

# Robust fractional programming

Bram L. Gorissen

*Tilburg University, Department of Econometrics and Operations Research*

*5000 LE Tilburg, Netherlands*

`b.l.gorissen@tilburguniversity.edu`

## Abstract

We extend Robust Optimization to fractional programming, where both the objective and the constraints contain uncertain parameters. Earlier work did not consider uncertainty in both the objective and the constraints, or did not use Robust Optimization. Our contribution is threefold. First, we provide conditions to guarantee that either a globally optimal solution, or a sequence converging to the globally optimal solution, can be found by solving one or more convex optimization problems. Second, we identify two cases for which an exact solution can be obtained by solving a single optimization problem: (1) when uncertainty in the numerator is independent from the uncertainty in the denominator, and (2) when the denominator does not contain an optimization variable. Third, we show that the general problem can be solved with an (iterative) root finding method. The results are demonstrated on a return-on-investment maximization problem, data envelopment analysis, and mean-variance optimization. We find that the robust optimal solution is only slightly more robust than the nominal solution. As a side-result, we use Robust Optimization to show that two existing methods for solving fractional programs are dual to each other.

## 1 Introduction

A fractional program (FP) is an optimization problem, where the objective is a fraction of two functions. It can be used for an economical trade-off such as maximizing return/investment, maximizing return/risk or minimizing cost/time (Schaible and Ibaraki, 1983). A comprehensive overview of FP papers, containing over 550 references which include many applications, is given by Schaible (1982). More up-to-date references can be found in (Stancu-Minasian, 2013), which also refers to six preceding bibliographies by the same author.

More often than not, the parameters in an optimization problem are affected by uncertainty. Robust Optimization (RO) is about solving optimization problems with uncertain data in a computationally tractable way (see, e.g., Ben-Tal et al. (2009); Bertsimas et al. (2011a)). The key concept is that a solution has to be feasible for all realizations of the uncertain parameters, which are assumed to reside in convex uncertainty regions.

Sometimes, the objective is the maximum of finitely many fractions, and the feasible region is a convex set. This is called a generalized FP (see, e.g., Barros et al. (1996); Crouzeix and Ferland (1991) for solution methods). A generalized FP with infinitely many fractions in the objective was solved by Lin and Sheu (2005) using a cutting plane method, that uses a set of finitely many fractions that is extended in each step. They do not mention that their method can be used to deal with uncertain data, and do not use existing results from RO. Our work can be seen as an alternative approach, where we also deal with uncertainty in the constraints.

The Lagrange dual of a robust FP was studied by Jeyakumar et al. (2013), extending a result by Beck and Ben-Tal (2009). They assume that the uncertainty in the numerator of the objective is independent from the uncertainty in the denominator. The dual is tractable when the numerator, denominator and constraints are linear, and the uncertainty regions are ellipsoidal or finite sets of scenarios. In this paper, we focus on the primal problem. Nevertheless, in Section 4.2, we obtain and extend the list of tractable duals.

The aim of this paper is to combine FP and RO, to provide a comprehensive overview of the solution methods, and to investigate the improvement of RO on numerical examples. First, we provide conditions that guarantee that a globally optimal solution, or a sequence that converges to the globally optimal solution can be found by solving one or more convex problems. The importance of these conditions is illustrated with a numerical example from literature. Second, we identify two cases for which an exact solution can be obtained by solving a single optimization problem. Third, we show that the general problem can be solved with an (iterative) root finding method.

In Section 2, we outline two existing solution methods for FPs, and present a new result showing that the two approaches are each others duals. In Section 3, we present existing results in RO, that will be used for FPs as well. Our main contribution is given in Section 4. The results are demonstrated on a return-on-investment maximization problem, data envelopment analysis, and mean-variance optimization in Section 5.

## 2 Solving Nonrobust Fractional Programs

In this section, we present two existing methods to solve FPs, and show that these methods are dual to each other. To the best of our knowledge, this duality result is new. Consider the following general formulation of an FP:

$$(\text{FP}) \quad \min_{\mathbf{x} \in \mathbb{R}^n} \quad \frac{f(\mathbf{x})}{g(\mathbf{x})} \quad \text{s.t.} \quad h_i(\mathbf{x}) \leq 0, \quad \forall i \in I.$$

We will assume that the constraint index set  $I$  is finite, that  $f$  is convex and non-negative and that  $g$  is concave and positive over the feasible region. When the functions  $f$ ,  $g$  and  $h_i$  are affine, (FP) is a *linear* fractional program:

$$(\text{LFP}) \quad \min_{\mathbf{x} \in \mathbb{R}^n} \quad \frac{b_0 + \mathbf{b}^\top \mathbf{x}}{c_0 + \mathbf{c}^\top \mathbf{x}} \quad \text{s.t.} \quad d_{0i} + \mathbf{d}_i^\top \mathbf{x} \leq 0, \quad i \in I.$$

Charnes and Cooper show that (LFP) can be reformulated as an (equivalent) LP, by making the substitutions  $\mathbf{y} = \mathbf{x}/(c_0 + \mathbf{c}^\top \mathbf{x})$  and  $t = 1/(c_0 + \mathbf{c}^\top \mathbf{x})$  (Charnes and Cooper, 1962):

$$(\text{CC-LFP}) \quad \min_{t \in \mathbb{R}_+, \mathbf{y} \in \mathbb{R}^n} \quad b_0 t + \mathbf{b}^\top \mathbf{y} \quad \text{s.t.} \quad d_{0i} t + \mathbf{d}_i^\top \mathbf{y} \leq 0, \forall i \in I, \quad c_0 t + \mathbf{c}^\top \mathbf{y} = 1.$$

An optimal solution of (LFP) is obtained from an optimal solution of (CC-LFP) by computing  $\mathbf{x} = \mathbf{y}/t$ . Schaible (1974) shows that a similar substitution ( $\mathbf{y} = \mathbf{x}/g(\mathbf{x})$ ,  $t = 1/g(\mathbf{x})$ ) transforms (FP) into an equivalent convex programming problem:

$$(\text{Schaible-FP}) \quad \min_{t \in \mathbb{R}_{++}, \mathbf{y} \in \mathbb{R}^n} \quad t f\left(\frac{\mathbf{y}}{t}\right) \quad \text{s.t.} \quad t g\left(\frac{\mathbf{y}}{t}\right) \geq 1, \quad t h_i\left(\frac{\mathbf{y}}{t}\right) \leq 0, \forall i \in I.$$

This is indeed a convex problem, since the perspective function  $p(\mathbf{y}, t) := t f(\mathbf{y}/t)$  is jointly convex on  $\mathbb{R}^n \times \mathbb{R}_+$  when  $f$  is convex on  $\mathbb{R}^n$ . Furthermore, an optimal  $\mathbf{x}$  is obtained from  $\mathbf{x} = \mathbf{y}/t$ . Schaible also shows that, if the constraint  $t g(\mathbf{y}/t) \geq 1$  is formulated as an equality constraint:  $t g(\mathbf{y}/t) = 1$ , it is not necessary to require  $f$  to be positive (Schaible, 1974).

A different solution approach uses the auxiliary parameterized optimization problem  $F(\alpha)$ , defined as  $\min_{\mathbf{x} \in \mathbb{R}^n} \{f(\mathbf{x}) - \alpha g(\mathbf{x}) : h_i(\mathbf{x}) \leq 0 \forall i \in I\}$ . The objective value of (FP) is at least  $\alpha$  if and only if  $F(\alpha) \geq 0$  (Dinkelbach, 1967). So, the objective of (FP) equals the largest  $\alpha$  such that  $F(\alpha) \geq 0$ :

$$(\text{P-FP}) \quad \max_{\alpha \in \mathbb{R}_+} \{\alpha : \min_{\mathbf{x} \in X} \{f(\mathbf{x}) - \alpha g(\mathbf{x})\} \geq 0\},$$

where  $X$  denotes the feasible region  $\{\mathbf{x} \in \mathbb{R}^n : h_i(\mathbf{x}) \leq 0, \forall i \in I\}$ . The usual way of solving this problem is by finding the root of  $F$ , since the corresponding  $\mathbf{x}$  is optimal for (FP) (Dinkelbach, 1967). This is usually done with a Newton-like algorithm, where there is some freedom in choosing the next iteration point (Chen et al., 2009). The root of  $F$  is unique, since  $F$  is monotonically decreasing in  $\alpha$ . The parameteric program  $F(\alpha)$  is convex when  $g$  is affine or when  $f$  is non-negative on the feasible region (since then, only non-negative values for  $\alpha$  needs to be considered). For these cases, the Newton method to find the root of  $F$  was described by Dinkelbach (1967), which creates a monotonically decreasing sequence, that converges superlinearly and often (local) quadratically to a root of  $F$  (Schaible, 1976).

We now show that these approaches are dual to each other when  $f$  is non-negative. The proof for affine  $g$  and possibly negative  $f$  is similar.

**Theorem 2.1** Assume that  $f$  is convex and non-negative on  $X$ ,  $g$  is concave on  $X$ ,  $X$  is closed and convex, and the optimal value of (FP) is attained. Then (Schaible-FP) and (P-FP) are dual to each other, and strong duality holds.

**Proof.** First note that the following reformulation is equivalent to (P-FP):

$$(\text{RP-FP}) \quad \max_{\alpha \in \mathbb{R}_+} \{\alpha : f(\mathbf{x}) - \alpha g(\mathbf{x}) \geq 0, \forall \mathbf{x} \in X\}.$$

The remainder of the proof is based on the theory “primal worst is dual best” introduced by Beck and Ben-Tal (2009). They assume that  $X$  is compact and convex. Since we have assumed  $X$  to be closed and convex and since an optimal solution of (FP) is attained, compactness can be achieved by intersecting  $X$  with a box that includes the optimal solution.

