb{R}^{m_\psi}$

- **Contextual Optimization**:
  - Optimizes a policy, $\boldsymbol{x} : \mathbb{R}^{m_\psi} \to \mathcal{X}$
    - I.e., action $x \in \mathcal{X}$ is adapted to the observed context ψ
  - **Contextual Stochastic Optimization** problem minimizes the expected cost of running the policy over the joint distribution of (ψ, ξ):

$$\min_{\boldsymbol{x}(\cdot)} \mathbb{E}[c(\boldsymbol{x}(\psi), \xi)] \Leftrightarrow \quad \boldsymbol{x}^*(\psi) \in \arg\min_{x \in \mathcal{X}} \mathbb{E}[c(x, \xi)|\psi] \text{ a.s.}$$

What is conditional robust optimization?

- We introduce a **Conditional Robust Optimization** model for solving contextual optimization problems in a risk-averse setting:

$$\text{(CRO)} \qquad \boldsymbol{x}^*(\psi) \in \arg\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}(\psi)} c(x, \xi), \forall\, \psi \in \mathcal{V}$$

  where $\mathcal{U}(\psi)$ is a **conditional uncertainty set** designed to contain with high probability the realization of ξ conditionally on observing ψ.

Estimate-then-Optimize with continuous adaptation

- We consider a continuously adapted conditional ellipsoidal set:

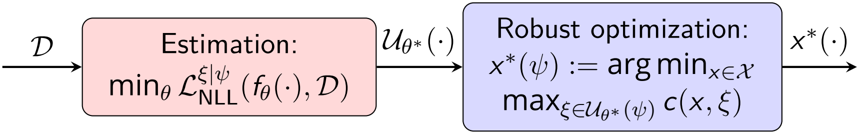
$$\mathcal{U}_\theta(\psi) := \left\{ \xi \in \mathbb{R}^{m_\xi} : (\xi - \mu_\theta(\psi))^T \Sigma_\theta^{-1}(\psi)(\xi - \mu_\theta(\psi)) \leq R_\theta \right\},$$

Estimate-then-Optimize with continuous adaptation

- We consider a continuously adapted conditional ellipsoidal set:

$$\mathcal{U}_\theta(\psi) := \{\ \xi \in \mathbb{R}^{m_\xi} : (\xi - \mu_\theta(\psi))^T \Sigma_\theta^{-1}(\psi)(\xi - \mu_\theta(\psi)) \leq R_\theta \}\,,$$

- Given a data set $\mathcal{D} = \{(\psi_1, \xi_1), (\psi_2, \xi_2) \ldots (\psi_N, \xi_N)\}$, an estimate-then-optimize (ETO) approach takes the form:



where $\mathcal{L}_{\text{NLL}}^{\xi|\psi}$ is the negative log likelihood for a conditional Gaussian density estimator (see Barratt and Boyd [2021]):

$$\xi \sim f_\theta(\psi) := \mathcal{N}(\mu_\theta(\psi), \Sigma_\theta(\psi))$$

and R_θ s.t. $\mathbb{P}_{\mathcal{D}}(\xi \in \mathcal{U}_\theta(\psi)) = 1 - \epsilon$

Related work in operations research literature

- Data-driven Distributionally Robust Optimization:
  - Moment-based DRO: D & Ye (> 2000 citations !!!), etc.
  - Divergence-based DRO: Ben-Tal et al. [2013], Duchi et al. [2021], etc.
  - Wassertein-based DRO: Mohajerin Esfahani and Kuhn [2018], Gao and Kleywegt [2023], etc.
- Contextual Stochastic Optimization:
  - Hannah et al. [2010], Bertsimas and Kallus [2020], ...: Conditional distribution estimation used to formulate and solve the CSO problem.
  - Donti et al. [2017], Elmachtoub and Grigas [2022], ...: End-to-end paradigm applied to solve the data driven CSO problem.
- Distributionally Robust CSO:
  - Bertsimas et al. [2022], McCord [2019], Wang and Jacquillat [2020], Kannan et al. [2020]: DRO approaches with ambiguity sets centered at the estimated conditional distribution
  - Nguyen et al. [2021],Esteban-Pérez and Morales [2022]: Wasserstein DRO for non-parametric CSO

