Annals of Operations Research

A data-driven robust approach to a problem of optimal replacement in maintenance

--Manuscript Draft--

Manuscript Number:	ANOR-D-24-03249
Full Title:	A data-driven robust approach to a problem of optimal replacement in maintenance
Article Type:	Original Research
Keywords:	Robust Markov Decision Processes, Data-Driven Uncertainty, Optimal Replacement, Stochastic Degradation Modeling.
Corresponding Author:	Sina Shahri Majarshin Eindhoven University of Technology: Technische Universiteit Eindhoven Eindhoven, North Brabant NETHERLANDS, KINGDOM OF THE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Eindhoven University of Technology: Technische Universiteit Eindhoven
Corresponding Author's Secondary Institution:	
First Author:	Sina Shahri Majarshin
First Author Secondary Information:	
Order of Authors:	Sina Shahri Majarshin
	Ahmadreza Marandi
	Claudia Fecarotti
	Geert-Jan van Houtum
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Maintenance strategies are pivotal in ensuring the reliability and performance of critical components within industrial machines and systems. However, accurately determining the optimal replacement time for such components under stress and deterioration remains a complex task due to inherent uncertainties and variability in operating conditions. In this paper, we propose a comprehensive approach based on Robust Markov Decision Processes (RMDP) to optimize component replacement decisions in machines with one critical component while addressing uncertainty in a structured manner. RMDP offers a robust framework for decision-making under uncertainty, allowing for the modeling of component degradation and variability in operating conditions. Our methodology uses data-driven ambiguity sets, including likelihood-based and Kullback-Leibler (KL)-based ambiguity sets, to capture and quantify uncertainty in the degradation process. We show the mathematical relationship between the KL-based and Likelihood-based ambiguity sets and provide statistical guarantees for the optimal cost. Through computational experiments, we demonstrate the effectiveness of our RMDP approach in identifying the optimal replacement time that minimizes the total maintenance cost while exhibiting greater stability compared to traditional methods.

A data-driven robust approach to a problem of optimal replacement in maintenance

Sina Shahri Majarshin[1*], Ahmadreza Marandi[1,2], Claudia Fecarotti[1], Geert-Jan van Houtum[1]

[1]Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands.
[2]Department of Management, University of British Columbia, Kelowna, Canada.

*Corresponding author(s). E-mail(s): s.shahri.majarshin@tue.nl;
Contributing authors: a.marandi@tue.nl; c.fecarotti@tue.nl;
g.j.v.Houtum@tue.nl;

Abstract

Maintenance strategies are pivotal in ensuring the reliability and performance of critical components within industrial machines and systems. However, accurately determining the optimal replacement time for such components under stress and deterioration remains a complex task due to inherent uncertainties and variability in operating conditions. In this paper, we propose a comprehensive approach based on Robust Markov Decision Processes (RMDP) to optimize component replacement decisions in machines with one critical component while addressing uncertainty in a structured manner. RMDP offers a robust framework for decision-making under uncertainty, allowing for the modeling of component degradation and variability in operating conditions. Our methodology uses data-driven ambiguity sets, including likelihood-based and Kullback-Leibler (KL)-based ambiguity sets, to capture and quantify uncertainty in the degradation process. We show the mathematical relationship between the KL-based and Likelihood-based ambiguity sets and provide statistical guarantees for the optimal cost. Through computational experiments, we demonstrate the effectiveness of our RMDP approach in identifying the optimal replacement time that minimizes the total maintenance cost while exhibiting greater stability compared to traditional methods.

Keywords: Robust Markov Decision Processes, Data-Driven Uncertainty, Optimal Replacement, Stochastic Degradation Modeling.

Conflicts of interest/Competing interests: The authors have no conflicts of interest to declare that are relevant to the content of this article.

1 Introduction

In industrial settings, the effective management of maintenance activities for critical components is essential for ensuring the reliability and efficiency of machinery and equipment. A key challenge in this domain lies in determining the optimal replacement time for components subject to deterioration, as it directly impacts system reliability, and overall performance (Teixeira et al, 2020). This paper addresses this challenge within the context of multi-component systems, where only the most critical component is monitored, and its progression of damage and deterioration serves as a primary indicator of system health. Specifically, we focus on optimizing replacement decisions for the critical component in the face of stochastic deterioration, periodic inspections, and state-dependent costs.

Many multi-component systems rely heavily on the functionality of one critical component to maintain operation. For example, in wind turbines, the gearbox, which transmits mechanical energy from the rotor to the generator, often acts as the critical component (Salameh et al, 2018). These systems are highly complex, and the failure of the critical component can lead to significant downtime, production losses, and expensive repairs (Lamghari-Idrissi et al, 2022). The progression of damage in this critical component directly influences the overall system's state, reflecting its cumulative degradation from wear, aging, and environmental conditions. Monitoring the degradation of the critical component is essential for making informed maintenance decisions, ensuring system reliability, and optimizing performance.

In this research, we specifically address the challenge of optimizing replacement decisions for the critical component within the system. The decision to replace the component or continue its operation is influenced by various factors, including the observed degradation level during periodic inspections, the associated operational costs, and the replacement cost. The deterioration of the component is inherently stochastic, meaning that the rate and extent of degradation vary over time and are subject to uncertainty. To assess the condition of the component and make timely maintenance decisions, periodic inspections are conducted at predetermined intervals. These inspections provide valuable information about the current state of the component and guide the decision-making process regarding whether to replace the component or allow it to continue operating.

The optimal replacement problem for components subject to stochastic degradation is usually addressed in the literature under the assumption that the degradation process can be modeled as a stochastic process with well defined parameters. Indeed, the single unit replacement problem has been modeled as a Markov Decision Process, which provide a valuable framework for sequential decision making. However, a drawback of such approaches is that they rely on the assumption that accurate

information is available to determine transition probabilities which accurately reflect the degradation process. In many real cases however, obtaining such precise estimates of the process parameters is challenging or even not possible. The errors in estimating the transition matrix could hugely influence the results attained by the MDP, which has been emphasized by Abbad et al (1992); Feinberg et al (2002); Mannor et al (2007), and Nilim and El Ghaoui (2005). This limitation underscores the need for more conservative and resilient decision-making approaches, particularly in contexts where robustness and risk-aversion is crucial.

To address this challenge, our research aims to utilize Robust Markov Decision Processes (RMDPs). Unlike traditional MDPs, which rely on known and fixed probability matrices for state transitions, RMDPs can leverage data to create conservative decision-making frameworks that account for uncertain transitions (Ramani and Ghate, 2022). Traditional MDP solutions are often sensitive to model parameters, requiring caution amidst changes or uncertainties. RMDPs introduce robustness by addressing these uncertainties through the use of ambiguity sets (Iyengar, 2005; Kumar et al, 2022). An ambiguity set is a collection of probability distributions that represent uncertainty or lack of complete knowledge about the true underlying distribution. By considering a range of plausible distributions and transition matrices within an ambiguity set, RMDPs optimize performance while steering clear of overly optimistic assumptions about the distributions that underestimate how expensive the total cost could be. This ensures a more reliable hedge against uncertainties. Additionally, using data-driven ambiguity sets, RMDPs can harness historical data, which in turn allows us to incorporate uncertainty into the optimization process and offer the potential for more effective and reliable maintenance strategies. This innovative approach marks a departure from conventional methodologies and holds promise for enhancing the reliability and efficiency of industrial maintenance practices.

While the application of RMDP to maintenance has been explored previously (Delage and Mannor, 2010; Goyal and Grand-Clément, 2023; Kim, 2016; Wiesemann et al, 2013), we extend the RMDP framework to a specific problem of optimal replacement in maintenance (see Section 3), providing a robust solution to this critical problem. Our work focuses on a finite horizon, where the end of the horizon represents the lifetime of the system or machine in which the critical component is installed. This approach allows for more precise planning and optimization over the entire lifespan of the system. Drawing inspiration from Nilim and El Ghaoui (2005), we introduce two types of ambiguity sets: one based on the distance from the maximum likelihood estimate (see (12) and its discussion) and another based on the Kullback-Leibler (KL) divergence (see (27) and its discussion). The novelty of our work lies in establishing a clear connection between KL-based and likelihood-based ambiguity sets, demonstrating how these two formulations are mathematically related. Next, we establish statistical bounds that provide confidence that the ambiguity set includes the true unknown distribution. This ensures that the actual cost incurred by applying the robust policy will not exceed a calculable value, offering practical assurance to decision-makers.

2 Literature review

Maintenance strategy optimization has been extensively reviewed in the literature, with key works highlighting advancements in inspection scheduling, degradation modeling, and system configurations (Arts et al, 2024; De Jonge and Scarf, 2020). These studies point out the ongoing challenge of ensuring system reliability while minimizing costs in complex and uncertain environments. De Jonge and Scarf (2020) provided a comprehensive review, categorizing strategies by inspection schedules, degradation processes, and system types. They emphasize the critical importance of addressing model uncertainty in stochastic degradation models, noting this area remains under-explored despite its practical significance. Arts et al (2024) offered a retrospective review of fifty years of advancements in maintenance optimization, introducing a maturity framework based on data availability and decision-making sophistication. Their work highlights recent advances in Condition-Based Maintenance (CBM), a degradation-based maintenance model that uses real-time condition data to monitor system health and make informed maintenance decisions. These advancements include the integration of machine learning and artificial intelligence techniques to improve degradation prediction and enhance decision-making accuracy.

CBM has emerged as a prominent strategy leveraging degradation data to optimize maintenance decisions. As defined by Arts (2017), CBM involves gauging the actual condition of a part and conducting maintenance based on this, either through periodic inspections or continuous monitoring via sensors or process control software. Alaswad and Xiang (2017) attributed CBM's growing popularity to advancements in degradation models, distinguishing between perfect and imperfect inspections. The latter motivates the use of Partially Observable Markov Decision Processes (POMDPs) for handling uncertainties. According to Quatrini et al (2020), CBM is regarded as the optimal maintenance strategy when failure or degradation could result in significant economic losses. The authors emphasize that implementing an effective CBM strategy can lead to substantial advantages for industrial companies, such as improved system uptime and decreased maintenance expenses. Building on these concepts, Teixeira et al (2020) highlighted the integration of Internet of Things (IoT) technologies into CBM, enabling more dynamic and adaptive maintenance strategies. However, their review focuses primarily on implementation aspects rather than theoretical foundations of maintenance decision models. While these reviews emphasize the importance of CBM strategies, they also highlight the need for decision-making tools to address the complexities and uncertainties in maintenance optimization.

MDPs are widely used in maintenance optimization to model stochastic degradation and guide sequential decision-making (Arts et al, 2024). One of the prominent approaches to developing a CBM strategy is through the application of MDPs (Arts, 2017). Early studies, such as the ones done by Derman (1963); Kolesar (1966); Ross (1969) focused on conditions for optimal threshold policies, assuming perfect knowledge of transition probabilities. Amari et al (2006) extended this approach by using MDPs to optimize inspection schedules and maintenance actions, accounting for

trade-offs between inspection costs and maintenance decisions. Their work demonstrated the flexibility of MDPs in CBM settings but still relied on the assumption of complete knowledge of system dynamics, limiting its robustness under uncertainty. To address imperfect information, Smallwood and Sondik (1973) and Rosenfield (1976) introduced POMDPs, where states must be inferred through costly inspections. Van Oosterom et al (2017) further expanded this framework to account for heterogeneous components with varying but known transition matrices and unobservable types, highlighting the complexity of systems that work with many types of components. Despite these advancements, most models assume complete knowledge of system parameters, limiting their applicability under real-world uncertainties. These gaps underscore the need for robust frameworks like RMDPs, which incorporate parameter uncertainty through ambiguity sets. Throughout this paper, we explore how RMDPs address these challenges, providing a robust foundation for CBM with unknown parameters.

The limitations of traditional MDPs in handling parameter uncertainty have led to the development of RMDPs. Mannor et al (2007) explored the impact of finite data on value function estimation in MDPs, highlighting how errors in probability estimation can significantly affect the resulting value function. This work underscored the need for more robust approaches to decision-making under uncertainty. Iyengar (2005) and Nilim and El Ghaoui (2005) established the foundation of modern RMDPs by introducing robust dynamic programming with ambiguity sets, such as divergences and likelihood regions, and emphasizing the rectangularity assumption for computational tractability. Rectangularity assumes independence across state-action pairs, and can result in conservative policies. Subsequent efforts, such as those by Delage and Mannor (2010); Xu and Mannor (2012), sought to address this conservatism. Delage and Mannor (2010) introduced chance-constrained MDPs, treating unknown parameters as random variables and formulating a distribution over them, while Xu and Mannor (2012) introduced Distributionally Robust MDPs (DRMDPs) that utilize probabilistic information and nested uncertainty sets. To further refine the trade-off between realism and computational feasibility, Wiesemann et al (2013) explored non-rectangular ambiguity sets, offering hierarchical approximations for infinite-horizon problems. Mannor et al (2016) expanded this line of inquiry by introducing K-rectangular ambiguity sets, which relax the independence assumption of rectangularity while preserving tractability. This innovation marked a key step in broadening the applicability of RMDPs to more complex, interdependent systems.

Recent advancements in RMDP theory have focused on reducing conservatism and improving computational efficiency, Goyal and Grand-Clément (2023) allowed for dependency in transition probabilities by expressing them as a linear function of a factor matrix, Grand-Clément and Petrik (2023) introduced a convex optimization formulation for RMDPs, providing new solution methods. Ou and Bi (2024) provided a comprehensive review of robust optimization in sequential decision-making under uncertainty. They discussed the limitations of traditional MDPs, the development of RMDPs, and the importance of the rectangularity assumption and its relaxations.

