Title: Optimal Replacement in the Proportional Hazards Model with semi-Markovian Covariate Process and Continuous Monitoring

Index terms—Optimal replacement, proportional hazards model, semi-Markov process, threshold replacement policy, sensitivity analysis

SUMMARY

This paper deals with the optimal replacement problem for general deteriorating systems in the proportional hazards model with a semi-Markovian covariate process, which we assume is under continuous monitoring. The form of an optimal policy is identified and a method is developed to find the parameters of the optimal policy. A numerical example and sensitivity analysis provide some insights for this problem.

CONCLUSIONS

Allowing the covariate process to be semi-Markovian endows our method with great capability and flexibility to model real world situations. Though the optimality of a threshold replacement policy to minimize the long-run average cost per unit time has been established previously in a more general setting, the policy evaluation step in an iterative algorithm to identify optimal threshold values poses computational challenges. To overcome them, we use conditioning to derive an explicit expression of the objective in terms of the set of state-dependent threshold ages for replacement. The iterative algorithm was customized for our model to find the optimal threshold ages. A three-state example illustrates the computational procedure as well as the effects of different sojourn time distributions of the covariate process on the optimal policy and cost. The results show that when the sojourn time distribution of the covariate process is difficult to estimate, viewing the process as a Markov process is not a bad option. Sensitivity analysis on an instance indicates that the variance of the scale parameter in the baseline hazard function accounts for most of the resulting variability in the cost.

1 INTRODUCTION

This article concerns a condition-based maintenance (CBM) problem for general deteriorating systems. The goal of CBM is to improve decision-making, compared to classical preventive maintenance, by exploiting available information about the system's condition. This work was motivated by the need to improve the management of aging assets such as high-voltage power transformers. As explained by Wang et al. [1], "As transformers age, their internal conditions degrades, which increases the risk of failure. Failures are usually triggered by severe conditions, such as lightning strikes, switching transients, short-circuits, or other incidents. When the transformer is new, it has sufficient electrical and mechanical strength to withstand unusual system conditions. As transformers age, their insulation strength can degrade to the point that they cannot withstand system events such as short-circuit faults or transient overvoltages." Increasingly, sensors and smart chips are deployed together with communication technology to stream data on expensive assets such as power distribution transformers. The condition information of the transformers in the field can be returned in real time to a central location for continuous monitoring [2]. The high cost of transformer failure motivates our study of how to use these data to decide when to preventively replace a transformer.

CBM models of the system's lifetime differ according to their approaches of utilizing the condition information. Many researchers assume that the system failure process can be described adequately by a multi-state deteriorating model because of the tractability of the resulting mathematical problems, and extensive research has been done with such models [3-6]. Grall et al. [7] presented a continuous time, continuous state mathematical model for a gradually deteriorating system and optimized both the inspection schedule and replacement policy. The issues of imperfect monitoring in the state-based preventive maintenance were considered in [8, 9]. In contrast to the multi-state deteriorating models, Toscano and Lyonnet [10] proposed a dynamic failure rate model that predicts the reliability of the system in real time by taking into account the past and present operating conditions.

Another valuable and increasingly prevalent model for estimating the risk of system failure when it is subject to condition monitoring is the proportional hazards model (PHM) [11], where the condition information is considered as a vector of covariates, each representing a certain measurement. The PHM combines a baseline hazard function which accounts for the aging degradation along with a component that takes the covariates into account to improve the prediction of failure. For transformers, the covariates could include remotely monitored variables such as acoustic and electrical signals caused by partial discharge, moisture or gases in the insulating oil, or other quantities that indicate the condition of the insulation [1]. The PHM has been applied in a variety of industrial sectors such as pulp and water; coal plant; nuclear plant refueling; military land armored vehicle; construction industry backhoes; marine diesel engines; and turbines in a nuclear plant [12]. A

PHM framework was developed in [13] to estimate failure rates of systems with two operational modes, particularly the electrical generating units.

Several papers have been published to optimize the decision-making step in the PHM setting. Makis and Jardine [14] investigated the optimal replacement policy for systems in the PHM with a Markov covariate process and periodic monitoring, and they showed that the optimal replacement policy is of a control limit type in terms of the hazard function. Banjevic and Jardine [15] extended Makis and Jardine's model by relaxing the monotonicity assumption of the hazard function and they developed methods for parameter estimation in the PHM as well. The same model was extended in [16] by assuming the information obtained at inspection epochs is imperfect; that is, the condition information of the system is only partially observed. Wu and Ryan [17] removed the discrete-time approximation of the continuous time covariate process in [14] which could lead to a counter-intuitive result when comparing the cost of policies with different monitoring intervals. They presented a new recursive procedure to obtain the optimal policy and assess the value of different condition monitoring schemes. All of these papers assumed the covariate processes to be Markov processes and under periodic monitoring.

An attempt to apply the PHM based replacement models to the power transformers under continuous monitoring exposed the gaps between the literature and practice. So far, how to utilize the continuous monitored information in the PHM has not been addressed. In addition, a Markovian model may not be appropriate for the covariate process. Many of the monitored variables are continuous. For analysis, it is convenient to discretize their combined values into a finite number of states. Requiring that times between transitions among those states are exponentially distributed is an added approximation that could introduce significant error. Therefore, we adopt a semi-Markov covariate process with general transition times.

