Contextual chance-constrained and risk averse optimization

Bernardo K. Pagnoncelli

SKEMA Business School, Lille, France

Joint work with Hamed Rahimian (Clemson), Domingo Ramirez (PUC-Chile) and Arturo Cifuentes (PUC-Chile)

Workshop on robust and stochastic optimization methods
November 19th, 2021

Outline

Outline

Decision making under uncertainty

- There are many frameworks for decision making under uncertainty
- I will focus on Stochastic Programming
- In the classical formulations, the c.d.f. of the random elements of the model is assumed to be known
- In this talk we will focus on static (as opposed to dynamic) problems

Stochastic Programming in today's data-rich world

In classical approaches, the contextual information (i.e., covariates/features/attributes) associated with the random parameters of interest, ξ, is *either* totally ignored or it is encapsulated in ξ

QUESTIONS:

- Could we have leveraged the information on the historical covariates along with the information on the future covariates in our favor?
- How can we include the contextual information within the stochastic programming framework?
  - Observe data $(X^i, \xi^i)_{i=1}^N$: Covariates X^i along with the realizations of random parameter of ξ^i

More examples

Farming/Agriculture

ξ: Crops yield
Covariates: precipitation, pesticides, humidity, ...

Finance

ξ: Return
Covariates: Google searches/news on war, oil price, political comments, unemployment, ...

Recent papers

- Ascarza '18 argues the churn problem (prediction) must take sensitivity to the intervention into account (prescription).
- In Muñoz et al. '20 the authors solve a predictive-prescriptive Cournot strategic producer partaking problem, consider a *problem-aware* loss function.
- In Wang et al. the authors consider MDPs where some parameters need to be estimated (transition function, transition probability)

Decision-focused learning, problem-aware loss function, predict-then-optimize, end-to-end model learning, objective mismatch...

Contextual stochastic programming

Two-stage stochastic programming

$$\min_{u \in U} \{cu + \mathbb{E}\left[Q(u, \xi) | \; X = x\right]\},$$

$$Q(u, \xi) = \min_{y \in Y} \{qy \mid Tu + Wy \geq h\}$$

Expected value constraints (chance constraints, risk)

$$\max_{u \in U} f(u)$$

s.t.

$$E\{G(u, \xi) \leq 0 \mid X = x\} \geq \alpha$$

Earlier Work: Hannah et al. '10, Donti et al. '17, Bertsimas & Van Parys, '17, Elmachtoub & Grigas, '17, Deng et al., '18, Deng and Sen, '18, Ban & Rudin, '19, Bertsimas & McCord ('18, '19), Ho & Hanasusanto '19, Larsen et al. '19, Bertsimas & Kallus '19, Cohen et al. '20, Rohit et al. '20

A portfolio problem

$$\begin{aligned} \Gamma^* = \max_{u \in U} \quad & (1+\bar{r})^T u \\ \text{s.t.} \quad & P\left((1+r)^T u \geq v \mid X = x\right) \geq 1-\alpha, \end{aligned}$$

where $\bar{r} = (0.0145, 0.0083), v = 0.8, \alpha = 0.1$ and $U = \{u \in \mathbb{R}^2 \mid u_1 + u_2 = 1, u_1 \geq 0, u_2 \geq 0\}$.

The random vector $r \sim N(1+\bar{r}, \Sigma(x))$:

$$\Sigma(x) = \begin{cases} \begin{pmatrix} 0.02900 & 0.02051 \\ 0.02051 & 0.01819 \end{pmatrix} & \text{if } x = 0 \text{ (bull)}, \\ \begin{pmatrix} 0.04799 & 0.02051 \\ 0.02051 & 0.02859 \end{pmatrix} & \text{if } x = 1 \text{ (bear)}. \end{cases}$$

Solution

- SAA problem with 10,000 samples: 5,000 with $x = 0$ and 5,000 with $x = 1$:

$$(u_1^*, u_2^*) = (0.57113, 0.42887), \quad \Gamma_N^* = 1.01186.$$

- When $x = 0$ the SAA solution is feasible and performs well.
- When $x = 1$ we have

$$\begin{aligned} P\left(r^T u \geq v \mid X = x_1\right) &= P\left(r^T(0.57113, 0.42887) \geq v \mid X = 1\right) \\ &= 1 - \Phi(-1.20409) = .88572 < 1 - \alpha = 0.9 \end{aligned}$$

INFEASIBLE!!!