Their review also highlighted how RMDP theory has given rise to related techniques such as safe learning and dynamic risk measures.

Our research focuses on applying RMDP to specific replacement problems within CBM, addressing uncertainties in transition probabilities while utilizing perfect inspection information. We explore the finite horizon case for RMDP, establishing a connection between Kullback-Leibler (KL) divergence-based and likelihood-based ambiguity sets, and demonstrating their mathematical relationship. Our approach provides statistical bounds for the expected total cost when robust policies are implemented throughout the lifetime of the machine. This work applies established RMDP methodologies to develop optimal replacement policies within the framework of CBM. By adapting these methods to the context of CBM, we provide a balanced approach that integrates robustness and practicality in maintenance decision-making, offering implementable tools for real-world scenarios.

3 Problem description and assumptions

Let us consider a machine or a system with a critical component subject to degradation. The system has a finite lifespan and is inspected periodically at predetermined points in time to assess the state of the critical component. The objective is to determine the optimal times for replacing the component or allowing it to continue operating, aiming to minimize the expected costs associated with degradation-related impacts on production quality and replacement actions. At each inspection time, if replacement is deemed necessary based on the observed degradation level, it is executed immediately to prevent potential failures. Failures occurring between inspections remain undetected until the next scheduled inspection. To be able to analyze the system and different maintenance policies, we will introduce the following notation.

The set of all inspection times is denoted by $\mathcal{T} = \{0, 1, \ldots, N-1\}$ where $N < \infty$ denotes the end of the time horizon. The condition of the component at time $t \in \mathcal{T} \cup \{N\}$ is represented by the random variable X_t, reflecting its health and performance level. This random variable ranges from an initial state of 0, indicating the component is new and fully functional, to a terminal state l, signifying failure. The set of the component's degradation levels (states) is denoted by $\mathcal{S} = \{0, 1, ..., l\}$, which is an ordered set. We also assume that the initial state of the component is 0, meaning $X_0 = 0$. At each inspection time, two possible actions can be taken: "replace" or "do nothing", represented by the set $\mathcal{A} = \{0, 1\}$. The action at time $t \in \mathcal{T}$ is denoted by $a_t \in \mathcal{A}$, where $a_t = 1$ indicates the decision to replace, and $a_t = 0$ indicates the decision to do nothing. There are two types of costs associated with the system that we study in this paper. The operational cost and the action cost. The operational cost, denoted by the function $c : \mathcal{S} \mapsto \mathbb{R}_+$, is a non-decreasing function of the component's state, capturing the relationship between its condition and the associated expenses. The operational cost is incurred at inspection times and reflects the potential loss in production quality due to the degradation of the critical component, as well as the upkeep required to maintain the continuous functionality

and performance of the system. Furthermore, the operational cost at the end of the machine's life must also be accounted for to reflect the potential loss in quality during the period between the last inspection (at time $N-1$) and the end of the machine's lifetime. This ensures that the impact of the component's degradation on production quality is fully considered for the entire duration of the system's operational life. The action cost, denoted by the function $r : \mathcal{A} \mapsto \{0, R\}$, with $R \in \mathbb{R}_+$, represents the cost of replacing the component or doing nothing. For $a_t = 1$, the replacement cost is $r(a_t) = R$, conversely for $a_t = 0$, $r(a_t) = 0$.

Let P^a be the transition matrix when action $a \in \mathcal{A}$ is chosen. Each entry P^a_{ij} in the matrix represents the probability of transitioning from state $i \in S$ at one inspection time to state $j \in \mathcal{S}$ at the next one. For instance, if the component is currently in state i, the transition probability matrix provides insights into the probability of remaining in the same state or transitioning to a worse state due to degradation. Let us note that P^0 is the transition matrix when the action is to do nothing, and we denote this by P. For $a = 1$, the transition matrix P^1, is such that all rows are identical to the first row of the probability transition matrix P. Specifically, for all $i \in \mathcal{S}$, $P^1_{ij} = P_{0j}$. We further assume that $P_{ij} = 0$ when $i > j$, meaning the component's deterioration cannot decrease. Moreover, if the component is in the failed state, it will remain there with probability 1 unless it is replaced, meaning $P_{ll} = 1$.

In this paper, we use the following condition for the probability matrix to derive our analytical results.

Condition 3.1. *The probability transition matrix P is such that for each $k \in S$, $\sum_{j=k}^{l} P_{ij}$ is increasing in i, implying that the more deteriorated the component gets, the more likely it will be for the component to get further deteriorated.*

As stated before, the goal is to optimally determine the action, whether to replace or do nothing at each inspection time $t \in \mathcal{T}$.

4 The solution approach

Given the problem setting outlined in Section 3, we first explain the traditional MDP approach used to find the optimal replacement policy, where it is assumed the probability transition matrix P is fully known in advance. This problem has been extensively studied in the literature, see, e.g., Derman (1963); Kolesar (1966); Ross (1983). Following this, we explain the RMDP approach, where the probability transition matrix is assumed to lie in a so-called ambiguity set.

4.1 The MDP approach

Throughout this section, we assume that the transition matrix P is known. Given the transition matrix P, the expected total cost function when $X_0 = 0$ is denoted by

$C_N(0)$ and is of the form

$$C_N(0) = \mathbb{E}\Big(\sum_{t=0}^{N-1} c(X_t) + r(a_t) + c(X_N) | X_0 = 0 \Big). \tag{1}$$

The goal is to solve the optimization problem $\min_{\pi \in \Pi} C_N(0)$, where π denotes a policy (a strategy that specifies the action to be taken at each inspection point t), and Π denotes the set of all acceptable policies. In the following lemma inspired by Ross (1983), we show how to solve $\min_{\pi \in \Pi} C_N(0)$ when P is known and deterministic.

Lemma 4.1. *Let us denote the $(k+1)$st row of the matrix P as P_k, then the optimization problem $\min_{\pi \in \Pi} C_N(0)$ is solved by the following Bellman equation (Ross, 1983):*

$$V_t(i) = c(i) + \min \Big\{ R + P_0^\top V_{t+1}, P_i^\top V_{t+1} \Big\}, \tag{2}$$

where $V_t(i)$ denotes the expected total cost when the state at time $t \in \mathcal{T}$ is $i \in \mathcal{S}$ and there are $N - t$ many time steps to go. Similarly, $V_N(i)$ is the terminating state with $V_N(i) = c(i)$, and V_t denotes the vector of $V_t(i)$'s for all $i \in \mathcal{S}$.

The proof of the lemma 4.1 is in the Appendix. In the following theorem, which is inspired by and closely related to Derman (1963) and Chapter 2 in Ross (1983), we show that the optimal policy for $\min_{\pi \in \Pi} C_N(0)$ is characterized as a threshold policy if the condition 3.1 holds for matrix P.

Theorem 1. *If the probability transition matrix P satisfies the condition 3.1 then, there exists a $i_t^* \in \mathcal{S}$, such that the optimal policy at time $t \in \mathcal{T}$ for the optimization problem, $\min_{\pi \in \Pi} C_N(0)$, is to replace, when $i \geq i_t^*$ and is to do nothing when $i < i_t^*$, with i_t^* given by*

$$i_t^* = \min \big\{ i : P_i^\top V_{t+1} \geq R + P_0^\top V_{t+1} \big\}, \tag{3}$$

where V_t denotes the vector of $V_t(i)$'s for all $i \in \mathcal{S}$ as in (2) and P_k denotes the $(k+1)$st row of the matrix P.

The detailed proof of this theorem can be found in the Appendix. A. From Theorem 1, it is concluded that the optimal policy for $\min_{\pi \in \Pi} C_N(i)$ is a threshold policy of the form, $\pi^* = (a_0^*, a_1^*, ..., a_{N-1}^*)$, where the optimal action at time $t \in \mathcal{T}$ when the component is in state $i \in \mathcal{S}$ is

$$a_t^*(i) = \begin{cases} 1 \text{ if } R \leq P_i^\top V_{t+1} - P_0^\top V_{t+1}, \\ \\ 0 \text{ otherwise.} \end{cases} \tag{4}$$

4.2 The RMDP approach

In RMDP, instead of assuming that the transition probability matrix is known, it is assumed to lie within an ambiguity set that captures the uncertainty in the transition probabilities. Given this ambiguity, to insure robustness, the worst-case transition matrix from this set is selected which maximizes the expected total cost. Let us explicitly characterize this maximization problem. Suppose there is a set $\mathcal{P}$ of stochastic matrices, where $P_{(t)} \in \mathcal{P}$ represents the transition matrix at time $t \in \mathcal{T}$. We denote an $N-$tuple of these matrices by $\eta = (P_{(t)})_{t\in\mathcal{T}}$, describing the probabilities governing state transitions over time. The collection of all such $N-$tuples is denoted by $\mathcal{H}$, and is defined as:

$$\mathcal{H} = \mathcal{P} \times \mathcal{P} \times ... \times \mathcal{P} = \mathcal{P}^N.$$

Now we can formulate the RMDP problem as:

$$\phi_N(\Pi, \mathcal{H}) = \min_{\pi\in\Pi} \max_{\eta\in\mathcal{H}} C_N(i), \tag{5}$$

where we minimize the expected cost on the worst-case transition matrix. An approach to solve problem (5) is to transform it into a recursive function, similar to the Bellman equation of problem (2). To facilitate this, we employ the concept of rectangular ambiguity. As explained by Iyengar (2005); Nilim and El Ghaoui (2005); Ou and Bi (2024), the rectangular ambiguity condition defines the ambiguity set as a structured rectangle, with each element constrained to its own specific bounds. This property simplifies uncertainty modeling by decoupling the elements into a clear and structured region. The ambiguity set $\mathcal{P}$ has the rectangularity property if it could be represented as the ambiguity over the rows of transition probability matrices. Let us denote the projection of the ambiguity set in the $(i+1)$st row by $\mathcal{P}_i$, i.e., $\mathcal{P}_i := \{P_i : P \in \mathcal{P}\}$, for $i \in \mathcal{S}$. Similar to Wiesemann et al (2013), we state that $\mathcal{P}$ is rectangular if

$$\mathcal{P} = \mathcal{P}_0 \times \mathcal{P}_1 \times ... \times \mathcal{P}_l. \tag{6}$$

Rectangularity basically implies that, optimizing over one row does not affect optimizing over another row. Now we can formally state the robust Bellman equation (Nilim and El Ghaoui, 2005).

Theorem 2. *If the rectangularity condition holds for the ambiguity set $\mathcal{P}$ as stated in (6), then the optimization problem (5) could be solved by the following robust Bellman equation*

$$W_t(i) = c(i) + \min\left\{R + \max_{P_0\in\mathcal{P}_0} P_0^\top W_{t+1}, \max_{P_i\in\mathcal{P}_i} P_i^\top W_{t+1}\right\} \quad i \in \mathcal{S}, t \in \mathcal{T}, \tag{7}$$

where $W_t(i)$ denotes the worst-case total expected cost when the state at time $t \in \mathcal{T}$ is $i \in \mathcal{S}$ and there are $N - t$ many time steps to go. Moreover, W_t denotes the vector of $W_t(i)$'s for all $i \in \mathcal{S}$.

Proof. The theory is proven mainly by using Theorem 1 in Nilim and El Ghaoui (2005), the rectangularity condition, and the structure of the simplified MDP recursion in (2). □

The robust Bellman equation in (7) clearly illustrates a solution procedure specifically designed for our optimal replacement problem. The challenge at this point is to solve

$$\max_{P_i \in \mathcal{P}_i} P_i^\top W_{t+1}. \tag{8}$$

The optimization problem (8) is a convex optimization problem if $\mathcal{P}_i$ is a convex set. In the next section, we discuss two data-driven ambiguity sets and show how we can solve (8).

5 Data-driven ambiguity sets

Throughout this section, we construct data-driven ambiguity sets for the deterioration probability transition matrix P, analyze their properties, and give a proper solution procedure for the optimization problem (8). To structure the ambiguity set, we assume there is a historical data that shows at different inspection times what the degradation level of the system was. Using the data, we count the number of transitions from state $i \in \mathcal{S}$ to state $j \in \mathcal{S}$ denoted as n_{ij}.

In what follows, similar to Nilim and El Ghaoui (2005) but only slightly different in formulation, we introduce two types of ambiguity sets, one based on the distance from the maximum likelihood estimation (see (12) and its discussion) and the other based on the Kullback-Leibler (KL) divergence (see (27) and its discussion). We then establish a connection between KL-based and likelihood-based ambiguity sets, outlining how these two formulations are mathematically related. By leveraging the relationship between these ambiguity sets, we introduce a hybrid statistical bound that uses available data to determine the most effective ambiguity set and robustness factor for each row of the transition matrix. This approach allows for greater flexibility and efficiency in addressing uncertainty. Then, in Section 6, we demonstrate how these statistical bounds impose restrictions on the worst-case expected total cost, enhancing the practical utility of our robust optimization framework.

