European Journal of Operational Research 309 (2023) 102–116

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing, Transportation and Logistics

Data-driven remanufacturing planning with parameter uncertainty

Zhicheng Zhu[a], Yisha Xiang[b,*], Ming Zhao[c], Yue Shi[d]

[a] Department of Industrial, Manufacturing, and Systems Engineering, Texas Tech University, Lubbock, TX 79409, USA
[b] Department of Industrial Engineering, University of Houston, Houston, TX 77004, USA
[c] Department of Business Administration, University of Delaware, Newark, DE 19716, USA
[d] Department of Management Science & Engineering, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Article history:
Received 29 November 2021
Accepted 16 January 2023
Available online 19 January 2023

Keywords:
Maintenance
Remanufacturing planning
Robust Markov decision process
Control-limit policy

ABSTRACT

We consider the problem of remanufacturing planning in the presence of statistical estimation errors. Determining the optimal remanufacturing timing, first and foremost, requires modeling of the state transitions of a system. The estimation of these probabilities, however, often suffers from data inadequacy and is far from accurate, resulting in serious degradation in performance. To mitigate the impacts of the uncertainty in transition probabilities, we develop a novel data-driven modeling framework for remanufacturing planning in which decision makers can remain robust with respect to statistical estimation errors. We model the remanufacturing planning problem as a robust Markov decision process, and construct ambiguity sets that contain the true transition probabilities with high confidence. We further establish structural properties of optimal robust policies and provide insights for remanufacturing planning. A computational study on the NASA turbofan engine shows that our data-driven robust decision framework consistently yields better out-of-sample reward and higher reliability of the performance guarantee, compared to the nominal model that uses the maximum likelihood estimates of the transition probabilities without considering parameter uncertainty.

1. Introduction

The manufacturing industry is a major consumer of materials and energy and imposes a significant impact on environment. Sustainable manufacturing with improved environmental performance has drawn great attentions from governments, companies and scientific communities. In the past decade, remanufacturing has emerged as one of the critical elements for developing a sustainable manufacturing industry (Ijomah et al., 2007). Remanufacturing is an overhaul process whereby used or broken-down products, referred to as “cores”, are restored to a like-new condition with an extended lifetime (Östlin et al., 2009). During this process, the cores pass through a number of operations including inspection, dismantling, part reprocessing, repair, replacement and reassembly. The performance of the remanufactured cores is expected to meet the desired product standards similar to the original product, but is not considered a new product in its first life.

Comparing to manufacturing a new product, remanufacturing can reduce up to 80% of energy consumption and carbon dioxide emissions (Sutherland et al., 2008), and 40–65% of manufacturing costs (Ford & Despeisse, 2016). Remanufacturing is being practiced across various sectors like automotive, aerospace, electrical and electronic equipment (EEE), medical equipment, and machinery (Russell & Nasr, 2019; Yang, 2020; Zhang et al., 2021). However, the growth of the remanufacturing industry faces several critical challenges. One major challenge faced by remanufacturers is managing the inherent uncertainty in cores’ conditions (Örsdemir et al., 2014; Yang et al., 2020), which is largely attributed to the current *reactive* end-of-life remanufacturing approach. Many cores collected at the end of a product’s life are no longer remanufaturable due to the lack of adequate technologies to restore them to like-new conditions. To overcome this barrier, much attention has been received in designing optimal acquisition decisions such as acquiring more cores than the demand or purchase cores in sorted grades, which allows a remanufacturer to be more selective and remanufacture only those items that are in the best condition (Galbreth & Blackburn, 2010; Örsdemir et al., 2014). While these acquisition strategies may work well for electronic and electrical equipment including consumer electronics, ink and toner cartridges, and white goods, it is much less applicable to several major sectors of remanufacturing such as aerospace, heavy duty and off-road (HDOR) equipment, where bulk purchase is rarely an option.

Further exacerbating the issues brought by cores’ uncertain conditions is that contrary to the conventional wisdom that

* Corresponding author.
E-mail addresses: zhicheng.zhu@ttu.edu (Z. Zhu), yxiang4@uh.edu (Y. Xiang), mzhao@udel.edu (M. Zhao), yue.shi.mse@whu.edu.cn (Y. Shi).

https://doi.org/10.1016/j.ejor.2023.01.031
0377-2217/

remanufacturing reduces environmental impacts, it can, in fact, lead to negative outcomes due to heavy damage. Several studies have shown that in some cases, remanufacturing actually consumes more energy than manufacturing a new product (Chandler, 2011; Gutowski et al., 2011). A natural question that arises is: Can we identify the optimal timing for remanufacturing prior to the product's life end when it is still remanufacturable and worth the effort? The focus of this paper is to investigate a *proactive* remanufacturing planning policy that is more viable for remanufacturing of large, capital equipment.

Two critical enablers of proactive remanufacturing for equipment in industries such as aerospace, HDOR, are: condition monitoring technologies and service-based contracts. Due to the mission-critical and capital-intensive nature of these equipment, they are monitored by various sensors and their conditions can be assessed by analyzing the collected sensor data. Moreover, manufacturers of equipment in these sectors have been increasingly offering service-based purchasing agreements. Through these agreements, the manufacturers have access to the status of the product, and can determine when to remanufacture equipment rather than wait until the product fails. An example of such agreement is the *GoldCare* provided by Boeing, which is an integrated service providing asset management, engineering, maintenance and support for airline customers (Parker et al., 2015).

In this paper, we provide a novel data-driven modeling framework for remanufacturing planning. In particular, we address the robustness of the planning decision threatened by the inherent data inadequacy in sensor data. The optimal planning decision involves suggesting the optimal action, such as no intervention, remanufacturing, or scrapping, at different system states, and therefore it is required to first and foremost estimate the transition dynamics of a system. The underlying transition probabilities (sometimes referred to as the true transition probabilities), which govern the condition evolving process of a system, are typically unknown and need to be estimated from data. The estimation is typically subject to large statistical errors due to noises and incorrect information contained in the sensor data. This data deficiency poses a critical question to decision makers: How does uncertainty in model parameters translate into uncertainty in the performance of interest? The decision makers must assess whether any observed nominal improvement in the environmental and economic effects resulted from remanufacturing at certain states is likely to be a true improvement, suggesting remanufacturing in those states, or conversely, a consequence of the parameter uncertainties due to statistical estimation errors, favoring remanufacturing when it causes negative effects. Note that the nominal improvement here refers to the improvement obtained from the planning model that uses the maximum likelihood estimates without considering parameter uncertainty. The assessment the decision maker needs to make here corresponds to the so-called "Optimizer's curse" phenomenon if we obtain an optimal decision based on a given dataset and evaluate its performance on a different dataset, then the resulting out-of-sample performance is often disappointing. To mitigate the impacts of the uncertainty in model parameters, we construct an ambiguity set that contains the true transition probabilities with high confidence using historical data and formulate the remanufacturing planning problem as a robust Markov decision process (MDP) that helps remanufacturers hedge against the worst transition probabilities.

We further establish structural properties of optimal robust policies for decision making in remanufacturing planning. We show that the optimal robust policies are of control-limit type with respect to both the condition of the equipment and the cumulative number of remanufacturing processes. Control limit refers to some threshold that delineates the upper or lower limit of the range of some action. These key properties provide useful managerial insights that support remanufacturers' robust decision making, allow us to reduce the search effort for determining the optimal policy, and facilitate easy implementation in practice. In addition, based on the monotone structure of the optimal robust policies, we develop a monotone value iteration algorithm to reduce computational efforts. Computational studies using simulated operational data of NASA turbo fan engine are conducted to demonstrate the optimal robust policies and to investigate the out-of-sample performance of the resulting optimal robust remanufacturing policies. We further derive data-driven solutions to improve the out-of-sample performance.

The main contributions of this paper are threefold. First, we develop a robust remanufacturing planning framework that helps remanufacturers to mitigate the effects of statistical estimation errors caused by limited data and/or errors contained in the data. Our study represents an initial attempt to prescribe optimal robust planning policies that help remanufacturers remain robust with respect to statistical estimation errors. Second, we establish sufficient conditions that ensure the optimal robust remanufacturing policies are of control-limit type. Few papers in the robust MDP literature have characterized the properties of optimal robust policies. The control-limit remanufacturing policies are appealing because of its easiness of implement and the computational efficiency. Lastly, we present a comprehensive computational study to demonstrate the utility of the proposed method and examine the impacts of parameter uncertainties. In the computational study, to enhance the out-of-sample performance, we have further developed data-driven decisions that perform well under the most adverse distribution within a certain statistical distance (e.g., phi-divergence, Wasserstein distance) from a nominal distribution constructed from the training samples in the computational study. The goal of this data-driven decision-making is to learn a decision from finitely many training operational data that will perform well on unseen data.

The remainder of this paper is organized as follows. Section 2 reviews relevant literature on remanufacturing planning and sequential decision making with parameter uncertainty. In Section 3, we develop the data-driven robust remanufacturing planning model. Section 4 establishes conditions to ensure the optimal robust policies are of control-limit type. In Section 5, we present a computational study using simulated operational data of NASA's turbofan engines. Section 6 concludes this study and suggests future research directions.

2. Literature review

Our study is related to two streams of the literature: remanufacturing planning and sequential decision-making with parameter uncertainty.

2.1. Remanufacturing planning

Due to limited data and/or the noises and incorrect information contained in the data, remanufacturing planning is typically confronted with the *internal uncertainty* in addition to the *external uncertainty*. The internal uncertainty is due to the stochastic nature of a component or system's condition evolution and the external uncertainty is due to the deviation of the estimates from their true values. Existing works on remanufacturing timing decisions often either ignore both types of uncertainties in transition dynamics of a remanufacturing system or only focus on the internal variation. For example, Song et al. (2015) determine remanufacturing timing based on a *deterministic* degradation process charaterized by residual strength factors. Wang et al. (2016) recommend remanufacturing based on online monitoring: Products are remanufactured when it reaches the limit condition beyond which the product is no longer remanufacturable. External variation is largely ignored in these works, and hence, remanufacturing could be blindly suggested even if it might lead to increased negative environmental or economic impacts, resulting in the robustness of remanufacturing planning decisions in question.

Remanufacturing planning decision bears a close resemblance to maintenance planning which aims to determine the optimal timing of preventive maintenance. In this paper, we model the transition dynamics using Markov models; the most relevant works in maintenance optimization literature are the ones that model maintenance problems using MDP (e.g., Elwany et al., 2011; Kim & Makis, 2013; Kurt & Kharoufeh, 2010). Most maintenance optimization models that are formulated as an MDP, however, assume that the cost parameters and the transition kernel are known, and hence, cannot provide satisfactory out-of-sample performances when future realizations deviate from the predicted ones. One of the few papers that consider ambiguity in transition probabilities is by Kim (2016). In his paper, Kim (2016) considers a failing system whose underlying state is unobservable and accounts uncertainties in both posterior distributions and transition probabilities. Our work contributes to the maintenance literature where very few papers have examined the impacts of the parameter uncertainty in the decisions and the performance of interest.

Several recent papers consider parameter uncertainty in maintenance decision making, using a Bayesian approach. For example, Fouladirad et al. (2018) study time-based replacement policies when parameters of the time-to-failure distribution are unknown and investigate the asymptotic distribution of the optimal replacement decision and the optimal average cost. de Jonge et al. (2015) and Omshi et al. (2020) use a Bayesian approach to model the parameter uncertainty and adjust maintenance decisions based on posterior distributions. The Bayesian approach, while providing a natural choice for learning parameter values, presents challenges in specifying an appropriate prior distribution when prior information of unknown parameters is limited. Our paper differs from these papers in two aspects. First, we consider an alternative approach to model parameter uncertainty. We use data-driven methods to construct ambiguity sets that contain true parameters with high confidence, which allow a decision maker to hedge against the worst-case parameters. Moreover, our objective is to find optimal robust policies that maximize the total reward under the worst transition probabilities, whereas the aforementioned three papers focus on quantifying the uncertainty in the optimal average cost rate and adjusting maintenance policies when more information becomes available. Second, we formulate the problem as a sequential decision process and these three papers model the maintenance decision problem as a renewal process.

2.2. Sequential decision making with parameter uncertainty

Early works on the MDPs with parameter uncertainty, including Satia & Lave (1973); Silver (1963); White & El-Deib (1986) and White & Eldeib (1994), formulate the uncertainty in either a game-theoretic or Bayesian approach. The game-theoretic approach assumes that the uncertainty about the transition probabilities is encoded by describing the set of all transition probability rows. Hence, when the decision maker makes a decision for a given state, the nature, who plays an adversarial role, observes the decision, and selects a transition probability row from the set to minimize the reward. Satia & Lave (1973) use the game-theoretic formulation to model the transition uncertainty in MDP and proposed a policy iteration procedure to solve the problem. White & Eldeib (1994) further develop a modified policy iteration-based algorithm for the MDP with imprecise transition probabilities. The Bayesian approach, first introduced by Silver (1963), assumes a known priori probability distribution of each transition probability row. The transition probabilities can be updated along the Bellman's equations. Dirichlet priors are a common choice of modeling the uncertainty in transition probabilities (Delage & Mannor, 2010).