Additionally, Beck and Ben-Tal assume that the constraint in (RP-FP) is convex in  $\alpha$  and concave in  $\mathbf{x}$ , which indeed holds. For fixed  $\mathbf{x}$ , (RP-FP) is an LP with the following dual:

$$(D-LP) \quad \min_{t \in \mathbb{R}_+} \{tf(\mathbf{x}) : tg(\mathbf{x}) \geq 1\}.$$

While (RP-FP) is robust since the constraint has to hold for all  $\mathbf{x}$  in  $X$ , the constraint in the optimistic counterpart of (D-LP) has to hold for a single  $\mathbf{x}$ :

$$(OD-LP) \quad \min_{t \in \mathbb{R}_+, \mathbf{x} \in X} \{tf(\mathbf{x}) : tg(\mathbf{x}) \geq 1\}.$$

(RP-FP) and (OD-LP) are dual to each other (Beck and Ben-Tal, 2009). Strong duality holds, since  $(\mathbf{x}, t)$  is a Slater point for (OD-LP) for sufficiently large  $t$ . It is obvious that (OD-LP) and (Schaible-FP) are equivalent, since  $t = 0$  is infeasible for (OD-LP). ■

### 3 Robust Optimization

There are currently two generic methods that deal with an infinite number of constraints. The first method is applicable to both linear and nonlinear constraints, while the second method can only be applied to robust LPs.

The first (“constraint wise”) approach uses conic duality (Ben-Tal et al., 2009) or Fenchel duality (Ben-Tal et al., 2015). The vector  $\mathbf{x}$  in  $\mathbb{R}^n$  satisfies the following infinite number of constraints:

$$h_i(\mathbf{a}_i, \mathbf{x}) \leq 0, \forall \mathbf{a}_i \in \mathbb{R}^L : \tau_{ij}(\mathbf{a}_i) \leq 0, \forall j \in J,$$

if and only if there exists  $u_{ij} \in \mathbb{R}_+$  and  $\mathbf{v}_{ij} \in \mathbb{R}^L$  ( $j \in J$ ,  $J$  being a finite set), such that  $\mathbf{x}$  satisfies the following convex constraint:

$$\sum_{j \in J} u_{ij} \tau_{ij}^* \left( \frac{\mathbf{v}_{ij}}{u_{ij}} \right) - (h_i)_* \left( \sum_{j \in J} \mathbf{v}_{ij}, \mathbf{x} \right) \leq 0, \quad (1)$$

where  $\tau_{ij}^*(\mathbf{s}) = \sup_{\mathbf{a}_i \in \mathbb{R}^L} \{\mathbf{s}^\top \mathbf{a}_i - \tau_{ij}(\mathbf{a}_i)\}$  and  $(h_i)_*(\mathbf{s}, \mathbf{x}) = \inf_{\mathbf{a}_i \in \mathbb{R}^L} \{\mathbf{s}^\top \mathbf{a}_i - h_i(\mathbf{a}_i, \mathbf{x})\}$  are the convex and concave conjugates of  $\tau_{ij}$  and  $h_i$ , respectively. This approach requires the constraint to be concave in  $\mathbf{a}_i$ , the functions  $\tau_{ij}$  to be convex, and  $\text{ri}(\text{dom } h_i(\cdot, \mathbf{x})) \cap \text{ri}(\mathcal{U}_i) \neq \emptyset$  for all  $\mathbf{x} \in \mathbb{R}^n$ , where  $\mathcal{U}_i := \{\mathbf{a}_i \in \mathbb{R}^L : \tau_{ij}(\mathbf{a}_i) \leq 0, \forall j \in J\}$ . This approach yields a tractable formulation for many constraints and many uncertainty sets (see Tables 1 and 2 in (Ben-Tal et al., 2015)), even if the conjugates do not have closed-form expressions. To give an impression of the broad applicability of this method, let us cite some examples, for which it provides a tractable reformulation. For uncertainty sets, one could have a norm-bounded (e.g., box or ball), polyhedral or conic representable set, or a generic set defined by (convex) power functions, exponential functions, negative logarithms, or any function for which the convex conjugate exists. Constraints could be linear or quadratic in the uncertain parameter. There are some operations on functions that preserve the availability of a tractable expression of the conjugate. One is multiplication with a non-negative scalar: the concave conjugate of  $tf(\mathbf{a}_0, \mathbf{y}/t)$  (with respect to  $\mathbf{a}_0$ ) for  $t \geq 0$ , is the perspective of the concave conjugate of  $f$ :

$tf_*(\mathbf{s}/t, \mathbf{y}/t)$ . Another one is when  $h_i$  is the sum of two functions:  $h_i = h_{i1} + h_{i2}$ . Suppose closed-form conjugates exist for  $h_{i1}$  and  $h_{i2}$  separately; then:

$$(h_i)_*(\mathbf{s}, \mathbf{x}) = \max_{\mathbf{s}_1 \in \mathbb{R}^L, \mathbf{s}_2 \in \mathbb{R}^L} \{(h_{i1})_*(\mathbf{s}_1, \mathbf{x}) + (h_{i2})_*(\mathbf{s}_2, \mathbf{x}) : \mathbf{s}_1 + \mathbf{s}_2 = \mathbf{s}\}. \quad (2)$$

When substituting (2) into (1), the max operator in (2) may be omitted, since if the resulting constraint holds for some  $\mathbf{s}_1$  and  $\mathbf{s}_2$ , then it surely holds for the maximum. This example shows that a closed-form expression for the conjugate is indeed not always required.

The second method solves any robust LP with a convex uncertainty region via its Lagrange dual (Gorissen et al., 2014). The transformation from the primal to the dual is a three step procedure. First, the dual of the nonrobust LP is formulated, where the uncertain parameters are assumed to be known. Second, instead of enforcing the constraints for all realizations of the uncertain parameters (“robust counterpart”), the constraints of the dual have to hold for a single realization of the uncertain parameters (“optimistic counterpart”). So, the uncertain parameters are added to the set of optimization variables. The last step is to reformulate the nonconvex optimistic counterpart to an equivalent convex optimization problem. The optimal solution of the resulting problem can be translated to an optimal solution of the original robust LP via the KKT vector.

In the remainder, we provide reformulations based on (1), but the reader should be aware that the other approach may be useful when all functions involved are linear.

## 4 Solving Robust Fractional Programs

In this section, we show how to solve (R-FP). It is our aim to obtain Robust Counterparts (RCs), that can be solved with existing Robust Optimization methods. First, we formulate conditions that give raise to convex optimization problems. Under these conditions, a globally optimal solution can be found by solving a single convex optimization problem (Sections 4.2 and 4.3). These conditions also guarantee that, in the general case (Section 4.4), a root finding method produces a sequence of convex optimization problems, whose solutions converge to a globally optimal solution. The results of this section are summarized in Tables 1-3.

### 4.1 Robust Formulation and Assumptions

The uncertain parameters, denoted by  $\mathbf{a}_i$ , are assumed to lie in sets  $\mathcal{U}_i \subset \mathbb{R}^L$ , which we define using functions  $\tau_{ij} : \mathbb{R}^L \rightarrow \mathbb{R}$ :

$$\mathcal{U}_i := \{\mathbf{a}_i \in \mathbb{R}^L : \tau_{ij}(\mathbf{a}_i) \leq 0, \forall j \in J\}, \quad (3)$$

where  $J$  is a finite set. In the RC of (FP), the constraints have to be satisfied by all realizations of the uncertain parameters:

$$\begin{aligned} \text{(R-FP)} \quad \min_{\alpha \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n} \quad & (\alpha) \quad \text{s.t.} \quad \frac{f(\mathbf{a}_0, \mathbf{x})}{g(\mathbf{a}_0, \mathbf{x})} - \alpha \leq 0, \forall \mathbf{a}_0 \in \mathcal{U}_0 \\ & h_i(\mathbf{a}_i, \mathbf{x}) \leq 0, \forall \mathbf{a}_i \in \mathcal{U}_i, \forall i \in I. \end{aligned} \quad (4)$$

Note that the uncertainty is specified constraint-wise, which is possible even if the parameters in different constraints are correlated (Ben-Tal et al., 2009, p. 12). We make the following assumptions:

- (a)  $\tau_{ij}$  are convex and the sets  $\mathcal{U}_i$  are convex and compact,
- (b)  $f$  and  $h_i$  are convex in  $\mathbf{x}$  for every fixed value of  $\mathbf{a}_i$  in  $\mathcal{U}_i$ ,
- (c)  $g$  is concave in  $\mathbf{x}$  for every fixed value of  $\mathbf{a}_i$  in  $\mathcal{U}_i$ ,
- (d)  $f$  and  $h_i$  are concave in  $\mathbf{a}_i$  for every feasible  $\mathbf{x}$ ,
- (e)  $g$  is convex in  $\mathbf{a}_i$  for every feasible  $\mathbf{x}$ ,
- (f)  $g(\mathbf{a}_0, \mathbf{x}) > 0$  for every  $\mathbf{a}_0$  in  $\mathcal{U}_0$  and every feasible  $\mathbf{x}$ , and
- (g)  $f(\mathbf{a}_0, \mathbf{x}) \geq 0$  for at least one  $\mathbf{a}_0 \in \mathcal{U}_0$  and for every feasible  $\mathbf{x}$ .