Outline

Empirical risk minimization

- The idea is to replace u by $u(x)$, e.g.

$$\mathcal{D} = \left\{ u : \mathcal{X} \to \mathbb{R} \mid u(x) = u'x = \sum_{j=1}^{p} u^j x^j \right\}$$

- Cross terms such as $x_i x_k$, and x_i^2 can be introduced as well
- Main advantage: by solving the problem *once* it generates a response function u
- Drawbacks: has issues dealing with constraints, linear rule may not be suited for all problems, and it often needs regularization

Kernel optimization

- It is an idea with a long history in the field of statistics (Nadaraya and Watson '64).
- Given contextual data $\{(x_i, \xi_i)\}_{i=1}^N$ we want to estimate

$$m(x) = \mathbb{E}\left[\xi \mid x\right].$$

- Locally weighted average:

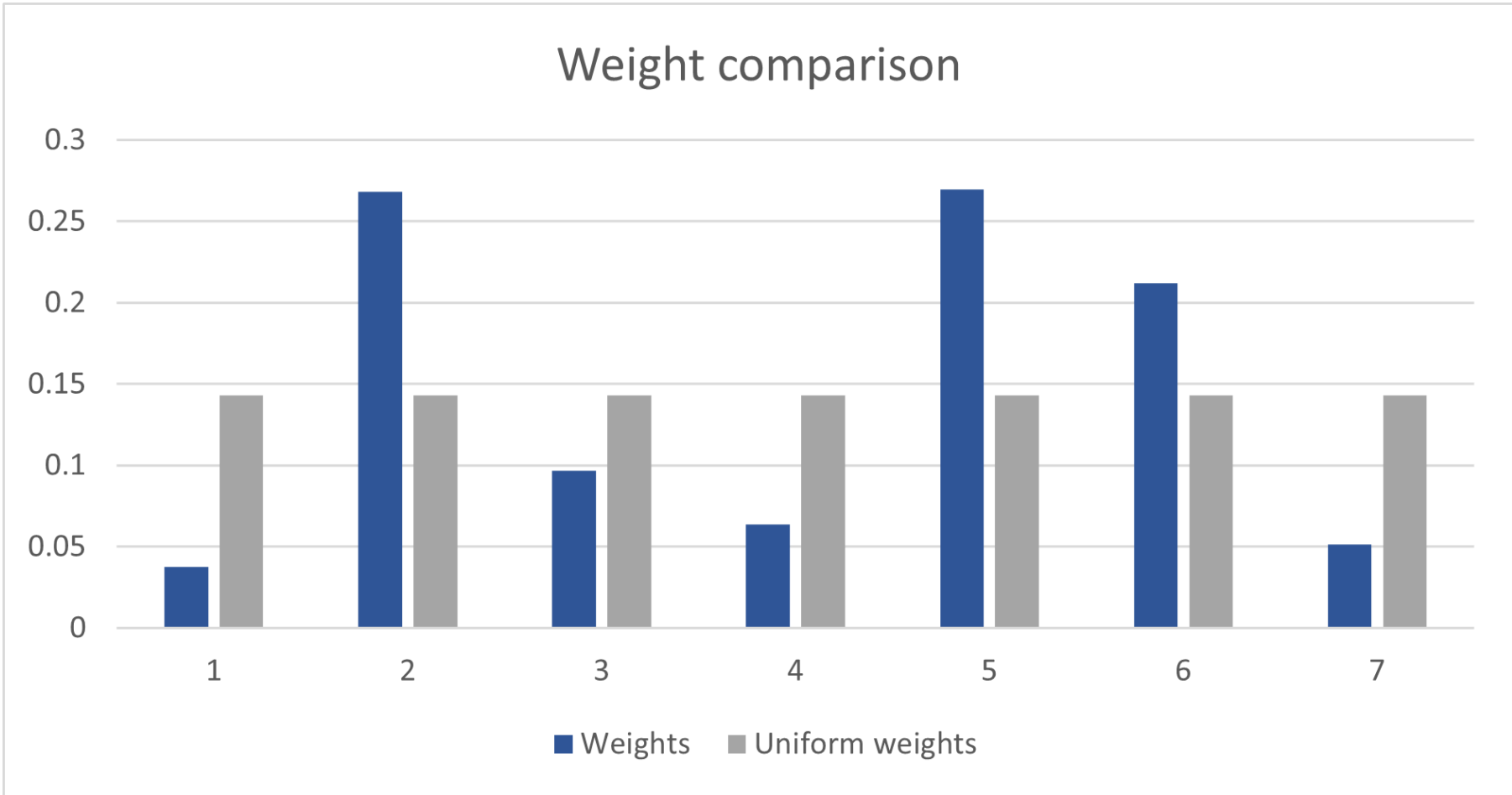
$$m_h(x) = \frac{\sum_{i=1}^N K_w(x - x_i)\xi_i}{K_w(x - x_i)}$$

- $K_w(\cdot) = K(\cdot/w)/w$, the parameter w is the bandwidth, and it has to be calibrated

Illustration - random data

Data	Att 1	Att 2	Att 3	ξ
1	-0.75	15.28	151.56	510
2	5.16	12.37	122.69	463
3	4.82	16.70	78.92	484
4	-3.94	17.60	145.58	651
5	0.03	9.98	97.60	498
6	2.99	11.55	91.14	499
7	0.20	14.08	71.41	579
x	2	13	110	???