Most of the early contributions, however, do not concern the construction of ambiguity sets. Inspired by the data-driven approaches, recent robust MDP works (Iyengar, 2005; Nilim & El Ghaoui, 2005; Wiesemann et al., 2013) have developed various methods to construct the uncertainty set of transition probabilities that contain the true transition probabilities with high confidence. Many statistical methods, such as likelihood constraints, deviation-type constraints and distance metrics (e.g., Wasserstein ball, ϕ-divergence balls), have been applied to construct an uncertainty set of transition probabilities with historical samples (Iyengar, 2005; Nilim & El Ghaoui, 2005; Wiesemann et al., 2013). Reformulation of robust MDPs with different types of ambiguity sets and the corresponding tractability have also been studied in the literature. Compared to the theoretical orientation of these works, our present work focuses more narrowly on developing methods for a specific problem class, establishing structural properties of optimal robust policies, and providing executable insights.

3. Robust remanufacturing planning problem

3.1. Model development

Consider remanufacturing planning of a single-component system that degrades during its operation. Because we focus on single-component systems, the words system and component are used interchangeably throughout the paper. The system is inspected at equally spaced discrete time epochs $\mathcal{T} = \{0, 1, \ldots\}$. Let $(\mathcal{S}, \mathcal{K})$ be the state space, where $\mathcal{S} = \{0, 1, 2, \ldots, S\}$ represents the set of condition states and $\mathcal{K} = \{0, 1, \ldots\}$ represents the set of cumulative numbers of completed remanufacturing activities. A larger value in $\mathcal{S}$ denotes a worse condition and the worst state s is an absorbing state, meaning the system is not operating properly and needs to be either remanufactured or scrapped. It should be noted that state $s \in \mathcal{S}$ can be a specific physical characteristic that reflects the condition of a system or a health index obtained from various sensor data to reflect the overall condition of a system. We consider a one-dimensional condition state because remanufacturing is a means for overhauling a system and in practice, when remanufacturing a system (e.g., an engine), the decision is typically based on the overall state of the system. If the state of a system is multi-dimensional, one can reduce the dimension of the data and obtain a one-dimensional health index and model the evolution process of this index. At each epoch, a decision maker observes the state of the system and then chooses an action from the set $\mathcal{A} = \{0, 1, 2\}$, where 0 means continuing operation to the next decision epoch, 1 means remanufacturing, which takes one decision period (i.e., duration between two consecutive decision epochs), and 2 means scrapping the component. Note that the scrap action takes the system to an absorbing state, denoted by Δ, in which case the system remains in the state Δ and the remanufacturing planning problem ends. The complete state space is thus $\mathcal{S} \times \mathcal{K} \cup \{\Delta\}$. The objective of the remanufacturing planning optimization is to maximize the total profit for a system during its lifetime, including extended lifetimes as a result of remanufacturing. This is practical for some applications. For example, some products (e.g., engines) have a long life span. During the lifetime of such a product, a new generation of products that use new, advanced technologies have often emerged. The user therefore typically purchases the new, upgraded product. The operational costs and gains of a product of the newer generation can be significantly different from the old, outdated ones, requiring a new remanufacturing planning policy. A notation list is provided in Table 1.

An important objective of remanufacturing is to minimize the negative environmental impacts while sustaining profitable

Table 1
Notation list.

$\mathcal{A}$	action space $\mathcal{A} = \{0, 1, 2\}$	$\mathcal{S}$	condition state space $\mathcal{S} = \{0, 1, 2, \ldots\}$
$\mathcal{K}$	cumulative number of completed remanufacturing activities $\mathcal{K} = \{0, 1, \ldots\}$	$\mathcal{U}$	ambiguity set
$\mathcal{T}$	planning horizon $\mathcal{T} = \{0, 1, 2\ldots\}$	θ	radius of ambiguity sets
$e(s,k)$	environmental cost of state (s,k)	$g(s,k)$	operational gain of state (s,k)
$r(s,k)$	reward of action 0; $r(s,k) = g(s,k) - e(s,k)$	c_r	remanufacturing cost
c_s	salvage value	$p(s'\|s,k)$	transition probability from states (s,k) to (s',k)
$\hat{p}(s'\|s,k)$	estimated transition probability	$\boldsymbol{P}$	transition probability matrix
$\hat{\boldsymbol{P}}$	estimated transition probability matrix	β	discount factor
$n(s'\|s)$	number of transitions from states s to s'		

growth. The direct environmental impacts of a remanufactured system are often measured by greenhouse gas emissions (e.g., CO_2, CH_4, N_2O, etc.) using life cycle assessment (LCA). Instead of using direct environmental impacts, we model the environmental effects using carbon cost, which is determined by the amount of carbon emissions and the carbon price, so that we have a single-objective problem, which is computationally efficient. As more market-based mechanisms, such as taxes on emissions, tradable emission allowances and deposit-refund schemes for harmful products (Abdallah et al., 2012), being designed and instituted, remanufacturing planning models that consider the carbon costs will become more relevant and applicable.

To model the profit of a remanufacturing system, we assume that during each decision period, the decision maker receives a gain $g(s,k)$ (e.g., production revenue) and incurs some environmental costs $e(s,k)$ if operation is not interrupted (i.e., $a = 0$). Note that when the system does not function properly in the worst condition s , the operational gain can be negative. The reward of keeping operation in one period is thus denoted by $r(s,k) = g(s,k) - e(s,k)$. If the decision is to remanufacture the component, a remanufacturing cost c_r, which comprises the manufacturing and carbon costs, is incurred. If the system is scrapped, a salvage value c_s is received. We assume that c_r and c_s are constants and do not depend on the condition of the system. We assume that remanufacturing cost (c_r) is a constant because remanufacturing process typically includes a number of operations, such as disassembly, cleaning, inspection, repair, replacement, and assembly. For many systems, the costs of most operations are fixed and the cost difference resulted from condition difference is negligible. When components are scrapped, there are two main mechanisms. One grades the component and prices the used component based on its condition, and the other one provides a fixed price. We consider the latter case in this study and assume that the scrap value is fixed. The system in the absorbing state Δ yields no operational gain, i.e., $r(\Delta) = 0$.

Although remanufacturing restores a component to like-new conditions, the system is not in an as-good-as-new state in its first life, and the expected value of the extended lifetime is typically shorter. To model this effect, we assume that a system's transition probabilities are dependent on the cumulative number of completed remanufacturing activities. We denote the transition probability matrix when the decision is to keep the system in operation by $\boldsymbol{P} = [p(s'|s,k)]_{s,s' \in \mathcal{S}, k \in \mathcal{K}}$ for a system that has been remanufactured k times. When $a = 0$ (the system is kept in operation), the system transitions from (s,k) to (s',k) with probability $p(s'|s,k)$. Note that when the system is kept in operation, the cumulative number of remanufacturing operations remains the same. We assume that the system can only transition to a state that is worse than the current state when the system is kept in operation ($a = 0$), that is, $p(s'|s,k) = 0$ for $s > s'$. We assume that remanufacturing brings the system to a like new condition (i.e., $s = 0$) but increments the cumulative number of remanufacturings by one. That is, when $a = 1$ (the system is remanufactured), the state of the system becomes $(0, k+1)$. This assumption is motivated by some practical applications. For example, the wall thickness of some piping system is a critical characteristic of its condition, and remanufacturing operation often adds additional materials and restores the thickness to the same level as a new system, but the remanufactured piping system usually deteriorates faster and has a shorter remaining useful life comparing to a brand-new system. We will further address the stochastic dominance relationship of transition behaviors under different k values when analyzing the structure of the optimal robust planning policies in Section 4. Due to limited data availability and statistical estimation errors, the transition probability of a remanufacturing system is fundamentally unknown. We construct an ambiguity set, denoted by $\mathcal{U}$, to model the uncertainty in the transition probability matrix $\boldsymbol{P}$. An appropriate ambiguity set should contain the underlying transition probability matrix with high confidence. Next we present an important assumption regarding the ambiguity set, which ensures deterministic and Markovian policies (Iyengar, 2005).

Assumption 1 (Rectangularity)**.** A robust MDP problem has a rectangular ambiguity set if the ambiguity set has the form $\mathcal{U} = \bigotimes_{s \in \mathcal{S}, k \in \mathcal{K}} \mathcal{U}_{sk}$ where $\bigotimes$ stands for the Cartesian product, and $\mathcal{U}_{sk}$ is the projection of $\mathcal{U}$ onto the parameters of state (s,k).

The implication of the rectangularity assumption is often interpreted in an adversarial setting (Iyengar, 2005; Nilim & El Ghaoui, 2005): The decision maker first chooses a policy π. Then an adversary observes π, and chooses a distribution that minimizes the reward. In this context, rectangularity is a form of an independence assumption: The choice of a particular distribution for a given state (s,k) does not limit the choices of the adversary of other states. There are two possible models to address the transition matrix uncertainty. One is the stationary uncertainty model where the worst-case transition probability matrix is chosen by the adversary once and for all, and remains fixed thereafter. The other one is the time-varying uncertainty model where the worst-case transition probability matrices can vary arbitrarily with time. In this paper, we consider the stationary worst-case distribution, that is, the choices of $\boldsymbol{p}(\cdot|s,k)$ are the same every time the state (s,k) is encountered. Note that there is no ambiguity in transitions in the period during which remanufacturing is conducted, since remanufacturing takes one period and there is no transition in that period. Because the optimal robust policies of the remanufacturing planning are Markovian and deterministic under the rectangularity assumption, we have the robust remanufacturing planning optimization model in the following recursive form:

$$V(s,k) = \sup_{a \in \mathcal{A}} w(s,k;a), \tag{RRmPO}$$

where

$$w(s,k;a) = \begin{cases} \inf_{\boldsymbol{P} \in \mathcal{U}} r(s,k) + \beta \sum_{s' \in \mathcal{S}} p(s'|s,k) V(s',k), & a = 0, \\ -c_r + \beta V(0, k+1), & a = 1, \\ c_s, & a = 2. \end{cases}$$

and $\beta \in [0, 1)$ is the discount factor.

3.2. Construction of ambiguity sets

The construction of ambiguity sets has been extensively studied. An ambiguity set $\mathcal{U}$ is considered *statistically* good if it is constructed with the asymptotic property $\liminf_{n\to\infty} P(P_0 \in \mathcal{U}) \geq 1-\alpha$, where P_0 is the true distribution and n is the number of samples (Lam, 2019). Methods that create ambiguity sets as confidence regions for P_0 include moment-based constraints, Wasserstein balls, ϕ-divergence balls. Among these ambiguity sets, moment-based ambiguity sets appear to display better tractability properties (Delage & Ye, 2010), but they do not consider any distributional information. Completely different distributions might have the same moments, consequently leading to overly conservative solutions. An attractive alternative is to define the ambiguity set as a ball in the space of probability distributions by using a probability distance function such as the ϕ-divergence or the Wasserstein metric. Such metric-based ambiguity sets contain all distributions that are close to a nominal or most likely distribution with respect to the prescribed probability metric. By adjusting the radius of the ambiguity set, both ϕ-divergence and the Wasserstein ambiguity sets allow decision makers to control the degree of conservatism. In this paper, we will first consider the use of ambiguity sets that are constructed as confidence sets using ϕ-divergence because (1) many ϕ-divergence have already been commonly used in statistics (e.g., the Kullback–Leibler distance, Burg entropy, and χ^2-distance), making them attractive to deal with data directly, and (2) ϕ-divergence sets preserve convexity, resulting in computationally tractable models. Robust models with Wasserstein ambiguity sets are more computationally involving, but it has been demonstrated that the worst-case expectation over a Wasserstein ambiguity set can be computed efficiently via convex optimization techniques for numerous loss functions of practical interest, and more importantly, Wasserstein ambiguity sets offer powerful out-of-sample performance guarantees (Esfahani & Kuhn, 2018; Hanasusanto & Kuhn, 2018). We will further extend our investigation of the structural properties of the optimal robust policies to Model (RRmPO) using Wassertein ambiguity sets.

4. Structure of the optimal robust policy

In this section, we investigate the structural properties of the optimal robust remanufacturing policies. We will focus our attention on control-limit policies. We establish sufficient conditions that ensure the existence of monotonically control-limit policies. The optimality of such structured policies is important because they are appealing to decision makers and enable efficient computation and are easy to implement. Our analysis will make significant use of the notion of the stochastic dominance, which helps establish stochastic dominance relationships for transition behaviors. Below, we define some stochastic order concepts that are used in our analysis.

Definition 1.

(a) A transition probability matrix $\boldsymbol{P} = [p(i|j)]_{i,j=0,1,\dots n}$ is said to be IFR (increasing failure rate) if $\sum_{i=m}^{n} p(i|j)$ is non-decreasing in j for all $m = 0, 1, \dots, n$.
(b) For two transition probability matrices $\boldsymbol{P}_1 = [p_1(i|j)]_{i,j=0,1,\dots n}$ and $\boldsymbol{P}_2 = [(p_2(i|j))]_{i,j=0,1,\dots n}$, we say $\boldsymbol{P}_1$ dominates $\boldsymbol{P}_2$, $\boldsymbol{P}_1 \succeq \boldsymbol{P}_2$, if $\sum_{i=m}^{n} p_1(i|j) \geq \sum_{i=m}^{n} p_2(i|j)$ for all $j, m = 0, 1, \dots, n$.