5.1 Likelihood-based Ambiguity

Likelihood-based ambiguity sets are constructed using likelihood functions to define confidence regions for the parameters of the probability transition matrix. To achieve this, we first derive the maximum likelihood estimation (MLE) for the Markov chain, following the well-established method discussed by Lehmann and Casella (1998). We then create a set based on the deviation of the log-likelihood of the transition probabilities from the MLE value. Given the log-likelihood function for a Markov chain,

denoted by $\ln L(P)$

$$\ln L(P) = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} n_{ij} \ln (P_{ij}), \tag{9}$$

the maximum likelihood estimation for the Markov chain, denoted by $\hat{P}$, is

$$\hat{P}_{ij} = \frac{n_{ij}}{\sum_{j \in \mathcal{S}} n_{ij}}, \text{ for } i, j \in \mathcal{S}. \tag{10}$$

For a detailed elaboration of this result we refer the reader to Teodorescu (2009). Using (9), the maximum log-likelihood value, denoted by β_{max}, is then equal to

$$\beta_{max} = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} n_{ij} \ln \left(\frac{n_{ij}}{\sum_j n_{ij}}\right). \tag{11}$$

The likelihood-based ambiguity set as described by Anderson and Goodman (1957) is as follows,

$$\mathcal{P}^{\text{MLE}} = \left\{P \in \mathbb{R}_+^{|\mathcal{S}| \times |\mathcal{S}|} : \sum_{j \in \mathcal{S}} P_{ij} = 1 \text{ for } i \in \mathcal{S}, \sum_{i,j \in \mathcal{S}} n_{ij} \ln (P_{ij}) \geq \beta\right\}, \tag{12}$$

where $\beta < \beta_{max}$ is called the robustness factor and its value represents the level of ambiguity. Determining how to compute a β appropriately is described in Section 6. Note that for any $i, j \in \mathcal{S}$, we have $n_{ij} = 0$ whenever $P_{ij} = 0$. Given that $\lim_{x \to 0^+} x \ln x = 0$, we impose the condition that for all $i, j \in \mathcal{S}$ where $P_{ij} = 0$, the following holds:

$$n_{ij} \ln P_{ij} = 0.$$

The set $\mathcal{P}^{\text{MLE}}$ is not rectangular. A common strategy involves decoupling the set through projection. This process entails considering each row of the ambiguity set separately, resulting in a larger ambiguity set containing $\mathcal{P}^{\text{MLE}}$. This decoupling and enlargement of the ambiguity set leads to an over-approximation of the uncertainty space, resulting in obtaining a policy that is potentially more robust.

Proposition 3. *The non-rectangular ambiguity set in (12) could be over-approximated and decoupled into sets of the following form for $i \in S$:*

$$\mathcal{P}_i^{MLE}(\beta_i) = \left\{P_i \in \mathbb{R}_+^{|S|} : \sum_{j \in S} P_{ij} = 1, \sum_{j \in S} n_{ij} \ln(P_{ij}) \geq \beta_i\right\}, \tag{13}$$

with β_i defined as:

$$\beta_i = \beta - \sum_{k \neq i} \sum_{j \in S} n_{kj} \left[\ln (n_{kj}) - \ln \left(\sum_{j \in \mathcal{S}} n_{kj}\right)\right].$$

We have added an elaboration of this proposition in the Appendix A.

5.1.1 Solving the inner worst-case problem with $\mathcal{P}_i^{\text{MLE}}$

In this subsection, we show how to solve the optimization problem (8), given the rectangular ambiguity set $\mathcal{P}_i^{\text{MLE}}$. Throughout the rest of this section, in order to make the notation less tedious, we drop the time index and write W instead of W_{t+1}. Moreover, since we are solving (8) for a given $i \in \mathcal{S}$, we drop the index i in P_i, β_i, and use P, β instead. Let $n_i = \sum_{j \in \mathcal{S}} n_{ij}$. We intend to address the following optimization problem:

$$\max_{P \in \mathcal{P}_i^{\text{MLE}}(\beta)} P^\top W. \tag{14}$$

Theorem 4. *The optimal value of the optimization problem (14) is the same as the optimal value of the optimization problem* $\inf_{\mu \geq W_m} \inf_{\lambda \in \mathbb{R}_+} h^{MLE}(\lambda, \mu)$*, where the function* $h^{MLE} : \mathbb{R}^2 \mapsto \mathbb{R}$ *is defined as:*

$$h^{MLE}(\lambda, \mu) = \mu - \lambda\big(\sum_{j \in S} n_{ij} + \beta\big) + \lambda \sum_{j \in S} n_{ij} \ln\big(\frac{\lambda n_{ij}}{\mu - W_j}\big) \tag{15}$$

and W_j *denotes the* $(j+1)$*st element of the vector* W*, and* $W_m = \max_{j \in S} W_j$*.*

Proof. Let us first construct the Lagrangian function associated with (14),

$$\mathcal{L}(P, \lambda, \mu, \xi) = P^\top W + \lambda\big(\sum_{j \in S} n_{ij} \ln(P_j) - \beta\big) + \mu(1 - P^\top \mathbf{1}) + \xi^\top P, \tag{16}$$

where $\mathbf{1}$ is the vector of all ones and, $\lambda \in \mathbb{R}_+$, $\mu \in \mathbb{R}$, and $\xi \in \mathbb{R}_+^{|\mathcal{S}|}$. The maximization problem (14) is an optimization problem with linear function and convex feasible region, since $\sum_{j \in S} n_j \ln(P_j)$ is concave in P_j, and the constraints $P \in \mathbb{R}_+^{|S|}$, and $\sum_{j \in S} P_j = 1$ are linear. Consequently, since the interior of the feasible region is not empty, we can approach this by solving the dual $Opt(D)$, where

$$Opt(D) := \inf\Big\{ \sup_{P \in \mathbb{R}^{|\mathcal{S}|}} \mathcal{L}(P, \lambda, \mu, \xi) : \lambda \in \mathbb{R}_+, \mu \in \mathbb{R}, \xi \in \mathbb{R}_+^{|\mathcal{S}|} \Big\}.$$

To obtain $Opt(D)$, we first analytically solve the inner supremum. Since the Lagrangian function is concave in P_j, the root of its first derivative gives us an optimal solution, if there exists one, and hence the optimal solution denoted by P^* is

$$P^* := \Big\{ P \in \mathbb{R}^{|\mathcal{S}|} : \frac{\partial \mathcal{L}(P, \lambda, \mu, \xi)}{\partial P_j} = W_j + \lambda \frac{n_{ij}}{P_j} + \xi_j - \mu = 0, \text{ for all } j \in \mathcal{S} \Big\}.$$

Consequently, the optimal solution is $P_j^* = \frac{\lambda n_{ij}}{\mu - \xi_j - W_j}$, with the restrictions that $\lambda > 0$, and $\mu - \xi_j - W_j > 0$ for all $j \in \mathcal{S}$. After substituting the solution in the Lagrangian

function, we have

$$h(\lambda,\mu,\xi) := \sup_{P\in\mathbb{R}^{|\mathcal{S}|}} \mathcal{L}(P,\lambda,\mu,\xi) = \mu - \lambda(n_i+\beta) + \lambda\sum_{j\in\mathcal{S}} n_{ij}\ln(\frac{\lambda n_{ij}}{\mu-\xi_j-W_j}),$$

and $Opt(D)$ is,

$$\inf\Big\{h(\lambda,\mu,\xi) : \lambda>0, \mu\in\mathbb{R}, \xi\in\mathbb{R}_+^{|\mathcal{S}|}, \mu-\xi_j-W_j>0 \text{ for all } j\in\mathcal{S}\Big\}. \tag{17}$$

The function $h(\lambda,\mu,\xi)$ is increasing in ξ_j, therefore given the constraints,

$$\mu - W_j > \xi_j \geq 0 \text{ for all } j\in\mathcal{S},$$

$\arg\min_{\xi_j\geq 0} h(\lambda,\mu,\xi) = 0$ for all $j\in\mathcal{S}$. Consequently, we have,

$$Opt(D) = \inf\Big\{h(\lambda,\mu,\mathbf{0}) : \lambda>0, \mu>W_j \text{ for all } j\in\mathcal{S}\Big\}, \tag{18}$$

where $\mathbf{0}$ is the vector of all zeros. By defining the function $h^{\text{MLE}}(\lambda,\mu) := h(\lambda,\mu,\mathbf{0})$ and noting that the constraints $\mu>W_j$ for all $j\in\mathcal{S}$ can be written as $\mu>W_m$, we rewrite (18) as,

$$\inf\big\{h^{\text{MLE}}(\lambda,\mu) : \lambda>0, \mu>W_m\big\}, \tag{19}$$

and the theorem is proven. □

In the following theorem, our objective is to further simplify the optimization problem (19) by reducing it to a one-dimensional minimization problem, which can be efficiently solved using a bisection method.

Theorem 5. *The function $h^{MLE}(\lambda,\mu)$ is jointly convex, and we have,*

$$\inf_{\mu>W_m,\lambda>0} h^{MLE}(\lambda,\mu) = \inf_{\mu>W_m} h_\lambda^{MLE}(\mu),$$

where the function $h_\lambda^{MLE}(\mu)$ is defined as,

$$h_\lambda^{MLE}(\mu) = \mu - n_i\lambda(\mu), \tag{20}$$

and the function $\lambda(\mu)$ is

$$\lambda(\mu) = \exp\Big(\frac{\beta}{n_i} - \sum_{j\in\mathcal{S}}\frac{n_{ij}}{n_i}\ln(\frac{n_{ij}}{\mu-W_j})\Big). \tag{21}$$

Proof. First, the gradient vector will give us the following equations.

$$\frac{\partial h^{\text{MLE}}(\lambda,\mu)}{\partial \lambda} = \sum_j n_{ij} \ln\left(\frac{n_{ij}\lambda}{\mu - W_j}\right) - \beta,$$
$$\frac{\partial h^{\text{MLE}}(\lambda,\mu)}{\partial \mu} = 1 - \lambda \sum_j \frac{n_{ij}}{\mu - W_j}.$$

Next, we state the hessian matrix and show that it is positive semi-definite (PSD):

$$\nabla^2 h^{\text{MLE}}(\lambda,\mu) = \begin{bmatrix} \frac{n_i}{\lambda} & -\sum_j \frac{n_{ij}}{\mu - W_j} \\ -\sum_j \frac{n_{ij}}{\mu - W_j} & \lambda \sum_j \frac{n_{ij}}{(\mu - W_j)^2} \end{bmatrix}.$$

By Jensen's inequality (Bennish, 2003) it is evident that,

$$\sum_j \frac{n_{ij}/n_i}{(\mu - W_j)^2} \geq \Big(\sum_j \frac{n_{ij}/n_i}{\mu - W_j} \Big)^2.$$

Hence, $\det(\nabla^2 h^{\text{MLE}}(\lambda,\mu)) \geq 0$, so the function $h^{\text{MLE}}(\lambda,\mu)$ is convex in (λ,μ). Now, by setting $\frac{\partial h^{\text{MLE}}(\lambda,\mu)}{\partial \lambda}$ equal to zero and solving for λ we have,

$$\lambda(\mu) \in \operatorname*{arginf}_{\lambda > 0} h^{\text{MLE}}(\lambda,\mu)$$

which is of the form stated in (21). After substituting $\lambda(\mu)$ in the function $h^{\text{MLE}}(\lambda,\mu)$, and some manipulations we have:

$$\inf_{\lambda>0, \mu > W_m} h^{\text{MLE}}(\lambda,\mu) = \inf_{\mu > W_m} h^{\text{MLE}}(\lambda(\mu),\mu) = \inf_{\mu > W_m} h_\lambda^{\text{MLE}}(\mu). \tag{22}$$

and the expression in (20) is achieved. □

The function $h_\lambda^{\text{MLE}}(\mu)$ is convex in μ (see, e.g, page 88 of Boyd and Vandenberghe (2004)). Hence a bisection algorithm can be employed to solve

$$\mu_* = \operatorname*{arginf}_{\mu \geq W_m} h_\lambda^{\text{MLE}}(\mu). \tag{23}$$

For this purpose, we need to establish an upper bound μ_u^{MLE} for μ such that $\mu_* \in [W_m, \mu_u^{\text{MLE}}]$. Inspired by Nilim and El Ghaoui (2005), the following theorem provides a value for μ_u^{MLE}.

