CrossMark

Lot sizing with storage losses under demand uncertainty

Stefano Coniglio[1] · Arie M. C. A. Koster[2] ·
Nils Spiekermann[2]

Abstract We address a variant of the single item lot sizing problem affected by proportional storage (or inventory) losses and uncertainty in the product demand. The problem has applications in, among others, the energy sector, where storage losses (or storage deteriorations) are often unavoidable and, due to the need for planning ahead, the demands can be largely uncertain. We first propose a two-stage robust optimization approach with second-stage storage variables, showing how the arising robust problem can be solved as an instance of the deterministic one. We then consider a two-stage approach where not only the storage but also the production variables are determined in the second stage. After showing that, in the general case, solutions to this problem can suffer from acausality (or anticipativity), we introduce a flexible affine rule approach which, albeit restricting the solution set, allows for causal production plans. A hybrid robust-stochastic approach where the objective function is optimized in expectation, as opposed to in the worst-case, while retaining robust optimization guarantees of feasibility in the worst-case, is also discussed. We conclude with an application to

✉ Stefano Coniglio
s.coniglio@soton.ac.uk

Arie M. C. A. Koster
koster@math2.rwth-aachen.de

Nils Spiekermann
spiekermann@math2.rwth-aachen.de

[1] Department of Mathematical Sciences, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom

[2] Lehrstuhl II für Mathematik, RWTH Aachen University, Pontdriesch 10-12, 52062 Aachen, Germany

⌀ Springer

heat production, in the context of which we compare the different approaches via computational experiments on real-world data.

1 Introduction

Lot Sizing (LS) is a fundamental problem in a large part of modern production planning systems. In its basic version, given a demand for a single good over a finite time horizon, the problem calls for a production plan which minimizes storage, production, and set-up costs, while keeping production and storage levels within the prescribed lower and upper bounds.

In the paper, we focus on a variant of LS where the storage suffers from proportional losses and the product demands are subject to uncertainty.[1] This variant suits the case of many applications in the energy sector where a small portion of the energy that is stored is lost over time and the demands (of, e.g., heat, as in the application that we will consider) are often not known in advance. This is especially relevant when the decision maker has to commit to a production plan ahead of when it becomes operational.

To cope with uncertainty, we develop a robust optimization approach. After observing that a fully first-stage approach is not viable, we adopt an approach with first-stage set-up and production variables and second-stage storage variables. For it, we establish a reduction showing that this problem can be solved as an instance of its deterministic counterpart with suitably (re)defined demands and storage upper bounds. Then, we consider the case where the second-stage decisions comprise not just the storage but also the production variables, while still retaining first-stage set-up variables. After showing that solutions to this problem may be acausal (or anticipative), i.e., that they may contain production variables whose value, at time t, can only be determined by knowing the realization of the demand at time $t' > t$, we introduce an affine rule approach. This technique, although restricting the solution set, allows for causal production plans. We then illustrate a set of computational experiments in the context of an application to heat production.

The paper, which is an extended version of Coniglio et al. (2016), is organized as follows. In Sect. 2, we summarize some key results on the deterministic version of LS and present some relevant previous work on robust optimization. The version of LS with storage losses that we tackle in this paper is formally introduced in Sect. 3, together with its uncertainty aspect. In Sect. 4, we introduce the robust counterpart of the problem with second-stage inventory variables and present our reduction. The approach with second-stage production and storage variables is proposed in Sect. 5,

---

[1] The variant admits the classical version of LS as a special case.

while, in Sect. 6, we present a hybrid stochastic-robust extension. Computational results are reported and illustrated in Sect. 7. Concluding remarks are drawn in Sect. 8.

2 Previous work

We give a brief account of the most relevant works on different versions of LS, some also encompassing the uncertain case.

2.1 Deterministic LS

In the deterministic setting, LS has been extensively studied in many variants. Among the different versions, we mention that with *backlogging*, in which shortages in the inventory are allowed (or, said differently, unmet demand can be postponed to the future, at a cost), the *uncapacitated* one, in which production has no predefined bounds, and that with production *capacities*, either static (constant over time) or time-dependent, where lower and upper bounds on the production are present. The storage is usually assumed to have either a zero or a static lower bound—in the latter case, the problem can be often brought to the case of a zero lower bound with a reformulation. Storage upper bounds are often assumed to be static, although many algorithms can be easily adapted to the time-dependent case. For an extensive account on LS and many of its variants, we refer the reader to the monograph (Pochet and Wolsey 2006).

LS is known to be in $\mathcal{P}$ for the case with linear costs, complete conservation (i.e., no losses in the storage), zero or a static storage lower bounds, possibly nonzero and time-dependent storage upper bounds, and no production bounds, as shown in Atamtürk and Küçükyavuz (2008). A similar result holds for nonnegative and nondecreasing concave cost functions, complete conservation, production bounds that are constant over time, and unrestricted storage, as shown in Hellion et al. (2012). For a polynomial time algorithm for the case with storage losses and nondecreasing concave costs, but no storage or production bounds, see Hsu (2000). As to $\mathcal{NP}$-hard cases, a number of examples are provided in Florian et al. (1980). These include the case of linear as well as fixed production costs, no inventory costs, no storage bounds, and no lower production bounds, but time-dependent production upper bounds.

2.2 LS with uncertain demands

Classical approaches to handle uncertainties in LS are, historically, stochastic in nature, dating back as early as 1960 (Scarft 1960). The idea is to first assign a probability distribution to the uncertain demands and, then, to tackle the problem by looking for a solution of minimum expected cost, which is feasible with high probability.

Unfortunately, as pointed out in Liyanage and Shanthikumar (2005), even when the distribution is estimated within sufficient precision from historical data, such methods can yield solutions which, when implemented with the demand that realizes in practice, can be substantially more costly than those that were predicted with the stochastic approach (or infeasible, depending on how the problem is formulated). Moreover, and

regardless of the accuracy of the estimation, such stochastic techniques are, in many cases, intrinsically doomed to suffer from the curse of dimensionality (Bertsimas and Thiele 2006). This is because they usually require a computing time which is, at least, linear in the size of the probability space (assumed to be discrete) to which the realized demand belongs—such space is, typically, exponential in the length of the time horizon. For an alternative way of modeling stochastic demands, via so-called service-level constraints, see Bookbinder and Tan (1988), Tempelmeier and Herpers (2011), and Tempelmeier (2013). Also see Calafiore (2008) for a stochastic programming approach where feasibility is enforced in expectation.

A different option, often more affordable from a computational standpoint, is resorting to a *robust optimization* approach. The idea is of looking for a solution which is feasible for every realization of the uncertain demand belonging to a given *uncertainty set* and which, among all such feasible solutions, is of minimum cost in the *worst case*. *Two-stage robust optimization* is an extension of this approach in which the variables are partitioned in two sets. The *first-stage* (or *here-and-now*) variables are required to take a value independent of the realization of the uncertain parameters, while the *second-stage* (or *wait-and-see*) variables are allowed to vary as a function of the specific realization that takes or has taken place, depending on whether the problem has a dynamical aspect or not.

Two seminal papers applying a robust optimization approach to LS are Bertsimas and Thiele (2004) and Bertsimas and Thiele (2006). The authors address the uncapacitated version of LS with backlogging, static storage, and production costs, with demands subject to a so-called Γ-*robustness* model of uncertainty (Bertsimas and Sim 2003, 2004). When applied to LS, the idea of Γ-robustness is of assuming that the uncertain parameters, i.e., the product demand for LS, (i) belong to symmetric intervals and (ii) given an integer Γ, the total number of time steps in which the uncertain demand deviates from its nominal value to one of the extremes of the interval is bounded by Γ, in any constraint of the problem.

Among other results, the authors of Bertsimas and Thiele (2004, 2006) show that the Γ-robust counterpart of the variant of LS they consider (with backlogging) can be solved as an instance of the deterministic problem with modified demands, also for the case where production bounds are in place. Notice that, for this result to hold, bounds on the storage cannot be enforced. Unfortunately, such bounds are necessary in many applications, such as, e.g., those in the energy sector, where backlogging is not tolerable as the demand of energy (e.g., heat) must be typically satisfied when issued.

3 The problem: LS with storage losses and uncertain product demands

Consider a single product and a time horizon $T = \{1, \ldots, n\}$. For each time step $t \in T$, let $d_t \geq 0$ be the product demand, $q_t \geq 0$ the production variable, and $u_t \geq 0$ the storage variable corresponding to the value of the storage at the end of time step t. Let $z_t \in \{0, 1\}$ be an indicator variable equal to 1 if production is active at time $t \in T$, that is, if $q_t > 0$, and let the constant $u_0 \geq 0$ represent the initial storage value. Let also c_t^i, c_t^v, and c_t^c be, respectively, the inventory, production, and fixed set-up cost

(the latter is incurred when production is active) at time $t \in T$. We consider time-dependent bounds $\underline{U}_t \leq u_t \leq \overline{U}_t$ on the storage, as well as time-dependent bounds $\underline{Q}_t \leq q_t \leq \overline{Q}_t$ on production, when active.