Assumption 2. Let $\hat{\boldsymbol{P}}(\cdot|\cdot, k)$ denote the nominal transition probability matrix for a system that has been remanufactured k times,

(a) $\hat{\boldsymbol{P}}(\cdot|\cdot, k)$ is IFR for all $k \in \mathcal{K}$.
(b) $\hat{\boldsymbol{P}}(\cdot|\cdot, k+1) \succeq \hat{\boldsymbol{P}}(\cdot|\cdot, k)$ for all $k \in \mathcal{K}$.

The nominal transition probability matrix in Assumption 2 refers to the transition probability matrix that is obtained using the conventional maximum likelihood estimation (MLE) method. Assumption 2(a) implies that, given the cumulative number of completed remanufacturing activities k, the system in a worse state at the current epoch is more likely than the other to be found in a worse condition at the next epoch. Assumption 2(b) imposes a first-order stochastic dominance relationship among the system's deterioration matrices corresponding to different remanufacturing histories. More explicitly, given two systems with the same condition but different remanufacturing histories, the system with a larger k is more likely to get worse than the other during operation. Additional assumption is made regarding the operational gains, environmental costs, and the salvage value.

Assumption 3.

(a) The operational gain $g(s,k)$ is non-increasing in $s \in \mathcal{S}$ and $k \in \mathcal{K}$, and the carbon cost $e(s,k)$ is non-decreasing in $s \in \mathcal{S}$ and $k \in \mathcal{K}$;
(b) The reward at state s , the salvage value c_s and the discount factor β satisfy the following condition: $\dfrac{r(S,0)}{1-\beta} < c_S$.

Assumption 3 (a) implies that as the number of completed remanufacturing activities increases and its condition worsens, the gain decreases and the carbon cost increases. For example, an engine in a worse state usually incurs higher maintenance costs, and consumes more gasoline or electricity, which leads to a higher environmental cost. Assumption 3(b) ensures that the decision of no intervention (i.e., $a = 0$) is excluded when a system is at the worst state for all $k \in \mathcal{K}$ because it is not practical that the system stays in the worst condition s for an infinitely long time. This unrealistic scenario is eliminated by assuming that the total expected reward from doing nothing at state $(S,0)$, computed as $\sum_{t=0}^{\infty} \beta^t r(S,0) = \dfrac{r(S,0)}{1-\beta}$, is less than the salvage value. Since $r(S,0) \geq r(S,k)$ for all $k > 0$, the condition also eliminates the no-intervention option for state (S,k) for all $k > 0$.

4.1. Remanufacturing planning with ϕ-divergence ambiguity sets

We first analyze the structure of the optimal robust policies under ϕ-divergence ambiguity sets. The ϕ-divergence between two vectors $\boldsymbol{p} = (p_1, \dots, p_m)^T$ and $\boldsymbol{q} = (q_1, \dots, q_m)^T$ is defined by Ben-Tal et al. (2013)

$$I_\phi(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right), \tag{1}$$

where the ϕ-divergence function $\phi(t)$ satisfies $\phi(t)$ is convex on $t \geq 0$, $\phi(1) = 0$, and when $q_i = 0$, the terms of (1) are interpreted as $0\phi(b/0) = b\lim_{t\to\infty}(\phi(t)/t)$ for $b > 0$, and $0\phi(b/0) = 0$ for $b = 0$. We are interested in transition probability distributions and denote the nominal distribution by $\hat{\boldsymbol{P}}$ (for notational convenience, we drop the notation of k). Given a radius θ, the ambiguity set is as follows:

$$\mathcal{U}_s = \left\{ \boldsymbol{p}_s : I_\phi\left(\boldsymbol{p}_s, \hat{\boldsymbol{p}}_s\right) \leq \theta, \sum_{s' \in \mathcal{S}} p_s(s') = 1,\ p_s(s') \in [0,1],\ s' \in \mathcal{S} \right\}. \tag{2}$$

Next, we provide reformulations and establish conditions that ensure control-limit type policies. We first reduce the bi-level problem (RRmPO) to a single-level problem by applying the Lagrangian dual theory, and then investigate the structure of the robust value function, which is necessary for establishing control-limit robust remanufacturing planning policies.

Proposition 1. *For Model (RRmPO) with ϕ-divergence ambiguity sets, $w(s,k;0)$ can be reformulated as*

$$w(s,k;0) = \sup_{\mu>0,\lambda} r(s,k) + \lambda - \mu\theta - \mu \sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k)\phi^*\left(\frac{\lambda - \beta V(s',k)}{\mu}\right), \quad (3)$$

where $\phi^*(b) = \sup_{t\geq 0} bt - \phi(t)$ is the conjugate function of $\phi(t)$.

Proof. See Appendix A.1. □

Among all ϕ-divergence ambiguity sets, the Kullback–Leibler ambiguity set has received most attention in robust optimization. Let $n(s'|s)$ be the number transitions observed from state s to state s', and let $N_s = \sum_{s'\in\mathcal{S}} n(s'|s)$ denote the total number of transitions observed from state s $N_s = \sum_{s'\in\mathcal{S}} n(s'|s)$. It has been shown that the normalized estimated Kullback–Leibler distance $2N_s I_\phi(\boldsymbol{p}_s, \hat{\boldsymbol{p}}_s)$ asymptotically follows a $\chi^2_{|\mathcal{S}|-1}$ distribution (Ben-Tal et al., 2013). In the following corollary, we show how to construct the worst transition probability distribution in a Kullback–Leibler ambiguity set.

Corollary 1. *For Model (RRmPO) with Kullback–Leibler ambiguity sets, $w(s,k;0)$ can be reformulated as*

$$w(s,k;0) = \sup_{\mu>0} r(s,k) - \mu \log\left(\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k)\exp\left(\frac{-\beta V(s',k)}{\mu}\right)\right) - \mu\theta, \quad (4)$$

and the worst-case distribution is

$$p^*(s'|s,k) = \frac{\hat{p}(s'|s,k)\exp\left(\frac{-\beta V(s',k)}{\mu^*_{sk}}\right)}{\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k)\exp\left(\frac{-\beta V(s',k)}{\mu^*_{sk}}\right)}, \quad (5)$$

where μ^*_{sk} is the optimal solution of the dual problem (4) given s and k.

Proof. See Appendix A.2. □

Proposition 2. *For Model (RRmPO) with ambiguity sets constructed using ϕ-divergence, the value function $V(s,k)$ is non-increasing in $s \in \mathcal{S}$ and $k \in \mathcal{K}$.*

Proof. See Appendix A.3. □

Based on Proposition 2, we further establish conditions that ensure control-limit robust policy structures, that is, the remanufacturing decisions are of control-limit type with respect to the condition of the system and the cumulative number of completed remanufacturing activities.

Theorem 1. *For Model (RRmPO) with ϕ-divergence ambiguity sets, there exists a cumulative number of completed remanufacturing activities $k^* \in \mathcal{K}$, and operation states $\zeta_{\rm rm}(k), \zeta_{\rm scrap}(k) \in \mathcal{S}$ such that for $k < k^*$*

$$a(s,k) = \begin{cases} 0 & \text{if } s < \zeta_{\rm rm}(k), \\ 1 & \text{if } s \geq \zeta_{\rm rm}(k), \end{cases}$$

and for $k \geq k^*$

$$a(s,k) = \begin{cases} 0 & \text{if } s < \zeta_{\rm scrap}(k), \\ 2 & \text{if } s \geq \zeta_{\rm scrap}(k). \end{cases}$$

Proof. See Appendix A.4. □

Theorem 1 shows that when has $k < k^*$, the optimal decision is either wait until the next period or remanufacture, and the system is remanufactured when the condition is equal to or exceeds the remanufacturing limit $\zeta_{\rm rm}(k)$. When the cumulative number of completed remanufacturing activities reaches the threshold k^*, the optimal decision is either wait until the next period or scrap and there exists a scrapping threshold $\zeta_{\rm scrap}(k)$. This implies that remanufacturing is not always optimal —it is not recommended after being conducted certain number of times. Note that $k^* = 0$ is a special case that remanufacturing is not optimal for all $k \in \mathcal{K}$. The structure of $\zeta_{\rm rm}(k)$ and $\zeta_{\rm scrap}(k)$ is examined in the next theorem.

Theorem 2. *Consider Model (RRmPO) with ϕ-divergence-based ambiguity set. Then, the following holds:*

(a) *If $\frac{\beta r(0,0)}{1-\beta} - \beta c_s \leq r(s,k) - r(s,k+1)$, $\zeta_{\rm rm}(k)$ is non-increasing in k, $k < k^*$, and $\zeta_{\rm rm}(k^*-1) \geq \zeta_{\rm scrap}(k^*)$.*
(b) *$\zeta_{\rm scrap}(k)$ is non-increasing in k, $k \geq k^*$.*

Proof. See Appendix A.5. □

The first part of Theorem 2(a) implies that the optimal robust policy is monotone with respect to $k \in \mathcal{K}$ for all $k < k^*$. That is, a remanufacturer tends to remanufacture earlier as the system goes through more remanufacturing processes. The second part of Theorem 2(a) ($\zeta_{\rm rm}(k^*-1) \geq \zeta_{\rm scrap}(k^*)$) indicates that if the optimal action is remanufacture ($a = 1$) at some s when $k = k^* - 1$, then the optimal action is scrap ($a = 2$) for all $s' \geq s$ when $k = k^*$. Theorem 2(b) shows that the remanufacturer should scrap early as k increases for all $k \geq k^*$. Therefore, the optimal robust remanufacturing policy has the appealing monotone structure with respect to k. Note that the condition in Theorem 2(a) is restrictive. We will show that most violations do not change the monotone structure of $\zeta_{\rm rm}(k)$ and $\zeta_{\rm scrap}(k)$ in Section 5.3.1 through computational studies.

4.2. Remanufacturing planning with Wasserstein ambiguity sets

In this section, we show that the optimal robust policies are of control-limit type for Model (RRmPO) with Wasserstein-based ambiguity sets under similar conditions. The Wasserstein distance of two distributions can be viewed as the minimum transportation cost for moving the probability mass from one distribution to the other. The Wasserstein ambiguity set contains all distributions that are sufficiently close to the empirical distribution with respect to the Wasserstein metric. Given N independently and identically distributed training samples, the true distribution P_0 belongs to the Wasserstein ambiguity set around the empirical distribution $\hat{P}_N$ with confidence $1-\alpha$ if its radius is a sublinearly growing function of $\log(1/\alpha)/N$ (Esfahani & Kuhn, 2018).

Let $(\mathcal{S}, d)$ be a metric space with metric d, and $\mathcal{F}(\mathcal{S})$ be the set of all probability distributions defined on $\mathcal{S}$. Given a radius θ and a state $s \in \mathcal{S}$ (for notational purpose, we drop the notation of k), the ambiguity set of the Wasserstein ball centered on $\hat{\boldsymbol{p}}_s \in \mathcal{F}(\mathcal{S})$ is

$$\mathcal{U}_s = \left\{\boldsymbol{p}_s : W_m^m(\boldsymbol{p}_s, \hat{\boldsymbol{p}}_s) \leq \theta^m, \sum_{s'\in\mathcal{S}} p_s(s') = 1, p_s(s') \in [0,1], s' \in \mathcal{S}\right\}, \quad (6)$$

where $W_m^m(\boldsymbol{p}_s, \hat{\boldsymbol{p}}_s)$ is the Wasserstein distance between $\boldsymbol{p}_s$ and $\hat{\boldsymbol{p}}_s$ with order m. The Wasserstein distance $W_m^m(\boldsymbol{p}_s, \hat{\boldsymbol{p}}_s)$ can be described as

$$W_m^m(\boldsymbol{p}_s, \hat{\boldsymbol{p}}_s) = \left\{\begin{matrix} \min_{\gamma\in\mathcal{F}(\mathcal{S}\times\mathcal{S})} \sum_{(x,y)\in\mathcal{S}\times\mathcal{S}} d(x,y)^m \gamma(x,y) \\ \text{s.t.} \sum_{y\in\mathcal{S}} \gamma(x,y) = p_s(x), \forall x \in \mathcal{S}, \sum_{x\in\mathcal{S}} \gamma(x,y) = \hat{p}_s(y), \forall y \in \mathcal{S} \end{matrix}\right\}. \quad (7)$$

To establish conditions that ensure the special structure of optimal robust policies for Model (RRmPO) with Wasserstein ambiguity sets, we similarly reformulate the bi-level problem into a single-level problem, and then show that the value function is non-increasing in $s \in \mathcal{S}$ and $k \in \mathcal{K}$.

