International Journal of Production Research, 2016
http://dx.doi.org/10.1080/00207543.2016.1181282

How heterogeneity influences condition-based maintenance for gamma degradation process

Linmiao Zhang[a*], Yong Lei[b] and Houcai Shen[c]

[a]*Department of Industrial & Systems Engineering, National University of Singapore, Singapore, Singapore;* [b]*State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou, China;* [c]*Department of Management Science and Engineering, Nanjing University, Nanjing, China*

(*Received 28 August 2015; accepted 8 April 2016*)

In many applications, units from the same population exhibit heterogeneity that they degrade with different rates due to random factors. This article studies how this heterogeneity in degradation influences condition-based maintenance (CBM) policy. Many CBM polices are developed based on gamma process because it is popularly used to characterise monotone degradation processes. In this study, we also model the unit's degradation by gamma process. To account for the heterogeneity among units' degradation, we incorporate a random effect parameter in the gamma process. Then the optimal policy for CBM is obtained through Markov decision process. We show that when heterogeneity exists, the transition probability of degradation state depends on both unit's age and observed degradation level. And consequently, the optimal maintenance policy is a monotone control limit policy. We conduct extensive numerical experiments to validate and demonstrate our findings in depth.

Keywords: condition-based maintenance; gamma process; heterogeneity; Markov decision process

1. Introduction

In many engineering applications, irreversible damage gradually occurs along system's usage. Such damage is often additive and leads to continuous degradation of system's physical condition. When cumulative degradation reaches a certain level, the system malfunctions and results in a failure. In need for predicting system's remaining useful life and determining maintenance actions, a lot of prognostic models (Singpurwalla 1995; Nikulin et al. 2010) have been developed to characterise the degradation process. Among all those models, gamma process (Abdel-Hameed 1975) has been most widely used in engineering applications for its properties such as monotone sample path and independent increment, see Van Noortwijk and Klatter (1999), Wang (2014), for example, and Van Noortwijk (2009) for a thorough review. Nevertheless, in certain applications, units from the same population may have different degradation features because of environmental or operational reasons. Consider submarine pipelines as an example. Affected by environmental factors such like temperature, stress and wave loads, fatigue cracks are likely to propagate with different rates (Gangloff 2005) in different segments of the pipeline system. In other applications, heterogeneity among population has been reported as well, see Lawless and Crowder (2004), Gebraeel et al. (2005), Liao and Tian (2013), Ye et al. (2014), for example. Such heterogeneity in degradation challenges maintenance decision, as adaptive maintenance policy is required to account for units' different degradation patterns.

In this study, we consider condition-based maintenance (CBM) policy. CBM determines maintenance action based on real-time state. Maintenance is only conducted when necessary, therefore maintenance resources are optimised. This makes CBM often outperforms time-based maintenance, see Wang (2002) for a summary of different maintenance policies. In recent years, advanced sensor technologies enable CBM being implemented at lower costs. Therefore CBM attracts intensive attention and have been applied more widely, see Chen et al. (2011), Xia et al. (2013), for example, and Jardine, Lin, and Banjevic (2006) for a review. More specifically, several studies of CBM based on gamma process model have been developed (Abdel-Hameed 1987; Abdel-Hameed 1995; Grall, Bérenguer, and Dieulle 2002; Liao, Elsayed, and Chan 2006). In Abdel-Hameed's (Abdel-Hameed 1987; Abdel-Hameed 1995) studies, a system's degradation level is revealed by scheduled inspection and modelled by gamma process. The system malfunctions if degradation process reaches a failure threshold, which is defined as a failure. A failure can only be detected by inspection, and if failure is detected, a corrective maintenance must be performed. Otherwise upon each inspection, the decision-maker can decide whether to execute a preventive maintenance. Both CM and PM could renew the system's degradation state. Figure 1 demonstrates such CBM policy.

Most of current CBM studies assume fixed degradation process for all units across population. However, this assumption is no longer valid when heterogeneity in degradation is present. From statistical aspect, with heterogeneity among population,

*Corresponding author. Email: zhanglinmiao@u.nus.edu

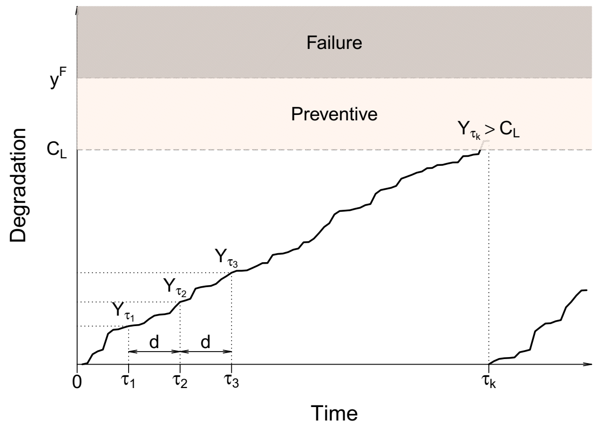


Figure 1. Condition-based maintenance policy. τ_i denotes inspection time, Y_{τ_i} denotes system state at each inspection, C_L denotes preventive control limit, and y^F denotes failure threshold.

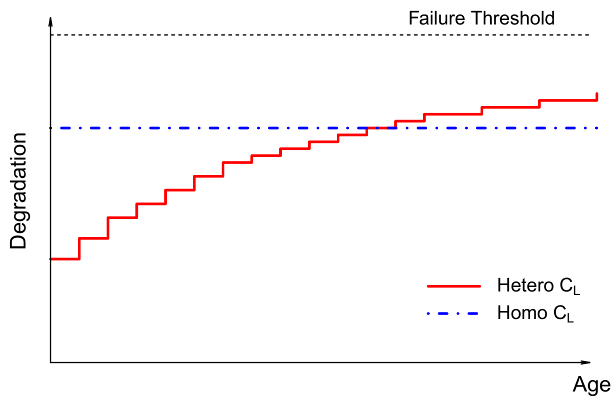


Figure 2. Preventive control limits when heterogeneity is present/absent.

the units' life-time distributions are not identical. Therefore, sharing same degradation model and preventive control limit becomes unreasonable. Intuitively speaking, units with faster degradation rates are expected to be replaced earlier, while those degrade slower could be replaced later. In this sense, with scheduled inspection, the preventive control limit is supposed to vary as shown in Figure 2.

In recent literature, there exists a few studies incorporating heterogeneity into statistical prognostic models. For instance, Ye et al. (2013) modelled the heterogeneous effects as a frailty term to study how heterogeneity affects field failure and accelerated tests. Xu, Hong, and Jin (forthcoming) used random effects and shape-restricted splines to model the dynamic covariates that caused heterogeneity. Zheng et al. (2016) proposed a state-space model to account for multiple sources of heterogeneity. And in terms of the popularly used gamma degradation process, Lawless and Crowder (2004) suggested implanting a random effect parameter into gamma process model to account for heterogeneous degradation. With proper prior distribution assigned to the random effect parameter, analytical conditional distribution on the gamma increment can be obtained once observations are collected.

This modified gamma process is capable to characterise heterogeneity in degradation among populations. Unfortunately, most of these studies are focused on developing prognostic models to estimate the system's remaining useful life (RUL). In terms of maintenance policy for heterogeneously degrading population, especially based on gamma process, the studies are quite limited. Finding an optimal CBM policy based on such heterogeneous gamma degradation model is challenging because the degradation process becomes non-stationary and age dependent.

In this paper, we study how heterogeneity in degradation influences CBM policy. Following Lawless and Crowder (2004), we apply gamma process with random effect parameter to model units' heterogeneous degradation. We formulate maintenance decision as a Markov decision process (MDP) to obtain optimal maintenance policy in terms of discounted costs, including maintenance cost, downtime cost and inspection cost. Based on the structural properties of the optimal policy, we focus on investigating how the costs and control limit change when heterogeneity is considered. The rest of this article is organised as follows. Section 2 formulates the CBM as MDP problem. Section 3 reviews the gamma process with heterogeneity and studies the influences on CBM policy when heterogeneity is present. Numerical demonstrations are presented in Section 4. And Section 5 concludes this study.