In this paper, we allow the covariate process to be a semi-Markov process and assume it is under continuous monitoring with perfect observation. The detailed mathematical model is developed in Section 2. By identifying our model as a special case of the one described in [18], we show in Section 3 that if the hazard function of the system is non-decreasing, then the optimal replacement policy is of the control limit type with respect to the hazard function, and may be uniquely defined by a set of state-dependent threshold ages for replacement. To compute the optimal policy and optimal cost, we derive an explicit expression for our objective function in terms of the policy parameters by conditioning arguments in Section 4. The iterative procedure developed by Bergman [18] is specified for our model to find the optimal threshold ages. In Section 5, numerical examples are provided for illustration and sensitivity analysis is performed on a specific instance to demonstrate how the variations in the input parameters would affect the long-run average cost. Section 6 concludes the paper and contains a discussion on the possible directions of future research.

2 MODEL DESCRIPTION

We assume the system deteriorates with time and is subject to random failure. Upon failure, the system is instantaneously replaced by a new one and the process renews. The hazard function of the system increases with the system's age as well as with the values of covariates that reflect the operating condition of the current system.

For simplicity, we consider only one covariate. To account for both the age effect and the condition information in the system's hazard function, the PHM is employed to describe the failure process of the system. That is, the hazard function of the system at time t can be expressed as

$$h(t, Z(t)) = h_0(t)\psi(Z(t)), t \geq 0 \qquad (1)$$

where $h_0(t)$ is a baseline hazard function dependent only on the age of the system and $\psi(\cdot)$ is a link function dependent only on the value of the covariate. We assume $Z = \{Z(t), t \geq 0\}$ is a continuous-time semi-Markov process with a finite state space $S = \{0,1,...,n-1\}$, which depicts the evolution of this covariate. Note that, due to the assumption of instantaneous replacement, no failure state is defined. It follows that the conditional survivor function is given by

$$S(t;Z) = P(T > t \mid Z_s, 0 \leq s \leq t) \equiv \exp\left(-\int_0^t h_0(s)\psi(Z_s)ds\right), t \geq 0\,. \qquad (2)$$

In the PHM, the failure of the system can happen in any state at any time.

For the process *Z*, state 0 represents the condition of a new system and states $1, 2, ..., n-1$ reflect the increasingly deteriorating working condition of the system. The Z process evolves as a pure birth process; i.e., whenever a transition occurs, the state of the system always increases by one. State $n-1$ is absorbing. Replacement is instantaneous and the system returns to state 0 upon replacement. Let X_k be the sojourn time of the *Z* process in state k. For $k \leq n-2$, X_k follows an arbitrary distribution since *Z* is semi-Markovian, while X_{n-1} is infinite. For notational convenience, define $S_k = \sum_{i=0}^{k} X_i, k = 0,1,..,n-2$, which is the age at which the system moves from state k to state $k+1$. Let the joint pdf and cdf of $S_0, S_1, ..., S_k$, for $k = 0,...,n-2$, be

$$g_k(s_0, s_1, ..., s_k) \text{ and } G_k(s_0, s_1, ..., s_k)$$

respectively, for $0 < s_0 < s_1 < ... < s_k$.

We assume the Z process is under continuous monitoring, which means that its state is known at every point of time. Continuous monitoring usually involves an upfront investment in hardware and software installation, and each inspection action costs nothing thereafter. Because this upfront cost does not affect the optimal policy that minimizes the long-run average cost, we do not include the cost of continuous monitoring in our objective function.

Utilizing the classical cost structure, assume each planned replacement costs $C > 0$ and each failure replacement incurs an additional cost $K > 0$. Let T be the time to failure of the system and T_d be a stopping time based on the continuous observation of Z_t and the system's age. Define the replacement rule δ_{Td} : Replace at failure or at T_d, whichever occurs first. Then according to the theory of renewal reward processes [19], the long run average cost per unit time can be expressed as

$$\phi(T_d) = \frac{C + KP(T_d \geq T)}{E\left[\min\{T, T_d\}\right]} \tag{3}$$

where $P(T_d \geq T)$ is the probability of failure replacement and $E\left[\min\{T, T_d\}\right]$ is the expected replacement time. The objective of this paper is to find an optimal replacement policy that minimizes the long-run average cost per unit time for systems with semi-Markovian covariate process and continuous inspection.

To summarize, we adopt the following notation in this paper:

t : The age of the current system.

$Z = \{Z_t, t \geq 0\}$: A right continuous semi-Markov process that reflects the condition of the system at age t with $Z_0 = 0$; in general, the effect of the operating environment on the system.

$h_0(t)$: The baseline hazard rate, which depends only on the age of the system.

$\psi(Z_t)$: A link function that depends on the state of the stochastic process Z .

X_k : The sojourn time of the Z process in state k .

S_k : The age at which the system state changes from k to $k+1$, $k=0,1,...,n-2$.

$g_k(s_0, s_1, ..., s_k)$: The joint pdf of $S_0, S_1, ..., S_k$, $k=0,...,n-2$.

$G_k(s_0, s_1, ..., s_k)$: The joint cdf of $S_0, S_1, ..., S_k$, $k=0,...,n-2$.

T : The time to failure of the system.

T_d : A stopping time dependent on the age of the system and Z_t .

δ_d : A replacement policy that replaces at failure or at T_d , whichever occurs first.

C : The replacement cost without failure, $C>0$.

K : The additional cost for a failure replacement, $K>0$.

In addition, we state the following basic assumptions:

1. The system must be kept in working order at all times. Replacement is instantaneous.
2. The baseline hazard rate, $h_0(t)$, is a non-decreasing function of the system age; that is, the system deteriorates with time.
3. The link function, $\psi(Z_t)$, is a non-decreasing function with $\psi(0)=0$.
4. The practice of continuous monitoring influences neither the covariate process Z nor the system failure process.