Related work in data-driven robust optimization literature

- Estimate then RO:
  - Goerigk and Kurtz [2023], Johnstone and Cox [2021]: learns non-contextual uncertainty sets using deep learning, and conformal prediction.
  - Chenreddy et al. [2022] learns a contextual uncertainty set using an integrated clustering then classification approach, Blanquero et al. [2023] constructs contextual ellipsoidal sets by making normality assumptions
  - Sun et al. [2024] solves a robust contextual LP problem by calibrating a predictive model to match robust objectives
  - Ohmori [2021], Sun et al. [2023]: calibrates a box/ellipsoidal set to cover the realizations of a kNN/residual-based conditional distribution.
- End-to-end RO
  - Wang et al. [2023] learns non-contextual sets to maximize performance across a set of randomly drawn parameterized robust constrained problems while ensuring constraint satisfaction guarantees w.r.t the joint distribution over perturbance and robust problems
  - Costa and Iyengar [2023]: proposes a distributionally robust end-to-end system that integrates point prediction and robustness tuning to the portfolio construction problem

Our contributions

- Limitation of existing CRO approaches:
  1. Training disregards entirely the out-of-sample performance of the solution obtained from robust optimization.

Our contributions

- Limitation of existing CRO approaches:
  1. Training disregards entirely the out-of-sample performance of the solution obtained from robust optimization.
  2. While the calibration process encourages **marginal coverage**:

     $$\mathbb{P}(\xi \in \mathcal{U}(\psi)) \geq 1 - \epsilon \quad ✔$$

     it does not promote **conditional coverage** over all ψ:

     $$\mathbb{P}(\xi \in \mathcal{U}(\psi)|\psi) \geq 1 - \epsilon \text{ a.s.} \quad ✖$$

Desirable coverage properties for $\mathcal{U}(\psi)$

The field of conformal prediction identifies two important properties for conditional uncertainty sets

- Marginal coverage property: $\mathbb{P}(\xi \in \mathcal{U}(\psi)) \geq 1 - \epsilon$
- Conditional coverage property: $\mathbb{P}(\xi \in \mathcal{U}(\psi)|\psi) \geq 1 - \epsilon$ a.s.

E.g., target coverage $1 - \epsilon = 90\%$:

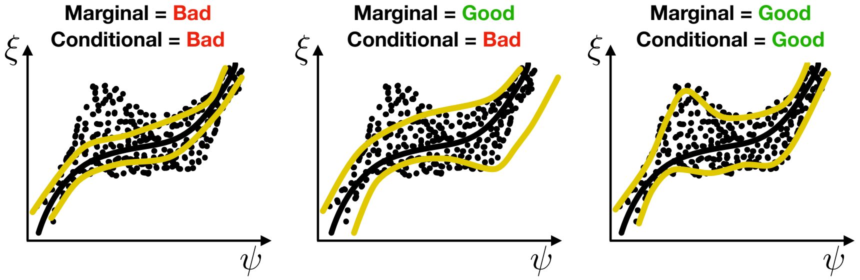


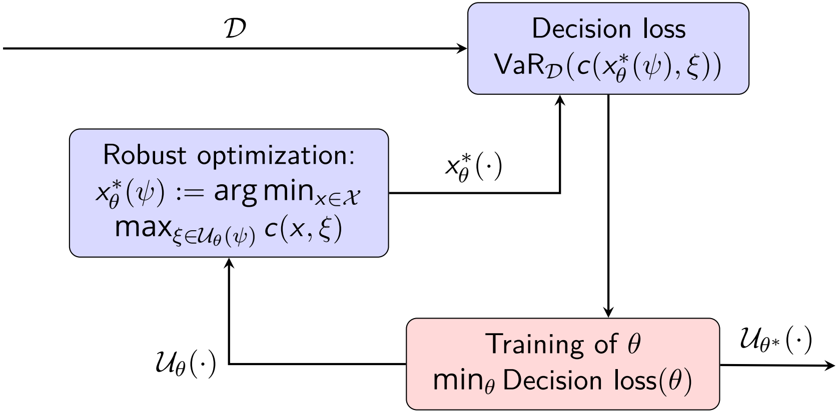
Image from Angelopoulos and Bates, A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification, CoRR, 2021.