Theorem 6. *Suppose μ_* is defined as (23) and set,*

$$\mu_u^{MLE} := \frac{W_m \exp(\frac{\beta_m - \beta}{n_i}) - \bar{W}}{\exp(\frac{\beta_m - \beta}{n_i}) - 1}, \tag{24}$$

where $\bar{W} = \sum_k \frac{n_{ik}}{n_i} W_k$ *and* β_m*, for* $i \in \mathcal{S}$*, is defined as,*

$$\beta_m = \sum_{j \in \mathcal{S}} n_{ij}(\ln(n_{ij}) - \ln(n_i)).$$

Then $\mu_u^{MLE} > \mu_*$*.*

Proof. Given that $h_\lambda^{\text{MLE}}(\mu)$ is convex, any μ for which $\frac{dh_\lambda^{\text{MLE}}(\mu)}{d\mu} > 0$ gives us an upper bound on the set of optimal solutions. Hence we are looking for a solution to

$$\frac{dh_\lambda^{\text{MLE}}(\mu)}{d\mu} = 1 - \lambda(\mu) \sum_{j \in \mathcal{S}} \frac{n_{ij}}{\mu - W_j} > 0,$$

which is simplified to

$$\lambda(\mu) \sum_{j \in \mathcal{S}} \frac{n_{ij}}{\mu - W_j} < 1.$$

Taking logarithm from each side, we have,

$$\ln(\lambda(\mu)) + \ln(\sum_{j \in \mathcal{S}} \frac{n_{ij}}{(\mu - W_j)}) < 0. \tag{25}$$

The function $\ln(\lambda(\mu))$ can be simplified:

$$\begin{aligned}
\ln(\lambda(\mu)) &= \frac{1}{n_i}\Big(\beta - \sum_{k \in \mathcal{S}} n_{ik} \ln(n_{ik}) + \sum_{k \in \mathcal{S}} n_{ik} \ln(\mu - W_k)\Big) \\
&= \frac{1}{n_i}\Big(\beta - \beta_m - n_i \ln(n_i) + \sum_{k \in \mathcal{S}} n_{ik} \ln(\mu - W_k)\Big).
\end{aligned}$$

Substituting this into (25) and defining

$$g(\mu) := \frac{\beta - \beta_m}{n_i} + \sum_{k \in \mathcal{S}} \frac{n_{ik}}{n_i} \ln(\mu - W_k) + \ln \sum_{j \in \mathcal{S}} \frac{n_{ij}}{n_i} \frac{1}{\mu - W_k},$$

we want to find a μ_u^{MLE} such that $g(\mu_u^{\text{MLE}}) < 0$. We bound the function $g(\mu)$ from above by another function $\tilde{g}(\mu)$ and find the root of $\tilde{g}(\mu)$. Let us consider the transformation $\mu = W_m + a$ and solve for $a > 0$ such that $\tilde{g}(W_m + a) = 0$. By Jensen's inequality (Bennish, 2003) and the fact that $W_k \leq W_m$ for any $k \in \mathcal{S}$, we can write

$$\begin{aligned}
g(\mu) &\leq \frac{\beta - \beta_m}{n_i} + \ln \sum_{k \in \mathcal{S}} \frac{n_{ik}}{n_i}(W_m + a - W_k) + \ln \sum_{j \in \mathcal{S}} \frac{n_{ij}}{n_i} \frac{1}{W_m + a - W_k} \\
&\leq \frac{\beta - \beta_m}{n_i} + \ln(W_m + a - \bar{W}) + \ln(1/a).
\end{aligned}$$

Now we are looking for a solution to

$$\tilde{g}(W_m + a) := \frac{\beta - \beta_m}{n_i} + \ln(W_m + a - \bar{W}) + \ln(1/a) = 0, \tag{26}$$

which is

$$\mu_u^{\text{MLE}} = \frac{W_m \exp(\frac{\beta_m - \beta}{n_i}) - \bar{W}}{\exp(\frac{\beta_m - \beta}{n_i}) - 1}.$$

□

Now we can use a bisection algorithm in order to solve (23).

5.2 Kullback-Leibler Ambiguity

The Kullback-Leibler (KL) divergence quantifies the difference between two probability distributions by measuring the information lost when one distribution is used to approximate another. For two discrete probability distributions Q_1 and Q_2, with the same support set $\mathcal{S}$, the KL divergence is defined as:

$$D_{\text{KL}}(Q_1 \parallel Q_2) = \sum_{i \in \mathcal{S}} Q_1(i) \ln \frac{Q_1(i)}{Q_2(i)},$$

where $Q_1(i)$ and $Q_2(i)$ represent the probabilities assigned by Q_1 and Q_2, respectively, to the event indexed by i. If we denote the empirical transition probabilities for a particular state i as Q_i, and the true (unknown) transition probabilities for that state as P_i, then $D_{\text{KL}}(Q_i \parallel P_i)$ serves as a plausible metric for quantifying the divergence between the empirical distribution Q_i and the true distribution P_i. It is easy to see that the KL-divergence is a non-negative value that could go to infinity:

$$\sup \Big\{ \sum_{j \in \mathcal{S}} Q_{ij} \ln \frac{Q_{ij}}{P_{ij}} : \sum_{j \in \mathcal{S}} P_{ij} = 1, P_i \in \mathbb{R}_+^{|\mathcal{S}|} \Big\} = \infty,$$
$$\inf \Big\{ \sum_{j \in \mathcal{S}} Q_{ij} \ln \frac{Q_{ij}}{P_{ij}} : \sum_{j \in \mathcal{S}} P_{ij} = 1, P_i \in \mathbb{R}_+^{|\mathcal{S}|} \Big\} = 0.$$

In our setting, a KL-based ambiguity set represents uncertainty in transition probabilities within an MDP. We assume that the ambiguity between the rows of P is uncoupled and we construct the ambiguity set for each row independently and therefore, construct the ambiguity set such that it satisfies the rectangularity condition. We now define the KL-based ambiguity set for the transition probabilities of state i. For ease of notation, similar to 5.1.1, we use P instead of P_i and Q instead of Q_i to denote the distribution from state i to any other state $j \in \mathcal{S}$. Formally, the ambiguity set for state i is given by:

$$\mathcal{P}_i^{\text{KL}}(\beta) = \Big\{ P \in \mathbb{R}_+^{|\mathcal{S}|} : \sum_{j \in \mathcal{S}} Q_j \ln \frac{Q_j}{P_j} \leq \beta, \sum_{j \in \mathcal{S}} P_j = 1 \Big\}. \tag{27}$$

Consequently, problem (8) reads as:

$$\sup_{P\in\mathcal{P}_i^{\mathrm{KL}}(\beta)} P^\top W \tag{28}$$

In the following theorem, we explain how the likelihood-based and the KL-based ambiguity sets are related. This relation exploits the fact that the empirical distribution for transitions from state $i \in \mathcal{S}$ to state $j \in \mathcal{S}$ assigns a mass of $\frac{1}{n_i}$ to each data point n_{ij}, leading to $Q_j = \frac{n_{ij}}{n_i}$.

Theorem 7. *Let us denote the robustness factor β in the likelihood-based ambiguity set as β^{MLE}, and in the KL-based ambiguity set as β^{KL}. The ambiguity sets $\mathcal{P}_i^{MLE}(\beta^{MLE})$ and $\mathcal{P}_i^{KL}(\beta^{KL})$ are equivalent and*

$$\sup_{P\in\mathcal{P}_i^{KL}(\beta^{KL})} P^T W = \sup_{P\in\mathcal{P}_i^{MLE}(\beta^{MLE})} P^T W, \tag{29}$$

if the following equality holds.

$$\beta^{MLE} = \sum_{j\in\mathcal{S}} n_{ij} \ln\left(\frac{n_{ij}}{n_i}\right) - n_i\beta^{KL}. \tag{30}$$

Proof. Let us rewrite (27), knowing that $Q_j = \frac{n_{ij}}{n_i}$,

$$\begin{aligned}
\mathcal{P}_i^{\mathrm{KL}}(\beta^{\mathrm{KL}}) &= \Big\{P \in \mathbb{R}_+^{|\mathcal{S}|} : \sum_{j\in\mathcal{S}} P_j = 1, \sum_{j\in\mathcal{S}} Q_j \ln(Q_j) - \sum_{j\in\mathcal{S}} Q_j \ln(P_j) \le \beta^{\mathrm{KL}}\Big\},\\
&= \Big\{P \in \mathbb{R}_+^{|\mathcal{S}|} : \sum_{j\in\mathcal{S}} P_j = 1, \sum_{j\in\mathcal{S}} Q_j \ln(P_j) \ge \sum_{j\in\mathcal{S}} Q_j \ln(Q_j) - \beta^{\mathrm{KL}}\Big\},\\
&= \Big\{P \in \mathbb{R}_+^{|\mathcal{S}|} : \sum_{j\in\mathcal{S}} P_j = 1, \sum_{j\in\mathcal{S}} n_{ij} \ln(P_j) \ge \sum_{j\in\mathcal{S}} n_{ij} \ln(\frac{n_{ij}}{n_i}) - n_i\beta^{\mathrm{KL}}\Big\}.
\end{aligned}$$

Comparing the last equality with the likelihood-based ambiguity set in (13), it is clear that the KL-based and likelihood-based ambiguity sets are equivalent, provided that the relationship in (30) holds. Furthermore, since both optimization problems in (14) and (28) have the same objective function, the equivalence of their respective ambiguity sets implies that the two problems are identical. Thus, the theorem is proven. □

Building on Theorems 4, 5, and the results of Theorem 7, we can conclude that the optimization problem (28) is equivalent to the one-dimensional problem:

$$\inf\left\{h_\lambda^{\mathrm{KL}}(\mu) : \mu \ge W_m\right\}, \tag{31}$$

where the function $h_\lambda^{\mathrm{KL}}(\mu)$ is defined as

$$h_\lambda^{\mathrm{KL}}(\mu) = \mu - e^{-\beta} \prod_{j \in \mathcal{S}} (\mu - W_j)^{Q_j}. \tag{32}$$

This transformation reduces the optimization problem in (28) to a one-dimensional optimization over μ, which can be efficiently solved using a bisection algorithm. In order to apply a bisection algorithm, we need to establish an upper bound μ_u^{KL} for μ, ensuring that the optimal solution falls within $[W_m, \mu_u^{\mathrm{KL}}]$. By leveraging the result of Theorem 6 and the equivalency result in Theorem 7, we attain the following upper bound,

$$\mu_u^{\mathrm{KL}} = \frac{W_m e^\beta - \bar{W}}{e^\beta - 1}. \tag{33}$$

6 Statistical bounds and determining β

In this section, we elaborate on how to attain the robustness factor for each ambiguity set and how, in doing so, we find statistical bounds on the total worst-case optimal cost. By statistical bounds on $\phi_N(\Pi, \mathcal{H})$, introduced in (5), we mean to be able to say with some degree of confidence that unexpected events leading to a total cost higher than $\phi_N(\Pi, \mathcal{H})$ will not occur. To do so, let us denote the true probability transition matrix by $\mathbf{P}$ and the optimal robust policy by:

$$\pi^{\mathrm{R}} = \arg\min_{\pi \in \Pi} \max_{\eta \in \mathcal{H}} C_N(i). \tag{34}$$

Lemma 6.1. *Let $\eta^P := (\boldsymbol{P})_{t \in \mathcal{T}}$ be the N-tuple containing the true transition matrix $\boldsymbol{P}$, and $\phi_N(\pi^R, \eta^P)$ be the total expected cost when $\boldsymbol{P}$ is in effect across the entire horizon and the optimal robust policy $\pi^R \in \Pi$ is applied. If $\mathbb{P}(\boldsymbol{P} \in \mathcal{P}) \geq (1-\alpha)$, then $\phi_N(\pi^R, \eta^P)$ is upper bounded by the expected worst-case optimal cost with a confidence level of at least $1-\alpha$, meaning,*

$$\mathbb{P}\Big(\phi_N(\pi^R, \eta^P) \leq \phi_N(\Pi, \mathcal{H})\Big) \geq (1-\alpha). \tag{35}$$

Proof. Since the ambiguity set is independent of stage, $\mathcal{H} = \mathcal{P} \times \mathcal{P} \times \cdots \times \mathcal{P}$, we can see that $\mathbf{P} \in \mathcal{P}$ is simply an equivalent expression for $\eta^{\mathbf{P}} \in \mathcal{H}$. Consequently, from (5) for any $\pi \in \Pi$, we have

$$\mathbb{P}\big(\phi_N(\pi, \eta^{\mathbf{P}}) \leq \phi_N(\pi, \mathcal{H})\big) \geq \mathbb{P}\big(\eta^{\mathbf{P}} \in \mathcal{H}\big).$$

Now, if $\mathbf{P} \in \mathcal{P}$, then for any policy $\pi \in \Pi$, we have

$$\phi_N(\pi, \eta^{\mathbf{P}}) \leq \phi_N(\pi, \mathcal{H}).$$

Since $\pi^{\mathrm{R}} \in \Pi$, it is evident from (34) that $\phi_N(\pi^{\mathrm{R}}, \eta^{\mathbf{P}}) \leq \phi_N(\Pi, \mathcal{H})$, also holds when $\mathbf{P} \in \mathcal{P}$, and we have

$$\mathbb{P}\Big(\phi_N(\pi^{\mathrm{R}}, \eta^{\mathbf{P}}) \leq \phi_N(\Pi, \mathcal{H})\Big) \geq \mathbb{P}(\mathbf{P} \in \mathcal{P}),$$

which completes the proof. □

Consequently, if we can adjust the ambiguity set $\mathcal{P}$, such that $\mathbb{P}\big(\mathbf{P} \in \mathcal{P}\big) \geq (1-\alpha)$ holds, then our desired statistical bound in (35) is achieved. We explain this adjustment in detail in the following two subsections.

6.1 Adjusting the Likelihood-based Ambiguity

The robustness factor in the likelihood-based ambiguity, $\mathcal{P}^{\mathrm{MLE}}$, explained in (12), is denoted by β^{MLE}. The optimal policy will get more conservative by decreasing β^{MLE} and less conservative by increasing it. We want to be able to link β^{MLE} to $\alpha \in (0,1)$ such that we can attain $\mathbb{P}(\mathbf{P} \in \mathcal{P}^{\mathrm{MLE}}) \geq (1-\alpha)$, meaning:

$$\mathbb{P}\Big(\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}} \tilde{n}_{ij} \ln\left(\mathbf{P}_{ij}\right) \geq \beta^{\mathrm{MLE}}\Big) \geq (1-\alpha), \tag{36}$$

where $\tilde{n}_{ij}$ is the random variable representing the number of transitions from state $i \in \mathcal{S}$ to $j \in \mathcal{S}$. In order to write β^{MLE} as a function of α such that (36) holds, we employ the Likelihood Ratio Test (LRT), which is thoroughly explained by Casella and Berger (2001) and Lehmann and Casella (1998). For completeness, we provide a brief definition of LRT below.