3 THE FORM OF THE OPTIMAL REPLACEMENT POLICIES

Bergman [18] investigated the optimal replacement problem under a very general failure model, in which the hazard rate $h(\bullet)$ of system failure time is non-decreasing and completely determined by a general stochastic process $\{X(t), t \geq 0\}$. The process $X(t)$ could be a stochastic vector process such that each of its components is non-decreasing. It is assume that $X(t)$ is under continuous monitoring. Under the classical cost structure described above, Bergman showed that the optimal replacement policy is of the control limit type and the optimal stopping time has the following form

$$T_d^* = \inf\{t \geq 0 : Kh\left(X(t)\right) \geq d^*\} \tag{4}$$

where $d^* = \phi(T_d^*)$ is the optimal cost. If the set in (4) is empty, then $T_d^* = \infty$, which means replacement only at failure.

In addition, Bergman proved the following proposition, which leads to an iterative algorithm that produces a sequence converging to optimal cost.

Proposition 1: *Choose any positive* d_0 *and set iteratively*

$$T_n = \inf\{t \geq 0, h\left(X(t)\right) \geq d_n / K\} \tag{5}$$

$$d_{n+1} = \phi(T_n), \quad n = 0, 1, 2, \cdots. \tag{6}$$

Then $\lim_{n\to\infty} d_n = d^*$.

The PHM with semi-Markovian covariate process and continuous monitoring presented in Section 2 is a special case of the general failure model defined by Bergman, where the age of the system could be regarded as one component of the stochastic process $X(t)$ and the covariate process $Z(t)$ as another component of $X(t)$. Thus it follows that the optimal stopping time is

$$T_d^* = \inf\{t \geq 0, h\left(t, Z(t)\right) \geq d^* / K\} \tag{7}$$

where $h(t, Z(t)) = h_0(t)\psi(Z(t))$ as in (1).

The optimal replacement policy defined by (7) may be explained as: Replace at failure or when the hazard rate of the system reaches or exceeds a certain level. Essentially, this is a control-limit policy with respect to the hazard rate. In our model, if we know the form of the baseline hazard function and the link function, then for a certain state, equation (7) determines a unique threshold time for replacement in each state because the hazard rate function is monotonic in time. Hence the optimal replacement policy is uniquely defined by *n* threshold ages for replacement. For example, consider a system with a three-state *Z* process. As illustrated in Figure 1, the control limit d^*/K for hazard rate fixes the planned replacement ages t_0, t_1, t_2 for state 0, 1, 2 respectively, and since the link function increases with system's state, we have $t_0 > t_1 > t_2$.

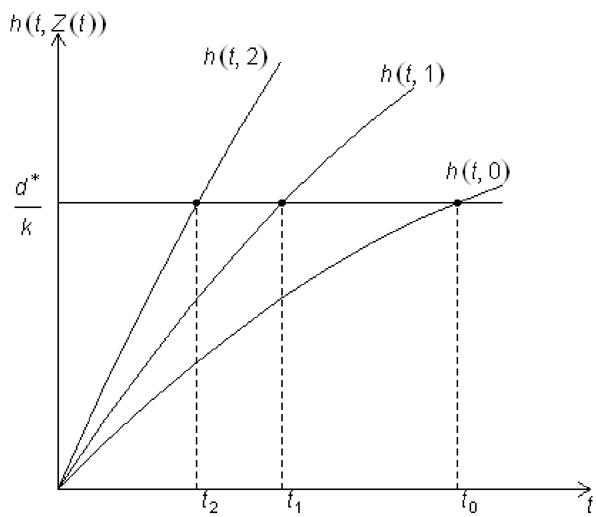

Figure 1

We henceforth restrict our attention to a special class of replacement policies, in which a policy is composed of *n* threshold times for replacement and we denote it as $\delta_d = \{t_0, t_1, ..., t_{n-1}\}$, $t_0 > t_1 > ... > t_{n-1}$, where t_i is the threshold age for replacement if system is in state *i*. Clearly the optimal policy in (7) falls within this class. Our objective is to compute the optimal threshold values.

With the form of the optimal policy known and the iterative algorithm, there is still one barrier in the way to obtain the optimal policy and cost for our model, which is the evaluation of equation (6); that is, how to compute the corresponding cost for a given stopping rule. From equation (3), this involves finding the failure probability $P(T_d \geq T)$ and expected time to replacement $E\left[\min\{T,T_d\}\right]$ for a given policy δ_d. An explicit expression for our objective function in terms of the policy parameters $t_0,t_1,...,t_{n-1}$ is necessary. We address this issue in Section 4.

4 EXPLICIT EXPRESSION OF THE LONG-RUN AVERAGE COST

For notational convenience, define $W_d = W(t_0,t_1,...,t_{n-1}) = E\left[\min\{T,T_d\}\right]$ as the expected life of the system and define $Q_d = Q(t_0,t_1,...,t_{n-1}) = P\left(T \leq T_d\right)$ as the probability of failure under policy $\delta_d = \{t_0,t_1,...,t_{n-1}\}$. In what follows, we show that it is possible to explicitly represent W_d, Q_d as functions of $t_0,t_1,.,t_{n-1}$ by conditioning on the time instants that the system changes states; that is, on $S_0,S_1,...,S_{n-2}$. For simplicity, we take the system with a three state covariate process as an illustration. The situations with more states could be generalized accordingly.

In accordance with the survivor function in (2), define the conditional cdf of system failure time T as follows by conditioning on S_0 and S_1, where s_0 and s_1 are realizations of S_0 and S_1 respectively, and $s_0 < s_1$. Let $F(t;s_0,s_1) \equiv P(T \leq t \mid S_0 = s_0, S_1 = s_1)$.

