DSO 670 Final Project

Yingxiao Ye
University of Southern California, Center for Artificial Intelligence in Society, yingxiao.ye@usc.edu

1. Introduction

In this project, I investigated the proposed method and numerical experiments in Ben-Tal et al. (2013). The problem I investigated and the experiment settings all follow those in the Section 6.4 in Ben-Tal et al. (2013) unless otherwise specified. The are in total four experiments. In each experiment, I considered two types of multi-item newsvendor problems. The first is to maximize the sum of the expected profits,

$$\max_{Q} \sum_{j} \sum_{i} p_i^{(j)} \bar{r}_j \left(Q_j, i\right),$$

and the second is to maximize the minimal expected profit,

$$\max_{Q} \min_{j} \sum_{i} p_i^{(j)} \bar{r}_j \left(Q_j, i\right).$$

The concrete robust and nonrobust models can be found in EC.2 and EC.3.

The main motivation is to answer the following questions.

1. How do the choice of ρ and the p-vectors in evaluation affect the results?
2. Why does this paper sample p-vector from a confidence set rather than a known true distribution?
3. How does the budget affect the quality of the solutions?
4. What's the difference between using the different ϕ-divergence functions?

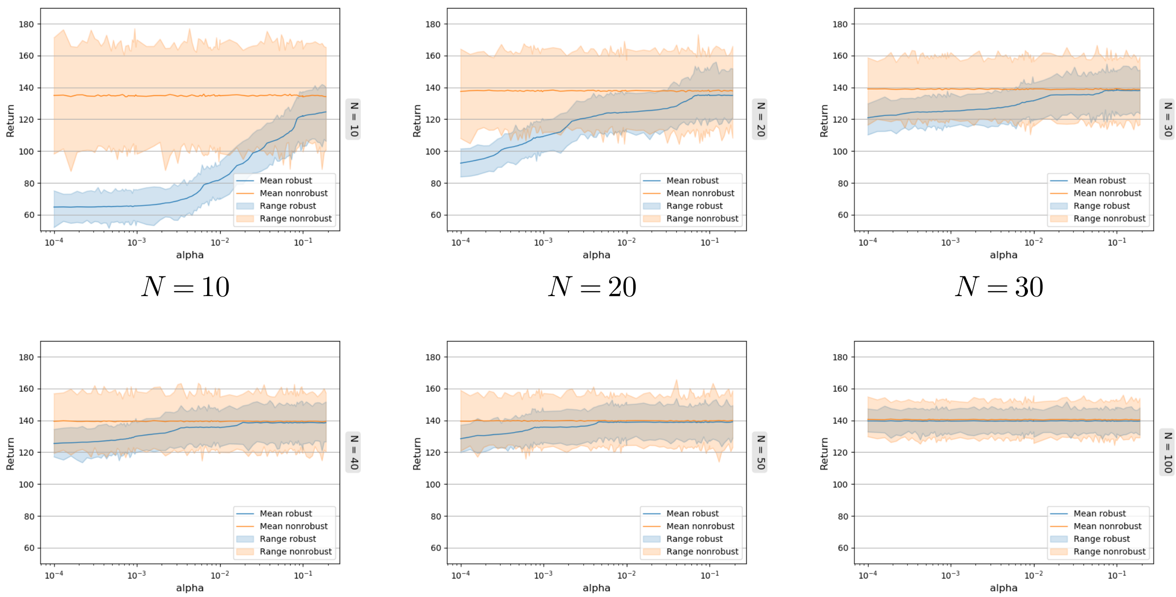
In all the following figures, the blue(orange) curve is the mean of the objective values of robust(nonrobust) solution corresponding to the sampled p-vectors. The blue(orange) area is the range of the objective values of robust(nonrobust) solution in simulation. The y-axis is the profit return value. The default setting of the experiments are same to those in the Section 6.4 in Ben-Tal et al. (2013).

2. Experiment 1: the Two alpha's

Objective:

Understand how the choice of the confidence level α corresponding to ρ (see equation (7) in the paper) and the confidence level $\bar{\alpha}$ corresponding to $\bar{\rho}$ in the simulation affect the performance.

Setup:



$N = 10$ $N = 20$ $N = 30$

$N = 40$ $N = 50$ $N = 100$

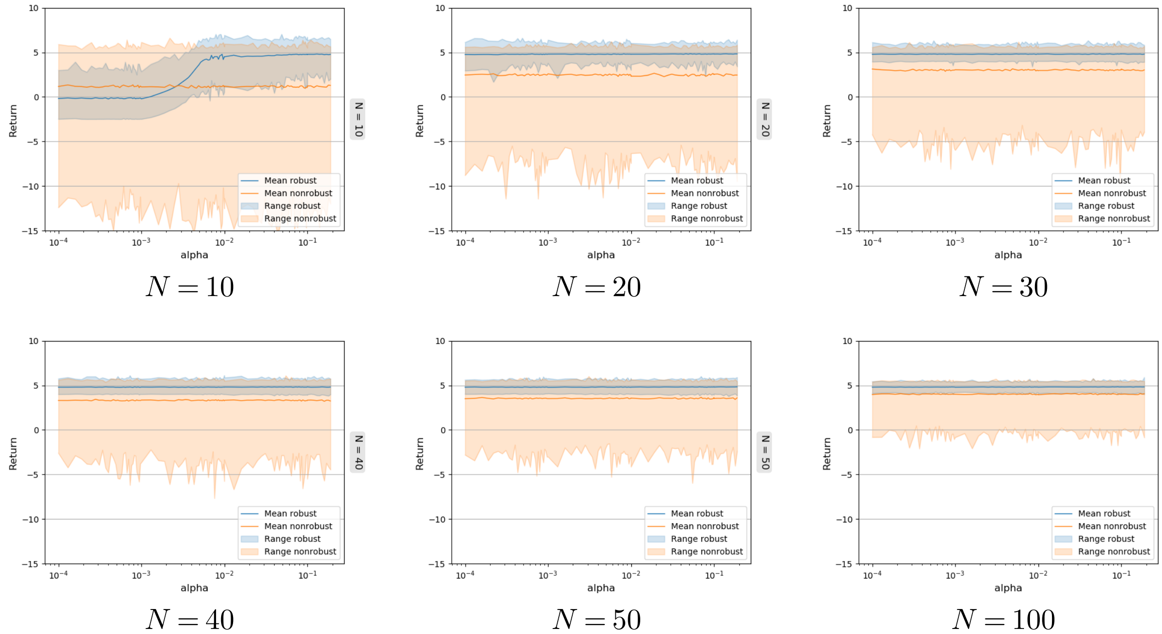
Figure 1 **Sum of return corresponding to the two solutions with different α for different sample size N.**

The settings are the same to those in Section 6.4 in the paper, except that multiple αs and $\bar{\alpha}$s are considered. I considered the range of [0.0001, 0.2] for both α and $\bar{\alpha}$.

