On the Sparsity of Optimal Linear Decision Rules in Robust Inventory Management

Haihao Lu
Booth School of Business
University of Chicago, haihao.lu@chicagobooth.edu

Bradley Sturt
Department of Information and Decision Sciences
University of Illinois at Chicago, bsturt@uic.edu

We consider the widely-studied class of production-inventory problems from the seminal work of Ben-Tal et al. (2004) on linear decision rules in robust optimization. We prove that there always exists an optimal linear decision rule for this class of problems in which the number of nonzero parameters in the linear decision rule is equal to a small constant times the number of parameters in a static decision rule. This result demonstrates that the celebrated performance of linear decision rules in such robust inventory management problems can be obtained without sacrificing the simplicity of static decision rules. From a practical standpoint, our result lays a theoretical foundation for the growing stream of literature on harnessing sparsity to develop practicable algorithms for computing optimal linear decision rules in operational planning problems with many time periods. Our proof is based on a principled analysis of extreme points of linear programming formulations, and we show that our proof techniques extend to other fundamental classes of robust optimization problems from the literature.

Key words: Linear programming; robust optimization; linear decision rules.

1. Introduction

Over the past two decades, robust optimization has emerged as a leading approach in operations research and management science for sequential decision-making under uncertainty. One of the major reasons for its popularity is computational: adaptive robust optimization problems are often amenable to efficient approximations in complex, real-world operational planning problems. The successful approximation techniques for adaptive robust optimization typically rely on restricting the control policies to a simple functional form, with the most popular restriction being to linear decision rules.

The success of linear decision rules in addressing real-world problems is very impressive. On the empirical side, linear decision rules have been found to exhibit strong performance in a myriad of high-stakes robust optimization applications such as disaster response, personalized healthcare, sustainable energy management, transportation routing, and many others [28, 24, 14, 27, 16]. On the theoretical side, a burgeoning literature has established that linear decision rules are provably optimal control policies in many classes of robust optimization problems [7, 9, 23, 20, 1, 28, 15, 19]. The aforementioned papers all build upon the seminal work of Ben-Tal et al. [4], which showed that optimal linear decision rules for robust optimization can be computed in polynomial time even in "fairly complicated models with high-dimensional state spaces and many stages" [4, p. 374].

However, the impressive performance of linear decision rules comes at a price. Compared to simpler classes of control policies such as static decision rules, linear decision rules are represented using a significantly greater number of parameters. Specifically, the number of parameters in linear decision rules grows quadratically in the number of stages of the robust optimization problem, whereas the number of parameters for static decision rules grows linearly in the number of stages. Because the number of parameters for representing linear decision rules can be enormous, computing optimal linear decision rules can require "the solution of monolithic and often dense optimization problems" [18, p.814] which can exceed computer memory in robust optimization problems with as few as seventy-five stages [18, p.827]. This is problematic because real-world applications routinely have many hundreds of stages, particularly when a discrete-time robust optimization problem is approximating a continuous-time operational planning problem.

To get around this, a recent stream of research has advocated for imposing sparsity constraints onto linear decision rules in robust optimization problems. The driving insight behind this stream of research is that if many of the parameters of linear decision rules are forced to be equal to zero, then the problem of optimizing the remaining parameters can often be solved more efficiently. Numerical studies have shown that applying this insight can yield tremendous improvements in computation times with only small losses in performance in applications such as newsvendor networks in pharmaceutical supply chains [3, §4], unit commitment problems in power systems [25, §4.3], and location-transportation problems [2], and column-generation techniques have been proposed for identifying the subset of parameters of linear decision rules to set to zero [29]. The harnessing of sparsity thus stands as one of the most promising directions for efficiently computing optimal linear decision rules for the sizes of robust optimization problems that arise in industry.

Ultimately, the success of using sparsity to develop faster and more memory-efficient algorithms will hinge on whether robust optimization problems have optimal linear decision rules that are sparse. Indeed, if a robust optimization problem does not have optimal linear decision rules that are sparse, then deploying a sparsity-imposing algorithm can risk leading to control policies with unexpectedly and undesirably suboptimal performance. Moreover, if the number of zero parameters in optimal linear decision rules is not a significant proportion of the total parameters, then the benefits of developing sparsity-imposing algorithms can be limited, and research efforts may be better spent on developing alternative algorithmic techniques such as those based on primal-dual saddle point formulations or online convex optimization [26, 21, 6]. To the best of our knowledge, the fundamental question of whether a significant number of zero parameters can be expected in optimal linear decision rules in any class of robust optimization problems with many time periods has remained open.

In this paper, we resolve this open question by revisiting the widely-studied class of production-inventory problems from the seminal work of Ben-Tal et al [4] on linear decision rules in robust optimization. This class of production-inventory problems served as the key illustration in [4] that optimal linear decision rules can be computed in polynomial time and provide excellent performance in realistic and complex robust inventory management problems. It has since become one of the most popular classes of problems in the robust optimization literature, serving as a test bed for the complex real-world applications in which linear decision rules are routinely used; see [12] and references therein. The class of production-inventory problems involves a firm which dynamically determines production quantities at multiple factories over a selling season, a single product with uncertain demand that lies in interval uncertainty sets, and complex business constraints that link the inventory levels and production decisions across multiple periods of time.

Our main contribution is a sparsity guarantee for optimal linear decision rules in this class of production-inventory problems. To state our result, we recall for any instance of such production-inventory problems with E factories and T time periods that the number of parameters for representing linear decision rules is equal to $\frac{1}{2}ET(T+1) = \mathcal{O}(ET^2)$. For this class of problems, our main result establishes that there always exists an optimal linear decision rule in which the number of nonzero parameters is at most equal to $2 + 8E + 10T + 6ET = \mathcal{O}(ET)$ whenever an optimal linear decision rule for the instance exists. In other words, although the number of parameters for representing linear decision rules grows quadratically in the number of time periods, our result shows for the first time that the minimum number of nonzero parameters for representing optimal linear decision rules grows only *linearly* in the number of time periods. We complement our bounds with an explicit characterization of the optimal linear decision rules that are sparse, and we demonstrate via numerical experiments that our results are indicative of practice.

Our sparsity guarantees can ultimately be viewed as valuable to the practice of robust optimization for two key reasons. First, our results prove for the first time that sparsity can be a fundamental property of optimal linear decision rules in practically-important classes of robust optimization problems when the number of time periods is large. Our results thus provide the first theoretical foundation for the validity and potential impact of the growing stream of research that is based on harnessing sparsity to develop faster and more memory-efficient algorithms for computing optimal linear decision rules in real-world applications of robust optimization [3, 25, 2, 29]. Second, our results establish that the celebrated performance of linear decision rules in such robust inventory management problems can be obtained without sacrificing the simplicity or interpretability of static decision rules. We remark that our results are not exclusive to the class of production-inventory problems from Ben-Tal et al. [4]; indeed, we show in §5 that our proof techniques can be extended to other widely-studied classes of robust optimization problems, such as inventory management problems with lead times and the class of dynamic newsvendor problems from [5, 9, 23].

The rest of this paper is organized as follows. In §2, we present a background on linear decision rules in robust optimization and production-inventory problems. In §3, we state our main result and present an overview of our proof. In §4, we illustrate the practical implications of our main result via numerical experiments. In §5, we discuss extensions of our main result to other classes of robust optimization problems. In §6, we conclude and discuss future directions of research. All technical proofs can be found in the supplemental appendices.

Notation. We let $\mathbb{R}$ denote the real numbers, we use boldface letters like $\mathbf{x}$ to non-scalar quantities like vectors and matrices, and we let $\|\mathbf{x}\|_0$ denote the number of nonzeros in $\mathbf{x}$. Given an optimization problem $\min_{\mathbf{x}} f(\mathbf{x})$, we say that a solution $\bar{\mathbf{x}}$ is feasible for the optimization problem if and only if the cost satisfies $f(\bar{\mathbf{x}}) < \infty$. It follows from this notation that $\min_{\mathbf{x}} f(\mathbf{x}) < \infty$ if and only if an optimization problem $\min_{\mathbf{x}} f(\mathbf{x})$ has a nonempty feasible region.

2. Background and Problem Setting

2.1. Robust Optimization and Linear Decision Rules

We consider robust optimization problems faced by firms in which decisions $\mathbf{x}_1, \ldots, \mathbf{x}_T \in \mathbb{R}^n$ are made sequentially over a planning horizon of T discrete stages. In the beginning of each stage $t \in [T] \equiv \{1, \ldots, T\}$, the firm observes a uncertain variable $\zeta_t \in \mathbb{R}$ that is chosen adversarially from an uncertainty set denoted by $\mathcal{U}_t$. The goal of the firm is to choose decisions sequentially, that is, adapting to the uncertain variables observed in the past stages, to minimize the firm's cost under an adversarial choice of the uncertain variables. Such problems are denoted generically by

$$\max_{\zeta_1 \in \mathcal{U}_1} \min_{\mathbf{x}_1 \in \mathbb{R}^n} \cdots \max_{\zeta_T \in \mathcal{U}_T} \min_{\mathbf{x}_T \in \mathbb{R}^n} C(\mathbf{x}_1, \ldots, \mathbf{x}_T, \zeta_1, \ldots, \zeta_T), \tag{RO}$$

where we adopt the convention that the cost $C(\mathbf{x}_1, \ldots, \mathbf{x}_T, \zeta_1, \ldots, \zeta_T) \in \mathbb{R} \cup \{\infty\}$ is equal to infinity if the chosen decisions are infeasible for the given realization of the uncertain variables. For comprehensive introductions to the many applications of adaptive robust optimization problems of the form (RO), we refer the reader to excellent survey papers such as [13, 17, 30, 10].

Our work focuses on one of the most popular approximation methods for solving (RO), denoted by the optimization problem (LDR). Proposed in the seminal work of Ben-Tal et al. [4], (LDR) aims to obtain a computationally tractable approximation of (RO) by restricting the decisions in each stage to be a linear function of the uncertain variables observed in the past. Formally, the set of optimal linear decision rules for (RO) is defined as the set of optimal solutions for

$$\min_{\mathbf{y}_{t1}, \ldots, \mathbf{y}_{tt} \in \mathbb{R}^n : \forall t \in [T]} \; \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} C\left(\sum_{s=1}^{1} \mathbf{y}_{1s} \zeta_s, \ldots, \sum_{s=1}^{T} \mathbf{y}_{Ts} \zeta_s, \, \zeta_1, \ldots, \zeta_T \right). \tag{LDR}$$

Speaking intuitively, the goal of (LDR) is to obtain the best parameters for the linear decision rules, denoted by $\mathbf{y}_{t1},\ldots,\mathbf{y}_{tt} \in \mathbb{R}^n$ for each $t \in [T]$. Given the parameters of the linear decision rules, the decision to make on each stage $t \in [T]$ is computed as $\mathbf{x}_t = \sum_{s=1}^{t} \mathbf{y}_{ts}\zeta_s$. It follows from the above notation that the number of parameters used for representing linear decision rules is given by $\sum_{t=1}^{T} nt = \frac{1}{2}nT(T+1) = \mathcal{O}(nT^2)$.

2.2. The Production-Inventory Problem

Our work focuses on a widely-studied class of production-inventory problems from [4]. The class of problems considers a firm with a central warehouse and E factories that aim to satisfy uncertain demand for a single product over a selling season. The selling season of the firm's product is discretized into T time periods, which are spaced equally over the selling season. In each time period $t \in [T]$, the firm sequentially performs the following three steps:

1. The firm replenishes the inventory level at the central warehouse by producing additional products at their factories. Let $x_{te} \geq 0$ denote the number of product units that the firm decides to produce at each of the factories $e \in [E] \equiv \{1,\ldots,E\}$ at a per-unit cost of $c_{te} \geq 0$. Each factory produces the additional units with zero lead time, and the additional units are stored immediately in the central warehouse.
2. The firm observes the customer demand at the central warehouse. The demand at the central warehouse is denoted by $\zeta_{t+1} \in \mathcal{U}_{t+1} \equiv [\underline{D}_{t+1}, \bar{D}_{t+1}]$, which must be satisfied immediately without backlogging from the inventory in the central warehouse. The lower and upper bounds in the uncertainty set, denoted by $\underline{D}_{t+1} < \bar{D}_{t+1}$, capture the minimum and maximum level of customer demand that the firm anticipates receiving in each time period t.
3. The firm verifies that the remaining inventory in the warehouse lies within a pre-specified interval given by $[V_{\min}, V_{\max}]$. Specifically, the remaining inventory level in the central warehouse at the end of each time period $t \in [T]$ must satisfy
$$V_{\min} \leq v_1 + \sum_{\ell=1}^{t}\sum_{e=1}^{E} x_{\ell e} - \sum_{s=2}^{t+1} \zeta_s \leq V_{\max},$$
where v_1 is the initial inventory level in the central warehouse at the beginning of the selling horizon, $\sum_{\ell=1}^{t}\sum_{e=1}^{E} x_{\ell e}$ is the cumulative number of product units that have been produced at the factories up through time period t, and $\sum_{s=2}^{t+1}\zeta_s$ is the cumulative customer demand that has been observed at the central warehouse up through time period t.

In addition to the satisfying the constraints on the inventory level in the central warehouse at the end of each time period, the firm's production decisions must satisfy $x_{te} \leq p_{te}$ in each time period $t \in [T]$ and each factory $e \in [E]$, where p_{te} is the maximum production level for factory e in time period t, and the firm's total production quantity across the selling season for each factory $e \in [E]$

must satisfy $\sum_{t=1}^{T} x_{te} \le Q_e$, where Q_e is the maximum total production level for factory e. The goal of the firm is to satisfy the customer demand at minimal cost while satisfying production and warehouse constraints.