2. Maintenance model and optimisation

2.1 *Maintenance assumptions*

We consider CBM for units subject to stochastic degradation. The degradation is modelled by stochastic process which has independent non-negative increment, such as gamma process. The maintenance policy employs periodic inspections to check the unit's degradation level. A fixed inspection cost c_i is incurred each time we conduct inspection. We assume the inspection is perfect, i.e. it reveals the unit's state instantly without measurement error.

Once degradation level exceeds the pre-specified failure threshold y^F, the unit is considered as failure. The failure is not self-announcing and can only be detected by inspection. Before the failure is detected, downtime cost c_d is incurred per time unit to account for quality loss.

Upon each inspection, if failure is detected, corrective maintenance (CM) must be executed immediately with fixed cost c_f; otherwise the decision-maker can choose either performing a preventive maintenance (PM) with fixed cost c_p ($c_p < c_f$), or not taking any actions. We also assume both CM and PM are perfect, i.e. they instantly restore the unit to as-good-as-new state. Moreover, we incorporate a discounting factor e^{-rt} for any costs incurred at time t, where r is a fixed constant. We are interested in the optimal maintenance policy that minimises the total discounted operational cost, including inspection cost, maintenance cost and downtime costs. This maintenance model is also used in other studies (Speijker et al. 2000; Chen et al. 2015).

2.2 *Markov decision process*

Suppose at current inspection, the unit's age since last maintenance is τ, and the unit's degradation level is revealed as Y_τ. In next subsection we will show Y_τ is Markovian. We let (τ, Y_τ) together form a discrete time continuous state Markov chain. To find optimal policy that minimise total operational cost, we formulate the maintenance decision as an infinite horizon MDP. An inspection node is also a decision epoch, at which the decision-maker has to choose among three actions {PM, CM, No Action}:

- If $Y_\tau \geq y^F$, the decision-maker has no choice but choose CM. A corrective maintenance cost c_f is incurred. The unit restores to the state (0,0) after CM.
- If $Y_\tau < y^F$ and the decision-maker chooses PM, the unit also restores to state (0,0). In this case, a preventive maintenance cost c_p is counted.
- If $Y_\tau < y^F$ and the decision-maker chooses No Action, no maintenance cost is incurred. However, as degradation persists, the unit may fail at a random time T before next inspection at time $\tau + d$. The corresponding discounted (relative to current age τ) downtime cost is function of T:

$$\rho(T|\tau) = \begin{cases} \displaystyle\int_T^{\tau+d} c_d\, e^{-r(t-\tau)}\, \mathrm{d}t, & \tau \leq T \leq \tau + d, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

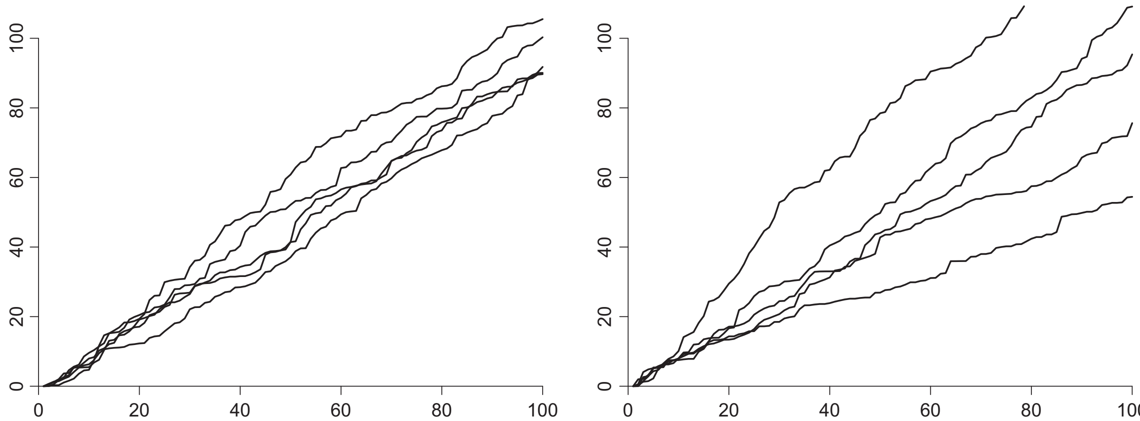
which is also random. If we know the distribution of failure-time T, we can obtain the expected downtime cost:

$$\begin{aligned} W_d(\tau, Y_\tau) &= \mathrm{E}[\rho(T)|\tau, Y_\tau] \\ &= \int_\tau^{\tau+d} \rho(t)\, \mathrm{d}[1 - \tilde{F}(t|\tau, Y_\tau)], \end{aligned} \tag{2}$$

where $\tilde{F}(t|\tau, Y_\tau) = \mathrm{P}(T > t|T > \tau, Y_\tau)$ denotes the failure-time distribution.

The above discussion can be summarised in the MDP framework by solving Bellman equation (Puterman 2009):

$$V_d(\tau, Y_\tau) = \begin{cases} \min\{e^{-rd} U_d(\tau, Y_\tau) + W_d(\tau, Y_\tau), c_p + V_d(0, 0)\}, & Y_\tau < y^F, \\ c_f + V_d(0, 0), & Y_\tau \geq y^F, \end{cases} \tag{3}$$



(a) $\omega^{-1} = 1$ (b) $\omega^{-1} \sim \mathrm{Ga}(2, 2^{-1})$

Figure 3. Sample paths with or without random effect.

where $\tau = 0, d, 2d, \ldots$, $Y_\tau \in \mathbb{R}^+$, $V_d(\tau, Y_\tau)$ is the value function that represents the minimum total discounted cost with initial state (τ, Y_τ), and $U_d(\tau, Y_\tau) = \mathrm{E}[V_d(\tau + d, Y_{\tau+d})|\tau, Y_\tau]$ is the expected value function after one-period transition from state (τ, Y_τ). It is noted that $V_d(0, 0)$ hence represents the minimum total costs, including PM, CM and downtime costs in long-term operation, for a brand new unit.

In addition to the maintenance and downtime costs, whatever action the decision-maker chooses, inspection cost is always incurred. Given inspection interval d, the total discounted inspection cost is:

$$S(d) = \sum_{k=0}^{\infty} c_i e^{-rkd} = \frac{c_i}{1 - e^{-rd}}, \tag{4}$$

which monotonically decreases as d increases.

With the formulations of all costs, optimal maintenance policy can be obtained by a two-step approach. We first fix d, solve $V_d(0, 0)$ using value iteration algorithm (Puterman 2009) and obtain corresponding optimal action. Then we find an inspection interval d such that $S(d) + V_d(0, 0)$ is minimised. Our objective is to study how heterogeneity influences the structure of such optimal policy.

3. Influence of heterogeneity on CBM policy

3.1 *Gamma process with heterogeneity*

To study the effects of heterogeneity in degradation, we first provide a background of gamma process with heterogeneity. The gamma process is a natural choice to model degradation processes in which irreversible damage gradually occurs, such as fatigue crack growth (Singpurwalla 1995). A gamma process describes a non-negative-valued sequence $\{Y_t, t > 0\}$ such that:

(1) It starts from zero at time 0, i.e. $Y_0 = 0$,
(2) the increments in disjoint time intervals $\Delta Y_t = Y_{t+\Delta} - Y_t$ are independent,
(3) ΔY_t follows a gamma distribution $\mathrm{Ga}(\Lambda(t + \Delta) - \Lambda(t), \omega)$ with mean $\omega(\Lambda(t + \Delta) - \Lambda(t))$ and variance $\omega^2(\Lambda(t + \Delta) - \Lambda(t))$,

where $\Lambda(t)$ is the transformed time scale function that monotonically increases, and $\Lambda(0) = 0$. Intuitively, given $\Lambda(t)$, the parameter ω controls both the degradation rate and volatility.