Results:

Figure 1 and Figure 2 show that in the two problem types the profit values of fixed $\bar{\alpha}$ and different α for given sample size N. The x-axis shows the α values. We can see that when the sample size is large, the choice of the confidence level has almost no effect on the expected profit. Though for any given N, a larger α leads to a better performance in profit, especially when N is small, this is because increasing α will enlarge the size of the confidence set and thus have a lower confidence level. Therefore, this “better” performance is not desired. Lam (2019) provides a way to select ρ to obtain a better statistic guarantee for the empirical divergence-based DRO problem, but the computation needed to calculate the ρ is much more complicated than expected, so here I skip the experiment on investigating this choice of ρ.

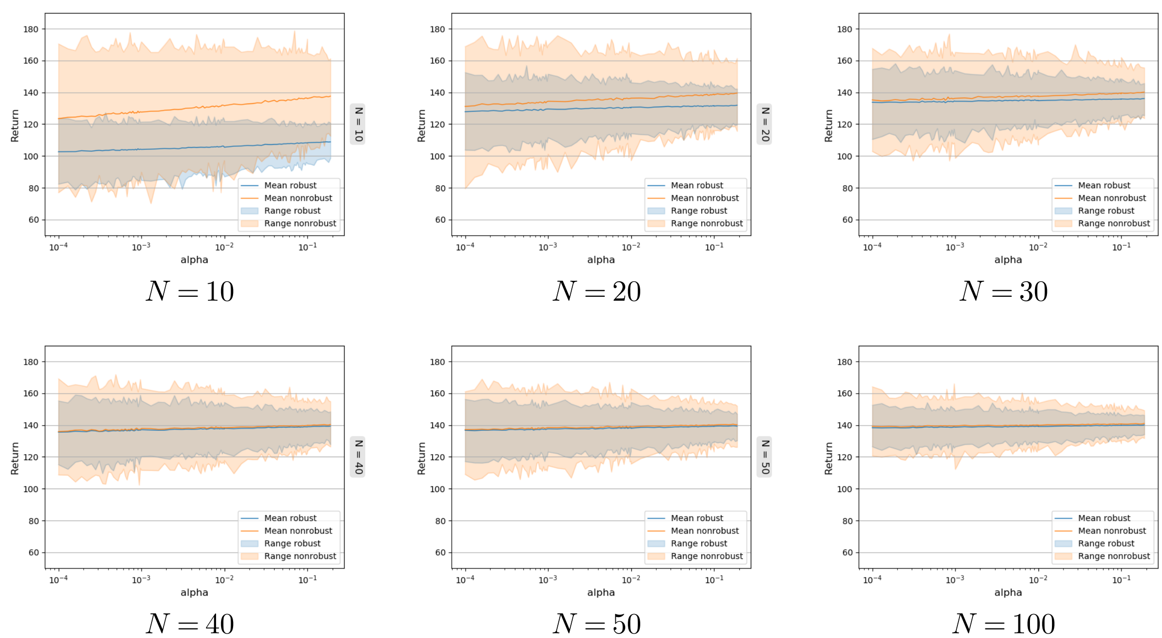
Figure 3 and Figure 4 show that in the two problem types the profit values of fixed α and different $\bar{\alpha}$ for given sample size N. The x-axis shows the $\bar{\alpha}$ values. Note that $\bar{\alpha}$ is only related to the simulation in the evaluation, and has nothing to do with the model. We can see, in both two problem types, when the $\bar{\alpha}$ is small, which means we have a higher confidence level and larger confidence set, the robust solutions performs better, and this improvement is more obvious when N is small. I conclude that the robust model performs better when the true distribution of the demand may not be close to the empirical distribution, or rather, the robust model is more robust to the true distribution, especially when the sample size is small.



$N = 10$ $N = 20$ $N = 30$

$N = 40$ $N = 50$ $N = 100$

Figure 2 **Worst case return corresponding to the two solutions with different α for different sample size N.**



$N = 10$ $N = 20$ $N = 30$

$N = 40$ $N = 50$ $N = 100$

Figure 3 **Sum of return corresponding to the two solutions with different $\bar{\alpha}$ for different sample size N.**

3. Experiment 2: A Different Simulation

Objective:

In the paper, they first observed the empirical distribution q_N, and then in evaluation, p-vectors are sampled in a 95% confidence set of the empirical distribution q_N. And thus, these p-vectors are used to evaluate the solutions. The idea behind this is that the true distribution should be in some

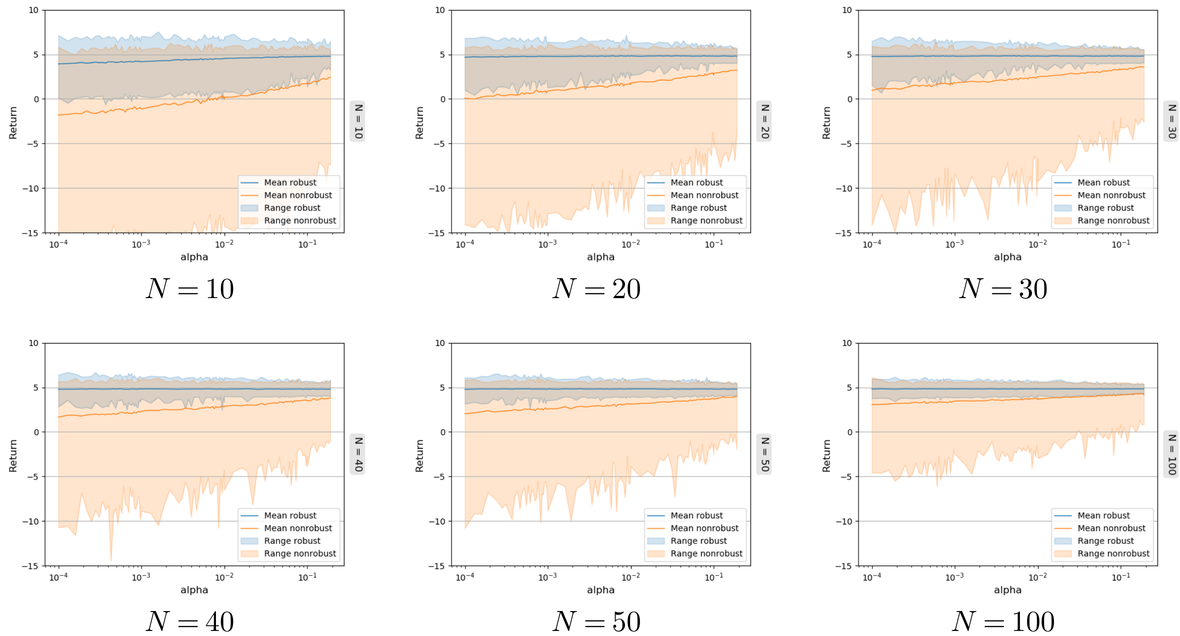


Figure 4 Worst case of return corresponding to the two solutions with different $\bar{\alpha}$ for different sample size N.