We observe that the above class of production-inventory problems from Ben-Tal et al. [4] is a special case of (RO) in which the cost function has the form

$$C(\mathbf{x}_1,\dots,\mathbf{x}_T,\zeta_1,\dots,\zeta_{T+1}) = \sum_{t=1}^{T}\sum_{e=1}^{E} c_{te}x_{te} \tag{1a}$$

$$\text{subject to} \quad \sum_{t=1}^{T} x_{te} \le Q_e \qquad \forall e\in[E] \tag{1b}$$

$$0 \le x_{te} \le p_{te} \qquad \forall e\in[E], t\in[T] \tag{1c}$$

$$V_{\min} \le v_1 + \sum_{\ell=1}^{t}\sum_{e=1}^{E} x_{\ell e} - \sum_{s=2}^{t+1}\zeta_s \le V_{\max} \qquad \forall t\in[T] \tag{1d}$$

and where the uncertainty sets have the form $\mathcal{U}_1 \equiv [\underline{D}_1, \bar{D}_1], \dots, \mathcal{U}_{T+1} \equiv [\underline{D}_{T+1}, \bar{D}_{T+1}]$ with $\underline{D}_1 = \bar{D}_1 = 1$ and $\underline{D}_{t+1} < \bar{D}_{t+1}$ for each time period $t\in[T]$. We use the convention that the cost function evaluates to (1a) if the constraints (1b)-(1d) are satisfied and equals infinity otherwise.[1] As a result, optimal linear decision rules for production-inventory problems can be obtain by solving (LDR), which we observe can be written as

$$\begin{aligned}
&\underset{\mathbf{y}_{t,1},\dots,\mathbf{y}_{t,t}\in\mathbb{R}^E:\,\forall t\in[T]}{\text{minimize}} && \max_{\zeta_1\in\mathcal{U}_1,\dots,\zeta_{T+1}\in\mathcal{U}_{T+1}} \left\{ \sum_{t=1}^{T}\sum_{e=1}^{E} c_{te}\left(\sum_{s=1}^{t} y_{t,s,e}\zeta_s\right)\right\} \\
&\text{subject to} && \sum_{t=1}^{T}\left(\sum_{s=1}^{t} y_{t,s,e}\zeta_s\right) \le Q_e \qquad \forall e\in[E] \\
& && 0 \le \left(\sum_{s=1}^{t} y_{t,s,e}\zeta_s\right) \le p_{te} \qquad \forall e\in[E],\ t\in[T] \\
& && V_{\min} \le v_1 + \sum_{\ell=1}^{t}\sum_{e=1}^{E}\left(\sum_{s=1}^{\ell} y_{\ell,s,e}\zeta_s\right) - \sum_{s=2}^{t+1}\zeta_s \le V_{\max} \qquad \forall t\in[T] \\
& && \forall \zeta_1\in\mathcal{U}_1,\dots,\zeta_{T+1}\in\mathcal{U}_{T+1}.
\end{aligned} \tag{LDR-1}$$

3. Main Result

In this section, we present the main result of the paper on the sparsity of optimal linear decision rules for the class of production-inventory problems. Throughout this section, we assume that:

Assumption 1.

a. (LDR) *is feasible and the optimal objective value for* (LDR) *is finite.*

[1] We recall that (RO) and (LDR) involve cost functions in which the number of stages with decisions is equal to the number of stages with uncertain variables. The cost function (1a)-(1d) can be easily modified to match this format by adding a dummy decision variable $\mathbf{x}_{T+1}\in\mathbb{R}^E$ and the constants $p_{T+1,e} = c_{T+1,e} = 0$ for each $e\in[E]$.

b. The uncertainty sets are intervals $\mathcal{U}_t = [\underline{D}_t, \bar{D}_t]$ *with* $\underline{D}_1 = \bar{D}_1 = 1$ *and* $\underline{D}_t < \bar{D}_t$ *for all* $t \geq 2$.

The first assumption states that there exists optimal linear decision rules for the robust optimization problem. The second assumption imposes that the uncertain variables in the robust optimization problem are chosen from interval uncertainty sets. In the production-inventory problem, this second assumption corresponds to the fact that the customer demand is uncertain but bounded in each stage $t \geq 2$. The claim that $\underline{D}_1 = \bar{D}_1 = 1$ ensures without loss of generality that linear decision rules can have a nonzero offset.

In our main result of the paper, presented below as Theorem 1, we establish that there always exists a sparse optimal linear decision rule for the class of production-inventory problems (LDR-1):

THEOREM 1. *Consider a cost function of the form* (1a)-(1d) *and let Assumption 1 hold. Then there exists an optimal solution* $\bar{\mathbf{y}}$ *for* (LDR) *which satisfies* $\|\bar{\mathbf{y}}\|_0 \leq 2 + 8E + 10T + 6ET$.

Recall that the production-inventory problem with linear decision rules (LDR-1) has $\frac{1}{2}ET(T+1) = \mathcal{O}(ET^2)$ parameters. Theorem 1 shows that if (LDR-1) has an optimal solution, then there always exists optimal linear decision rules with $\mathcal{O}(TE)$ nonzero parameters. As far as we are aware, this is the first theoretical result to show the existence of sparse optimal linear decision rules for a practically-important class of adaptive robust optimization problems with many time periods. Furthermore, we notice that the number of parameters for representing static decision rules in production-inventory problems is equal to TE. Theorem 1 essentially shows that the complexity of optimal linear decision rules is at the same level as the complexity of static decision rules, which provides a new perspective for the success of linear decision rule: namely, the fundamental reason that linear decision rules exhibit superior performance to static decision rules is not due to the enormous parameter space of linear decision rules, but rather a very small cardinality of significant parameters that are contained only in the linear decision rules.

In the remainder of this section, we present a high-level roadmap for the proof of Theorem 1. Our proof is based on a new understanding of the extreme points of the feasible regions of linear decision rules for a general class of adaptive robust optimization problems. We postpone the detailed proof of Theorem 1 to Appendix B. Our theorem and proof techniques in this section are not limited to the class of production-inventory problems from Ben-Tal et al. [4], and we present extensions of our results to other robust inventory management problems in §5.

In greater detail, our proof of Theorem 1 focuses on a broad class of adaptive robust optimization problems which includes production-inventory problems as a special case. This class of adaptive robust optimization problems is characterized by cost functions of the form

$$\begin{aligned} C(\mathbf{x}_1,\ldots,\mathbf{x}_T,\zeta_1,\ldots,\zeta_T) = &\sum_{t=1}^{T} \mathbf{a}_{0,t}^\intercal \mathbf{x}_t - \sum_{t=1}^{T} b_{0,t}\zeta_t \\ \text{subject to} \quad &\sum_{t=1}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{x}_t - \sum_{t=1}^{T} b_{i,t}\zeta_t \le c_i \quad \forall i \in [m], \end{aligned} \tag{C-G}$$

where the parameters of the cost function are $\mathbf{a}_{0,t},\ldots,\mathbf{a}_{m,t} \in \mathbb{R}^n$ and $b_{0,t},\ldots,b_{m,t} \in \mathbb{R}$ for each stage $t \in [T]$ and $c_i \in \mathbb{R}$ for each constraint $i \in [m]$. For robust optimization problems (RO) with cost functions of the form (C-G), we observe that (LDR) can be reformulated by its epigraph as

$$\begin{aligned} &\underset{\substack{c_0 \in \mathbb{R} \\ \mathbf{y}_{t,1},\ldots,\mathbf{y}_{t,t} \in \mathbb{R}^n:\, \forall t \in [T]}}{\text{minimize}} && c_0 \\ &\text{subject to} && \max_{\zeta_1 \in \mathcal{U}_1,\ldots,\zeta_T \in \mathcal{U}_T} \left\{ \sum_{t=1}^{T} \mathbf{a}_{i,t}^\intercal \left(\sum_{s=1}^{t} \mathbf{y}_{t,s}\zeta_s \right) - \sum_{t=1}^{T} b_{i,t}\zeta_t \right\} \le c_i \; \forall i \in \{0,\ldots,m\}. \end{aligned} \tag{LDR-G}$$

The decision variables in the above optimization problem include the parameters of the linear decision rules as well as an epigraph decision variable $c_0 \in \mathbb{R}$. For this general class of problems, we will assume in addition to Assumption 1 that:

Assumption 2. *If $C(\mathbf{x}_1,\ldots,\mathbf{x}_T,\zeta_1,\ldots,\zeta_T) < \infty$, then the decisions satisfy $\mathbf{x}_1,\ldots,\mathbf{x}_T \ge \mathbf{0}$.*

The above assumption essentially stipulates that the constraints of (C-G) ensure that feasible decisions for the robust optimization problem satisfy $\mathbf{x}_t \ge \mathbf{0}$ for each stage $t \in [T]$. This is a reasonable assumption because, in practice, the decisions usually refer to levers like order quantities or prices of products that must take nonnegative values. For example, in the class of production-inventory problems from Ben-Tal et al. [4], the decisions $\mathbf{x}_t \in \mathbb{R}^E$ refer to the production quantities at the factories at time period t, and the constraint (1c) guarantees that the production levels are nonnegative.

Equipped with the above notation, for a general class of adaptive robust optimization problems, our proof of Theorem 1 is split into three major steps, organized below as Lemmas 1, 2, and 3. In our first step, denoted below by Lemma 1, we characterize the feasible region of (LDR-G) and show that the set of feasible solutions of (LDR-G) is a polyhedron with at least one extreme point.

Lemma 1. *Let Assumptions 1 and 2 hold. Then the set of feasible solutions to* (LDR-G) *is a nonempty polyhedron with at least one extreme point.*

Let us make two observations about the above lemma. First, since (LDR-G) has a linear objective function and a polyhedral feasible region, we observe that (LDR-G) is a linear optimization problem. Therefore, whenever this linear optimization problem has at least one extreme point and has a finite optimal objective value, there must exist an optimal solution for (LDR-G) that is an extreme point of its feasible set (see, e.g., [8, Theorem 2.7]). Second, it follows from routine arguments in linear programming that if $(\bar{\mathbf{y}}, \bar{c}_0)$ is an extreme point of (LDR-G), then the value of $\bar{c}_0$ is uniquely determined by the value of $\bar{\mathbf{y}}$.[2] Therefore, we will for the sake of simplicity omit $\bar{c}_0$ when referring to an extreme point $(\bar{\mathbf{y}}, \bar{c}_0)$ of the set of feasible solutions to (LDR-G).

In the second step of our proof of Theorem 1, we provide an explicit characterization of the extreme points for the set of feasible solutions of (LDR-G). This key step, which is presented below as Lemma 2, reveals that the extreme points of the set of feasible solutions of (LDR-G) can always be represented as the unique solution of a certain decomposable system of equations.

LEMMA 2. *Let Assumption 1 hold, and let* $(\bar{\mathbf{y}}, \bar{c}_0)$ *be an extreme point of the feasible set of* (LDR-G)*. Then there exists an index set* $\mathcal{I}^{\bar{\mathbf{y}}} \subseteq \{0, \ldots, m\}$*, an index set* $\mathcal{T}_i^{\bar{\mathbf{y}}} \subseteq [T]$ *for each* $i \in \mathcal{I}$*, and a hyperplane* $(\boldsymbol{\alpha}_i^{\bar{\mathbf{y}}}, \beta_i^{\bar{\mathbf{y}}})$ *for each* $i \in \mathcal{I}^{\bar{\mathbf{y}}}$ *such that* $\bar{\mathbf{y}}$ *is the unique solution of the following system:*

$$\sum_{s=1}^{T} \sum_{t=s}^{T} \boldsymbol{\alpha}_{i,t,s}^{\bar{\mathbf{y}}} \cdot \mathbf{y}_{t,s} = \beta_i^{\bar{\mathbf{y}}} \qquad \forall i \in \mathcal{I}^{\bar{\mathbf{y}}}, \tag{HARD}$$

$$\sum_{t=s}^{T} \mathbf{a}_{i,t} \cdot \mathbf{y}_{t,s} = b_{i,s} \qquad \forall i \in \mathcal{I}^{\bar{\mathbf{y}}}, s \in \mathcal{T}_i^{\bar{\mathbf{y}}}. \tag{EASY}$$

Let us provide an interpretation of this second step. In a nutshell, Lemma 2 establishes that every extreme point of the feasible set of (LDR-G) is the unique solution to a linear system that can be decomposed into two types of equations: a small number of hard equations and a large number of easy equations. Indeed, the first type of equations (HARD) is defined by hyperplanes $(\boldsymbol{\alpha}_i^{\bar{\mathbf{y}}}, \beta_i^{\bar{\mathbf{y}}})$ which are functions of $\bar{\mathbf{y}}$, the extreme point of the set of feasible solutions of (LDR-G). We refer to this first type of equations by the moniker (HARD) because the structure of the hyperplanes $(\boldsymbol{\alpha}_i^{\bar{\mathbf{y}}}, \beta_i^{\bar{\mathbf{y}}})$ cannot be analyzed independently of extreme point $\bar{\mathbf{y}}$. In contrast, the second type of equations (EASY) is defined by hyperplanes that are independent of $\bar{\mathbf{y}}$, and so the structure of the second type of equations can be analyzed statically using the structure of the underlying robust optimization problem. The number of equations in (HARD) is at most equal to $m + 1 = \mathcal{O}(m)$, where m is the number of constraints in (C-G), whereas the number of equations in (EASY) is at most equal to $(m+1)T = \mathcal{O}(mT)$. Hence, when the number of stages is large, we observe that there can be significantly more equations of type (EASY) than of type (HARD).

[2] Suppose that $(\bar{\mathbf{y}}, \bar{c}_0)$ is an extreme point of the set of feasible solutions for (LDR-G). Then we readily observe that the value $\bar{c}_0$ must satisfy $\bar{c}_0 = \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} \left\{ \sum_{t=1}^{T} \mathbf{a}_{i,t}^\intercal \left(\sum_{s=1}^{t} \mathbf{y}_{t,s} \zeta_s \right) - \sum_{t=1}^{T} b_{i,t} \zeta_t \right\}$.

The third and final step in our proof of Theorem 1 is to show that the unique solution to the system of equations (HARD)-(EASY) is sparse when (C-G) is equal to the cost function of the production-inventory problem from lines (1a)-(1d). Specifically, it turns out in many practical problems including the production-inventory problem from Ben-Tal et al. [4] that the system of equations (HARD)-(EASY) can be massaged into an instance of a system of equations denoted below (S-1)-(S-3), in which the number of equations in line (S-1) grows linearly in the number of stages of the robust optimization problem. Through this insight, the following Lemma 3 proves that the number of nonzeros in every extreme point of (LDR-G) in problems such as the production-inventory problem from Ben-Tal et al. [4] grows linearly with respect to the number of stages.