The last assumption is not necessary if  $g$  is biaffine, i.e., when  $g$  is affine in each parameter when the other parameter is fixed, of which we show the consequences in Section 4.5. In robust *linear* programming, the assumption that  $\mathcal{U}_i$  is compact and convex is made without any loss of generality (Ben-Tal et al., 2009, p. 12). For robust FP, compactness is not a restriction since the functions  $h_i$  are continuous and the constraints do not contain strict inequalities. So, the problem remains unchanged if the uncertainty region is replaced with its closure. However, requiring  $\mathcal{U}_i$  to be convex is a restriction (unless  $f$ ,  $g$  and  $h$  are affine in  $\mathbf{a}_i$ ), that is necessary for using existing results in RO.

Assumptions (d) and (e) are made solely because they are required by generic methods to derive a tractable RC. There are some examples where the RC can be derived even though these conditions are not fulfilled, e.g., if the uncertainty region is the convex hull of a limited number of points and the constraint is convex in the uncertain parameter, for a conic quadratic program with implementation error, or when the  $\mathcal{S}$ -lemma or a sums of squares result can be applied (Ben-Tal et al., 2009, 2015; Ben-Tal and den Hertog, 2014; Bertsimas et al., 2011b). In these cases, assumptions (d) and (e) are not necessary.

In literature, a problem is solved, that does not satisfy these requirements (Lin and Sheu, 2005). While the authors claim that this did not affect their computations, and that they find the global optimum, in Appendix A we show that their solution is suboptimal.

## 4.2 Special Case: Uncertainty in the Numerator is Independent of the Uncertainty in the Denominator

Suppose that the uncertainty in the numerator of the objective is decoupled from the uncertainty in the denominator:

$$\begin{aligned}
 \text{(R-S1)} \quad \min_{\alpha \in \mathbb{R}_+, \mathbf{x} \in \mathbb{R}^n} \quad & (\alpha) \quad \text{s.t.} \quad \frac{f(\mathbf{a}_0, \mathbf{x})}{g(\mathbf{a}'_0, \mathbf{x})} - \alpha \leq 0, \quad \forall \mathbf{a}_0 \in \mathcal{U}_0, \quad \forall \mathbf{a}'_0 \in \mathcal{U}'_0 \\
 & h_i(\mathbf{a}_i, \mathbf{x}) \leq 0, \quad \forall \mathbf{a}_i \in \mathcal{U}_i, \quad \forall i \in I.
 \end{aligned}$$

We claim that (R-S1) is equivalent to the RC of the Schaible reformulation:

$$\begin{aligned}
(\text{R-Schaible-S1}) \quad & \min_{\alpha \in \mathbb{R}_+, t \in \mathbb{R}_{++}, \mathbf{y} \in \mathbb{R}^n} \quad (\alpha) \quad \text{s.t.} \quad tf\left(\mathbf{a}_0, \frac{\mathbf{y}}{t}\right) - \alpha \leq 0, \forall \mathbf{a}_0 \in \mathcal{U}_0 \\
& tg\left(\mathbf{a}'_0, \frac{\mathbf{y}}{t}\right) \geq 1, \forall \mathbf{a}'_0 \in \mathcal{U}'_0 \\
& th_i\left(\mathbf{a}_i, \frac{\mathbf{y}}{t}\right) \leq 0, \forall \mathbf{a}_i \in \mathcal{U}_i, \forall i \in I.
\end{aligned} \tag{5}$$

Clearly, an optimal solution of (R-Schaible-S1) exists, for which  $t = \sup_{\mathbf{a}'_0 \in \mathcal{U}'_0} 1/g(\mathbf{a}'_0, \mathbf{y}/t)$ . Equivalence between (R-S1) and (R-Schaible-S1) readily follows from the substitution  $\mathbf{x} = \mathbf{y}/t$ , that converts a feasible solution of one problem to a feasible solution of the other problem.

This result extends (Jeyakumar et al., 2013). They provide the dual of (R-S1), and show that strong duality holds. In case  $f$ ,  $g$  and  $h_i$  are linear, they show that the dual of (R-S1) is tractable, when the uncertainty region is ellipsoidal, or consists of a finite set of scenarios. The resulting problems can also be obtained from our work, by applying the solution method by Gorissen et al. (2014) to (R-Schaible-S1). In addition to ellipsoids or scenarios, our method works with any convex uncertainty region, such as a polyhedral set or a conic quadratic representable set. A similar result was found by Kaul et al. (1986), but that result is wrong (see Appendix B).

We provide a reformulation of (R-Schaible-S1) if the RO method using conjugates is used (eq. (1)). The resulting equivalent problem becomes:

$$\begin{aligned}
\min \quad & (\alpha) \quad \text{s.t.} \quad \sum_{j \in J} u_{0j} \tau_{0j}^* \left( \frac{\mathbf{v}_{0j}}{u_{0j}} \right) - tf_* \left( \frac{\sum_{j \in J} \mathbf{v}_{0j}}{t}, \frac{\mathbf{y}}{t} \right) - \alpha \leq 0 \\
& \sum_{j \in J} u_{0'j} \tau_{0'j}^* \left( \frac{\mathbf{v}_{0'j}}{u_{0'j}} \right) - tg_* \left( \frac{\sum_{j \in J} \mathbf{v}_{0'j}}{t}, \frac{\mathbf{y}}{t} \right) \geq 1 \\
& \sum_{j \in J} u_{ij} \tau_{ij}^* \left( \frac{\mathbf{v}_{ij}}{u_{ij}} \right) - t(h_i)_* \left( \frac{\sum_{j \in J} \mathbf{v}_{ij}}{t}, \frac{\mathbf{y}}{t} \right) \leq 0, \forall i \in I \\
& \alpha \in \mathbb{R}_+, t \in \mathbb{R}_{++}, \mathbf{u} \in \mathbb{R}^{(|I|+2) \times |J|}, \mathbf{v} \in \mathbb{R}^{(|I|+2) \times |J| \times L}, \mathbf{y} \in \mathbb{R}^n.
\end{aligned}$$

### 4.3 Special Case: the Denominator Does Not Depend on the Optimization Variable $\mathbf{x}$

If the optimization variables do not appear in the denominator, (R-FP) is equivalent to (cf. Ben-Tal et al., 2015, Ex. 30):

$$\begin{aligned}
(\text{R-S2}) \quad & \min_{\alpha \in \mathbb{R}_+, \mathbf{x} \in \mathbb{R}^n} \quad (\alpha) \quad \text{s.t.} \quad f(\mathbf{a}_0, \mathbf{x}) - \alpha g(\mathbf{a}_0) \leq 0, \forall \mathbf{a}_0 \in \mathcal{U}_0 \\
& h_i(\mathbf{a}_i, \mathbf{x}) \leq 0, \forall \mathbf{a}_i \in \mathcal{U}_i, \forall i \in I.
\end{aligned}$$

Note that  $g$  indeed does not depend on  $\mathbf{x}$ . (R-S2) can be solved via the following equivalent convex reformulation using (1), and standard techniques for the conjugate of the sum of two

functions:

$$\begin{aligned}
\min \quad (\alpha) \quad \text{s.t.} \quad & \sum_{j \in J} t_{0j} \tau_{0j}^* \left( \frac{\mathbf{v}_{0j}}{t_{0j}} \right) - f_* \left( \mathbf{s} + \sum_{j \in J} \mathbf{v}_{0j}, \mathbf{x} \right) + \alpha g^* \left( \frac{\mathbf{s}}{\alpha} \right) \leq 0 \\
& \sum_{j \in J} t_{ij} \tau_{ij}^* \left( \frac{\mathbf{v}_{ij}}{t_{ij}} \right) - (h_i)_* \left( \sum_{j \in J} \mathbf{v}_{ij}, \mathbf{x} \right) \leq 0, \quad \forall i \in I \\
& \alpha \in \mathbb{R}_+, \mathbf{t} \in \mathbb{R}_+^{(|I|+1) \times |J|}, \mathbf{s} \in \mathbb{R}^L, \mathbf{v} \in \mathbb{R}^{(|I|+1) \times |J| \times L}, \mathbf{x} \in \mathbb{R}^n.
\end{aligned}$$