Then for $t \leq s_0$,

$$F(t;s_0,s_1) = F_0(t) \equiv 1 - \exp\left(-\psi\left(0\right)\int_0^t h_0\left(u\right)du\right).$$

For $s_1 > t > s_0$

$$F(t;s_0,s_1) = F_1(t;s_0) \equiv 1 - \exp\left(-\psi\left(0\right)\int_0^{s_0} h_0\left(u\right)du - \psi\left(1\right)\int_{s_0}^{t} h_0\left(u\right)du\right).$$

For $t > s_1$

$$F(t;s_0,s_1)=F_2(t;s_0,s_1)\equiv 1-\exp\left(-\psi(0)\int_0^{s_0}h_0(u)\,du-\psi(1)\int_{s_0}^{s_1}h_0(u)\,du-\psi(2)\int_{s_1}^{t}h_0(u)\,du\right)$$

.

Again conditioning on S_0 and S_1, there will be five different cases based on the relative positions among t_2,t_1,t_0 and s_0,s_1, as shown below. Note $t_2<t_1<t_0$ and $s_0<s_1$. Under each case, the expressions of W_d and Q_d can be derived easily.

Let

$$W(t_0,t_1,t_2;s_0,s_1)\equiv E\left(\min\{T,T_d\}\mid S_0=s_0,S_1=s_1\right)$$

$$Q(t_0,t_1,t_2;s_0,s_1)\equiv P(T\le T_d\mid S_0=s_0,S_1=s_1).$$

By the Law of Iterated Expectation [20],

$$W(t_0,t_1,t_2;s_0,s_1)=E\left(E\left(\min\{T,T_d\}\mid S_0,S_1,T\right)\mid S_0=s_0,S_1=s_1\right).$$

Case 0: If $s_0>t_0$, then

$$\min\{T,T_d\}=\begin{cases}T & \text{if } T\le t_0\\ t_0 & \text{if } T>t_0\end{cases}$$

$$W(t_0,t_1,t_2;s_0,s_1)=W_0(t_0)\equiv\int_0^{t_0}t\,dF_0(t)+t_0\left[1-F_0(t_0)\right]$$

$$Q(t_0,t_1,t_2;s_0,s_1)=Q_0(t_0)\equiv F_0(t_0)$$

Case 1: If $t_1<s_0<t_0$, then

$$\min\{T,T_d\}=\begin{cases}T & \text{if } T\le s_0\\ s_0 & \text{if } T>s_0\end{cases}$$

$$W(t_0,t_1,t_2;s_0,s_1)=W_1(s_0)\equiv\int_0^{s_0}t\,dF_0(t)+s_0\left[1-F_0(s_0)\right]$$

$$Q(t_0,t_1,t_2;s_0,s_1)=Q_1(s_0)\equiv F_0(s_0)$$

Case 2: If $s_0<t_1$, $s_1>t_1$, then

$$\min\{T,T_d\}=\begin{cases}T & \text{if } T\le t_1\\ t_1 & \text{if } T>t_1\end{cases}$$

$$W(t_0,t_1,t_2;s_0,s_1) = W_2(s_0,t_1) \equiv \int_0^{s_0} tdF_0(t) + \int_{s_0}^{t_1} tdF_1(s_0,t) + t_1\left[1 - F_1(s_0,t_1)\right]$$

$$Q(t_0,t_1,t_2;s_0,s_1) = Q_2(s_0,t_1) \equiv F_1(s_0,t_1)$$

Case 3: If $s_0 < t_1$, $t_2 < s_1 < t_1$, then

$$\min\{T,T_d\} = \begin{cases} T & \text{if } T \le s_1 \\ s_1 & \text{if } T > s_1 \end{cases}$$

$$W(t_0,t_1,t_2;s_0,s_1) = W_3(s_0,s_1) \equiv \int_0^{s_0} tdF_0(t) + \int_{s_0}^{s_1} tdF_1(s_0,t) + s_1\left[1 - F_1(s_0,s_1)\right]$$

$$Q(t_0,t_1,t_2;s_0,s_1) = Q_3(s_0,s_1) \equiv F_1(s_0,s_1)$$

Case 4: If $s_0 < t_1$, $s_1 < t_2$, then

$$\min\{T,T_d\} = \begin{cases} T & \text{if } T \le t_2 \\ t_2 & \text{if } T > t_2 \end{cases}$$

$$W(t_0,t_1,t_2;s_0,s_1) = W_4(s_0,s_1,t_2) \equiv \int_0^{s_0} tdF_0(t) + \int_{s_0}^{s_1} tdF_1(s_0,t) + \int_{s_1}^{t_2} tdF_2(s_0,s_1,t) + t_2\left[1 - F_2(s_0,s_1,t_2)\right]$$

$$Q(t_0,t_1,t_2;s_0,s_1) = Q_4(s_0,s_1,t_2) \equiv F_2(s_0,s_1,t_2)\,.$$

With the above five cases at hand, by another application of the Law of Iterated Expectation,

$$\begin{aligned} W_d &= E\left[E\left(\min\{T,T_d\} \mid S_0,S_1\right)\right] = \int_{t_0}^{\infty} W_0(t_0)g_0(s_0)ds_0 + \int_{t_1}^{t_0} W_1(s_0)g_0(s_0)ds_0 \\ &+ \int_{t_1}^{\infty}\int_0^{t_1} W_2(s_0,t_1)g_1(s_0,s_1)ds_0ds_1 + \int_{t_2}^{t_1}\int_0^{s_1} W_3(s_0,s_1)g_1(s_0,s_1)ds_0ds_1 + \int_0^{t_2}\int_0^{s_1} W_4(s_0,s_1,t_2)g_1(s_0,s_1)ds_0ds_1 \end{aligned} \tag{8}$$

$$\begin{aligned} Q_d &= E\left[P(T \le T_d \mid S_0,S_1)\right] = \int_{t_0}^{\infty} Q_0(t_0)g_0(s_0)ds_0 + \int_{t_1}^{t_0} Q_1(s_0)g_0(s_0)ds_0 \\ &+ \int_{t_1}^{\infty}\int_0^{t_1} Q_2(s_0,t_1)g_1(s_0,s_1)ds_0ds_1 + \int_{t_2}^{t_1}\int_0^{s_1} Q_3(s_0,s_1)g_1(s_0,s_1)ds_0ds_1 + \int_0^{t_2}\int_0^{s_1} Q_4(s_0,s_1,t_2)g_1(s_0,s_1)ds_0ds_1 \end{aligned} \tag{9}$$

