Mathematical Programming manuscript No.
(will be inserted by the editor)

Supplementary Material

Adrián Esteban-Pérez · Juan M. Morales

1 Additional numerical experiments, case $\alpha > 0$. Portfolio optimization

In this section, we present and discuss some numerical results for the case $\mathbb{Q}(\widetilde{\Xi}) = \alpha > 0$. For this purpose, we use the same portfolio allocation problem described in the main manuscript. To this end, we assume instead that the feature vector lives in an uncertainty set $\mathscr{Z}$ such that $\mathbb{Q}(\widetilde{\Xi}) > 0$. In particular, we consider $\mathscr{Z} := \{\mathbf{z} \in \mathbb{R}^3 : \|\tilde{\mathbf{z}}\|_\infty \leqslant r\}$, with $\tilde{\mathbf{z}}$ being the standardized feature vector. Thus, we have that $\widetilde{\Xi}$ is given by

$$\widetilde{\Xi} := \{(\mathbf{z}, \mathbf{y}) \in \mathbb{R}^{3+6} : \|\tilde{\mathbf{z}}\|_\infty \leqslant r\}$$

We take $r = 0.6$ for the simulation experiments. We draw 50 000 samples from the true joint data-generating distribution through the explicit form of $\mathbf{y}/\mathbf{z}$ given in the main text. We then use the conditional empirical distribution made up of those samples falling within $\widetilde{\Xi}$, specifically, 7306 data points, as a proxy of the true conditional distribution $\mathbb{Q}_{\widetilde{\Xi}}$. Consequently, we have that $\mathbb{Q}(\widetilde{\Xi}) \approx 0.14612$. We wish to solve the following optimization problem

$$\min_{(\mathbf{x},\beta')\in X} \mathbb{E}\left[\beta' + \frac{1}{\delta}\left(-\langle \mathbf{x}, \mathbf{y}\rangle - \beta'\right)^+ - \lambda \langle \mathbf{x}, \mathbf{y}\rangle \;\middle|\; (\mathbf{z}, \mathbf{y}) \in \widetilde{\Xi}\right] \tag{1}$$

with the rest of the parameters being equal to the values taken in the instance $\alpha = 0$.

We also compare here four data-driven approaches to solve problem (1), namely:

- Our two approaches, i.e., problem $\mathrm{P}_{(\alpha,\widetilde{\rho}_N)}$ with $\alpha := \mathbb{Q}(\widetilde{\Xi})$ (that is approximately equal to 0.14612, as we have just mentioned), denoted as "DROTRIMM1" and problem $\mathrm{P}_{(\alpha_N,\widetilde{\rho}_N)}$, where $\alpha_N := \widehat{\mathbb{Q}}_N(\widetilde{\Xi})$ is an estimate of α. We refer to this approach as "'DROTRIMM2." In principle, this would be the natural approach that a decision-maker with no knowledge of α would use.
- A sample average approximation (SAA) method that works with the samples falling in $\widetilde{\Xi}$.

Adrián Esteban-Pérez, Juan M. Morales (corresponding author)
Universidad de Málaga, Andalucía Tech, Departamento de Matemática Aplicada, 29071 Málaga, España
E-mail: adrianesteban@uma.es, juan.morales@uma.es

– The aforementioned SAA method followed by a standard Wasserstein-metric-based DRO approach to robustify it, which we call "SAADRO".

As in the previous numerical experiments, we employ a similar bootstrapping procedure based on the available data sample to tune the robustness parameter that each method j, with $j \in \{\text{DROMTRIMM1, DROTRIMM2, SAADRO}\}$, uses. More specifically, for each $j \in \{\text{DROMTRIMM1, DROTRIMM2, SAADRO}\}$ and a given value of reliability $1-\beta \in (0,1)$ (in our numerical experiments, we set β to 0.15), we seek an estimator $param_N^{\beta,j}$ that leads to the best out-of-sample performance, while guaranteeing the desired level of confidence $1-\beta$. For each sample of size N, we use the following algorithm to derive $param_N^{\beta,j}$ and the corresponding portfolio solution:

1. We construct *kboot* resamples (with replacement) of size N, each playing the role of a different training dataset. In our experiments we use $kboot = 50$. Moreover, we build a validation dataset (per resample) from those data points from the original sample of size N that fall in $\widetilde{\Xi}$, but which have not been involved in the resample. We only consider resamples from which we can build a validation set of at least one data point. Furthermore, unlike DROTRIMM1 and DROTRIMM2, SAADRO can only be implemented if we have at least one data point falling within $\widetilde{\Xi}$ in the training set (the same occurs with SAA). Thus, we implicitly assume that the source sample has no fewer than two data points in $\widetilde{\Xi}$.
2. For each resample $k = 1,\ldots,kboot$ and each candidate value for *param* (taken from the discrete set $\{b\cdot 10^c \;:\; b\in\{0,\ldots,9\},\; c\in\{-3,-2,-1,0\}\}$), we compute a solution by method j with parameter *param* on the k-th resample. The resulting optimal decision is denoted as $\widehat{x}_N^{j,k}(param)$ and its corresponding objective value as $\widehat{J}_N^{j,k}(param)$. Thereafter, we calculate the out-of-sample performance $J(\widehat{x}_N^{j,k}(param))$ of the data-driven solution $\widehat{x}_N^{j,k}(param)$ over the validation dataset.
3. From among the candidate values for *param* such that $\widehat{J}_N^{j,k}(param)$ exceeds the value $J(\widehat{x}_N^{j,k}(param))$ in at least $(1-\beta)\times kboot$ different resamples, we take as $param_N^{\beta,j}$ the one yielding the best cost performance averaged over the *kboot* resamples.
4. Finally, we compute the solution given by method j with parameter $param_N^{\beta,j}$, $\widehat{x}_N^j := \widehat{x}_N^j(param_N^{\beta,j})$ and the respective certificate $\widehat{J}_N^j := \widehat{J}_N^j(param_N^{\beta,j})$.

Figure 1 shows the box plots pertaining to the out-of-sample disappointment and performance associated with each of the considered data-driven approaches for various sample sizes. The box plots have been obtained from 200 independent runs per sample size N. The SAA method provides portfolios that, in expectation, perform reasonably well, especially when the sample size is large enough. However, SAA definitely fails to ensure the desired level of reliability. As for the three approaches that incorporate robustness in the decision-making, DROTRIMM1 and DROTRIMM2 seem to systematically identify reliable portfolios with a better expected performance than those given by SAADRO.

To investigate the ability of SAADRO, DROTRIMM1 and DROTRIMM2 to identify good portfolios, we provide Figure 2, which is analogous to Figure 3 in the case

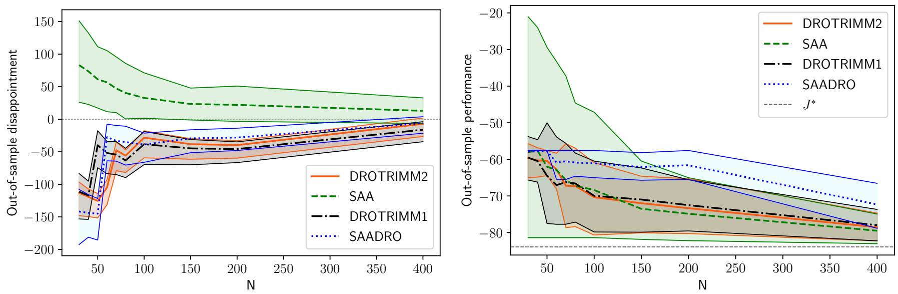

Fig. 1: Portfolio problem with features: Performance metrics. Case $\alpha > 0$ and $\delta = 0.5$, $\lambda = 0.1$

$\alpha = 0$. Observe that both DROTRIMM1 and DROTRIMM2 guarantee reliability for smaller values of their robustness parameter than SAADRO. This gives the former a competitive advantage over the latter, essentially because it appears that a better out-of-sample performance (in expectation) is, in general, aligned with a lower distributional robustness (this finding is consistent with the fact that the unreliable SAA solution performs fairly well in terms of the weighted mean-risk asset returns). To be more precise, taking a small sample size N (say 50) and an equal value for each of their robustness parameters, DROTRIMM1 and DROTRIMM2 deliver portfolios with an actual expected cost (and variance) that is lower than or approximately equal to that of the portfolios provided by SAADRO. They do so for any value of their robustness parameter. Furthermore, when N is increased, even though there exists a range of values of the robustness parameter for which SAADRO also identifies portfolios with a good performance out of sample, these are discarded by the method because they do not comply with the reliability specification. For instance, take $N = 400$. SAADRO needs a radius larger than 0.2-0.3 to ensure reliability. However, for these values of the Wasserstein-ball radius, the portfolios given by SAADRO result in an actual expected cost above -70. On the other hand, DROTRIMM2 guarantees reliability with a value of its robustness parameter above 0.003-0.004, for which, in addition, it provides solutions with an actual expected cost below -77.