Proposition 3. *For Model (RRmPO) with the Wasserstein-distance-based ambiguity set, $w(s,k;0)$ can be reformulated as*

$$w(s,k;0) = \sup_{\mu>0} -\mu\theta^m + \mu \sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k) \min_{s''\in\mathcal{S}} \left\{ d(s',s'')^m + \frac{V(s'',k)}{\mu} \right\}, \quad (8)$$

Proof. See Appendix A.6. □

Proposition 4. *For Model (RRmPO) with ambiguity sets constructed using Wasserstein distance, the value function $V(s,k)$ is non-increasing in $s \in \mathcal{S}$ and $k \in \mathcal{K}$.*

Proof. See Appendix A.7. □

Since the value function is non-increasing in $s \in \mathcal{S}$ and $k \in \mathcal{K}$, we can similarly show that structural properties in Theorems 1 and 2 hold for Wasserstein-distance-based ambiguity sets under the same conditions. The theorems and proofs are omitted here.

4.3. Solution methodology

Model (RRmPO) can be solved using robust value iteration (Iyengar, 2005). We further develop an efficient algorithm for finding optimal robust policies with control-limit structures. If the optimal robust policy is of control-limit type with respect to $s \in \mathcal{S}$ and $k \in \mathcal{K}$, then Model (RRmPO) can be more efficiently solved by the monotone robust value iteration (Algorithm 1)

Algorithm 1 Monotone robust value iteration.

1: **Initialization:**
$V(s,k), a^*(s,k) \leftarrow 0, \bar{V}(s,k) \leftarrow M, \forall (s,k) \in \mathcal{S}\times\mathcal{K}, \epsilon > 0$
2: **while** $||\bar{\boldsymbol{V}} - \boldsymbol{V}|| \geq \frac{(1-\beta)\epsilon}{4\beta}$ **do**
3: $\quad \bar{\boldsymbol{V}} \leftarrow \boldsymbol{V}$, $A(s,k) \leftarrow \{0,1,2\} \forall (s,k) \in \mathcal{S}\times\mathcal{K}$
4: $\quad$ **for** $(s,k) \in \mathcal{S}\times\mathcal{K}$ **do**
5: $\quad\quad V(s,k) \leftarrow \max_{a\in A(s,k)} w(s,k;a)$,
$a^*(s,k) \leftarrow \arg\max_{a\in A(s,k)} w(s,k;a)$
6: $\quad\quad$ **if** $s+1 \in \mathcal{S}$ **then**
7: $\quad\quad\quad A(s+1,k) \leftarrow \{a : a \geq a^*(s,k)\}$
8: $\quad\quad$ **end if**
9: $\quad\quad$ **if** $k+1 \in \mathcal{K}$ **then**
10: $\quad\quad\quad A(s,k+1) \leftarrow \{a : a \geq a^*(s,k)\}$
11: $\quad\quad$ **end if**
12: $\quad$ **end for**
13: **end while**
14: **return** $\boldsymbol{V}, \boldsymbol{a}^*$

This modified algorithm differs from robust value iteration in Iyengar (2005) in that the action space A becomes smaller with increasing s and k. Specifically, given a state (s,k) and its optimal robust solution $a^*(s,k)$, we reduce the action space of state $(s+1,k)$ in step 7 based on Theorem 1 and reduce the action space of state $(s,k+1)$ in step 10 based on Theorem 2. For example, if the optimal action for a given state (s,k) is to remanufacture (i.e., $a(s,k) = 1$), then the optimal action for any state (s',k') where $s' > s, k' = k$, the optimal action is to remanufacture based on Theorem 1. Similarly, if the optimal action for a given state (s,k) is to remanufacture (i.e., $a(s,k) = 1$), then the optimal action for any state (s',k') where $s' = s, k' > k$, the optimal action is to remanufacture based on Theorem 2. In the worst case, $A(s,k)$ remains the same for all $s \in \mathcal{S}$ and $k \in \mathcal{K}$ and computational effort is the same as that of the robust value iteration algorithm; however, when the control limits exist, the sets $A(s,k)$ will decrease in size as s and k increase and hence the number of actions which need to be evaluated in step 5 is reduced; at some state (s,k), the action set may only contain a single element, and no further optimization is necessary since that action will be optimal for all states $(s',k'), s' \geq s, k' \geq k$. Therefore, this algorithm achieves a better computational efficiency than the robust value iteration when the optimal robust policy has a monotone structure. The inner problem $w(s,k;0)$ in step 5 can be solved by employing a numerical search for its dual problem by taking the advantage that both dual problems are concave in their decision variables. Note that the time complexity of the robust value iteration algorithm for a ϵ-optimal robust policy is $O(C|\mathcal{S}|\log(R/\epsilon)/\log(1/\beta))$ (Iyengar, 2005), where C is the cost of computing inner minimization problem $w(s,k;0)$, and R is the upper bound of the reward function. Because the time complexity of solving the inner minimization problem is polynomial for both ϕ-divergence and Wasserstein-distance-based ambiguity sets, the runtime of solving a robust MDP does not increase much compared with solving a nominal MDP.

5. Computational study

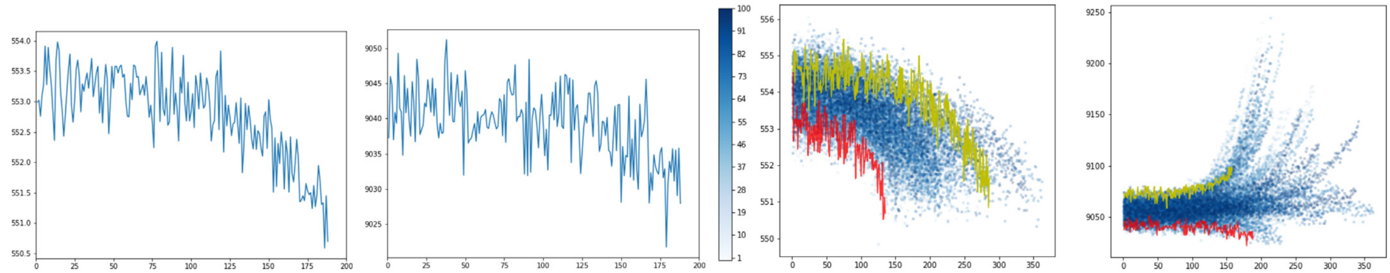
5.1. System model description

We use the operational data simulated using the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) software (Frederick et al., 2007) developed at NASA to demonstrate our robust remanufacturing planning model and examine the performance of the optimal robust remanufacturing policies. The C-MAPSS offers 14 inputs and can produce a number of outputs for analysis.

The dataset used in this study pertains to a single failure mode and a single operating condition, and consists of 100 units which are run to failure. Note that end-of-life can be subjectively determined as a function of operational thresholds that can be measured; these thresholds depend on user specifications to determine safe operational limits. For illustration purposes, we arbitrarily choose four features and plot the time series of these features for a randomly selected unit and all units (Fig. 1). From Fig. 1, we can see that the data contains a lot of noises. Various sources can contribute to noises, and the main sources of noise are manufacturing and assembly variations, process noise, and measurement noise to name a few important ones (Saxena et al., 2008). Due to the large amount of noises and limited real-world operational data available, there often exists a high level of uncertainties in transition probabilities of the turbofan engines, and operators and manufacturers are in great need of robust remanufacturing planning.

5.2. Construction of the ambiguity set

It is typically desirable to reduce the dimensionality of the data and reconstruct them from a lower dimensional samples. We therefore use the principal component analysis method to compress the high-dimensional sensor outputs and use the first principal component that accounts for the largest variability of data (approximately 70% on average) as the health indicator. We further discretize the obtained health indicator into 7 intervals, representing 7 condition states, as recommended by Moghaddass & Zuo (2014). The nominal transition probability is estimated using the maximum likelihood method, i.e., $\hat{p}(s'|s) = \sum_{i=1}^{m} n_i(s'|s) / \sum_{i=1}^{m} \sum_{s'\in\mathcal{S}} n_i(s'|s)$, where $n_i(s'|s)$ is the number of transitions from state s to s' for unit i, and m is the total number of units in a sample. We construct the ambiguity sets as de-



(a) Feature 7 (b) Feature 9 (c) Feature 11 (d) Feature 13

Fig. 1. Illustrations of raw sensor data sequences. (a) and (b), time series of the selected features of unit 6. (c) and (d), time series of the selected features of all units. Solid lines are the time series of the unit that has the most maximum (yellow line) and minimum (red line) points. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

scribed in Section 3.2. It has been reported in the literature that remanufactured components/systems are like-new but have reduced lifetimes (Östlin et al., 2009). However, the data simulated using the C-MAPSS software do not contain operational data after remanufacturing. To model the reduction in lifetime after remanufacturing, we modify the nominal transition probability matrix obtained for new turbofan engines (i.e., $k = 0$) for each k (the number of completed remanufacturing operations). Specifically, we assume that the mean time to failure of a system is reduced by approximately 7% each time it is remanufactured. This percentage is arbitrarily chosen. We then adjust the nominal transition probability matrix to achieve this reduction by trial and error. If post-remanufacturing operational data are available (i.e., the transition histories) for all k, then for each number of completed remanufacturing operations, one can repeat the estimation procedure described in Section 5.2 to obtain the nominal (empirical) transition matrix, which is the center of the ball that contains all possible transition matrices. The radius of the ball can be determined by either choosing the desired confidence level (as described in Section 4) or using the data-driven approach which uses out-of-sample tests to select the best-performing radius (as described in 5.3.3).

5.3. Experiments

Next, we demonstrate the structure of the optimal robust remanufacturing policy and examine the out-of-sample performance of the optimal robust policies. We arbitrarily choose cost parameters that satisfy Assumption 3 in all the following experiments: $g(s,k) = 4 - 0.25s - 0.25k$, $e(s,k) = 1 + 0.25s + 0.25k$, $c_r = 2$, and $c_s = 0.5$. We follow the convention in the MDP works that arbitrarily select a value of the discount factor no less than 0.8 (Delage & Mannor, 2010; Goh et al., 2018; Wiesemann et al., 2013). Thus, the discount factor β is 0.9 for all the following experiments. We use the Kullback–Leiber distance to demonstrate the performance of general ϕ-divergence. For Wasserstein distance, we consider order $m = 1$. In our experiments, the nominal policy refers to the optimal remanufacturing policy obtained using the nominal transition probabilities (i.e., the MLE estimates) which does not consider parameter uncertainties, and we refer to this approach as the nominal approach.

5.3.1. Policy structures

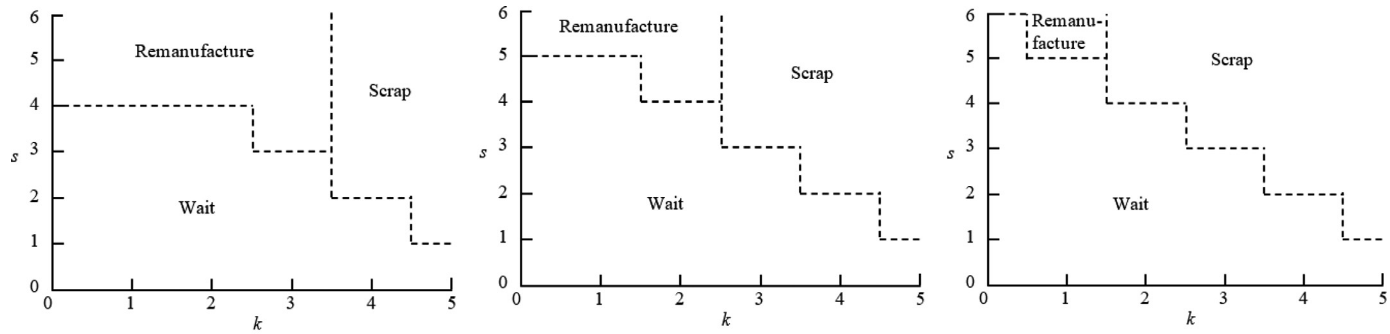
We have established conditions to ensure control-limit policies for Model (RRmPO) with ϕ-divergence ambiguity sets and for Model (RRmPO) with Wasserstein ambiguity sets. For illustration purposes, we show the structure of optimal robust policies for Kullback–Leibler ambiguity sets and Wasserstein ambiguity sets. As Fig. 2 shows, the remanufacturing policies exhibit control-limit structure. We can also see that as θ increases, the remanufacturing threshold $\zeta_{\rm rm}(k)$ increases and k^* decreases (i.e., the scrap action is performed earlier). This implies that when parameter uncertainty is large, a decision maker needs to be cautious about remanufacturing used products and to consider scrapping at an earlier stage. This is because (1) the remanufacturing cost may not be offset by the subsequent operational gains due to large parameter uncertainties and (2) securing the fixed salvage value better hedges against uncertainties in future gains.

As stated earlier, the condition of Theorem 2(a) is restrictive and difficult to satisfy. We further examine whether the optimal robust policies are still of control-limit type when this condition is violated. We test a total of 5000 instances and the generation of the test instances is described in Appendix B.1. Out of the 3060 test instances that violate the condition of Theorem 2(a), only 209 (i.e., approximately 6.8%) instances violate the monotone structure. Therefore, we believe that a control-limit policy with respect to k can be obtained in most practical cases even when the condition that guarantees it is violated.