3.1 The deterministic case

The (DETerministic) variant of LS with storage losses tackled in this paper can be cast as the following Mixed-Integer Linear Programming (MILP) problem:

$$(\text{LS-DET}) \quad \min \quad \eta \tag{1a}$$

$$\text{s.t.} \quad \eta \geq \sum_{t \in T} \left(c_t^v q_t + c_t^c z_t + c_t^i u_t \right) \tag{1b}$$

$$\alpha_t u_{t-1} + q_t = u_t + d_t \qquad\qquad \forall t \in T \tag{1c}$$

$$\underline{U}_t \leq u_t \leq \overline{U}_t \qquad\qquad \forall t \in T \tag{1d}$$

$$z_t \underline{Q}_t \leq q_t \leq z_t \overline{Q}_t \qquad\qquad \forall t \in T \tag{1e}$$

$$\eta \in \mathbb{R} \tag{1f}$$

$$z_t \in \{0, 1\} \qquad\qquad \forall t \in T \tag{1g}$$

$$u_t, q_t \geq 0 \qquad\qquad \forall t \in T. \tag{1h}$$

For notational consistency with the remainder of the paper, the problem is written in *epigraph* form, with the objective function moved into Constraint (1b). Constraints (1c) are balance constraints, which only differ from those in the standard LS problem for the (possibly time-dependent) *conservation factor* $\alpha_t \in (0, 1]$. They account for the fact that, at the beginning of each time period $t \in T$, a fraction $(1 - \alpha_t)$ of the previously stored product is lost, whereas, at the end of time period t, the remaining product is conserved. Constraints (1d) and (1e) impose lower and upper bounds on storage and production. Constraints (1f)–(1h) specify the nature of the decision variables.

3.2 The uncertain case

Let us assume an *uncertain* product demand d, belonging to an *uncertainty set D* —as better illustrated in Sect. 7, we will construct this set from a mixture of historical data and forecasts.

In the combinatorial optimization literature, two types of uncertainty sets are predominantly employed: *finite uncertainty sets* and *polyhedral uncertainty sets*. In the former case, D is assumed to be a finite set, albeit possibly extremely large.[2] In the latter, D is assumed to be a bounded polyhedron. The Γ-*robustness* approach used in Bertsimas and Sim (2004) falls into this category. Its uncertainty set, the Γ *uncertainty set*, corresponds to assuming, for each time period $t \in T$, that the uncertain demand d_t belongs to a symmetric interval $[\bar{d}_t - \hat{d}_t, \bar{d}_t + \hat{d}_t]$ centered around a nom-

---

[2] Some authors adopt the expression *discrete*, as opposed to finite. See, e.g., Büsing et al. (2011).

inal value $\bar{d}_t$ and that, in each constraint of the problem, the demand could deviate from $\bar{d}_t$ by $\pm \hat{d}_t$ for, at most, Γ time steps.

4 Robust LS with first-stage production and second-stage storage

As a first approach to the problem, we assume, as natural in a number of applications, that the storage variables u could be determined in a *wait-and-see* fashion as a function of the realization of the uncertain demand d, while all the other variables are required to a take a *here-and-now* value independently of the realization. This corresponds to adopting a two-stage setting with second-stage storage variables, as better explained in the following.[3] In terms of the production and set-up variables, this approach can be regarded as an *open-loop* approach (see Calafiore 2008), in which the value of the production variables at time t does not depend on the realization of d up to time t. Although we will consider more flexible approaches in the next section, this one is of interest since, as shown in this section, it allows for solving the corresponding robust problem very efficiently, as an instance of the deterministic problem, obtained after modifying a few of its givens.

The problem with second-stage u, which we report in full for notational convenience, reads:

$$(\text{LS-2RO}_u) \quad \min \quad \eta \tag{2a}$$

$$\text{s.t.} \quad \eta \geq \sum_{t \in T} \left(c_t^v q_t + c_t^c z_t + c_t^i u_t(d) \right) \qquad \forall d \in D \tag{2b}$$

$$\alpha_t u_{t-1}(d) + q_t = u_t(d) + d_t \qquad \forall t \in T, d \in D \tag{2c}$$

$$\underline{U}_t \leq u_t(d) \leq \overline{U}_t \qquad \forall t \in T, d \in D \tag{2d}$$

$$z_t \underline{Q}_t \leq q_t \leq z_t \overline{Q}_t \qquad \forall t \in T \tag{2e}$$

$$\eta \in \mathbb{R} \tag{2f}$$

$$z_t \in \{0, 1\} \qquad \forall t \in T \tag{2g}$$

$$q_t \geq 0 \qquad \forall t \in T \tag{2h}$$

$$u_t(d) \geq 0 \qquad \forall t \in T, d \in D. \tag{2i}$$

It only differs from LS-DET in that, here, we have a vector $u(d)$ for each realization $d \in D$.

As for the classical version of LS without storage losses, see, e.g., Bertsimas and Thiele (2006), Gaussian elimination shows that, in any feasible solution, the storage variables u are uniquely defined as a function of d and q:

---

[3] Note that a completely first-stage approach, calling for a solution (η, z, q, u) which is feasible for *all* realizations of the uncertain demand $d \in D$, which is obtainable by solving the formulation for LS-DET after imposing Constraints (1c) for all realizations $d \in D$, yields an infeasible problem as soon as $|D| > 1$, even with a single time period.

$$u_t(d) = \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{it} (q_i - d_i) \quad \forall t \in T, \tag{3}$$

with:

$$\alpha_{mn} := \prod_{k=m+1}^{n} \alpha_k.$$

A compact formulation for this problem is thus obtained by substituting, in the previous formulation, the right-hand side of Eq. (3) for $u(d)$.

With the next theorem, we show how LS-2RO$_u$ can be reduced to an instance of LS-DET with identical givens except for a different demand d' and different storage upper bounds $\overline{U}'$, defined as follows:

$$d'_t := \max_{d \in D} \left\{ d_t - \sum_{i=1}^{t-1} \alpha_{it} \left(d'_i - d_i \right) \right\} \qquad \forall t \in T \tag{4a}$$

$$\Delta_t := \max_{d \in D} \left\{ \sum_{i=1}^{t} \alpha_{it} \left(d'_i - d_i \right) \right\} \qquad \forall t \in T. \tag{4b}$$

$$\overline{U}'_t := \overline{U}_t - \Delta_t \qquad \forall t \in T \tag{4c}$$

The idea is that, with this definition, the newly introduced demand d'_t specifies a lower bound on the product which has to be available at time $t \in T$ to ensure that every demand $d \in D$ can be met. The value Δ_t reduces the storage upper bounds of the transformed problem to prevent that, if a large production is realized but a large deficit in demand occurs (an event which would result in an overflow of storage), the storage upper bound $\overline{U}_t$ is not exceeded. Formally:

Theorem 1 *Assuming d'_t and Δ_t are well defined for each $t \in T$, LS-2RO$_u$ can be reduced, with respect to the first-stage variables (η, z, q), to an instance of LS-DET with $d = d'$ and $\overline{U} = \overline{U}'$, as defined in Eq. (4).*

Proof First note that, due to the assumptions on D, the values for d'_t and $\overline{U}'_t$ can be computed iteratively from $t = 1$ to $t = |T|$. Let $u_t(d)$ be defined as in Eq. (3) and let:

$$\eta := \sum_{t \in T} \left(c_t^v q_t + c_t^c z_t \right) + \max_{d \in D} \left\{ \sum_{t \in T} c_t^i u_t(d) \right\}. \tag{5a}$$

For a given (z, q), adopting $d = d'$ and $\overline{U} = \overline{U}'$, we now show that $(\eta, z, q, u(d))$, with $u(d)$ and η defined as above, is feasible for LS-2RO$_u$ if and only if it is feasible for LS-DET with demand d' and storage upper bounds $\overline{U}'$. Following the derivations reported in "Appendix A", we deduce:

$$d'_t \geq 0 \qquad \forall t \in T \tag{6a}$$

$$u_t = \min_{d \in D} \{u_t(d)\} \qquad\qquad \forall t \in T \qquad (6b)$$

$$u_t + \Delta_t = \max_{d \in D} \{u_t(d)\} \qquad\qquad \forall t \in T \qquad (6c)$$

$$\eta = \sum_{t \in T} \left(c_t^v q_t + c_t^c z_t + c_t^i u_t \right) + \underbrace{\max_{d \in D} \left\{ \sum_{t \in T} c_t^i \sum_{i=1}^{t} \alpha_{it} \left(d_i' - d_i \right) \right\}}_{\text{const}}. \qquad (6d)$$

What remains to be shown is that Constraints (2d) are satisfied by $u(d)$ if and only if Constraints (1d) are satisfied by u. This is done by observing that, for all $t \in T$, the following holds true:

$$\underline{U}_t \le u_t(d) \ \ \forall d \in D \Leftrightarrow \underline{U}_t \le \min_{d \in D}\{u_t(d)\} \Leftrightarrow \underline{U}_t \le u_t$$
$$u_t(d) \le \overline{U}_t \ \ \forall d \in D \Leftrightarrow \max_{d \in D}\{u_t(d)\} \le \overline{U}_t \Leftrightarrow u_t + \Delta_t \le \overline{U}_t \Leftrightarrow u_t \le \overline{U}_t'.$$

Since the objective functions moved to Constraints (2b) and (1b) are equal up to a constant additive term, as highlighted in Eq. (6d), we deduce that (η, z, q) is optimal for LS-2RO$_u$ if and only if it is optimal for LS-DET.   □

We remark that the reduction in Theorem 1 allows us to cast LS-2RO$_u$ as an instance of LS-DET with static lower bounds on u. This can be quite relevant as, to the best of our knowledge, most of the algorithms for LS-DET, such as the one by Atamtürk and Küçükyavuz (2008), require a storage with either zero or static lower bounds.