LEMMA 3. *Let* $\mathbf{P}_1 \in \mathbb{R}^{m_1 \times n}$, $\mathbf{P}_2 \in \mathbb{R}^{m_2 \times n}$, $\mathbf{q} \in \mathbb{R}^{m_1}$, *and* $\mathcal{A} \subseteq [n]$, *where* $m_1 \le m_2 \le n$. *Suppose that there is a unique* $\bar{\mathbf{z}} \in \mathbb{R}^n$ *that satisfies the system of equations*

$$\mathbf{P}_1 \mathbf{z} = \mathbf{q} \tag{S-1}$$

$$\mathbf{P}_2 \mathbf{z} = \mathbf{0} \tag{S-2}$$

$$z_j = 0 \quad \forall j \in \mathcal{A}, \tag{S-3}$$

and suppose that each column of $\mathbf{P}_2$ *has at most one nonzero entry, that is,* $\sum_{i=1}^{m_2} \mathbb{I}\{p_{2,i,j} \neq 0\} \le 1$ *for each* $j \in [n]$. *Then the unique solution* $\bar{\mathbf{z}}$ *has at most* $2m_1$ *nonzero entries, that is,* $\|\bar{\mathbf{z}}\|_0 \le 2m_1$.

In summary, our proof of Theorem 1 considers a more general class of cost functions (C-G) and looks at the epigraph formulation (LDR-G) of the corresponding problem of computing optimal linear decision rules. Under reasonable assumptions, we show in Lemma 1 that (LDR-G) is indeed a linear optimization problem with at least one extreme point in its feasible region, thus establishing the existence of an optimal extreme point. Furthermore, we show in Lemma 2 that any extreme point of the feasible region is the unique solution to a linear system with two types of equations (HARD) and (EASY), where the number of equations in (HARD) is significantly smaller than the number of equations in (EASY). By utilizing the structure of the production-inventory problem, we conclude by showing in Lemma 3 that the number of nonzeros in every extreme point grows at most linearly with respect to the number of stages of the robust optimization problem. Our formal proof of the theorem is found in Appendix B.

4. Numerical Experiments

To investigate the practical implications of Theorem 1, we perform numerical experiments that generalize those from [4, 12]. Specifically, our numerical experiments focus on instances of (LDR-1) in which the customer demand and production costs of a new product follow a cyclic pattern due to seasonality over a selling horizon of one year. Given a discretization of the selling season into T stages, the customer demand in each stage $t \in \{2,\dots,T+1\}$ is given by

$$\phi_t = 1 + 0.5\sin\left(\frac{2\pi(t-2)}{T}\right), \qquad \theta_t = 0.2, \qquad \mathcal{U}_t = \left[\frac{1000(1-\theta)\phi_t}{T/24}, \frac{1000(1+\theta)\phi_t}{T/24}\right],$$

where parameters ϕ_t and θ_t are interpreted, respectively, as a phase parameter, which captures seasonality, and a demand parameter, which controls the radius of the uncertainty sets. Given E factories available to the firm, the production costs and capacities for each stage $t \in [T]$ and each factory $e \in [E]$ are

$$c_{te} = \left(1 + \frac{e-1}{E-1}\right)\phi_t, \qquad p_{te} = \frac{567}{(T/24)\,(E/3)}, \qquad Q_e = \frac{13600}{E/3},$$

and the capacities and initial inventory at the central warehouse are

$$V_{\min} = 500, \quad V_{\max} = 2000, \quad v_1 = 500.$$

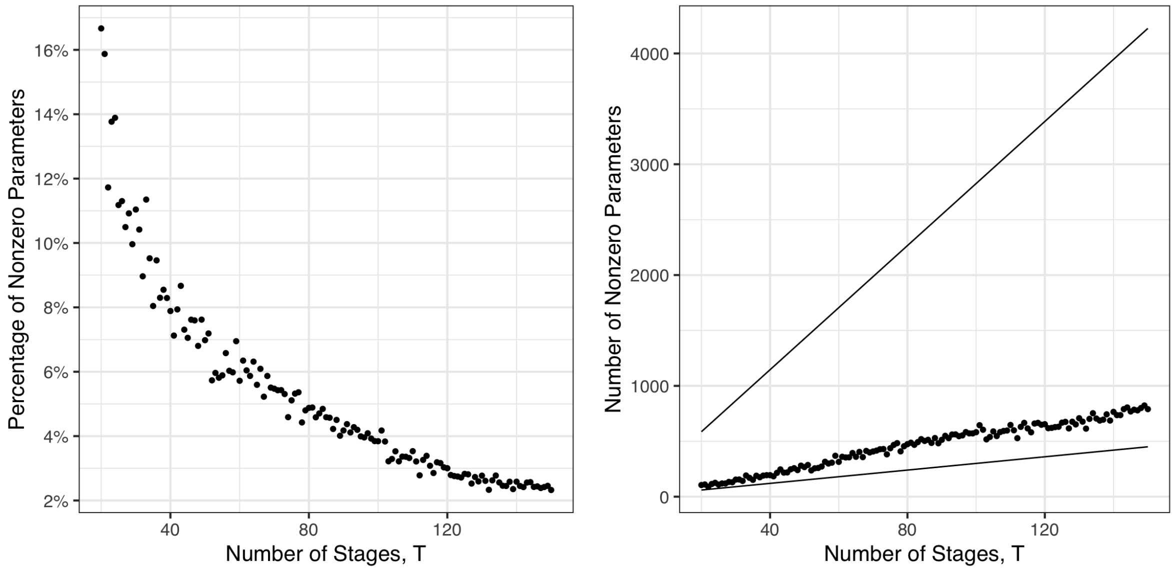
In the case of $T = 24$ and $E = 3$, our parameters are equivalent to those of [4, §5] and [12, Table 1]. We follow this setup from [4, 12] to capture a realistic setting for the parameters of the problem, and our generalization of the setup from the prior work allows us to explore the impact of T on sparsity. We are particularly interested in settings where T grows large, in which case the robust optimization problem serves as an approximation of a continuous-review ordering system. For each value of T and E, we compute the optimal linear decision rules by solving the linear programming formulation from [4, Equation (39)] using primal simplex method. Additional computational results and managerial interpretations of the optimal linear decision rules are presented in Appendix D.

In Figure 1, we present the results of numerical experiments with $E = 3$ factories and varying numbers of periods T. The results of our numerical experiments provide two key takeaways.

Our first takeaway from Figure 1 is that the level of sparsity of the optimal linear decision rules obtained in the numerical experiments is very significant when the number of periods is large. Indeed, the left plot in Figure 1 shows that the number of nonzero parameters in the optimal linear decision rules decreases to 3% of the total number of parameters when inventory levels and production decisions are made twice per week over a selling horizon of one year. From a practical perspective, such a low density level of nonzeros in the optimal linear decision rules provides a clear motivation for the growing stream of research that harnesses sparsity to develop faster and more memory-efficient algorithms in various applications [3, 25, 2, 29]. The level of sparsity also leads to control policies that are nearly as interpretable as static decision rules; see Appendix D for more details.

Our second takeaway from Figure 1 is that Theorem 1 is found to be predictive of the level of sparsity in the optimal linear decision rules. In the right plot of Figure 1, we present the number

Figure 1 Sparsity of optimal linear decision rules for production-inventory problem, $E = 3$.



Note. Each point represents the optimal linear decision rules computed for the corresponding number of stages T and for $E = 3$ factories. Left figure shows the percentage of parameters of optimal linear decision rules which are nonzero. Right figure shows the number of nonzero parameters in optimal linear decision rules compared to the upper bound from Theorem 1 (top solid black line) and the number of parameters in static decision rules (bottom solid black line).

of nonzero parameters in the optimal linear decision rules obtained using primal simplex method, the upper bound from Theorem 1, and the number of parameters in static decision rules (which serves as a natural lower bound). Indeed, the multiplicative gap between the upper bound from Theorem 1 and the lower bound from the static rule is very narrow (it is roughly a factor of 11). As we can see, the numbers of nonzero parameters in the optimal linear decision rules stay in this a narrow band and grow linearly in the number of stages. This result not only validates our theory but also showcases a novel perspective on linear decision rules and static decision rules: while there is an enormous parameter space (with $O(ET^2)$ parameters) for representing linear decision rules, the optimal linear decision rules only need the same order of parameters as static decision rules (which have $O(ET)$ parameters).

5. Extensions

In this section, we show that our proof techniques can be extended to establish sparsity results for other classes of robust optimization problems that have received extensive study in the operations literature. These extensions thus demonstrate that the sparsity results and proof techniques in this paper are not exclusive to the class of production-inventory problems from Ben-Tal et al. [4], and these extensions provide a starting point for using the developments in this paper to establish sparsity guarantees for other classes of robust optimization problems.

5.1. Inventory Management with Lead Times

As discussed in the literature, the class of production-inventory problems admits a number of natural variations. For example, a common variation introduces lead-times to the factories, whereby the production quantity $x_{te} \geq 0$ at factory $e \in [E]$ on period $t \in [T]$ will not become available at the central warehouse until stage $t + \delta_e$, with δ_e denoting the lead time for factory $e \in [E]$. The resulting cost function for this generalization of (1a)-(1d) can thus be written as

$$
\begin{aligned}
C(\mathbf{x}_1, \ldots, \mathbf{x}_T, \zeta_1, \ldots, \zeta_{T+1}) = & \sum_{e=1}^{E} \sum_{t=1}^{T} c_{te} x_{te} && && \text{(2a)} \\
\text{subject to} \quad & \sum_{t=1}^{T} x_{te} \leq Q_e && \forall e \in [E] && \text{(2b)} \\
& 0 \leq x_{te} \leq p_{te} && \forall e \in [E], t \in [T] && \text{(2c)} \\
& V_{\min} \leq v_1 + \sum_{e=1}^{E} \sum_{\ell=1}^{t-\delta_e} x_{\ell e} - \sum_{s=2}^{t+1} \zeta_s \leq V_{\max} && \forall t \in [T]. && \text{(2d)}
\end{aligned}
$$

By applying the proof techniques developed in §3, it can be shown that the $\mathcal{O}(TE)$ sparsity result for optimal linear decision rules is retained:

THEOREM 2. *Consider a cost function of the form (2a)-(2d) and let Assumption 1 hold. Then there exists an optimal solution $\bar{\mathbf{y}}$ for (LDR) that satisfies $\|\bar{\mathbf{y}}\|_0 \leq 2 + 8E + 10T + 6ET$.*

5.2. Dynamic Newsvendor Problems

A significant research effort in the robust optimization literature has been dedicated to dynamic newsvendor problems with interval uncertainty sets; see [5, 9, 23]. This class of dynamic newsvendor problems is characterized by a single factory and nonlinear convex cost functions which capture the holding and backorder costs for inventory at the warehouse. Specifically, the cost function for these dynamic newsvendor problems is given by

$$
C(x_1, \ldots, x_T, \zeta_1, \ldots, \zeta_{T+1}) = \sum_{t=1}^{T} \left(c_t x_t + h_t \left[v_1 + \sum_{s=1}^{t} x_s - \sum_{s=2}^{t+1} \zeta_s \right]^+ + b_t \left[-v_1 - \sum_{s=1}^{t} x_s + \sum_{s=2}^{t+1} \zeta_s \right]^+ \right) \tag{3a}
$$

$$
\text{subject to} \quad 0 \leq x_t \leq p_t \quad \forall t \in [T], \tag{3b}
$$

where the firm begins at the start of the selling season with an initial inventory of v_1 units of product, and, in each stage $t \in [T]$, the firm can decide to produce an additional $x_t \in [0, p_t]$ units of product from a single factory at a cost of c_t per unit. The customer demands for the product are denoted by $\zeta_2, \ldots, \zeta_{T+1} \in \mathbb{R}$, and the holding and backorder costs for the inventory at the end of each period $t \in [T]$ are given by h_t and b_t.

A fundamental result for this class of dynamic newsvendor problems [9, Theorem 3.1] is that if Assumption 1 holds, then the optimal objective values for (RO) and (LDR) with cost function (3a)-(3b) are equal, that is, linear decision rules are optimal control policies. However, with the exception of guarantees on whether of the parameters of optimal linear decision rules are positive or negative ([9, Proposition 5.1], [23, Lemma 5]), the structure of optimal linear decision rules for the production quantities in dynamic newsvendor problems has remained unknown. Using our proof techniques from §3, we establish for the first time that optimal linear decision rules for the class of dynamic newsvendor problems are not only optimal for (RO); they are also sparse.

THEOREM 3. *Consider a cost function of the form* (3a)-(3d) *and let Assumption 1 hold. Then there exists an optimal solution* $\bar{\mathbf{y}}$ *for* (LDR) *that satisfies* $\|\bar{\mathbf{y}}\|_0 \le 10 + 12T$.

6. Conclusion and Future Research

In this paper, we proved that there exist sparse optimal linear decision rules for the widely-studied class of production-inventory problems from Ben-Tal et al. [4], and that the number of nonzero parameters in sparse optimal linear decision rules is at the same level as the number of parameters in static decision rules. We also showed similar results for generalizations of the production-inventory problem as well as for dynamic newsvendor problems. By establishing practical upper bounds on the level of sparsity and by providing an explicit characterization of optimal linear decision rules that are sparse (Lemmas 2 and 3), our work offers new motivations to firms for using linear decision rules in real-world applications and opens up new research directions at the interface between the practice and theory of robust optimization. These research directions include harnessing the structure of sparse optimal linear decision rules to design effective algorithms for computing optimal linear decision rules in real-world large-scale applications, studying the tradeoffs between interpretability of linear decision rules and Pareto efficiency [22, 11], and analyzing the implications of sparsity of linear decision rules on time inconsistency in risk-averse planning problems.

References

[1] Amir Ardestani-Jaafari and Erick Delage. Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research*, 64(2):474–494, 2016.