We further investigate the structure of the optimal robust policy when remanufacturing costs and salvage values are state-dependent. We conducted a numerical study that considers more parameter values to examine whether the control-limit structures in Theorems 1 and 2, still exist. We assume $c_r(s,k)$ and $c_s(s,k)$ are linear with respect to s and k. Suppose $c_r(s,k) = d_r + a_r s + b_r k$ and $c_s(s,k) = d_s + a_s s + b_s k$, where a_r, b_r, d_r, a_s, b_s, and d_s are parameters. The ranges of parameters are all bounded by 0 and 2. The operational gain function is the same as the one used in other experiments, i.e., $r(s,k) = 3 - 0.5s - 0.5k$. There are approximately 20% cases that violate Theorems 1 and 2 for Kullback-Leibler-distance-based ambiguity sets. For Wasserstein-distance-based ambiguity sets, there are about 25.7% cases that violate Theorem 1 and 26.8% cases that violate Theorem 2. This shows that in the majority of cases with state-dependent costs, the optimal robust policies are still of the control-limit type.

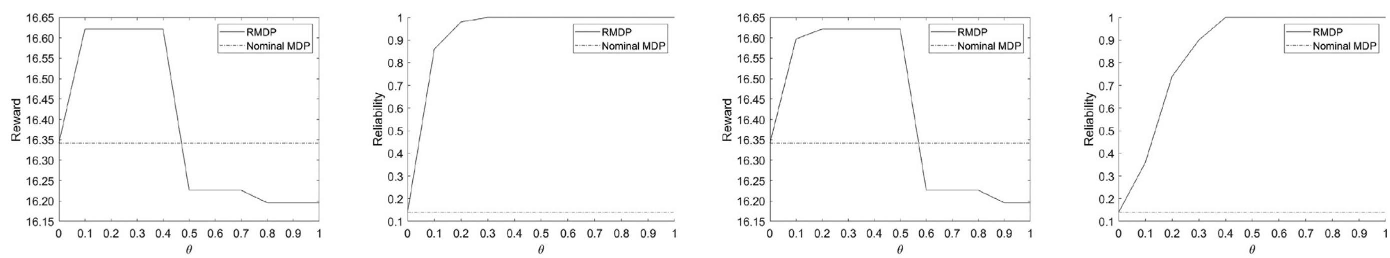
5.3.2. Impact of the parameter uncertainty

We first conduct experiments to investigate the impact of the parameter uncertainty on the out-of-sample performance. We sample a training set $\mathcal{N}$ from the data set to obtain nominal transition probability $\hat{\boldsymbol{p}}$ using the maximum likelihood estimator in Section 5.2, where $m = |\mathcal{N}|$. The optimal robust policies of Model (RRmPO) with ambiguity sets constructed under different hyperparameter values using the training dataset, $\pi_{\mathcal{N}}(\theta)$ (θ determines the size of an ambiguity set), are then implemented in a test dataset $\mathcal{M}$ to assess the out-of-sample performance. We examine two performance measurements: the average reward and the reliability of performance guarantees. The average reward is defined as $\bar{v}_{\mathcal{N}}(\theta) = \sum_{i \in \mathcal{M}} v_i(\pi_{\mathcal{N}}(\theta))/|\mathcal{M}|$, where $v_i(\pi_{\mathcal{N}}(\theta))$ is the expected



(a) $\theta = 0$ (b) $\theta = 0.5$ (c) $\theta = 1$

Fig. 2. Optimal robust policies for different θs.



(a) (b) (c) (d)

Fig. 3. Out-of-sample reward $\bar{\nu}_{\mathcal{N}}(\theta)$ and reliability $\Pr\{\bar{\nu}_{\mathcal{N}}(\theta) \geq V_{\mathcal{N}}(\theta)\}$ as a function of ψ. (a) and (b) Kullback–Leibler ambiguity set. (c) and (d) Wasserstein ambiguity set.

reward of robust policy $\pi_{\mathcal{N}}(\theta)$ for test sample $i \in \mathcal{M}$ when the system is brand new ($s = 0, k = 0$). The reliability is defined as the proportion of the event $\nu_i(\pi_{\mathcal{N}}(\theta)) \geq V_{\mathcal{N}}(\theta)$ for all $i \in \mathcal{M}$, where $V_{\mathcal{N}}(\theta)$ is the in-sample value of $V(0, 0)$ given θ.

Figure 3 depicts the experiment results when the size of the training dataset is 5 ($|\mathcal{N}| = 5$) and the size of the test dataset is 50 ($|\mathcal{M}| = 50$). The value iteration algorithm and the monotone robust value iteration algorithm (Algorithm 1) are used to efficiently obtain the nominal policy and the robust policy, respectively. From Fig. 3(a), we observe that the average reward of the robust policy is better than that of the nominal policy when θ is not too large. As θ increases, the average reward of the robust policy deteriorates because the robust policy is too conservative. The empirical reliability in Fig. 3(b) is in general non-decreasing in θ, and the reliability of the performance guarantee under the robust approach is much higher than that under the nominal approach. We also find that the out-of-sample average reward using a robust approach is better as long as the reliability of the performance guarantee is noticeably smaller than 1 and deteriorates when it is close to 1. Figure 3(c) and (d) present the out-of-sample performance and the reliability of Model (RRmPO) with the Wasserstein-distance-based ambiguity sets, respectively. Similar patterns are observed. Results of this experiment provide an empirical justification of adopting a robust remanufacturing approach, especially when the size of the dataset is small.

5.3.3. Remanufacturing planning driven by out-of-sample performance

From the previous experiment on the impact of the parameter uncertainty, it is shown that different hyperparameter θ values may lead to robust remanufacturing policies with different out-of-sample performance $\bar{\nu}_{\mathcal{N}}(\theta)$. It is desired to select a θ that maximizes the average award $\bar{\nu}_{\mathcal{N}}(\theta)$. This, however, requires the true transition probability that is not precisely known. We select the optimal θ via validation using the training data. Specifically, we randomly select 60% of the training dataset $\mathcal{N}$ for training and the remaining 40% of the training data is used for validation. Using newly formed training dataset to construct the ambiguity sets, Model (RRmPO) is solved for a finite number of candidate hyperparameter θ. We then use the validation dataset to evaluate the out-of-sample performance of $\pi_{\mathcal{N}}(\theta)$, select the optimal θ^* as the one that maximizes $\bar{\nu}_{\mathcal{N}}(\theta)$ of the validation set, and report $\pi_{\mathcal{N}}(\theta^*)$ as the data-driven solution.

Figure 4(a) shows the mean value of the out-of-sample performance $\bar{\nu}_{\mathcal{N}}(\theta^*)$ as a function of the sample size $|\mathcal{N}|$. We also observe that both out-of-sample and in-sample performances exhibit asymptotic consistency. Figure 4(b) shows the mean of the reliability of the guaranteed performance under different sample sizes. We can see that the robust policy significantly outperforms the nominal one, particularly when the training data is scarce. As more data become available, the optimal robust policy converges to the nominal policy, and so does the performance of the robust policy. Figure 4(c) reports in-sample estimate $V_{\mathcal{N}}(\theta)$. We can see that the nominal approach is over-optimistic while the robust approaches act on the cautious side.

5.3.4. Comparison with the alternative Bayesian approach

To demonstrate the performance of the parameter uncertainty modeling approach used in this study as a viable alternative, we further compare the performance of the Bayesian approach and the proposed robust approach in remanufacturing planning. Specifically, we compare the out-of-sample performance of the Bayesian approach for prior distributions that are randomly chosen with that of the proposed robust approach. For computational efficiency, we consider Dirichlet priors because the Dirichlet distribution is the conjugate prior for the multinomial distribution. Let $Dir(\xi_s \boldsymbol{\lambda}_s)$ be the Dirichlet distribution given state $s \in \mathcal{S}$ and $k = 0$ with parameter $\xi_s \boldsymbol{\lambda}_s$, where $\xi_s \in \mathbb{R}^+$ and $\boldsymbol{\lambda}_s \in \mathbb{R}^{|\mathcal{S}|}$ satisfying $\boldsymbol{\lambda}_s^{\mathrm{T}} \mathbf{1} = 1$ and

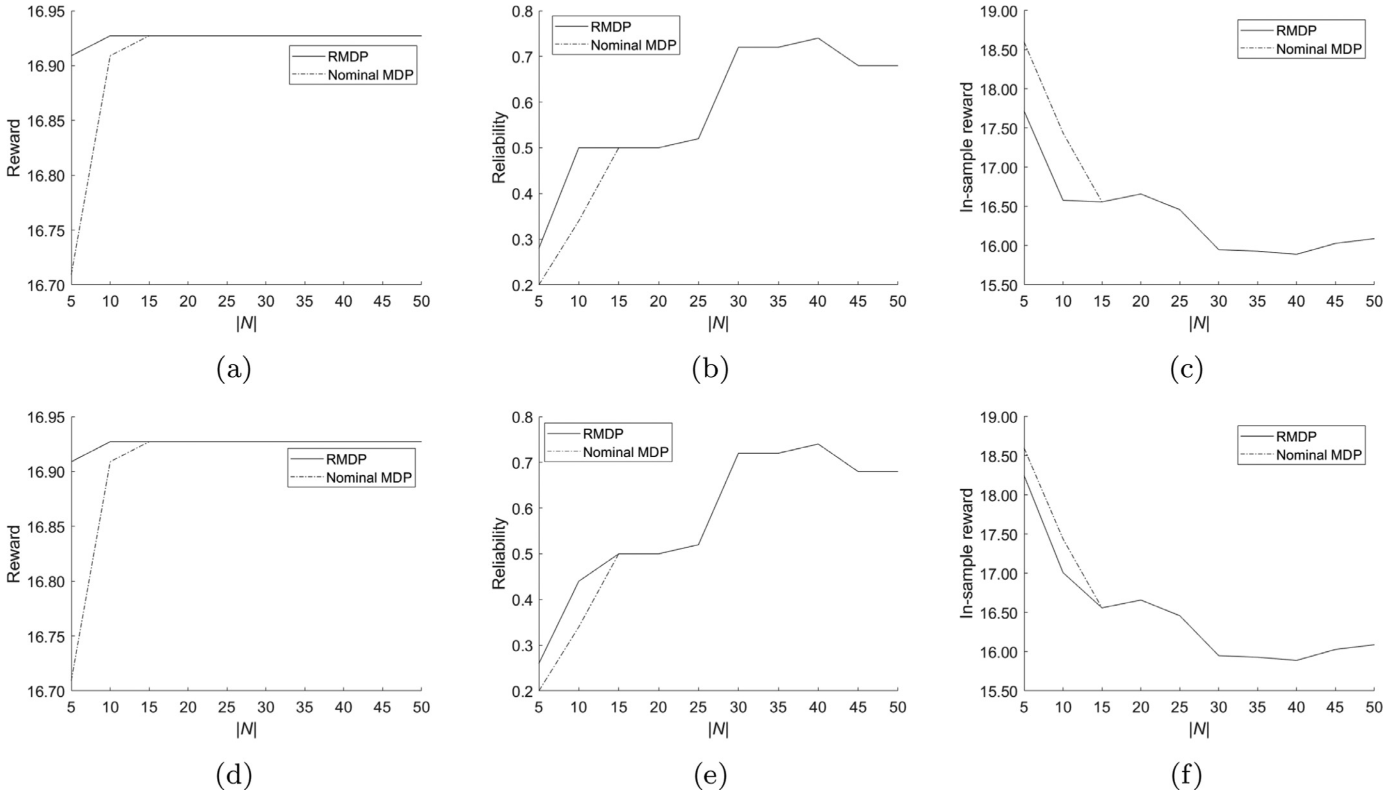


Fig. 4. Out-of-sample reward $\bar{v}_{\mathcal{N}}(\theta^*)$, reliability $\Pr\{\bar{v}_{\mathcal{N}}(\theta^*) \geq V_{\mathcal{N}}(\theta^*)\}$, and in-sample reward $V_{\mathcal{N}}(\theta^*)$ as a function of $|\mathcal{N}|$. (a)–(c) Kullback–Leibler ambiguity set. (d) and (f) Wasserstein ambiguity set.

Table 2
Average out-of-sample reward with respect to 50 priors and hyperparameter θ.

Training sample size	Bayesian approach		Robust approach		
	ξ	Mean	θ	Mean (KL)	Mean (Wass.)
1	5	16.12	0.1	16.62	16.34
	10	14.56	0.5	16.62	16.62
	20	14.56	1.0	16.23	16.20
5	5	16.62	0.1	16.62	16.60
	10	16.62	0.5	16.23	16.62
	20	16.12	1.0	16.20	16.20

$\boldsymbol{\lambda}_s \geq 0$. We use the same random priors for all states $(s, k) \in \mathcal{S} \times \mathcal{K}$. We consider three levels of $\xi \in \{5, 10, 20\}$, and a larger value implies a smaller variance of priors when randomly generated. For each level of variance ξ, we randomly generate $\boldsymbol{\lambda}$. For each prior, the posterior distribution is obtained using the training data based on Bayes' theorem. The posterior predictive transition probabilities are then served as the transition probabilities in an MDP, which is solved to obtain the optimal policy. We report the average reward of all 50 optimal policies in the test set. The same training set and the test set in Section 5.3.2 are used in this experiment. Table 2 compares the results of the Bayesian approach and the robust approach for two different training sample sizes. From Table 2, we can see that the two approaches have similar performances in some cases and that the robust approach has a slightly better average reward than the Bayesian one when data are limited. In particular, when ξ increases, the performance of the Bayesian approach decreases, especially when training sample size is small. This is because a smaller ξ leads to more sparse priors (i.e., higher variance), which can reduce the chance of concentrating on transition probabilities that are largely deviated from the true one.