Presentation overview

Outline

(Single) Task-based Set (TbS) training

A task-based approach learns the estimator by trying to minimize the decision loss, e.g. the portfolio risk based on VaR



Decision loss relaxation and derivatives

- Decision loss $\text{VaR}_{\mathcal{D}}(c(x_\theta^*(\psi), \xi))$ suffers from multiple local optima.

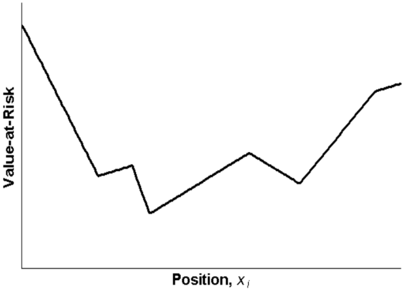


Figure 3: Simulation-based trade risk profile

Image from Mausser and Rosen, Beyond VaR: from measuring risk to managing risk, CIFEr, 1999.

Decision loss relaxation and derivatives

- Decision loss $\text{VaR}_{\mathcal{D}}(c(x^*_\theta(\psi), \xi))$ suffers from multiple local optima.

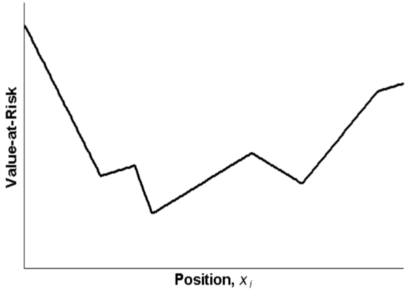


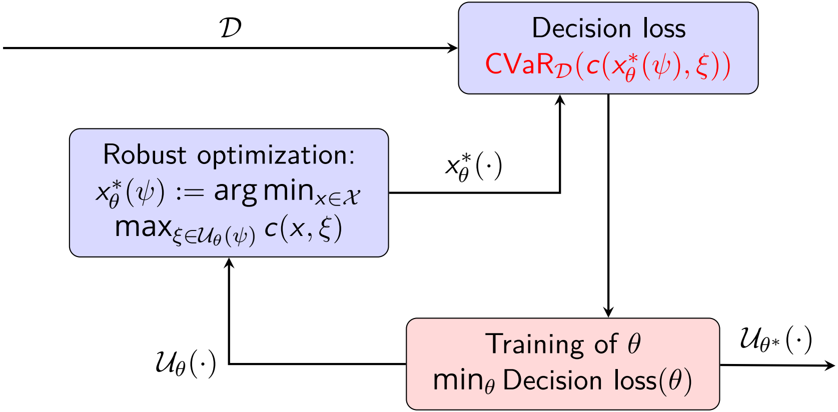
Figure 3: Simulation-based trade risk profile

Image from Mausser and Rosen, Beyond VaR: from measuring risk to managing risk, CIFEr, 1999.

- We therefore replace it with upper bound $\text{CVaR}_{\mathcal{D}}(c(x^*_\theta(\psi), \xi))$.

$$\frac{\partial \text{CVaR}_{i\sim N}(y_i)}{\partial y_i} = v_i(y) \text{ with } \boldsymbol{v}(\boldsymbol{y}) \in \underset{\boldsymbol{v}\in\mathbb{R}^M_+ : \mathbb{1}^T\boldsymbol{v}=1, \boldsymbol{v}\leq((1-\alpha)N)^{-1}}{\text{argmax}} \boldsymbol{v}^T\boldsymbol{y}$$

Decision loss relaxation and derivatives



Robust optimization reformulation and derivatives

- We assume that $c(x, \xi)$ is convex in x and concave in ξ, while $\mathcal{X}$ is a convex set.