From a computational complexity perspective, Theorem 1 allows us to establish a direct relationship between LS-2RO$_u$ and LS-DET:

Corollary 1 *Given an uncertainty set D over which a linear function can be optimized in polynomial time, the reduction in Theorem 1 is a polynomial reduction. As such, LS-ROB2-u is in $\mathcal{P}$ (respectively, $\mathcal{NP}$-hard) if and only if the corresponding version of LS-DET is in $\mathcal{P}$ (respectively, $\mathcal{NP}$-hard).*

As a consequence of Corollary 1, LS-ROB2-u is in $\mathcal{P}$ for all the polynomially solvable cases of LS-DET that we reported in Sect. 2, provided that their algorithm allows for the introduction of time-dependent upper bounds $\overline{U}$ on u. This is, for instance, the case of the problem studied in Hellion et al. (2012). Note that the result still holds if we introduce additional constraints on z and q or assumptions on the givens (except for d and $\overline{U}$). It is also valid for $\overline{U}_t = \infty$ and for not necessarily nonnegative demands d.

The assumptions on the uncertainty set in Corollary 1 subsume the cases of many robustness models, including finite uncertainty sets, polyhedral uncertainty sets (such as the Γ-robustness one), and ellipsoidal uncertainty sets, such as those employed in Ben-Tal et al. (2009). The advantages in terms of computing times of this robust optimization approach are better illustrated in Sect. 7.

5 Robust LS with second-stage production and storage variables

In a number of applications, it is reasonable to assume that not only the storage variables but also the production variables could be adjusted (possibly, in real time) in a second-stage fashion, after the uncertain demand reveals.

Still retaining first-stage set-up variables, we can cast the problem with explicit second-stage variables $u(d)$ and $q(d)$, for all $d \in D$, as this (infinite, for a continuous D) MILP:

$$(\text{LS-2RO}_{q,u}) \quad \min \quad \eta \tag{7a}$$

$$\text{s.t.} \quad \eta \geq \sum_{t \in T} \left(c_t^v q_t(d) + c_t^c z_t + c_t^i u_t(d) \right) \qquad \forall d \in D \tag{7b}$$

$$\alpha_t u_{t-1}(d) + q_t(d) = u_t(d) + d_t \qquad \forall t \in T, d \in D \tag{7c}$$

$$\underline{U}_t \leq u_t(d) \leq \overline{U}_t \qquad \forall t \in T, d \in D \tag{7d}$$

$$z_t \underline{Q}_t \leq q_t(d) \leq z_t \overline{Q}_t \qquad \forall t \in T, d \in D \tag{7e}$$

$$\eta \in \mathbb{R} \tag{7f}$$

$$z_t \in \{0, 1\} \qquad \forall t \in T \tag{7g}$$

$$u_t(d), q_t(d) \geq 0 \qquad \forall t \in T, d \in D. \tag{7h}$$

LS-2RO$_{q,u}$ suffers from an important drawback: *acausality* (or *anticipativity*). Indeed:

Proposition 1 *In the general case, optimal solutions to LS-2RO$_{q,u}$ can be acausal, with $q_t(d)$ depending, for some $t \in T$, from $d_{t'}$ for $t' > t$.*

Proof Let us consider a finite uncertainty set case with $D = \{d^1, d^2\}$, where $d^1 = (1, 3, 1)$ and $d^2 = (1, 1, 3)$, with $u_0 = 0$, $c^i = (1, 1, 1)$, $\underline{U} = \underline{Q} = (0, 0, 0)$, $\overline{U} = (2, 2, 2)$, $\overline{Q} = (2, 2, 2)$, $c^v = (1, 1, 1)$, and $c^c = (0, 0, 0)$. Since the set-up costs are identically zero, we can assume w.l.o.g. $z_t = 1$ for all $t \in T$. It follows that LS-2RO$_{q,u}$ can be solved by solving two independent instances of LS-DET, one for $d = d^1$ and one for $d = d^2$:

– for $d^1 = (1, 3, 1)$, the unique optimal production vector is $q^1 = (2, 2, 1)$;
– for $d^2 = (1, 1, 3)$, the unique optimal production vector is $q^2 = (1, 2, 2)$.

Note that $q_1^1 \neq q_1^2$, even though $d_1^1 = d_1^2 = 1$. To be able to decide whether, at time $t = 1$, $q_1 := q_1^1 = 2$ or $q_1 := q_1^2 = 1$, we need to know d_2 (so to be able to determine whether $d = d^1$ or $d = d^2$). The (unique) optimal solution to the problem is, thus, acausal. □

5.1 Affine decision rules

To prevent acausality, we resort to *affine rules*, a technique which has been successfully adopted in a number of optimization and control problems, see, e.g., Ben-Tal et al. (2004), Calafiore (2008), and Bertsimas and Georghiou (2015).

While still allowing $q(d)$ to be a function of d, the idea is to restrict the feasible region of the problem by requiring $q_t(d)$ to be, for each $t \in T$, an affine function of the uncertain demand vector d. We impose:

$$q_t(d) := \sum_{j \in T: j \le t} \xi_{jt} d_j + \Xi_t, \tag{8}$$

which stipulates that $q_t(d)$ be composed of two parts: a first-stage part Ξ_t, which, as in LS-2RO$_u$, does not depend on the realization of d, and a second-stage part $\sum_{j \in T: j \le t} \xi_{jt} d_j$, which does, but only for components of d up to time t.

Notice that, with this approach, the coefficients ξ_{it} and Ξ_t must be determined as first-stage variables. As it is clear, this approach generalizes the case of LS-2RO$_u$ with a first-stage q, as the latter is obtained when requiring $\xi_{jt} = 0$ for all $j, t \in T$.

Overall, we obtain the following (infinite, for a nonfinite D) MILP with Affine Rules (AR):

(LS-2RO$_{q,u}$-AR)

$$\min \quad \eta \tag{9a}$$

$$\text{s.t. } \eta \ge \sum_{t \in T} \left(c_t^v q_t(d) + c_t^c z_t + c_t^i u_t(d) \right) \qquad \forall d \in D \tag{9b}$$

$$\alpha_t u_{t-1}(d) + q_t(d) = u_t(d) + d_t \qquad \forall t \in T, \ d \in D \tag{9c}$$

$$\underline{U}_t \le u_t(d) \le \overline{U}_t \qquad \forall t \in T, \ d \in D \tag{9d}$$

$$z_t \underline{Q}_t \le q_t(d) \le z_t \overline{Q}_t \qquad \forall t \in T, \ d \in D \tag{9e}$$

$$q_t(d) = \sum_{j \in T: j \le t} \xi_{jt} d_j + \Xi_t \qquad \forall t \in T, \ d \in D \tag{9f}$$

$$\eta \in \mathbb{R} \tag{9g}$$

$$z_t \in \{0, 1\} \qquad \forall t \in T \tag{9h}$$

$$\xi_{jt}, \Xi_t \in \mathbb{R} \qquad \forall t, j \in T \tag{9i}$$

$$q_t(d), u_t(d) \ge 0 \qquad \forall t \in T, \ d \in D. \tag{9j}$$

To arrive at a finite MILP, we can first derive a semi-infinite formulation by substituting the right-hand sides of Eq. (3) and Constraints (9f) for variables $u(d)$ and $q(d)$. For finite scenario uncertainty sets, the problem thus obtained can be rewritten as a finite MILP by duplicating the constraints involving the demand vector d exactly $|D|$ times. For the case of Γ-robustness, we can obtain a reformulation of polynomial size via the dualization procedure introduced by Bertsimas and Sim (2004). We report the formulation thus obtained in "Appendix B". Note that, since the feasible region of the problem belongs to a higher-dimensional space than that of LS-2RO$_u$, we cannot hope for a result similar to that of Theorem 1.

5.2 A regularization approach

As mentioned in Caramanis (2006), affine rules add an affine regression subproblem to the robust counterpart of a deterministic problem.

Let $q_t^*(d)$ be the (assumed unique, for simplicity) optimal production value at time t for LS-2RO$_{q,u}$ (in which the second-stage q is not restricted by affine rules and, thus, it can be anticipative), expressed as a function of the realization $d \in D$. If LS-2RO$_{q,u}$-AR admits a solution with coefficients ξ_{jt} and Ξ_t satisfying:

$$\underbrace{\sum_{j \in T: j \leq t} \xi_{jt} d_j + \Xi_t}_{q_t(d)} = q_t^*(d) \quad \forall t \in T, \forall d \in D, \tag{10}$$

$q_t(d)$ can then take the same value that it takes in LS-2RO$_{q,u}$ and, thus, the affine rules pose no restriction to the problem. If Eq. (10) cannot be satisfied for all $t \in T$ and $d \in D$, LS-2RO$_{q,u}$-AR calls for coefficients which minimize the objective function loss that is incurred when adopting $q_t(d) = \sum_{j \in T} \xi_{jt} d_j + \Xi_t$ instead of $q_t^*(d)$.