Table 3
Parameter bounds in sensitivity analysis.

d	a	b	c_r	c_s
(0, 12)	(0, 2)	(0, 2)	(0, 5)	(0, 3)

5.3.5. Sensitivity analysis

In this section, we conduct more experiments to examine the performance of the robust policy under different parameter values. Specifically, we assume $r(s, k) = g(s, k) - e(s, k) = d - as - bk$, and test 1000 instances where parameters d, a, b, c_r, and c_s are drawn randomly from uniform distributions. The parameters of the uniform distributions are provided in Table 3.

We first examine the performance of the robust approach under different cost parameters and θs. Table 2 summarizes the percentages of test instances where the robust policy is no worse than the nominal policy given θ. From Table 4, we can see that there is a very high chance that the robust policy is no worse than the nominal policy when θ is small, and this chance decreases as θ increases because the robust policy can be overly conservative. This is consistent with the conclusion in Section 5.3.2.

We further compare the robust policy and the nominal policy given different training sample sizes. In this new experiment, we similarly test sample sizes of 5, 10, and 15. For all 1000 test instances of each training sample size, the robust policies are no worse than the nominal ones. This agrees with our conclusion in Section 5.3.3.

Table 5 compares the performance of the robust approach and Bayesian approach with different hyperparameter ξ and θ. From Table 3, we can see that when θ is small, the robust policy has a higher percentage of outperforming the policy using Bayesian approach.

Table 4
Percentage of cases where robust policy is no worse than the nominal policy.

θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
KL	96.6%	90.7%	84.7%	80.4%	77.9%	75.5%	73.2%	70.9%	69.7%	60.9%
Wass	99.9%	98.9%	93.6%	96.4%	79.7%	74.9%	72.9%	70.3%	68.8%	67.4%

Table 5
Percentage of instances where robust policy is no worse than the Bayesian policy.

ξ	5			10			20		
θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
KL	59.6%	43%	37.6%	62.4%	45.6%	39.6%	69.1%	55.1%	47.3%
Wass	52.3%	43.9%	36.9%	55.9%	47.5%	38.8%	63%	56.7%	46.5%

6. Conclusion and future work

In this paper, we consider the problem of remanufacturing planning in the presence of parameter uncertainty. We formulate the problem as a robust Markov decision process in which the true transition probability is unknown but lies in an ambiguity set with high confidence. Two distance-metric based ambiguity sets are considered: ϕ-divergence and Wasserstein distance. We investigate the structure of the optimal robust policies and establish conditions to ensure the policies are of control-limit type. We also establish sufficient conditions for some of the intuitive results seen in our computational study. We demonstrate the structure of the optimal robust policies via a computational study using the simulated operational data of the turbofan engine operated by NASA, investigate the out-of-sample performance, and derive the data-driven solutions to improve the out-of-sample performance.

In this paper, we consider a remanufacturing planning problem with a scrap action that takes a system to an absorbing state, in which case the remanufacturing planning problem ends. It is worth considering a replacement action taking the system to a new state in the remanufacturing decision-making process to ensure the continuity of business in the future. Remanufacturing cost and salvage value are assumed to be constants in this study; extending our model to incorporate state-dependent remanufacturing cost and salvage value in remanufacturing planning is a natural future extension of this work. Moreover, at each decision epoch, decision makers make new observation about the system, and an important question that arises is how the information that becomes available in the decision process can be leveraged to resolve some ambiguity, so that the optimal robust policies are not overly conservative. In addition, an implicit assumption made in this paper is that the states of a system are directly observable (i.e., the sensor data reveal the underlying state of the system with certainty). In practice, many systems are not directly observable and the states have to be inferred from signals collected. Future work will investigate the partially observable Markov decision process with parameter uncertainty.

Acknowledgment

This work is supported in part by the U.S. National Science Foundation under award 2305486.

Appendix A

A1. Proof of Proposition 1

Since Proposition 1 applies to all $(s,k)\in\mathcal{S}\times\mathcal{K}$, we drop (s,k) in the value function when proving this proposition for the notational convenience. The value function defined in Model (RRmPO) involves solving an inner problem for any given $s\in\mathcal{S}$ and $k\in\mathcal{K}$ as follows

$$w(s,k;0) = \min\ r+\beta\sum_{s'\in\mathcal{S}} p(s')V(s')$$
$$\text{s.t.}\quad \sum_{s'\in\mathcal{S}} p(s')=1,\ \sum_{s'\in\mathcal{S}}\hat{p}(s')\phi\left(\frac{p(s')}{\hat{p}(s')}\right),\ p(s') \geq 0,\ s'\in\mathcal{S}. \leq\theta \tag{A.1}$$

The Lagrangian dual problem of (A.1) is

$$\max_{\lambda\ \text{free},\mu\geq 0} L(\lambda,\mu)$$

where the Lagrangian dual objective function is

$$\begin{aligned} L(\lambda,\mu) &= \min_{\boldsymbol{p}\geq 0} L(\lambda,\mu,\boldsymbol{p})\\ &= \min_{\boldsymbol{p}\geq 0} r+\beta\sum_{s'\in\mathcal{S}} p(s')V(s')+\lambda\left(1-\sum_{s'\in\mathcal{S}}p(s')\right)\\ &\quad +\mu\left(\sum_{s'\in\mathcal{S}}\hat{p}(s')\phi\left(\frac{p(s')}{\hat{p}(s')}\right)-\theta\right)\\ &= \min_{\boldsymbol{t}\geq 0} r+\lambda-\mu\theta+\mu\sum_{s'\in\mathcal{S}}\left(\frac{t_{s'}(\beta V(s')-\lambda)}{\mu}+\phi(t_{s'})\right)\hat{p}(s') \end{aligned} \tag{A.2}$$

$$\begin{aligned} &= r+\lambda-\mu\theta-\mu\sum_{s'\in\mathcal{S}}\left(\max_{\boldsymbol{t}\geq 0}\frac{t_{s'}(\lambda-\beta V(s'))}{\mu}-\phi(t_{s'})\right)\hat{p}(s')\\ &= r+\lambda-\mu\theta-\mu\sum_{s'\in\mathcal{S}}\hat{p}(s')\phi^*\left(\frac{\lambda-\beta V(s')}{\mu}\right), \end{aligned} \tag{A.3}$$

where $\phi^*(a)=\sup_{t\geq 0} at-\phi(t)$ and Eq. (A.2) is implied by the change of decision variables $\boldsymbol{t}=\boldsymbol{p}/\hat{\boldsymbol{p}}$. Because the terms r, λ, and $\mu\theta$ in Eq. (A.2) are independent of the decision variable $\boldsymbol{t}$, and there is no constraint on each entry of $\boldsymbol{t}$, we obtain Eq. (A.3) by switching $\min_{\boldsymbol{t}\geq 0}$ and $\sum_{s'\in\mathcal{S}}$, and replacing $\min f$ by $-\max -f$.

A2. Proof of Corollary 1

The value function defined in (RRmPO) involves solving an inner problem for any given $s\in\mathcal{S}$ and $k\in\mathcal{K}$ as follows

$$w(s,k;0) = \min\ r(s,k)+\beta\sum_{s'\in\mathcal{S}} p(s'|s,k)V(s',k)$$
$$\text{s.t.}\quad \sum_{s'\in\mathcal{S}} p(s'|s,k)=1,\ \sum_{s'\in\mathcal{S}}\log\left(\frac{p(s'|s,k)}{\hat{p}(s'|s,k)}\right)p(s'|s,k) \leq\theta,\ p(s'|s,k)\geq 0,\ s'\in\mathcal{S}. \tag{A.4}$$

The Lagrangian dual problem of (A.4) is

$$\max_{\lambda \text{ free}, \mu \geq 0} L(\lambda, \mu) \quad \text{s.t. } L(\lambda, \mu) = \min_{\boldsymbol{p}(\cdot|s,k) \geq 0} L(\lambda, \mu, \boldsymbol{p}(\cdot|s,k))$$

where the Lagrangian function is

$$\begin{aligned} L(\lambda, \mu, \boldsymbol{p}(\cdot|s,k)) &= r(s,k) + \beta \sum_{s' \in \mathcal{S}} p(s'|s,k) V(s',k) \\ &\quad + \lambda \left(1 - \sum_{s' \in \mathcal{S}} p(s'|s,k)\right) \\ &\quad + \mu \left(\sum_{s' \in \mathcal{S}} p(s'|s,k) \log \left(\frac{p(s'|s,k)}{\hat{p}(s'|s,k)} \right) - \theta \right) \\ &= r(s,k) + \lambda - \mu\theta + \sum_{s' \in \mathcal{S}} \Big(\beta V(s',k) - \lambda \\ &\quad + \mu \log \left(\frac{p(s'|s,k)}{\hat{p}(s'|s,k)} \right) \Big) p(s'|s,k). \end{aligned}$$

The strong duality holds because $\hat{\boldsymbol{p}}(\cdot|s,k)$ is a strictly feasible solution to the problem (A.4) and the Slater condition holds. The first order conditions of the Lagrangian function give

$$\begin{aligned} &\frac{\partial L(\lambda, \mu, \boldsymbol{p}(\cdot|s,k))}{\partial p(s'|s,k)} = \beta V(s',k) - \lambda + \mu \log \left(\frac{p(s'|s,k)}{\hat{p}(s'|s,k)} \right) \\ &\quad + \mu = 0, \ \forall s' \in \mathcal{S} \\ &\quad \Rightarrow p(s'|s,k) = \hat{p}(s'|s,k) \exp \left(\frac{-\beta V(s',k) + \lambda - \mu}{\mu} \right), \\ &\quad \forall s' \in \mathcal{S}. \end{aligned} \tag{A.5}$$

By substituting (A.5) into the Lagrangian function, the dual problem becomes

$$\begin{aligned} \max_{\lambda \text{ free}, \mu \geq 0} L(\lambda, \mu) &= r(s,k) + \lambda - \mu\theta \\ &\quad - \exp \left(\frac{\lambda - \mu}{\mu} \right) \mu \sum_{s' \in \mathcal{S}} \hat{p}(s'|s,k) \\ &\quad \exp \left(\frac{-\beta V(s',k)}{\mu} \right). \end{aligned}$$

Again, the first order conditions give

$$\begin{aligned} \frac{\partial L(\lambda, \mu)}{\partial \lambda} &= 1 - \exp \left(\frac{\lambda - \mu}{\mu} \right) \sum_{s' \in \mathcal{S}} \hat{p}(s'|s,k) \\ &\quad \exp \left(\frac{-\beta V(s',k)}{\mu} \right) = 0 \\ &\quad \Rightarrow \lambda = -\mu \log \left(\sum_{s' \in \mathcal{S}} \hat{p}(s'|s,k) \exp \left(\frac{-\beta V(s',k)}{\mu} \right) \right) + \mu. \end{aligned} \tag{A.6}$$

The dual problem can be rewritten as

$$\max_{\mu \geq 0} L(\mu) = r(s,k) - \mu \log \left(\sum_{s' \in \mathcal{S}} \hat{p}(s'|s,k) \exp \left(\frac{-\beta V(s',k)}{\mu} \right) \right) - \mu\theta.$$

By combining (A.5) and (A.6), we have the worst-case transitional probabilities as

$$p^*(s'|s,k) = \frac{\hat{p}(s'|s,k) \exp\left(-\beta V(s',k)/\mu^*_{sk}\right)}{\sum_{s'' \in \mathcal{S}} \hat{p}(s''|s,k) \exp\left(-\beta V(s'',k)/\mu^*_{sk}\right)}, \quad \forall s' \in \mathcal{S}.$$

where μ^*_{sk} is the optimal solution of the dual problem with given s and k.

A3. Proof of Proposition 2

Let $V^n(s,k) = \max_{a \in \mathcal{A}} w^n(s,k;a)$ denote the value function at the nth iteration of the robust value iteration algorithm in Section 4.3 (Iyengar, 2005). We will show that $V^n(s,k)$ is non-increasing in $s \in \mathcal{S}$ and $k \in \mathcal{K}$ for any integer $n \geq 0$ by induction. Then, the theorem follows because the robust value iteration algorithm converges to an optimal policy.