Robust optimization reformulation and derivatives

- We assume that $c(x, \xi)$ is convex in x and concave in ξ, while $\mathcal{X}$ is a convex set.
- Using Fenchel duality, one can follow Ben-Tal et al. [2015] to reformulate the robust optimization problem as:

$$x^*_\theta(\psi) := \arg\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}_\theta(\psi)} c(x, \xi) = \arg\min_{v, x \in \mathcal{X}} \underbrace{\delta^*(v|\mathcal{U}_\theta(\psi)) - c_*(x, v)}_{f(x, v, \mathcal{U}_\theta(\psi))}$$

where the support function

$$\delta^*(v|\mathcal{U}_\theta(\psi)) := \sup_{\xi \in \mathcal{U}_\theta(\psi)} \xi^T v = \mu^T v + \sqrt{v^T \Sigma v}$$

while the partial concave conjugate function is defined as

$$c_*(x, v) := \inf_{\xi} v^T \xi - c(x, \xi)$$

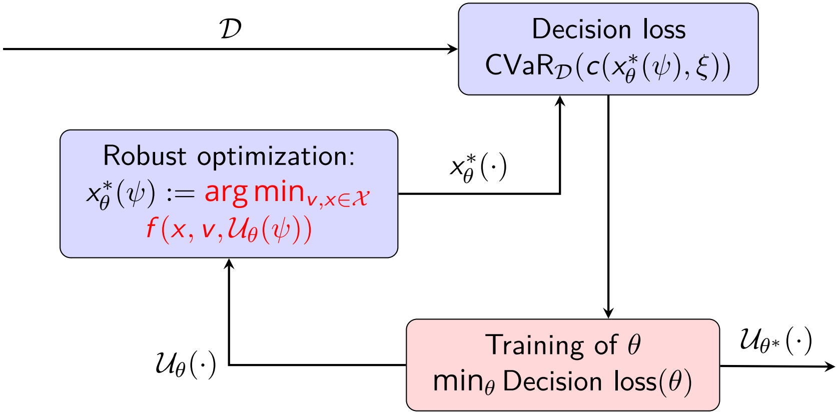
Robust optimization reformulation and derivatives

- We assume that $c(x, \xi)$ is convex in x and concave in ξ, while $\mathcal{X}$ is a convex set.
- Using Fenchel duality, one can follow Ben-Tal et al. [2015] to reformulate the robust optimization problem as:

$$x^*_\theta(\psi) := \arg\min_{x \in \mathcal{X}} \max_{\xi \in \mathcal{U}_\theta(\psi)} c(x, \xi) = \arg\min_{v, x \in \mathcal{X}} \underbrace{\delta^*(v|\mathcal{U}_\theta(\psi)) - c_*(x, v)}_{f(x,v,\mathcal{U}_\theta(\psi))}$$

- The derivatives of $x^*_\theta(\psi) := \arg\min_{v, x \in \mathcal{X}} f(x, v, \mathcal{U}_\theta(\psi))$ w.r.t. θ can be obtained using implicit differentiation (see Blondel et al. [2022])

Robust optimization reformulation and derivatives



Comparative study with GMM environment

- $(\psi, \xi) \in \mathbb{R}^2 \times \mathbb{R}^2$ drawn from a joint Gaussian mixture model with two modes
- Data: 600 points for training, 400 for validation, 1000 for test
- Targeted confidence level of 90%
- Average is calculated over 10 runs

Comparative study with GMM environment

- $(\psi, \xi) \in \mathbb{R}^2 \times \mathbb{R}^2$ drawn from a joint Gaussian mixture model with two modes
- Data: 600 points for training, 400 for validation, 1000 for test
- Targeted confidence level of 90%
- Average is calculated over 10 runs

	ETO-ACPS	ETO-DbS	TbS	
Avg. CVaR	1.69 ± 0.05	1.64 ± 0.07	**1.03** ± 0.10	
Avg. VaR	1.12 ± 0.04	1.07 ± 0.02	**0.72** ± 0.07	
Avg. marginal cov.	**91%** $\pm 1.4\%$	$85\% \pm 7.8\%$	$23\% \pm 6.1\%$	

Outline

Second-task: Conditional coverage

Lemma

An uncertainty set $\mathcal{U}_\theta(\psi)$ *has an a.s. conditional coverage of* $1-\epsilon$ *if and only if*

$$\mathcal{L}_{CC}(\theta) := \mathbb{E}[\,(\mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi) - (1-\epsilon))^2\,] = 0$$