Weights



	Sample average	Weighted average
ξ	526.28	501.82

Examples of Kernels

$$\textbf{Naive Kernel:} \quad K(x) = 1/2 \; \mathbb{1}\{\|x\| \leq 1\}$$

$$\textbf{Gaussian Kernel:} \quad K(x) = 1/(\sqrt{2\pi}) \; e^{-\|x\|^2/2}$$

$$\textbf{Epanechnokov:} \quad K(x) = (1 - \|x\|^2)\mathbb{1}\{\|x\| \leq 1\}$$

$$\textbf{Quartic:} \quad K(x) = (1 - \|x\|^2)^2\mathbb{1}\{\|x\| \leq 1\}$$

$$\textbf{Tri-cubic:} \quad K(x) = (1 - \|x\|^3)^3\mathbb{1}\{\|x\| \leq 1\}$$

For categorical variables we need special kernels proposed in Aitchison & Aitken '76, and their ideas are implemented in the `np` package in R (Hayfield & Racine '08).

Drawbacks: Kernels with categorical variables come with more parameters to calibrate, and the resulting kernel is the product of different types of kernels $\rightarrow$ instability.

Weighting functions

- Similar to kernel methods:

$$P\{G(u,\xi) \leq 0 \mid X = x\} = \mathbb{E}\left[\mathbb{1}_{(-\infty,0)}(G(u,\xi)) \mid X = x\right]$$
$$\approx \sum_{i=1}^{N} w_i(x)\mathbb{1}_{(-\infty,0)}(G(u,\xi_i))$$

- Weights can be given by
  1. k-NN: $w_i(x) = 1/k, w_i(x) = N^{1/2}$
  2. CART: $w_i(x) = 1/b$ for b instances that constitute the leaf where x belongs to in the tree
  3. Random forest: Same, averaging over different trees
- The general structure is similar to that of SAA
- Can we obtain theoretical results for sums of weighted random variables (as opposed to $1/N$?)

Outline

Lower bound

Theorem I (Rahimian and P.)

Consider a sequence of pairs of i.i.d. random vectors $\{(x_i, \xi_i)\}_{i=1}^N$ on a probability space $(\mathcal{X} \times \Xi, \mathcal{F}_\mathcal{X} \times \mathcal{F}_\Xi, P)$, where $x_i \in \mathbb{R}^p$ and $\xi_i \in \mathbb{R}$. For the positive values of the function Q_x, where $Q_x : \mathbb{R}^p \to \mathbb{R}$ is a continuous function, for a fixed $X = x$. Define the array of weights $w_i(x)$, $i = 1, \ldots, N$ as

$$w_i(x) := \frac{Q_x(X_i)}{\sum_{j=1}^N Q_x(X_j)}, \ i = 1, \ldots, N, \tag{1}$$

and assume at least one $Q_x(x_i)$ is positive. If $\alpha > \epsilon$, we have

$$P^N(z_\alpha^N \leq z_\epsilon^* | X = x) \geq 1 - \exp\left\{ -\frac{2(\alpha - \epsilon)^2}{\sum_{i=1}^N w_i(x)^2} \right\}.$$

Corollary I

k independent of N:

If $w_i(x) = 1/k$, from some value integer value of $k \leq N$ we have

$$\exp\left\{-2k(\alpha - \epsilon)^2\right\}. \tag{2}$$

No asymptotic convergence.

$k = N^{1/2}$:

In this case we have

$$\exp\left\{-2\sqrt{N}(\alpha - \epsilon)^2\right\},$$

which goes to zero. If we want a lower bound with confidence $1 - \delta$ we need at least

$$N \geq \frac{1}{4(\alpha - \epsilon)^4}(\log(1/\delta))^2$$

data points.

Feasibility result I

Theorem II (Rahimian and P.)

Assume that the feasible set $U \subset \mathbb{R}^n$ of CCP is finite, and let $\alpha \in [0, \epsilon)$. For a covariate vector $x \in \mathbb{R}^p$, we have

$$P_{\xi|X}(U_\alpha^N \subseteq U_\epsilon \mid X = x) \geq 1 - |U \backslash U_\epsilon| \exp\left\{ -\frac{2(\epsilon - \alpha)^2}{\sum_{i=1}^N w_i(x)^2} \right\},$$

where

$$U_\epsilon = \{u \in U \mid p(u) \geq 1 - \epsilon\}$$

is the feasible set of C-CCP.