We set the initial value as $V^0(s,k) = 0$ for all $s \in \mathcal{S}$ and $k \in \mathcal{K}$. First, we show that $V(s,k)$ is non-increasing in $s \in \mathcal{S}$ for all $k \in \mathcal{K}$. Because $V^0(s,k) = 0$ for all $s \in \mathcal{S}$, the induction holds at the initial iteration. Assume that $V^n(s,k)$ is non-increasing in $s \in \mathcal{S}$ for $n = 1, \ldots, N-1$. Let $s', s \in \mathcal{S}$ with $s' > s$ and λ^*_{sk} and μ^*_{sk} be the optimal solution of the dual problems for any give state $(s,k) \in \mathcal{S} \times \mathcal{K}$. We consider two cases at iteration N. If $a = 0$, for ϕ-divergence, we have

$$\begin{aligned} w^N(s',k;0) &= \max_{\mu > 0, \lambda} r(s',k) + \lambda - \mu\theta \\ &\quad - \mu \left(\sum_{s'' \in \mathcal{S}} \hat{p}(s''|s',k) \phi^* \left(\frac{\lambda - \beta V^{N-1}(s'',k)}{\mu} \right) \right) \\ &= r(s',k) + \lambda^*_{s'k} - \mu^*_{s'k}\theta \\ &\quad - \mu^*_{s'k} \left(\sum_{s'' \in \mathcal{S}} \hat{p}(s''|s',k) \phi^* \left(\frac{\lambda^*_{s'k} - \beta V^{N-1}(s'',k)}{\mu^*_{s'k}} \right) \right) \\ &\leq r(s,k) + \lambda^*_{s'k} - \mu^*_{s'k}\theta \\ &\quad - \mu^*_{s'k} \left(\sum_{s'' \in \mathcal{S}} \hat{p}(s''|s',k) \phi^* \left(\frac{\lambda^*_{s'k} - \beta V^{N-1}(s'',k)}{\mu^*_{s'k}} \right) \right) \end{aligned} \tag{A.7}$$

$$\begin{aligned} &\leq r(s,k) + \lambda^*_{s'k} - \mu^*_{s'k}\theta \\ &\quad - \mu^*_{s'k} \left(\sum_{s'' \in \mathcal{S}} \hat{p}(s''|s,k) \phi^* \left(\frac{\lambda^*_{s'k} - \beta V^{N-1}(s'',k)}{\mu^*_{s'k}} \right) \right) \\ &\leq \max_{\mu > 0, \lambda} r(s,k) + \lambda - \mu\theta \\ &\quad - \mu \left(\sum_{s'' \in \mathcal{S}} \hat{p}(s''|s,k) \phi^* \left(\frac{\lambda - \beta V^{N-1}(s'',k)}{\mu} \right) \right) \\ &\quad = w^N(s,k;0) \end{aligned} \tag{A.8}$$

The inequality (A.7) holds because $r(s',k) \leq r(s,k)$. The inequality (A.8) follows Lemma 4.7.2 in Puterman (2014) because $\boldsymbol{P}(\cdot|\cdot,k)$ is IFR and $\phi^*\left(\left(\lambda - \beta V^{N-1}(s,k)\right)/\mu\right)$ is non-decreasing in s due to $V^{N-1}(s,k)$ is non-increasing in s by the induction hypothesis.

If $a = 1$, we have $w^N(s,k;1) = w^N(s',k;1) = -c_r + \beta V^{N-1}(0,k+1)$. Therefore, $w^N(s,k;1)$ is non-increasing in s given k. Similarly, since $w^N(s,k;2) = w^N(s',k;2) = c_s$, $w^N(s,k;2)$ is also non-increasing in s given k. Since $V^N(s,k) = \max_{a \in \mathcal{A}} w^N(s,k;a) \geq \max_{a \in \mathcal{A}} w^N(s',k;a) = V^N(s',k)$, the induction hypothesis holds at iteration N.

Next, we show that $V(s,k)$ is non-increasing in $k \in \mathcal{K}$, $\forall s \in \mathcal{S}$. Because $V^0(s,k) = 0$, $\forall k \in \mathcal{K}$, the induction holds at the initial iteration. Assume for any $s \in \mathcal{S}$, $V^n(s,k)$ is non-increasing in $k \in \mathcal{K}$ for $n = 0, \ldots, N-1$. We consider two cases at iteration N. If $a = 0$, we have

$$\begin{aligned} &w^N(s,k+1;0) \\ &\quad = \max_{\mu > 0, \lambda} r(s,k+1) + \lambda - \mu\theta \\ &\quad - \mu \left(\sum_{s' \in \mathcal{S}} \hat{p}(s'|s,k+1) \phi^* \left(\frac{\lambda - \beta V^{N-1}(s',k+1)}{\mu} \right) \right) \end{aligned}$$

$$= r(s,k+1) + \lambda^*_{s,k+1} - \mu^*_{s,k+1}\theta - \mu^*_{s,k+1}\left(\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k+1)\phi^*\left(\frac{\lambda^*_{s,k+1} - \beta V^{N-1}(s',k+1)}{\mu^*_{s,k+1}}\right)\right)$$
$$\le r(s,k) + \lambda^*_{s,k+1} - \mu^*_{s,k+1}\theta - \mu^*_{s,k+1}\left(\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k+1)\phi^*\left(\frac{\lambda^*_{s,k+1} - \beta V^{N-1}(s',k)}{\mu^*_{s,k+1}}\right)\right) \tag{A.9}$$

$$\le r(s,k) + \lambda^*_{s,k+1} - \mu^*_{s,k+1}\theta - \mu^*_{s,k+1}\left(\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k)\phi^*\left(\frac{\lambda^*_{s,k+1} - \beta V^{N-1}(s',k)}{\mu^*_{s,k+1}}\right)\right)$$
$$\le \max_{\mu>0,\lambda}\ r(s,k) + \lambda - \mu\theta - \mu\log\left(\sum_{s'\in\mathcal{S}} \hat{p}(s''|s,k)\phi^*\left(\frac{\lambda - \beta V^{N-1}(s',k)}{\mu}\right)\right)$$
$$= w^N(s,k;0) \tag{A.10}$$

The inequality (A.9) holds because $r(s,k+1)\le r(s,k)$ and $V^{N-1}(s,k+1)\le V^{N-1}(s,k)$ by the induction hypothesis. The inequality (A.10) follows Lemma 4.7.2 in Puterman (2014) because $\boldsymbol{P}(\cdot|\cdot,k+1)\succeq \boldsymbol{P}(\cdot|\cdot,k)$ by Assumption 2(b) and $\phi^*\big(\big(\lambda-\beta V^{N-1}(s,k)\big)/\mu\big)$ is non-decreasing in s due to $V^{N-1}(s,k)$ is non-increasing in s by the induction hypothesis.

If $a=1$, we have $w^N(s,k;1) = -c_r + \beta V^{N-1}(0,k+1) \ge -c_r + \beta V^{N-1}(0,k+2) = w^N(s,k+1;1)$. Therefore, $w^N(s,k;1)$ is non-increasing in k for all $s\in\mathcal{S}$. Similarly, since $w^N(s,k;2) = w^N(s,k+1;2) = c_s$, $w^N(s,k;2)$ is also non-increasing in k for all $s\in\mathcal{S}$. Since $V^N(s,k) = \max_{a\in\mathcal{A}} w^N(s,k;a) \ \ge \max_{a\in\mathcal{A}} w^N(s,k+1;a) = V^N(s,k+1)$, the induction hypothesis holds at iteration N.

A4. Proof of Theorem 1

We first show that the optimal policy is of control-limit type for all $k\in\mathcal{K}$. Let $s' > s$. We consider two cases: (i) If $a^*(s,k)=1$, then $V(s,k) = w(s,k;1) = -c_r + \beta V(0,k+1) = w(s',k;1) \le V(s',k)$. Because $V(s,k)$ is non-increasing in s for all $k\in\mathcal{K}$, $V(s,k)\ge V(s',k)$. Thus, we have $V(s',k) = w(s',k;1)$ and $a^*(s',k)=1$. (ii) If $a^*(s,k)=2$, then $V(s,k) = w(s,k;2) = c_s = w(s',k;2)$, and by Proposition 2, $V(s,k)\ge V(s',k)$, we have $V(s',k) = w(s',k;2)$ and $a^*(s',k)=2$.

Next, we show the existence of the threshold k^*. This is equivalent to show that if $a^*(s,k)=2$ for some k, then $a^*(s,k+1)=2$. Since $V(s,k) = w(s,k;2) = c_s = w(s,k+1;2) \le V(s,k+1)$ and $V(s,k)\ge V(s,k+1)$, we have $V(s,k+1) = w(s,k+1;2)$ and hence $a^*(s,k+1)=2$.

A5. Proof of Theorem 2

We first prove that $\zeta_{\rm rm}(k)$ is non-increasing in k, $\forall k\in\{0,\ldots,k^*-1\}$. This is equivalent to show that $a^*(s,k+1)=1$ if $a^*(s,k)=1$ $\forall k\in\{0,\ldots,k^*-2\}$. We prove this by contradiction.

Suppose $a^*(s,k)=1$ but $a^*(s,k+1)=0$ for some $s\in\mathcal{S}$ and $k\in\{0,\ldots,k^*-2\}$. Then, we have $w(s,k;1)\ge w(s,k;0)$, $w(s,k+1;1) < w(s,k+1;0)$ and hence,

$$w(s,k;1) - w(s,k+1;1) > w(s,k;0) - w(s,k+1;0). \tag{A.11}$$

For ϕ-divergence, the right hand side (RHS) of Eq. (A.11) be rewritten as

$$\text{RHS} = r(s,k) + \max_{\mu>0,\lambda}\Big(\lambda - \mu\theta - \mu\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k)\phi^*\left(\frac{\lambda-\beta V(s',k)}{\mu}\right)\Big)$$
$$- r(s,k+1) - \max_{\mu>0,\lambda}\Big(\lambda - \mu\theta - \mu\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k+1)\phi^*\left(\frac{\lambda-\beta V(s',k+1)}{\mu}\right)\Big)$$
$$\ge r(s,k) + \lambda^*_{s,k+1} - \mu^*_{s,k+1}\theta - \mu^*_{s,k+1}\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k)\phi^*\left(\frac{\lambda^*_{s,k+1}-\beta V(s',k)}{\mu^*_{s,k+1}}\right)$$
$$- r(s,k+1) - \Big(\lambda^*_{s,k+1} - \mu^*_{s,k+1}\theta - \mu^*_{s,k+1}\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k+1)\phi^*\left(\frac{\lambda^*_{s,k+1}-\beta V(s',k+1)}{\mu^*_{s,k+1}}\right)\Big)$$
$$\ge r(s,k) - r(s,k+1) + \mu^*_{s,k+1}\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k+1)\phi^*\left(\frac{\lambda^*_{s,k+1}-\beta V(s',k)}{\mu^*_{s,k+1}}\right) - \mu^*_{s,k+1}\sum_{s'\in\mathcal{S}} \hat{p}(s'|s,k)\phi^*\left(\frac{\lambda^*_{s,k+1}-\beta V(s',k)}{\mu^*_{s,k+1}}\right)$$
$$\ge r(s,k) - r(s,k+1), \tag{A.12}$$

where inequality (A.12) follows Lemma 4.7.2 in Puterman (2014) because $\phi^*\left(\dfrac{\lambda-\beta V(s,k)}{\mu}\right)$ is non-decreasing in $s\in\mathcal{S}$ and $\hat{\boldsymbol{P}}(\cdot|\cdot,k+1)\succeq \hat{\boldsymbol{P}}(\cdot|\cdot,k)$ in Assumption 2. The left hand side (LHS) of Eq. (A.11) be rewritten as

$$\text{LHS} = -c_r + \beta V(0,k+1) + c_r - \beta V(0,k+2) \le \beta V(0,k+1) - \beta c_s \le \frac{\beta r(0,0)}{1-\beta} - \beta c_s, \tag{A.13}$$

where the first inequality holds because $V(0,k+2)\ge w(s,k+2;2) = c_s$, and the second inequality holds because $V(0,k+1)\le \sum_{t=0}^{\infty}\beta^t r(0,0) = r(0,0)/(1-\beta)$. By (A.12) and (A.13), we have $\beta r(0,0)/(1-\beta) - \beta c_s \ge r(s,k) - r(s,k+1)$, which violates condition in Theorem 2(a) and implies that $a^*(s,k+1)=1$ if $a^*(s,k)=1$.

Now we show that $\zeta_{\rm rm}(k^*-1)\ge \zeta_{\rm scrap}(k^*)$. This is equivalent to show that $a^*(\zeta_{\rm rm}(k^*-1),k^*)=2$. From the proof above, we can easily show that $w(\zeta_{\rm rm}(k^*-1),k^*;1)\ge w(\zeta_{\rm rm}(k^*-1),k^*;0)$. By the definition of k^*, there exists a s' that $a^*(s',k^*)=2$. Therefore, $w(s,k^*;2)\ge w(s,k^*;1)$ for all $s\in\mathcal{S}$. Thus, we have $w(\zeta_{\rm rm}(k^*-1),k^*;2)\ge w(\zeta_{\rm rm}(k^*-1),k^*;1)\ge w(\zeta_{\rm rm}(k^*-1),k^*;0)$, which shows $a^*(\zeta_{\rm rm}(k^*-1),k^*)=2$. It is straightforward that $\zeta_{\rm scrap}(k)$ is non-increasing in $k\in\mathcal{K}$ because $a^*(s,k+1)=2$ if $a^*(s,k)=2$ as shown in the proof of Theorem 1.