Second-task: Conditional coverage

Lemma

An uncertainty set $\mathcal{U}_\theta(\psi)$ has an a.s. conditional coverage of $1-\epsilon$ if and only if

$$\mathcal{L}_{CC}(\theta) := \mathbb{E}[\,(\mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi) - (1-\epsilon))^2\,] = 0$$

$\mathcal{L}_{\mathsf{CC}}(\theta)$ can be approximated using:

$$\widehat{\mathcal{L}}_{\mathsf{CC}}(\theta) := \mathbb{E}_{\mathcal{D}}[(g_{\phi^*(\theta)}(\psi) - (1-\epsilon))^2]$$

where $g_{\phi^*(\theta)}(\psi) \approx \mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi)$ is obtained using logistic regression of membership variable $y(\psi, \xi; \theta) := \mathbb{1}\{\xi \in \mathcal{U}_\theta(\psi)\}$ on ψ.

Second-task: Conditional coverage

Lemma

An uncertainty set $\mathcal{U}_\theta(\psi)$ has an a.s. conditional coverage of $1-\epsilon$ if and only if

$$\mathcal{L}_{CC}(\theta) := \mathbb{E}[\,(\mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi) - (1-\epsilon))^2\,] = 0$$

$\mathcal{L}_{\mathsf{CC}}(\theta)$ can be approximated using:

$$\widehat{\mathcal{L}}_{\mathsf{CC}}(\theta) := \mathbb{E}_{\mathcal{D}}[(g_{\phi^*(\theta)}(\psi) - (1-\epsilon))^2]$$

where $g_{\phi^*(\theta)}(\psi) \approx \mathbb{P}(\xi \in \mathcal{U}_\theta(\psi)|\psi)$ is obtained using logistic regression of membership variable $y(\psi, \xi; \theta) := \mathbb{1}\{\xi \in \mathcal{U}_\theta(\psi)\}$ on ψ.

- I.e., letting the augmented data set

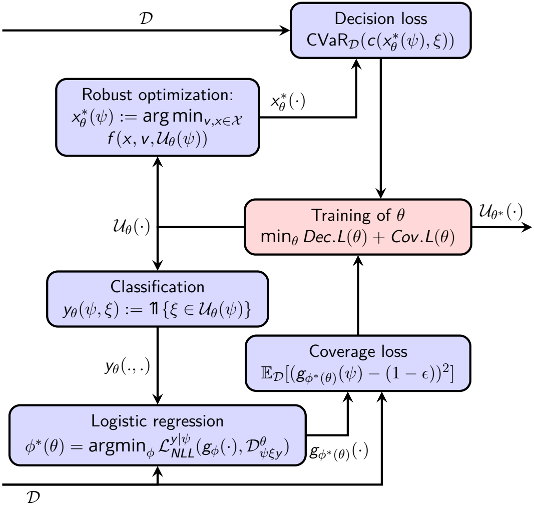
$$\mathcal{D}^\theta_{\psi\xi y} := \{(\psi_1, \xi_1, y(\psi_1, \xi_1; \theta)), \ldots, (\psi_N, \xi_N, y(\psi_N, \xi_N; \theta))\},$$

one solves $\phi^*(\theta) \in \mathrm{argmin}_\phi\, \mathcal{L}^{y|\psi}_{NLL}(g_\phi(\cdot), \mathcal{D}^\theta_{\psi\xi y})$ with

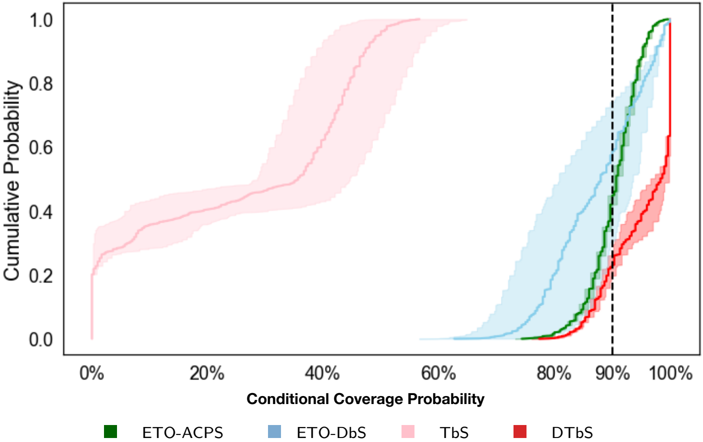
$$g_\phi(\psi) := \frac{1}{1+\exp^{\phi^T\psi+\phi_0}}$$