Corollary

Number of data points N

If we have $k-$NN with $k = N^{1/2}$ we get exponential convergence to 1 as the data size grows. Moreover, if N satisfies

$$N \geq \left(\log\left(\frac{|U \backslash U_\epsilon|}{\delta}\right)\right)^2 \frac{1}{4(\epsilon - \alpha)^4},$$

then with probability $1 - \delta$ a solution of problem DDC-CCP will be feasible to problem C-CCP.

Feasibility result II

Theorem III (Rahimian and P.)

Let $\alpha \in [0, \epsilon), \beta \in (0, \epsilon - \alpha)$ and $\eta > 0$. Under mild assumptions, for a covariate vector $x \in \mathbb{R}^p$, we have

$$P_{\xi|X}(U^N_{\alpha,\eta} \subseteq U_\epsilon \mid X = x) \geq 1 - \lceil 1/\beta \rceil \lceil 2LD/\eta \rceil^{d_u} \exp\left\{ -\frac{2(\epsilon - \alpha - \beta)^2}{\sum_{i=1}^N w_i(x)^2} \right\},$$

where $U^N_{\alpha,\eta}$ is the modified feasible set of CCD-CCP and U_ϵ is the feasible set of C-CCP.

Corollary

Number of data points N

If we have $k-$NN with $k = N^{1/2}$, and ϵ, α such that $\alpha < \epsilon$ and $\beta \in (0, \gamma)$, e.g., $\beta = (\epsilon - \alpha)/2$, we have that a feasible solution to DDC-CCP is feasible to C-CCP with confidence at least $1 - \delta$ if

$$N \geq \frac{4}{(\epsilon - \alpha)^2} \left(\log \frac{1}{\delta} + d_u \log \left\lceil \frac{2LD}{\eta} \right\rceil + \log \left\lceil \frac{2}{\epsilon - \alpha} \right\rceil \right)^2 .$$

Outline

Portfolio selection: history

- Any person or institution who wants to invest faces the same problem: "How should I allocate my funds?"
- There have been several rule-of-thumb approaches to investing, and one of the most popular is the Equally Weighted (EW) approach
- Markowitz '52 published a seminal paper introducing the trade-off between risk (variance) and return.
- Even though Markowitz's work was revolutionary, it had its own pitfalls: bringing the methodology to practice proved challenging

Portfolio selection: today

Over the last years, there have been several advances in the discipline. Examples include the introduction of risk measures (coherence, the CVaR), the use of copulas to estimate joint distribution functions, and the advent of AI and Machine Learning that is starting to permeate into finance.

Main issues to tackle:

- Low signal-to-noise ratio, specifically of yearly returns.
- Use of ML techniques to incorporate contextual information (features).

Mathematical Formulation

$$
\begin{aligned}
\max_{w} \quad & \mathbb{E}\left[w^\top r \mid X = x\right] \\
\text{s.t.} \quad & \mathrm{CVaR}_\alpha\left[-w^\top r \mid X = x\right] \leq \gamma \\
& w^\top \mathbb{1} = 1 \\
& w \geq 0
\end{aligned}
$$

To solve the problem we need the conditional distribution $P_{r|X=x}$, i.e., distribution of $(r \mid X = x)$, which in most cases is not available

Asset Classes (Jan '07–Jan '21)

Stocks	
VTMSX	Total Stock Market Index
VEIEX	Emerging Markets Stock Index
Bonds	
VBMFX	Total Bond Market Index
VIPSX	Inflation Protected Securities
Real Estate	
VGSIX	U.S. Real Estate Investments

Sixth asset class: a risk-free investment with a 0.5% annual return.

Feature Selection

The goal: Select features that capture the current state of the world and that can be helpful to estimate future returns

Pre selected 20 potential features picked by several investment professionals.

Applied a feature selection algorithm (PFA) to choose the best-performing features

Final selection consist of U.S. unemployment rate, the St. Louis Fed financial stress index, the Euro-area package holidays index, and the manufacturers' inventories-to-sales ratio[1].

[1] The respective mnemonics are UNRATE, STLFSI2, CP0960EZ19M086NEST and MNFCTRIRSA.