A6. Proof of Proposition 3

For notational convenience, we drop s and k. The value function defined in (RRmPO) involves solving an inner problem for any given $s\in\mathcal{S}$ and $k\in\mathcal{K}$ as follows

$$w(s,k;0) = \min\ r + \beta\sum_{s'\in\mathcal{S}} p(s')V(s')$$
$$\text{s.t.}\quad \sum_{s'\in\mathcal{S}} p(s') = 1,\ W_m^m\big(\boldsymbol{p},\hat{\boldsymbol{p}}\big)\le\theta^m,\ p(s')\ge 0, s'\in\mathcal{S}. \tag{A.14}$$

The Lagrangian dual problem is

$$v_D = \max_{\mu \geq 0} L(\mu) \quad \text{s.t. } L(\mu) = \min_{\boldsymbol{p} \geq 0, \sum_{s'} p(s')=1} L(\mu, \boldsymbol{p})$$

where the Lagrangian function is

$$L(\mu, \boldsymbol{p}) = \sum_{s'} p(s')V(s') + \mu\left(W_m^m(\boldsymbol{p}, \hat{\boldsymbol{p}}) - \theta^m\right)$$
$$= -\mu\theta^m + \sum_{s'} p(s')V(s') + \mu W_m^m(\boldsymbol{p}, \hat{\boldsymbol{p}}).$$

For dual objective function, we have

$$L(\mu) = \min_{\boldsymbol{p} \geq 0, \sum_{s'} p(s')=1} -\mu\theta^m + \sum_{s'} p(s')V(s') + \mu W_m^m(\boldsymbol{p}, \hat{\boldsymbol{p}}) \tag{A.15}$$

$$= -\mu\theta^m + \min_{\boldsymbol{p} \geq 0, \sum_{s'} p(s')=1} \Bigg\{ \sum_{s'} p(s')V(s') + \mu \max_{u \in L^1(\boldsymbol{p}), v \in L^1(\hat{\boldsymbol{p}})} \Bigg\{ \sum_{x \in \mathcal{S}} u(x)p(x) + \sum_{y \in \mathcal{S}} v(y)\hat{p}(y) : v(y) \leq \min_{x \in \mathcal{S}} \{d(x, y)^m - u(x)\}, \forall y \in \mathcal{S} \Bigg\} \Bigg\} \tag{A.16}$$

Let $u = -\boldsymbol{V}/\mu$ for $\mu > 0$, then $u \in L^1(\boldsymbol{p})$ and $v(\cdot) = \min_{x \in \mathcal{S}}\{d(x, \cdot)^m - u(x)\} \in L^1(\hat{\boldsymbol{p}})$. Thus,

$$L(\mu) \geq -\mu\theta^m + \min_{\boldsymbol{p} \geq 0} \Bigg\{ \mu \sum_{y \in \mathcal{S}} \hat{p}(y) \min_{x \in \mathcal{S}} \{d(x, y)^m + V(x)/\mu\} \Bigg\}$$
$$= -\mu\theta^m + \mu \sum_{y \in \mathcal{S}} \hat{p}(y) \min_{x \in \mathcal{S}} \{d(x, y)^m + V(x)/\mu\} = L'(\mu)$$

Based on Theorem 1 of Gao & Kleywegt (2016), there exists an optimizer μ^* such that $L(\mu^*) = L'(\mu^*)$ and $v_P = v_D$, i.e., strong duality holds.

A7. Proof of Proposition 4

To prove $V(s, k)$ is non-increasing in $s \in \mathcal{S}$, the key step is to show $w^N(s', k; 0) \leq w^N(s, k; 0)$ at iteration N. Therefore, it suffices to show inequality (A.8) holds for Wasserstein-distance-based ambiguity sets, which is equivalent to show $g(y) = \min_x g(x, y)$ is non-increasing in y, where $g(x, y) = d(x, y) + V(x)$ (k is dropped here). To see this, let $y_1 < y_2$ and $x_a = \arg\min_x g(x, y_a)$ for $a = 1, 2$. If $y_2 \leq x_1$, we have $g(y_2) \leq g(x_1, y_2) = d(x_1, y_2) + V(x_1) < d(x_1, y_1) + V(x_1) = g(y_1)$. Otherwise if $y_2 > x_1$, we have $g(y_2) \leq g(y_2, y_2) = V(y_2) \leq V(x_1) < d(x_1, y_1) + V(x_1) = g(y_1)$.

Since $g(y)$ is non-increasing in y, the inequality (A.10) holds for Wasserstein-distance ambiguity sets, which leads to $w^N(s, k+1; 0) \leq w^N(s, k; 0)$. Therefore, $V(s, k)$ is non-increasing in $k \in \mathcal{K}$.

Appendix B

B1. Experiment parameters

The following table provides the experiment parameters used in the experiment that examines the existence of control limit policies when the condition of Theorem 2(a) is violated. Note that for the ease of parameter control, we redefine the reward as $r(s, k) = a_0 - a_1 k - a_2 s$. Parameter values are drawn from their respective uniform distributions.

a_0	a_1	a_2	c_r	c_s	θ	β
$U(10, 50)$	$U(1, 15)$	$U(1, 15)$	$U(0, 10)$	$U(0, 10)$	$U(0, 2)$	$U(0.01, 0.99)$

B2. Data-driven decision process

Algorithm B.1 Data-driven decision-making process.

Input: Sensor data collected by continuous monitoring the operation of units after the kth remanufacturing before the $(k+1)$th remanufacturing, $k = 0, 1, \ldots$

1: Data Processing
(a): Data Compression. Apply a data compression technique (e.g., PCA) to reduce the high-dimensional data to one dimension to represent the overall health of a component; discretize the health into several intervals representing different discrete states.
(b): Data Preparation. Separate the entire dataset into a training set $\mathcal{N}$ and a test set $\mathcal{M}$
(c): Parameter Estimation. Use a subset of the training dataset $\mathcal{N}$ to obtain the nominal transition probability $\hat{\boldsymbol{P}}_k$
2: Hyperparameter Tuning
for each θ **do**
(a): Construct the ambiguity set using the nominal transition probability $\hat{\boldsymbol{P}}_k$ and the radius θ.
(b): Solve the robust MDP under the constructed ambiguity set and obtain a robust policy $\pi(\theta)$
(c): Evaluate the out-of-sample performance $\bar{v}(\theta)$ of $\pi(\theta)$ using the remaining data in set $\mathcal{N}$
end for
3: Obtain Data-Driven Solution
(a): Select the optimal θ^* out of all candidate θ's in Step 2 that maximizes $\bar{v}(\theta)$
(b): Obtain the data-driven solution $\pi(\theta^*)$
4: *Evaluate the out − of − sample performance of the data−driven solution*
for each unit $i \in \mathcal{M}$ **do**
(a): Obtain the nominal transition probability $\hat{\boldsymbol{P}}_k^i$ for each k
(b): Obtain the reward v_i by implementing $\pi(\theta^*)$ for unit i
end for
(c): Calculate the average reward $\pi(\theta^*)$ for all $i \in \mathcal{M}$

References

Abdallah, T., Farhat, A., Diabat, A., & Kennedy, S. (2012). Green supply chains with carbon trading and environmental sourcing: Formulation and life cycle assessment. *Applied Mathematical Modelling, 36*(9), 4271–4285.
Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science, 59*(2), 341–357.
Chandler, D. L. (2011). When is it worth remanufacturing?
Delage, E., & Mannor, S. (2010). Percentile optimization for Markov decision processes with parameter uncertainty. *Operations Research, 58*(1), 203–213.
Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research, 58*(3), 595–612.
Elwany, A. H., Gebraeel, N. Z., & Maillart, L. M. (2011). Structured replacement policies for components with complex degradation processes and dedicated sensors. *Operations Research, 59*(3), 684–695.
Esfahani, P. M., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming, 171*(1–2), 115–166.
Ford, S., & Despeisse, M. (2016). Additive manufacturing and sustainability: An exploratory study of the advantages and challenges. *Journal of Cleaner Production, 137*, 1573–1587.
Fouladirad, M., Paroissin, C., & Grall, A. (2018). Sensitivity of optimal replacement policies to lifetime parameter estimates. *European Journal of Operational Research, 266*(3), 963–975.
Frederick, D. K., DeCastro, J. A., & Litt, J. S. (2007). User's guide for the commercial modular aero-propulsion system simulation (C-MAPSS).
Galbreth, M. R., & Blackburn, J. D. (2010). Optimal acquisition quantities in remanufacturing with condition uncertainty. *Production and Operations Management, 19*(1), 61–69.
Gao, R., & Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. arXiv:1604.02199.

Goh, J., Bayati, M., Zenios, S. A., Singh, S., & Moore, D. (2018). Data uncertainty in Markov chains: Application to cost-effectiveness analyses of medical innovations. *Operations Research, 66*(3), 697–715.
Gutowski, T. G., Sahni, S., Boustani, A., & Graves, S. C. (2011). Remanufacturing and energy savings. *Environmental science and Technology, 45*(10), 4540–4547.
Hanasusanto, G. A., & Kuhn, D. (2018). Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research, 66*(3), 849–869.
Ijomah, W. L., McMahon, C. A., Hammond, G. P., & Newman, S. T. (2007). Development of design for remanufacturing guidelines to support sustainable manufacturing. *Robotics and Computer-Integrated Manufacturing, 23*(6), 712–719.
Iyengar, G. N. (2005). Robust dynamic programming. *Mathematics of Operations Research, 30*(2), 257–280.
de Jonge, B., Dijkstra, A. S., & Romeijnders, W. (2015). Cost benefits of postponing time-based maintenance under lifetime distribution uncertainty. *Reliability Engineering and System Safety, 140*, 15–21.
Kim, M. J. (2016). Robust control of partially observable failing systems. *Operations Research, 64*(4), 999–1014.
Kim, M. J., & Makis, V. (2013). Joint optimization of sampling and control of partially observable failing systems. *Operations Research, 61*(3), 777–790.
Kurt, M., & Kharoufeh, J. P. (2010). Optimally maintaining a Markovian deteriorating system with limited imperfect repairs. *European Journal of Operational Research, 205*(2), 368–380.
Lam, H. (2019). Recovering best statistical guarantees via the empirical divergence-based distributionally robust optimization. *Operations Research, 67*(4), 1090–1105.
Moghaddass, R., & Zuo, M. J. (2014). An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process. *Reliability Engineering and System Safety, 124*, 92–104.
Nilim, A., & El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Operations Research, 53*(5), 780–798.
Omshi, E. M., Grall, A., & Shemehsavar, S. (2020). A dynamic auto-adaptive predictive maintenance policy for degradation with unknown parameters. *European Journal of Operational Research, 282*(1), 81–92.
Örsdemir, A., Kemahlıoğlu-Ziya, E., & Parlaktürk, A. K. (2014). Competitive quality choice and remanufacturing. *Production and Operations Management, 23*(1), 48–64.
Östlin, J., Sundin, E., & Björkman, M. (2009). Product life-cycle implications for remanufacturing strategies. *Journal of Cleaner Production, 17*(11), 999–1009.
Parker, D., Riley, K., Robinson, S., Symington, H., Tewson, J., Jansson, K., Ramkumar, S., & Peck, D. (2015). Remanufacturing market study.
Puterman, M. L. (2014). *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons.
Russell, J., & Nasr, N. (2019). Value-retention processes within the circular economy. In *Remanufacturing in the circular economy* (pp. 1–29). Wiley Online Library.
Satia, J. K., & Lave, R. E., Jr. (1973). Markovian decision processes with uncertain transition probabilities. *Operations Research, 21*(3), 728–740.
Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 international conference on prognostics and health management* (pp. 1–9). IEEE.
Silver, E. A. (1963). Markovian decision processes with uncertain transition probabilities or rewards. *Technical report*. Massachusetts Inst of Tech Cambridge Operations Research Center.
Song, S., Liu, M., Ke, Q., & Huang, H. (2015). Proactive remanufacturing timing determination method based on residual strength. *International Journal of Production Research, 53*(17), 5193–5206.
Sutherland, J. W., Adler, D. P., Haapala, K. R., & Kumar, V. (2008). A comparison of manufacturing and remanufacturing energy intensities with application to diesel engine production. *CIRP Annals, 57*(1), 5–8.
Wang, Y., Hu, J., Ke, Q., & Song, S. (2016). Decision-making in proactive remanufacturing based on online monitoring. *Procedia CIRP, 48*, 176–181.
White, C. C., III, & El-Deib, H. K. (1986). Parameter imprecision in finite state, finite action dynamic programs. *Operations Research, 34*(1), 120–129.
White, C. C., III, & Eldeib, H. K. (1994). Markov decision processes with imprecise transition probabilities. *Operations Research, 42*(4), 739–749.
Wiesemann, W., Kuhn, D., & Rustem, B. (2013). Robust Markov decision processes. *Mathematics of Operations Research, 38*(1), 153–183.
Yang, L., Hu, Y., & Huang, L. (2020). Collecting mode selection in a remanufacturing supply chain under cap-and-trade regulation. *European Journal of Operational Research, 287*(2), 480–496.
Yang, S. (2020). Global challenges and market transformation in support of remanufacturing. In *Remanufacturing in the circular economy: Operations, engineering and logistics* (pp. 169–207). Wiley Online Library.
Zhang, Z., Matsubae, K., & Nakajima, K. (2021). Impact of remanufacturing on the reduction of metal losses through the life cycles of vehicle engines. *Resources, Conservation and Recycling, 170*, 105614.