[2] Amir Ardestani-Jaafari and Erick Delage. The value of flexibility in robust location–transportation problems. *Transportation Science*, 52(1):189–209, 2018.

[3] Chaithanya Bandi, Eojin Han, and Omid Nohadani. Sustainable inventory with robust periodic-affine policies and application to medical supply chains. *Management Science*, 65(10):4636–4655, 2019.

[4] Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[5] Aharon Ben-Tal, Boaz Golany, Arkadi Nemirovski, and Jean-Philippe Vial. Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing & Service Operations Management*, 7(3):248–271, 2005.

[6] Aharon Ben-Tal, Elad Hazan, Tomer Koren, and Shie Mannor. Oracle-based robust optimization via online learning. *Operations Research*, 63(3):628–638, 2015.

[7] Dimitris Bertsimas and Vineet Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming*, 134(2):491–531, 2012.

[8] Dimitris Bertsimas and John N Tsitsiklis. *Introduction to Linear Optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

[9] Dimitris Bertsimas, Dan A Iancu, and Pablo A Parrilo. Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010.

[10] Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[11] Dimitris Bertsimas, Stefan ten Eikelder, Dick den Hertog, and Nikolaos Trichakis. Pareto adaptive robust optimality via a fourier-motzkin elimination lens. *arXiv preprint arXiv:2012.04419*, 2020.

[12] Frans JCT de Ruiter, Ruud Brekelmans, and Dick den Hertog. The impact of the existence of multiple adjustable robust solutions. *Mathematical Programming*, 160(1):531–545, 2016.

[13] Erick Delage and Dan A Iancu. Robust multistage decision making. In *The Operations Research Revolution*, pages 20–46. INFORMS, 2015.

[14] Stefan Eikelder, Ali Ajdari, Thomas Bortfeld, Dick den Hertog, et al. Adjustable robust treatment-length optimization in radiation therapy. *arXiv preprint arXiv:1906.12116*, 2019.

[15] Omar El Housni and Vineet Goyal. On the optimality of affine policies for budgeted uncertainty sets. *Mathematics of Operations Research*, 2021.

[16] Charles Gauvin, Erick Delage, and Michel Gendreau. Decision rule approximations for the risk averse reservoir management problem. *European Journal of Operational Research*, 261(1):317–336, 2017.

[17] Angelos Georghiou, Daniel Kuhn, and Wolfram Wiesemann. The decision rule approach to optimization under uncertainty: methodology and applications. *Computational Management Science*, 16(4):545–576, 2019.

[18] Angelos Georghiou, Angelos Tsoukalas, and Wolfram Wiesemann. Robust dual dynamic programming. *Operations Research*, 67(3):813–830, 2019.

[19] Angelos Georghiou, Angelos Tsoukalas, and Wolfram Wiesemann. On the optimality of affine decision rules in robust and distributionally robust optimization. *Available at Optimization Online*, 2021.

[20] Chrysanthos E Gounaris, Wolfram Wiesemann, and Christodoulos A Floudas. The robust capacitated vehicle routing problem under demand uncertainty. *Operations Research*, 61(3):677–693, 2013.

[21] Nam Ho-Nguyen and Fatma Kılınç-Karzan. Online first-order framework for robust convex optimization. *Operations Research*, 66(6):1670–1692, 2018.

[22] Dan A Iancu and Nikolaos Trichakis. Pareto efficiency in robust optimization. *Management Science*, 60(1):130–147, 2014.

[23] Dan A Iancu, Mayank Sharma, and Maxim Sviridenko. Supermodularity and affine policies in dynamic robust optimization. *Operations Research*, 61(4):941–956, 2013.

[24] Hongming Li, Erick Delage, Ning Zhu, Michael Pinedoc, and Shoufeng Ma. Distributional robustness and inequity mitigation in disaster preparedness of humanitarian operations. *Available at Optimization Online*, 2022.

[25] Alvaro Lorca, X Andy Sun, Eugene Litvinov, and Tongxin Zheng. Multistage adaptive robust optimization for the unit commitment problem. *Operations Research*, 64(1):32–51, 2016.

[26] Krzysztof Postek and Shimrit Shtern. First-order algorithms for robust optimization problems via convex-concave saddle-point lagrangian reformulation. *arXiv preprint arXiv:2101.02669*, 2021.

[27] Sajad Aliakbari Sani, Olivier Bahn, and Erick Delage. Affine decision rule approximation to immunize against demand response uncertainty in smart grids' capacity planning. *European Journal of Operational Research*, 2020.

[28] David Simchi-Levi, Nikolaos Trichakis, and Peter Yun Zhang. Designing response supply chain against bioattacks. *Operations Research*, 67(5):1246–1268, 2019.

[29] Shanshan Wang and Erick Delage. A column generation scheme for distributionally robust multi-item newsvendor problems. *Available at Optimization Online*, 2021.

[30] İhsan Yanıkoğlu, Bram L Gorissen, and Dick den Hertog. A survey of adjustable robust optimization. *European Journal of Operational Research*, 277(3):799–813, 2019.

Technical Proofs and Additional Results

Appendix A: Proofs of Lemmas 1, 2, and 3

Throughout the proofs in Appendix A, we will denote the set of feasible solutions to (LDR-G) by

$$\mathcal{Y} := \left\{ (\mathbf{y}, c_0) : \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} \left\{ \sum_{t=1}^{T} \mathbf{a}_{i,t}^{\intercal} \left(\sum_{s=1}^{t} \mathbf{y}_{t,s} \zeta_s \right) - \sum_{t=1}^{T} b_{i,t} \zeta_t \right\} \le c_i \;\; \forall i \in \{0, \ldots, m\} \right\}.$$

We remark that it follows readily from Assumption 1 that $\mathcal{Y}$ is a nonempty convex polyhedron.

Proof of Lemma 1. Our proof of Lemma 1 is based on contradiction. Indeed, for the sake of developing a contradiction, suppose that the set $\mathcal{Y}$ does not have an extreme point. Since the set $\mathcal{Y}$ is a polyhedron, it follows from the supposition that $\mathcal{Y}$ has no extreme points that $\mathcal{Y}$ must contain a line. In other words, there must exist a solution $(\mathbf{y}^0, c_0^0) \in \mathcal{Y}$ and a direction $(\mathbf{d}, r) \neq (\mathbf{0}, 0)$ such that $(\mathbf{y}^0, c_0^0) + \alpha(\mathbf{d}, r) \in \mathcal{Y}$ for all $\alpha \in \mathbb{R}$.

We begin by showing under the supposition that $\mathcal{Y}$ has no extreme points that $\mathbf{d} \neq \mathbf{0}$. Indeed, we recall from Assumption 1 that the optimal objective value of (LDR) is finite. Moreover, we recall that $(\mathbf{y}^0, c_0^0) + \alpha(\mathbf{d}, r) \in \mathcal{Y}$ for all $\alpha \in \mathbb{R}$. Since the objective value of $(\mathbf{y}^0, c_0^0) + \alpha(\mathbf{d}, r)$ is equal to $c_0^0 + \alpha r$, it must be the case that $r = 0$. Therefore, it follows from the fact that $(\mathbf{d}, r) \neq (\mathbf{0}, 0)$ that $\mathbf{d} \neq \mathbf{0}$.

Next, it follows from Assumption 2 that $(\mathbf{y}^0, c_0^0) + \alpha(\mathbf{d}, r) \in \mathcal{Y}$ for all $\alpha \in \mathbb{R}$ must satisfy

$$\sum_{s=1}^{t} \left(\mathbf{y}_{t,s}^0 + \alpha \mathbf{d}_{t,s} \right) \zeta_s \ge \mathbf{0} \qquad \forall \alpha \in \mathbb{R},\; t \in [T],\; \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T.$$

Since the above inequalities hold for all $\alpha \in \mathbb{R}$, it follows from algebra that

$$\sum_{s=1}^{t} \mathbf{d}_{t,s} \zeta_s = \mathbf{0} \qquad \forall t \in [T],\; \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T. \tag{EC.1}$$

Recall that the uncertainty sets are intervals of the form $\mathcal{U}_1 \triangleq [\underline{D}_1, \bar{D}_1], \ldots, \mathcal{U}_T \triangleq [\underline{D}_T, \bar{D}_T]$, where $\underline{D}_1 = \bar{D}_1 = 1$ and $\underline{D}_t < \bar{D}_t$ for all $t \in \{2, \ldots, T\}$. Therefore, we observe that the equalities on line (EC.1) imply that the equality $\mathbf{d}_{t,s} = \mathbf{0}$ must hold for all $s \in \{2, \ldots, T\}$ and $t \in \{s, \ldots, T\}$. Moreover, it follows from line (EC.1), from the fact that $\underline{D}_1 = \bar{D}_1 = 1$, and from the fact that $\mathbf{d}_{t,s} = \mathbf{0}$ for all $s \in \{2, \ldots, T\}$ and $t \in \{s, \ldots, T\}$ that the equality $\mathbf{d}_{t,1} = -\sum_{s=2}^{t} \mathbf{d}_{t,s} \zeta_s = \mathbf{0}$ must hold for all $t \in \{1, \ldots, T\}$. We have thus shown that $\mathbf{d} = \mathbf{0}$, which contradicts the supposition that the set of optimal solutions $\mathcal{Y}$ has a line. This concludes our proof that $\mathcal{Y}$ has at least one extreme point. $\square$

Proof of Lemma 2. Let $(\bar{\mathbf{y}}, \bar{c}_0)$ denote an extreme point of the set $\mathcal{Y}$ for (LDR-G). We first observe from the definitions of the uncertainty sets $\mathcal{U}_1 = [\underline{D}_1, \bar{D}_1], \ldots, \mathcal{U}_T = [\underline{D}_T, \bar{D}_T]$ and from algebra that $\mathcal{Y}$ can be written equivalently as

$$\begin{aligned}
\mathcal{Y} &= \left\{ (\mathbf{y}, c_0) : \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} \left\{ \sum_{t=1}^{T} \mathbf{a}_{i,t}^{\intercal} \left(\sum_{s=1}^{t} \mathbf{y}_{t,s} \zeta_s \right) - \sum_{t=1}^{T} b_{i,t} \zeta_t \right\} \le c_i \;\; \forall i \in \{0, \ldots, m\} \right\} \\
&= \left\{ (\mathbf{y}, c_0) : \max_{\zeta_1 \in \mathcal{U}_1, \ldots, \zeta_T \in \mathcal{U}_T} \left\{ \sum_{s=1}^{T} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^{\intercal} \mathbf{y}_{t,s} \right) \zeta_s \right\} \le c_i \;\; \forall i \in \{0, \ldots, m\} \right\} \\
&= \left\{ (\mathbf{y}, c_0) : \sum_{s=1}^{T} \max \left\{ \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^{\intercal} \mathbf{y}_{t,s} \right) \underline{D}_s,\; \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^{\intercal} \mathbf{y}_{t,s} \right) \bar{D}_s \right\} \le c_i \;\; \forall i \in \{0, \ldots, m\} \right\} \\
&= \left\{ (\mathbf{y}, c_0) : \sum_{s=1}^{T} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^{\intercal} \mathbf{y}_{t,s} \right) D_s^* \le c_i \;\; \forall D_s^* \in \left\{ \underline{D}_s, \bar{D}_s \right\},\; s \in [T],\; \text{and } i \in \{0, \ldots, m\} \right\}.
\end{aligned}$$

Since $(\bar{\mathbf{y}}, \bar{c}_0)$ is an extreme point of the set $\mathcal{Y}$, and since the set $\mathcal{Y}$ is a polyhedron, we observe that $(\bar{\mathbf{y}}, \bar{c}_0)$ is a basic feasible solution of $\mathcal{Y}$. In other words, $(\bar{\mathbf{y}}, \bar{c}_0)$ must be the unique solution to the system of constraints in $\mathcal{Y}$ which are active constraints at $(\bar{\mathbf{y}}, \bar{c}_0)$. Let $\mathcal{I}$ denote the subset of $\{0, \ldots, m\}$ corresponding to the active constraints at the basic feasible solution $(\bar{\mathbf{y}}, \bar{c}_0)$, that is, let

$$\mathcal{I} \triangleq \{0\} \cup \left\{ i : \text{there exists } D_s^* \in \left\{ \underline{D}_s, \bar{D}_s \right\} \text{ for each stage } s \in [T] \text{ such that } \sum_{s=1}^{T} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} \right) D_s^* = c_i \right\}.$$

Since $(\bar{\mathbf{y}}, \bar{c}_0)$ is an element of $\mathcal{Y}$, we observe for each $i \in \mathcal{I}$ that the equality

$$\sum_{s=1}^{T} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} \right) D_s^* = \begin{cases} c_i, & \text{if } i \neq 0, \\ \bar{c}_0, & \text{if } i = 0 \end{cases}$$

is satisfied whenever the following equality holds for each $s \in [T]$:

$$D_s^* = \begin{cases} \bar{D}_s, & \text{if } -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} > 0, \\ \underline{D}_s, & \text{if } -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} < 0, \\ \underline{D}_s \text{ or } \bar{D}_s & \text{if } -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} = 0. \end{cases}$$

Therefore, we conclude from the above observation that the set of active constraints at the basic feasible solution $(\bar{\mathbf{y}}, \bar{c}_0)$ is given by the system of equalities

$$\begin{aligned} &\sum_{s \in \mathcal{T}_i^{>}} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) \bar{D}_s + \sum_{s \in \mathcal{T}_i^{<}} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) \underline{D}_s + \sum_{s \in \mathcal{T}_i^{=}} \left(-b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} \right) D_s^* = c_i \\ &\quad \forall i \in \mathcal{I} \text{ and } D_s^* \in \left\{ \underline{D}_s, \bar{D}_s \right\} \forall s \in \mathcal{T}_i^{=}, \end{aligned} \tag{EC.2}$$

where we define the disjoint index sets $\mathcal{T}_i^{>}$, $\mathcal{T}_i^{=}$, $\mathcal{T}_i^{<}$ for each $i \in \mathcal{I}$ as

$$\mathcal{T}_i^{>} \triangleq \left\{ s : -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} > 0 \right\}, \quad \mathcal{T}_i^{=} \triangleq \left\{ s : -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} = 0 \right\},$$
$$\mathcal{T}_i^{<} \triangleq \left\{ s : -b_{i,s} + \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \bar{\mathbf{y}}_{t,s} < 0 \right\}.$$

We observe that the system of equalities (EC.2) can be rewritten as

$$\begin{aligned} \sum_{s \in \mathcal{T}_i^{>}} \left(-b_{0,s} + \sum_{t=s}^{T} \mathbf{a}_{0,t}^\intercal \mathbf{y}_{t,s} \right) \bar{D}_s + \sum_{s \in \mathcal{T}_i^{<}} \left(-b_{0,s} + \sum_{t=s}^{T} \mathbf{a}_{0,t}^\intercal \mathbf{y}_{t,s} \right) \underline{D}_s &= c_i && \forall i \in \mathcal{I} \\ \sum_{t=s}^{T} \mathbf{a}_{i,t}^\intercal \mathbf{y}_{t,s} &= b_{i,s} && \forall i \in \mathcal{I}, s \in \mathcal{T}_i^{=}, \end{aligned}$$

which concludes our proof of Lemma 2. □

Proof of Lemma 3. The overarching idea of the proof of Lemma 3 is that the equations $\mathbf{P}_2 \mathbf{z} = \mathbf{0}$ from (S-2) can be eliminated by eliminating the corresponding variables.