Following Lawless and Crowder (2004), to account for heterogeneity, we let ω be a random effect parameter. we assume ω^{-1} follows a gamma distribution $\mathrm{Ga}(\delta, \gamma^{-1})$ prior to any observations. Different units have independent realisations of ω, therefore their degradation patterns are heterogeneous. Figure 3 demonstrates the difference of sample paths between using constant and random ω.

Suppose now the degradation level Y_t is observed along times t_j, $j = 1, 2, \ldots, n$. Let $Y_j = Y_{t_j}$ and $\Lambda_j = \Lambda(t_j)$ for notation simplicity. Given the observations $\mathbf{Y}_n \equiv [Y_1, Y_2, \ldots, Y_n]$ and corresponding time $\mathbf{t}_n \equiv [t_1, t_2, \ldots, t_n]$, applying

Bayes's rule leads to the result that $\langle\omega^{-1}|\mathbf{Y}_n, \mathbf{t}_n\rangle$ still follows gamma distribution with updated parameters Ga$(\Lambda_n + \delta, (Y_n + \gamma)^{-1})$. Using the conditional distribution of $\langle\omega^{-1}|\mathbf{Y}_n, \mathbf{t}_n\rangle$, we can obtain the conditional distribution of Y_{n+1} by marginalisation:

$$\begin{aligned} f(Y_{n+1} = y|\mathbf{Y}_n, \mathbf{t}_n) &= \int f(Y_{n+1} = y|\omega^{-1}, \mathbf{Y}_n, \mathbf{t}_n) f(\omega^{-1}|\mathbf{Y}_n, \mathbf{t}_n)\,\mathrm{d}\omega^{-1} \\ &= \frac{(y - Y_n)^{\Lambda_{n+1}-\Lambda_n-1}(Y_n + \gamma)^{\Lambda_n+\delta}}{B(\Lambda_{n+1} - \Lambda_n, \Lambda_n + \delta)(y + \gamma)^{\Lambda_{n+1}+\delta}}, \end{aligned} \tag{5}$$

Where $y > Y_n$ and $B(\Lambda_{n+1} - \Lambda_n, \Lambda_n + \delta) = \Gamma(\Lambda_{n+1} - \Lambda_n)\Gamma(\Lambda_n + \delta)/\Gamma(\Lambda_{n+1} + \delta)$ is the beta function. It is readily shown from Equation (5) that $\{Y_t\}$ is Markovian since $\langle Y_{n+1}|\mathbf{Y}_n, \mathbf{t}_n\rangle$ only depends on current observation Y_n and t_n. Therefore, we could drop vector forms of $\mathbf{Y}_n$ and $\mathbf{t}_n$ in the remainder.

In addition, when failure threshold is specified as y^F, given the fact that the unit is not failed at time t_n, we can obtain the conditional distribution of failure-time T:

$$\mathrm{P}(T > t_{n+1}|T > t_n, Y_n) = \mathrm{P}(Y_{n+1} < y^F|Y_n < y^F, t_n). \tag{6}$$

It is noted from Equation (5) that

$$\left(\frac{\Lambda_n + \delta}{\Lambda_{n+1} - \Lambda_n}\right)\left(\frac{y - Y_n}{Y_n + \gamma}\right) \sim \mathcal{F}_{2\Lambda_{n+1}-2\Lambda_n, 2\Lambda_n+2\delta}, \tag{7}$$

where $\mathcal{F}_{d_1,d_2}$ denotes F-distribution with parameters d_1, d_2. Thus, combining Equations (6) and (7) we obtain:

$$\mathrm{P}(T > t_{n+1}|T > t_n, Y_n) = F\left(\frac{(\Lambda_n + \delta)(y^F - Y_n)}{(\Lambda_{n+1} - \Lambda_n)(Y_n + \gamma)}\right), \tag{8}$$

where F is the cumulative distribution function of $\mathcal{F}_{2\Lambda_{n+1}-2\Lambda_n, 2\Lambda_n+2\delta}$. Using Equation (8) we can easily obtain the expected down time cost W_d discussed in Equation (2).

3.2 *Why heterogeneity matters*

If degradation is modelled by gamma process without heterogeneity, the degradation increments are independently and identically distributed as long as the time increments (after scale transformed) are equal. Mathematically, $\mathrm{P}(Y_{t_1+\Delta_1} - Y_{t_1} \le y) = \mathrm{P}(Y_{t_2+\Delta_2} - Y_{t_2} \le y)$ holds for any t_1, t_2 as long as $\Lambda(t_1 + \Delta_1) - \Lambda(t_1) = \Lambda(t_2 + \Delta_2) - \Lambda(t_2)$. This implies that whatever historical observations we record, the expected degradation rates in future remain the same. However, this property is no longer valid when heterogeneity is present.

In Section 3.1, we describe modelling heterogeneity using random effect parameter ω. Since we update the distribution of ω with historical observations, the degradation rate therefore also depends on the historical data. In detail, at age t, a unit with a higher degradation level (larger y_t) means it is likely to degrade faster. Thus, we expect this unit would have a larger degradation increment in the future. We formally state this intuition in Lemma 1 by means of stochastic order:

LEMMA 1 *When heterogeneity is present,* $\langle Y_{t+\Delta} - Y_t|Y_t\rangle$ *is stochastically non-decreasing in* Y_t*, i.e.* $\langle Y_{t+\Delta} - Y_t|Y_t = y_1\rangle \prec \langle Y_{t+\Delta} - Y_t|Y_t = y_2\rangle$ *provided* $y_1 \le y_2$*,*

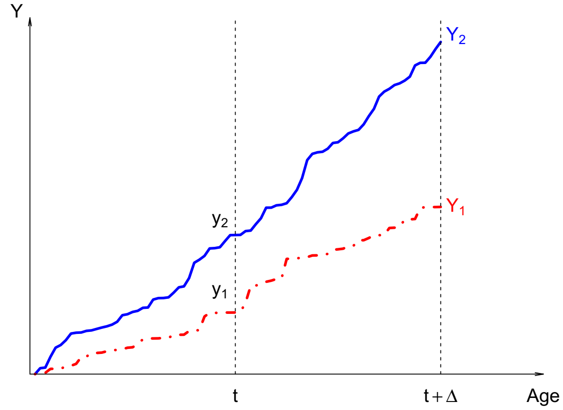
where $\prec$ denotes 'stochastic smaller than'. A random variable X is said to be stochastic smaller than a random variable Y if $P(X \le t) \ge P(Y \le t) \quad \forall t$. On the other hand, if a unit reaches a given degradation level y at a later time t, it is likely to degrade slower. In this case, we expect its future degradation increments to be smaller. This intuition can also be stated by stochastic order, as in Lemma 2:

LEMMA 2 *When heterogeneity is present,* $\langle Y_{t+\Delta} - Y_t|Y_t\rangle$ *is stochastically non-decreasing in* t*, i.e.* $\langle Y_{t_1+\Delta_1} - Y_{t_1}|Y_{t_1} = y\rangle \prec \langle Y_{t_2+\Delta_2} - Y_{t_2}|Y_{t_2} = y\rangle$ *provided* $t_1 > t_2$ *and* $\Lambda(t_1 + \Delta_1) - \Lambda(t_1) \le \Lambda(t_2 + \Delta_2) - \Lambda(t_2)$.

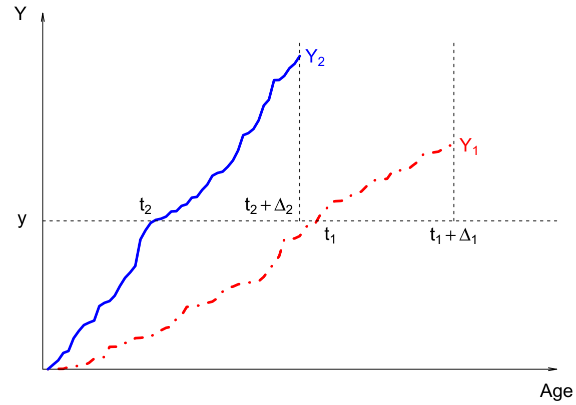
The proofs of Lemmas 1 and 2 are provided in Appendices 1 and 2. Figure 4 graphically illustrates the idea of these two Lemmas.