Selecting the weights

- Given historical data $D_N := \{(x_i, r_i)\}_{i=1}^N$ we use a weight function $w_i(x)$ to estimate the conditional distribution $P_{r|X=x}$
- Intuitively, given the current state of the world x, the weight of each observation will be the its distance to x
- The proposed weight function $w_i(x)$ is a local learner model based on the Nadaraya-Watson kernel regression:

$$w_i(x) = \frac{K_b(x - x_i)}{\sum_{i=1}^N K_b(x - x_i)}, \quad i = 1, \ldots, N$$

Data-Driven Formulation

$$
\begin{aligned}
\max_{z,v_i,\eta} \quad & z^\top \sum_{i=1}^{N} w_i(x) r_i \\
\text{s.t.} \quad & \eta + \tfrac{1}{1-\alpha} \sum_{i=1}^{N} w_i(x) v_i \leq \gamma \\
& v_i \geq -z^\top r_i - \eta, \quad i = 1, \ldots, N \\
& z^\top \mathbb{1} = 1 \\
& z, v \geq 0
\end{aligned}
$$

Results

- Two methods reported, NORTA with Features (NwF) and NORTA without features (Nw/oF)
- Equally Weighted (EW) reported for benchmarking purposes
- 1,000 synthetic years generated per investment window

NwF surface

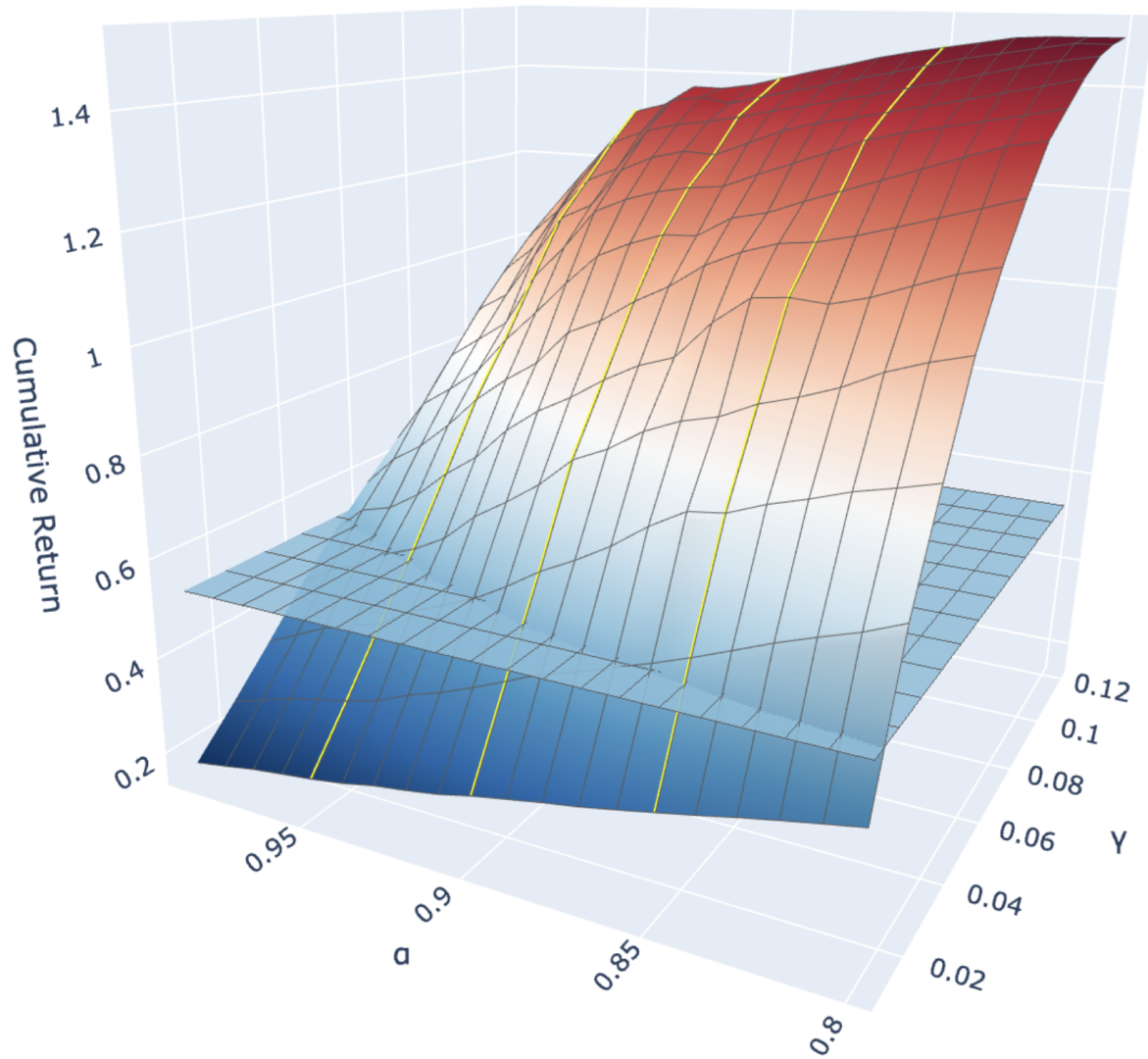


Figure: 3-D efficient frontier for NwF with bandwidth $b = 2$ for different values of α and γ.