We begin by making several assumptions without loss of generality. First, we recall that $\mathbf{P}_2$ is a matrix with more columns than rows. We henceforth assume without loss of generality that $\mathbf{P}_2$ has linearly independent rows (otherwise we may remove the rows which are linearly dependent without changing the set of feasible solutions to the system of equations (S-1)-(S-3)). We also assume for each row $i \in [m_2]$ of the matrix $\mathbf{P}_2$ that there exists a column $j \in [n] \setminus \mathcal{A}$ such that $p_{2,i,j} \neq 0$ (otherwise we may remove row i without changing the set of feasible solutions to the system of equations (S-1)-(S-3)).

We will also use the following notation. For each row $i \in [m_2]$ of the matrix $\mathbf{P}_2$, let $j_i \in [n] \setminus \mathcal{A}$ denote the smallest column for which $p_{2,i,j} \neq 0$, and let $\mathcal{J}_i \triangleq \{j \in [n] \setminus \{j_i\} : p_{2,i,j} \neq 0\}$ denote the remaining columns for which the i-th row of $\mathbf{P}_2$ has nonzero entries. Finally, let $\mathcal{J} \triangleq \{j_1, \ldots, j_{m_2}\}$, in which case it follows from the assumption that $\sum_{i=1}^{m_2} \mathbb{I}\{p_{2,i,j} \neq 0\} \leq 1$ for each $j \in [n]$ that $\mathcal{J}, \mathcal{J}_1, \ldots, \mathcal{J}_{m_2}$ are disjoint sets.

We now perform a substitution of variables to eliminate the constraints (S-2) from the system of equations. Specifically, by performing the substitution $z_{j_i} = -\sum_{j \in \mathcal{J}_i} z_j$ for each $i \in [m_2]$, we conclude that $(\bar{z}_j : j \in [n] \setminus \mathcal{J})$ is the unique solution to the following system of equations:

$$\sum_{j \in [n] \setminus \mathcal{J}} \mathbf{P}_{1,j} z_j + \sum_{i=1}^{m_2} \mathbf{P}_{1,j_i} \left(-\sum_{j \in \mathcal{J}_i} z_j \right) = \mathbf{q}$$
$$z_j = 0 \quad \forall j \in \mathcal{A}.$$

Since $(\bar{z}_j : j \in [n] \setminus \mathcal{J})$ is the unique solution to the above system, and since $\mathbf{q} \in \mathbb{R}^{m_1}$, we conclude that $(\bar{z}_j : j \in [n] \setminus \mathcal{J})$ must have at most m_1 nonzero entries. Moreover, it follows from the fact that $\mathcal{J}, \mathcal{J}_1, \ldots, \mathcal{J}_{m_2}$ are disjoint sets and the fact that $z_{j_i} = -\sum_{j \in \mathcal{J}_i} z_j$ for each $i \in [m_2]$ that at most m_1 of the entries $z_{j_1}, \ldots, z_{j_{m_2}}$ are nonzero. This concludes our proof that $\|\bar{\mathbf{z}}\|_0 \leq 2m_1$. □

Appendix B: Proofs of Theorem 1 and Theorem 2

We observe that Theorem 1 is a special case of Theorem 2 from §5 in which $\delta_e = 0$ for all factories $e \in [E]$. Therefore, it suffices to present the proof of Theorem 2.

Proof of Theorem 2. Our proof is split into three steps, corresponding to Lemmas 1, 2, and 3.

Step 1: We begin in the first step of our proof of Theorem 2 by applying Lemma 1 for the epigraph formulation of (LDR) for cost functions given by (2a)-(2d). To do this, we first rewrite the cost function from lines (2a)-(2d) to match the format used by Lemmas 1 and 2. Indeed, we recall that Lemmas 1 and 2 involve cost functions in which the number of stages with decisions is equal to the number of stages with uncertain variables. Thus, by introducing dummy decision variables $\mathbf{x}_{T+1} \in \mathbb{R}^E$, we observe that the cost function on lines (2a)-(2d) can be equivalently written as

$$C(\mathbf{x}_1, \ldots, \mathbf{x}_{T+1}, \zeta_1, \ldots, \zeta_{T+1}) = \sum_{e=1}^{E} \sum_{t=1}^{T+1} c_{te} x_{te} \tag{2a}$$
$$\text{subject to} \quad \sum_{t=1}^{T+1} x_{te} \leq Q_e \qquad \forall e \in [E] \tag{2b}$$
$$0 \leq x_{te} \leq p_{te} \qquad \forall e \in [E], t \in [T+1] \tag{2c}$$
$$V_{\min} \leq v_1 + \sum_{e=1}^{E} \sum_{\ell=1}^{t-\delta_e} x_{\ell e} - \sum_{s=2}^{t+1} \zeta_s \leq V_{\max} \qquad \forall t \in [T], \tag{2d}$$

where we define $p_{T+1,e} \equiv 0$ and $c_{T+1,e} \equiv 0$ for each factory $e \in [E]$. After adding the dummy decision variables $\mathbf{x}_{T+1}$, we observe that the above cost function matches the format used by Lemmas 1 and 2.

Before proceeding onward, let us make two brief remarks about the convention used in the above cost function. First, we remark that the constraint (2c) implies that the dummy decision variables $\mathbf{x}_{T+1}$ will always be equal to zero. Second, we remark that the decisions x_{te} for each $t \geq T - \delta_e + 1$ are unnecessary in the above formulation, in the sense that these decisions could always be set to zero without loss of generality.

To see why the decisions are unnecessary, we notice that the decision x_{te} for each $t \geq T - \delta_e + 1$ does not appear in constraint (2d); therefore, since $c_{te} \geq 0$, we observe that x_{te} can at optimality take its minimum value that is allowed by constraints (2b) and (2c). Although they are not necessary, the decisions x_{te} for each $t \geq T - \delta_e + 1$ are included in the above cost function to simplify the notation in the rest of the proof.

We next derive the optimization problem (LDR) corresponding to the above cost function (2a)-(2d). For the sake of clarity, we will show each of the intermediary algebra steps in this derivation. Indeed, we first observe from algebra that line (2a) can be written with linear decision rules as

$$\sum_{e=1}^{E}\sum_{t=1}^{T+1} c_{te}x_{te} = \sum_{e=1}^{E}\sum_{t=1}^{T+1} c_{te}\left(\sum_{s=1}^{t} y_{t,s,e}\zeta_s\right) = \sum_{s=1}^{T+1}\left(\sum_{t=s}^{T+1}\sum_{e=1}^{E} c_{te}y_{t,s,e}\right)\zeta_s,$$

where the first equality comes from using linear decision rules and the second equality follows from algebra. We observe that the left-hand side of line (2b) for each $e \in [E]$ can be written with linear decision rules as

$$\sum_{t=1}^{T+1} x_{te} = \sum_{t=1}^{T+1}\left(\sum_{s=1}^{t} y_{t,s,e}\zeta_s\right) = \sum_{s=1}^{T+1}\left(\sum_{t=s}^{T+1} y_{t,s,e}\right)\zeta_s,$$

where the first equality comes from using linear decision rules and the second equality follows from algebra. We observe that the decision in line (2c) for each $e \in [E]$ and $t \in [T+1]$ can be written with linear decision rules as

$$x_{te} = \sum_{s=1}^{t} y_{t,s,e}\zeta_s + \sum_{s=t+1}^{T+1} 0\zeta_s,$$

where the equality comes from using linear decision rules. Finally, we observe that the inventory in line (2d) for each $t \in [T]$ can be written with linear decision rules as

$$\begin{aligned}
& v_1 + \sum_{e=1}^{E}\sum_{\ell=1}^{t-\delta_e} x_{\ell e} - \sum_{s=2}^{t+1}\zeta_s \\
&= v_1 + \sum_{e=1}^{E}\sum_{\ell=1}^{t-\delta_e}\left(\sum_{s=1}^{\ell} y_{\ell,s,e}\zeta_s\right) - \sum_{s=2}^{t+1}\zeta_s \\
&= v_1 + \sum_{s=1}^{t}\left(\sum_{\ell=s}^{t}\sum_{e\in[E]:\delta_e\leq t-\ell} y_{\ell,s,e}\right)\zeta_s - \sum_{s=2}^{t+1}\zeta_s \\
&= v_1 + \left(\sum_{\ell=1}^{t}\sum_{e\in[E]:\delta_e\leq t-\ell} y_{\ell,1,e}\right)\zeta_1 + \sum_{s=2}^{t}\left(-1 + \sum_{\ell=s}^{t}\sum_{e\in[E]:\delta_e\leq t-\ell} y_{\ell,s,e}\right)\zeta_s - \zeta_{t+1} + \sum_{s=t+2}^{T+1} 0\zeta_s,
\end{aligned}$$

where the first equality comes from using linear decision rules and the second and third equalities follow from algebra. Combining the above steps, we conclude that the epigraph form of the optimization problem (LDR) with cost function (2a)-(2d) is equivalent to

$$\begin{aligned}
&\underset{\substack{c_0\in\mathbb{R}\\ \mathbf{y}_{t,1},\dots,\mathbf{y}_{t,t}\in\mathbb{R}^E:\,\forall t\in[T+1]}}{\text{minimize}} && c_0\\
&\text{subject to} && \sum_{s=1}^{T+1}\left(\sum_{t=s}^{T+1}\sum_{e=1}^{E} c_{te}y_{t,s,e}\right)\zeta_s \le c_0\\
&&& \sum_{s=1}^{T+1}\left(\sum_{t=s}^{T+1} y_{t,s,e}\right)\zeta_s \le Q_e \qquad \forall e\in[E]\\
&&& 0\le \sum_{s=1}^{t} y_{t,s,e}\zeta_s + \sum_{s=t+1}^{T+1} 0\zeta_s \le p_{te} \quad \forall e\in[E],\, t\in[T+1]\\
&&& V_{\min}\le v_1 + \left(\sum_{\ell=1}^{t}\sum_{e\in[E]:\delta_e\le t-\ell} y_{\ell,1,e}\right)\zeta_1\\
&&& \qquad + \sum_{s=2}^{t}\left(-1+\sum_{\ell=s}^{t}\sum_{e\in[E]:\delta_e\le t-\ell} y_{\ell,s,e}\right)\zeta_s - \zeta_{t+1} + \sum_{s=t+2}^{T+1} 0\zeta_s \le V_{\max} \quad \forall t\in[T]\\
&&& \forall \zeta_1\in\mathcal{U}_1,\dots,\zeta_{T+1}\in\mathcal{U}_{T+1}.
\end{aligned} \tag{LDR-2}$$

With the above notation, we are now ready to invoke Lemma 1. Indeed, we recall from the statement of Theorem 2 that Assumption 1 holds for cost function (2a)-(2d). Furthermore, the constraint in (2c) guarantees Assumption 2 holds. Therefore, it follows from Lemma 1 that the set of feasible solutions to (LDR-2) is a nonempty polyhedron with at least one extreme point.