As typical in many applications, D is usually an estimate of the "hidden" uncertainty set, let us call it $\tilde{D}$, to which the realizations of the uncertain demand belong. In that case, one should look for coefficients which, rather than just minimizing the worst-case objective function value on demands $d \in D$, are also likely to achieve a good objective function value on realizations $d \in \tilde{D} \setminus D$, thus enjoying good *generalization properties*. In a number of machine learning works, see Bishop (2006), it is observed that, both in theory and in practice, a better generalization is obtained by introducing a *regularization* term which limits the magnitude of the coefficients of the function that is being learned—those in Eq. (8) in our case.

For these reasons, we also consider, in the computational experiments, an alternative approach encompassing a restriction on the norm of the vector of coefficients $\xi_t = (\xi_{1t}, \dots, \xi_{jt}, \dots, \xi_{tt})$. More precisely, we consider, together with the original approach with unrestricted ξ_t, a version where the coefficients ξ_{jt} are restricted to the interval $[-1, 1]$. This corresponds to introducing, for all $t \in T$, the ∞-norm constraint $||\xi_t||_\infty \leq 1$ on the vector of coefficients ξ_t adopted in each affine rule.

6 Hybrid robust-stochastic approach

A robust approach is, arguably, bound not to be viable in a free market where the decision maker is facing a number of not risk-averse competitors. While interested in solutions which are feasible even in the worst-case (which is particularly relevant in the energy sector, as in the application considered in the next section), minimizing the objective function in expectation, as opposed to in the worst-case, is likely to offer solutions which, by being less risk-averse, can be more profitable in practice.

As customary in many results in robust optimization, such as, e.g., those in Bertsimas and Sim (2004) on Γ-robustness, let us assume that $d \in D$ is a symmetrically distributed random variable with $E[d] = \bar{d}$. With this, the following holds:

Proposition 2 *Let $D \subseteq \mathbb{R}^n$ be a set centrally symmetric about $\bar{d}$, $d \in D$ a random variable with density function $p : D \to [0, 1]$ symmetric about $\bar{d}$, and $g : D \to \mathbb{R}$ an affine function. Then, $E[g(d)] = \int_{x \in D} p(x)g(x)dx = g(\bar{d})$.*

Proof Let $D' := \{d - \bar{d} : d \in D\}$. Since D' is centrally symmetric about 0, $D' = -D'$ holds. Let $p'(x) := p(x + \bar{d})$ for $x' \in D'$. Then, p' is symmetric about 0. Let f be a linear function. We have:

$$\int_{x' \in D'} p'(x')f(x')dx' = -\int_{x' \in D'} -p'(x')f(x')dx'$$

$$= -\int_{x' \in D'} p'(x')f(-x')dx' \text{ (by the linearity of } f)$$

$$= -\int_{(-x') \in (-D')} p'(-x')f(x')dx' \text{ (subs. } -x' \text{ for } x')$$

$$= -\int_{(-x') \in (-D')} p'(x')f(x')dx' \text{ (by the symm. of } p')$$

$$= -\int_{x' \in D'} p'(x')f(x')dx' \text{ (due to } -D' = D').$$

Thus, we conclude: $\int_{D'} p(x')g(x')dx' = 0$. Since $g(x) - g(0)$ is a linear function, we deduce:

$$E[g(d)] = \int_{x \in D} p(x)g(x)dx$$

$$= \int_{x \in D} p(x)\big(g(x) - g(0)\big)dx + \int_D p(x)g(0)dx$$

$$= \int_{x' \in D'} p(\bar{d} + x')\big(g(\bar{d} + x') - g(0)\big)dx' + g(0)$$

$$= \int_{x' \in D'} p'(x')\big(g(\bar{d} + x') - g(0)\big)dx' + g(0)$$

$$= \underbrace{\int_{x' \in D'} p'(x')\big(g(x') - g(0)\big)dx'}_{=0} + \underbrace{\int_{x' \in D'} p'(x')(g(\bar{d}) - g(0))dx'}_{=g(\bar{d}) - g(0)} + g(0)$$

$$= g(\bar{d}),$$

which concludes the proof.   □

For LS-2RO$_{q,u}$ (the problem with second-stage u and q and affine rules), the objective function constraint of the hybrid stochastic-robust formulation reads:

$$\eta \geq E\left[\sum_{t \in T}\left(c_t^c z_t + c_t^v\left(\sum_{j \in T}\xi_{jt}d_j + \Xi_t\right)\right.\right.$$
$$\left.\left.+c_t^i\left(\alpha^t u_0 + \sum_{i=1}^{t}\alpha^{t-i}\left(\sum_{j \in T}\xi_{ji}d_j + \Xi_i - d_i\right)\right)\right)\right]$$
$$= \sum_{t \in T}\left(c_t^c z_t + c_t^v\left(\sum_{j \in T}\xi_{jt}\bar{d}_j + \Xi_t\right) + c_t^i\left(\alpha^t u_0 + \sum_{i=1}^{t}\alpha^{t-i}\left(\sum_{j \in T}\xi_{ji}\bar{d}_j + \Xi_i - \bar{d}_i\right)\right)\right).$$

In the next section, we will apply this approach for the case of Γ-robustness.

For a finite D and assuming a uniform distribution with first moment $\bar{d}$, the objective can be restated as:

$$\eta \geq E\left[\sum_{t \in T}\left(c_t^c z_t + c_t^v\left(\sum_{j \in T}\xi_{jt}d_j + \Xi_t\right)\right.\right.$$
$$\left.\left.+c_t^i\left(\alpha^t u_0 + \sum_{i=1}^{t}\alpha^{t-i}\left(\sum_{j \in T}\xi_{ji}d_j + \Xi_i - d_i\right)\right)\right)\right]$$
$$= \frac{1}{|D|}\sum_{d \in D}\sum_{t \in T}\left(c_t^c z_t + c_t^v\left(\sum_{j \in T}\xi_{jt}d_j + \Xi_t\right)\right.$$
$$\left.+c_t^i\left(\alpha^t u_0 + \sum_{i=1}^{t}\alpha^{t-i}\left(\sum_{j \in T}\xi_{ji}d_j + \Xi_i - d_i\right)\right)\right)$$
$$= \sum_{t \in T}\left(c_t^c z_t + c_t^v\left(\sum_{j \in T}\xi_{jt}\bar{d}_j + \Xi_t\right)\right.$$
$$\left.+c_t^i\left(\alpha^t u_0 + \sum_{i=1}^{t}\alpha^{t-i}\left(\sum_{j \in T}\xi_{ji}\bar{d}_j + \Xi_i - \bar{d}_i\right)\right)\right).$$

In the next section, we adopt this approach for the case of finite uncertainty sets which, by construction (see further), are not symmetric in our application.

We remark that the hybrid approach is different from applying affine rules in a stochastic optimization setting, as, here, all constraints, except for the objective function one, are enforced in the worst case. We also remark that the hybrid stochastic-robust approach differs from the purely robust one only when affine rules are employed. This is because, if q is first-stage, the objective function reads: $\sum_{t \in T}(c_t^v q_t + c_t^c z_t + c_t^i(\alpha_{0t}u_0 + \sum_{i=1}^{t}\alpha_{it}(q_i - d_i)))$. After rearranging the terms, it becomes: $\sum_{t \in T}(c_t^v q_t + c_t^c z_t + c_t^i\alpha_{0t}u_0 + c_t^i\sum_{i=1}^{t}\alpha_{it}q_i - c_t^i\sum_{i=1}^{t}\alpha_{it}d_i)$. Since this function only depends on d in the additive term $c_t^i\sum_{i=1}^{t}\alpha_{it}d_i$ which involves no variables, the additive term can be dropped without affecting the optimality of the

solutions that are found. The same holds in expectation over d, where, by linearity, we have: $\sum_{t \in T} (c_t^v q_t + c_t^c z_t + c_t^i \alpha_{0t} u_0 + c_t^i \sum_{i=1}^{t} \alpha_{it} q_i - c_t^i \sum_{i=1}^{t} \alpha_{it} E[d_i])$, with d, here a random vector, only appearing in the additive term $c_t^i \sum_{i=1}^{t} \alpha_{it} E[d_i]$. Since the latter does not involve any variables, it can be dropped.