Definition 1. *(Likelihood Ratio Test) Let $x_1, x_2, \ldots, x_n$ be observed values of a random sample from a population and $L(\theta : x)$ be the likelihood function, then the likelihood ratio test statistic for testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$ is*

$$\lambda(x) = \frac{\sup_{\theta\in\Theta_0} L(\theta : x)}{\sup_{\theta\in\Theta_0\cup\Theta_1} L(\theta : x)}.$$

LRT is any test that has a rejection region of the form $\{x : \lambda(x) \leq c\}$, where c is any value between 0 and 1.

Note that $\lambda(x)$ is the realization for the random variable

$$\Lambda(\tilde{x}) = \frac{\sup_{\theta\in\Theta_0} L(\theta : \tilde{x})}{\sup_{\theta\in\Theta_0\cup\Theta_1} L(\theta : \tilde{x})}.$$

The following key theorem is a direct consequence of the results of Bartlett (1951); Christoffersen (1998) and the theorem 10.3.3 in Casella and Berger (2001).

Theorem 8. *Let $\mathcal{X}^2_d$ be a chi-square distributed random variable with d degrees of freedom, and $\mathcal{X}^2_{d,\alpha}$ be defined by,*

$$\mathcal{X}^2_{d,\alpha} = \inf\big\{x \in \mathbb{R}_+ : \mathbb{P}(\mathcal{X}^2_d \leq x) = \alpha\big\}. \tag{37}$$

If we have a random sample such that $\tilde{n}_{ij}$ shows the number of transitions from state $i \in \mathcal{S}$ to $j \in \mathcal{S}$ and $\tilde{n} = \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}} \tilde{n}_{ij}$ then under some regularity conditions for a significance level α we have:

$$\lim_{\tilde{n}\to\infty} \mathbb{P}\Big(\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}} \tilde{n}_{ij} \ln\left(\boldsymbol{P}_{ij}\right) \leq \tilde{\beta}_{max} - \frac{\mathcal{X}^2_{l(l+1)/2,1-\alpha}}{2}\Big) = \alpha, \tag{38}$$

where $\tilde{\beta}_{max}$ is denoted as the log-likelihood evaluated at the maximum likelihood estimator,

$$\tilde{\beta}_{max} = \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}} \tilde{n}_{ij} \ln\left(\frac{\tilde{n}_{ij}}{\sum_j \tilde{n}_{ij}}\right).$$

Proof. Let us assume that $\theta \in \mathbb{R}_+^{|S|\times|S|}$ denotes the parameter of interest, which here is the probability transition matrix. The logarithm of the likelihood ratio test statistic for testing $H_0 : \theta = \mathbf{P}$ vs $H_1 : \theta \neq \mathbf{P}$ is of the form:

$$\ln(\lambda(n)) = \sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}} n_{ij} \ln(\mathbf{P}_{ij}) - \beta_{max}, \tag{39}$$

where $\ln(\lambda(n))$ is a realization of $\ln(\Lambda(\tilde{n}))$. From Bartlett (1951) we have that under some regularity conditions, stated in section 10.6.2 in Casella and Berger (2001), as the sample size goes to infinity, the random variable $-2\ln(\Lambda(\tilde{n}))$, asymptotically converges to a chi-square distribution with $\frac{l(l+1)}{2}$ degrees of freedom. Consequently, for a large enough sample, we have

$$\mathbb{P}\big(-2\ln(\Lambda(\tilde{n})) \leq \mathcal{X}^2_{l(l+1)/2,(1-\alpha)}\big) = 1-\alpha,$$

From (39) we can write,

$$\mathbb{P}\big(-2\ln(\Lambda(\tilde{n})) \leq \mathcal{X}^2_{l(l+1)/2,1-\alpha}\big) = \mathbb{P}\big(\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}} \tilde{n}_{ij} \ln(\mathbf{P}_{ij}) \geq \tilde{\beta}_{max} - \frac{\mathcal{X}^2_{l(l+1)/2,1-\alpha}}{2}\big),$$

hence, as the sample size grows to infinity,

$$\mathbb{P}\Big(\sum_{i\in\mathcal{S}}\sum_{j\in\mathcal{S}} \tilde{n}_{ij} \ln(\mathbf{P}_{ij}) \geq \tilde{\beta}_{max} - \frac{\mathcal{X}^2_{l(l+1)/2,1-\alpha}}{2}\Big) \to 1-\alpha.$$

□

To ensure the validity of Theorem 8, the Markovian degradation process must satisfy ergodicity, requiring both irreducibility and aperiodicity of the transition matrix (Bartlett, 1951). This condition is met through our data collection process, where we assume the historical data on the degradation level of the component is obtained by inspecting the component and only performing the corrective maintenance policy. In other words, we inspect the condition of the component, and when it reaches the failed state $l \in \mathcal{S}$, it is immediately replaced with a new component, and no replacements occur before state $l \in \mathcal{S}$. This data collection process ensures a continuous cycle of degradation and renewal, contributing to the ergodic nature of the process. We further assume all deterioration levels are possible, implying $P_{ij} \neq 0$ for $i \leq j$, reflecting realistic degradation paths. These assumptions ensure the Markov chain is ergodic, converging to a stationary distribution and satisfying the regularity conditions required for Theorem 8.

Now from (36) and (38), we can easily see that the realized value for the robustness factor in the likelihood-based ambiguity is

$$\beta^{\text{MLE}} = \beta_{max} - \frac{\mathcal{X}^2_{l(l+1)/2,1-\alpha}}{2}. \tag{40}$$

Note that β^{MLE} in (40) is the robustness factor in the ambiguity set (12). It is easy to see that the robustness factor for the $(i+1)$st row in the ambiguity set (13) is

$$\begin{aligned}\beta_i^{\text{MLE}} &= \beta^{\text{MLE}} - \beta_{max} + \sum_{j\in\mathcal{S}} n_{ij} \ln(\frac{n_{ij}}{n_i}) \\ &= -\frac{\mathcal{X}^2_{l(l+1)/2,1-\alpha}}{2} + \sum_{j\in\mathcal{S}} n_{ij} \ln(\frac{n_{ij}}{n_i}).\end{aligned}$$

Rewriting the ambiguity set (13), after some simple manipulations, and the definition of the empirical distribution, $Q_{ij} = \frac{n_{ij}}{n_i}$, we get the following set

$$\mathcal{P}_i^{\text{MLE}}(\beta_i^{\text{MLE}}) = \big\{P_i \in \mathbb{R}_+^{|S|} : \sum_{j\in S} P_{ij} = 1,\ \ 2n_i \sum_{j\in S} Q_{ij} \ln(\frac{Q_{ij}}{P_{ij}}) \leq \mathcal{X}^2_{l(l+1)/2,1-\alpha}\big\}. \tag{41}$$

As the sample size grows, the empirical distribution approaches the true underlying distribution. With an infinitely large dataset, the set (41) shows that the optimal solution for problem (14) converges to the empirical distribution. This occurs because the feasible region in the likelihood ambiguity set becomes increasingly tighter around the empirical distribution.

A practitioner in an industry can select a significance level α and construct the robustness factor β^{MLE} to ensure a $100 \times (1-\alpha)\%$ or higher guarantee that, under perfect information, the optimal decision would not incur costs greater than what is currently paid. This statistical bound holds only if the data is gathered from the

true transition matrix. In other words, the data must come from components with characteristics very similar to those being analyzed or planned for. It is important to note that this structure for β^{MLE} is recommended only when ample data is available; otherwise, convergence to the chi-square distribution may not occur, and the statistical guarantees could fail.

6.2 Adjusting the KL-based Ambiguity

As in the previous subsection our goal is to construct the robustness factor in the KL-based ambiguity set β^{KL}, such that $\mathbb{P}\big(\mathbf{P} \in \mathcal{P}\big) \geq (1-\alpha)$ holds, so that ultimately the statistical bound in (35) holds. It is important to note that the KL-based ambiguity sets are constructed over the rows separately and not over the entire matrix. Denoting the $(i+1)$st row of the true probability transition matrix by $\mathbf{P}_i \in \mathbb{R}_+^{|\mathcal{S}|}$, we have:

$$\mathbb{P}\big(\mathbf{P} \in \mathcal{P}\big) = \mathbb{P}\big(\mathbf{P}_0 \in \mathcal{P}_0, \mathbf{P}_1 \in \mathcal{P}_1, \ldots, \mathbf{P}_{l-1} \in \mathcal{P}_{l-1}\big). \tag{42}$$

Benefiting from the rectangularity (6) we can write:

$$\mathbb{P}(\mathbf{P} \in \mathcal{P}) = \prod_{k=0}^{l-1} \mathbb{P}(\mathbf{P}_k \in \mathcal{P}_k). \tag{43}$$

It is easy to see that if we structure $\mathcal{P}_k$'s such that $\mathbb{P}\big(\mathbf{P}_k \in \mathcal{P}_k\big) \geq (1-\alpha)^{1/l}$ for all $k \in \{0, 1, \ldots, l-1\}$ then $\mathbb{P}\big(\mathbf{P} \in \mathcal{P}\big) \geq (1-\alpha)$ holds. Hence we only have to find the robustness factor for the $(k+1)$st row, β_k^{KL}, such that:

$$\mathbb{P}\Big(D(Q_k||\mathbf{P}_k) \leq \beta_k^{\text{KL}}\Big) \geq (1-\alpha)^{1/l}, \tag{44}$$

where $Q_k \in \mathbb{R}_+^{|\mathcal{S}|}$ is the empirical distribution for the $(k+1)$th row. Directly from Csiszar (1998); Mardia et al (2019) we state the following concentration result,

$$\mathbb{P}\Big(D(Q_k||\mathbf{P}_k) \geq \beta_k^{\text{KL}}\Big) \leq \binom{n_k + l - k}{l-k} e^{-n_k \beta_k^{\text{KL}}}. \tag{45}$$

with n_k denoting the number of transitions from state k to any other state. Knowing (45), we can write β_k^{KL} as a function of α such that (44) holds:

$$\beta_k^{\text{KL}} = \frac{1}{n_k} \ln\Big(\frac{\binom{n_k+l-k}{l-k}}{1-(1-\alpha)^{1/l}}\Big). \tag{46}$$

It is evident from (46) that as the sample size $n_k \in \mathbb{N}$ increases, the robustness factor β_k^{KL} for the KL-based ambiguity set decreases. Similar to the likelihood-based ambiguity set, this behavior indicates that larger datasets lead to more reliable empirical results. Consequently, the robustness factor adjusts, compelling the optimal solution of problem (8) to converge towards the empirical distribution. This convergence reflects

the increased confidence in the empirical estimates as more data becomes available.

The KL-based ambiguity set is advantageous for limited data scenarios, as its bounds do not depend on distributional convergence. However, it can produce overly conservative and costly policies due to high risk aversion. The next theorem will show that with a sufficiently large number of transitions from a state $i \in \mathcal{S}$, the likelihood ambiguity consistently yields more cost-effective policies than the KL-based approach.

Theorem 9. *Suppose β_i^{KL} and β_i^{MLE} are the robustness factors for the $(i+1)$th row of the KL-based and likelihood-based ambiguity sets, respectively. Both of these parameters create confidence bounds with at least $(1-\alpha)\%$ confidence level. Let n_i denote the number of transition from state $i \in S$ to any other state. Moreover, let n_{s_i} be defined as,*

$$n_{s_i} = \min\left\{ n_i \in \mathbb{N} : \; 2\ln\left(\frac{\binom{n_i+l-k}{l-k}}{1-(1-\alpha)^{1/l}} \right) \geq \mathcal{X}^2_{l(l+1)/2,1-\alpha} \right\}. \tag{47}$$

Provided that the convergence result in theorem (8) holds, if $n_i \geq n_{s_i}$, then

$$\sup_{P \in \mathcal{P}_i^{KL}(\beta_i^{KL})} P^T W \geq \sup_{P \in \mathcal{P}_i^{MLE}(\beta_i^{MLE})} P^T W. \tag{48}$$

Proof. From Theorem 7, we have that when

$$\beta_i^{\text{MLE}} > \sum_{j \in \mathcal{S}} n_{ij} \ln(\frac{n_{ij}}{n_i}) - n_i \beta_i^{\text{KL}}, \tag{49}$$

the feasible region for the likelihood-based inner problem gets smaller and since the objective function in both the KL-based and the likelihood-based inner problems is linear, the optimal value of (28) is greater than or equal to the optimal value of (14). Now let us rewrite the robustness factor β_i^{MLE}, using the statistical bound in the theorem (8):

$$\beta_i^{\text{MLE}} = -\frac{\mathcal{X}^2_{l(l+1)/2,1-\alpha}}{2} + \sum_{j \in \mathcal{S}} n_{ij} \ln(\frac{n_{ij}}{n_i}).$$

Simply substituting the result in the inequality (49), we have:

$$2n_i \beta_i^{\text{KL}} > \mathcal{X}^2_{l(l+1)/2,1-\alpha}. \tag{50}$$

Using the concentration result in (46) we can derive the inequality result below:

$$2\ln\left(\frac{\binom{n_i+l-i}{l-i}}{1-(1-\alpha)^{1/l}} \right) > \mathcal{X}^2_{l(l+1)/2,1-\alpha}. \tag{51}$$

The left hand side in the inequality (51) is increasing in n_i, hence the inequality in (48) holds for any $n_i \in \mathbb{N}$ that satisfies (51). □

Since the inequality (48) holds for any $i \in \mathcal{S}$ with $n_i \in \mathbb{N}$ and $\alpha \in (0,1)$ such that the condition (51) holds, it could be easily seen from the robust Bellman equation (7) that if $n_i \geq n_{s_i}$ for all $i \in \mathcal{S}$, then employing the KL-based ambiguity set would lead to a larger robust total cost. Hence, as a result of Theorem (9), when there is insufficient data or significant uncertainty in verifying convergence to the *chi-square* distribution as indicated in Theorem (8), it is prudent to use the KL-based robust policy instead of the likelihood-based policy.