Step 2: In the second step of our proof of Theorem 2, we use Lemma 2 to characterize the structure of extreme points for the feasible set of (LDR-2). Indeed, let $(\bar{\mathbf{y}},\bar{c}_0)$ denote an extreme point of the set of feasible solutions of (LDR-2). Since Assumption 1 holds, it follows from Lemma 2 that there exists

- index sets $\mathcal{I}^{2b,\bar{\mathbf{y}}}\subseteq[E]$, $\underline{\mathcal{I}}^{2c,\bar{\mathbf{y}}},\bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}}\subseteq[T+1]\times[E]$, and $\underline{\mathcal{I}}^{2d,\bar{\mathbf{y}}},\bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}}\subseteq[T]$;
- index sets $\mathcal{T}^{2a,\bar{\mathbf{y}}}\subseteq[T+1]$, $\mathcal{T}_e^{2b,\bar{\mathbf{y}}}\subseteq[T+1]$ for each $e\in\mathcal{I}^{2b,\bar{\mathbf{y}}}$, $\underline{\mathcal{T}}_{t,e}^{2c,\bar{\mathbf{y}}}\subseteq[T+1]$ for each $(t,e)\in\underline{\mathcal{I}}^{2c,\bar{\mathbf{y}}}$, $\bar{\mathcal{T}}_{t,e}^{2c,\bar{\mathbf{y}}}\subseteq[T+1]$ for each $(t,e)\in\bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}}$, $\underline{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}}\subseteq[T+1]$ for each $t\in\underline{\mathcal{I}}^{2d,\bar{\mathbf{y}}}$, and $\bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}}\subseteq[T+1]$ for each $t\in\bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}}$;
- hyperplanes $(\boldsymbol{\alpha}^{2a,\bar{\mathbf{y}}},\beta^{2a,\bar{\mathbf{y}}})$, $(\boldsymbol{\alpha}_e^{2b,\bar{\mathbf{y}}},\beta_e^{2b,\bar{\mathbf{y}}})$ for each $e\in\mathcal{I}^{2b,\bar{\mathbf{y}}}$, $(\underline{\boldsymbol{\alpha}}_{t,e}^{2c,\bar{\mathbf{y}}},\underline{\beta}_{t,e}^{2c,\bar{\mathbf{y}}})$ for each $(t,e)\in\underline{\mathcal{I}}^{2c,\bar{\mathbf{y}}}$, $(\bar{\boldsymbol{\alpha}}_{t,e}^{2c,\bar{\mathbf{y}}},\bar{\beta}_{t,e}^{2c,\bar{\mathbf{y}}})$ for each $(t,e)\in\bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}}$, $(\underline{\boldsymbol{\alpha}}_t^{2d,\bar{\mathbf{y}}},\underline{\beta}_t^{2d,\bar{\mathbf{y}}})$ for each $t\in\underline{\mathcal{I}}^{2d,\bar{\mathbf{y}}}$, and $(\bar{\boldsymbol{\alpha}}_t^{2d,\bar{\mathbf{y}}},\bar{\beta}_t^{2d,\bar{\mathbf{y}}})$ for each $t\in\bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}}$

such that $\bar{\mathbf{y}}$ is the unique solution to the following system of equations:

$$\boldsymbol{\alpha}^{2a,\bar{\mathbf{y}}}\cdot\mathbf{y}=\beta^{2a,\bar{\mathbf{y}}} \tag{HARD-2a}$$

$$\boldsymbol{\alpha}_e^{2b,\bar{\mathbf{y}}}\cdot\mathbf{y}=\beta_e^{2b,\bar{\mathbf{y}}} \qquad \forall e\in\mathcal{I}^{2b,\bar{\mathbf{y}}} \tag{HARD-2b}$$

$$\underline{\boldsymbol{\alpha}}_{t,e}^{2c,\bar{\mathbf{y}}}\cdot\mathbf{y}=\underline{\beta}_{t,e}^{2c,\bar{\mathbf{y}}} \qquad \forall (t,e)\in\underline{\mathcal{I}}^{2c,\bar{\mathbf{y}}} \tag{HARD-2c-LB}$$

$$\bar{\boldsymbol{\alpha}}_{t,e}^{2c,\bar{\mathbf{y}}}\cdot\mathbf{y}=\bar{\beta}_{t,e}^{2c,\bar{\mathbf{y}}} \qquad \forall (t,e)\in\bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}} \tag{HARD-2c-UB}$$

$$\underline{\boldsymbol{\alpha}}_t^{2d,\bar{\mathbf{y}}}\cdot\mathbf{y}=\underline{\beta}_t^{2d,\bar{\mathbf{y}}} \qquad \forall t\in\underline{\mathcal{I}}^{2d,\bar{\mathbf{y}}} \tag{HARD-2d-LB}$$

$$\bar{\boldsymbol{\alpha}}_t^{2d,\bar{\mathbf{y}}}\cdot\mathbf{y}=\bar{\beta}_t^{2d,\bar{\mathbf{y}}} \qquad \forall t\in\bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}} \tag{HARD-2d-UB}$$

$$\sum_{t=s}^{T+1}\sum_{e=1}^{E} c_{te}y_{t,s,e}=0 \qquad \forall s\in\mathcal{T}^{2a,\bar{\mathbf{y}}} \tag{EASY-2a}$$

$$\sum_{t=s}^{T+1} y_{t,s,e} = 0 \qquad \forall e \in [E],\ s \in \mathcal{T}_e^{2b,\bar{\mathbf{y}}} \tag{EASY-2b}$$

$$y_{t,s,e} = 0 \qquad \forall e \in [E],\ t \in [T+1],\ s \in \mathcal{T}_{t,e}^{2c,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_{t,e}^{2c,\bar{\mathbf{y}}} \text{ if } s \le t \tag{EASY-2c-i}$$

$$0 = 0 \qquad \forall e \in [E],\ t \in [T+1],\ s \in \mathcal{T}_{t,e}^{2c,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_{t,e}^{2c,\bar{\mathbf{y}}} \text{ if } s \ge t+1 \tag{EASY-2c-ii}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \le t-\ell} y_{\ell,s,e} = 0 \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \text{ if } s = 1 \tag{EASY-2d-i}$$

$$\sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \le t-\ell} y_{\ell,s,e} = 1 \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \text{ if } s \in \{2,\dots,t\} \tag{EASY-2d-ii}$$

$$0 = 1 \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \text{ if } s = t+1 \tag{EASY-2d-iii}$$

$$0 = 0 \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \text{ if } s \ge t+2. \tag{EASY-2d-iv}$$

Let us make two observations about the above system of equations. First, we recall from its construction that there is exactly one solution to the above system, namely, $\bar{\mathbf{y}}$. Therefore, we readily observe that there must be no constraints of the form (EASY-2d-iii), i.e., the inequality $s \le t$ must hold for all $t \in [T]$ and all $s \in \mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}}$. Second, we observe that we can remove the constraints of type (EASY-2c-ii) and (EASY-2d-iv) without loss of generality. Hence, we will drop the constraints (EASY-2d-iii), (EASY-2c-ii), and (EASY-2d-iv) in our subsequent analysis.

Step 3: In the third step of our proof of Theorem 2, we use Lemma 3 to show that every extreme point $(\bar{\mathbf{y}}, \bar{c}_0)$ of the feasible set of (LDR-2) is sparse.

To motivate our usage of Lemma 3, let us make several observations about the system of equations corresponding to the extreme point $(\bar{\mathbf{y}}, \bar{c}_0)$. First, we observe that there are $\mathcal{O}(TE)$ equations in each of the lines (HARD-2a), (HARD-2b), (HARD-2c-LB), (HARD-2c-UB), (HARD-2d-LB), (HARD-2d-UB), (EASY-2a), (EASY-2b), and (EASY-2d-i). Second, we recall from our discussion at the end of Step 2 that the equations in lines (EASY-2c-ii), (EASY-2d-iii) and (EASY-2d-iv) can be dropped without loss of generality. Third, we observe that the equations in line (EASY-2c-i) will ultimately impose sparsity into the solution the system of equations. Finally, we observe that there are up to $\mathcal{O}(ET^2)$ equations on line (EASY-2d-ii).

With the goal of applying Lemma 3 to the above system of equations, we perform algebraic manipulations on the equations in (EASY-2d-ii). Indeed, we first define the following index sets:

$$\mathscr{S}^{\bar{\mathbf{y}}} \triangleq \bigcup_{t=1}^{T} \left(\left(\mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \right) \cap \{2,\dots,t\} \right),$$
$$\mathscr{T}_s^{\bar{\mathbf{y}}} \triangleq \left\{ t \in [T] : s \in \left(\mathcal{T}_t^{2d,\bar{\mathbf{y}}} \cup \bar{\mathcal{T}}_t^{2d,\bar{\mathbf{y}}} \right) \cap \{2,\dots,t\} \right\} \quad \forall s \in \mathscr{S}^{\bar{\mathbf{y}}}.$$

With the above notation, we readily observe that (EASY-2d-ii) can be written equivalently as

$$\sum_{\ell=s}^{t} \sum_{e \in [E]: \delta_e \le t-\ell} y_{\ell,s,e} = 1 \quad \forall s \in \mathscr{S}^{\bar{\mathbf{y}}},\ t \in \mathscr{T}_s^{\bar{\mathbf{y}}}. \tag{EASY-2d-ii}$$

For each $s \in \mathscr{S}^{\bar{\mathbf{y}}}$, let the elements of $\mathscr{T}_s^{\bar{\mathbf{y}}}$ be indexed in ascending order by $t_{s,1}^{\bar{\mathbf{y}}} < \ldots < t_{s,|\mathscr{T}_s^{\bar{\mathbf{y}}}|}^{\bar{\mathbf{y}}}$. With this notation, we observe that (EASY-2d-ii) can be written equivalently as the following system of equations:

$$\sum_{\ell=s}^{t_{s,1}^{\bar{\mathbf{y}}}} \sum_{e \in [E]: \delta_e \le t_{s,1}^{\bar{\mathbf{y}}} - \ell} y_{\ell,s,e} = 1 \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}}} \tag{EASY-2d-ii'}$$

$$\sum_{\ell=s}^{t^{\bar{\mathbf{y}}}_{s,k+1}} \sum_{e\in[E]:\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} - \sum_{\ell=s}^{t^{\bar{\mathbf{y}}}_{s,k}} \sum_{e\in[E]:\delta_e\le t^{\bar{\mathbf{y}}}_{s,k}-\ell} y_{\ell,s,e} = 1-1=0 \qquad \forall s\in\mathscr{S}^{\bar{\mathbf{y}}},\ k\in\{1,\dots,|\mathscr{T}^{\bar{\mathbf{y}}}_s|-1\}. \tag{EASY-2d-ii”}$$

In particular, it follows from algebra that (EASY-2d-ii”) is equivalent to

$$\sum_{\ell=t^{\bar{\mathbf{y}}}_{s,k}+1}^{t^{\bar{\mathbf{y}}}_{s,k+1}} \sum_{e\in[E]:\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} + \sum_{\ell=s}^{t^{\bar{\mathbf{y}}}_{s,k}} \sum_{e\in[E]:t^{\bar{\mathbf{y}}}_{s,k}-\ell+1\le\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} = 0 \quad \forall s\in\mathscr{S}^{\bar{\mathbf{y}}},\ k\in\{1,\dots,|\mathscr{T}^{\bar{\mathbf{y}}}_s|-1\}. \tag{EASY-2d-ii”}$$

To simplify our notation, we now compactly represent the constraints from lines (HARD-2a), (HARD-2b), (HARD-2c-LB), (HARD-2c-UB) (HARD-2d-LB), (HARD-2d-UB), (EASY-2a), (EASY-2b), (EASY-2d-i) and (EASY-2d-ii’) using the index set $\mathscr{I}^{\bar{\mathbf{y}}}$ and hyperplanes $(\boldsymbol{\alpha}^{\bar{\mathbf{y}}}_i,\beta^{\bar{\mathbf{y}}}_i)$ for each $i\in\mathscr{I}^{\bar{\mathbf{y}}}$, where

$$\begin{aligned}
|\mathscr{I}^{\bar{\mathbf{y}}}| &= \underbrace{1}_{\text{(HARD-2a)}} + \underbrace{|\mathcal{I}^{2b,\bar{\mathbf{y}}}|}_{\text{(HARD-2b)}} + \underbrace{|\mathcal{I}^{2c,\bar{\mathbf{y}}}|}_{\text{(HARD-2c-LB)}} + \underbrace{|\bar{\mathcal{I}}^{2c,\bar{\mathbf{y}}}|}_{\text{(HARD-2c-UB)}} + \underbrace{|\mathcal{I}^{2d,\bar{\mathbf{y}}}|}_{\text{(HARD-2d-LB)}} + \underbrace{|\bar{\mathcal{I}}^{2d,\bar{\mathbf{y}}}|}_{\text{(HARD-2d-UB)}} \\
&\quad + \underbrace{|\mathcal{T}^{2a,\bar{\mathbf{y}}}|}_{\text{(EASY-2a)}} + \underbrace{\sum_{e=1}^{E}|\mathcal{T}^{2b,\bar{\mathbf{y}}}_e|}_{\text{(EASY-2b)}} + \underbrace{\left|\left\{t\in[T]:\ 1\in\mathcal{T}^{2d,\bar{\mathbf{y}}}_t\cup\bar{\mathcal{T}}^{2d,\bar{\mathbf{y}}}_t\right\}\right|}_{\text{(EASY-2d-i)}} + \underbrace{|\mathscr{S}^{\bar{\mathbf{y}}}|}_{\text{(EASY-2d-ii’)}} \\
&\le \underbrace{1}_{\text{(HARD-2a)}} + \underbrace{E}_{\text{(HARD-2b)}} + \underbrace{(T+1)E}_{\text{(HARD-2c-LB)}} + \underbrace{(T+1)E}_{\text{(HARD-2c-UB)}} + \underbrace{T}_{\text{(HARD-2d-LB)}} + \underbrace{T}_{\text{(HARD-2d-UB)}} \\
&\quad + \underbrace{T+1}_{\text{(EASY-2a)}} + \underbrace{(T+1)E}_{\text{(EASY-2b)}} + \underbrace{T}_{\text{(EASY-2d-i)}} + \underbrace{T-1}_{\text{(EASY-2d-ii’)}} \\
&= 1+4E+5T+3ET.
\end{aligned}$$

With the above notation, we have shown that $\bar{\mathbf{y}}$ is the unique solution to the following system of equations:

$$\boldsymbol{\alpha}^{\bar{\mathbf{y}}}_i\cdot\mathbf{y}=\beta^{\bar{\mathbf{y}}}_i \quad \forall i\in\mathscr{I}^{\bar{\mathbf{y}}} \tag{HARD-combined}$$

$$\sum_{\ell=t^{\bar{\mathbf{y}}}_{s,k}+1}^{t^{\bar{\mathbf{y}}}_{s,k+1}} \sum_{e\in[E]:\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} + \sum_{\ell=s}^{t^{\bar{\mathbf{y}}}_{s,k}} \sum_{e\in[E]:t^{\bar{\mathbf{y}}}_{s,k}-\ell+1\le\delta_e\le t^{\bar{\mathbf{y}}}_{s,k+1}-\ell} y_{\ell,s,e} = 0 \quad \forall s\in\mathscr{S}^{\bar{\mathbf{y}}},\ k\in\{1,\dots,|\mathscr{T}^{\bar{\mathbf{y}}}_s|-1\}. \tag{EASY-2d-ii”}$$

$$y_{t,s,e}=0 \quad \forall e\in[E],\ t\in[T+1],\ s\in\mathcal{T}^{2c,\bar{\mathbf{y}}}_{t,e}\cup\bar{\mathcal{T}}^{2c,\bar{\mathbf{y}}}_{t,e} \text{ if } s\le t \tag{EASY-2c-i}$$

We next apply Lemma 3 to the above system of equations by noticing (HARD-combined), (EASY-2d-ii”), and (EASY-2c-i) follow the same structure of (S-1), (S-2) and (S-3), respectively. Furthermore, we notice that there is no overlapping index of y in (EASY-2d-ii”), thus each column of the corresponding $\mathbf{P}_2$ has at most one nonzero entry. Since $\bar{\mathbf{y}}$ is the unique solution to the above system of equations, it follows from Lemma 3 that the number of nonzero entries in $\bar{\mathbf{y}}$ satisfies

$$\|\bar{\mathbf{y}}\|_0 \le 2\,|\mathcal{I}| \le 2\,(1+4E+5T+3ET),$$

Since the above inequality holds for every extreme point $(\bar{\mathbf{y}},\bar{c}_0)$, since we have proven that (LDR-2) has at least one extreme point, and since we observe from Assumption 1 that (LDR-2) has an optimal solution which is an extreme point, our proof of Theorem 2 is complete. □

Appendix C: Proofs of Theorem 3

Proof of Theorem 3. Our proof of Theorem 3 follows a similar organization to that of Theorem 2. Specifically, the proof of Theorem 2 is split into three steps which correspond to Lemmas 1, 2, and 3. In contrast to the proof of Theorem 2, the proof of Theorem 3 requires the introduction of auxiliary decision rules to account for nonlinear cost functions.