sense close to the empirical distribution, and in the paper, they used the confidence set to limit the "divergence" between the true distribution and the empirical distribution. However, since in numerical experiment, we can actually know the true distribution exactly, we can sample from the true distribution to get the empirical q_N, and again, sample from the true distribution to obtain p-vectors to evaluate the solutions. In this experiment, I would like to see how things will happen if both q_N and p-vectors are computed by the samples generated from a known true distribution.

Setup:

Assume the true distribution is given by the q_N in Table 6 in Ben-Tal et al. (2013). For each given sample size N, generate N samples from the true distribution and use these samples to compute the new empirical distribution $\tilde{q}_N$. Then follow the same step in Ben-Tal et al. (2013) with this new $\tilde{q}_N$. After obtaining the robust and nonrobust solutions, again, sample from the true distribution with given sample size N. Repeat the sampling process 10,000 times, and in each time, compute the empirical distribution as one p-vector. Thus, we have 10,000 p-vectors, and then use them to do the evaluation.

Results:

Figure 5 and Figure 6 show the performance of the solutions corresponding to the two sampling process in two problem types. We can see, in general, the performance corresponding to the two sampling process are qualitatively similar. However, the curves corresponding to the true distribution sample are less smooth, and the minimal profit of the robust solution corresponding to the true distribution in first problem type is closer to the curve of the nonrobust solution when N

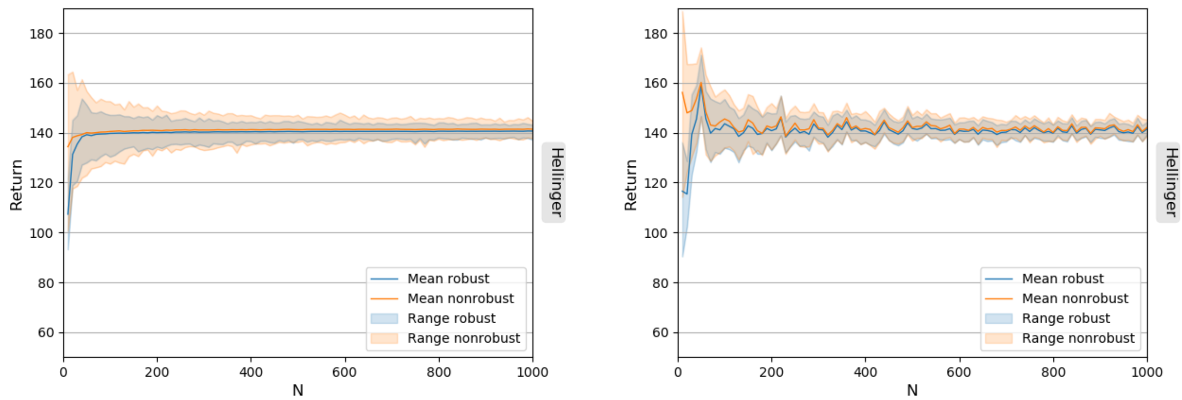


Figure 5 Sum of profit performance of two different sampling process.
(Left: sample from a confidence set. Right: sample from a true distribution.)

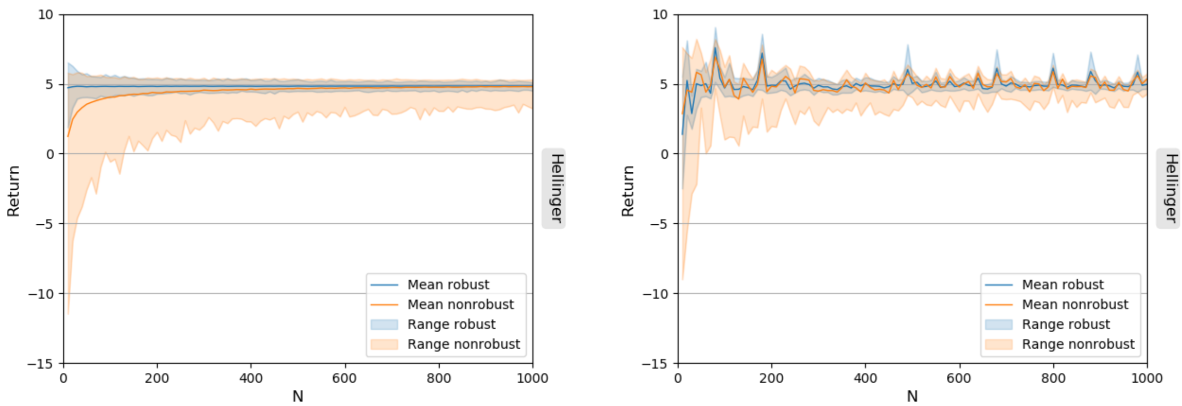


Figure 6 Worst case profit performance of two different sampling process.
(Left: sample from a confidence set. Right: sample from a true distribution.)

is large. Also, in the second problem type, the robust solution always outperforms the nonrobust solution when sampling from the confidence set, while the robust solution may perform worse than the nonrobust solution when sampling from the true distribution.