Double Task-based Set (DTbS) training

We train $\mathcal{U}_\theta(\psi)$ using the two tasks: produce good decision + produce good conditional coverage:



Comparative study with GMM environment



(See uncertainty set animation (url))

Comparative study with GMM environment

- $(\psi, \xi) \in \mathbb{R}^2 \times \mathbb{R}^2$ drawn from a joint Gaussian mixture model with two modes
- Data: 600 points for training, 400 for validation, 1000 for test
- Targeted confidence level of 90%
- Average is calculated over 10 runs

	ETO-ACPS	ETO-DbS	TbS	
Avg. CVaR	1.68 ± 0.04	1.66 ± 0.06	**1.05** ± 0.09	
Avg. VaR	1.12 ± 0.04	1.07 ± 0.02	**0.72** ± 0.07	
Avg. marginal cov.	**91%** $\pm 1.4\%$	$85\% \pm 7.8\%$	$23\% \pm 6.1\%$	

Comparative study with GMM environment

- $(\psi, \xi) \in \mathbb{R}^2 \times \mathbb{R}^2$ drawn from a joint Gaussian mixture model with two modes
- Data: 600 points for training, 400 for validation, 1000 for test
- Targeted confidence level of 90%
- Average is calculated over 10 runs

	ETO-ACPS	ETO-DbS	TbS	!! DTbS !!
Avg. CVaR	1.68 ± 0.04	1.66 ± 0.06	**1.05** ± 0.09	**1.07** $\pm$ 0.09
Avg. VaR	1.12 ± 0.04	1.07 ± 0.02	**0.72** ± 0.07	**0.75** $\pm$ 0.10
Avg. marginal cov.	**91%** $\pm 1.4\%$	$85\% \pm 7.8\%$	$23\% \pm 6.1\%$	$92\% \pm 1.5\%$

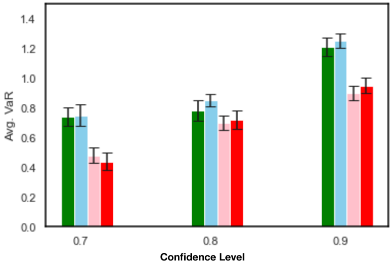
Portfolio optimization with market data

- Contextual info: Trading volume, volatility index (VIX), 10-year treasury yield index (TNX), oil index (CLF), S&P 500 (GSPC), gold price (GC=F), Dow Jones (DJI).
- Market data from Yahoo! Finance: 70 different stocks during period from 01/01/2012 to 31/12/2019 (2017-2019 reserved for test).
- Target confidence level of 70%, 80%, or 90%

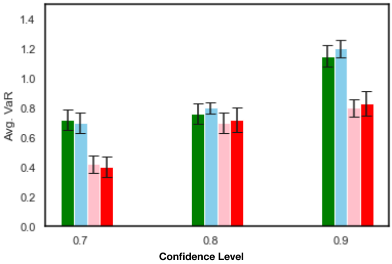
Model	Marginal coverage					
	2018			2019		
	70%	80%	90%	70%	80%	90%
ETO-ACPS	68%	78%	87%	71%	78%	89%
ETO-DbS	59%	75%	87%	61%	76%	86%
TbS	23%	24%	29%	26%	30%	32%
DTbS	71%	80%	93%	69%	78%	92%

Portfolio optimization with market data

- Contextual info: Trading volume, volatility index (VIX), 10-year treasury yield index (TNX), oil index (CLF), S&P 500 (GSPC), gold price (GC=F), Dow Jones (DJI).
- Market data from Yahoo! Finance: 70 different stocks during period from 01/01/2012 to 31/12/2019 (2017-2019 reserved for test).