Step 1: We begin in the first step of our proof of Theorem 3 by applying Lemma 1 to an epigraph formulation of (LDR) for cost function (3a)-(3b). Indeed, we observe that (RO) with cost function (3a)-(3b) can be reformulated as

$$
\begin{aligned}
\underset{\mathbf{x},\mathbf{z}}{\text{minimize}} \quad & \max_{\zeta_1 \in \mathcal{U}_1,\ldots,\zeta_{T+1}\in\mathcal{U}_{T+1}} \left\{ \sum_{t=1}^{T+1} c_t x_t(\zeta_1,\ldots,\zeta_t) + \sum_{t=1}^{T+1} z_t(\zeta_1,\ldots,\zeta_t) \right\} \\
\text{subject to} \quad & z_{t+1}(\zeta_1,\ldots,\zeta_{t+1}) \ge h_t\left(v_1 + \sum_{s=1}^{t} x_s(\zeta_1,\ldots,\zeta_s) - \sum_{s=2}^{t+1} \zeta_s \right) && \forall t \in [T] \\
& z_{t+1}(\zeta_1,\ldots,\zeta_{t+1}) \ge -b_t\left(v_1 + \sum_{s=1}^{t} x_s(\zeta_1,\ldots,\zeta_s) - \sum_{s=2}^{t+1} \zeta_s \right) && \forall t \in [T] \\
& 0 \le x_t(\zeta_1,\ldots,\zeta_t) \le p_t && \forall t \in [T+1] \\
& z_1(\zeta_1) \ge 0 \\
& \quad \forall \zeta_1 \in \mathcal{U}_1,\ldots,\zeta_{T+1} \in \mathcal{U}_{T+1},
\end{aligned}
$$

where the auxiliary decision rule $z_{t+1}(\zeta_1,\ldots,\zeta_{t+1})$ captures the holding and backorder costs in each period $t \in [T]$, $p_{T+1} = 0$, $c_{T+1} = 1$, and $x_{T+1}(\zeta_1,\ldots,\zeta_{T+1})$ and $z_1(\zeta_1)$ are dummy decision rules that can always be identically equal to zero at optimality. We note that the dummy decision rules $x_{T+1}(\zeta_1,\ldots,\zeta_{T+1})$ and $z_1(\zeta_1)$ have been introduced into the above optimization problem to match the setting of Lemmas 1 and 2, in which the number of decisions is constant in each stage (in this case, the number of decisions in each stage is $n = 2$). We also readily observe from inspection that any feasible solution of the above optimization problem will satisfy $x_t(\zeta_1,\ldots,\zeta_t), z_t(\zeta_1,\ldots,\zeta_t) \ge 0$ for all stages $t \in [T+1]$ and all realizations $\zeta_1 \in \mathcal{U}_1,\ldots,\zeta_{T+1} \in \mathcal{U}_{T+1}$.

It follows from [9, Theorem 3.1] and Assumption 1 that the auxiliary decision rules $z_t(\zeta_1,\ldots,\zeta_t)$ and the production decision rules $x_t(\zeta_1,\ldots,\zeta_t)$ for all $t \in [T+1]$ in the above optimization problem can be replaced with linear decision rules without any loss of optimality. Hence, we conclude that the optimization problems (RO) and (LDR) with cost function (3a)-(3b) are equivalent to one another and can be written as

$$
\begin{aligned}
\underset{\substack{y_{t,1},\ldots,y_{t,t}\in\mathbb{R}\ \forall t\in[T+1] \\ w_{t,1},\ldots,w_{t,t}\in\mathbb{R}\ \forall t\in[T+1]}}{\text{minimize}} \quad & \max_{\zeta_1 \in \mathcal{U}_1,\ldots,\zeta_{T+1}\in\mathcal{U}_{T+1}} \left\{ \sum_{t=1}^{T+1} c_t \left(\sum_{s=1}^{t} y_{t,s}\zeta_s \right) + \sum_{t=1}^{T+1} \left(\sum_{s=1}^{t} w_{t,s}\zeta_s \right) \right\} \\
\text{subject to} \quad & \sum_{s=1}^{t+1} w_{t+1,s}\zeta_s \ge h_t \left(v_1 + \sum_{\ell=1}^{t} \left(\sum_{s=1}^{\ell} y_{\ell,s}\zeta_s \right) - \sum_{s=2}^{t+1} \zeta_s \right) && \forall t \in [T] \\
& \sum_{s=1}^{t+1} w_{t+1,s}\zeta_s \ge -b_t \left(v_1 + \sum_{\ell=1}^{t} \left(\sum_{s=1}^{\ell} y_{\ell,s}\zeta_s \right) - \sum_{s=2}^{t+1} \zeta_s \right) && \forall t \in [T] \\
& 0 \le \sum_{s=1}^{t} y_{t,s}\zeta_s \le p_t && \forall t \in [T+1] \\
& w_{1,1}\zeta_1 \ge 0 \\
& \quad \forall \zeta_1 \in \mathcal{U}_1,\ldots,\zeta_{T+1} \in \mathcal{U}_{T+1}.
\end{aligned}
$$

By rearranging terms in the above optimization problem, by adding an epigraph decision variable $c_0 \in \mathbb{R}$, and by noticing that $\sum_{\ell=t+1}^{t} h_t y_{\ell,s} = -\sum_{\ell=t+1}^{t} b_t y_{\ell,s} = 0$ for each period $t \in [T]$, we observe that the above optimization problem can be written equivalently as

$$
\begin{aligned}
\underset{\substack{c_0 \in \mathbb{R} \\ y_{t,1},\ldots,y_{t,t} \in \mathbb{R}\ \forall t \in [T+1] \\ w_{t,1},\ldots,w_{t,t} \in \mathbb{R}\ \forall t \in [T+1]}}{\text{minimize}} \quad & c_0 \\
\text{subject to} \quad & \sum_{s=1}^{T+1} \left(\sum_{t=s}^{T+1} (c_t y_{t,s} + w_{t,s}) \right) \zeta_s \le c_0 && && \text{(3a)} \\
& \sum_{s=1}^{t+1} \left(\sum_{\ell=s}^{t} h_t y_{\ell,s} - h_t - w_{t+1,s} \right) \zeta_s + \sum_{s=t+2}^{T+1} 0 \zeta_s \le -h_t v_1 && \forall t \in [T] && \text{(3b)} \\
& \sum_{s=1}^{t+1} \left(\sum_{\ell=s}^{t} -b_t y_{\ell,s} + b_t - w_{t+1,s} \right) \zeta_s + \sum_{s=t+2}^{T+1} 0 \zeta_s \le b_t v_1 && \forall t \in [T] && \text{(3c)} \\
& 0 \le \sum_{s=1}^{t} y_{t,s} \zeta_s + \sum_{s=t+1}^{T+1} 0 \zeta_s \le p_t && \forall t \in [T+1] && \text{(3d)} \\
& \sum_{s=1}^{t} w_{t,s} \zeta_s + \sum_{s=t+1}^{T+1} 0 \zeta_s \ge 0 && \forall t \in [1] && \text{(3e)} \\
& \quad \forall \zeta_1 \in \mathcal{U}_1, \ldots, \zeta_{T+1} \in \mathcal{U}_{T+1}.
\end{aligned}
$$

With the above notation, we are now ready to invoke Lemma 1. Indeed, we recall from the statement of Theorem 3 that Assumption 1 holds for cost function (3a)-(3b). Therefore, it follows from the above reasoning, the constraints in (3b)-(3e), and Lemma 1 that the set of feasible solutions to the above optimization problem is a nonempty polyhedron with at least one extreme point.

Step 2: In the second step of our proof of Theorem 3, we use Lemma 2 to characterize the structure of extreme points for the feasible set of the above optimization problem. Indeed, let $(\bar{\mathbf{y}}, \bar{\mathbf{w}}, \bar{c}_0)$ denote an extreme point of the set of feasible solutions of the above optimization problem. Then it follows readily from Lemma 2 that there exists

- index sets $\mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T]$, $\underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$, and $\mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [1]$;
- index sets $\mathcal{T}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$, $\mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $\underline{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $\bar{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, and $\mathcal{T}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \subseteq [T+1]$ for each $t \in \mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}$;
- hyperplanes $(\boldsymbol{\alpha}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}})$, $(\boldsymbol{\alpha}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $(\boldsymbol{\alpha}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $(\underline{\boldsymbol{\alpha}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \underline{\boldsymbol{\gamma}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \underline{\beta}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, $(\bar{\boldsymbol{\alpha}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \bar{\boldsymbol{\gamma}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \bar{\beta}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}$, and $(\boldsymbol{\alpha}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $t \in \mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}$

such that $(\bar{\mathbf{y}}, \bar{\mathbf{w}})$ is the unique solution to the following system of equalities:

$$\boldsymbol{\alpha}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \boldsymbol{\gamma}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \beta^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3a}$$

$$\boldsymbol{\alpha}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \boldsymbol{\gamma}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \beta_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3b}$$

$$\boldsymbol{\alpha}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \boldsymbol{\gamma}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \beta_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3c}$$

$$\bar{\boldsymbol{\alpha}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \bar{\boldsymbol{\gamma}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \bar{\beta}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3d-UB}$$

$$\underline{\boldsymbol{\alpha}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \underline{\boldsymbol{\gamma}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \underline{\beta}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3d-LB}$$

$$\boldsymbol{\alpha}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \boldsymbol{\gamma}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{w} = \beta_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \qquad \forall t \in \mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{HARD-3e}$$

$$\sum_{t=s}^{T+1} (c_t y_{t,s} + w_{t,s}) = 0 \qquad \forall s \in \mathcal{T}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{EASY-3a}$$

$$\sum_{\ell=s}^{t} h_t y_{\ell,s} - h_t - w_{t+1,s} = 0 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t+1 \tag{EASY-3b-i}$$

$$0 = 0 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \ge t+2 \tag{EASY-3b-ii}$$

$$-\sum_{\ell=s}^{t} b_t y_{\ell,s} + b_t - w_{t+1,s} = 0 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t+1 \tag{EASY-3c-i}$$

$$0 = 0 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \ge t+2 \tag{EASY-3c-ii}$$

$$y_{t,s} = 0 \qquad \forall t \in [T+1],\ s \in \bar{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \underline{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t \tag{EASY-3d-i}$$

$$0 = 0 \qquad \forall t \in [T+1],\ s \in \bar{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \underline{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \ge t+1 \tag{EASY-3d-ii}$$

$$w_{t,s} = 0 \qquad \forall t \in [1], s \in \mathcal{T}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t \tag{EASY-3e-i}$$

$$0 = 0 \qquad \forall t \in [1], s \in \mathcal{T}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \ge t+1. \tag{EASY-3e-ii}$$

Step 3: In the third step of our proof of Theorem 3, we apply Lemma 3 to every extreme point $(\bar{\mathbf{y}}, \bar{\mathbf{w}}, \bar{c}_0)$. With the goal of Lemma 3 in mind, we first observe from (EASY-3b-i) and (EASY-3c-i) that any solution to the above system of equations must satisfy the following equalities:

$$w_{t+1,s} = \sum_{\ell=s}^{t} h_t y_{\ell,s} - h_t \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t+1$$

$$w_{t+1,s} = -\sum_{\ell=s}^{t} b_t y_{\ell,s} + b_t \qquad \forall t \in [T],\ s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t+1.$$

We can thus eliminate the variables $w_{t+1,s}$ for each $t \in [T]$ and $s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}$ with $s \le t+1$ in the above system of equations by the substitution

$$w_{t+1,s} \leftarrow \begin{cases} \sum_{\ell=s}^{t} h_t y_{\ell,s} - h_t, & \text{if } t \in [T] \text{ and } s \in \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}, \\ -\sum_{\ell=s}^{t} b_t y_{\ell,s} + b_t, & \text{if } t \in [T] \text{ and } s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \setminus \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}. \end{cases}$$

With this substitution, we observe that (EASY-3b-i) and (EASY-3c-i) can be replaced with

$$\sum_{\ell=s}^{t} y_{\ell,s} = 1 \qquad \forall t \in [T], s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}},\ s \le t+1 \tag{EASY-3bc-i}$$

We next perform algebraic manipulations on the equations in (EASY-3bc-i). Indeed, we first define the following sets:

$$\mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \triangleq \bigcup_{t=1}^{T} \left(\left(\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \right) \cap \{1, \ldots, t+1\} \right),$$

$$\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \triangleq \left\{ t \in [T] : s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \{1, \ldots, t+1\} \right\} \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}.$$

With the above notation, we observe that the constraints (EASY-3bc-i) can be rewritten as

$$\sum_{\ell=s}^{t} y_{\ell,s} = 1 \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}},\ t \in \mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}.$$