7 Computational results

We report on a set of computational results obtained with the models and methods that we have proposed and discussed. As an application, we consider a problem arising within the project *Robuste Optimierung der Stromversorgungsysteme* (Robust Optimization of Power Supply Systems), funded by the German *Bundesministerium für Wirtschaft und Energie* (Federal Ministry for Economic Affairs and Energy, BMWi). Originally, this is a Coproduction of Heat and Power (CHP) problem in which fuel is transformed into heat and power (at a fixed heat-to-power ratio) by a CHP device. When assuming that there is no internal power demand and that all the power which is produced can be sold to the market, the problem reduces to an instance of LS with heat as the unique good. Since heat can be stored (as hot water) for free, we set $c_t^i = 0$ for all $t \in T$. The cost of fuel is transferred to the static costs c^v, together with the profit generated by the amount of power that is sold. More precisely, given a static fuel cost c^f per unit of heat, a heat-to-fuel ratio v, a heat-to-power ratio ρ, and a static market profit c^m per unit of power, c^v is set to $c^v := c^f v - c^m \rho$.

7.1 Instances and setup

We consider a dataset spanning a period of 2 years (with some missing months) with hourly time steps. The first half of the data set (338 days) is used as *training set*, the second half (232 days) as *testing set*. The heat demand is taken from historical data of a portion of Frankfurt (of around 50000 households). For all $t \in T$, the storage bounds are set to $\underline{U}_t = 0$ MWh and $\overline{U}_t = 120$ MWh, while we set the production bounds to $\underline{Q}_t = 37.5$ MWh and $\overline{Q}_t = 125$ MWh. We also set $u_0 = 36$ MWh. The market prices for the power market are taken from EPEX SPOT (the European Power Exchange).

We consider two cases: finite uncertainty sets and Γ-robustness, with Γ uncertainty sets. Both sets are built based on the training set as well as on a heat demand forecast provided by our industrial partner ProCom GmbH. This forecast is generated in a two-stage fashion, with an autoregressive component and a neural network one, relying on temperature and calendar events as main influence factors.

The finite uncertainty sets D are constructed as follows. From the original forecast $\bar{d}$ for the current day, we single out the (up to 70) days from the set of historical time series (the training set) in which demand and forecast are closest in 1-norm. After computing the forecast error between the two, we create a realization where such error is added to the forecast of the current day (for which the problem is being solved).

For the Γ-robustness approach, the hours of the training set days are partitioned into three groups: *morning*, *midday*, and *evening*. Similarly, the demand forecasts are partitioned in the groups *low*, *medium*, and *high*. This way, we can associate each

pair (t, d_t) with a category in {morning, midday, evening} × {low, medium, high}. We consider all the hours in a given category over all the training days and look at their deviation (i.e., the forecast error) between the historical demand and the forecast $\bar{d}_t$. We then take the 95%-quantiles w.r.t. positive and negative deviations and let the deviation $\hat{d}_t$ at time t be equal to their maximum.

In our experiments, we do not introduce constraints on the end-of-horizon value of $u_{|T|}$. If required by the application though, it is possible to introduce a lower and/or upper bound on $u_{|T|}$, solve the problem thus obtained and, in case of infeasibility, adjust those values, iterating until two values yielding a feasible solution are found.[4] See Calafiore (2008) for more details on a similar procedure.

The experiments are run on an Intel i7-3770 3.40 GHz machine with 32 GB RAM using CPLEX 12.6 and AMPL as modeling language. All the instances are solved to optimality within default precision.

7.2 Realized robustness

While we optimize over uncertainty sets built with a combination of training set and forecast for the current day, we evaluate the quality of the results on the testing set (i.e., out of sample) in terms of *realized robustness*, simulating a real-world situation in which a solution found via robust optimization is implemented with a realized demand $\tilde{d}$ not necessarily belonging to the uncertainty set D. For a given $\tilde{d}$, we first transfer any violation occurring in the lower and upper bound constraints imposed on u and q to the balance constraints. This is achieved by redefining $q(\tilde{d})$ and $u(\tilde{d})$, for all $t \in T$, as follows:

$$q'_t(\tilde{d}) := \min \left\{ \max \left\{ \sum_{j \in T} \xi_{jt} \tilde{d}_j + \Xi_t, z_t \underline{Q}_t \right\}, z_t \overline{Q}_t \right\} \tag{11a}$$

$$u'_t(\tilde{d}) := \min \left\{ \max \left\{ \alpha_t u'_{t-1} + q'_t(\tilde{d}) - \tilde{d}_t, \underline{U}_t \right\}, \overline{U}_t \right\}, \tag{11b}$$

where $u'_0 := u_0$. This way, the lower and upper bound constraints on storage and production are always satisfied. We then define the violation of the balance constraints for the newly introduced $q'(\tilde{d})$ and $u'(\tilde{d})$ as:

$$v'_t(\tilde{d}) := \alpha_t u'_{t-1}(\tilde{d}) + q'_t(\tilde{d}) - \tilde{d}_t - u'_t(\tilde{d}).$$

For the case of a first-stage q, we simply let:

---

[4] Empirically, we observe that the solutions that we obtain tend to reduce the value of the storage at time $t = |T|$ to a small nonzero quantity, produced in previous periods $t < |T|$ with low production costs, which is then used to cope with potential positive deviations in the demand due to uncertainty—if no such deviations realize, the remaining unutilized product is stored in $u_{|T|}$.

$$q'_t(\tilde{d}) := q_t,$$

for all $t \in T$, while adopting the same definition as in Eq. (11b) for $u'(\tilde{d})$.

We remark that, if $v'(\tilde{d}) \leq 0$ and $\tilde{d} + v'(\tilde{d}) \geq 0$, the solution with $q'(\tilde{d})$ and $u'(\tilde{d})$ corresponds to what a practitioner would arguably do when an infeasibility is detected: production and storage are adjusted so to stay within their prescribed bounds, while the infeasibility is transferred, if possible, to the customer, by satisfying only the reduced demand $\tilde{d} + v'(\tilde{d})$ (remember that $v'(\tilde{d}) \leq 0$ is assumed to hold here), as opposed to the original $\tilde{d}$.

For each instance of the problem (i.e., for each realization $\tilde{d}$), we evaluate the quality of a robust solution in terms of how it performs when implemented with the given realization $\tilde{d}$ in terms of (total) *realized violation*:

$$v(\tilde{d}) := \sum_{t \in T} \left| v'_t(\tilde{d}) \right|,$$

and *realized cost*:

$$c(\tilde{d}) = \sum_{t \in T} \left(c_t^v q'_t(\tilde{d}) + c_t^c z_t(\tilde{d}) + c_t^i u'_t(\tilde{d}) \right).$$

Note that the realized cost can, in general, be quite different from the objective function value achieved by an optimal robust solution. Indeed, the latter accounts for either the worst-case or the expected cost over D (depending on whether a worst-case or a stochastic objective function is employed), as opposed to the former, which amounts to the cost that realizes for a given $\tilde{d}$.

The adoption of affine rules, although advantageous in terms of the extra flexibility it offers, introduces an element of so-called *nervousness* to the solutions as, for different realizations of $\tilde{d}$, the realized production values $q'_t(\tilde{d})$ are likely to be quite different. As mentioned in Tempelmeier and Herpers (2011), a large nervousness can have a negative impact on other managerial decisions which are taken after the lot sizing problem has been solved. To quantify this phenomenon, we also report a *cumulated measure of nervousness*, defined as:

$$n(\tilde{d}) = \sum_{t \in T} \left| q_t(\bar{d}) - q'_t(\tilde{d}) \right|,$$

where $\bar{d}$ is the nominal value of d (the forecast).

7.3 Computational study

When affine rules are in place, all the experiments with either $|D| = 0$ or $\Gamma = 0$ (that is, those relying only on the forecast) are obtained by letting $\xi_{jt} = 0$ for all $j, t \in T$.

When presenting our results, we report the sum over all the instances in our data set of the realized violation, realized nervousness, and realized cost, for each value of $|D|$ or Γ. The charts that we report show the violation-versus-cost and nervousness-versus-cost curves, thus allowing for, visually, assessing the tradeoff between these quantities

that is achieved with the different approaches. The curves are obtained by joining each pair of points obtained with two consecutive values of $|D|$ or Γ. Computing times are reported in Tables 1, 2, and 3.