Assume the marginal pdfs of sojourn times X_0, X_1 are $f_{X_0}(\bullet)$ and $f_{X_1}(\bullet)$ respectively. It follows that the pdf of S_0 is

$$g_0(s_0) = f_{X_0}(s_0)\,. \tag{10}$$

Also, note that the event $S_0 = s_0, S_1 = s_1$ is equivalent to the event $X_0 = s_0, X_1 = s_1 - s_0$. Hence the joint pdf of S_0, S_1 is

$$g_1(s_0, s_1) = f_{X_0}(s_0) f_{X_1}(s_1 - s_0), \quad 0 < s_0 < s_1 . \tag{11}$$

So far, we have obtained the integral expressions of W_d and Q_d in terms of the policy parameters t_0, t_1, t_2 for three-state system. For an *n*-state system, there are $2n-1$ different cases. Thus the expression for W_d consists of $2n-1$ terms, each of which is an *n*-fold integral. The expression of Q_d is similar. To explicitly write out the $(2n-1)$ *n*-fold integrals seems to be a formidable task. However, thanks to the connection between *n* state situation and $(n+1)$ state situation, this task is reduced to something tractable. In fact, for the $(n+1)$ state situation, the expression of W_d has $2n+1$ cases, the first $2n-2$ cases of which are exactly the same as those of the W_d expression for the *n* state situation and the last three cases of which are divisions of the last case of *n* state situation by the new transition instant, S_n. Therefore, we can build on the expressions of W_d and Q_d for *n*-state system by adding one state at a time. To illustrate this process, we show the formulation of two-state system in APPENDIX I.

Based on the explicit expressions of W_d and Q_d we just derived and Proposition 1, we describe the following iterative algorithm, which may be employed to find the optimal policy parameters and the optimal cost simultaneously.

Algorithm I

1. Initialize the iteration counter $m = 0$. Choose an arbitrary replacement policy and let d_0 equal the cost of the chosen policy.
2. For d_m, use (5) to find the threshold time t_i^m for replacement if the system state is in state i, i.e.,

$$t_i^m = \inf\left\{t \geq 0 : h_0(t)\psi(i) = d_m / K\right\}, i \in S . \tag{12}$$

3. Use the replacement policy $\delta_m = \{t_0^m, t_1^m, .., t_{n-1}^m\}$ obtained in step 2 and equations (3), (8) and (9) to update $d_{m+1} = \phi(\delta_m)$.

4. If $d_{m+1} = d_m$, stop with $d^* = d_{m+1}$ and $\delta^* = \{t_0^*, t_1^*, ..., t_{n-1}^*\} = \{t_0^m, t_1^m, ..., t_{n-1}^m\}$; otherwise, set $m \leftarrow m+1$ and go to step 2.

5 NUMERICAL EXAMPLE AND SENSITIVITY ANALYSIS

5.1 Numerical Example

To illustrate our model and the procedure to construct the optimal policy for a three-state deteriorating system, we consider the following numerical example. Assume the baseline hazard function is a Weibull hazard function given by

$$h_0(t) = \frac{bt^{b-1}}{a^b}$$

with $a = 1$ and $b = 2$ and suppose that $\psi(Z_t) = \exp(cZ_t)$ with $c = 2$. Assume $C = 5$ and $K = 25$. Since the forms of $h_0(t)$ and $\psi(Z_t)$ are predefined, the PHM here is parametric rather than semi-parametric as described in [11].

Suppose the semi-Markov process Z has three states $\{0,1,2\}$ and the sojourn times X_0 and X_1 are independent identically distributed Weibull random variables with mean 1. The Weibull distribution is chosen here as it includes the exponential distribution as a special case, which allows convenient comparisons between systems with Markovian and semi-Markovian covariate processes. Assume the pdf of X_i is

$$f_{X_i}(x_i) = \frac{\beta x_i^{\beta-1}}{\eta^\beta} \exp\left[-\left(\frac{x_i}{\eta} \right)^\beta \right], \quad 0 < x_i, \; i = 0,1.$$

Assume $\beta = 1.5$ and $\eta = 1.1077$ so that the mean of X_i equals 1.

As in Algorithm I, we initialize $d_0 = (C+K)/E(T)$, which is the cost of the policy that replaces only at failure. The mean time to failure $E(T)$ could be obtained from equation (8) by setting $t_0 = t_1 = t_2 = \infty$. In this way we find $E(T) = 0.6813$ and $d_0 = 44.0335$.

The complete results are shown in Table 1. The iterative algorithm converges after five iterations to the optimal average cost $d^* = 23.4364$. The algorithm was implemented in *Mathematica*® for precise and efficient numerical evaluation of multiple integrals.