3.3 *Influence on the costs*

Lemmas 1 and 2 establish a framework in which we could compare different units' degradation rates based on historical observations. In this subsection, we study how heterogeneity in degradation rate affects minimum total cost (Equation (3)). For notation simplicity, we fix inspection interval, hence drop the subscript d in this subsection.



(a) Lemma 1



(b) Lemma 2

Figure 4. Illustration of the stochastic order when heterogeneity is present, in both figures $Y_1 \prec Y_2$.

We first investigate how heterogeneity Influences downtime cost (Equation (2)). Since the downtime cost is only incurred if a failure occurs before next inspection, its quantity is determined by the distribution of failure-time. It can be easily seen that no matter whether heterogeneity exits, when the current degradation level is larger, the unit is likely to fail at an earlier time before next inspection and thus incurs higher downtime cost. As a result $W(\tau, Y_\tau)$ increases with Y_τ. On the other hand, if given the same degradation level y:

- when heterogeneity does not exists, whenever the unit reaches y, its future degradation rate will not change and its failure-time distribution is identical. Thus, $W(\tau, Y_\tau)$ will not change with τ.
- when heterogeneity exists, if the unit reaches y at a later age, it degrades slower and is expected to have a later failure-time. Thus, $W(\tau, Y_\tau)$ decreases with τ.

These discussions can be summarised as Proposition 1 with detailed proof provided in Appendix 3.

PROPOSITION 1 *When heterogeneity is absent, the expected downtime cost $W(\tau, Y_\tau)$ is a constant of τ, and non-decreasing function of Y_τ; while if heterogeneity is present, $W(\tau, Y_\tau)$ is non-increasing function of τ, and non-decreasing function of Y_τ.*

Similarly, we could establish such monotone property of the cost function $U(\tau, Y_\tau)$ and $V(\tau, Y_\tau)$. Recall that $U_d(\tau, Y_\tau) = \mathrm{E}[V_d(\tau + d, Y_{\tau+d})|\tau, Y_\tau]$. The results are summarised in Proposition 2 with proof provided in Appendix 4.

PROPOSITION 2 *When heterogeneity is absent, the value functions $V(\tau, Y_\tau)$ is constants of τ, and non-decreasing function of Y_τ; while if heterogeneity is present, $V(\tau, Y_\tau)$ is non-increasing function of τ, and non-decreasing function of Y_τ.*

3.4 *Influence on the control limit*

Propositions 1 and 2 enables us to explore the structure of the optimal maintenance policy. Because of the monotone property of the cost functions, the optimal policy can be shown to have simplistic structure. In this subsection, we mainly study how heterogeneity influences the structure of optimal policy.

We first describe the optimal maintenance policy. The optimal action at each inspection can be determined through the value function (Equation (3)). More specifically, given $y < y^F$, the optimal action a in state (τ, Y_τ) is:

$$a(\tau, Y_\tau) = \begin{cases} \text{PM}, & e^{-rd}U(\tau, Y_\tau) + W(\tau, Y_\tau) > c_p + V(0, 0), \\ \text{No Action}, & \text{otherwise}. \end{cases}$$

This is equivalent to define a preventive control limit $C_L(\tau) = \{Y_\tau : e^{-rd}U(\tau, Y_\tau) + W(\tau, Y_\tau) = c_p + V(0, 0)\}$, which is a function of τ.

- Without heterogeneity, Propositions 1 and 2 show that $W(\tau, Y_\tau)$ and $U(\tau, Y_\tau)$ are constants of τ. Therefore, we expect a constant control limit C_L at all ages.

Table 1. Cost parameters in numerical experiments.

Item	c_f	c_d	c_p	c_i	r
Value	60	5	15	0.25	0.01

- With heterogeneity present, both $W(\tau, Y_\tau)$ and $U(\tau, Y_\tau)$ decrease with τ because a unit reaches high degradation level earlier is more likely to fail in the future. We would like to screen those units with higher degradation rate earlier to reduce failure risk. Consequently, we expect a non-decreasing control limit $C_L(\tau)$ in τ.

The above intuitions are summarised in Theorem 1:

THEOREM 1 *Given periodic inspection interval d, when heterogeneity is absent, the optimal maintenance policy that minimises $V(0, 0)$ is a constant control limit policy; while if heterogeneity is present, the optimal maintenance policy is a monotone control limit policy.*

The corresponding proof is provided in Appendix 5. Next, we use numerical examples to demonstrate such influences brought by heterogeneity.

4. Numerical experiment

4.1 *Parameter settings*

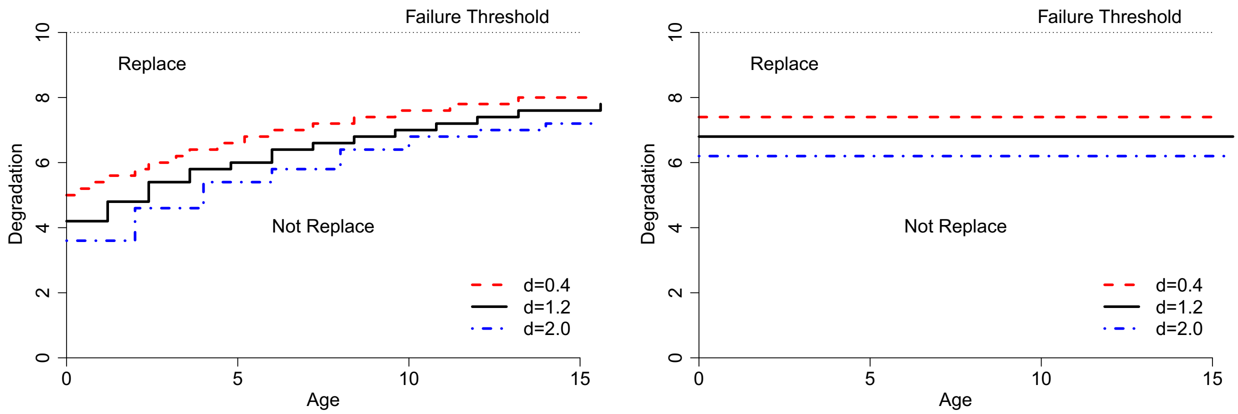
Let the unit degrade following gamma process with parameters $\Lambda(t) = t$ and ω. When heterogeneity is present, ω is random effect parameter, and ω^{-1} follows gamma distribution $\text{Ga}(\delta, \gamma^{-1})$ with $\delta = 8, \gamma = 8$. In comparison when heterogeneity is absent, we fix ω^{-1} as constant $\omega^{-1} = \delta \times \gamma^{-1} = 1$, which is the expected value of $\text{Ga}(\delta, \gamma^{-1})$. The unit fails when its degradation level reaches the threshold $y^F = 10$. To apply MDP to determine the optimal maintenance policy, we discretise the continuous degradation state space into finite values to make the computation feasible. This discretisation is a common practice in other similar studies (Elwany, Gebraeel, and Maillart 2011; Chen et al. 2015). Intuitively speaking, if we set the state space to more discrete values, we will receive a more smooth maintenance control limit. A more smooth control limit leads to an improved maintenance accuracy. But this improvement comes at a price that we need more computation time because the size of the state-transition matrix gets much larger. From our industrial experience, because the failure is degradation induced, rounding the control limit to a near value may not affect the total cost too much. Therefore we suggest that practitioners can choose the discretisation strategy as a trade-off between accuracy and computation budget. For demonstration purpose, in this study we set the degradation state space between 0 to 10, and discretise it to 50 values.