4. Experiment 3: Different Budget

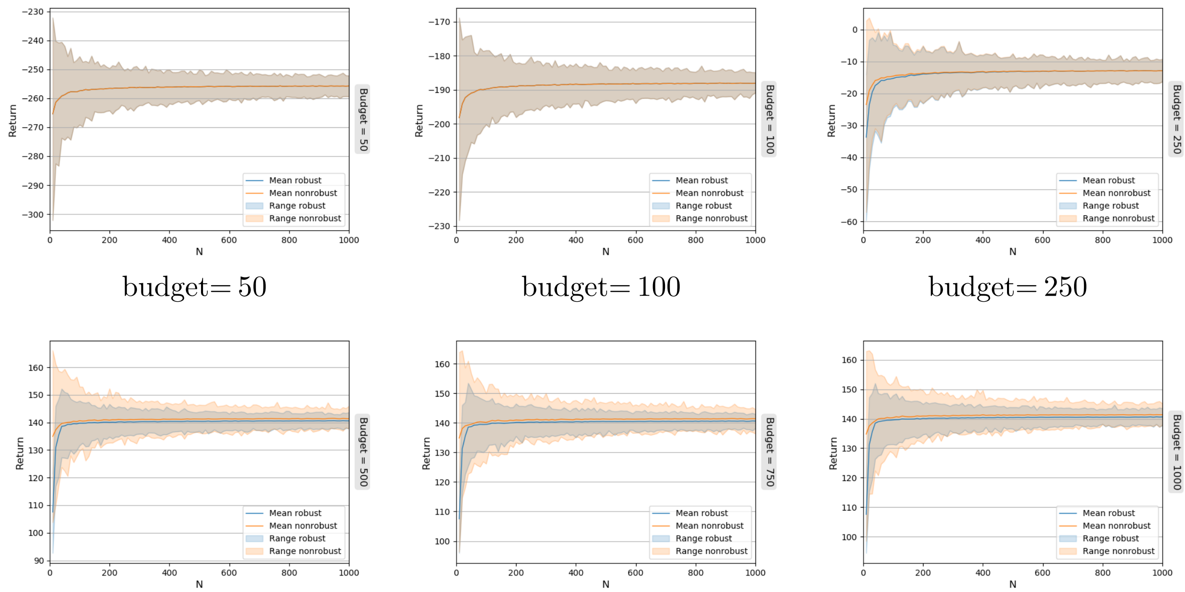
Objective:

In the paper, budget is a fixed parameter. I would like to figure out how the performance of the robust and nonrobust solution will change with different budgets.

Setup:

The settings are the same to those in Section 6.4 in the paper, except that multiple budgets are considered.

Results:



budget= 50 budget= 100 budget= 250

budget= 500 budget= 750 budget= 1000

Figure 7 Sum of return with different budgets.

Figure 7 and Figrue 8 show the performance of the two solutions with different budget in two problem types. We can find that in both problem types (sum of profits and worst case profit), when the budget is large (≥ 50), there is almost no difference between the performance corresponding to different budgets. However, when we have limited budget, we find that both robust and nonrobust solutions have very similar behavior. Actually, in the first problem type, when the budget is no more than 100, the robust solution and nonrobust solution coincide with each other. One explanation for this is that when the budget is very low, both models have less freedom to choose which and how many items to order. And thus given the limited budget, they may choose the same items to order and the order quantities are also the same. Therefore, in real-world case, when the budget is very low, there is no need to use the robust model. Instead, the nonrobust model is good enough and can be solved much faster.

5. Experiment 4: Different ϕ-Divergence

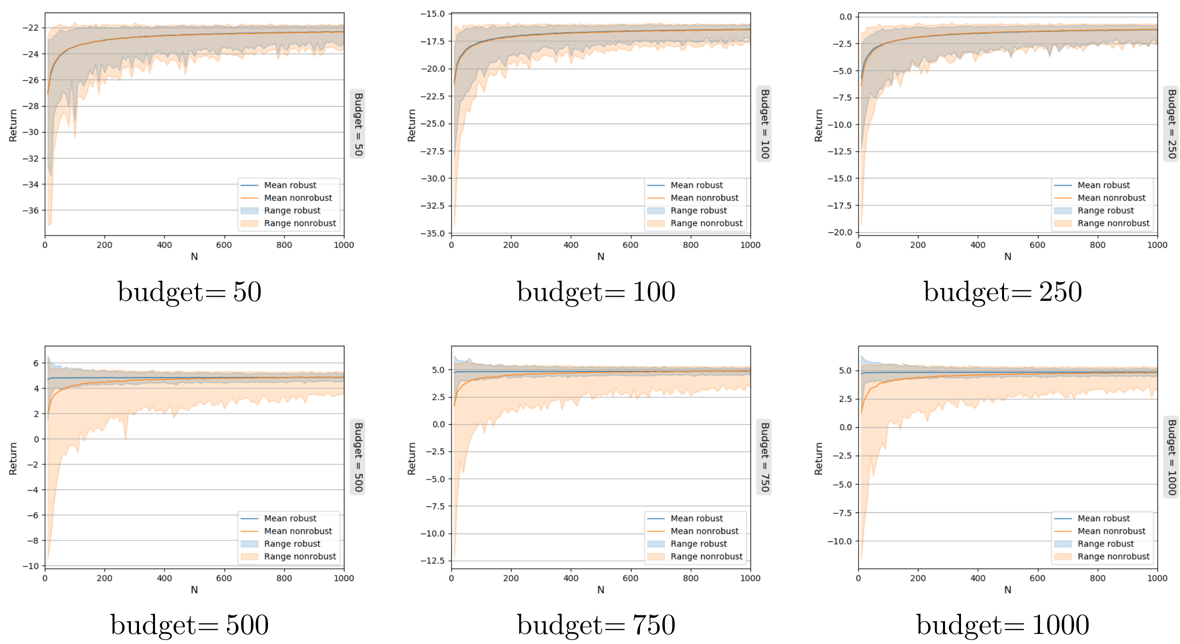
Objective:

In the above experiments, only Hellinger distance is used. Actually, different ϕ-divergence functions may lead to a different result. Though results for different divergence functions are qualitatively similar, I investigate the quantitative performance here.

Setup:

Use the default setting in the paper but different ϕ-divergence functions are used.

Results:



budget= 50 budget= 100 budget= 250

budget= 500 budget= 750 budget= 1000

Figure 8 Worst case return with different budgets.

Since the results for different divergence functions are qualitatively similar, the figures are not shown in this section, but they can be found in the appendix(Figure EC.4, Figure EC.6).

Table 1 and Table 2 show the results of the comparison between robust model with different phi-divergences and nonrobust model in two problem types. The tables contains the improvement of nonrobust model and robust models with χ^2-distance and modified χ^2-distance against the robust model with Hellinger distance in both average of the expected profit as well as the minimal expected profits over the sampled p-vectors. In Tabel 1, when N is large, the difference between the models in both the mean and minimum are very small, while in Table 2, the difference in mean is still small, but robust models have much better minimal profit. Among the three divergences, χ^2-distance performs the worst when the sample size is small, but when N is large, it performs much better. In general, modified χ^2-distance have the best performance except in the minimal profit in second problem type, where Hellinger distance is the best.