#### 4.4 General Case

We now show how to solve the general problem (R-FP) using the following parametric problem:

$$\begin{aligned}
F(\alpha) := \min_{\mathbf{x}, w} \quad (w) \quad \text{s.t.} \quad & f(\mathbf{a}_0, \mathbf{x}) - \alpha g(\mathbf{a}_0, \mathbf{x}) \leq w, \quad \forall \mathbf{a}_0 \in \mathcal{U}_0 \\
& h_i(\mathbf{a}_i, \mathbf{x}) \leq 0, \quad \forall \mathbf{a}_i \in \mathcal{U}_i, \quad \forall i \in I,
\end{aligned} \tag{6}$$

which is a convex optimization problem, since we only have to solve it for  $\alpha \in \mathbb{R}_+$ . Let  $\alpha^*$  be a root of  $F$ . Lin and Sheu show that an optimal solution of (R-FP) is the minimizer  $\mathbf{x}$  of  $F(\alpha^*)$ , if the feasible region for  $\mathbf{x}$  is compact (Lin and Sheu, 2005). We assume from now on that the constraint functions  $h_i$  define a compact feasible region. Moreover, they show that  $F(\alpha) < 0$  if and only if  $\alpha > \alpha^*$ . Lin and Sheu do not use results from RO to arrive at the deterministic reformulation (7). Instead, they replace the set  $\mathcal{U}_0$  with a finite set, approximate  $F(\alpha)$  with an entropic regularization method, and iteratively generate a sequence  $\tilde{\alpha}_k$  that converges to  $\alpha^*$ . The approximation becomes more accurate as the root of  $F$  is approached. The reason why they approximate  $F(\alpha)$  is because they claim that computing its value is difficult. Our approach is to solve  $F(\alpha)$  using RO, which inherently produces tractable problems. The following convex reformulation using (1) is equivalent:

$$\begin{aligned}
F(\alpha) = \min \quad (w) \quad \text{s.t.} \quad & \sum_{j \in J} t_{0j} \tau_{0j}^* \left( \frac{\mathbf{v}_{0j}}{t_{0j}} \right) - f_* \left( \mathbf{s} + \sum_{j \in J} \mathbf{v}_{0j}, \mathbf{x} \right) + \alpha g^* \left( \frac{\mathbf{s}}{\alpha}, \mathbf{x} \right) \leq w \\
& \sum_{j \in J} t_{ij} \tau_{ij}^* \left( \frac{\mathbf{v}_{ij}}{t_{ij}} \right) - (h_i)_* \left( \sum_{j \in J} \mathbf{v}_{ij}, \mathbf{x} \right) \leq 0, \quad \forall i \in I \\
& \mathbf{t} \in \mathbb{R}_+^{(|I|+1) \times |J|}, \mathbf{s} \in \mathbb{R}^L, \mathbf{v} \in \mathbb{R}^{(|I|+1) \times |J| \times L}, w \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n.
\end{aligned} \tag{7}$$

Since  $F$  is monotonically decreasing in  $\alpha$ , as for FPs and generalized FPs, existing root-finding methods can be used. We mention a few of these that produce a sequence  $\{\alpha_k\}$  which converges to  $\alpha^*$ :

(a) The bisection method. Bounds on the interval that contain  $\alpha^*$  are:

$$\alpha_{LB} := \min_{\mathbf{x} \in \mathbb{R}^n} \{ f(\bar{\mathbf{a}}_0, \mathbf{x}) / g(\bar{\mathbf{a}}_0, \mathbf{x}) : h_i(\mathbf{a}_i, \mathbf{x}) \leq 0, \quad \forall \mathbf{a}_i \in \mathcal{U}_i, \quad \forall i \in I \} \tag{8}$$

$$\alpha_{UB} := \sup_{\mathbf{a}_0 \in \mathcal{U}_0} f(\mathbf{a}_0, \mathbf{x}) / g(\mathbf{a}_0, \mathbf{x}), \tag{9}$$



where (8) is computed for a fixed  $\bar{\mathbf{a}}_0$  from  $\mathcal{U}_0$ , and (9) is computed for some  $\mathbf{x}$  that is (robust) feasible, i.e., for an  $\mathbf{x}$  that satisfies (4). These bounds can be computed relatively easily using the Schaible reformulation. If the lower bound (8) is hard to compute due to the “for all” quantifier in the constraints, it may be computed for fixed  $\bar{\mathbf{a}}_i$  from  $\mathcal{U}_i$ . Since  $F(\alpha_{LB}) \geq 0$  and  $F(\alpha_{UB}) \leq 0$ , and since  $F$  is clearly nonincreasing,  $\alpha^*$  lies in  $[\alpha_{LB}, \alpha_{UB}]$ . The middle point of this interval is  $\alpha_k := 0.5(\alpha_{LB} + \alpha_{UB})$ . By evaluating  $F(\alpha_k)$ , the width of the interval that contains  $\alpha^*$  can be halved: if  $F(\alpha_k) > 0$ , then set  $\alpha_{LB} = \alpha_k$ , otherwise set  $\alpha_{UB} = \alpha_k$ . By increasing  $k$  by 1 and repeating this procedure, a series  $\{\alpha_k\}$  is constructed, that converges to  $\alpha^*$ .

- (b) The Dinkelbach type algorithm by Crouzeix et al. (1985), adjusted for infinitely many ratios. The method starts with  $k = 0$  and  $\alpha_k = \sup_{\mathbf{a}_0 \in \mathcal{U}_0} f(\mathbf{a}_0, \mathbf{x})/g(\mathbf{a}_0, \mathbf{x})$  for some feasible  $\mathbf{x}$ . Then  $F(\alpha_k)$  is computed, with maximizer  $\mathbf{x}_k$ . If  $F(\alpha_k) < 0$ , then the next  $\alpha$  is determined by  $\alpha_{k+1} := \max_{\mathbf{a}_0 \in \mathcal{U}_0} f(\mathbf{a}_0, \mathbf{x}_k)/g(\mathbf{a}_0, \mathbf{x}_k)$ . Computing  $\alpha_{k+1}$  requires solving an FP. The method proceeds by increasing  $k$  by 1, and again computing  $F(\alpha_k)$ . If the feasible region for  $\mathbf{x}$  is compact, then the series  $\{\alpha_k\}$  converges linearly to  $\alpha_k$ .
- (c) The same as method (b), except that the right-hand side of (6) is multiplied with  $g(\mathbf{a}_0, \mathbf{x}_k)$ :  $f(\mathbf{a}_0, \mathbf{x}) - \alpha g(\mathbf{a}_0, \mathbf{x}) \leq w g(\mathbf{a}_0, \mathbf{x}_k)$ . This may increase the speed of convergence for the same complexity of computation (Crouzeix et al., 1986).
- (d) The same as method (b) or (c), except that the  $\mathbf{a}_0$  that maximizes  $F(\alpha_k)$  is used to compute  $\alpha_{k+1}$ , instead of solving a new optimization problem. The worst case  $\mathbf{a}_0$  in the computation of  $F(\alpha_k)$  can be recovered without much computational effort. Thus,  $\alpha_{k+1} = f(\mathbf{a}_0, \mathbf{x}_k)/g(\mathbf{a}_0, \mathbf{x}_k)$  is computed more efficiently than in the method (b) or (c). Additional work is required to ensure convergence of  $\{\alpha_k\}$  to  $\alpha^*$  (Crouzeix and Ferland, 1991, Section 5).

Let  $\mathbf{x}_k$  be the maximizer of  $F(\alpha_k)$ . If a root finding method finds the root in a finite number of steps, then an exact solution of (R-FP) is found. Otherwise, Crouzeix et al. show that, if the sequence  $\{\alpha_k\}$  converges to  $\alpha^*$ , then any convergent subsequence of  $\{\mathbf{x}_k\}$  converges to the optimal solution  $\mathbf{x}^*$  of (R-FP) (Crouzeix et al., 1985, Theorem 4.1c).

## 4.5 Consequences when the Denominator is Biaffine

The assumption that the numerator is positive, ensures that the objective value of (R-FP) is positive over the feasible region. Consequently, we could assume  $\alpha \in \mathbb{R}_+$ ; this would produce a convex optimization problems. If  $g$  is biaffine, then the resulting problems are also convex for  $\alpha < 0$ . We shall discuss the results to each of the three aforementioned cases separately. For the *first special case* (Section 4.2), the restriction that the numerator is positive may be dropped only if the denominator does not contain an uncertain parameter. Then, (R-S1) and (R-Schaible-S1) are equivalent if  $\alpha \in \mathbb{R}_+$  is replaced with  $\alpha \in \mathbb{R}$  and (5) is stated as an equality (cf. Schaible (1974)). The reason why the denominator may not contain an uncertain parameter, is because  $t = 1/g(\mathbf{x}, \mathbf{a}_{01})$  is not possible for multiple  $\mathbf{a}_{01}$ .

For the *second special case* (Section 4.3), the denominator only depends on  $\mathbf{a}_0$ , so “bi-affine” in the title of this subsection should be read as “affine”. When (R-S2) is solved for  $\alpha \in \mathbb{R}$ , the restriction that the numerator is positive, may be dropped.

For the *general case* (Section 4.4), no changes need to be made to drop the restriction that the numerator is positive.

Table 1: Tractable cases when uncertainty in the numerator is independent of the uncertainty in the denominator.  $\ell$  denotes an affine function.

$f$	$g$	$h_i$	$\text{sgn}(f)$	(R-Schaible-S1)
$f(\mathbf{a}_0, \mathbf{x})$	$g(\mathbf{a}'_0, \mathbf{x})$	$h_i(\mathbf{a}_i, \mathbf{x})$	$\geq 0$	no modifications
$f(\mathbf{a}_0, \mathbf{x})$	$\ell(\mathbf{x})$	$h_i(\mathbf{a}_i, \mathbf{x})$	any	$\alpha \in \mathbb{R}$ , (5) as an equality

Table 2: Tractable cases when the denominator does not depend on  $\mathbf{x}$ .  $\ell$  denotes an affine function.

$f$	$g$	$h_i$	$\text{sgn}(f)$	(R-S2)
$f(\mathbf{a}_0, \mathbf{x})$	$g(\mathbf{a}_0)$	$h_i(\mathbf{a}_i, \mathbf{x})$	$\geq 0$	no modifications
$f(\mathbf{a}_0, \mathbf{x})$	$\ell(\mathbf{a}_0)$	$h_i(\mathbf{a}_i, \mathbf{x})$	any	$\alpha \in \mathbb{R}$

Table 3: Tractable cases for the general case.  $\ell$  denotes a biaffine function.