Table 1 Results for LS-2RO$_u$ with finite uncertainty sets and Γ-robustness, obtained by the application of Theorem 1

	Finite uncertainty set							
$|D|$	0	10	20	30	40	50	60	70
Viol.	3802	1562	1226	1097	1017	999	969	965
Costs	2675	2900	2953	2996	3017	3037	3054	3063
Time	15	15	15	15	15	15	15	15

	Γ-robustness						
Γ	0	1	2	3	4	5	6
Viol.	3802	2635	1982	1507	1202	972	882
Costs	2675	2741	2807	2877	2951	3026	3105
Time	15	15	15	15	15	15	14

Table 2 Results for LS-2RO$_{q,u}$-AR with finite uncertainty sets

$|D|$	0	10	20	30	40	50	60	70
Worst-case objective: restricted coefficients ξ_{tj}								
Viol.	3802	1600	1136	874	629	612	481	425
Nerv.	0	7241	7319	6889	6866	6717	6652	6682
Costs	2675	2759	2758	2773	2788	2793	2800	2806
Time	16	1541	7108	12026	19424	27493	34566	41562
Worst-case objective: unrestricted coefficients ξ_{tj}								
Viol.	3802	9841	6635	3205	1857	1347	1383	1242
Nerv.	0	28417	27549	20714	17819	16205	15309	14718
Costs	2675	2788	2810	2803	2811	2801	2811	2813
Time	15	862	4580	11247	19564	27369	34417	39647
Stochastic objective: restricted coefficients ξ_{tj}								
Viol.	3802	1376	851	582	456	378	329	271
Nerv.	0	6552	6303	6359	6378	6357	6315	6385
Costs	2675	2697	2711	2720	2726	2730	2734	2737
Time	15	1445	5885	12117	16543	23101	29357	35916
Stochastic objective: unrestricted coefficients ξ_{tj}								
Viol.	3802	3854	3157	946	684	540	441	402
Nerv.	0	9612	10854	7933	7427	7317	7192	7332
Costs	2675	2689	2694	2705	2712	2714	2719	2721
Time	15	767	2842	6975	9971	13792	17358	21161

Table 3 Results for LS-2RO$_{q,u}$-AR with Γ uncertainty sets

Γ	0	1	2	3	4	5	6	7	8	9	10	11	12
Worst-case objective: restricted coefficients ξ_{tj}													
Viol.	3802	2048	1258	906	658	554	504	479	482	429	428	422	411
Nerv.	0	4619	5857	6000	6168	6296	6193	6089	6095	6030	6016	5958	5935
Costs	2609	2736	2786	2830	2867	2896	2917	2930	2944	2956	2968	2975	2983
Time	18	552	955	1481	1983	2304	2724	2787	2898	2974	2994	3011	3000
Worst-case objective: unrestricted coefficients ξ_{tj}													
Viol.	3802	1928	1194	977	651	581	495	477	493	426	431	421	406
Nerv.	0	4555	5918	6051	6314	6359	6271	6179	6085	6018	6028	5997	5924
Costs	2675	2736	2785	2831	2866	2896	2916	2930	2944	2957	2968	2976	2983
Time	19	534	1098	1792	2365	2848	3291	3627	3676	3677	3721	3724	3587
Stochastic objective: restricted coefficients ξ_{tj}													
Viol.	3802	1041	246	159	122	67	54	53	53	53	53	53	54
Nerv.	0	4606	5108	5255	5473	5655	5716	5726	5781	5790	5792	5813	5818
Costs	2609	2744	2786	2815	2833	2843	2848	2853	2854	2857	2858	2859	2860
Time	18	303	449	655	743	799	829	862	876	874	866	863	867
Stochastic objective: unrestricted coefficients ξ_{tj}													
Viol.	3802	1040	362	134	93	70	54	53	53	53	53	53	54
Nerv.	0	4138	4893	5236	5511	5645	5749	5750	5792	5798	5809	5819	5829
Costs	2675	2750	2786	2815	2831	2842	2847	2853	2854	2856	2858	2859	2859
Time	18	296	499	713	910	1048	1189	1235	1297	1289	1328	1281	1301

Fig. 1 Results for LS-2RO$_u$, obtained by the application of Theorem 1, comparing finite uncertainty sets to Γ-robustness with increasing values of $|D|$ and Γ
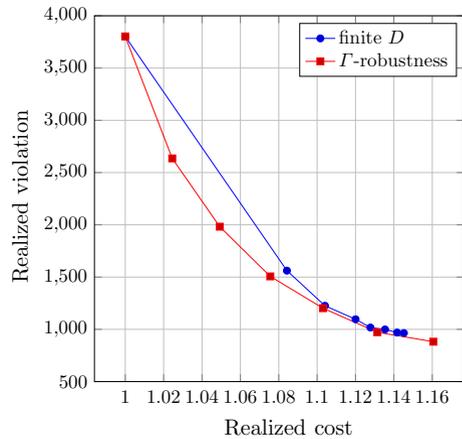
7.3.1 First-stage production and second-stage storage variables (LS-2RO$_u$)

The results for LS-2RO$_u$ (the approach with first-stage production variables q and second-stage storage variables u) are obtained by solving the problem as an instance of LS-DET, relying on Theorem 1. We let $|D| = 0, 10, 20, 30, 50, 60, 70$ for the finite uncertainty set case and $\Gamma = 0, 1, 2, 3, 4, 5, 6$ for Γ-robustness. We do not go beyond $\Gamma = 6$ as, for larger values, the problem becomes infeasible for at least a few instances. The results are reported in Table 1, with an illustration in terms of realized violations and costs provided in Fig. 1. Nervousness is not reported as, due to q being first-stage, it is equal to zero for all instances.

Overall, we observe that the adoption of Γ-robustness provides slightly better results, with a curve, as seen in Fig. 1, which dominates that for finite scenarios, providing smaller realized violations and costs, especially for small values of $|D|$ and Γ. Interestingly, as these two parameters increase, the two approaches become almost equivalent.

Although, as we will see in the remainder of the section, this approach is limited in terms of quality of the solutions, it comes at the advantage of being very fast from a computational standpoint and, by definition, of not suffering from nervousness.

7.3.2 Second-stage storage and production variables with affine rules (LS-2RO$_{q,u}$-AR)

We consider LS-2RO$_{q,u}$-AR (the two stage approach with second-stage u and q). We illustrate the results with finite uncertainty sets and Γ-robustness, considering both the cases of a worst-case objective function and a stochastic one, with either unrestricted coefficients ξ_{tj} or coefficients restricted to $\xi_{tj} \in [-1, 1]$ for all $j, t \in T$.

The results for the case of finite uncertainty sets are reported in Table 2 and illustrated in Fig. 2.

Considering violations and costs, when adopting a worst-case objective, we register highly "unstable" solutions when no restrictions are imposed in ξ_{jt}. The solutions that
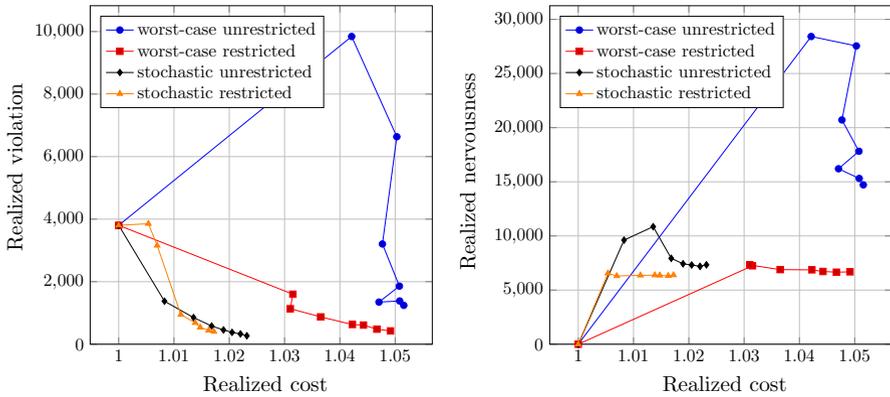
Fig. 2 Violation and nervousness VS realized costs for LS-2RO$_{q,u}$ with finite scenarios and affine rules, with a worst-case or stochastic objective

are obtained when imposing $\xi_{tj} \in [-1, 1]$ for all $j, t \in T$ are much more stable and largely dominate the previous ones. The introduction of a stochastic objective yields very large improvements w.r.t. both realized cost and violation, while also introducing a good deal of stability even for small values of $|D|$. The restriction of the coefficients provides an improvement also with a stochastic objective, but not as substantial as with a worst-case one.

Notice that, with $|D| = 10, 20$, the results are substantially worse than those obtained for larger values of $|D|$. This is a consequence of *overfitting*. Indeed, if $|D| \ll |T|$, there are infinitely many (with $|D| - |T|$ degrees of freedom) affine functions satisfying Eq. (10) and, as such, we are likely to obtain solutions with a large fitting error on demand vectors $\tilde{d} \notin D$, such as all those in the testing set.

We remark that, although $|D| = 30 > 24 = |T|$, and, thus, the fitting subproblem (assuming the vectors $d \in D$ are in general position) is well-posed, $|D| = 30$ seems to be still insufficient to suitably capture the structure of the "hidden" uncertainty set $\tilde{D}$ as, for $|D| \geq 40$, much better results are obtained. Overall, as expected, the quality of the results on the testing set increases with the size of the finite uncertainty set D.

As to nervousness, we observe larger values when unrestricted coefficients are used, as opposed to the restricted case, and when the objective function is minimized in the worst-case, as opposed to in expectation. Not surprisingly, higher values of nervousness are observed in presence of overfitting ($|D| < |T|$), which then start declining for $|D| \geq 30$. Overall, the values of nervousness tend to decrease for larger values of $|D|$ in all the configurations, although not completely monotonically. In all approaches but the worst-case unrestricted one, the nervousness levels seem to converge for larger $|D|$. Although it is unclear whether the nervousness values are converging for the worst-case unrestricted approach, they are clearly getting smaller for larger values of $|D|$.