(a) 2018



(b) 2019

ETO-ACPS ETO-DbS TbS DTbS

Outline

Concluding remarks

- We introduced a new conditional robust optimization approach for solving risk averse contextual optimization problems.
- In CRO, deep neural networks can be used to:
  - Represent richly structured uncertainty sets (see Goerigk and Kurtz [2023], Chenreddy et al. [2022])
  - Adapt uncertainty set continuously to covariates (this talk)
- Two types of training objectives:
  - Decision performance: Producing decisions that achieve low VaR/CVaR
  - Statistical performance: achieving the right marginal/conditional coverage

Thank you Professor Ye
for your mentoring, guidance, and support
throughout all these years.

Bibliography I

Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification, 2022.

Shane Barratt and Stephen Boyd. Covariance prediction via convex optimization, 2021.

Aharon Ben-Tal, Dick den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

Aharon Ben-Tal, Dick Hertog, and Jean-Philippe Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Math. Program.*, 149(1–2):265–299, feb 2015.

Dimitris Bertsimas and Nathan Kallus. From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044, 2020.

Dimitris Bertsimas, Christopher McCord, and Bradley Sturt. Dynamic optimization with side information. *European Journal of Operational Research*, 2022.

Rafael Blanquero, Emilio Carrizosa, and Nuria Gómez-Vargas. Contextual uncertainty sets in robust linear optimization. 2023.

Mathieu Blondel, Quentin Berthet, Marco Cuturi, Roy Frostig, Stephan Hoyer, Felipe Llinares-López, Fabian Pedregosa, and Jean-Philippe Vert. Efficient and modular implicit differentiation. *Advances in neural information processing systems*, 35:5230–5242, 2022.

Abhilash Reddy Chenreddy, Nymisha Bandi, and Erick Delage. Data-driven conditional robust optimization. In *Advances in Neural Information Processing Systems*, volume 35, pages 9525–9537, 2022.

Bibliography II

Giorgio Costa and Garud N Iyengar. Distributionally robust end-to-end portfolio construction. *Quantitative Finance*, 23(10):1465–1482, 2023.

Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

Priya Donti, Brandon Amos, and J Zico Kolter. Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

John C. Duchi, Peter W. Glynn, and Hongseok Namkoong. Statistics of robust optimization: A generalized empirical likelihood approach. *Mathematics of Operations Research*, 46(3): 946–969, 2021.

Adam N Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68 (1):9–26, 2022.

Adrián Esteban-Pérez and Juan M. Morales. Distributionally robust stochastic programs with side information based on trimmings. *Mathematical Programming*, 195(1):1069–1105, 2022.

Rui Gao and Anton Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *Math. Oper. Res.*, 48(2):603–655, 2023.

Marc Goerigk and Jannis Kurtz. Data-driven robust optimization using deep neural networks. *Computers and Operational Research*, 151(C), 2023.

Bibliography III

Lauren Hannah, Warren Powell, and David Blei. Nonparametric density estimation for stochastic optimization with an observable state variable. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.

Chancellor Johnstone and Bruce Cox. Conformal uncertainty sets for robust optimization, 2021.

Rohit Kannan, Güzin Bayraksan, and James R Luedtke. Residuals-based distributionally robust optimization with covariate information. *arXiv preprint arXiv:2012.01088*, 2020.

Christopher George McCord. *Data-driven dynamic optimization with auxiliary covariates*. PhD thesis, Massachusetts Institute of Technology, 2019.

Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.*, 171(1–2):115–166, 2018.

Viet Anh Nguyen, Fan Zhang, Jose Blanchet, Erick Delage, and Yinyu Ye. Robustifying conditional portfolio decisions via optimal transport, 2021.

Shunichi Ohmori. A predictive prescription using minimum volume k-nearest neighbor enclosing ellipsoid and robust optimization. *Mathematics*, 9(2):119, 2021.

Chunlin Sun, Linyu Liu, and Xiaocheng Li. Predict-then-calibrate: A new perspective of robust contextual lp, 2023.

Chunlin Sun, Linyu Liu, and Xiaocheng Li. Predict-then-calibrate: A new perspective of robust contextual lp. *Advances in Neural Information Processing Systems*, 36, 2024.

Bibliography IV

Irina Wang, Cole Becker, Bart Van Parys, and Bartolomeo Stellato. Learning for robust optimization. *arXiv preprint arXiv:2305.19225*, 2023.

Kai Wang and Alex Jacquillat. From classification to optimization: A scenario-based robust optimization approach. *Available at SSRN 3734002*, 2020.