$f$	$g$	$h_i$	$\text{sgn}(f)$
$f(\mathbf{a}_0, \mathbf{x})$	$g(\mathbf{a}_0, \mathbf{x})$	$h_i(\mathbf{a}_i, \mathbf{x})$	$\geq 0$
$f(\mathbf{a}_0, \mathbf{x})$	$\ell(\mathbf{a}_0, \mathbf{x})$	$h_i(\mathbf{a}_i, \mathbf{x})$	any

## 5 Numerical Examples

In this section, we test our method on three examples: a multi-item newsvendor problem (Section 5.1), mean-variance optimization (Section 5.2), and data envelopment analysis (Section 5.3).

### 5.1 Multi-item Newsvendor Example

In Gorissen et al. (2014), a multi-item newsvendor problem is solved by minimizing the investment cost under the condition that at least a certain expected profit is made.

We show how to directly optimize the expected return on investment for this example. Let us first recapitulate the problem. The newsvendor buys  $Q_i$  units of item  $i$  at the beginning of the day. Each item has its associated ordering cost  $c_i$ , selling price  $v_i$ , salvage price  $r_i$ , and unsatisfied demand loss  $l_i$ . We assume  $r_i \leq v_i + l_i$ . During the day the newsvendor faces a demand  $d_i$ , resulting in a profit of  $\min\{v_i Q_i + l_i(Q_i - d_i) - c_i Q_i, v_i d_i + r_i(Q_i - d_i) - c_i Q_i\}$ . The demand is not known in advance, but there are finitely many demand scenarios  $d_{is}$  ( $s$  in  $S$ ) that occur with (uncertain) probability  $p_{is}$ , independently of other items.

The problem of maximizing expected return on investment can be formulated as:

$$\begin{aligned}
(\text{R-NV}) \quad & \max_{\mathbf{Q} \in \mathbb{R}_+^{|I|}, \mathbf{u} \in \mathbb{R}^{|I| \times |S|}} \min_{\mathbf{p}_i \in \mathcal{U}_i} \frac{\sum_{i \in I} \sum_{s \in S} p_{is} u_{is}}{\sum_{i \in I} c_i Q_i} \\
\text{s.t.} \quad & u_{is} + (c_i - r_i) Q_i \leq d_{is} (v_i - r_i), \forall i \in I, \forall s \in S \\
& u_{is} + (c_i - v_i - l_i) Q_i \leq -d_{is} l_i, \forall i \in I, \forall s \in S,
\end{aligned}$$

where  $u_{is}$  is the contribution to the profit of item  $i$  in scenario  $s$ , and the convex and compact uncertainty regions  $\mathcal{U}_i$  are defined using the Matusita distance, which is a  $\phi$ -divergence measure (Ben-Tal et al., 2013):

$$\mathcal{U}_i = \left\{ \mathbf{p}_i \in \mathbb{R}_+^{|S|} : \sum_{s \in S} p_{is} = 1, \sum_{s \in S} |(\hat{p}_{is})^\alpha - (p_{is})^\alpha|^{1/\alpha} \leq \rho, \forall i \in I \right\}.$$

Note that the assumptions (a)-(e) and (g) are always fulfilled, and that assumption (e) is fulfilled if at least one item is bought. (R-NV) can be classified under the first special case (Section 4.2). Since all functions are affine and the denominator is certain, the Schaible reformulation and the Charnes-Cooper reformulation are equivalent. (R-NV) is therefore equivalent to:

(R-CC-NV)

$$\begin{aligned}
& \max_{\mathbf{Q} \in \mathbb{R}_+^{|I|}, t \in \mathbb{R}_+, \mathbf{u} \in \mathbb{R}^{|I| \times |S|}} \min_{\mathbf{p}_i \in \mathcal{U}_i} \sum_{i \in I} \sum_{s \in S} p_{is} u_{is} \\
\text{s.t.} \quad & u_{is} + (c_i - r_i) Q_i \leq d_{is} (v_i - r_i) t, \forall i \in I, \forall s \in S \\
& u_{is} + (c_i - v_i - l_i) Q_i \leq -d_{is} l_i t, \forall i \in I, \forall s \in S \\
& \sum_{i \in I} c_i Q_i = 1,
\end{aligned}$$

where  $\mathbf{Q}$  and  $\mathbf{u}$  in (R-CC-NV) have to be divided by  $t$  to obtain the  $\mathbf{Q}$  and  $\mathbf{u}$  in (R-NV). (R-CC-NV) is a linear program with a convex uncertainty region that we solve via its dual, as outlined in the introduction. The last of the reformulation steps, a substitution, is not necessary since the uncertainty only appears in the objective. Let  $x_{is}$ ,  $y_{is}$  and  $z$  be the dual variables of (CC-NV); then the optimistic dual (OD-CC-NV) is given by:

$$\begin{aligned}
(\text{OD-CC-NV}) \quad & \min \quad z \\
\text{s.t.} \quad & x_{is} + y_{is} = p_{is}, \forall i \in I, \forall s \in S \\
& \sum_{s \in S} \{(c_i - r_i)x_{is} + (c_i - v_i - l_i)y_{is}\} + c_i z \geq 0, \forall i \in I \tag{10} \\
& \sum_{i \in I} \sum_{s \in S} -d_{is}(v_i - r_i)x_{is} + d_{is}l_i y_{is} \geq 0 \tag{11} \\
& \sum_{s \in S} p_{is} = 1, \forall i \in I \\
& \sum_{s \in S} |(\hat{p}_{is})^\alpha - (p_{is})^\alpha|^{1/\alpha} \leq \rho, \forall i \in I \\
& \mathbf{p} \in \mathbb{R}_+^{|S|}, \mathbf{x} \in \mathbb{R}_+^{|I| \times |S|}, \mathbf{y} \in \mathbb{R}_+^{|I| \times |S|}, z \in \mathbb{R}.
\end{aligned}$$

The optimal value of (OD-CC-NV) is the robust expected return on investment. The corresponding optimal order quantities  $\mathbf{Q}$  can be derived from the KKT vector of (OD-CC-NV), by dividing its elements associated with (10) by the element corresponding to (11). This is the same as dividing  $\mathbf{Q}$  by  $t$  in (R-CC-NV) to undo the Charnes-Cooper transformation.

We solve the problem for the same data as Gorissen et al. (2014) with AIMMS 3.11 (Paragon Decision Technology, the Netherlands) and KNITRO 7.0 (Zienna Optimization LLC, USA) with its default settings. Computation errors for negative  $p_{is}$  were avoided by using  $|p_{is}|^\alpha$  instead of  $(p_{is})^\alpha$ . We take  $\rho = 0$  to obtain the nominal solution, while  $\rho = 0.03$  for the robust solution. Solutions were obtained in less than 0.01 seconds. When the probabilities are as expected ( $\rho = 0$ ), the expected return on investment of the nominal solution is 0.297, while for the robust solution it is 0.285. When  $\rho = 0.03$  and the worst case probabilities occur for the nominal solution, i.e., the probabilities that minimize the expected return on investment for the nominal solution, the objective value drops to 0.211, while for the robust solution it drops to 0.214. So, the solution indeed becomes more robust, but the difference with the nominal solution is small. We verify if the decision maker could have done better, if he knew beforehand which probability vector realizes. This done by optimizing the nominal model ( $\rho = 0$ ), while setting the probability estimates  $\hat{p}_{is}$  equal to the worst case probabilities for the robust solution. This gives the so-called perfect hindsight solution. The objective value is as low as 0.214. So, even though the robust objective could deteriorate substantially, there is no other solution that performs better.

## 5.2 Mean-variance Optimization

We are to present an example that involves a trade-off between mean and variance. This trade-off is commonly used in portfolio optimization, including the Modern Portfolio Theory (MPT) founded by Markowitz (1952), where the goal is to select the right mix of assets. In contrast to MPT, we do not impose that the expected returns on the assets and the covariance matrix are fully known. Instead, we assume that finitely many scenarios  $s$  (in  $S$ ) for the future can be identified along with unknown probabilities of occurrence  $p_s$ , which are estimated by  $\hat{p}_s$ . The return of asset  $i$  in scenario  $s$  is a constant  $r_{is}$ , so when  $x_i$  units of money are invested in asset  $i$ , the return in scenario  $s$  is given by  $u_s = \sum_{i \in I} r_{is} x_i$  (possibly negative). The expected return and variance are given by:

$$\begin{aligned} \mathbb{E}(\text{return}) &:= \sum_{s \in S} p_s u_s \\ \text{Var}(\text{return}) &:= \sum_{s \in S} p_s (u_s - \mathbb{E}(\text{return}))^2 \end{aligned} \tag{12}$$

$$= \sum_{s \in S} p_s u_s^2 - \left( \sum_{s \in S} p_s u_s \right)^2. \tag{13}$$

To remain in the minimization framework, the objective is to minimize the variance-to-mean ratio (or the dispersion index). The robust optimization problem is given by:

$$\begin{aligned}
(\text{R-I}) \quad & \min_{\mathbf{x} \in \mathbb{R}_+^{|I|}, \mathbf{u} \in \mathbb{R}^{|S|}} \max_{\mathbf{p} \in \mathcal{U}} \frac{\sum_{s \in S} p_s u_s^2 - (\sum_{s \in S} p_s u_s)^2}{\sum_{s \in S} p_s u_s} \\
& \text{s.t.} \quad u_s = \sum_{i \in I} r_{is} x_i, \quad \forall s \in S \\
& \quad \sum_{i \in I} x_i = C \\
& \quad \sum_{s \in S} p_s u_s > 0, \quad \forall \mathbf{p} \in \mathcal{U}.
\end{aligned}$$