Analogous experiments for the case of Γ-uncertainty sets and Γ-robustness are reported in Table 3 and illustrated in Fig. 3.
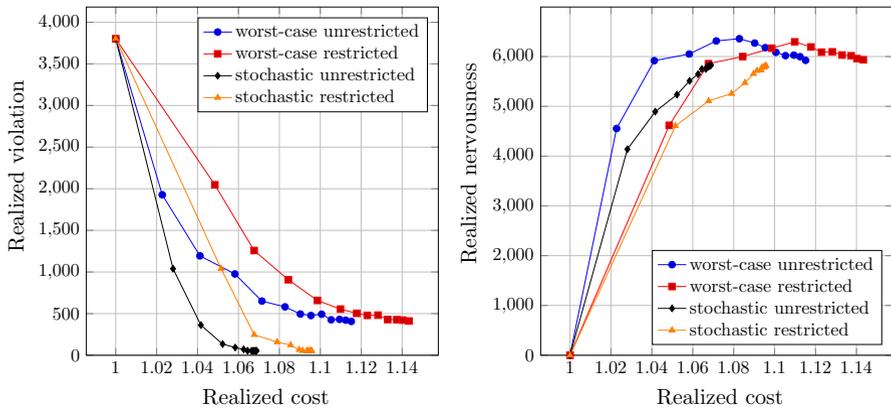
Fig. 3 Violation and nervousness VS realized costs for LS-2RO$_{q,u}$ with Γ-robustness and affine rules, with a worst-case or stochastic objective

Considering violations and costs, the results seem quite stable, even with unrestricted coefficients. This is possibly due to the fact that Γ uncertainty sets are continuous and full-dimensional and, as a consequence, the set of points on top of which the affine rules are constructed always contains (except for Γ = 0) sufficiently many data points in general position and, thus, overfitting cannot happen. What is more, we see that, in this case, restricting the coefficients hinders the quality of the solutions. As expected, we also observe that the results with a stochastic objective function dominate those with a worst-case one.

As to nervousness, we observe that, although its value increases substantially with small values of Γ, it seems to converge to a constant for a larger Γ. This value is comparable to that obtained with the three more "stable" approaches with finite uncertainty sets (worst-case restricted, stochastic unrestricted, and stochastic restricted).

7.3.3 A note on computing times

We remark that, for each setting, the hybrid stochastic-robust approach requires, for bigger uncertainty sets (larger $|D|$ or larger Γ), much shorter computing times as opposed to the approach with a worst-case objective. For Γ-robustness and with the largest values of Γ, the difference in computing time can amount to a factor as large as eight.

We also observe that the Γ-robustness approach is much faster than the finite uncertainty sets one. This is, most likely, due to the fact that, with a finite D, we have to replicate each constraint containing d exactly $|D|$ times whereas, with Γ-robustness, we can employ the compact reformulation reported in "Appendix B".

As it is clear when comparing Tables 1, 2, and 3, the method with first-stage production is faster than the other ones employing affine rules by almost two orders of magnitude.

Fig. 4 Hybrid stochastic-robust approaches with unrestricted coefficients with either finite uncertainty sets or Γ-robustness, also compared to the approach with first-stage q and second-stage u with Γ-robustness
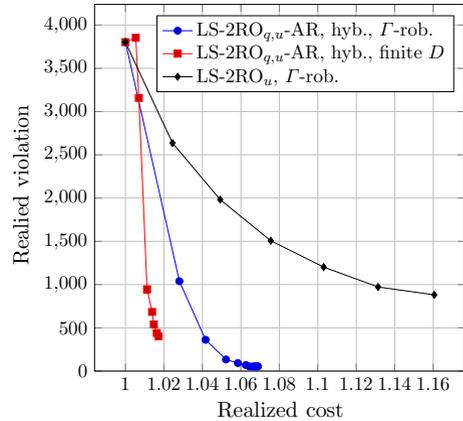
7.3.4 Comparison of the best methods

We conclude with a final comparison of the best method, the one with second-stage production and affine rules employing a stochastic objective with unrestricted coefficients in the two cases of finite uncertainty sets and Γ-robustness. For the sake of comparisons, we also consider LS-2RO$_u$ (that with first-stage production) with Γ-robustness (which performs better than the one with finite uncertainty sets).

The results are reported in Fig. 4. Here, LS-2RO$_u$ with Γ-robustness is clearly dominated by the other two. While, with finite uncertainty sets, the affine rules fare rather well in terms of costs, the approach with Γ-robustness yields solutions with substantially smaller violations, differently from the method with a finite uncertainty set, whose solutions converge, experimentally, to a total violation of 400, without further improvements for a larger $|D|$. Assuming the larger computational effort that is required to construct solutions with this method can be undertaken, it should be preferred due to offering higher quality solutions.

8 Concluding remarks

We have considered a variant of the lot sizing problem with storage losses subject to demand uncertainty. We have developed two two-stage robust optimization approaches, both with first-stage set-up variables, one with second-stage storage variables but first-stage production variables, and another one in which production and storage variables are both second-stage. For greater tractability, we have introduced affine rules to tackle the second problem, also encompassing a regularization aspect. We have then proposed a hybrid stochastic-robust approach with a stochastic objective function and robustness in all its constraints, arguably less conservative than the purely robust method with a worst-case objective.

Computational experiments on a heat production problem have shown that the two-stage approach with second-stage production and storage variables outperforms that with first-stage production variables. The solutions that we have obtained in this set-

ting with Γ-robustness are better in terms of cost and violations than those obtained with a finite uncertainty set, and are also much faster to be computed from a computational perspective. We have found that the hybrid stochastic-robust approach yields the overall best results, resulting in smaller costs and smaller violations when compared to the one with a worst-case objective, and also being, with Γ-robustness, much faster than the former.

Appendix A: Derivations in the proof of Theorem 1

For $t > 1$, define $\check{d} \in D$ such that the following holds:

$$d'_{t-1} = \max_{d \in D} \left\{ d_{t-1} - \left(\sum_{i=1}^{t-2} \alpha_{i,t-1} \left(d'_i - d_i \right) \right) \right\} = \check{d}_{t-1} - \left(\sum_{i=1}^{t-2} \alpha_{i,t-1} \left(d'_i - \check{d}_i \right) \right).$$

As a consequence, we deduce:

$$
\begin{aligned}
d'_t &= \max_{d \in D} \left\{ d_t - \sum_{i=1}^{t-1} \alpha_{i,t} \left(d'_i - d_i \right) \right\} \\
&= \max_{d \in D} \left\{ d_t - \alpha_t \left(\sum_{i=1}^{t-2} \alpha_{i,t-1} \left(d'_i - d_i \right) + d'_{t-1} - d_{t-1} \right) \right\} \\
&= \max_{d \in D} \left\{ d_t - \alpha_t \left(d'_{t-1} - \left(d_{t-1} - \sum_{i=1}^{t-2} \alpha_{i,t-1} \left(d'_i - d_i \right) \right) \right) \right\} \\
&\geq \check{d}_t - \alpha_t \left(d'_{t-1} - \left(\check{d}_{t-1} - \sum_{i=1}^{t-2} \alpha_{i,t-1} \left(d'_i - \check{d}_i \right) \right) \right) \\
&= \check{d}_t - \alpha_t \left(d'_{t-1} - d'_{t-1} \right) \\
&= \check{d}_t \geq 0.
\end{aligned}
$$

With this, we can then show:

$$
\begin{aligned}
\min_{d \in D} \{ u_t(d) \} &= \min_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} (q_i - d_i) \right\} \\
&= \min_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t-1} \alpha_{i,t} (q_i - d_i) + (q_t - d_t) \right\}
\end{aligned}
$$

$$= \min_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t-1} \alpha_{i,t} q_i - \sum_{i=1}^{t-1} \alpha_{i,t} d_i + q_t - d_t + \sum_{i=1}^{t-1} \alpha_{i,t} d_i' - \sum_{i=1}^{t-1} \alpha_{i,t} d_i' \right\}$$

$$= \min_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t-1} \alpha_{i,t} q_i - \sum_{i=1}^{t-1} \alpha_{i,t} d_i' + q_t - d_t + \sum_{i=1}^{t-1} \alpha_{i,t} d_i' - \sum_{i=1}^{t-1} \alpha_{i,t} d_i \right\}$$

$$= \min_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t-1} \alpha_{i,t} \left(q_i - d_i' \right) + q_t - d_t + \sum_{i=1}^{t-1} \alpha_{i,t} \left(d_i' - d_i \right) \right\}$$

$$= \alpha_{0t} u_0 + \sum_{i=1}^{t-1} \alpha_{i,t} \left(q_i - d_i' \right) + q_t + \min_{d \in D} \left\{ -d_t + \sum_{i=1}^{t-1} \alpha_{i,t} \left(d_i' - d_i \right) \right\}$$

$$= \alpha_{0t} u_0 + \sum_{i=1}^{t-1} \alpha_{i,t} \left(q_i - d_i' \right) + q_t - \max_{d \in D} \left\{ d_t - \sum_{i=1}^{t-1} \alpha_{i,t} \left(d_i' - d_i \right) \right\}$$

$$= \alpha_{0t} u_0 + \sum_{i=1}^{t-1} \alpha_{i,t} \left(q_i - d_i' \right) + \left(q_t - d_t' \right)$$

$$= \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} \left(q_i - d_i' \right) = u_t.$$