Moreover, let the elements of $\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$ be denoted by $t_{s,1}^{\bar{\mathbf{y}},\bar{\mathbf{w}}} < \ldots < t_{s,|\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}|}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$. With this notation, we readily observe using algebra that the constraints (EASY-3bc-i) can be replaced with the following equivalent constraints:

$$\sum_{\ell=s}^{t_{s,1}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}} y_{\ell,s} = 1 \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \tag{EASY-3bc-i'}$$

$$\sum_{\ell=t_{s,k}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}+1}^{t_{s,k+1}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}} y_{\ell,s} = 0 \qquad \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}},\ k \in \{1,\ldots,|\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}|-1\}. \tag{EASY-3bc-i''}$$

To simplify our notation, we now compactly represent the constraints from lines (HARD-3a), (HARD-3b), (HARD-3c), (HARD-3d-UB), (HARD-3d-LB), (HARD-3e), (EASY-3a), and (EASY-3bc-i') using the index set $\mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$ and hyperplanes $(\boldsymbol{\alpha}_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}}, \boldsymbol{\gamma}_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}}, \beta_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}})$ for each $i \in \mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$[3], where

$$\begin{aligned}
|\mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}| &= \underbrace{1}_{\text{(HARD-3a)}} + \underbrace{|\mathcal{I}^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3b)}} + \underbrace{|\mathcal{I}^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3c)}} + \underbrace{|\bar{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3d-UB)}} + \underbrace{|\underline{\mathcal{I}}^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3d-LB)}} + \underbrace{|\mathcal{I}^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(HARD-3e)}} + \underbrace{|\mathcal{T}^{3a,\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(EASY-3a)}} + \underbrace{|\mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}|}_{\text{(EASY-3bc-i')}} \\
&\le \underbrace{1}_{\text{(HARD-3a)}} + \underbrace{T}_{\text{(HARD-3b)}} + \underbrace{T}_{\text{(HARD-3c)}} + \underbrace{T+1}_{\text{(HARD-3d-UB)}} + \underbrace{T+1}_{\text{(HARD-3d-LB)}} + \underbrace{1}_{\text{(HARD-3e)}} + \underbrace{T+1}_{\text{(EASY-3a)}} + \underbrace{T}_{\text{(EASY-3bc-i')}} \\
&\le 5 + 6T.
\end{aligned}$$

It follows from the above notation, from the fact that we have used substitution to eliminate the variable $w_{t+1,s}$ for each $t \in [T]$ and $s \in \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}$ with $s \le t+1$, and from the fact that (EASY-3b-ii), (EASY-3c-ii), (EASY-3d-ii), and (EASY-3e-ii) can be eliminated without loss of generality that $\left(\bar{\mathbf{y}}, \left(\bar{w}_{t+1,s} : t \in [T],\ s \notin \left(\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}\right) \cap \{1,\ldots,t+1\}\right)\right)$ is the unique solution to the following system of equations:

$$\begin{aligned}
&\boldsymbol{\alpha}_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \cdot \mathbf{y} + \sum_{t=1}^{T} \sum_{s \notin \left(\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cap \{1,\ldots,t+1\}\right)} \gamma_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}} w_{t+1,s} = \beta_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}} && \forall i \in \mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}} \\
&y_{t,s} = 0 && \forall t \in [T+1],\ s \in \bar{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \underline{\mathcal{T}}_t^{3d,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t \qquad \text{(EASY-3d-i)} \\
&w_{1,1} = 0 && \forall t \in [1], s \in \mathcal{T}_t^{3e,\bar{\mathbf{y}},\bar{\mathbf{w}}} \text{ if } s \le t \qquad \text{(EASY-3e-i)} \\
&\sum_{\ell=t_{s,k}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}+1}^{t_{s,k+1}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}} y_{\ell,s} = 0 && \forall s \in \mathscr{S}^{\bar{\mathbf{y}},\bar{\mathbf{w}}},\ k \in \{1,\ldots,|\mathscr{T}_s^{\bar{\mathbf{y}},\bar{\mathbf{w}}}|-1\}.
\end{aligned} \tag{EASY-3bc''}$$

Notice that any $y_{\ell,s}$ only appear once in (EASY-3bc''). It thus follows from Lemma 3 that the number of nonzero entries in $\bar{\mathbf{y}}$ satisfies

$$\|\bar{\mathbf{y}}\|_0 \le \left\|\left(\bar{\mathbf{y}}, \left(\bar{w}_{t+1,s} : t \in [T],\ s \notin \left(\mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}}\right) \cap \{1,\ldots,t+1\}\right)\right)\right\|_0 \le 2\,|\mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}| \le 10 + 12T,$$

which concludes our proof of Theorem 3. □

[3] The vector $\boldsymbol{\gamma}_i^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$ in each hyperplane $i \in \mathscr{I}^{\bar{\mathbf{y}},\bar{\mathbf{w}}}$ contains an element corresponding to the variable $w_{t+1,s}$ for each $t \in [T]$ and $s \notin \mathcal{T}_t^{3b,\bar{\mathbf{y}},\bar{\mathbf{w}}} \cup \mathcal{T}_t^{3c,\bar{\mathbf{y}},\bar{\mathbf{w}}}$.

Appendix D: Additional Numerical Results

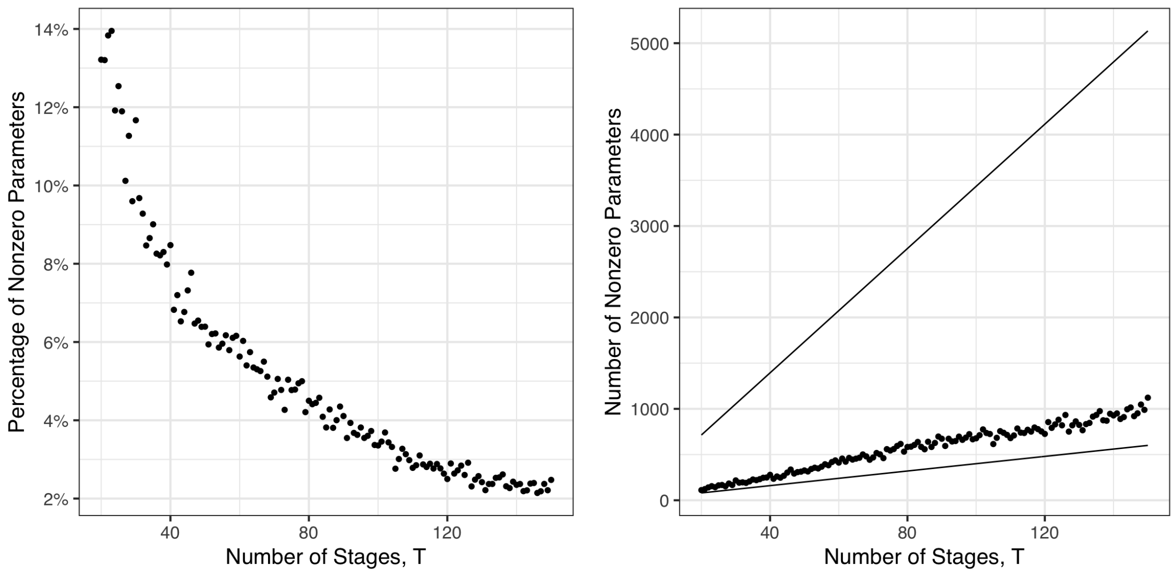
We begin by showing an example of optimal linear decision rules for the production-inventory problem as described in §4 with the number of periods is $T = 24$ and and the number of factories is $E = 2$:

$$
\begin{aligned}
x_{1,1} &= 850.5 & x_{1,2} &= 850.5 \\
x_{2,1} &= 850.5 & x_{2,2} &= 255.083 + 0.496\zeta_2 \\
x_{3,1} &= 850.5 & x_{3,2} &= -1511.555 + 0.504\zeta_2 + 1.262\zeta_3 \\
x_{4,1} &= 850.5 & x_{4,2} &= -413.185 - 0.262\zeta_3 + \zeta_4 \\
x_{5,1} &= 850.5 & x_{5,2} &= -786.933 + \zeta_5 \\
x_{6,1} &= 850.5 & x_{6,2} &= -869.115 + \zeta_6 \\
x_{7,1} &= 850.5 & x_{7,2} &= -396.863 + 0.335\zeta_7 \\
x_{8,1} &= 850.5 & x_{8,2} &= -949.5 + \zeta_8 \\
x_{9,1} &= 850.5 & x_{9,2} &= -929.055 + \zeta_9 \\
x_{10,1} &= 850.5 & x_{10,2} &= -1547.528 + 0.338\zeta_7 + \zeta_{10} \\
x_{11,1} &= 267.915 + 0.327\zeta_7 & x_{11,2} &= -974.764 + \zeta_{11} \\
x_{12,1} &= 850.5 & x_{12,2} &= -995.209 + \zeta_{12} \\
x_{13,1} &= 543.709 + 0.226\zeta_{13} & x_{13,2} &= -699 + 0.774\zeta_{13} \\
x_{14,1} &= 850.5 & x_{14,2} &= -388.417 + 0.486\zeta_{14} \\
x_{15,1} &= 850.5 & x_{15,2} &= 49.5 \\
x_{16,1} &= -1094.209 + \zeta_{15} + \zeta_{16} & x_{16,2} &= 0 \\
x_{17,1} &= 74.764 + \zeta_{17} & x_{17,2} &= 0 \\
x_{18,1} &= 850.5 & x_{18,2} &= 0 \\
x_{19,1} &= 655.992 + 0.286\zeta_{18} & x_{19,2} &= 0 \\
x_{20,1} &= -460.111 + 0.514\zeta_{14} + 0.714\zeta_{18} + 0.334\zeta_{19} & x_{20,2} &= 0 \\
x_{21,1} &= -340.23 + 0.666\zeta_{19} + \zeta_{20} + 0.286\zeta_{21} & x_{21,2} &= 0 \\
x_{22,1} &= -1701 + 0.955\zeta_{11} + 0.515\zeta_{21} + \zeta_{22} & x_{22,2} &= 0 \\
x_{23,1} &= -688.706 + 0.833\zeta_2 + 0.558\zeta_{23} & x_{23,2} &= 2551.5 - 0.833\zeta_2 - 0.955\zeta_{11} \\
x_{24,1} &= 0 & x_{24,2} &= 0.
\end{aligned}
$$

Because the optimal linear decision rules are sparse, we can draw a number of managerial insights regarding these control policies. Indeed, we recall that the maximum production level at each factory $e \in \{1, 2\}$ in each period $t \in [24]$ is equal to $p_{te} = 567/((T/24) \times (E/3)) = 850.5$. We thus observe in the first half of the selling season that the optimal linear decision rules selects the maximum production level from the first factory and accounts for variability in customer demand using the second factory. In the second half of the selling season, the optimal linear decision rules stop using the second factory and account for variability in customer demand using the first factory. More broadly, the above optimal linear decision rules exemplify that sparsity leads to linear decision rules which retain interpretability akin to that of static decision rules.

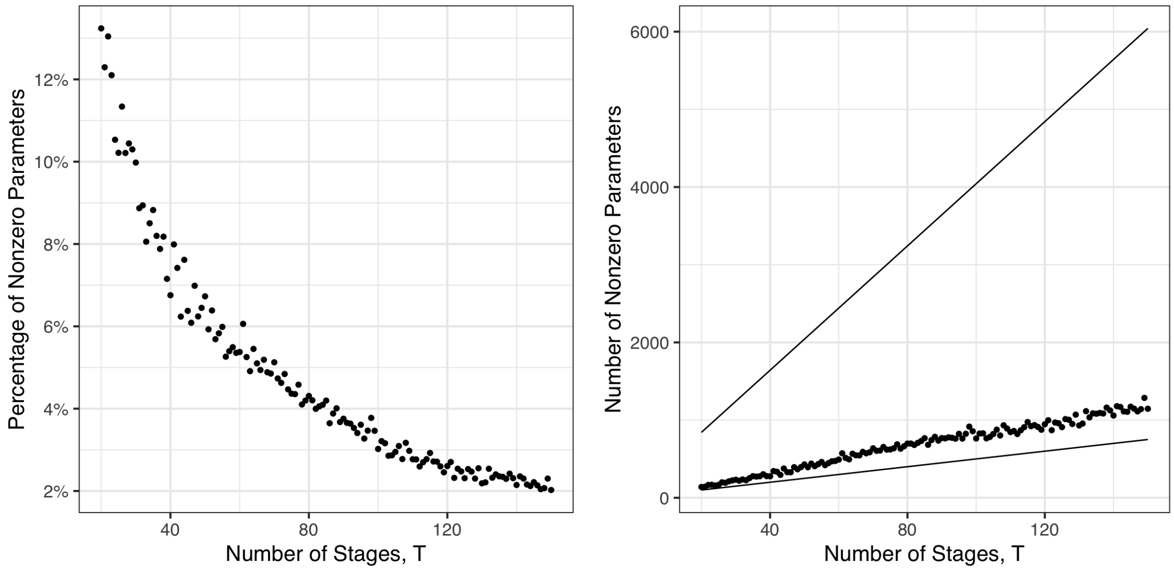
We conclude Appendix D by presenting additional numerical results to those from Figure 1 for when the number of factories is $E = 4$ and $E = 5$. The results, shown in Figures EC.1 and EC.2, show that the findings from Figure 1 are not exclusive to the case with $E = 3$ factories.

Figure EC.1 Sparsity of optimal linear decision rules for production-inventory problem, $E = 4$.



Note. Each point represents the optimal linear decision rules computed for the corresponding number of stages T and for $E = 4$ factories. Left figure shows the percentage of parameters of optimal linear decision rules which are nonzero. Right figure shows the number of nonzero parameters in optimal linear decision rules compared to the upper bound from Theorem 1 (top solid black line) and the number of parameters in static decision rules (bottom solid black line).

Figure EC.2 Sparsity of optimal linear decision rules for production-inventory problem, $E = 5$.



Note. Each point represents the optimal linear decision rules computed for the corresponding number of stages T and for $E = 5$ factories. Left figure shows the percentage of parameters of optimal linear decision rules which are nonzero. Right figure shows the number of nonzero parameters in optimal linear decision rules compared to the upper bound from Theorem 1 (top solid black line) and the number of parameters in static decision rules (bottom solid black line).