The last two constraints ensure that  $C$  units of money are invested, and that this model has a feasible solution only if the expected profit is positive. The numerator is convex in  $u_s$  (from (12)) and concave in  $p_s$  (from (13)). The denominator is clearly concave in  $p_s$  and convex in  $u_s$ . Moreover, the numerator is non-negative and the denominator is positive on the feasible region. For the uncertainty region we use the modified  $\chi^2$ -distance as a  $\phi$ -divergence measure, which can be justified by statistical theory (Ben-Tal et al., 2013):

$$\mathcal{U}_0 := \left\{ \mathbf{p} \in \mathbb{R}_+^{|S|} : \sum_{s \in S} p_s = 1, \sum_{s \in S} \frac{(p_s - \hat{p}_s)^2}{\hat{p}_s} \leq \rho \right\}.$$

(R-I) is not one of the special cases, so we solve this problem using the general method. In order to formulate (7) explicitly, we first derive some conjugate functions. The conjugate for  $f$  is from Ben-Tal et al. (2015, Ex. 25).

$$\begin{aligned}
f(\mathbf{p}, \mathbf{u}) &:= \sum_{s \in S} p_s u_s^2 - \left( \sum_{s \in S} p_s u_s \right)^2 & f_*(\mathbf{v}, \mathbf{u}) &= \sup_z \left\{ -\frac{z^2}{4} : u_s^2 + u_s z = v_s, \quad \forall s \in S \right\} \\
g(\mathbf{p}, \mathbf{u}) &:= \sum_{s \in S} p_s u_s & g^*(\mathbf{v}, \mathbf{u}) &= \begin{cases} 0, & \text{if } v_s = u_s, \quad \forall s \in S \\ \infty, & \text{otherwise} \end{cases} \\
\tau_{i1}(\mathbf{p}) &:= \max_s \{-p_s\} & \tau_{i1}^*(\mathbf{v}) &= \begin{cases} 0, & \text{if } v_s \leq 0, \quad \forall s \in S \text{ and } \sum_{s \in S} v_s \geq -1 \\ \infty, & \text{otherwise} \end{cases} \\
\tau_{i2}(\mathbf{p}) &:= \sum_{s \in S} p_s - 1 & \tau_{i2}^*(\mathbf{v}) &= \begin{cases} 1, & \text{if } v_s = 1, \quad \forall s \in S \\ \infty, & \text{otherwise} \end{cases} \\
\tau_{i3}(\mathbf{p}) &:= 1 - \sum_{s \in S} p_s & \tau_{i3}^*(\mathbf{v}) &= \begin{cases} -1, & \text{if } v_s = -1, \quad \forall s \in S \\ \infty, & \text{otherwise} \end{cases} \\
\tau_{i4}(\mathbf{p}) &:= \sum_{s \in S} \frac{(p_s - \hat{p}_s)^2}{\hat{p}_s} - \rho & \tau_{i4}^*(\mathbf{v}) &= \rho + \sum_{s \in S} \hat{p}_s \left( \frac{1}{4} v_s^2 + v_s \right).
\end{aligned}$$

Plugging in these formulas in (7) yields  $F(\alpha) =$

$$\min \quad w \quad \text{s.t.} \quad t_{02} - t_{03} + \rho t_{04} + \sum_{s \in S} \hat{p}_s \left( \frac{(v_{04})_s^2}{4t_{04}} + (v_{04})_s \right) + \frac{z^2}{4} \leq w \quad (14)$$

$$\begin{aligned} (v_{01})_s &\leq 0 \quad (v_{11})_s \leq 0, \quad \forall s \in S \\ \sum_{s \in S} (v_{01})_s &\geq -t_{01} \quad \sum_{s \in S} (v_{11})_s \geq -t_{11} \\ u_s^2 + u_s z &= u_s \alpha + (v_{01})_s + t_{02} - t_{03} + (v_{04})_s, \quad \forall s \in S \\ t_{12} - t_{13} + \rho t_{14} + \sum_{s \in S} \hat{p}_s \left( \frac{(v_{14})_s^2}{4t_{14}} + (v_{14})_s \right) &< 0 \\ u_s + (v_{11})_s + t_{12} - t_{13} + (v_{14})_s &= 0, \quad \forall s \in S \\ u_s &= \sum_{i \in I} r_{is} x_i, \quad \forall s \in S \\ \sum_{i \in I} x_i &= C \\ \mathbf{t} \in \mathbb{R}_+^{2 \times 4}, \mathbf{u} \in \mathbb{R}^{|S|}, \mathbf{v} \in \mathbb{R}^{2 \times 4}, w \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^{|I|}, z \in \mathbb{R}. \end{aligned} \quad (15)$$

This problem is not convex because of the product  $u_s z$  in (15). Similar to Yanıkoğlu et al. (2013, Theorem 1) the problem can be made convex by replacing  $t_{02} - t_{03}$  in (14) with  $u_s^2 + u_s z - u_s \alpha - (v_{01})_s - (v_{04})_s$  and omitting (15) from the problem. Constraint (14) then becomes:

$$\left( u_s + \frac{z}{2} \right)^2 - u_s \alpha - (v_{01})_s - (v_{04})_s + \rho t_{04} + \sum_{s' \in S} \hat{p}_{s'} \left( \frac{(v_{04})_{s'}^2}{4t_{04}} + (v_{04})_{s'} \right) \leq w, \quad \forall s \in S, \quad (16)$$

which is jointly convex in all variables. In order to improve the tractability and accuracy of computing  $F(\alpha)$ , we cast it as a conic quadratic problem. The only complicating terms are  $(v_{04})_s^2/(4t_{04})$ , which can be reformulated using a standard trick. Constraint (16) is satisfied if and only if there exists auxiliary variables  $y_s$  such that the following inequalities are satisfied:

$$\begin{aligned} \left( u_s + \frac{z}{2} \right)^2 - u_s \alpha - (v_{01})_s - (v_{04})_s + \rho t_{04} + 2 \sum_{s' \in S} \hat{p}_{s'} (y_{s'} + (v_{04})_{s'}) &\leq w, \quad \forall s \in S \\ \left\| \begin{pmatrix} (2v_{04})_s \\ y_s - 4t_{04} \end{pmatrix} \right\|_2 &\leq y_s + 4t_{04}, \quad \forall s \in S. \end{aligned}$$

The problem (R-I) can now be solved by determining the root of  $F$ .

We perform a numerical analysis on 10 items and a generated data set of 50 scenarios. In order to incorporate correlations, we first construct a covariance matrix  $\mathbf{A}\mathbf{A}^\top$ , where  $\mathbf{A}$  is a  $10 \times 10$  matrix whose entries are uniformly and independently distributed on  $[-0.5, 0.5]$ . Then, to reflect the idea that a higher risk gives a higher expected return, a vector of expected returns  $\boldsymbol{\mu}$  is constructed with a linear mapping on the variances of the items. The mapping is constructed such that the item with the smallest variance gets an expected return of 0.01, and the item with the largest variance gets an expected return of 0.20. Finally, the scenarios are drawn, each from a multivariate normal distribution with the constructed mean  $\boldsymbol{\mu}$  and

covariance  $\mathbf{A}\mathbf{A}^\top$ . We solve the model for  $\rho = 1$ ,  $\hat{p}_s = 0.02$  for all  $s$ , and  $C = 100$  to obtain a robust solution using YALMIP (Löfberg, 2012) and MOSEK (Mosek ApS, Denmark) with their default settings. For this value of  $\rho$ , the probabilities can vary, on average, between 0 and 0.04. Additionally, we solve the same problem for  $\rho = 0$ , i.e., when  $p_s = \hat{p}_s$ , to obtain a nonrobust solution. We use bisection search on the interval determined by (8) and (9) and stopped when the interval width was less than  $10^{-10}$ . One step in bisection search takes around 2 seconds, of which around 7% is spent by MOSEK.

The convergence of the bisection method turns out to be adequate. Let  $\mathbf{x}_i$  denote the solution in iteration  $i$ , and let  $\mathbf{x}^*$  denote the final solution. The initial search interval is  $[0.70, 20.09]$ . The solution  $\mathbf{x}_1$ , obtained from solving  $F((0.70+20.09)/2)$ , is far from optimal:  $\|\mathbf{x}_1 - \mathbf{x}^*\|_\infty \approx 2.2$ . In each three or four iterations,  $\mathbf{x}_i$  gains one extra digit of accuracy. After 22 iterations, the accuracy has improved to  $\|\mathbf{x}_{22} - \mathbf{x}^*\|_\infty \approx 4.1 \cdot 10^{-7}$ . The algorithm terminates after 37 iterations, with no apparent improvement following the 22<sup>nd</sup> iteration. Since  $\|\mathbf{x}^*\|_2 \approx 44.9$ , the error after 22 iterations is relatively small.

When  $p_s = \hat{p}_s$  for all  $s$ , the mean-variance ratio of the nominal solution (which is 6.34) is indeed lower than that of the robust solution (which is 6.45). For both solutions we determined the worst case  $\hat{\mathbf{p}}$  and the corresponding objective value. The objective of the robust solution (which is 18.62) is slightly better than that of the nominal solution (which is 18.98). This shows that uncertainty may cause a factor three deterioration of the objective value. Relative to this large difference, the difference between the two solutions is small. So, the nominal solution performs quite well for this example. For the worst case probabilities for the robust solution, we have computed the optimal portfolio as if these solutions were known beforehand (perfect hindsight solution). The objective value equals that of the robust solution. So, even though the robust objective could deteriorate substantially, there is no other solution that performs better.