7 Computational Results

In this section, we present the results of numerical experiments to evaluate the performance of our robust approach, leading to likelihood-based and KL-based policies. We focus on expected total costs and variability, considering different components, data noise levels, robustness factors, and data availability. For clarity and brevity, we refer to the likelihood-based policy as the MLE-based policy in the plots and tables.

We examine the following problem setting. Consider a multi-component machine whose overall functionality relies on the performance of a key component. The deterioration of this component significantly affects machine efficiency. The machine operates for 2.5 years (30 months), with monthly inspections at the beginning of each month. After evaluating the deterioration level of the key component during each inspection, a decision must be made on whether to replace the component. The component's deterioration is categorized into six states, with 0 representing a new and fully functional state, and 5 indicating severely damaged (failed) state. The operational cost associated with the component at a given deterioration state $i \in \mathcal{S}$ is modeled by the function

$$c(i) := \frac{36}{125} i^3,$$

where higher deterioration levels lead to increased costs. This sharply increasing function for the operational cost is motivated by the fact that the significance of a robust policy becomes particularly evident when the operational cost in the failed state is much higher than in other states, making operation in the failed state much more expensive than in the other states. In addition, the cost to replace the component is fixed at 17 units. The true probability transition matrix for deterioration from state to state is not known, but instead we have some data available, where the deterioration level of the component from month to month is recorded, and the component is replaced only when it reaches the failed state.

We present the results of comprehensive computational experiments conducted to evaluate the performance of various policies under different noise types and training data sizes. Performance is assessed based on total cost across a spectrum of significance levels $\alpha \in (0,1)$, allowing for a nuanced analysis of policy robustness. In our analysis,

we examine three component types: "Fragile", "Moderate" and "Sturdy", each with their true associated transition matrices, which is stated in Table 1. We consider a component to be fragile if its average time to failure is two months or shorter, and sturdy if its average time to failure is five months or longer. These transition matrices reflect the underlying dynamics of the system being modeled, defining the probabilities of transitions between deterioration levels.

Table 1: Transition Matrices for Different Components

Component	**Transition Matrix**
Fragile $P^{\mathrm{f}} =$	$\begin{bmatrix} 0.0500 & 0.0750 & 0.1000 & 0.1000 & 0.2250 & 0.4500 \\ 0.0000 & 0.1000 & 0.1000 & 0.1000 & 0.2000 & 0.5000 \\ 0.0000 & 0.0000 & 0.1250 & 0.1250 & 0.2500 & 0.5000 \\ 0.0000 & 0.0000 & 0.0000 & 0.1667 & 0.1667 & 0.6667 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.2500 & 0.7500 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$
Moderate $P^{\mathrm{m}} =$	$\begin{bmatrix} 0.1667 & 0.1667 & 0.1667 & 0.1667 & 0.1667 & 0.1667 \\ 0.0000 & 0.2000 & 0.2000 & 0.2000 & 0.2000 & 0.2000 \\ 0.0000 & 0.0000 & 0.2500 & 0.2500 & 0.2500 & 0.2500 \\ 0.0000 & 0.0000 & 0.0000 & 0.3333 & 0.3333 & 0.3333 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.5000 & 0.5000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$
Sturdy $P^{\mathrm{s}} =$	$\begin{bmatrix} 0.5500 & 0.1500 & 0.1000 & 0.1000 & 0.0500 & 0.0500 \\ 0.0000 & 0.4500 & 0.2500 & 0.1500 & 0.1000 & 0.0500 \\ 0.0000 & 0.0000 & 0.4000 & 0.2000 & 0.2000 & 0.2000 \\ 0.0000 & 0.0000 & 0.0000 & 0.3500 & 0.3000 & 0.3500 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.4500 & 0.5500 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$

Ideally, ample and dependable data closely approximating the true matrices would be available. However, in real-world scenarios, this is often not the case. When limited data is available, our understanding of the true transition matrix becomes imperfect, introducing noise into the data. We generate training data by simulating transitions based on the true matrix and applying noise to mimic real-world data collection challenges. We consider two types of noise: unbiased and biased. Unbiased noise occurs when we are confident that the data accurately reflects a component similar to the one being studied, and the empirical transition matrix's average time to failure is close to that of the true matrix. Conversely, biased noise occurs when we believe the data suggests a sturdier component than it actually is—e.g., the observed average time to failure is 10% longer than the true value, indicating the real component is more fragile. This bias might stem from limited data, data collection methods, or data drawn from a component with similar but not identical characteristics.

To start with our experiments, based on the true transition matrices in Table 1, we first generate training data by simulating transitions from state to state and apply biased and unbiased noise as described above for each of the three component types. Based on the training data set, for each component type, each combination of noise

level (unbiased vs biased) and each $\alpha \in (0, 1)$, we construct the Likelihood-based and KL-based ambiguity sets, and derive the corresponding optimal robust policies. We refer to these robust policies as MLE-based robust policy and KL-based robust policy, respectively. Similarly, based on the training data set we build the empirical transition matrix and solve the classic MDP as detailed in Theorem 4.1 (Subsection 4.1) to derive the corresponding optimal policy which we refer to as the "Empirical policy".

We are interested in investigating how the derived policies would work once applied to the real components, for which the underlying degradation behaviour is represented by the true transition matrices in Table 1. To do this, we employ two key criteria. First, we implement each policy into the expected total cost function in (1), where the true transition matrix as detailed in Table 1, is used for each component type, and compute its expected cost, providing a theoretical benchmark for policy performance. Second, For each component type, each policy type (MLE-based, KL-based and Empirical), and each significance level $\alpha \in (0, 1)$, we use Monte Carlo simulation to simulate 10,000 trajectories, where a trajectory represents the observed transitions and replacements which occur throughout the operational lifespan of the machine to which the component is affixed. From these simulations, we compute the standard deviation of the total costs, allowing us to assess the variability and consistency of each policy under uncertainty.

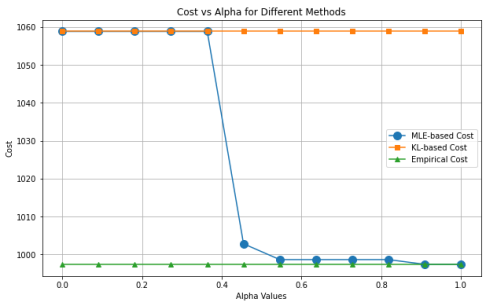
This comprehensive methodology enables us to analyze how varying levels of robustness influence policy outcomes and overall costs for different noise types. By examining both the expected costs and the standard deviation of costs, we can evaluate the effectiveness and consistency of each policy in managing component deterioration under uncertain conditions. Table 2 provides an overview of this experiment, summarizing the key variables and objectives.

In the following, we present the results of the policy performance evaluation. For brevity, the expected total cost and standard deviation of the total cost associated with the likelihood-based policy are abbreviated as MLE-based cost and MLE-based stdev in all of the figures in this section. Figure 1, shows the expected total cost (operational and replacement costs) and corresponding standard deviation for the fragile component as a function of the robustness level controlled by $\alpha \in (0, 1)$, when the size of the training data set is 3000 observations. The results are reported for both the unbiased (Figures 1a, 1b) and biased (Figures 1c, 1d) scenarios. Similarly, Figures 2 and 3 show the expected total costs and standard deviation for the Moderate component and the Sturdy component, respectively, and have the training data size of 3000 observations.

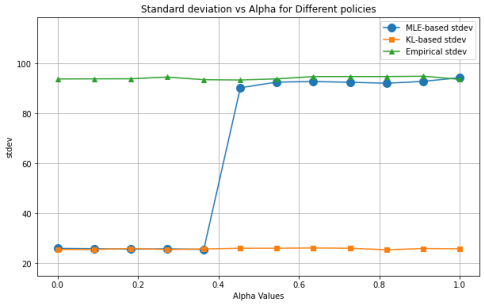
Figure 1 illustrates the performance of the Fragile component (Table 1), highlighting that the MLE-based policy exhibits greater sensitivity to $\alpha \in (0, 1)$ compared to the KL-based policy. In the case of unbiased training data (Figures 1a and 1b), the KL-based policy achieves a significant reduction in standard deviation, approximately 70% lower, while incurring only a 6% increase in expected total costs (Figure

Table 2: Experiment 1 Overview, Impact of α on Cost and Variability Across Component Types and Noise Levels

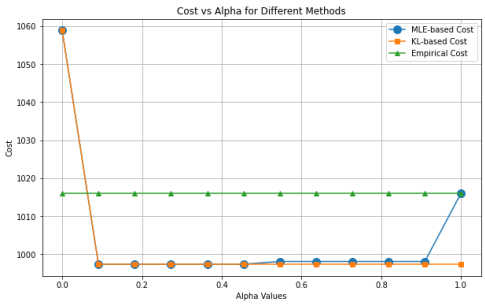
Component Type	Data	Metrics Analyzed	Plots Provided
Fragile	unbiased biased	• Exp Cost • Std. Dev	For each component and data type: 1. Exp Cost vs. $\alpha \in (0, 1)$ 2. Std. Dev vs. $\alpha \in (0, 1)$ Compare exp total cost of policies: • KL-based • MLE-based • Empirical
Moderate	unbiased biased		
Sturdy	unbiased biased		



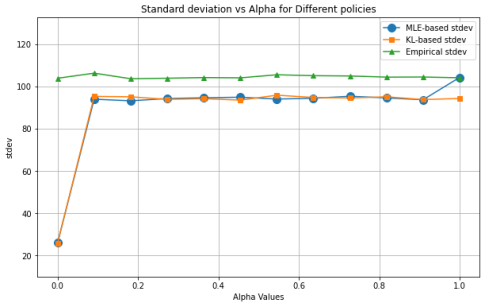
(a) Expected total cost for unbiased data



(b) Standard deviation for unbiased data



(c) Expected total cost for biased data

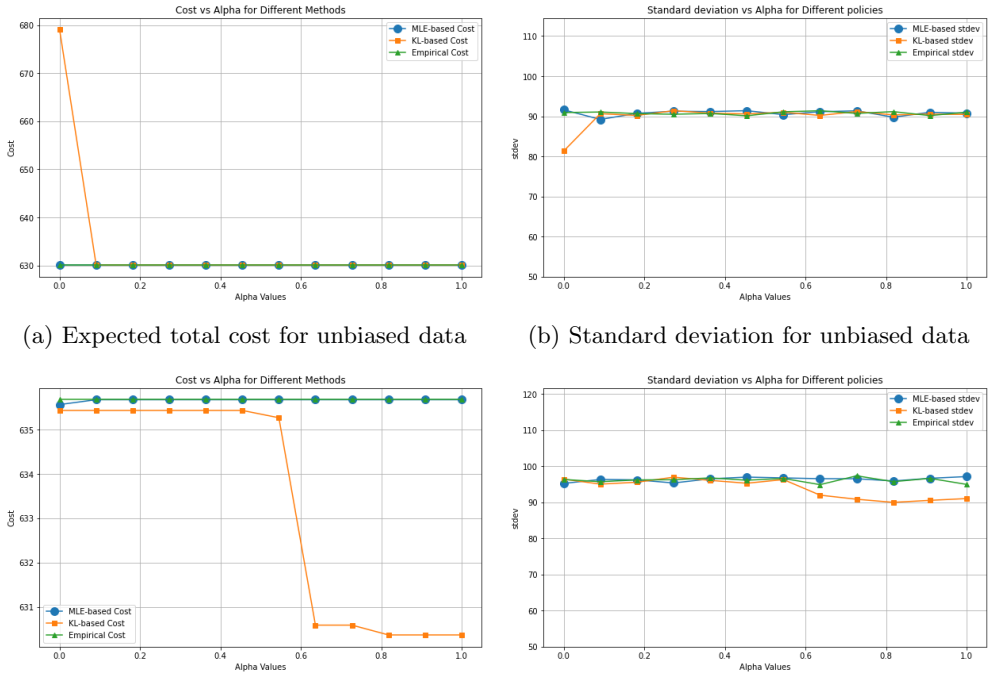


(d) Standard deviation for biased data

Fig. 1: Fragile component, training data size 3000.

1a). The MLE-based policy shows sensitivity to $\alpha \in (0, 1)$, becoming more robust as $\alpha \in (0, 1)$ decreases and aligning with the KL-based policy for $\alpha \in (0, 1)$ values below 0.35. For biased training data (Figures 1c and 1d), both robust policies generally outperform the empirical policy in terms of expected total cost and standard deviation. However, this advantage diminishes for $\alpha \in (0, 1)$ values below 0.1, where the ambiguity sets become overly conservative, resulting in higher expected costs compared to the empirical policy. This finding suggests that when data is limited or

biased, indicating a more durable component than reality, selecting $\alpha \in (0, 1)$ values close to 1 prioritizes cost-effectiveness, while values near 0 favor risk aversion. It is worth noting that a cost of 1058.87 is reached when the policy is not to replace the component throughout the machine's lifetime. By breaking down the average total cost into average replacement cost and average operational cost, we find out that under both unbiased and biased conditions, robustness results in fewer replacements. The closer the significance level $\alpha \in (0, 1)$ is to 0, the fewer the replacements, ultimately reaching a policy of no replacements at all.