As discussed in Section 2.1, CM refers to the maintenance action taken when a unit encounters unexpected failure. This failure does not necessarily cause the unit shutting down, but does critically affect the production and can damage the unit's physical condition. And for most of the time the reparation requires vendor's intervention and assistance. Therefore, the CM cost c_f is usually very high. In addition to CM cost, the downtime cost c_d accounts for the economic loss when the unit is in failure status. For example, in semiconductor manufacturing, if a chamber (unit) is working in failure status, then all wafers processed from this chamber will encounter yield loss which we refer to as the downtime cost. This downtime cost is accumulated along time till the chamber is repaired and back to production, thus it is supposed to be calculated per unit time.

On the contrary, PM is scheduled and performed before the unit enters failure status. PM action involves replacing a few machine components, and is usually performed by company's own equipment team. Therefore, the PM cost c_p is much lower than the CM cost c_f. Similarly, the inspection is also conducted by company's own employees and its cost c_i is even lower than any maintenance costs. For demonstration purpose we set all cost parameters according to their orders, and summarise them in Table 1.

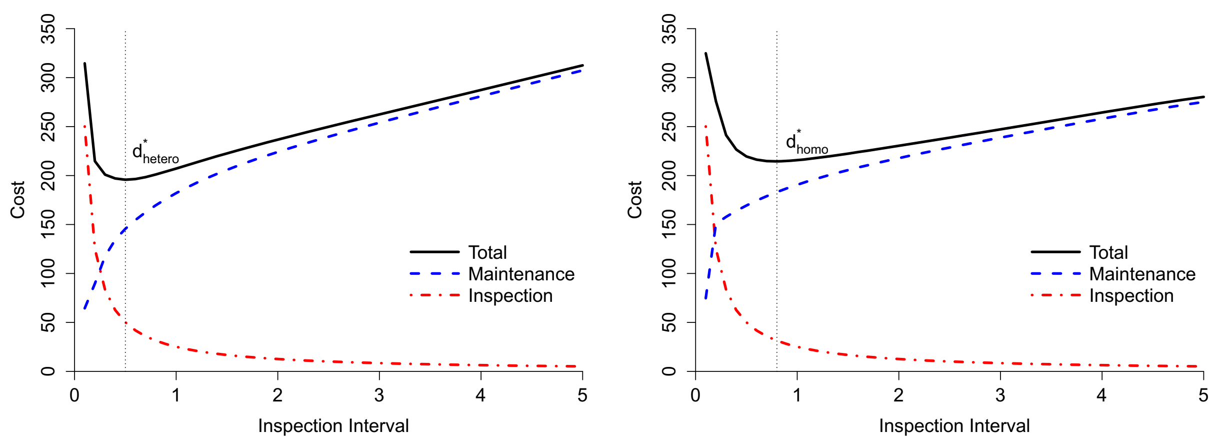
4.2 *Optimal maintenance policy*

We first study the optimal maintenance policy given all parameter settings. The optimal policy can be obtained using the value iteration algorithm. To make the computation feasible, we use a time horizon that is long enough to approximate infinite horizon. Figure 5 shows the optimal maintenance policies for both cases with different inspection interval d. Regardless of d, Figure 5(a) demonstrates that when heterogeneity exists, the optimal control limit is monotonically increasing along unit's age. In contrast, without heterogeneity, the optimal control limit is a constant as shown in Figure 5(b). Moreover, as we can observe in both cases, when smaller d is used, the maintenance policy become less conservative with higher control limits.



(a) Policy for heterogeneous degradation (b) Policy for homogeneous degradation

Figure 5. Optimal maintenance policies when different inspection intervals are used.



(a) Heterogeneous case (b) Homogeneous case

Figure 6. Cost functions at different inspection intervals.

4.3 *Optimal inspection interval*

Next, we study optimal inspection intervals in both cases. The optimal inspection interval d^* can be found by minimising inspection cost $S(d)$ and value function $V_d(0, 0)$:

$$d^* = \arg\min_d S(d) + V_d(0, 0), \tag{9}$$

which is only a one-dimensional optimisation problem. For any given d, $V_d(0, 0)$ can be found using value iteration algorithm. Derivative free search methods can be applied to find d^*

It is obviously shown from Equation (4) that $S(\delta)$ is decreasing in δ. Intuitively speaking, less frequent inspections incur lower inspection cost. In addition, we can also see that $V_d(0, 0)$ is an increasing function of d. In particular, when $d \to \infty$, only down time cost will be counted towards the total maintenance cost while $S(d) \to 0$. Figure 6 demonstrates how $S(d)$ and $V_d(0, 0)$ change in d in both heterogeneous and homogeneous cases. As we can observe from two figures, while $S(d)$ approaches to 0 and $V_d(0, 0)$ approaches to a constant, their sum has a minimum at d^*, which is the optimal inspection interval. More specifically, with heterogeneity $d^*_{\text{hetero}} = 0.5$, while without heterogeneity $d^*_{\text{homo}} = 0.8$. This implies that when heterogeneity is present, more frequent inspections are expected to track the degradation in order to reduce failure and downtime costs.

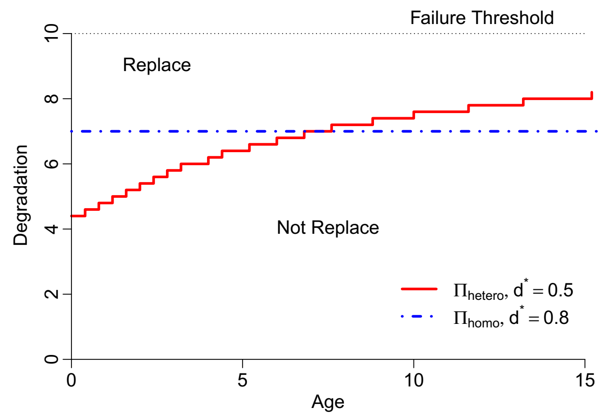


Figure 7. Optimal preventive control limits with/without considering heterogeneity.

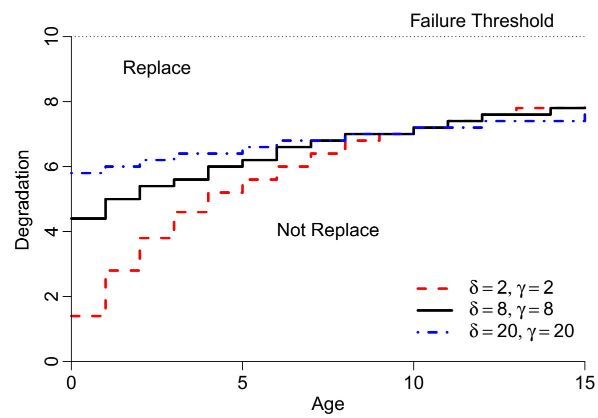


Figure 8. Sensitivity of preventive control limits to different random effect parameters.

4.4 *Policy comparison*

We are also interested in the consequence if we fail to consider the heterogeneity when it does exist. To study this scenario, we design the following simulation study. We generate degradation data with heterogeneity in population. Using the optimal intervals obtained in Section 4.3, we construct two control limits Π_{homo}, Π_{hetero} as shown in Figure 7. When we use Π_{homo} as control limit for heterogeneity population, the expected total cost is 214.51. On the contrary, using Π_{hetero} the expected total cost reduces to 195.73. This justifies the necessity of considering heterogeneity when establishing maintenance policy.

4.5 *Sensitivity analysis of random effect parameter*

Last but not least, we conduct sensitivity analysis to study how optimal maintenance policy varies in random effect parameter. We fix the inspection interval $d = 1$. Table 2 summarises the random effect parameters we set during sensitivity analysis. Recall that $\mathrm{E}[\omega^{-1}] = \delta \times \gamma^{-1}$, $\mathrm{Var}[\omega^{-1}] = \delta \times \gamma^{-2}$. This shows that in all settings, expected degradation rates keep identical. Smaller values of δ and γ indicate larger variance of ω, thus the *Low* case has a larger heterogeneity. Figure 8 shows how control limits vary when different parameters are used.