6. Conclusions

I conclude these experiments by answering the four questions asked in Section 1.

1. For given sample size N, increasing α will lead to a better performance, especially when N is small. However, this actually decreases the confidence level, and thus is not desired. While decreasing the α will obtain higher confidence level, but results in low performance. On the other hand, the robust model is more valuable when the true distribution can be quite different from the empirical distribution, especially when N is small.

Table 1 Comparison between divergence functions for sum of profit.

N	Mean				Min			
	Hellinger Distance	χ^2-Distance	Modified χ^2-Distance	Nonrobust	Hellinger Distance	χ^2-Distance	Modified χ^2-Distance	Nonrobust
10	0%	-39.95%	+14.66%	+25.19%	0%	-40.58%	+6.75%	+8.02%
20	0%	-50.00%	+3.94%	+5.19%	0%	-52.11%	+2.90%	-1.12%
30	0%	-21.01%	+2.00%	+2.46%	0%	-18.06%	+3.71%	-1.66%
40	0%	-15.08%	-0.00%	+0.66%	0%	-15.81%	-1.22%	-4.13%
50	0%	-10.24%	-2.57%	+0.59%	0%	-9.16%	-0.78%	-3.90%
60	0%	-9.69%	+1.31%	+7.20%	0%	-8.98%	-0.10%	-2.85%
70	0%	-3.40%	+0.03%	+0.56%	0%	-2.29%	+0.47%	-2.98%
80	0%	-2.34%	-0.04%	+0.60%	0%	+0.74%	-0.49%	-1.81%
90	0%	-2.34%	+0.04%	+0.55%	0%	-2.77%	+0.92%	-2.59%
100	0%	-1.68%	-0.07%	+0.55%	0%	-3.24%	+0.93%	-2.15%
250	0%	+0.11%	-0.10%	+0.61%	0%	-1.66%	-0.01%	-1.59%
500	0%	+2.17%	+0.02%	+0.55%	0%	-0.29%	-0.84%	-0.77%
750	0%	+0.23%	+0.08%	+0.55%	0%	+0.75%	+0.85%	-0.38%
1000	0%	+0.12%	+0.04%	+0.53%	0%	+0.68%	+0.73%	+0.38%
Average	0%	-1.43%	+0.22%	+0.89%	0%	-1.69%	+0.11%	-0.82%

Table 2 Comparison between divergence functions for worst case profit.

N	Mean				Min			
	Hellinger Distance	χ^2-Distance	Modified χ^2-Distance	Nonrobust	Hellinger Distance	χ^2-Distance	Modified χ^2-Distance	Nonrobust
10	0%	-102.69%	-1.94%	-73.85%	0%	-230.68%	+21.74%	-711.40%
20	0%	-104.55%	+0.21%	-48.82%	0%	-173.34%	+7.83%	-291.29%
30	0%	-0.47%	-0.68%	-38.39%	0%	-5.71%	+2.19%	-217.05%
40	0%	-0.73%	-0.43%	-31.71%	0%	-3.01%	+0.35%	-194.52%
50	0%	-0.26%	+0.51%	-26.13%	0%	-0.13%	-0.21%	-164.05%
60	0%	-0.69%	+0.04%	-23.64%	0%	+1.28%	-3.11%	-142.52%
70	0%	-0.00%	+0.69%	-21.08%	0%	+2.82%	-7.12%	-172.83%
80	0%	-0.61%	+0.53%	-19.17%	0%	-2.21%	-3.17%	-123.03%
90	0%	-0.59%	+0.47%	-17.62%	0%	+2.92%	-10.53%	-97.05%
100	0%	-0.35%	+0.52%	-15.69%	0%	+5.26%	-2.59%	-114.64%
250	0%	-0.52%	+2.60%	-7.49%	0%	+2.01%	+0.16%	-80.93%
500	0%	+1.69%	+0.16%	-3.24%	0%	-8.67%	-0.54%	-33.11%
750	0%	+1.83%	+0.09%	-1.84%	0%	-6.35%	-4.81%	-29.04%
1000	0%	+1.55%	+0.23%	-0.81%	0%	-6.10%	+0.06%	-28.11%
Average	0%	-0.91%	+0.18%	-7.04%	0%	-9.87%	-0.79%	-62.89%

2. Sampling p-vectors from the true distribution or from the confidence set lead to the similar qualitative behavior. However, the sampling process used in the paper has more smooth curves, better robust solution performance, and is easier to interpret.
3. When the budget is very low, there is no benefit to use the robust model. Indeed, the robust model and nonrobust model will generate the same solutions.
4. When N is large, the difference between Hellinger Distance, χ^2-distance and modified χ^2-distance is quite small. To make the robust model perform best, the choice of ϕ-divergence function should depend on the sample size.

References

A Ben-Tal, D denHertdog, A De Waegenaere, B Melenberg, and G Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 55(2):341–352, 2013.

Henry Lam. Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Operations Research*, 67(4):1090–1105, 2019.

E-Companion

EC.1. Links to Code Review and Github Repository

Code review: `https://www.dropbox.com/s/99490mtgviet6bx/zoom_0.mp4?dl=0`

Github repository: `https://github.com/yeyingxiao/DSO670-RSoOPAbUP-reproduction.git`

EC.2. Sanity Check

Before the experiments, a sanity check is done to make sure the coded model is on its right way. In the sanity check, the Figure 2 and Figure 3 in Ben-Tal et al. (2013) are reproduced.

To do the sanity check, I considered the multi-item Newsvendor example. For the description of the multi-item Newsvendor problem and the setting of the experiments, please see Section 6.4 in Ben-Tal et al. (2013). The optimization problem to be solved is as follows.

$$\begin{aligned} &\max\ \varphi(\tau) \\ &\text{s.t. } -c_jQ_j+\textstyle\sum_i p_i^{(j)} f_{i,j}(Q_j)\geq \tau_j, \quad \forall j, \forall p^{(j)}\in U^{(j)} \\ &\qquad \textstyle\sum_j c_jQ_j\leq \gamma, \end{aligned} \tag{EC.1}$$

where $\varphi(\tau)=\sum_j \tau_j$ when optimizing the sum of profits, and $\varphi(\tau)=min_j\tau_j$ when optimizing the worst case profit. Note that this is different from those in the paper, since the paper used $\|\tau\|$ instead of $\varphi(\tau)$, which is actually incorrect. For the nonrobust model, assume q_N is the true probability vector, which means $p^{(j)}=q_N^{(j)}$. And thus, Problem EC.1 can be solved directly by Gurobi. As to the robust model, see Section EC.3 for the reformulations corresponding to different ϕ-divergences. And the reformulation is solved by Gurobi. Figure EC.1 shows the reproduced Figure 2 and Figure 3 in Ben-Tal et al. (2013). The behavior of the two figures are similar to those in Ben-Tal et al. (2013), so I conclude that the coded models are correct, and thus they will be used in the experiments.