Table 1 Illustration of the Computation Procedure with Weibull(1.2089, 1.5) Sojourn Time

m	d_m	t_0^m	t_1^m	t_2^m	$W(t_0^m, t_1^m, t_2^m)$	$Q(t_0^m, t_1^m, t_2^m)$	$\phi(t_0^m, t_1^m, t_2^m)$
0	44.0335	0.8807	0.1192	0.016	0.5618	0.3846	26.0157
1	26.0157	0.5203	0.0704	0.0095	0.4248	0.1998	23.5262
2	23.5262	0.4705	0.0637	0.0086	0.3958	0.1710	23.4365
3	23.4365	0.4687	0.0634	0.0086	0.3947	0.1700	23.4364
4	23.4364	0.4687	0.0634	0.0086	0.3947	0.1700	23.4364

To study the effect of the parameters of the Weibull sojourn time, we varied the shape parameter β from 0.8 to 2, and changed the scale parameter η accordingly to ensure the same mean sojourn time. Table 2 shows the optimal replacement policies and costs for various Weibull sojourn time distributions (STD). We also include coefficients of variation (CV) of the distributions in Table 2 to gain more insight. One interesting observation is that the optimal cost increases with the CV of the STD, which is reasonable because in practice larger variability always tends to boost the cost.

Another notable observation is that different STDs lead to different optimal policies and costs, even if they all follow Weibull and have the same mean. This observation implies a pitfall if we always model the covariate process as Markovian. Suppose the true STD is

Weibull(1.1077, 1.5). If we use the Markov model, then the best estimated STD is Weibull(1, 1), i.e., Exp(1), which would lead to a non-optimal replacement policy and boost the replacement cost. The cost errors for using policy parameters from the Markov model in other sojourn times are shown in Table 3. We can see that the error becomes smaller as the CV of the true STD gets closer to 1. In this example, those errors are relatively small, which means when the STD of the covariate process is unknown and hard to estimate, a Markov process might be a good candidate, and the investment for a good estimation of the STD would be only of marginal value. Besides, the Markov model could simplify the computation for the optimal policy because exponential STD would simplify the evaluation of the multiple integrals.

Table 2 Effect of Different Weibull Parameters on the Optimal Policy and Cost

Sojourn Time Distribution	Coefficient of Variation	t_0	t_1	t_2	$W(t_0,t_1,t_2)$	$Q(t_0,t_1,t_2)$	d^*
Weibull(0.7900, 0.7)	1.4624	0.5293	0.0716	0.0097	0.3281	0.1473	26.4652
Weibull(0.8826, 0.8)	1.2605	0.5125	0.0694	0.0094	0.3428	0.1514	25.6249
Weibull(1, 1)	1	0.4913	0.0665	0.0090	0.3646	0.1582	24.5645
Weibull(1.1077, 1.5)	0.6790	0.4687	0.0634	0.0086	0.3947	0.1700	23.4364
Weibull(1.1284, 2)	0.5227	0.4609	0.0624	0.0084	0.4088	0.1769	23.0469

Table 3 Cost errors for using policy parameters from Markov model

Sojourn Time Distribution	Coefficient of Variation	Absolute Error	Relative Error
Weibull(0.7900, 0.7)	1.4624	0.0453	0.171%
Weibull(0.8826, 0.8)	1.2605	0.0144	0.056%
Weibull(1.1077, 1.5)	0.6790	0.0185	0.079%
Weibull(1.1284, 2)	0.5227	0.0355	0.154%

Table 4 shows the optimal policies and costs for Lognormal STDs. Again all these distributions have the same mean 1. Table 4 confirms the conclusion that large CV has a negative effect on the optimal cost. Besides, comparing similar cases in Table 2 and Table 4 suggests that when having the same CV, the Lognormal sojourn time leads to a lower optimal cost than the Weibull sojourn time.

Table 4 Optimal Policy and Cost when Sojourn time is Lognormal

Sojourn time Distribution	Coefficient of Variation	t_0	t_1	t_2	$W(t_0,t_1,t_2)$	$Q(t_0,t_1,t_2)$	d^*
Lognor(-0.5, 1)	1.3108	0.4805	0.0650	0.0088	0.3691	0.1548	24.0264
Lognor(-0.3469, 0.833)	1	0.4680	0.0633	0.0086	0.3893	0.1645	23.4036
Lognor(-0.1922, 0.62)	0.6846	0.4585	0.0621	0.0084	0.4108	0.1770	22.9264
Lognor(-0.125, 0.5)	0.5329	0.4560	0.0617	0.0084	0.4192	0.1823	22.7990

5.2 Sensitivity Analysis

In the above numerical example, we assume all the model parameters are known. However, in practice, some of those parameters must be estimated from the lifetime data of the system. And the quality of the estimates will directly affect the validity of the replacement policy obtained. In this subsection, we investigate how the variations in the model parameters impact the long-run average cost and assess the relative importance of model parameters through sensitivity analysis. In particular, we evaluate three input parameters, which are *a*, *b* in the baseline hazard function, $h_0(t) = \frac{bt^{b-1}}{a^b}$, and *c* in the link function, $\psi(Z_t) = \exp(cZ_t)$. (For simplicity, we assume the forms of $h_0(t)$ and $\psi(Z_t)$ are known and all the other parameters are given and the same as in Subsection 5.1.) We choose Weibull(1.1077, 1.5) as the STD for the *Z* process.

Assume the true parameters values are $a = 2, b = 2, c = 2$ and assume their estimates follow the distributions below

$$\hat{a} \sim N(2, 0.4), \ \hat{b} \sim N(2, 0.4), \ \hat{c} \sim N(2, 0.4).$$

Performing the FAST sensitivity analysis method [21] with 1000 samples using software SimLab [22], we get the FAST first-order indexes as shown in Table 5. This index gives the expected reduction in the variance of the cost if an individual parameter is fixed. This table indicates that the scale parameter of the baseline hazard function, *a*, accounts for most of the variability in the output and therefore is the most important of the three parameters. It implies that if we can somehow reduce the variances of some input parameters' estimates by investing more, we should give parameter *a* the highest priority.