(a) Expected total cost for unbiased data

(b) Standard deviation for unbiased data

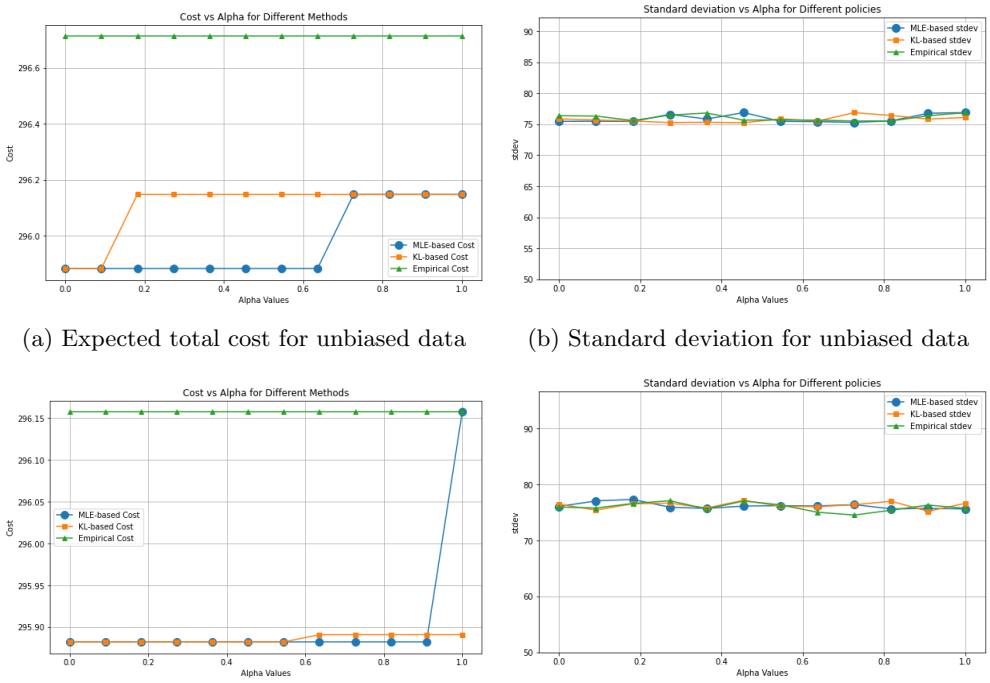
(c) Expected total cost for biased data

(d) Standard deviation for biased data

Fig. 2: Moderately durable component, training data size 3000.

For the Moderately durable component (Moderate in Table 1), Figure 2 illustrates different dynamics compared to the Fragile component. In unbiased data scenarios (Figures 2a and 2b), the KL-based policy becomes more expensive for $\alpha \in (0, 1)$ values below 0.1. As $\alpha \in (0, 1)$ approaches 0, the KL policy incurs 8% higher costs while achieving an 11% reduction in risk. In biased data scenarios (Figures 2c and 2d), the difference between robust and empirical policies narrows significantly, with variations of less than 0.8%. Interestingly, robustness for the moderately durable component behaves differently than for the fragile component. With biased training data, robustness leads to more replacements, whereas with unbiased data, it results

in fewer replacements.



(a) Expected total cost for unbiased data

(b) Standard deviation for unbiased data

(c) Expected total cost for biased data

(d) Standard deviation for biased data

Fig. 3: Sturdy component, training data size 3000.

For the Sturdy component (Table 1), as shown in Figure 3, the distinctions between policies become negligible. No significant variations are observed in terms of costs or standard deviations across the robust policies, suggesting that the choice of policy has limited impact on performance for highly durable components.

These findings highlight the complex interplay between component fragility, data bias, and the choice of $\alpha \in (0, 1)$ in determining the performance and risk profiles of maintenance policies. The results underscore the importance of carefully considering these factors when selecting and implementing robust maintenance strategies, particularly for components with varying degrees of fragility and in situations where data quality may be uncertain.

Building on the insights from our previous experiment, we now focus our attention on the Fragile component (Table 1) for our second experiment. This choice is motivated by the observation that the robust policies demonstrated the most significant impact and clearest differentiation for components prone to rapid deterioration.

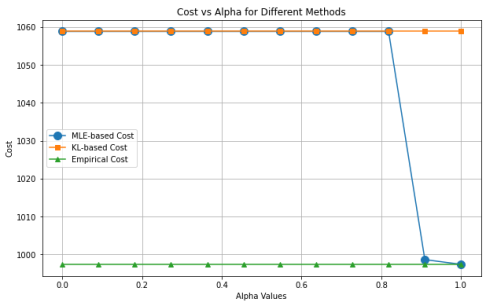
Table 3: Experiment 2 Overview: Impact of Training Data Size and α on Policy Performance for Fragile Component

Data Type	Data Size	Policies	Plotted	Analysis Focus
Unbiased	1,000 10,000 50,000	• KL-based • MLE-based • Empirical	For each data size and data type: • X-axis: $\alpha \in (0, 1)$ • Y-axis: - Exp Total Cost - Std.Dev	• Impact of α on: - Exp costs and cost variability • Impact of data size on: - Convergence behavior - Effectiveness of α
Biased	1,000 10,000 50,000			

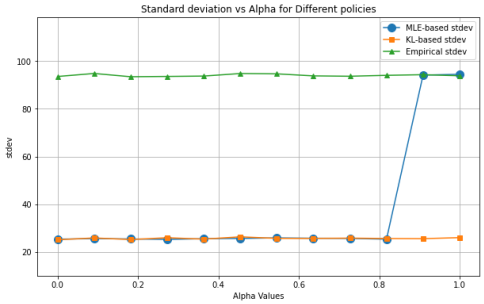
By concentrating on the Fragile component, we aim to highlight and analyze the nuanced performance differences between the KL-based, MLE-based, and Empirical policies under varying conditions of data availability and quality. In Experiment 2, an overview of which is displayed in Table 3, we investigate how training data size and robustness which is controlled by $\alpha \in (0, 1)$, influence policy performance specifically for the Fragile component.

We aim to validate a theoretical expectation regarding the behavior of ambiguity sets as the available data size increases. From results elaborated in Subsections 6.1 and 6.2, it is expected that as more data becomes available, our understanding of the component's behavior improves, leading to the ambiguity sets decreasing in size, reflecting increased certainty about the system's parameters and reduction in the need for robustness, causing robust policies to align more closely with the empirical policy. This hypothesis is tested in the second experiment, Table 3, analyzing the performance of KL-based, MLE-based, and Empirical policies across different data sizes (1000, 10000, and 50000) under both unbiased and biased data sets.

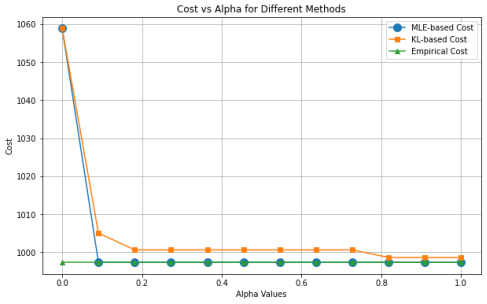
Figure 4 illustrates the performance of different policies under unbiased noise conditions. Figures 4a, 4c, and 4e show the expected total cost for data sizes of 1000, 10000, and 50000, respectively, while Figures 4b, 4d, and 4f present the standard deviations. As the data size increases, expected costs and standard deviations converge across all policies, supporting the hypothesis that larger datasets reduce ambiguity. For smaller datasets (e.g., 1000 samples), robust policies reduce cost variability significantly compared to the empirical policy, albeit with slightly higher expected costs. This highlights the trade-off of robustness under unbiased conditions, particularly when data is limited.



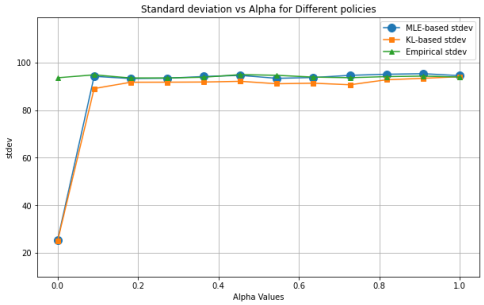
(a) Expected total cost, data size 1000



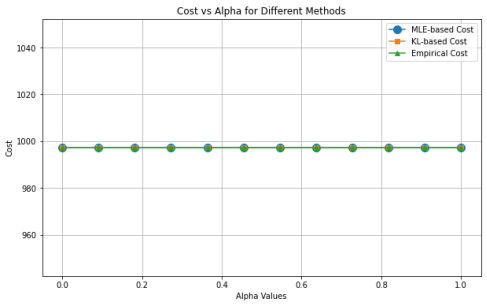
(b) Standard deviation, data size 1000



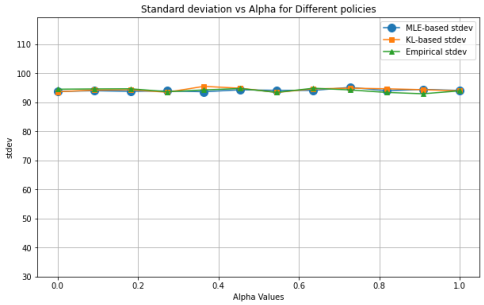
(c) Expected total cost, data size 10000



(d) Standard deviation, data size 10000



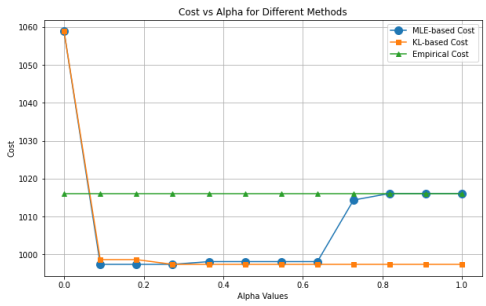
(e) Expected total cost, data size 50000



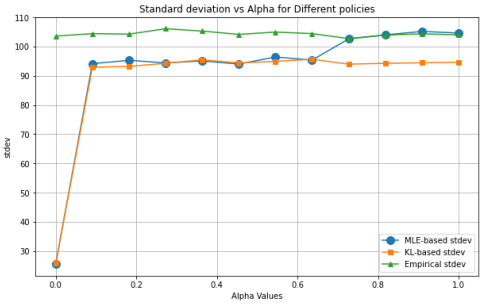
(f) Standard deviation, data size 50000

Fig. 4: Policy performance under unbiased noise for different training data sizes.

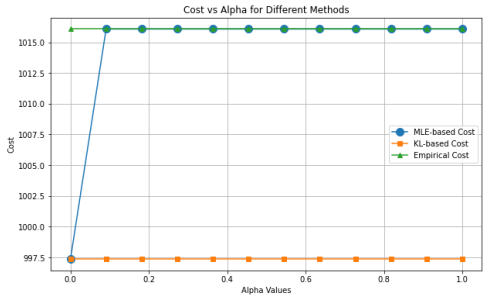
Figure 5 shows the performance of robust policies under biased noise for training data sizes of 1000, 10000, and 50000 samples. Unlike the unbiased data case, robust policies (particularly the KL-based policy) clearly outperform the empirical policy in hedging lower expected costs across all data sizes. This advantage is most pronounced for smaller datasets, as seen in the Figures 5a and 5c. However, in terms of variability, standard deviation, the robust policies provide only slight improvements over the empirical policy, as shown in Figures 5b, 5d, and 5f. This contrasts with the unbiased data case, where robust policies significantly reduced variability. As the data size



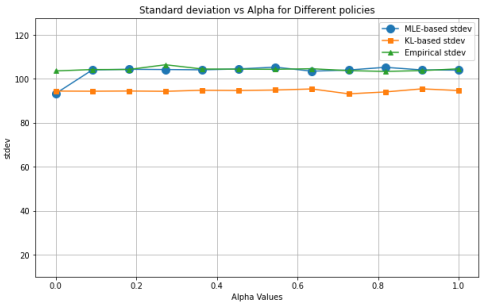
(a) Expected total cost, data size 1000



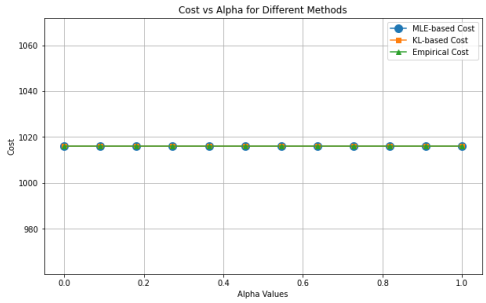
(b) Standard deviation, data size 1000



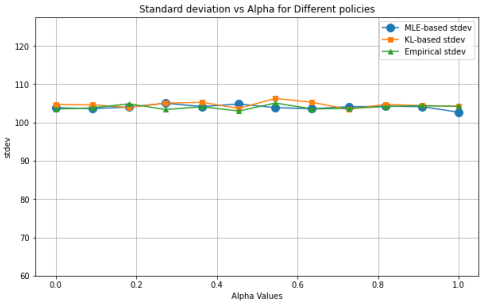
(c) Expected total cost, data size 10000



(d) Standard deviation, data size 10000



(e) Expected total cost, data size 50000



(f) Standard deviation, data size 50000

Fig. 5: Policy performance under biased noise for different training data sizes.

increases to 50000, all policies begin to converge in both cost and variability, reflecting the diminishing influence of robustness when sufficient data is available. These results highlight the effectiveness of robust policies in lowering costs under biased conditions, though their impact on reducing variability is less pronounced compared to the unbiased case.

The Figures 4 and 5 indicate that as the size of the training data increases, the policies converge and become increasingly similar, regardless of whether the data is biased or unbiased. Additionally, the effectiveness of $\alpha \in (0, 1)$ diminishes with larger data sizes. This suggests that when ample data is available, selecting an appropriate $\alpha \in (0, 1)$ becomes less critical, as policy performance shows reduced sensitivity to this parameter. Conversely, in scenarios with limited data, choosing a suitable $\alpha \in (0, 1)$ based on whether the goal is to minimize cost or variability is crucial, particularly for the likelihood-based policy.