To further support this finding, we conclude this section with Figure 3, which is similar to Figure 1. However, Figure 3 has been obtained through a different experiment, in which the value of the robustness parameter that each method uses has been *optimally* selected from the previously indicated discrete set. In other words, the results shown in that figure are those a decision-maker would obtain in the hypothetical case that the true conditional distribution $\mathbb{Q}_{\widetilde{\Xi}}$ could be used to tune the robustness parameters of the DRO methods. Therefore, these results correspond to the best solutions that can be obtained from SAADRO, DROTRIMM1 and DROTRIMM2, and confirm that our approaches (especially, DROTRIMM2) can potentially identify portfolios that significantly outperform those delivered by SAADRO under the same reliability requirement.

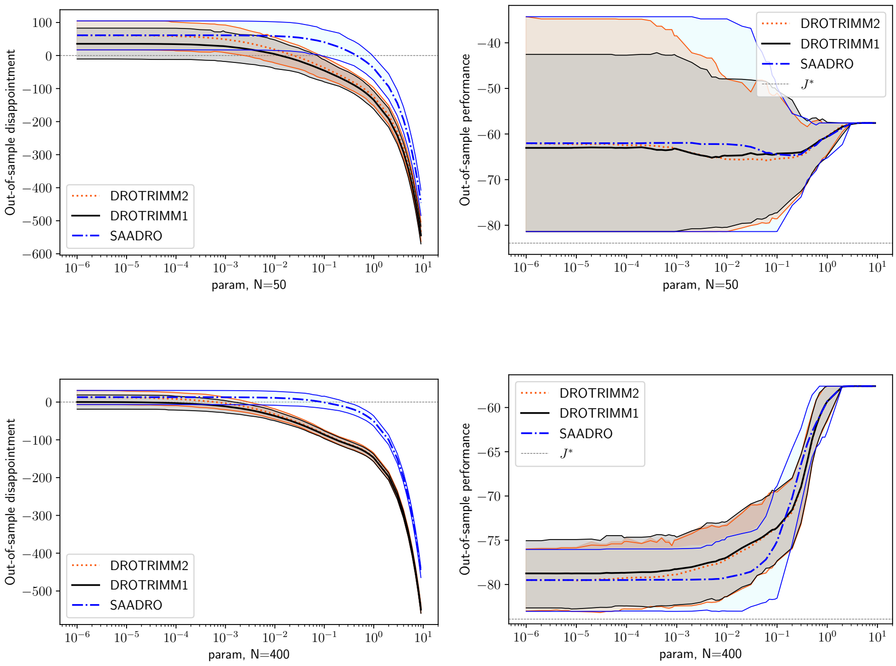

Fig. 2: Case $\alpha > 0$, impact of the robustness parameter with 200 training samples and $\delta = 0.5, \lambda = 0.1$

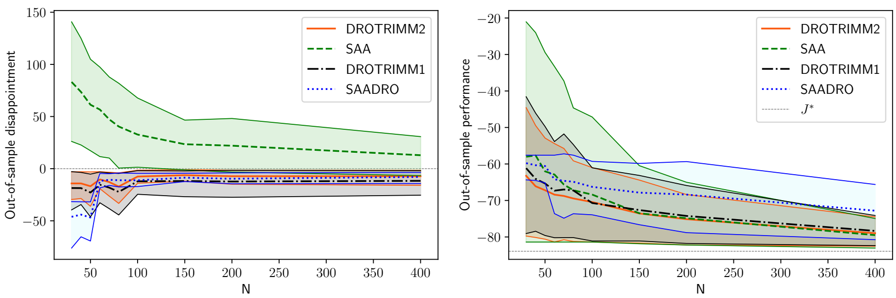

Fig. 3: Portfolio problem with features: Performance metrics under an optimal selection of the robustness parameters. Case $\alpha > 0$ and $\delta = 0.5,\ \lambda = 0.1$