2D Plane

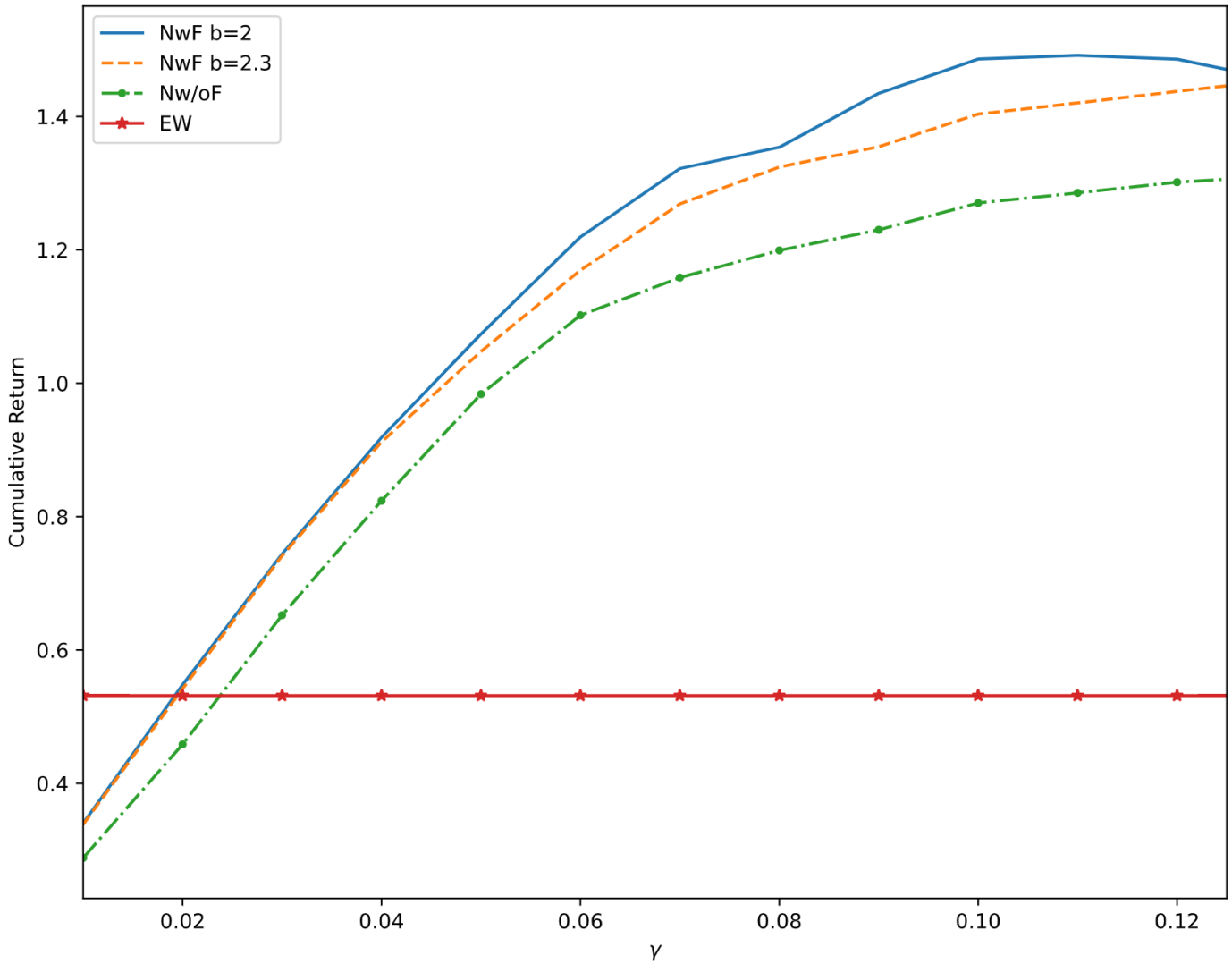


Figure: Efficient frontiers at $\alpha = 0.9$.