Notably, the conclusions reached by sensitivity analysis are case-specific and should not be generalized if the model parameters are changed.

Table 5 FAST first-order indexes

Parameters	First-order indexes on cost
a	0.3329
b	0.1069
c	0.0383

6 CONCLUSION

In this paper, we studied the optimal replacement problem for general deteriorating systems in the proportional hazards model with a semi-Markovian covariate process, which we assume is under continuous monitoring. Allowing the covariate process to be semi-Markovian endows our method with great capability and flexibility to model real world situations. To minimize the long-run average cost per unit time, first we identified our model as a special case of Bergman's model [18] so that the optimal replacement policy of our model is of the control limit type with respect to the hazard function. Given that an optimal policy may be uniquely defined by a set of state-dependent threshold ages for replacement, an explicit expression for the objective function was derived in terms of those ages by conditioning. Then the iterative procedure developed by Bergman was customized for our model to find the optimal threshold ages.

A numerical example with n=3 illustrates the computational procedure as well as the effects of different sojourn time distributions of the covariate process on the optimal policy and cost. The results show that larger variability in the sojourn time distributions (STD) tends to boost the replacement cost. However, some numerical results show that when the STD of the covariate process is difficult to estimate, viewing the process as a Markov process is not a bad option. Sensitivity analysis on an instance demonstrates that the variance of the scale parameter in the baseline hazard function accounts for most of the resulting variability in the cost and therefore the scale parameter is of the most importance among the three chosen parameters.

Possible extensions of the research could be, to 1) generalize the one-dimensional covariate to a multi-dimensional vector which would permit the Z process to evolve along multiple paths; 2) introduce the uncertainty of the monitoring process, that is, the partial observation problem to our current model; and 3) use a new model to relate the covariate information to system failure time distribution, such as a scale-accelerated failure time (SAFT) model [23].

APPENDIX I FORMULATION FOR SYSTEM WITH A TWO-STATE COVARIATE PROCESS

For the system with a two-state covariate process, there will be only one time instant, S_0, at which the system changes states. In the following, we show how to explicitly represent the expected life of the system $W_d = W(t_0, t_1) = E\left[\min\{T, T_d\}\right]$ and the probability of failure $Q_d = Q(t_0, t_1) = P\left(T \le T_d\right)$ under policy $\delta_d = \{t_0, t_1\}$ by conditioning on S_0.

Define the conditional cdf of system failure time T as follows.

$$F(t; s_0) \equiv P(T \le t \mid S_0 = s_0),$$

where s_0 is the realization of S_0.

Then for $t \le s_0$

$$F(t; s_0) = F_0(t) \equiv 1 - \exp\left(-\psi\left(0\right)\int_0^t h_0\left(u\right)du\right).$$

For $t > s_0$

$$F(t; s_0) = F_1(t; s_0) \equiv 1 - \exp\left(-\psi\left(0\right)\int_0^{s_0} h_0\left(u\right)du - \psi\left(1\right)\int_{s_0}^t h_0\left(u\right)du\right).$$

Let

$$W(t_0,t_1;s_0) \equiv E\left(\min\{T,T_d\} \mid S_0 = s_0\right)$$

$$Q(t_0,t_1;s_0) \equiv P(T \le T_d \mid S_0 = s_0)$$

By the Law of Iterated Expectation [20],

$$W(t_0,t_1;s_0) = E\left(E\left(\min\{T,T_d\} \mid S_0,T\right) \mid S_0 = s_0\right).$$

There will be three cases.

Case 0: If $s_0 > t_0$, then

$$\min\{T,T_d\} = \begin{cases} T & \text{if } T \le t_0 \\ t_0 & \text{if } T > t_0 \end{cases}$$

$$W(t_0,t_1;s_0) = W_0(t_0) \equiv \int_0^{t_0} t dF_0(t) + t_0\left[1 - F_0(t_0)\right]$$

$$Q(t_0,t_1;s_0) = Q_0(t_0) \equiv F_0(t_0)$$

Case 1: If $t_1 < s_0 < t_0$, then

$$\min\{T,T_d\} = \begin{cases} T & \text{if } T \le s_0 \\ s_0 & \text{if } T > s_0 \end{cases}$$

$$W(t_0,t_1;s_0) = W_1(s_0) \equiv \int_0^{s_0} t dF_0(t) + s_0\left[1 - F_0(s_0)\right]$$

$$Q(t_0,t_1;s_0) = Q_1(s_0) \equiv F_0(s_0)$$

Case 2: If $s_0 < t_1$, then

$$\min\{T,T_d\} = \begin{cases} T & \text{if } T \le t_1 \\ t_1 & \text{if } T > t_1 \end{cases}$$

$$W(t_0,t_1;s_0) = W_2(s_0,t_1) \equiv \int_0^{s_0} t dF_0(t) + \int_{s_0}^{t_1} t dF_1(s_0,t) + t_1\left[1 - F_1(s_0,t_1)\right]$$

$$Q(t_0,t_1;s_0) = Q_2(s_0,t_1) \equiv F_1(s_0,t_1)$$

Then by another application of the Law of Iterated Expectation,

$$W_d = E\left[E\left(\min\{T,T_d\} \mid S_0\right)\right] = \int_{t_0}^{\infty} W_0(t_0) g_0(s_0) ds_0 + \int_{t_1}^{t_0} W_1(s_0) g_0(s_0) ds_0 + \int_0^{t_1} W_2(s_0,t_1) g(s_0) ds_0 ,$$

$$Q_d = E\left[P(T \le T_d \mid S_0)\right] = \int_{t_0}^{\infty} Q_0(t_0) g_0(s_0) ds_0 + \int_{t_1}^{t_0} Q_1(s_0) g_0(s_0) ds_0 + \int_0^{t_1} Q_2(s_0,t_1) g_0(s_0) ds_0 .$$

Comparison with equations (8) and (9) shows the recursive nature of these expressions.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation under grant CNS-0540293.