As we can observe from the figure that when heterogeneity is larger, the control limit turns out more steep. This is because with larger heterogeneity, there is more uncertainty on the degradation rate. Our inference on the degradation heavily depends on observations. Thus, we expect to screen out those units with high degradation level in early ages. On the other hand, as heterogeneity reduces, we are more certain on the true degradation rate. Consequently, the control limit become less steep and approach to a constant.

Table 2. Summary of model parameters in sensitivity analysis.

	Low	Base	High
δ	2	8	20
γ	2	8	20

5. Conclusion

In this article, we investigated how heterogeneity in degradation influences condition-based maintenance. We use gamma process with random effect parameter to model heterogeneous degradation. The structure of optimal maintenance policy was analysed. We found the optimal policy is a monotone control limit policy with heterogeneity present. In contrast, when heterogeneity is absent, the optimal control limit is a constant. We conducted extensive numerical studies to demonstrate such influences of heterogeneity.

Our study can extend to scenarios when imperfect inspection, aperiodic inspection, or imperfect repairs are considered in the future.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Linmiao Zhang http://orcid.org/0000-0002-5884-1377

Funding

This work is partially supported by National Natural Science Foundation of China (NSFC) [grant 51475422]; Science Fund for Creative Research Groups of NSFC [number 51521064]; Prof. Shen is supported by NSFC [grant 71528003].

References

Abdel-Hameed, Mohamed. 1975. "A Gamma Wear Process." *IEEE Transactions on Reliability* 24 (2): 152–153.

Abdel-Hameed, Mohamed. 1987. "Inspection and Maintenance Policies of Devices Subject to Deterioration." *Advances in Applied Probability* 19 (4): 917–931.

Abdel-Hameed, Mohamed. 1995. "Inspection, Maintenance and Replacement Models." *Computers & Operations Research* 22 (4): 435–441.

Chen, Nan, Yong Chen, Zhiguo Li, Shiyu Zhou, and Cris Sievenpiper. 2011. "Optimal Variability Sensitive Condition-based Maintenance with a Cox PH Model." *International Journal of Production Research* 49 (7): 2083–2100.

Chen, Nan, Zhi-Sheng Ye, Yisha Xiang, and Linmiao Zhang. 2015. "Condition-based Maintenance using the Inverse Gaussian Degradation Model." *European Journal of Operational Research* 243 (1): 190–199.

Elwany, Alaa H., Nagi Z. Gebraeel, and Lisa M. Maillart. 2011. "Structured Replacement Policies for Components with Complex Degradation Processes and Dedicated Sensors." *Operations Research* 59 (3): 684–695.

Gangloff, R. 2005. *Environmental Cracking Corrosion Fatigue*, 302–321. West Conshohocken: ASTM International.

Gebraeel, Nagi Z., Mark A. Lawley, Rong Li, and Jennifer K. Ryan. 2005. "Residual-life Distributions from Component Degradation Signals: A Bayesian Approach." *IIE Transactions* 37 (6): 543–557.

Grall, Antoine, Christophe Bérenguer, and Laurence Dieulle. 2002. "A Condition-based Maintenance Policy for Stochastically Deteriorating Systems." *Reliability Engineering & System Safety* 76 (2): 167–180.

Jardine, Andrew K. S., Daming Lin, and Dragan Banjevic. 2006. "A Review on Machinery Diagnostics and Prognostics Implementing Condition-based Maintenance." *Mechanical Systems and Signal Processing* 20 (7): 1483–1510.

Lawless, Jerry, and Martin Crowder. 2004. "Covariates and Random Effects in a Gamma Process Model with Application to Degradation and Failure." *Lifetime Data Analysis* 10 (3): 213–227.

Liao, Haitao, Elsayed A. Elsayed, and Ling-Yau Chan. 2006. "Maintenance of Continuously Monitored Degrading Systems." *European Journal of Operational Research* 175 (2): 821–835.

Liao, Haitao, and Zhigang Tian. 2013. "A Framework for Predicting the Remaining Useful Life of a Single Unit under Time-varying Operating Conditions." *IIE Transactions* 45 (9): 964–980.

Nikulin, Mikhail Stepanovich, Nikolaos Limnios, N. Balakrishnan, Waltraud Kahle, and Catherine Huber-Carol. 2010. *Advances in Degradation Modeling: Applications to Reliability, Survival Analysis, and Finance*. Boston: Springer Verlag.

Puterman, Martin L. 2009. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: John Wiley & Sons.

Shaked, Moshe , and J. George Shanthikumar. 2007. *Stochastic Orders*. New York: Springer.

Singpurwalla, Nozer D. 1995. "Survival in Dynamic Environments." *Statistical Science* 10 (1): 86–103.

Speijker, Lennaert J. P., Jan M. Van Noortwijk, Matthijs Kok, and Roger M. Cooke. 2000. "Optimal Maintenance Decisions for Dikes." *Probability in the Engineering and Informational Sciences* 14 (01): 101–121.

Van Noortwijk, J. M. 2009. "A Survey of the Application of Gamma Processes in Maintenance." *Reliability Engineering & System Safety* 94 (1): 2–21.

Van Noortwijk, J. M., and H. E. Klatter. 1999. "Optimal Inspection Decisions for the Block Mats of the Eastern-Scheldt Barrier." *Reliability Engineering & System Safety* 65 (3): 203–211.

Wang, Hongzhou. 2002. "A Survey of Maintenance Policies of Deteriorating Systems." *European Journal of Operational Research* 139 (3): 469–489.

Wang, Wenbin. 2014. "A Scheduling Model for Systems with Task and Health Dependent Remaining Useful Life Prognostics." *International Journal of Production Research* 52 (19): 5764–5779.

Xia, Tangbin, Lifeng Xi, Xiaojun Zhou, and Jay Lee. 2013. "Condition-based Maintenance for Intelligent Monitored Series System with Independent Machine Failure Modes." *International Journal of Production Research* 51 (15): 4585–4596.

Xu, Zhibing, Yili Hong, and Ran Jin. Forthcoming. "Nonlinear General Path Models for Degradation Data with Dynamic Covariates." *Applied Stochastic Models in Business and Industry* 32 (2): 153–167.

Ye, Zhi-Sheng, Yili Hong, Yimeng Xie, et al. 2013. "How do Heterogeneities in Operating Environments Affect Field Failure Predictions and Test Planning?" *The Annals of Applied Statistics* 7 (4): 2249–2271.

Ye, Zhi-Sheng, Min Xie, Loon-Ching Tang, and Nan Chen. 2014. "Semiparametric Estimation of Gamma Processes for Deteriorating Products." *Technometrics* 56 (4): 504–513.

Zheng, J. F., X. S. Si, C. H. Hu, Z. X. Zhang, and W. Jiang. 2016. "A Nonlinear Prognostic Model for Degrading Systems with Three-source Variability." *IEEE Transactions on Reliability* PP (99):1–15.