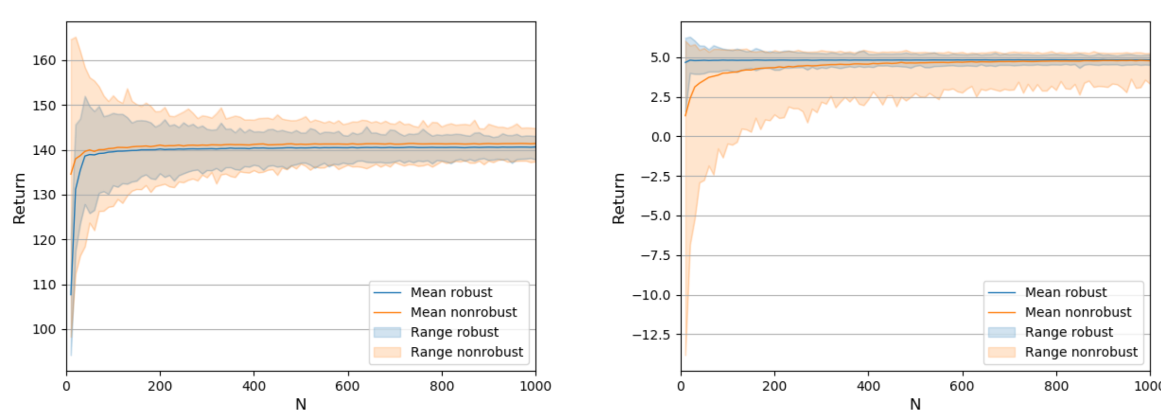


Figure EC.1 Sanity Check (Left: sum of profit. Right: worst profit)

EC.3. Reformulations of the Robust Models

The robust counterpart of Problem EC.1 is as following.

$$
\begin{aligned}
&\max\ \varphi(\tau)\\
&\text{s.t. } -c_jQ_j-\eta_j-\lambda_j\rho-\lambda_j\sum_{i=1}^{I}q_{i,N}^{(j)}\phi^*\left(\frac{-f_{i,j}(Q_j)-\eta_j}{\lambda_j}\right)\geq\tau_j,\quad\forall j=1,\dots,J\\
&\textstyle\sum_j c_jQ_j\leq\gamma\\
&\lambda\in\mathbb{R}_+^J,\ \eta\in\mathbb{R}^J,
\end{aligned}
\tag{EC.2}
$$

where $f_{i,j}\left(Q_j\right)=v_j\min\left\{d_i,Q_j\right\}+s_j\left(Q_j-d_i\right)^+-l_j\left(d_i-Q_j\right)^+$. Given the type of the ϕ-divergence, we can write ϕ^* explicitly. Also, since $v+l\geq s$, we can get rid of the $(\cdot)^+$ and $\min\{\cdot,\cdot\}$. Therefore, we can see in the following that if the ϕ-divergence is Hellinger distance, χ^2-distance or modified χ^2-distance, the problem can be reformulated as a convex SOCP problem.

EC.3.1. Hellinger Distance

$$
\begin{aligned}
&\max\ \varphi(\tau)\\
&\text{s.t. } \tau\in\mathbb{R}^J,\ Q\in\mathbb{R}_+^J,\ y\in\mathbb{R}^{I\times J},\ \lambda\in\mathbb{R}_+^J,\ \eta\in\mathbb{R}^J\\
&\quad -c_jQ_j-\eta_j-\lambda_j(\rho-1)-\tfrac{1}{4}\sum_{i=1}^{I}q_{i,N}^{(j)}y_{ij}\geq\tau_j,\quad\forall j=1,\dots,J\\
&\quad \sqrt{\lambda_j^2+\tfrac{1}{4}\left(y_{ij}-\lambda_j-f_{i,j}(Q_j)-\eta_j\right)^2}\leq\tfrac{1}{2}\left(y_{ij}+\lambda_j+f_{i,j}(Q_j)+\eta_j\right)\quad\forall i,\ j\\
&\quad \lambda_j+f_{i,j}(Q_j)+\eta_j\geq 0\quad\forall i,\ j\\
&\quad f_{i,j}(Q_j)\leq s_jQ_j-d_i(s_j-v_j)\quad\forall i=1,\dots,I\\
&\quad f_{i,j}(Q_j)\leq -l_jd_i+Q_j(l_j+v_j)\quad\forall i=1,\dots,I\\
&\quad \textstyle\sum_j c_jQ_j\leq\gamma,
\end{aligned}
\tag{EC.3}
$$