### 5.3 Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a tool to estimate the efficiency of different decision making units (DMUs), based on their inputs and outputs. DEA was originally introduced for not-for-profit companies, e.g., schools where inputs could be number of teacher hours and number of students per class, and outputs could be arithmetic scores and psychological tests of student attitudes, e.g., toward the community (Charnes et al., 1978). The applicability of DEA is not limited to nonprofit organizations. A reference list of more than 4,000 publications on DEA is given by Emrouznejad et al. (2008).

Let  $n_i$  and  $n_o$  denote the number of in- and outputs, respectively. The efficiency of a DMU is defined as the largest fraction of weighted outputs divided by weighted inputs, given that the efficiency of the other DMUs is at most 1:

$$(DEA) \quad \max_{\mathbf{u} \in \mathbb{R}_+^{n_o}, \mathbf{v} \in \mathbb{R}_+^{n_i}} \frac{\mathbf{u}^\top \mathbf{y}_0}{\mathbf{v}^\top \mathbf{x}_0} \quad \text{s.t.} \quad \frac{\mathbf{u}^\top \mathbf{y}_i}{\mathbf{v}^\top \mathbf{x}_i} \leq 1, \quad \forall i \in I,$$

where  $\mathbf{x}_i$  and  $\mathbf{y}_i$  are the vectors of inputs and outputs of DMU  $i$ , and  $\mathbf{u}$  and  $\mathbf{v}$  are the non-negative weights.

The inputs and outputs are model parameters that have to be acquired from each DMU and are affected by measurement errors. Especially when a single DMU represents a group

of smaller business units or is a pool of all activities in a certain region, and the inputs and outputs are aggregated, errors become practically inevitable. There have been many attempts to incorporate uncertainty in DEA. For an overview, e.g., see (Shokouhi et al., 2014). Since our focus is on RO, we only discuss the three papers that are relevant. The first only considers uncertain outputs (Sadjadi and Omrani, 2008). The second considers jointly uncertain inputs and outputs (Shokouhi et al., 2010). Unfortunately, the robust counterpart in the latter is constructed in an ad-hoc manner that results in a nonconvex formulation, for which it is not clear whether globally optimal solutions were found. The third considers either uncertain inputs or uncertain outputs (Wang and Wei, 2010). In the last two papers, a simulation study is performed to quantify the improvement offered by the robust solution. For each randomly drawn set of inputs and outputs, they compute the relative efficiencies with the  $\mathbf{u}$  and  $\mathbf{v}$  obtained from the robust solution. However, in our view, when the inputs and outputs are fully known, the relative efficiencies can only be computed by optimizing (DEA) for those known inputs and outputs. Our results are therefore different in two ways. First, we consider both uncertain inputs and uncertain outputs and solve the correct problem. Second, we perform a valid simulation study to verify whether the robust solution is better than the nominal solution.

In this section, we take the data from Shokouhi et al. (2010). In this data set there are five DMUs, two inputs and two outputs. The in- and outputs are uncertain, but known to reside in given intervals, given in Table 4.

In order to get in the minimization framework, the objective of (DEA) is replaced with its reciprocal. The optimal solution of the robust counterpart of (DEA) then corresponds to the reciprocal of the root of:

$$F(\alpha) = \min_{\mathbf{u} \in \mathbb{R}_+^{n_o}, \mathbf{v} \in \mathbb{R}_+^{n_i}, w \in \mathbb{R}} w \quad \text{s.t.} \quad \mathbf{v}^\top \mathbf{x}_0 - \alpha \mathbf{u}^\top \mathbf{y}_0 \leq w, \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathcal{U},$$

$$\mathbf{u}^\top \mathbf{y}_i \leq \mathbf{v}^\top \mathbf{x}_i, \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathcal{U}, \quad \forall i \in I.$$

Following Shokouhi et al. (2010), we take the Bertsimas and Sim uncertainty region:

$$\mathcal{U} = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{(|I|+1) \times n_i} \times \mathbb{R}^{(|I|+1) \times n_o} : x_{ij} = \bar{x}_{ij} + \zeta_{ij}^x \Delta \hat{x}_{ij}, \quad y_{ij} = \bar{y}_{ij} + \zeta_{ij}^y \Delta \hat{y}_{ij}, \\ \|\text{vec}(\boldsymbol{\zeta}^x, \boldsymbol{\zeta}^y)\|_\infty \leq 1, \quad \|\text{vec}(\boldsymbol{\zeta}^x, \boldsymbol{\zeta}^y)\|_1 \leq \Gamma\},$$

where  $\bar{x}_{ij}$  and  $\bar{y}_{ij}$  are the midpoints,  $\Delta \hat{x}_{ij}$  and  $\Delta \hat{y}_{ij}$  are the half-widths of the uncertainty intervals, and the  $\text{vec}$  operator stacks the columns of the matrix arguments into a single vector. For robust LP, this set has the property that when  $\Gamma$  is integer, it controls the number of uncertain elements that can deviate from their nominal values (Bertsimas and Sim, 2004). This property also holds for a robust LFP, since  $F(\alpha)$  is a robust LP.

The optimal weights  $\mathbf{u}$  and  $\mathbf{v}$  depend on the actual inputs and outputs. One may therefore be inclined to use Adjustable Robust Optimization (ARO) when  $\Gamma$  is larger than the dimensions of  $\mathbf{x}_i$  and  $\mathbf{y}_i$  added. Consequently,  $\mathbf{u}$  and  $\mathbf{v}$  are replaced by functions of the uncertain parameters. Unfortunately, even in the simple case of affine decision rules, this is often intractable. In the constraints of  $F(\alpha)$ ,  $\mathbf{u}$  and  $\mathbf{v}$  are multiplied with uncertain parameters, which yields constraints that are quadratic in the uncertain parameters. These can currently only be solved efficiently for ellipsoidal uncertainty sets.



Table 4: Data set for the DEA example of Section 5.3.

DMU <sub><i>i</i></sub>	Input 1	Input 2	Output 1	Output 2
1	[14,15]	[0.06, 0.09]	[157, 161]	[28, 40]
2	[4,12]	[0.16, 0.35]	[157, 198]	[21, 29]
3	[10,17]	[0.10, 0.70]	[143, 159]	[28, 35]
4	[12,15]	[0.21, 0.48]	[138, 144]	[21, 22]
5	[19,22]	[0.12, 0.19]	[158, 181]	[21, 25]

We use bisection search on the interval determined by (8) and (9) to determine the root of  $F(\alpha)$ , and stop when the interval width is less than  $10^{-4}$  (which turns out to be accurate enough for ranking the DMUs).  $F(\alpha)$  is computed using YALMIP and MOSEK with their default settings, and takes a few tenths of a second on a normal desktop computer, where MOSEK accounts for approximately 10% of that time. The time it takes to compute  $F(\alpha)$  turns out to be approximately constant, so independent of the remaining width of the interval and independent of the size of the uncertainty region  $\Gamma$ . The root of  $F(\alpha)$  is determined in a few seconds.

We computed the robust efficiencies of the DMUs for  $\Gamma$  ranging between 0 and 4 in steps of 0.1, since each constraint has at most four uncertain parameters. For  $\Gamma \leq 0.2$ , the list of DMUs ranked from most to least efficient, is 1, 2, 3, 5, 4. For  $\Gamma \geq 0.3$ , DMUs 3 and 5 switch positions. Hence, DMU 5 is more efficient than DMU 3 when  $\Gamma \geq 0.3$ . We have tried to verify this claim by running 100 simulations, where in each simulation we uniformly drew inputs and outputs from the uncertainty region, solved (DEA) for each set of inputs and outputs, and ranked the DMUs based on efficiency. In 76 out of 100 simulations, DMU 3 was more efficient than DMU 5. This result advocates against the use of RO in DEA, since it shows that for  $\Gamma = 0$  (i.e., the nonrobust solution), the ranking is better than for  $\Gamma \geq 1$ . We have also performed the simulation with more extreme data, by drawing the inputs and outputs only from the endpoints of their uncertainty intervals. This yielded similar results. Other experiments, where we used an ellipsoidal uncertainty region instead of the Bertsimas and Sim uncertainty region, or where we used the nominal objectives (based on  $\bar{\mathbf{x}}$  and  $\bar{\mathbf{y}}$ ) but kept the uncertain constraints, also yielded similar results.

## 6 Conclusions

We have shown how RO can be applied to FPs as a method to deal with uncertain data. The method has been tested on three problems. In all three examples, we observe that the nominal solution, which is obtained by solving the deterministic problem, is severely affected by uncertainty. Surprisingly, this also holds for the robust solution, and in none of the examples the robust solution offers a significant improvement; even when comparing worst case performance.

The first question that arises, is why the nominal solution performs so well. We try to answer this question for the mean-variance optimization problem, and note that the explanation for the multi-item newsvendor problem is similar. For a given solution, the worst

case for the mean is when the probability vector is a unit vector, that assigns unit weight to the scenario with lowest return. For the variance, the worst case is when the scenarios with the lowest and highest return each occur with probability 0.5. For a robust solution w.r.t. the mean value, the scenario with lowest return should be optimized, whereas for the variance, the returns in the scenarios with lowest and highest return should be close to each other. The nominal solution simultaneously maximizes the expected value and minimizes the variance. While not identical to the robust objective, it contains some aspects of it. For example, the mean is a weighted sum that contains the return for the scenario with lowest return.