Appendix 1. Proof of Lemma 1

Firstly, we consider conditional gamma increment $\langle Y_{t+\Delta} - Y_t = y|\omega\rangle \sim \mathrm{Ga}(\lambda, \omega)$, where $\lambda = \Lambda(t+\Delta) - \Lambda(t)$. For any $0 < \omega_1 < \omega_2$, we can obtain their likelihood ratio:

$$\begin{aligned} \frac{f(Y_{t+\Delta} - Y_t = y|\omega_2)}{f(Y_{t+\Delta} - Y_t = y|\omega_1)} &= \left\{ \frac{1}{\Gamma(\lambda)\omega_2^{\lambda}}\, y^{\lambda-1} e^{-\frac{y}{\omega_2}} \right\} \Big/ \left\{ \frac{1}{\Gamma(\lambda)\omega_1^{\lambda}}\, y^{\lambda-1} e^{-\frac{y}{\omega_1}} \right\} \\ &= \left(\frac{\omega_1}{\omega_2}\right)^{\lambda} \cdot e^{\left(\frac{1}{\omega_1} - \frac{1}{\omega_2}\right) y} \end{aligned} \tag{A1}$$

Since $\omega_1^{-1} > \omega_2^{-1}$, the right-hand side of Equation (A1) is increasing in y. Therefore, by definition of likelihood ratio order (Shaked and Shanthikumar 2007) we have:

(S1) $\langle Y_{t+\Delta} - Y_t|\omega_1\rangle \prec_{\mathrm{LR}} \langle Y_{t+\Delta} - Y_t|\omega_2\rangle$ provided $\omega_1 < \omega_2$,

where $\prec_{\mathrm{LR}}$ denotes 'smaller than in likelihood ratio order'.

Secondly, we consider conditional distribution of ω given observation Y_t at time t. In Section 3.1 we know $\langle \omega^{-1}|Y_t\rangle \sim \mathrm{Ga}(\Lambda_t + \delta, (Y_t + \gamma)^{-1})$. For any $y_1 < y_2$, we can obtain the likelihood ratio of $\langle \omega^{-1}|Y_t = y_1\rangle$ and $\langle \omega^{-1}|Y_t = y_2\rangle$:

$$\begin{aligned} \frac{f(\omega^{-1}|Y_t = y_1)}{f(\omega^{-1}|Y_t = y_2)} &= \frac{(y_1+\gamma)^{\Lambda_t+\delta}/\,\Gamma(\Lambda_t+\delta)\,\omega^{-(\Lambda_t+\delta-1)} e^{-(y_1+\gamma)\omega^{-1}}}{(y_2+\gamma)^{\Lambda_t+\delta}/\,\Gamma(\Lambda_t+\delta)\,\omega^{-(\Lambda_t+\delta-1)} e^{-(y_2+\gamma)\omega^{-1}}} \\ &= \left(\frac{y_1+\gamma}{y_2+\gamma}\right)^{\Lambda_t+\gamma} \cdot e^{(y_2-y_1)\omega^{-1}} \end{aligned} \tag{A2}$$

The right hand side of Equation (A2) is increasing in ω^{-1}. Thus $\langle \omega^{-1}|Y_t = y_2\rangle \prec_{\mathrm{LR}} \langle \omega^{-1}|Y_t = y_1\rangle$. According to Shaked and Shanthikumar (2007), we can show that

(S2) $\langle \omega|Y_t = y_1\rangle \prec_{\mathrm{LR}} \langle \omega|Y_t = y_2\rangle$ provided $y_1 < y_2$.

Thirdly, combining S1, S2 and according to Shaked and Shanthikumar (2007), we obtain:

$$\begin{aligned} \mathrm{P}(Y_{t+\Delta} - Y_t \le \tilde{y}|Y_t = y_1) &= \int \mathrm{P}(Y_{t+\Delta} - Y_t \le \tilde{y}|Y_t = y_1, \omega)\, \mathrm{d}F(\omega|Y_t = y_1) \\ &\ge \int \mathrm{P}(Y_{t+\Delta} - Y_t \le \tilde{y}|Y_t = y_2, \omega)\, \mathrm{d}F(\omega|Y_t = y_2) \\ &= \mathrm{P}(Y_{t+\Delta} - P_t \le \tilde{y}|Y_t = y_2), \quad \forall y_1 < y_2. \end{aligned} \tag{A3}$$

This proves Lemma 1. □

Appendix 2. Proof of Lemma 2

The proof of Lemma 2 is similar to that of Lemma 1. Firstly, we consider conditional gamma increment $\left\langle Y_{t+\Delta} - Y_t = y|\omega\right\rangle \sim \mathrm{Ga}(\lambda, \omega)$. Let $\lambda_1 = \Lambda(t_1 + \Delta_1) - \Lambda(t_1)$ and $\lambda_2 = \Lambda(t_2 + \Delta_2) - \Lambda(t_2)$. For any $\lambda_1 \le \lambda_2$ and $\omega_1 < \omega_2$, we obtain likelihood ratio:

$$\begin{aligned}\frac{f(Y_{t_2+\Delta_2} - Y_{t_2} = y|\omega_2)}{f(Y_{t_1+\Delta_1} - Y_{t_1} = y|\omega_1)} &= \left\{\frac{1}{\Gamma(\lambda_2)\omega_2^{\lambda_2}}\, y^{\lambda_2-1}e^{-\frac{y}{\omega_2}}\right\}\Big/\left\{\frac{1}{\Gamma(\lambda_1)\omega_1^{\lambda_1}}\, y^{\lambda_1-1}e^{-\frac{y}{\omega_1}}\right\}\\ &= \frac{\Gamma(\lambda_1)\omega_1^{\lambda_1}}{\Gamma(\lambda_2)\omega_2^{\lambda_2}}\cdot y^{\lambda_2-\lambda_1}\, e^{\left(\frac{1}{\omega_1}-\frac{1}{\omega_2}\right)y}\end{aligned} \tag{B1}$$

Since $\omega_2^{-1} < \omega_1^{-1}$, the right-hand side of Equation (B1) is increasing in y. As a result, we have:

(S3) $\left\langle Y_{t_1+\Delta_1} - Y_{t_1}|\omega_1\right\rangle \prec_{\mathrm{LR}} \left\langle Y_{t_2+\Delta_2} - Y_{t_2}|\omega_2\right\rangle$ provided $\omega_1 < \omega_2$.

Secondly, we consider conditional distribution of $\left\langle \omega^{-1}|Y_t\right\rangle$. Let $\Lambda_1 = \Lambda(t_1)$ and $\Lambda_2 = \Lambda(t_2)$. Given $Y_{t_1} = Y_{t_2} = y$, for any $t_1 > t_2$, we obtain the likelihood ratio:

$$\begin{aligned}\frac{f(\omega^{-1}|Y_{t_1} = y)}{f(\omega^{-1}|Y_{t_2} = y)} &= \frac{(y+\gamma)^{\Lambda_1+\delta}/\,\Gamma(\Lambda_1+\delta)\,\omega^{-(\Lambda_1+\delta-1)}e^{-(y+\gamma)\omega^{-1}}}{(y+\gamma)^{\Lambda_2+\delta}/\,\Gamma(\Lambda_2+\delta)\,\omega^{-(\Lambda_2+\delta-1)}e^{-(y+\gamma)\omega^{-1}}}\\ &= \frac{\Gamma(\Lambda_2)}{\Gamma(\Lambda_1)}\cdot(y+\gamma)^{\Lambda_1-\Lambda_2}\,\omega^{-(\Lambda_1-\Lambda_2)}\end{aligned} \tag{B2}$$

Since $\Lambda_1 > \Lambda_2$, the right hand side of Equation (B2) is increasing in ω^{-1}. Consequently, we obtain:

(S4) $\left\langle \omega|Y_{t_1} = y\right\rangle \prec_{\mathrm{LR}} \left\langle \omega|Y_{t_2} = y\right\rangle$ provided $t_1 > t_2$.