EC.3.2. χ^2 Distance

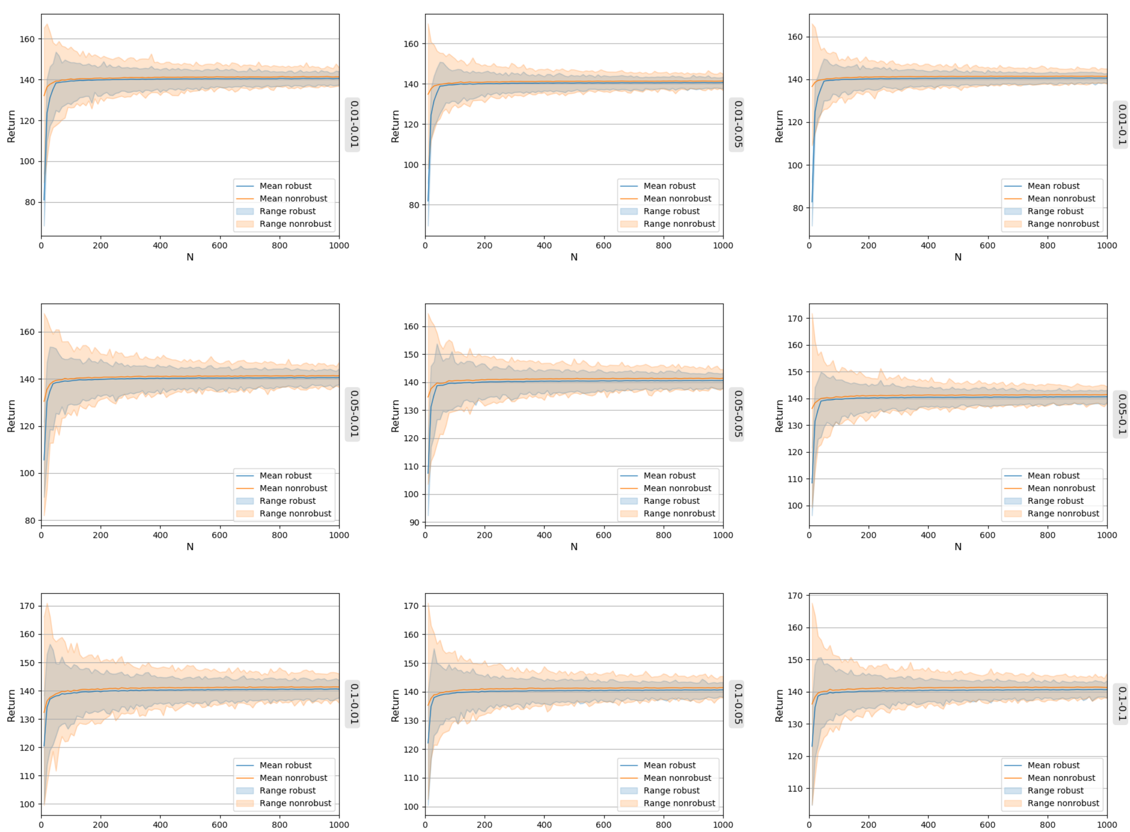
$$
\begin{aligned}
&\max\ \varphi(\tau)\\
&\text{s.t. } \tau\in\mathbb{R}^J,\ Q\in\mathbb{R}_+^J,\ y\in\mathbb{R}^{I\times J},\ \lambda\in\mathbb{R}_+^J,\ \eta\in\mathbb{R}^J\\
&\quad -c_jQ_j-\eta_j-\lambda_j(\rho+2)+2\sum_{i=1}^{I}q_{i,N}^{(j)}y_{ij}\geq\tau_j,\quad\forall j=1,\dots,J\\
&\quad \sqrt{y_{ij}^2+\tfrac{1}{4}\left(f_{i,j}(Q_j)+\eta_j\right)^2}\leq\tfrac{1}{2}\left(2\lambda_j+f_{i,j}(Q_j)+\eta_j\right)\quad\forall i,\ j\\
&\quad \lambda_j+f_{i,j}(Q_j)+\eta_j\geq 0\quad\forall i,\ j\\
&\quad f_{i,j}(Q_j)\leq s_jQ_j-d_i(s_j-v_j)\quad\forall i=1,\dots,I\\
&\quad f_{i,j}(Q_j)\leq -l_jd_i+Q_j(l_j+v_j)\quad\forall i=1,\dots,I\\
&\quad \textstyle\sum_j c_jQ_j\leq\gamma,
\end{aligned}
\tag{EC.4}
$$

EC.3.3. Modified χ^2 Distance

$$
\begin{aligned}
&\max \varphi(\tau) \\
&\text{s.t. } \tau \in \mathbb{R}^J,\ Q \in \mathbb{R}^J_+,\ y \in \mathbb{R}^{I\times J},\ z \in \mathbb{R}^{I\times J}_+,\ \lambda \in \mathbb{R}^J_+,\ \eta \in \mathbb{R}^J \\
&\qquad -c_j Q_j - \eta_j - \lambda_j(\rho - 1) - \tfrac{1}{4}\sum_{i=1}^{I} q^{(j)}_{i,N} y_{ij} \geq \tau_j, \quad \forall j = 1, \dots, J \\
&\qquad \sqrt{z_{ij}^2 + \tfrac{1}{4}\left(\lambda_j - y_{ij}\right)^2} \leq \tfrac{1}{2}\left(y_{ij} + \lambda_j\right) \quad \forall i,\ j \\
&\qquad 2\lambda_j \leq z_{ij} + f_{i,j}(Q_j) + \eta_j \quad \forall i,\ j \\
&\qquad f_{i,j}(Q_j) \leq s_j Q_j - d_i(s_j - v_j) \quad \forall i = 1, \dots, I \\
&\qquad f_{i,j}(Q_j) \leq -l_j d_i + Q_j(l_j + v_j) \quad \forall i = 1, \dots, I \\
&\qquad \textstyle\sum_j c_j Q_j \leq \gamma,
\end{aligned}
\tag{EC.5}
$$

EC.4. Additional Figures

Figure EC.2 shows different combinations of α and $\bar{\alpha}$ in the first problem type.



(a) (b) (c)

Figure EC.2 Each row has the same α and increasing $\bar{\alpha}$ and each column has the same $\bar{\alpha}$ and increasing α.