Similarly, we deduce:

$$\max_{d \in D} \{ u_t(d) \} = \max_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} (q_i - d_i) \right\}$$

$$= \max_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} q_i - \sum_{i=1}^{t} \alpha_{i,t} d_i \right\}$$

$$= \max_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} q_i - \sum_{i=1}^{t} \alpha_{i,t} d_i + \sum_{i=1}^{t} \alpha_{i,t} d_i' - \sum_{i=1}^{t} \alpha_{i,t} d_i' \right\}$$

$$= \max_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} q_i - \sum_{i=1}^{t} \alpha_{i,t} d_i' + \sum_{i=1}^{t} \alpha_{i,t} d_i' - \sum_{i=1}^{t} \alpha_{i,t} d_i \right\}$$

$$= \max_{d \in D} \left\{ \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} (q_i - d_i') + \sum_{i=1}^{t} \alpha_{i,t} (d_i' - d_i) \right\}$$

$$= \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} (q_i - d_i') + \max_{d \in D} \left\{ \sum_{i=1}^{t} \alpha_{i,t} (d_i' - d_i) \right\} = u_t + u_t'.$$

Lastly, we compute η:

$$\eta = \max_{d \in D} \left\{ \sum_{t \in T} c_t^i u_t(d) \right\}$$

$$= \max_{d \in D} \left\{ \sum_{t \in T} c_t^i \left(\alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t}(q_i - d_i) \right) \right\}$$

$$= \max_{d \in D} \left\{ (\sum_{t \in T} c_t^i \left(\alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} q_i - \sum_{i=1}^{t} \alpha_{i,t} d_i + \sum_{i=1}^{t} \alpha_{i,t} d_i' - \sum_{i=1}^{t} \alpha_{i,t} d_i' \right) \right\}$$

$$= \max_{d \in D} \left\{ \sum_{t \in T} c_t^i \left(\alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t} q_i - \sum_{i=1}^{t} \alpha_{i,t} d_i' + \sum_{i=1}^{t} \alpha_{i,t} d_i' - \sum_{i=1}^{t} \alpha_{i,t} d_i \right) \right\}$$

$$= \max_{d \in D} \left\{ \sum_{t \in T} c_t^i \left(\alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t}(q_i - d_i') + \sum_{i=1}^{t} \alpha_{i,t}(d_i' - d_i) \right) \right\}$$

$$= \max_{d \in D} \left\{ \sum_{t \in T} c_t^i \left(\alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t}(q_i - d_i') \right) + \sum_{t \in T} c_t^i \sum_{i=1}^{t} \alpha_{i,t}(d_i' - d_i) \right\}$$

$$= \sum_{t \in T} c_t^i \left(\alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{i,t}(q_i - d_i') \right) + \max_{d \in D} \left\{ \sum_{t \in T} c_t^i \sum_{i=1}^{t} \alpha_{i,t}(d_i' - d_i) \right\}$$

$$= \sum_{t \in T} c_t^i u_t + \text{const.}$$

Appendix B: MILP formulation for LS-2RO$_{q,u}$-AR and Γ-robustness

The MILP formulation with uncertainty intervals $[\bar{d}_t - \hat{d}_t, \bar{d}_t + \hat{d}_t]$ reads:

$$\min \quad \eta$$

$$\text{s.t.} \quad \eta \geq \sum_{t \in T} \left(c_t^v \Xi_t + c_t^c z_t + c_t^i \left(\alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{jt}(\Xi_i - \bar{d}_i) \right) \right) + \sum_{t \in T} \pi_t + \Gamma \pi_0 \qquad \forall t \in T$$

$$\pi_0 + \pi_t \geq \left(\sum_{i=t}^{|T|} c_i^v \xi_{ti} + \sum_{k \in T} c_k^i \sum_{i=t}^{k} \alpha_{ik} \xi_{ti} - \sum_{j=t}^{|T|} \alpha_{jt} c_j^i \right) \hat{d}_t \qquad \forall t \in T$$

$$\pi_0 + \pi_t \geq - \left(\sum_{i=t}^{|T|} c_i^v \xi_{ti} + \sum_{k \in T} c_k^i \sum_{i=t}^{k} \alpha_{ik} \xi_{ti} - \sum_{j=t}^{|T|} \alpha_{jt} c_j^i \right) \hat{d}_t \qquad \forall t \in T$$

$$\underline{U}_t \leq \alpha_{0t} u_0 + \sum_{i=1}^{t} \alpha_{it}(\Xi_i - \bar{d}_i) - \sum_{k=1}^{t} \rho_{tk} - \Gamma \rho_{t0} \qquad \forall t \in T$$

$$\overline{U}_t \geq \alpha^t u_0 + \sum_{i=1}^{t} \alpha_{it}(\Xi_i - \bar{d}_i) + \sum_{k=1}^{t} \rho_{tk} + \Gamma \rho_{t0} \qquad \forall t \in T$$

$$\rho_{t0} + \rho_{tk} \geq - \left(\sum_{i=k}^{t} \alpha_{it} \xi_{ki} - \alpha_{kt} \right) \hat{d}_t \qquad \forall t, k \in T : k \leq t$$

$$\rho_{t0} + \rho_{tk} \geq \left(\sum_{i=k}^{t} \alpha_{it} \xi_{ki} - \alpha_{kt} \right) \hat{d}_t \qquad \forall t, k \in T : k \leq t$$

$$z_t \underline{Q}_t \leq \Xi_t - \sum_{k=1}^{t} \phi_{tk} - \Gamma \phi_{t0} \qquad \forall t \in T$$

$$z_t \overline{Q}_t \geq \Xi_t + \sum_{k=1}^{t} \phi_{tk} + \Gamma \phi_{t0} \qquad \forall t \in T$$

$$\phi_{t0} + \phi_{tk} \geq \xi_{kt} \hat{d}_t \qquad \forall t, k \in T : k \leq t$$

$$\phi_{t0} + \phi_{tk} \geq -\xi_{kt} \hat{d}_t \qquad \forall t, k \in T : k \leq t$$

$$z_t \in \{0, 1\} \qquad \forall t \in T$$

$$\eta, \xi_{jt}, \Xi_t \in \mathbb{R} \qquad \forall t, j \in T$$

$$\pi_0, \pi_t, \rho_{t0}, \rho_{tk}, \phi_{t0}, \phi_{tk} \geq 0 \qquad \forall t, k \in T : k \leq t.$$

References

Atamtürk A, Küçükyavuz S (2008) An algorithm for lot sizing with inventory bounds and fixed costs. Oper Res Lett 36(3):297–299

Ben-Tal A, Goryashko A, Guslitzer E, Nemirovski A (2004) Adjustable robust solutions of uncertain linear programs. Math Program 99(2):351–376

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) Robust optimization. Princeton University Press, Princeton

Bertsimas D, Georghiou A (2015) Design of near optimal decision rules in multistage adaptive mixed-integer optimization. Oper Res 63(3):610–627

Bertsimas D, Sim M (2003) Robust discrete optimization and network flows. Math Program 98(1–3):49–71

Bertsimas D, Sim M (2004) The price of robustness. Oper Res 52(1):35–53

Bertsimas B, Thiele A (2004) A robust optimization approach to supply chain management. IPCO 2004:86–100

Bertsimas D, Thiele A (2006) A robust optimization approach to inventory theory. Oper Res 54(1):150–168

Bishop CM (2006) Pattern recognition. Mach Learn 128

Bookbinder JH, Tan J-Y (1988) Strategies for the probabilistic lot-sizing problem with service-level constraints. Manag Sci 34(9):1096–1108

Büsing C, Koster AMCA, Kutschka M (2011) Recoverable robust knapsacks: the discrete scenario case. Optim Lett 5(3):379–392

Calafiore GC (2008) Multi-period portfolio optimization with linear control policies. Automatica 44(10):2463–2473

Caramanis C (2006) Adaptable optimization: theory and algorithms. PhD thesis, Massachusetts Institute of Technology

Coniglio S, Koster AMCA, Spiekermann N (2016) On robust lot sizing problems with storage deterioration, with applications to heat and power cogeneration. Lecture notes in computer science. pp 1–12

Florian M, Lenstra JK, Rinnooy Kan AHG (1980) Deterministic production planning: algorithms and complexity. Manag Sci 26(7):669–679

Hellion F, Mangione F, Penz B (2012) A polynomial time algorithm to solve the single-item capacitated lot sizing problem with minimum order quantities and concave costs. Eur J Oper Res 222(1):10–16

Hsu VN (2000) Dynamic economic lot size model with perishable inventory. Manag Sci 46(8):1159–1169

Liyanage LH, Shanthikumar JG (2005) A practical inventory control policy using operational statistics. Oper Res Lett 33(4):341–348

Pochet Y, Wolsey LA (2006) Production planning by mixed integer programming. Springer, New York

Scarft H (1960) The optimality of (s, S) policies in the dynamic inventory problem. In: Arrow KJ, Karlin S, Suppes P (eds) Mathemtical methods in the social sciences, vol 1. Stanford University Press, p 196

Tempelmeier H (2013) Stochastic lot sizing problems. In: Handbook of stochastic models and analysis of manufacturing system operations. Springer, New York, pp 313–344

Tempelmeier H, Herpers S (2011) Dynamic uncapacitated lot sizing with random demand under a fillrate constraint. Eur J Oper Res 212(3):497–507