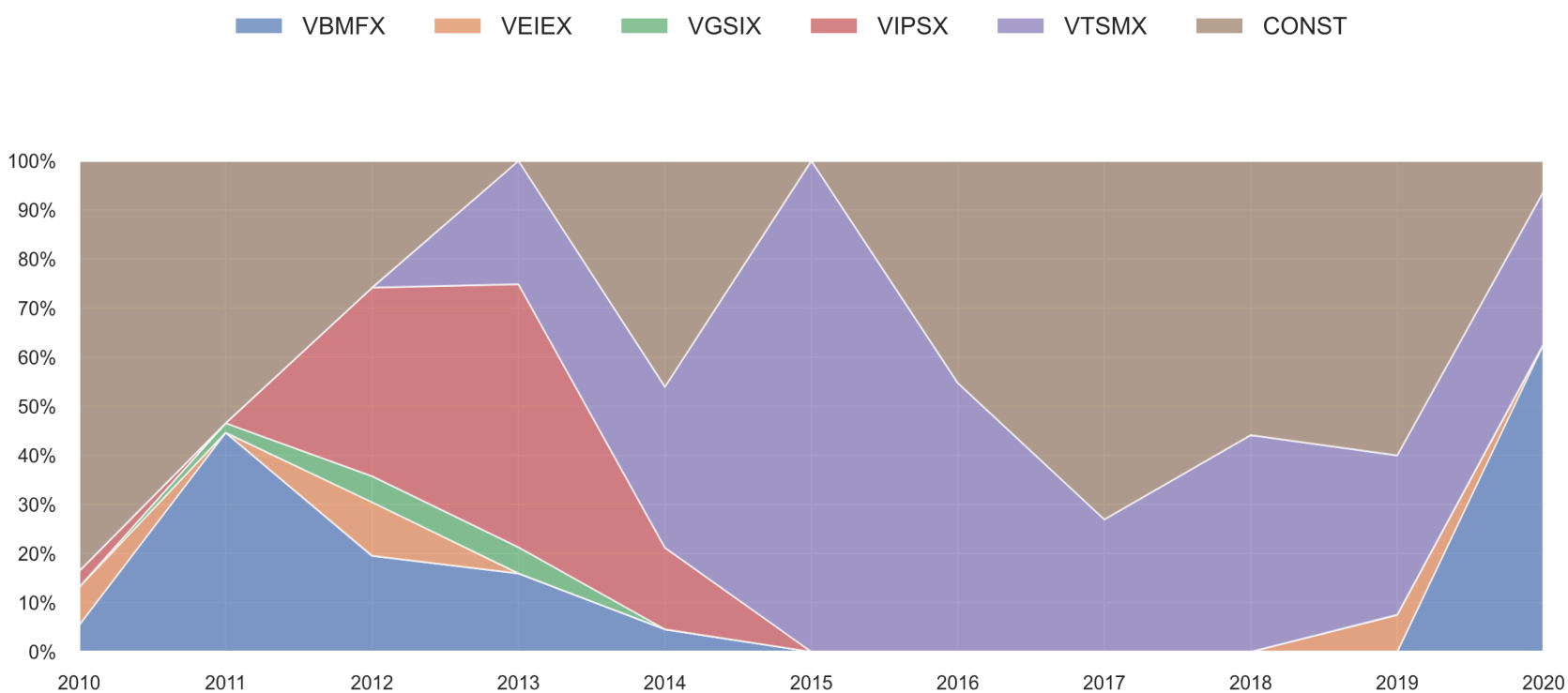
Diversification

Diversification index (DI) for different values of α and γ.

	$\alpha = 0.85$		$\alpha = 0.90$		$\alpha = 0.95$	
γ	**Nw/oF**	**NwF**	**Nw/oF**	**NwF**	**Nw/oF**	**NwF**
0.03	45%	47%	51%	53%	63%	58%
0.05	29%	28%	41%	42%	53%	53%
0.07	26%	22%	28%	28%	40%	40%
0.09	23%	19%	25%	22%	32%	33%
0.12	22%	14%	24%	18%	25%	23%

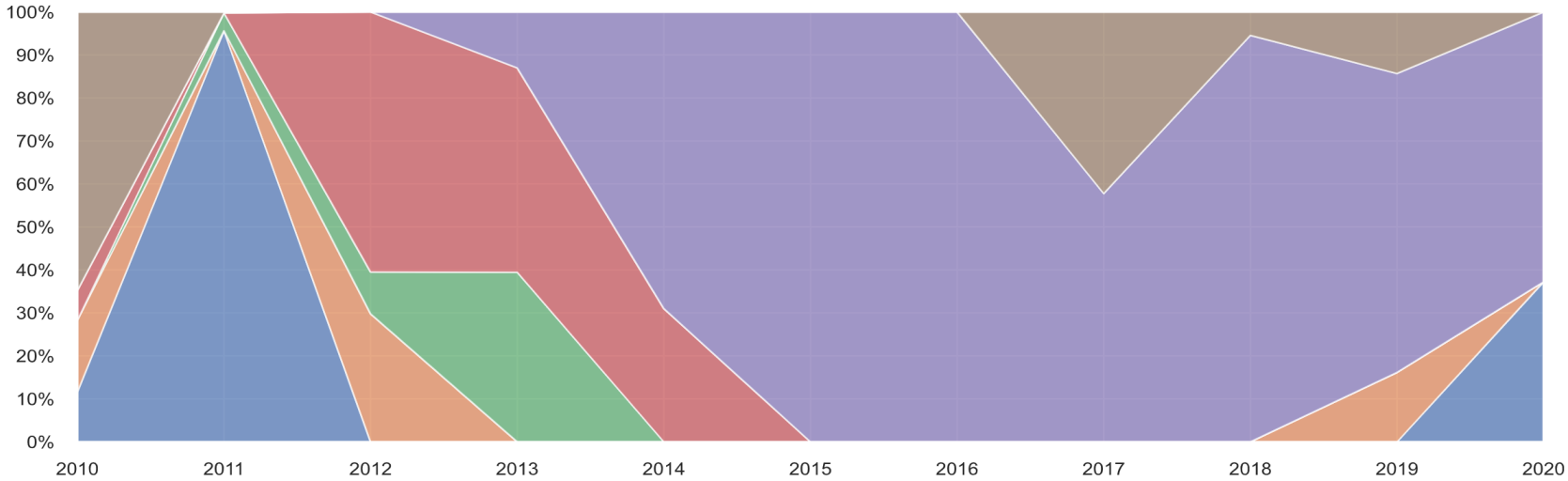
$$DI = \frac{1 - \sum_{i=1}^{m} z_i^2}{1 - 1/m}.$$