Thirdly, S3 and S4 collaboratively show that

$$\begin{aligned}\mathrm{P}(Y_{t_1+\Delta_1} - Y_{t_1} \le \tilde{y}|Y_{t_1} = y) &= \int \mathrm{P}(Y_{t_1+\Delta_1} - Y_{t_1} \le \tilde{y}|Y_{t_1} = y, \omega)\,\mathrm{d}F(\omega|Y_{t_1} = y)\\ &\ge \int \mathrm{P}(Y_{t_2+\Delta_2} - Y_{t_2} \le \tilde{y}|Y_{t_2} = y, \omega)\,\mathrm{d}F(\omega|Y_{t_2} = y)\\ &= \mathrm{P}(Y_{t_2+\Delta_2} - Y_{t_2} \le \tilde{y}|Y_{t_2} = y), \quad \forall t_1 > t_2.\end{aligned} \tag{B3}$$

This proves Lemma 2. □

Appendix 3. Proof of Proposition 1

Firstly, we prove the heterogeneous case. We first show that the expected downtime cost is non-decreasing in degradation level. In Section 3.1 we analyse the conditional failure-time distribution. Given current state (τ, Y_τ), the failure-time T has distribution:

$$\mathrm{P}(T < t|Y_\tau) = 1 - P(Y_t < y^F|Y_\tau). \tag{C1}$$

According to Lemma 1, $\langle Y_t|Y_\tau = y_1\rangle \prec \langle Y_t|Y_\tau = y_2\rangle$ when $y_1 < y_2$. This is equivalent to:

(S5) $\langle T|Y_\tau = y_2\rangle \prec \langle T|Y_\tau = y_1\rangle$ when $y_1 < y_2$.

Since Equation (1) shows that $\rho(T)$ is a decreasing function of T, according to Shaked and Shanthikumar (2007) we can conclude:

$$W(\tau, y_1) = \mathrm{E}[\rho(T)|Y_\tau = y_1] \le \mathrm{E}[\rho(T)|Y_\tau = y_2] = W(\tau, y_2), \quad \forall y_1 < y_2. \tag{C2}$$

This proves that $W(\tau, Y_\tau)$ is non-decreasing in Y_τ.

Similarly, we can show that the expected downtime cost is non-increasing in age. From Lemma 2, we know that $\left\langle Y_t|Y_{\tau_1} = y\right\rangle \prec \left\langle Y_t|Y_{\tau_2} = y\right\rangle$ when $t > \tau_1 > \tau_2$. Equivalently, we have:

(S6) $\left\langle T|Y_{\tau_2} = y\right\rangle \prec \left\langle T|Y_{\tau_1} = y\right\rangle$ when $\tau_1 < \tau_2$.

As a result, we can conclude that

$$W(\tau_2, y) = \mathrm{E}[\rho(T)|Y_{\tau_2} = y] \le \mathrm{E}[\rho(T)|Y_{\tau_1} = y] = W(\tau_2, y), \quad \forall \tau_1 < \tau_2. \tag{C3}$$

This proves that $W(\tau, Y_\tau)$ is non-increasing in τ.

Secondly, we consider the homogeneous case. Recall that $\left\langle Y_{t_1+\Delta_1} - Y_{t_1}\right\rangle$ and $\left\langle Y_{t_2+\Delta_2} - Y_{t_2}\right\rangle$ are same in distribution for any t_1, t_2 provided $\Lambda(t_1 + \Delta_1) - \Lambda(t_1) = \Lambda(t_2 + \Delta_2) - \Lambda(t_2)$ and constant ω. Given $y_1 < y_2$, we still have $\langle Y_t|Y_\tau = y_1\rangle \prec \langle Y_t|Y_\tau = y_2\rangle$. Thus S5 also holds, and $W(\tau, Y_\tau)$ is non-decreasing in Y_τ. Moreover, since age has no effect on degradation rate, $W(\tau, Y_\tau)$ remains constant in τ. This completes the proof. □

Appendix 4. Proof of Proposition 2

We first prove the heterogeneous case. We apply mathematical induction based on the value iteration algorithm (Puterman 2009). We denote $V^k(\tau, Y_\tau)$ as the value function at the kth iteration.

Basis: In the initialising step, we set $V^0(\tau, Y_\tau) = 0,\ \forall \tau, Y_\tau$. Obviously $V^0(\tau, Y_\tau)$ is non-increasing in τ and non-decreasing in Y_τ. Hence Proposition 2 holds for $V^0(\tau, Y_\tau)$.

Induction step: Assume such monotone properties hold for $V^k(\tau, Y_\tau)$, then at $(k+1)$th iteration, we have:

$$V^{k+1}(\tau, Y_\tau) = \begin{cases} \min\{e^{-rd} U_d(\tau, Y_\tau) + W(\tau, Y_\tau), c_p + V(0,0)\}, & Y_\tau < y^F, \\ c_f + V(0,0), & Y_\tau \geq y^F, \end{cases} \tag{D1}$$

When $Y_\tau < y^F$, since $V^k(\tau, Y_\tau)$ is non-decreasing in Y_τ and $\langle Y_{\tau+d} | Y_\tau = y \rangle$ is stochastically non-decreasing in y, $U^k(\tau, Y_\tau) = \mathrm{E}[V^k(\tau + d, Y_{\tau+d}) | \tau, Y_\tau]$ is also non-decreasing in Y_τ. Similarly, given the fact that $V^k(\tau, Y_\tau)$ is non-increasing in τ and $\langle Y_{\tau+d} | Y_\tau = y \rangle$ is non-increasing in τ, we can conclude that $U^k(\tau, Y_\tau)$ is non-increasing in τ. In addition, recall that $W(\tau, Y_\tau)$ is also non-increasing in τ and non-decreasing in Y_τ. With $c_p + V^k(0,0)$ being constant, $V^{k+1}(\tau, Y_\tau)$ is non-increasing in τ and non-decreasing in Y_τ as a consequence. On the other hand when $Y_\tau \geq y^F$, $V^{k+1}(\tau, Y_\tau) = c_f + V^k(0,0)$ is also a constant. Hence the monotone properties hold for any τ and Y_τ.

We now conclude that $V^k(\tau, Y_\tau)$ is non-increasing in τ and non-decreasing in Y_τ for all iteration k. Moreover, as $k \to \infty$, $V^k(\tau, Y_\tau)$ will converge to $V(\tau, Y_\tau)$ (Puterman 2009). Therefore the above monotone properties also hold for $V(\tau, Y_\tau)$.

In the homogeneous case, the proof is very similar. We apply the above mathematical induction again, but assume $V^k(\tau, Y_\tau)$ is constant in τ in the inductive step. Since $W(\tau, Y_\tau)$ is also constant in τ in this case according to Proposition 1, we easily show that $V^{k+1}(\tau, Y_\tau)$ is constant in τ.This leads to the conclusion that $V(\tau, Y_\tau)$ is constant in τ. This completes the proof of Proposition 2. □

Appendix 5. Proof of Theorem 1

We first show that when heterogeneity exists, the optimal maintenance policy is a monotone control limit policy. Recall that the preventive control limit is defined as

$$C_L(\tau) = \{Y_\tau : e^{-rd} U(\tau, Y_\tau) + W(\tau, Y_\tau) = c_p + V(0,0)\}. \tag{E1}$$

Assume there exists an $\pi < y^F$ such that $e^{-rd} U(\tau, \pi) + W(\tau, \pi) = c_p + V(0,0)$. Then $\forall \psi \geq \pi$, we have $e^{-rd} U(\tau, \psi) + W(\tau, \psi) \geq c_p + V(0,0)$ according to Propositions 1 and 2. This proves that such policy is a control limit policy.

In addition, assume there exits a ξ such that $e^{-rd} U(\xi, Y_\xi) + W(\xi, Y_\xi) = c_p + V(0,0)$. Then $\forall \zeta \leq \xi$, we have $e^{-rd} U(\zeta, Y_\xi) + W(\zeta, Y_\xi) \geq c_p + V(0,0)$ because both $U(\tau, Y_\tau)$ and $W(\tau, Y_\tau)$ are non-increasing in τ. This proves that the control limit $C_L(\tau)$ is non-decreasing in τ.

On the other hand in homogeneous case, since both $U(\tau, Y_\tau)$ and $W(\tau, Y_\tau)$ are constant in τ, then $\forall \zeta \neq \xi$, we have $e^{-rd} U(\zeta, Y_\xi) + W(\zeta, Y_\xi) = e^{-rd} U(\xi, Y_\xi) + W(\xi, Y_\xi) = c_p + V(0,0)$. This proves that the $C_L(\tau)$ is constant is τ. This completes the proof of Theorem 1. □