The second question that arises, is why there is a realization of the uncertain parameters in Sections 5.1 and 5.2, for which no solution can outperform the robust solution; even if the former is optimized as if the realization of the uncertain parameters are known beforehand. This turns out to be due to Sion's minimax theorem (Sion, 1958). The assumptions (a)-(g) ensure that  $f(\mathbf{a}_0, \mathbf{x})/g(\mathbf{a}_0, \mathbf{x})$  is quasi-convex in  $\mathbf{x}$  (for fixed  $\mathbf{a}_0$ ), quasi-concave in  $\mathbf{a}_0$  (for fixed  $\mathbf{x}$ ), and continuous, that the uncertainty set is compact and convex, and that the feasible set for  $\mathbf{x}$ , say  $X$ , is convex. Therefore, by Sion's minimax theorem,  $\max_{\mathbf{a}_0 \in \mathcal{U}} \min_{\mathbf{x} \in X} f(\mathbf{a}_0, \mathbf{x})/g(\mathbf{a}_0, \mathbf{x}) = \min_{\mathbf{x} \in X} \max_{\mathbf{a}_0 \in \mathcal{U}} f(\mathbf{a}_0, \mathbf{x})/g(\mathbf{a}_0, \mathbf{x})$ . This no longer holds when there is uncertainty in the constraints, since the feasible region  $X$  changes when the values for the uncertain parameters are known.

So, the robust solution is good in the sense that it cannot be improved in the worst case, even if the values of the uncertain parameters are known beforehand. On the other hand, the nominal solution performs well, at least in the examples studied. It shall be interesting to see the difference in real-life examples, especially with uncertainty in the constraints.

## Acknowledgments

We thank D. den Hertog (Tilburg University) for many useful ideas and comments, D. Iancu (Stanford University) for showing the formulation (RP-FP) used in Theorem 2.1, and the referee and editor for their valuable comments.

## References

- A. I. Barros, J. B. G. Frenk, S. Schaible, and S. Zhang. A new algorithm for generalized fractional programs. *Mathematical Programming*, [72\(2\):147–175](#), 1996.
- A. Beck and A. Ben-Tal. Duality in robust optimization: primal worst equals dual best. *Operations Research Letters*, [37\(1\):1–6](#), 2009.
- A. Ben-Tal and D. den Hertog. Hidden conic quadratic representation of some nonconvex quadratic optimization problems. *Mathematical Programming*, [143\(1-2\):1–29](#), 2014.
- A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

- A. Ben-Tal, D. den Hertog, A. M. B. de Waegenare, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, [59\(2\):341–357](#), 2013.
- A. Ben-Tal, D. den Hertog, and J.-Ph. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, [149\(1–2\):265–299](#), 2015.
- D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, [52\(1\):35–53](#), 2004.
- D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, [53\(3\):464–501](#), 2011a.
- D. Bertsimas, D. A. Iancu, and P. A. Parrilo. A hierarchy of near-optimal policies for multi-stage adaptive optimization. *IEEE Transactions on Automatic Control*, [56\(12\):2809–2824](#), 2011b.
- A. Charnes and W. W. Cooper. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, [9\(3–4\):181–186](#), 1962.
- A. Charnes, W. W. Cooper, and E. Rhodes. Measuring the efficiency of decision making units. *European Journal of Operational Research*, [2\(6\):429–444](#), 1978.
- H. J. Chen, S. Schaible, and R. L. Sheu. Generic algorithm for generalized fractional programming. *Journal of Optimization Theory and Applications*, [141\(1\):93–105](#), 2009.
- J.-P. Crouzeix and J. A. Ferland. Algorithms for generalized fractional programming. *Mathematical Programming*, [52\(1–3\):191–207](#), 1991.
- J.-P. Crouzeix, J. A. Ferland, and S. Schaible. An algorithm for generalized fractional programs. *Journal of Optimization Theory and Applications*, [47\(1\):35–49](#), 1985.
- J.-P. Crouzeix, J. A. Ferland, and S. Schaible. A note on an algorithm for generalized fractional programs. *Journal of Optimization Theory and Applications*, [50\(1\):183–187](#), 1986.
- W. Dinkelbach. On nonlinear fractional programming. *Management Science*, [13\(7\):492–498](#), 1967.
- A. Emrouznejad, B. R. Parker, and G. Tavares. Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*, [42\(3\):151–157](#), 2008.
- B. L. Gorissen, A. Ben-Tal, J. P. C. Blanc, and D. den Hertog. Deriving robust and globalized robust solutions of uncertain linear programs with general convex uncertainty sets. *Operations Research*, [62\(3\):672–679](#), 2014.
- V. Jeyakumar, G. Y. Li, and S. Srisatkunarajah. Strong duality for robust minimax fractional programming problems. *European Journal of Operational Research*, [228\(2\):331–336](#), 2013.

- R. N. Kaul, S. Kaur, and V. Lyall. Duality in inexact fractional programming with set-inclusive constraints. *Journal of Optimization Theory and Applications*, [50\(2\):279–288](#), 1986.
- J. Lin and R. Sheu. Modified Dinkelbach-type algorithm for generalized fractional programs with infinitely many ratios. *Journal of Optimization Theory and Applications*, [126\(2\):323–343](#), 2005.
- J. Löfberg. Automatic robust convex programming. *Optimization Methods and Software*, [27\(1\):115–129](#), 2012.
- H. M. Markowitz. Portfolio selection. *Journal of Finance*, [7\(1\):77–91](#), 1952.
- S. J. Sadjadi and H. Omrani. Data envelopment analysis with uncertain data: an application for Iranian electricity distribution companies. *Energy Policy*, [36\(11\):4247–4254](#), 2008.
- S. Schaible. Parameter-free convex equivalent and dual programs of fractional programming problems. *Zeitschrift für Operations Research*, [18\(5\):187–196](#), 1974.
- S. Schaible. Fractional programming. II, On Dinkelbach’s algorithm. *Management Science*, [22\(8\):868–873](#), 1976.
- S. Schaible. Bibliography in fractional programming. *Zeitschrift für Operations Research*, [26\(1\):211–241](#), 1982.
- S. Schaible and T. Ibaraki. Fractional programming. *European Journal of Operational Research*, [12\(4\):325–338](#), 1983.
- A. H. Shokouhi, A. Hatami-Marbini, M. Tavana, and S. Saati. A robust optimization approach for imprecise data envelopment analysis. *Computers & Industrial Engineering*, [59\(3\):387–397](#), 2010.
- A. H. Shokouhi, H. Shahriari, P. Agrell, and A. Hatami-Marbini. Consistent and robust ranking in imprecise data envelopment analysis under perturbations of random subsets of data. *OR Spectrum*, [36\(1\):133–160](#), 2014.
- M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, [8\(1\):171–176](#), 1958.
- I. M. Stancu-Minasian. A seventh bibliography of fractional programming. *Advanced Modeling and Optimization*, [15\(2\):309–386](#), 2013.
- K. Wang and F. Wei. Robust data envelopment analysis based MCDM with the consideration of uncertain data. *Journal of Systems Engineering and Electronics*, [21\(6\):981–989](#), 2010.
- İ. Yanıkoğlu, D. den Hertog, and J. P. C. Kleijnen. Adjustable robust parameter design with unknown distributions. *CentER Discussion Paper*, [2013\(022\)](#), 2013.

## A The Importance of Convexity Conditions

We provide a short example to stress the importance of the convexity/concavity conditions on  $f$  and  $g$ . The second numerical example by Lin and Sheu (2005) is:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{a \in [0,1]} \frac{a^2 x_1 x_2 + x_1^{2a} + a x_3^3}{5(a-1)^2 x_1^4 + 2x_2^2 + 4a x_3} \quad \text{s.t.} \quad 0.5 \leq x_i \leq 5, \quad i = 1, 2, 3.$$

This problem does not satisfy the convexity/concavity conditions from Section 4. Lin and Sheu claim that  $\mathbf{x} = (0.5 \quad 1.5 \quad 0.5)$  and  $a = 0$  is optimal with a value of 0.21 (reported as  $-0.21$ ), but  $\mathbf{x} = (0.5 \quad 5 \quad 0.5)$  and  $a = 1$  is a better solution (maybe still not optimal) since the corresponding value is 0.06.

## B On the Result by Kaul et al. (1986)

This appendix shows a mistake in the paper by Kaul et al. (1986). Essentially, they formulate the dual of:

$$\min_{\alpha \in \mathbb{R}_+, \mathbf{x} \in \mathbb{R}_+^n} \alpha \quad \text{s.t.} \quad \frac{b_0 + \mathbf{b}^\top \mathbf{x}}{c_0 + \mathbf{c}^\top \mathbf{x}} \leq \alpha, \quad \forall (b_0, \mathbf{b}) \in \mathcal{U}_1 \times \mathcal{U}_2, \quad \forall (c_0, \mathbf{c}) \in \mathcal{U}_3 \times \mathcal{U}_4, \quad \mathbf{A}\mathbf{x} \leq \mathbf{d}.$$

Note that  $\mathbf{x}$  is non-negative. In their Lemma 2.1, they claim that the worst case  $(c_0, \mathbf{c})$  does not depend on  $\mathbf{x}$ , and is given by  $c_0^* = \min_{c_0 \in \mathcal{U}_3} \{c_0\}$  and  $\mathbf{c}^*$ , with components  $c_i^* = \min_{\mathbf{c} \in \mathcal{U}_4} \{c_i\}$ . This implicitly assumes that  $\mathbf{c}^*$  is a member of  $\mathcal{U}_4$ , which is not always true. The mistake becomes clear in their numerical example, where they use  $\mathbf{c}^* = [4; 2]$ , which is not in the uncertainty set. Consequently, the proposed approach gives the wrong dual problem and a suboptimal solution. Our results in Section 4.2 can provide the correct dual problem under milder conditions on the uncertainty sets.