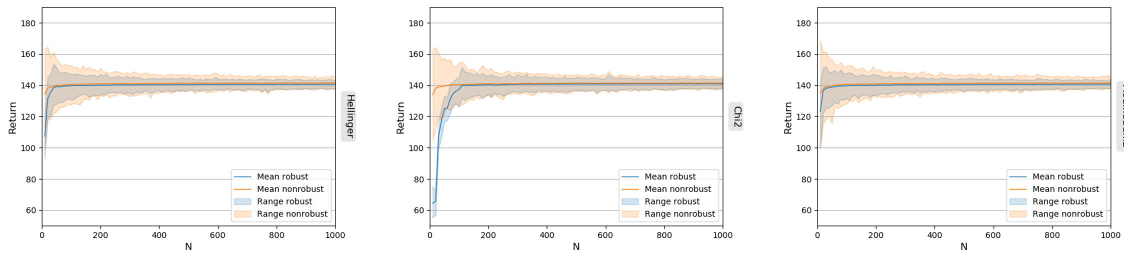
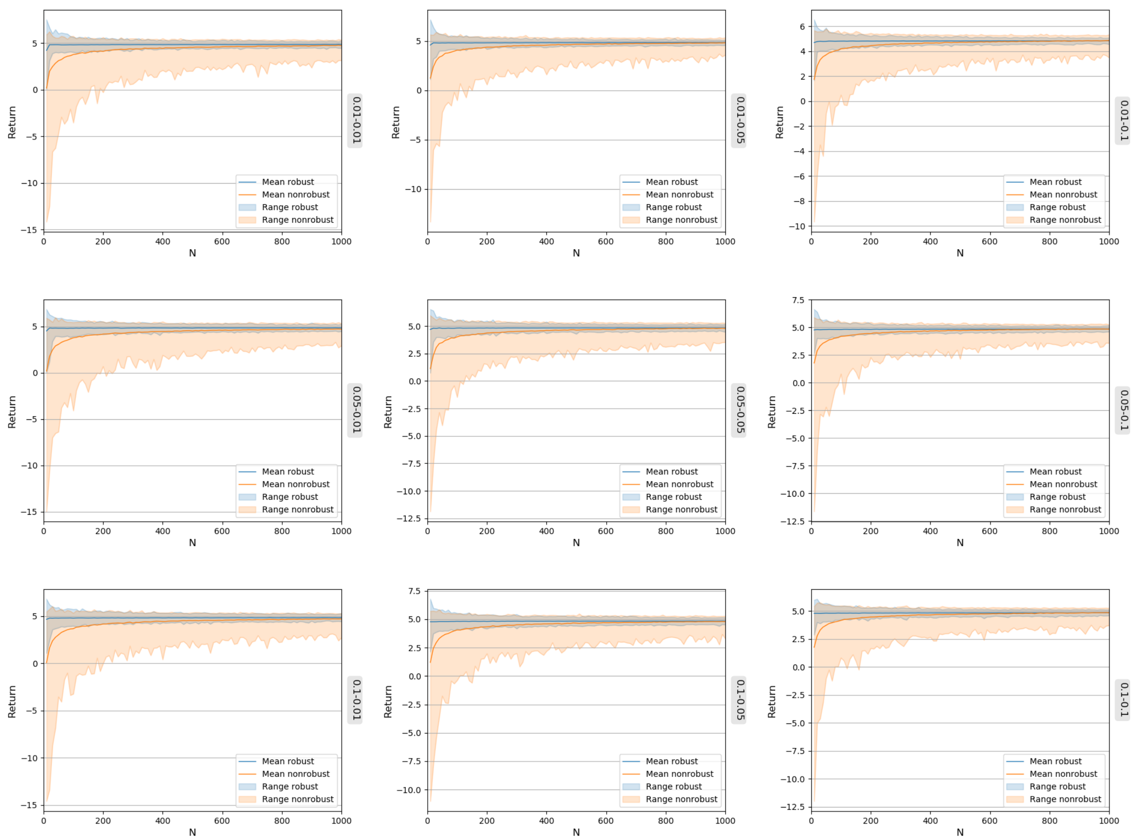


Figure EC.4 Sum of profit for different divergence functions. (sample from a confidence set)

Figure EC.3 shows different combinations of α and $\bar{\alpha}$ in the first problem type.



(a) (b) (c)

Figure EC.3 Each row has the same α and increasing $\bar{\alpha}$ and each column has the same $\bar{\alpha}$ and increasing α.

Figure EC.4, Figure EC.5, Figure EC.6 and Figure EC.7 show the performance corresponding to different phi-divergences with two sampling processes in two problem types.

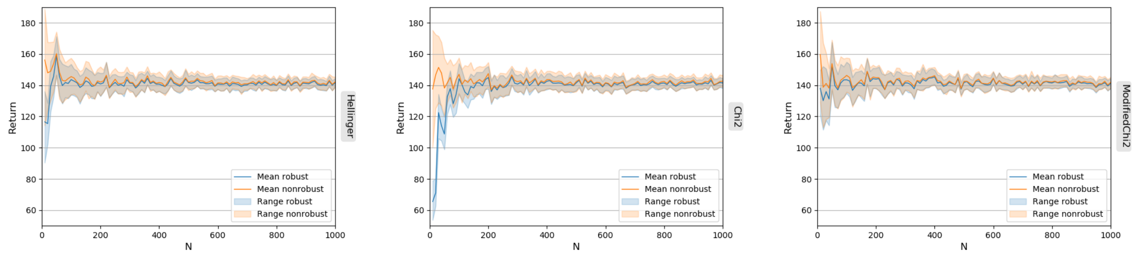


Figure EC.5 Sum of profit for different divergence functions. (sample from true distribution)

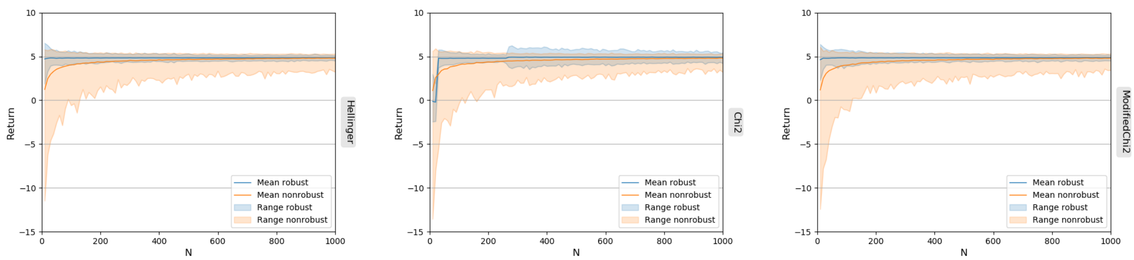


Figure EC.6 Worst case profit for different divergence functions. (sample from a confidence set)

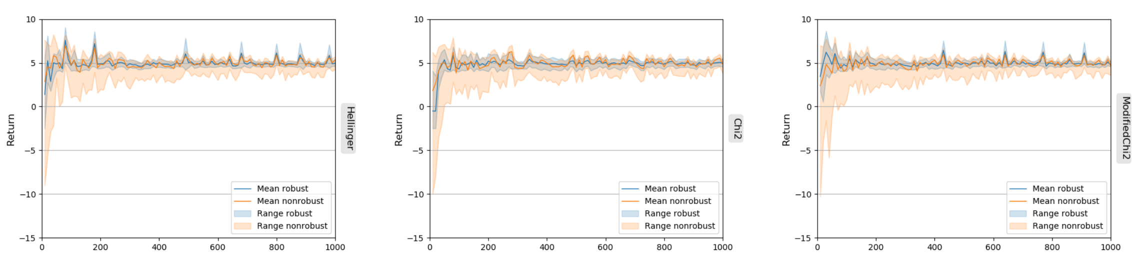


Figure EC.7 Worst case profit for different divergence functions. (sample from true distribution)