REFERENCES

[1] M. Wang, A. J. Vandermaar, and K. D. Srivastava, “Review of condition assessment of power transformers in service,” *IEEE Electrical Insulation Magazine,* vol. 18, no. 6, 2002.

[2] W. Q. Meeker, "Trends in the statistical assessment of reliability," Technical Report, Iowa State University, 2009.

[3] H. Mine, and H. Kawai, “An optimal inspection and replacement policy,” *IEEE Transactions on Reliability,* vol. 24, pp. 305-309, 1975.

[4] H. Mine, and H. Kawai, “An optimal inspection and replacement policy of a deteriorating system,” *Journal of Operations Research Society of Japan,* vol. 25, pp. 1-15, 1982.

[5] C. T. Lam, and R. H. Yeh, “Optimal replacement policies for multi-state deteriorating systems,” *Naval Research Logistics,* vol. 41, no. 33, pp. 303-315, 1994.

[6] D. Chen, and K. S. Trivedi, “Optimization for condition-based maintenance with semi-Markov decision process,” *Reliability Engineering & System Safety,* vol. 90, no. 1, pp. 25-29, 2005.

[7] A. Grall, L. Dieulle, C. Berenguer *et al.*, “Continuous-time predictive-maintenance scheduling for a deteriorating system,” *IEEE Transactions on Reliability,* vol. 51, no. 2, 2002.

[8] A. Barros, C. Berenguer, and A. Grall, “Optimization of replacement times using imperfect monitoring information,” *IEEE Transactions on Reliability,* vol. 52, no. 4, 2003.

[9] J. S. Ivy, and S. M. Pollock, “Marginally monotonic maintenance policies for a multi-state deteriorating machine with probabilistic monitoring, and silent failures,” *IEEE Transactions on Reliability,* vol. 54, no. 3, 2005.

[10] R. Toscano, and P. Lyonnet, “On-line reliability prediction via dynamic failure rate model,” *IEEE Transactions on Reliability,* vol. 57, no. 3, 2008.

[11] D. R. Cox, and D. Oakes, *Analysis of Survival Data,* London: Chapman & Hall, 1984.

[12] A. K. S. Jardine, and A. H. C. Tsang, *Maintenance, Replacement & Reliability: Theory and Applications*: Taylor & Francis Group, 2006.

[13] S. Gasmi, C. E. Love, and W. Kahle, "A general repair, proportional-hazards, framework to model complex repairable systems," *IEEE Transactions on Reliability,* vol. 52, no. 1, 2003.

[14] V. Makis, and A. K. S. Jardine, "Optimal replacement in the proportional hazards model," *INFOR,* vol. 30, no. 1, pp. 172-183, 1992.

[15] D. Banjevic, A. K. S. Jardine, V. Makis *et al.*, "A control-limit policy and software for condition-based maintenance optimization," *INFOR,* vol. 39, pp. 32-50, 2001.

[16] S. Ghasemi, S. Yacout, and M. S. Ouali, "Optimal condition based maintenance with imperfect information and the proportional hazards model," *International Journal of Production Research,* vol. 45, no. 4, pp. 989-1012, 2007.

[17] X. Wu, and S. M. Ryan, "Value of condition monitoring for optimal replacement in the proportional hazards model with continuous time degradation," *forthcoming in IIE Transactions*.

[18] B. Bergman, "Optimal replacement under a general failure model," *Adv. in Appl. Prob.,* vol. 10, no. 2, pp. 431-541, 1978.

[19] S. M. Ross, *Introduction to Probability Models*, 8th ed., San Diego, CA: Academic Press, 2003.

[20] P. Billingsley, *Probability and Measure*, New York, NY: John Wiley & Sons, Inc., 1995.

[21] A. Saltelli, S. Tarantola, F. Campolongo *et al.*, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, New York: John Wiley & Sons, 2004.

[22] European Commission. "SimLab - Sensitivity Analysis," December 3, 2009; http://simlab.jrc.ec.europa.eu/licenses/index.htm.

[23] W. Q. Meeker, and L. A. Escobar, *Statistical Methods for Reliability Data*, New York: Wiley, 1998.

Xiang Wu, Ph.D. Candidate
Industrial & Manufacturing Systems Engineering
3024 Black Engineering
Iowa State University
Ames, IA 50011-2164 USA
Phone: 515-509-1679
Email: xiangwu@iastate.edu

Sarah M. Ryan, Professor
Industrial & Manufacturing Systems Engineering
3004 Black Engineering Building
Iowa State University
Ames, IA 50011-2164 USA
Phone: 515-294-4347
Email: smryan@iastate.edu

Biographies

Xiang Wu is a Ph.D. candidate in the Department of Industrial and Manufacturing Systems Engineering at Iowa State University. He received his B.S. and M.S. degrees in Automation from Tsinghua University, Beijing, P. R. China. His research interests include stochastic modeling and optimization, reliability and maintenance, and production scheduling.

Sarah M. Ryan, a member of IEEE and senior member of IIE, is Professor and Director of Graduate Education in the Department of Industrial and Manufacturing Systems Engineering at Iowa State University. She teaches courses in stochastic modeling and optimization and conducts research on electric power systems, capacity expansion, and closed loop supply chains. Her work has been published in *IIE Transactions*, *Operations Research*, *Management Science* and several other archival journals.