8 Discussion and Conclusion

The RMDP method introduced in this study offers a novel approach to optimal replacement under uncertainty in maintenance. Its primary advantage lies in generating robust policies that perform well across various potential system behaviors, particularly for fragile components and in scenarios with limited or biased data. This is evidenced by the reduced cost variability compared to empirical policies, making it valuable in real-world scenarios where true transition probabilities are unknown or subject to change. A practical benefit of this method is the provision of statistical bounds on the expected total cost that the user will incur by the end of the machine's lifetime. As detailed in Section 6, this expected total cost is upper-bounded by a maximum expected cost that can be computed before policy implementation. This feature offers practitioners valuable foresight into potential implementation costs before actual deployment. It's important to note, however, that these bounds are most reliable when there is sufficient unbiased training data available. Under these conditions, the method offers decision-makers a valuable tool for risk assessment and budget planning in maintenance operations.

However, a key limitation of the current implementation is its static nature. The system does not update its understanding of the component's behavior as new transitions occur during policy implementation. This lack of adaptivity means that the system may continue to operate based on outdated or incomplete information, potentially missing opportunities for policy improvement as more data becomes available. Another significant disadvantage emerges in scenarios with severe data scarcity, particularly for the likelihood-based ambiguity set. In such cases, the statistical bounds used to construct robust policies may be unreliable, potentially leading to overly conservative policies or those that inadequately account for the true range of system behaviors. Moreover, extreme data scarcity can result in counterintuitive and non-threshold policies, such as recommending replacement at an intermediate state while suggesting no action in more deteriorated states.

Future work could enhance RMDP in maintenance scheduling by incorporating Bayesian Learning (Arts et al, 2024; Delage and Mannor, 2010; Drent et al, 2024; Kim, 2016), and Multi-Armed Bandit theory (Slivkins, 2024), enabling policies to adapt and learn from recent transitions during policy implementation. This would address the current static implementation, allowing real-time updates of transition

probabilities and policy refinement. Additionally, the issue of non-threshold policies in data-scarce scenarios could potentially be resolved by introducing a constraint in the inner optimization problems. This constraint would seek to find the closest probability distribution to the current solution that is stochastically dominated by the distribution of the next row, ensuring more intuitive and consistent threshold policies, an idea inspired by (Junová and Kopa, 2025). Additionally, exploring the use of Wasserstein-based ambiguity sets as an alternative to KL-based and likelihood-based sets could provide interesting insights into policy robustness and performance. These enhancements could lead to more adaptive, efficient, and practically applicable tools for robust maintenance scheduling in uncertain environments, while broadening the comparative analysis of different ambiguity set formulations.

Appendix A Proofs

Proof of Lemma 4.1. Utilizing the memoryless property of Markov chains and the principles of conditional probability, we can solve the problem $\min_{\pi\in\Pi} C_N(0)$ through the following recursive Bellman equation:

$$V_t(i) := c(i) + \min_{a\in\{0,1\}} \{r(a) + \sum_{j\in S} P_{ij}^a V_{t+1}(j)\}, \tag{A1}$$

where $\min_{\pi\in\Pi} C_N(0) = V_0(0)$. Since the action set is defined as $\mathcal{A} = \{0, 1\}$, with $r(0) = 0$ and $r(1) = 1$, the function represented in (2) can be readily derived. □

Proof of Theorem 1. If the condition (3.1) holds, It is proven that for any increasing function $f : \mathbb{R} \mapsto \mathbb{R}$ it holds that $\sum_j P_{ij} f(j)$ increases in i. Since $V_N(i) = c(i)$ and the operational cost function $c(i)$ is increasing in i, we can conclude that $V_{N-1}(i)$ is non-decreasing in i. Now assume that $V_{t+1}(j)$ is increasing in j, using the condition (3.1), $\sum_j P_{ij} V_{t+1}(j)$ increases in i, and $V_t(i)$ is non-decreasing in i. Now we can state that $V_t(i)$ is non-decreasing in i for all $t \in \mathcal{T}$. It follows from the structure of the Bellman equation in (2) that the optimal action is to replace if:

$$R + \sum_{j\in S} P_{0j} V_{t+1}(j) \leq \sum_{j\in S} P_{ij} V_{t+1}(j).$$

By induction we showed that $V_t(i)$ is non-decreasing in i and employing condition (3.1) again, we can state that $\sum_{j\in S} P_{ij} V_{t+1}(j)$ is increasing in i. Meaning if a replacement policy exists, then it will be a threshold policy. □

Proof of Proposition 3. The ambiguity set in Proposition 3 is attained similarly to Nilim and El Ghaoui (2005). To clarify this result, observe that:

$$\sum_{i,j\in\mathcal{S}} n_{ij} \ln(P_{ij}) = \sum_{j\in\mathcal{S}} n_{ij} \ln(P_{ij}) + \sum_{k\neq i} \sum_{j\in\mathcal{S}} n_{kj} \ln(P_{kj}).$$

Since the primarily interested in finding a lower bound for $\sum_{j\in\mathcal{S}} n_{ij}\ln(P_{ij})$, we can rewrite the ambiguity set in (12) as:

$$\mathcal{P}^{\text{MLE}} = \Big\{P \in \mathbb{R}_+^{|\mathcal{S}|\times|\mathcal{S}|} : \sum_{j\in\mathcal{S}} P_{ij} = 1 \text{ for } i \in \mathcal{S}, \sum_{j\in\mathcal{S}} n_{ij}\ln(P_{ij}) \geq \beta - \sum_{k\neq i, j\in\mathcal{S}} n_{kj}\ln(P_{kj})\Big\}.$$

It is easy to see, since $\hat{P}$ in (10) is the maximum likelihood estimator for $\ln L(P)$ in (9), the following inequality holds:

$$\sum_{k\neq i, j\in\mathcal{S}} n_{kj}\ln\Big(\frac{n_{kj}}{\sum_{j\in\mathcal{S}} n_{kj}}\Big) \geq \sum_{k\neq i, j\in\mathcal{S}} n_{kj}\ln(P_{kj}),$$

it follows that:

$$\beta - \sum_{k\neq i, j\in\mathcal{S}} n_{kj}\ln(P_{kj}) \geq \beta - \sum_{k\neq i, j\in\mathcal{S}} n_{kj}\ln\Big(\frac{n_{kj}}{\sum_{j\in\mathcal{S}} n_{kj}}\Big).$$

By setting a lower bound for the log-likelihood function of each row $i \in \mathcal{S}$, expressed as $\sum_{j\in\mathcal{S}} n_{ij}\ln(P_{ij})$, to

$$\beta_i = \beta - \sum_{k\neq i, j\in\mathcal{S}} n_{kj}\ln\Big(\frac{n_{kj}}{\sum_{j\in\mathcal{S}} n_{kj}}\Big),$$

we effectively expand the ambiguity sets corresponding to the individual rows. It becomes evident from (13) that the ambiguity sets for each row are uncoupled from each other, and the combined ambiguity set $\bigcup_{i\in\mathcal{S}} \mathcal{P}_i^{\text{MLE}}(\beta_i)$ over-approximates $\mathcal{P}^{\text{MLE}}$, ensuring that all points in $\mathcal{P}^{\text{MLE}}$ are included.

□

Declarations

Funding: The research of the first author was funded by the Eindhoven Artificial Intelligence Systems Institute, The Netherlands.
Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abbad M, Filar J, Bielecki T (1992) Algorithms for singularly perturbed limiting average Markov control problems. IEEE Transactions on Automatic Control 37(9):1421–1425

Alaswad S, Xiang Y (2017) A review on condition-based maintenance optimization models for stochastically deteriorating system. Reliability Engineering & System Safety 157:54–63

Amari S, McLaughlin L, Hoang Pham (2006) Cost-effective condition-based maintenance using markov decision processes. In: RAMS '06. Annual Reliability and Maintainability Symposium, 2006. IEEE, Newport Beach, CA, USA, pp 464–469

Anderson TW, Goodman LA (1957) Statistical Inference about Markov Chains. The Annals of Mathematical Statistics 28(1):89–110

Arts J (2017) Maintenance Modeling and Optimization. BETA publication

Arts J, Boute RN, Loeys S, et al (2024) Fifty years of maintenance optimization: Reflections and perspectives. European Journal of Operational Research p S0377221724005241

Bartlett MS (1951) The frequency goodness of fit test for probability chains. Mathematical Proceedings of the Cambridge Philosophical Society 47(1):86–95

Bennish J (2003) A Proof of Jensen's Inequality. Missouri Journal of Mathematical Sciences 15(1)

Boyd SP, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge, UK ; New York

Casella G, Berger RL (2001) Statistical inference, 2nd edn. Cengage Learning, USA

Christoffersen PF (1998) Evaluating Interval Forecasts. International Economic Review 39(4):841

Csiszar I (1998) The method of types [information theory]. IEEE Transactions on Information Theory 44(6):2505–2523

De Jonge B, Scarf PA (2020) A review on maintenance optimization. European Journal of Operational Research 285(3):805–824

Delage E, Mannor S (2010) Percentile Optimization for Markov Decision Processes with Parameter Uncertainty. Operations Research 58(1):203–213

Derman C (1963) On optimal replacement rules when changes of state are markovian. Mathematical optimization techniques 396:201–210

Drent C, Drent M, Arts J (2024) Condition-Based Production for Stochastically Deteriorating Systems: Optimal Policies and Learning. Manufacturing & Service Operations Management 26(3):1137–1156

Feinberg EA, Shwartz A, Hillier FS (eds) (2002) Handbook of Markov Decision Processes, International Series in Operations Research & Management Science, vol 40. Springer US, Boston, MA

Goyal V, Grand-Clément J (2023) Robust Markov Decision Processes: Beyond Rectangularity. Mathematics of Operations Research 48(1):203–226

Grand-Clément J, Petrik M (2023) On the convex formulations of robust Markov decision processes. ArXiv:2209.10187

Iyengar GN (2005) Robust Dynamic Programming. Mathematics of Operations Research 30(2):257–280

Junová J, Kopa M (2025) Measures of stochastic non-dominance in portfolio optimization. European Journal of Operational Research 321(1):269–283

Kim MJ (2016) Robust Control of Partially Observable Failing Systems. Operations Research 64(4):999–1014

Kolesar P (1966) Minimum Cost Replacement Under Markovian Deterioration. Management Science 12(9):694–706

Kumar N, Levy K, Wang K, et al (2022) Efficient Policy Iteration for Robust Markov Decision Processes via Regularization. ArXiv:2205.14327

Lamghari-Idrissi D, Van Hugten R, Van Houtum GJ, et al (2022) Increasing Chip Availability Through a New After-Sales Service Supply Concept at ASML. INFORMS Journal on Applied Analytics 52(5):460–470

Lehmann EL, Casella G (1998) Theory of point estimation, 2nd edn. Springer texts in statistics, Springer, New York

Mannor S, Simester D, Sun P, et al (2007) Bias and Variance Approximation in Value Function Estimates. Management Science 53(2):308–322

Mannor S, Mebel O, Xu H (2016) Robust MDPs with k -Rectangular Uncertainty. Mathematics of Operations Research 41(4):1484–1509

Mardia J, Jiao J, Tánczos E, et al (2019) Concentration Inequalities for the Empirical Distribution. URL http://arxiv.org/abs/1809.06522, arXiv:1809.06522

Nilim A, El Ghaoui L (2005) Robust Control of Markov Decision Processes with Uncertain Transition Matrices. Operations Research 53(5):780–798

Ou W, Bi S (2024) Sequential Decision-Making under Uncertainty: A Robust MDPs review. ArXiv:2404.00940

Quatrini E, Costantino F, Di Gravio G, et al (2020) Condition-Based Maintenance—An Extensive Literature Review. Machines 8(2):31

Ramani S, Ghate A (2022) Robust Markov Decision Processes with Data-Driven, Distance-Based Ambiguity Sets. SIAM Journal on Optimization 32(2):989–1017

Rosenfield D (1976) Markovian Deterioration with Uncertain Information. Operations Research 24(1):141–155

Ross SM (1969) A markovian replacement model with a generalization to include stocking. Management Science 15(11):702–715

Ross SM (1983) Introduction to Stochastic Dynamic Programming: Probability and Mathematical. Academic Press, Inc., USA

Salameh JP, Cauet S, Etien E, et al (2018) Gearbox condition monitoring in wind turbines: A review. Mechanical Systems and Signal Processing 111:251–264

Slivkins A (2024) Introduction to Multi-Armed Bandits. ArXiv:1904.07272

Smallwood RD, Sondik EJ (1973) The Optimal Control of Partially Observable Markov Processes over a Finite Horizon. Operations Research 21(5):1071–1088

Teixeira HN, Lopes I, Braga AC (2020) Condition-based maintenance implementation: a literature review. Procedia Manufacturing 51:228–235

Teodorescu I (2009) Maximum likelihood estimation for markov chains. arXiv:09054131

Van Oosterom C, Peng H, Van Houtum GJ (2017) Maintenance optimization for a Markovian deteriorating system with population heterogeneity. IISE Transactions 49(1):96–109

Wiesemann W, Kuhn D, Rustem B (2013) Robust Markov Decision Processes. Mathematics of Operations Research 38(1):153–183

Xu H, Mannor S (2012) Distributionally Robust Markov Decision Processes. Mathematics of Operations Research 37(2):288–300