Results

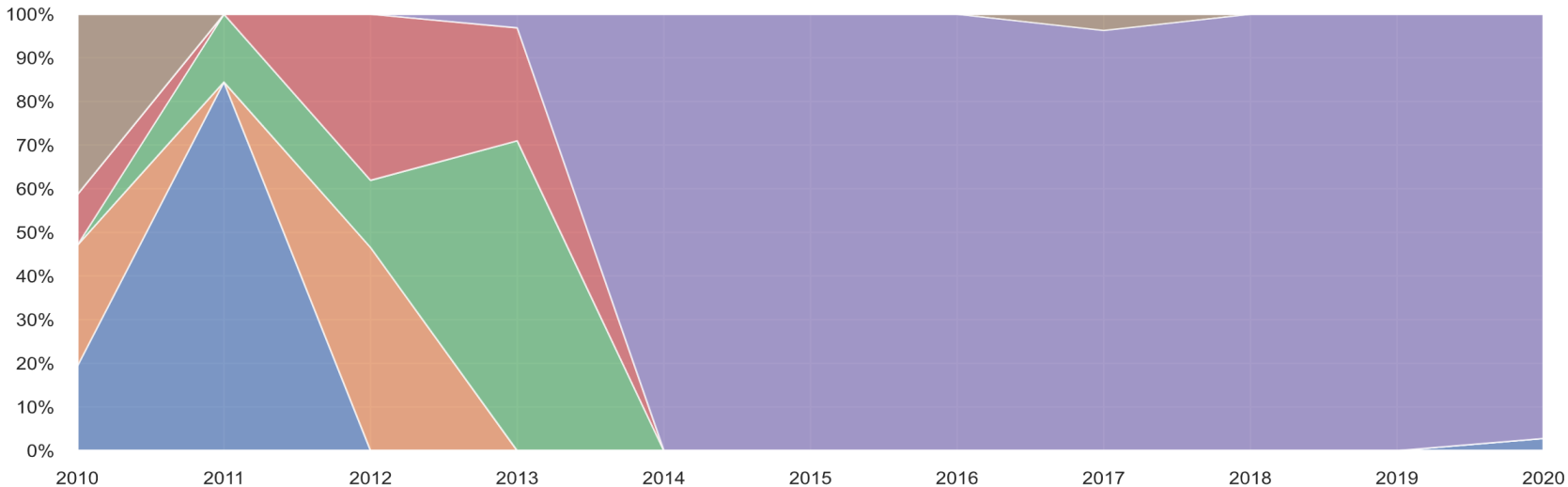


(a) $\gamma = 0.03$ and $\alpha = 0.95$.

Results



(a) $\gamma = 0.07$ and $\alpha = 0.95$.



(b) $\gamma = 0.12$ and $\alpha = 0.95$.

Figure: Asset allocation with the NwF approach.

Outline

Concluding remarks

- We discussed how to include contextual information in risk constrained problems
- We showed theoretical results that ensure feasibility for the CC as the number of data points grows (results hold for expected value constraints)
- We illustrated our findings in a portfolio selection problem under CVaR constraints
- Future work includes extensions to the multistage case, methodological developments exploring the link between prediction and prescription, and applications in transportation (urban mobility), energy (OPF problem), natural resources management...

Contact information

Thank you!

Bernardo Pagnoncelli

bernardo.pagnoncelli@skema.edu

This research has been partially supported by the Patrick and Amy McCarter Fellowship, 2018-2019 (IEMS, Northwestern University), and ANID grant number 1170178.